

Miguel Arias Estrada,  
Alexander Gelbukh  
(Eds.)

# **Avances en la Ciencia de la Computación**

2004

## **Prefacio**

Este libro contiene los artículos seleccionados de entre los trabajos presentados para los 10 talleres del Quinto Encuentro Internacional de Computación (ENC'04). En esta ocasión, para la edición 2004 del congreso, se partió de la experiencia de los talleres organizados el año pasado. Se hizo una especial invitación a algunos talleres a que repitieran la experiencia de organizar nuevamente un foro, y se invitó a algunos nuevos talleres en áreas que sabíamos hay gente trabajando pero que no existen aun una comunidad formada que pudiera compartir experiencias. Adicionalmente se convocó a los miembros de la Sociedad Mexicana de Ciencias de la Computación (SMCC) a que se propusieran nuevos talleres en áreas de interés y que pudieran haber pasado desapercibidas para nosotros.

El propósito de los talleres es servir como catalizador para la creación de foros especializados y unificados, y que en un futuro cercano permitan convertirse en comunidades especializadas con intereses de investigación comunes.

La experiencia de los talleres durante el ENC'03 mostró que existen comunidades en nuestro país que aun no se encuentran formadas pero que sobrepasan la masa crítica para mantener un foro de discusión interesante. En esta ocasión, la respuesta fue muy interesante: de seis talleres que hubo en 2003, ahora tendremos 10 talleres simultáneos.

De manera global, dentro de los 10 talleres se recibió un total de 121 artículos, de los cuales, después de un cuidadoso proceso de revisión y arbitraje, se seleccionaron 88 trabajos que se incluyen en estas memorias. El número de talleres demuestra la cantidad de grupos activos en el país que empiezan a agruparse para discutir los avances en cada uno de sus campos. En particular creemos que los talleres son un primer foro para identificar e integrar a las comunidades científicas de distintas áreas de la Computación con presencia en México y que sirva para consolidar dichas comunidades. Adicionalmente, se desea que los talleres atraigan a nuevos investigadores, investigadores jóvenes y estudiantes de posgrado a adentrarse en estas áreas de interés que, ciertamente, representan todas un gran potencial de investigación y desarrollo tecnológico en nuestro país. La evolución y crecimiento de los talleres permitirá catalizar la actividad científica de nuestro país, al mismo tiempo que crear oportunidades de difusión y colaboración con otros grupos internacionales.

Agosto 2004

Miguel Arias Estrada  
Alexander Gelbukh



## Contenido

---

### **Interacción humano-computadora: interacción – integración**

---

Introducción.....	3
Interfaz de Comunicación y Administración de Tareas (ICAT) para Grupos de Trabajo Utilizando Software Libre.....	5
<i>Ironelis Valdez Batista Juan Contreras-Castillo Gerardo Barreto Palacios</i>	
Integración de elementos de usabilidad en la ingeniería de requerimientos mediante Áncora.....	13
<i>Karen Cortés Verdín, Ángeles Sumano López, Dolores Vargas Cerdán</i>	
Desarrollo de un Marco de Trabajo para Juegos Matemáticos Multiusuarios en un Ambiente Distribuido.....	19
<i>Adriana Miramontes Rivera, Gabriel López-Morteo, Gilberto López</i>	
Diagnóstico Automático de Usabilidad de Sitios Web.....	28
<i>Fernando Sánchez Martínez, Oscar Mayora Ibarra</i>	
Estudio sobre el uso de gestos en una interfaz multimodal hombre-máquina simulada.....	37
<i>Luis Villaseñor Pineda &amp; Claudia Vázquez Valerdi</i>	
Diseño de Aplicaciones Interactivas utilizando el Paradigma de Patrones.....	47
<i>Jaime Muñoz Arteaga, Juan M. González Calleros y Jorge Aguilar Cisneros</i>	
Impacto de la usabilidad en la implantación de un ambiente de tesis digitales.....	56
<i>Lourdes Fernández Ramírez, J. Alfredo Sánchez, Alberto García</i>	
Evaluación de Maestros: Una Herramienta en Línea.....	62
<i>Gerardo Barreto, Ironelis Valdez Batista, Omar Álvarez, Sara Sandoval</i>	
Interfaz con reconocimiento de voz para la manipulación robótica.....	72
<i>Carlos Miranda-Palma, Oscar Mayora-Ibarra y Esmeralda Uraga-Serratos</i>	
Towards a Methodology to Envision and Evaluate Ubiquitous Computing.....	79
<i>Víctor M. González, Jesus Favela, and Marcela Rodríguez</i>	
Importancia de Modelos del Proceso Creativo en la IHM.....	86
<i>Rafael Pérez y Pérez, Fernando Gamboa</i>	
Diseño e Implementación de una Interfaz de Comunicación en Internet, Orientada a Personas con Discapacidad Visual.....	96
<i>Noemí Guerra, Román Gallardo, María Andrade, Ricardo Acosta</i>	
Hacia un Sistema de Diagnóstico y Compensación de Dislexia para Niños de Nivel Primaria.....	105
<i>Mario A Moreno Rocha, Rosario Peralta Calvo, Ricardo Ruiz Rodríguez</i>	

Creación de un diccionario electrónico de sexualidad.....	111
<i>Gerardo Sierra, Marlene Moreno, Eduardo Solares, Alfonso Medina</i>	

---

## **Interacciones en la realidad virtual y las simulaciones físicas**

---

Introduction .....	119
Enhanced Hair Paint .....	121
<i>Isaac Rudomín, Benjamín Hernández</i>	
Volume Preserved Deformation of Human Face with Hand .....	129
<i>Yohei Iwasaki, Isamu Mizuno, and Toyohisa Kaneko</i>	
Point Based Rendering and Displaced Subdivision for Interactive Animation of Crowds of Clothed Characters.....	139
<i>Isaac Rudomín and Erik Millán</i>	
Haptic Rendering using Velocity Driven Level of Detail.....	149
<i>Pavel Kolčárek and Jiří Sochor</i>	
ALICE: A Collaborative Virtual Reality Navigator .....	159
<i>Victor Theoktisto, Marta Fairén, and Isabel Navazo</i>	
Designing head tracking devices for your VR needs.....	170
<i>Isaac Rudomin, Marissa Diaz</i>	
Advanced Athroscopy Training Simulator insightMIST.....	178
<i>Jose M. Espadero, Sofia Bayona, Jose M. Fernández, Marcos García</i>	
Hardware-Accelerated Dynamic Volume Rendering for Real-Time Surgical Simulation .....	188
<i>Marco Agus, Andrea Giachetti, Enrico Gobbetti Gianluigi Zanetti, and Antonio Zorcolo</i>	

---

## **Tecnologías del lenguaje humano**

---

Introducción.....	199
Identificación de Antónimos en Textos Planos.....	203
<i>Cupertino Lucero, David Pinto y Héctor Jiménez</i>	
Some Tests in Text Categorization using Term Selection by DTP.....	212
<i>Edgar Moyotl, Héctor Jiménez</i>	
Herramientas de análisis para el Corpus Lingüístico en Ingeniería.....	219
<i>Gabriel Garduño, Gerardo Sierra, Alfonso Medina</i>	
Uso del Contexto para la Búsqueda de Respuestas en Español .....	227
<i>M. Pérez, M. Montes, A. López</i>	
Proyecto de preparación del corpus paralelo alineado español-inglés.....	235
<i>José Angel Vera, Grigori Sidorov</i>	
Uso de Lattices para la Recuperación de Términos.....	243
<i>Liset Fraguera, Héctor Jiménez, Gerardo Sierra</i>	

A Proposed Architecture for a Spoken Information Retrieval with Multimedia Databases.....	251
<i>Luis R. Salgado, Juan A. Nolasco, Pablo D. Díaz</i>	
Aplicando la Clasificación de Texto en la Extracción de Información.....	259
<i>Alberto Téllez, Manuel Montes, Luis Villaseñor</i>	
Comparación de dos métodos que determinan automáticamente el extracto de un texto .....	267
<i>Hilario Salazar, David Pinto &amp; Héctor Jiménez</i>	
Enhancing Flexibility of Named Entity Classifiers .....	274
<i>Thamar Solorio, Aurelio López</i>	
Estudio Comparativo de Traductores Automáticos para QA multilingüe .....	281
<i>Rita M. Aceves, Luis Villaseñor y Manuel Montes</i>	
Formalización del Sistema de Nombres Hispanos.....	289
<i>Noé Alejandro Castro, José Ángel Vera, Igor A. Bolshakov, Grigori Sidorov</i>	
Un Modelo de Representación basado en Sintagmas para Recuperación de Información .....	296
<i>Miguel Rodríguez, Héctor Jiménez y David Pinto</i>	
Pruebas con Algoritmos de Agrupamiento para generar una Base de Datos Léxica .....	304
<i>Sonia Paniagua, Héctor Jiménez y David Pinto</i>	
Unsupervised Learning of Word Combinations for Syntactic Disambiguation.....	311
<i>Tania Lugo, Alexander Gelbukh, Grigori Sidorov</i>	

---

## **Tecnología de objetos de aprendizaje**

---

Introducción.....	321
Mundo óseo para niños.....	323
<i>Misael Medina Hernández, Edgar Ortiz Sánchez, Marva Angélica Mora Lumbreras</i>	
Objetos de Aprendizaje Adaptativos .....	330
<i>José Mario García-Valdez, Alfredo Cristóbal-Salas</i>	
Propuesta de Metodología para elaboración de cursos en línea utilizando Objetos de Aprendizaje y Software Libre como entorno de aprendizaje.....	338
<i>Juan P Cardona S., Francisco Álvarez R.</i>	
Guías de diseño para el desarrollo de objetos de aprendizaje.....	347
<i>Jorge Aguilar Cisneros, Jaime Muñoz Arteaga, Saúl Pomares Hernández</i>	

Propuesta de implementación para generación de materiales multimedia en línea utilizando un meta etiquetador y componentes OO CSPR(I) versión 3.0 para resolver los problemas de contenido, control, reutilización y secuenciación dinámica .....	356
<i>Rubén Peredo Valderrama, Leandro Balladares Ocaña, Cutberto García Tamayo</i>	
Diseño e implantación de una interfaz interoperable para un patrimonio de recursos educativos basado en una red de acervos abiertos y distribuidos de objetos de aprendizaje.....	366
<i>Victor Germán Sánchez Arias, Salvador López, Rafael Morales</i>	
De LOM a XML y SQL, o cómo almacenar metadatos en un repositorio de objetos de aprendizaje.....	373
<i>Rafael Morales y Salvador López Mendoza</i>	

---

## **Ingeniería de software**

---

Introducción.....	385
La Ingeniería de Software como Proceso Permanente en la Enseñanza de la Computación.....	387
<i>Ricardo Ruiz-Rodríguez, Mario Alberto Moreno Rocha</i>	
Propuesta de una línea curricular de Ingeniería de Software en la Licenciatura en Sistemas Computacionales.....	393
<i>Francisco Álvarez R., Juan Cardona S. Alfredo Weitzenfeld R.</i>	
Hacia una Especificación del Marco Conceptual para la Formación de Recursos Humanos en Ingeniería de Software .....	399
<i>Gerardo Padilla, Carlos Montes de Oca; Luis Felipe Fernández M.</i>	
Evaluación de la utilización de Rational Rose con el Proceso Unificado de Desarrollo de Software .....	405
<i>Ángeles Arenas Valdés, Juan Manuel Fernández Peña, Ángeles Sumano López</i>	
El Desarrollo de la Industria de Software en Aguascalientes: Diagnostico y Oportunidades.....	411
<i>Tirso Velázquez Navarro, Héctor Sánchez Herrera, Jaime Muñoz Arteaga</i>	
Mejora de Calidad del Software Desarrollado por Publisoft de Aguascalientes en base al CMM.....	418
<i>José Guillermo Torres Tovar, Maribel Romo Contreras, Jaime Muñoz Arteaga</i>	
Administración Estratégica del Personal en base al People-CMM para la Mejora de la Productividad del Software.....	425
<i>Teresa Cervantes López, Jaime Muñoz Arteaga</i>	

Las Revisiones Técnicas Formales de los requerimientos y el análisis utilizando Áncora y el Proceso Unificado de Desarrollo de Software.....	433
<i>Juan Manuel Fernández Peña, María de Lourdes Hernández Rodríguez, María de los Ángeles Sumano López</i>	
Aplicando extensiones de UML para desarrollo de aplicaciones Web .....	439
<i>Alejandro Botello, Sandra Dinora Orantes</i>	
Elementos del Proceso Unificado de Desarrollo de Software aplicados al desarrollo del Sistema Integral de Gestión de Calidad y de Comercio Electrónico del Café-Veracruz (SIGC3).....	446
<i>María de Lourdes Hernández Rodríguez, Beatriz Gutiérrez Almaraz, Julio César Aguilar</i>	

---

### **Problemas de la optimización en la computación paralela**

---

Introduction .....	455
A X Window Proxy Server for Display Wall .....	457
<i>Jesus Verduzco, Bruno Raffin</i>	
A polynomial approximation scheme for the $1 \mid r_j \mid L_{\max}$ scheduling problem.....	465
<i>Alexander Lazarev, Ruslan Sadykov</i>	
Solution Algorithms for Total Tardiness Scheduling Problem.....	474
<i>Alexander Kvaratskhelia, Alexander Lazarev, Andrei Tchernykh</i>	
Dynamic Analysis and Trace Simulation for Data Parallel Programs in ParJava Environment .....	481
<i>Victor Ivannikov, Sergey Gaissaryan, Arutyun Avetisyan, Vartan Padaryan, Hennadiy Leontyev</i>	
Clusterweb – a WEB-based cluster management interface .....	489
<i>Arutyun Avetisyan, Oleg Samovarov, Dmitry Grushin, Andrey Ryzhov</i>	

---

### **Robótica móvil**

---

Introducción.....	499
Object Identification and Location by Using Simple Sonar Sensors .....	503
<i>Luis Enrique Arámbula Miranda</i>	
Una arquitectura de software para la integración de aplicaciones en robótica móvil.....	512
<i>José Antonio Pacheco-Sánchez, Roberto Valdivia-Beutelspacher, Oscar Mayora-Ibarra</i>	
Deterministic sampling: A framework for Probabilistic Roadmap Planners.....	522
<i>Abraham Sánchez, René Zapata, and David Pinto</i>	
Analyzing Multi-Robot Box-Pushing.....	530
<i>Angélica Muñoz Meléndez, Alexis Drogoul</i>	



Current work at the Robotics group of the Universidad de Guanajuato .....	540
<i>Victor Ayala Ramírez and Raúl E. Sánchez Yañez</i>	
Grupo de robótica del INAOE .....	545
<i>Angélica Muñoz Meléndez</i>	

---

## **Cómputo móvil**

---

Introducción.....	551
Service Discovery in Mobile Ad-hoc Networks by Extending the AODV Protocol.....	553
<i>Dante Arias-Torres, José Antonio García-Macias</i>	
Lucrn: A middleware for application development in spontaneous networks of Bluetooth devices .....	559
<i>J. Alfonso Briones-García, Silvana Bravo-Hernández, J. Oscar Olmedo-Aguirre</i>	
Arquitectura de Colaboración Mediante Dispositivos Móviles Aplicada a la Administración del Conocimiento .....	565
<i>Chadwick Carreto Arellano, Rolando Menchaca García, Rolando Menchaca-Méndez, Agustín Gutiérrez Tornos</i>	
Mobile Access to Digital Libraries via Generic Interfaces.....	571
<i>Nohema Castellanos, J. Alfredo Sánchez</i>	
µCronos: Un micro Núcleo con Soporte para Adaptabilidad Segura de Bajo Nivel .....	577
<i>Víctor D. Castillo Díaz, Rolando Menchaca-Méndez</i>	
Context-Aware Middleware for Pervasive and Mobile Ad Hoc Environments .....	583
<i>Hector A. Duran-Limon</i>	
Seamless IP micro-mobility.....	589
<i>Christian P. García-Martínez, J. Antonio García-Macias</i>	
Requirements analysis for ubiquitous computing system to be used during a congress .....	595
<i>Oscar-Iván Lepe-Aldama, Luis-Enrique Vizcarra-Corral, Edgar Arriola-Arreola, Angel Galindo-Rodríguez, Frankli-Augusto Pérez-Navidad</i>	
Administrador de Candados para Transacciones Anidadas sobre Dispositivos Móviles.....	601
<i>Jorge Martínez-Muñoz, Luis A. Gama-Moreno</i>	
Autonomous Agents for Supporting Mobility in AmI Environments.....	607
<i>Marcela Rodríguez Urrea, Jesús Favela , Edgar A. Martínez</i>	

---

## **Bioinformática y biología computacional**

---

Introducción.....	615
Clasificación No Sesgada de Proteínas con Redes Neuronales y Computación Evolutiva.....	617
<i>Angel Kuri, Martha Ortiz, Lucía Castellanos, Carlos Ponce, Christopher Gasca Oscar Herrera</i>	
Experimental Analysis of a Genetic Algorithm for the Sequencing by Hybridization Problem.....	623
<i>Luis C. González-Gurrola and Carlos A. Brizuela</i>	
Optimización de la Geno-tipificación de Papilomavirus Humano como un Problema de Selección de Características Utilizando PCA y RNA's.....	629
<i>M. Rodríguez-Muro, P. Gómez-Gil, C. Castañeda-Roldán, J. Garcés - Eisele, J. Lozano-Yécora, M. Osorio-Galindo</i>	
Using Graph-Based Concept Learning to the Predictive Toxicology Domain .....	635
<i>Ivan Olmos and Jesus A. Gonzalez</i>	
Stereo-sensitive Cortical Regions Identified Using Shape Discrimination from Stereopsis: an fMRI study .....	641
<i>Héctor-Gabriel Acosta-Mesa ,Nicandro Cruz-Ramírez, John Frisby , Ying Zheng, David Buckley, Janet Morris, and John Mayhew</i>	

---

## **Computación clínica e informática médica**

---

Introducción.....	649
Aplicación de una herramienta de realidad virtual colaborativa y adaptación de modelos gráficos en 3D para el apoyo del diagnóstico médico de lesiones óseas .....	651
<i>L. A. Cervantes Medina, M. A. García Ruiz</i>	
Apoyo a las actividades médicas a través de Servicios Web basados en HL7/CDA .....	657
<i>P. C. Santana Mancilla, L. Galicia Jiménez, A. I. Martínez García, J. A. García Macías</i>	
Clasificación no Paramétrica de la Distribución de Bordes Espectrales para la Valoración del Estado de Profundidad Anestésica .....	663
<i>B. D. Escamilla Alonso, O. Yáñez Suárez</i>	
Digital Image Processing of Functional Magnetic Resonance Images to Identify Stereo-sensitive Cortical Regions Using a Global Stereo Stimulus .....	669
<i>H. G. Acosta Mesa, N. Cruz Ramírez, J. Frisby, Y. Zheng, D. Buckley, J. Morris, J. Mayhew</i>	
Diseño de una Sala Virtual de Rayos X con Aplicaciones en Seguridad Radiológica.....	675
<i>C. R. Bustos Mendoza, M. A. García Ruiz, D. M. López Rosas, L. Hernández García</i>	

Experiencias de Implantación del Estándar DICOM.....	681
<i>J. A. Treviño, C. Feregrino Uribe</i>	
SAPeR: un sistema para la administración de pacientes en rehabilitación .....	687
<i>L. O. Escobedo Bravo, M. E. Tentori Espinosa, A. I. Martínez García, L. Galicia Jiménez, J. A. García Macías</i>	
SNIAC Sistema Nutricional Inteligente Asistido Por Computadora .....	693
<i>J. L. Ramírez Sánchez, J. A. Cruz Zamora</i>	
<b>Índice de autores</b> .....	<b>699</b>

# **Interacción humano-computadora: interacción – integración**

Oscar Mayora Ibarra  
Luis Villaseñor Pineda  
Carlos Alberto Reyes García  
(Editores)



## **Introducción**

La Interacción Humano-Computadora (IHC) es la disciplina que estudia el diseño, evaluación e implementación de sistemas informáticos interactivos para el uso de seres humanos, y el estudio de los fenómenos más importantes con los que está relacionado. En los últimos tiempos, esta área se ha convertido en una disciplina de suma importancia debido a que los sistemas computacionales están dejando de ser exclusivos de los expertos en computadoras y están pasando a ser una cuestión cotidiana para la gente en general. Es por eso que los sistemas que se construyan en el futuro próximo, deberán poder integrar al mismo tiempo funcionalidad y facilidad de usar para poder subsistir en el gran mercado global.

El objetivo general de este taller es identificar y reunir a la comunidad mexicana que trabaja en el área de IHC con el fin de intercambiar experiencias y presentar trabajos o propuestas de trabajo sobre tópicos referentes a la interacción humano – computadora. Así mismo, el taller fue una excelente oportunidad para informar acerca de la reciente reactivación del grupo de interés de ACM en México sobre IHC (SIG-CHI México) y la presentación del próximo Congreso Latinoamericano de HCI (CLIHC'05), México, octubre de 2005.

## **Comité Científico**

Copresidentes:

Oscar Mayora Ibarra, ITESM Cuernavaca, [omayora@itesm.mx](mailto:omayora@itesm.mx)

Luis Villaseñor Pineda, INAOE, [villasen@inaoep.mx](mailto:villasen@inaoep.mx)

Carlos Alberto Reyes García, INAOE, [kargaxxi@inaoep.mx](mailto:kargaxxi@inaoep.mx)

Alfredo Sánchez, UDLA, [alfredo@mail.udlap.mx](mailto:alfredo@mail.udlap.mx)

Manuel Romero, IMP, [mromeros@imp.mx](mailto:mromeros@imp.mx)

Víctor González, University of California, Irvine, [vmgyg@ics.uci.edu](mailto:vmgyg@ics.uci.edu)

Fernando Gamboa, UNAM, [gfer@aleph.cinstrum.unam.mx](mailto:gfer@aleph.cinstrum.unam.mx)

# Interfaz de Comunicación y Administración de Tareas (ICAT) para Grupos de Trabajo Utilizando Software Libre

Ironelis Valdez Batista Juan Contreras-Castillo Gerardo Barreto Palacios

Facultad de Telemática Universidad de Colima  
Av. Universidad 333 Col. Las Víboras Colima, Col. 28044 México  
ing\_valdez\_b@yahoo.com

## Abstract

In general, the Internet has been considered as a great database of global coverage more than as a place in which to carry out specific work. In this context, classic tools such as electronic mail and distribution lists are not capable of offering better support for the realization of group work through a network. Starting at this point, groupware tools for the Internet offer the users the basic pillars from which promote collaborative work via the Web. Based on this concept, we developed ICAT, a tool designed for the administration of and communication between work groups that are located in distributed places.

## Resumen

Normalmente, El Internet ha sido considerado como una gran base de datos de cobertura global más que como un lugar en donde llevar a cabo un trabajo concreto. En este contexto, las herramientas clásicas como el correo electrónico y las listas de distribución no son capaces de ofrecer un mayor soporte para la realización de trabajos en grupo a través de la Red. En este punto es donde se están desarrollando las aplicaciones Groupware para Internet, que ofrecen a los usuarios el pilar básico desde el cual promover el trabajo colaborativo a través del Web. Basado en este concepto, se diseña el ICAT, una herramienta de software diseñada para administrar y comunicar grupos de trabajo ubicados en lugares remotos.

**Palabras claves:** HCI, CSCW, ICAT.

## 1. Introducción

En los últimos años, el trabajo en grupos de desarrollo de software distribuidos ha generado la necesidad de compartir información dada la separación temporal o geográfica de los participantes. Sin embargo, la falta de herramientas de comunicación electrónica que permitan administrar las actividades de estos grupos e intercambiar información entre ellos constituye una limitante cada vez más fuerte.



Situación que orilla a la división del trabajo en fases que se manejan como partes separadas. Esto, de alguna manera representa un problema, dado que cada fase de un proyecto se relaciona con las posteriores y anteriores, por esto es tan necesaria la existencia de herramientas que permitan el monitoreo constante de las actividades o tareas que conforman un proyecto, y que, en muchos casos, son llevadas a cabo en puntos geográficamente distantes y que a la vez permitan una interacción rápida, informal y fluida entre los miembros del equipo de trabajo.

En este respecto, las tecnologías de groupware pueden servir como apoyo para la realización de estas tareas, ya que según Bannon (1993) CSCW puede ser concebido como un esfuerzo para entender la naturaleza y las características de trabajo cooperativo con el objetivo de diseñar tecnologías adecuadas basadas en la computadora.

## **2. Antecedentes**

El trabajo colaborativo se define como el proceso mediante el cual un conjunto de individuos realizan actividades relacionadas con la finalidad de lograr un objetivo específico, cumplir una meta o compartir una ideología común [1]. Según la teoría más reciente, el trabajo colaborativo persigue el desarrollo de conocimientos compartidos, la aceleración de los flujos de información, la coordinación de los flujos de recursos para producir economías de costos y tiempos en el desarrollo de proyectos [2]. Un proyecto se define como el esfuerzo temporal que se realiza para crear un producto o servicio único. El término temporal de la definición se refiere a que cada proyecto tiene una fecha de inicio y una fecha de término, y, el concepto único de la definición, se refiere a que el producto o servicio contienen elementos o rasgos que los distinguen de los productos o servicios ya existentes. Administrar un proyecto es mucho más que encontrar el camino crítico que conlleva a un buen término: además, incluye el control de fechas para las tareas, de modo de prever días feriados o vacaciones de personal; el uso de recursos no almacenables, tal como tiempo de trabajo de personas, equipos o instalaciones; el cálculo de fondos necesarios día a día para hacer frente a los pagos de materiales o de servicios contratados; el cálculo del costo global del proyecto para hacer presupuestos, etc [1].

En la actualidad existen distintas herramientas electrónicas para dar soporte al trabajo colaborativo, tales como: Project de Microsoft (<http://www.monografias.com/trabajos/msproyect4/msproyect4.shtml>), LeFiell (<http://www.lefiell.co.com/espmsg.htm>), Jabber (<http://www.jabber.com/index.cgi>) entre otros. Además, se utiliza el teléfono como medio de comunicación tradicional, lo que conlleva a un gasto excesivo, y en otro ámbito se utiliza el correo electrónico, para compartir ideas, información, archivos y demás [3]. La problemática en estas herramientas, es que no se integra en ellas el concepto de administración de proyectos y de comunicación instantánea, lo que conlleva a complejidad, más uso de recursos al tener que utilizar dos o más herramientas distintas y muchas veces incompatibilidad entre ellas. En éste sentido se hace necesaria la existencia de herramientas que permitan el monitoreo constante de las actividades o tareas que conforman un proyecto, y que, en muchos casos, son llevadas a cabo en puntos geográficamente distantes y que a la vez

permitan una interacción rápida, informal y fluida entre los miembros del equipo de trabajo. La mensajería instantánea es un de los métodos de colaboración textual más importantes que existen, dado que ofrece minimización de costos y tiempo. La minimización de costo en un entorno compartido se logra al compartir información, ideas y sugerencias, el cual permite al usuario acceder a la información e interactuar con sus compañeros de grupo, evitando la movilización del usuario. La comunicación en un grupo usando herramientas de comunicación instantánea es más fluida, más rica, no requiere de introducción y facilita la interacción entre todos los integrantes del grupo, por estas y otras razones el considerar herramientas comunicación instantáneas para grupos, asegura brindar interacciones oportunas y eficaces [4].

### **3. Solución propuesta**

Con miras a obtener una herramienta que integra la administración de tareas y la comunicación instantánea, se desarrolla el ICAT (Interfaz de Comunicación y Administración de tareas para Grupos de Trabajo utilizando software libre), una herramienta para administrar y comunicar integrantes de un grupo de trabajo, y que además esté desarrollada bajo el concepto de Open Source, en busca de obtener mayor crítica y retroalimentación, que genere riquezas sin la necesidad de adquirir licencias de software, conservando efectivamente los recursos, donde fueron generados, y donde más se necesitan.

El ICAT fue creado con el propósito de lograr una herramienta de comunicación instantánea inter e intra-grupal con capacidades de administración, asignación y monitoreo de actividades. Su objetivo principal es integrar grupos de trabajo cuyos participantes se encuentren distribuidos en lugares remotos. Surge por la necesidad de unir dos conceptos fundamentales en el trabajo colaborativo.- Administración de proyectos y comunicación instantánea. Esta herramienta consta de dos módulos principales: uno de comunicación instantánea y otro de asignación y monitoreo de tareas o actividades, de esta forma se integra de una manera amigable los dos conceptos antes citados.

Para estar claro con la definición y aplicación del ICAT, resulta necesario explicar un concepto que va muy de la mano con el criterio de dicha herramienta; Interacción Humano computadora (HCI). Interacción Humano computadora es una disciplina referida al diseño, evaluación y puesta en marcha de los sistemas de cálculo y operaciones interactivos de humano-máquina [5]. Una de las características de HCI, es el manejo de interfaces de grupo, que es precisamente lo que hace referenciar el ICAT con el HCI.

Para el diseño del ICAT se toman en cuenta las directrices citadas por Nielsen (2002), quien recomienda, en primer lugar, determinar el contenido de la página de inicio según las necesidades de las empresas. Pero, dado que el ICAT no está diseñado para una empresa en particular se decide dar prioridad a los elementos que se considera deben estar en la página de inicio para beneficio del usuario; en esta página se usan colores en distintos tonos, basados siempre en las recomendaciones de

Nielsen, que recomienda evitar colores fuertes a la vista del usuario. En esta página de inicio se encuentran también el logotipo de la herramienta en un tamaño razonable; el eslogan, para definir explícitamente las actividades del sitio; objetos de orientación para el usuario como calendario y hora actual, además los campos requeridos para entrar a la herramienta y el botón de ingreso.

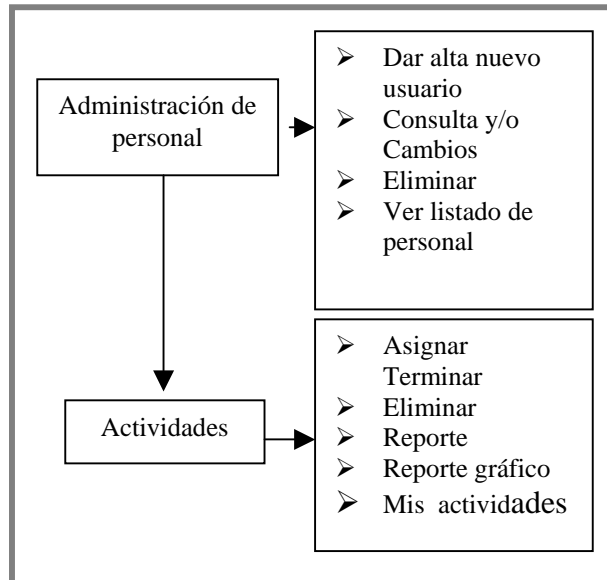
La herramienta está compuesta por dos módulos, uno de administración y asignación de actividades, y otro de comunicación, en éste último, además de permitir la interacción de los usuarios, permite el monitoreo de las actividades, mediante una gráfica, en la que el usuario ve el estado de cada una de las fases de las actividades del otro usuario. Actualmente se utiliza el e-mail (correo electrónico) así como también el MSN; pero de alguna manera estas siguen arrastrando la variable que se quería eliminar con la creación del ICAT: el costo; variable de mucho peso en las aplicaciones comerciales, costo que no directamente es por el pago de las licencias, sino por la poca o ninguna libertad con que cuenta el usuario para manejar y manipular dichas herramientas según sus necesidades. La implementación del ICAT se logra con la utilización de software disponibles en la red (Internet), los cuales son ampliamente difundidos en el área del desarrollo de sistemas por ser tecnología con la madurez suficiente para ser aplicados acertadamente en este proyecto (Tabla 1).

<b>Tipo</b>	<b>Nombre</b>	<b>Características</b>
Servidor www	Apache	Servidor web más empleado en el mundo, flexibilidad de configuración.
Base de datos	MySQL	Motor de base de datos relacional, compacto y eficiente.
Interfaz	HTML	Lenguaje de definición de formato y contenido.
Interfaz	JavaScript	Lenguaje de programación.
Interfaz	PHP	Lenguaje de programación.
Modelado	UML	Modelado del sistema.

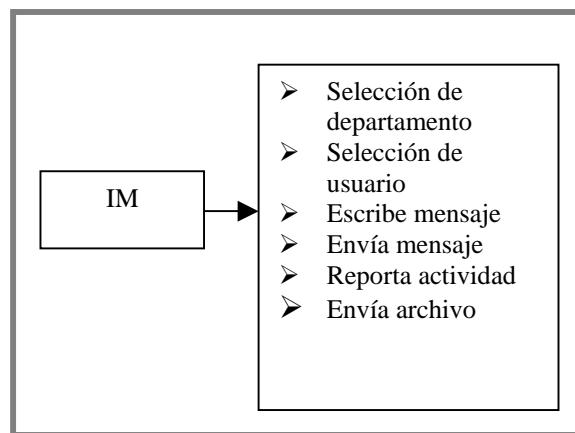
**Tabla1. Software utilizados para el desarrollo del ICAT**

La finalidad principal de este sistema es proporcionar un entorno para trabajo colaborativo que permita a los integrantes de un grupo de trabajo interactuar en tiempo real, asignar, administrar y monitorear las actividades pertenecientes a su grupo. La herramienta se ha dividido en dos módulos: asignación y monitoreo de las actividades y de Comunicación Instantánea. Es bueno aclarar, que cuando se dice que la herramienta está compuesta por dos módulos, ambos son complemento uno del otro, por lo que ninguno debe ser visto como parte separada.

A continuación se presentan los módulos que forman el ICAT y la información y opciones que los componen a cada uno (Figuras 1 y 2).



**Figura 1. Módulo de administración y administración de tareas.**



**Figura 2. Módulo de comunicación instantánea**

Con la unión de los dos módulos se obtiene las siguientes Figura 3, 4 y 5.



Figura 3. Ventana de inicio.

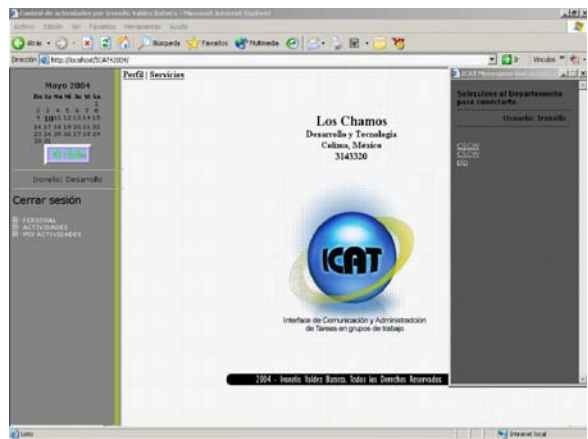


Figura 4. Sesión iniciada.

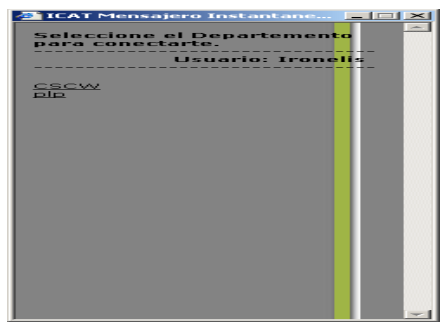


Figura 5. Ventana inicio IM.

El objetivo del módulo de mensajería instantánea del ICAT es; además, de permitir la interacción entre los integrantes de un grupo de trabajo, la visualización gráfica del estado o alcance de las actividades asignadas a un usuario. Esta opción se obtiene en tiempo real, dado que éste módulo cuenta con gráfica (Gantt) en la parte superior de cada ventana de conversación. En la que se muestra el estado de las actividades del usuario receptor. De modo que se tiene en una misma ventana la opción de interactuar y de monitorear alcance o atraso las actividades según el tiempo programado de inicio y fin de cada una (ver Figura 6). En ésta grafica, los elementos en rojo representan el inicio, azul con palomita los días reportados, y los grises la duración de la actividad.

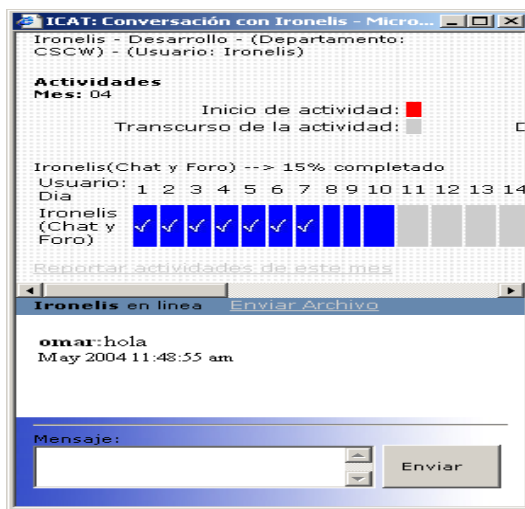


Figura 6. Ventana de Comunicación Instantánea.

### 3.1 Prueba informal de la herramienta

Una vez concluida la implementación y la revisión por el autor, para fines de probarla se lleva a cabo un caso de estudio con un empresa hipotética, en la que se realizó un Proyecto para ser terminado en dos semanas, con un número de 7 personas a las cuales se le sugirió hacer uso del ICAT como herramienta de trabajo. Finalmente, se aplicaron cuestionarios con fin de evaluar el diseño, utilidad y usabilidad de la herramienta. Los resultados menos favorables fueron con respecto al diseño, situación que no es para preocuparse del todo, dado que cada persona tiene una percepción distinta para el diseño; con relación a la utilidad y usabilidad, se puede concluir, que el objetivo que de crear una herramienta para administrar y comunicar grupos de trabajo se ha cumplido satisfactoriamente.

#### 4. Conclusiones

La información constituye una valiosa herramienta en la planeación de proyectos en entornos colaborativos, y mucho más la facilidad de contar con un medio que permite el intercambio de dicha información. Por esta razón se desarrolló el ICAT, una herramienta para trabajo de grupo en un entorno colaborativo, ésta hace posible la interacción entre los usuarios, el seguimiento de las tareas que conforman un proyecto, reportes gráficos de los avances de las actividades, etc.

#### 5. Agradecimientos

Cuando surge la idea de crear una herramienta multiplataforma que integre el concepto de comunicación instantánea y administración de tareas en grupos de trabajo, existe la limitante de que se cuenta con las ideas, pero no se tiene la forma de ordenarlas de modo que se puedan crear fases para ir evaluando la materialización de la idea inicial. Es ahí donde entran personas como el ingeniero Pedro Fernando Mir, el M.C Armando Román Gallardo, el Ing. Omar Chávez quienes han colaborado de manera muy estrecha en el desarrollo, prueba y evaluación de la herramienta ICAT.

#### 6. Referencias

- [1] H. Koontz and H. Weihrich, *Administración una perspectiva global*, 11 ed. México: Mc-Graw Hill, 1998.
- [2] A. DIAZ M., "E-business: Tecnología de información y redes de negocios," *Debates IESA*, vol. V, 2000.
- [3] V. M. González González, "«Dinámica de Grupos Electrónicos»," *Soluciones Avanzadas.*, vol. Año 7, ., pp. 22-24, 1999.
- [4] E. Issacs, A. Walendowski, and D. Ranganthan, "Hubbub: A sound-enhanced mobile messenger that supports awareness and opportunistic interactions," vol. 4, pp. 179-185, 2002.
- [5] Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, S. and, and Verplank, "ACM SIGCHI Curricula for Human-Computer Interaction," *ACM SIGCHI*, pp. 5, 2002.

#### Los autores

**Valdez Batista, Ironelis** ing\_valdez\_b@yahoo.com

Profesora de la Universidad Central del Este (UCE). Maestra en Ciencias, área Telemática. Universidad de Colima, Colima, Col. México.

**Contreras-Castillo, Juan** juancont@ucol.mx

Profesor-investigador de tiempo completo de la Facultad de Telemática (Universidad de Colima), Colima, Col. México.

**Barreto Palacios, Gerardo** grardbato@yahoo.mx

Estudiante de la Facultad de Telemática Universidad de Colima.

# Integración de Elementos de Usabilidad en la Ingeniería de Requerimientos mediante Áncora

Karen Cortés Verdín<sup>1</sup>, Ángeles Sumano López<sup>1</sup>, Dolores Vargas Cerdán<sup>2</sup>

<sup>1</sup>Facultad de Estadística e Informática, Universidad Veracruzana, Av. Xalapa s/n,  
<sup>2</sup>Órgano de Fiscalización Superior del Estado de Veracruz, Avenida Arco Sur s/n, Xalapa, Ver. [kcortes@uv.mx](mailto:kcortes@uv.mx), [asumano52@yahoo.com](mailto:asumano52@yahoo.com), [mdvargasc@hotmail.com](mailto:mdvargasc@hotmail.com)

## Abstract

Áncora is a methodology for Requirement Engineering with orientation to user involvement, which includes linguistics, psychosocial and planning aspects. As usability is a key quality factor for software development, some elements of Usability Engineering were included in Áncora's lifecycle. During the development of several case studies, it was possible to state the feasibility of introducing usability elements within the first stage of software development.

## Resumen

Áncora es una metodología de Ingeniería de Requerimientos con orientación al usuario que incluye aspectos lingüísticos, psico-sociales y de planeación. Debido a que la usabilidad es un factor clave para el desarrollo de software, algunos elementos de Ingeniería de Usabilidad se incluyeron dentro del ciclo de vida de Áncora. Durante el desarrollo de varios sistemas, fue posible determinar la factibilidad de introducir elementos de usabilidad en la primera etapa del desarrollo de software.

**Palabras clave:** Usabilidad, requerimientos, metas de usabilidad, evaluación, prototipo

## Introducción

La usabilidad es uno de los aspectos menos cuidados cuando se desarrolla un sistema de software, una de las razones principales es la falta de tiempo de los desarrolladores. Por lo anterior en este trabajo se muestra una forma de incluir la usabilidad desde la primera etapa del desarrollo de software, la de establecimiento de requerimientos.

Áncora es una metodología de Ingeniería de Requerimientos, abarca la obtención y análisis de requerimientos de software, que conjunta métodos computacionales, psico-sociales, de planeación y lingüísticos. Dada su orientación al usuario se decidió incluir en Áncora algunas estrategias que permitan determinar aspectos de usabilidad a la par que se realiza la especificación de requerimientos. Después de una revisión de métodos y técnicas para la definición de usabilidad propuestos por varios autores y de su aplicación en varios sistemas, en este trabajo se presenta la inclusión, en Áncora, de los elementos ya probados.

El trabajo se divide como sigue: en la sección uno se dan la definición de Ingeniería de Usabilidad (IU) y las etapas de su ciclo vida, en la sección dos una breve descripción de la metodología Áncora, en la sección tres se describe cómo perfilar al usuario *adaptando dos de los artefactos* de la Áncora, en la sección cuatro se muestra cómo establecer requerimientos y metas de usabilidad *agregando dos nuevos artefactos* a Áncora, en las secciones cuatro y cinco se señala *cómo se aplican* las metas de usabilidad en la construcción del prototipo y su respectiva evaluación.



## 1. Ingeniería de Usabilidad

Una característica importante que debe mostrar un sistema de software es la usabilidad. Un sistema de software puede tener varias características de calidad como: rendimiento, eficiencia, correctez y robustez; pero si no es fácil de usar, no será bien empleado por sus usuarios y por lo tanto, nunca podrá ser aprovechado en su totalidad.

La usabilidad determina hasta qué punto un sistema puede ser empleado por sus usuarios, de tal manera que los apoye en el desarrollo de sus actividades y en el logro de sus metas. Tanto Nielsen [3] como Shneiderman [4] definen la usabilidad, en términos de cinco factores o propiedades del sistema con los cuales interactuará el usuario:

1. Facilidad de aprendizaje, de manera que el usuario rápidamente pueda trabajar con el sistema.
2. Eficiencia, es decir que una vez que el usuario ha aprendido a usar el sistema, él pueda obtener un alto nivel de productividad.
3. Facilidad de retención. Las opciones y operaciones son fáciles de retener.
4. Baja incidencia de errores, o sea que el sistema tenga un bajo nivel de errores.
5. Satisfacción. Debe ser agradable de usar, de manera que los usuarios se sientan satisfechos al usarlo.

La IU consiste en una serie de actividades a desarrollar a lo largo del ciclo de vida del sistema con un especial énfasis en involucrar al usuario desde las primeras etapas del mismo [3]. Los aspectos más importantes para la usabilidad son: las tareas de los usuarios, sus características individuales y sus diferencias. La IU debe incorporarse al proceso de desarrollo del producto de software, pues al incluir factores humanos asegura el impacto positivo del producto en sus clientes. Las etapas de la IU son:

1. Conocimiento del usuario.	4. Diseño paralelo.	8. Elaboración de prototipos
2. Análisis competitivo	5. Diseño participativo	9. Diseño iterativo
3. Establecimiento de metas de usabilidad	6. Diseño coordinado de la interfaz total.	10. Retroalimentación del uso en campo
	7. Aplicación de lineamientos y heurísticas de análisis.	

## 2. Áncora

Áncora (**Análisis de Requerimientos Conducente al Reuso de Artefactos**) [5] es una metodología para la Ingeniería de Requerimientos, que se basa en la interacción con los involucrados en el desarrollo del sistema. Se ha utilizado en más de cincuenta proyectos de software con muy buenos resultados y con gran satisfacción por parte del usuario. El elemento que articula a la metodología de Áncora es el guión, en el que existen: escenas, utensilios y papeles (interpretados por los usuarios del sistema). Un ejemplo de guión es el de la Figura 1. El ciclo de vida de Ancora es el siguiente:

1. Entendimiento del Dominio y Contexto de la Aplicación. El objetivo de ésta es definir claramente las metas de la empresa, el dominio de la aplicación, las actividades que actualmente lleva a cabo en el sistema así como la justificación del nuevo sistema.
2. Recolección y clasificación de requerimientos. El producto de esta etapa es la propuesta computacional del nuevo sistema.

3. Reuso de requerimientos. Lo que se busca en esta etapa es dar al analista recursos de otros sistemas que ya están probados y cuyas especificaciones de requerimientos se encuentran ya en una base de datos de reuso.
4. Resolución de conflictos, priorización y validación de requerimientos. Para ello, se realizan la Reunión de Reflexión y Diseño.
5. Cierre. En donde se dan pautas al desarrollador de software para pasar a las siguientes etapas de desarrollo.

<p><b>Guión:</b> Diagnóstico Empresarial</p> <p><b>Papeles:</b>          AT = Asistente Técnico          CO = Consultor          EMP = Empresa          EM = Empresario o Cliente          DG = Director general          CUSEM</p> <p><b>Útiles:</b>          R1= Cédula de registro de consultor          DE = Directorio de Empresas          C1= Solicitud de consultoría          C2= Formato de seguimiento de consultoría          Cu = Cuestionario          C = Computadora          Ca = Cañón de proyección          RD = Reporte de Diagnóstico          R = Recomendaciones          P = Propuesta          IC = Indicadores de Consultoría</p> <p><b>Condiciones de entrada:</b>          EM quiere mejorar situación de su empresa          CO quiere o puede apoyar</p> <p><b>Condiciones de salida:</b>          EM establece relación de trabajo con CUSEM o la rechaza</p>	<p><b>Escena 1:</b> Identificación del usuario          USU ingresa nombre y contraseña          ¿nombre o contraseña incorrecta?          {USU sale o va a escena1}</p> <p>USU ingresa al sistema</p> <p><b>Escena 2:</b> Registrar empresa          AT captura datos de Empresa en DE</p> <p><b>Escena 3:</b> Registrar Consultor          AT captura R1</p> <p><b>Escena 4:</b> Registro de C1          AT captura C1</p> <p><b>Escena 5:</b> Registro de cita con EM          CO o AT capturan cita</p> <p><b>Escena 6:</b> Adecuación de CU          CO ingresa al sistema          CO revisa cuestionario y modifica</p> <p><b>Escena 7:</b> Registro de respuestas de CU          CO registra respuestas a Cu          CO guarda Cu contestado</p> <p><b>Escena 8:</b> Elaboración del RD          CO elabora RD con resultados del Cu          ¿Está incompleto?          { Ir a la escena 11}</p> <p><b>Escena 9:</b> Actualización respuestas Cu          CO registra o modifica respuestas a Cu          CO guarda Cu contestado</p> <p><b>Escena 10:</b> Elaboración del RD          CO elabora RD con resultados del Cu</p> <p><b>Escena 11:</b> Revisión del RD          CO entrega RD para revisión a DG          DG revisa RD y hace observaciones          CO realiza modificaciones al RD</p> <p><b>Escena 12:</b> Elaboración de R y P          CO elabora R y P para EM usando RD</p> <p><b>Escena 13:</b> Seguimiento          AT o CO registra C2</p> <p><b>Escena 14:</b> Obtención de Indicadores          AT o DG consultan/imprimen IC</p>
---	--

Figura 1. Guión de la propuesta computacional del Sistema de Diagnóstico Empresarial para una empresa de consultoría.

### 3. Introducción de actividades y herramientas de IU en Áncora

Durante la etapa 1 de Áncora, se llevaron a cabo algunas de las actividades de IU necesarias para determinar los roles y perfiles de usuarios. En la segunda etapa, de recolección y clasificación de requerimientos, al realizar el guión de la propuesta computacional (se ve un ejemplo en la Figura 1) se definen las metas de usabilidad del sistema y se dan pautas para la elaboración del prototipo rápido. Por último, en la etapa de resolución de conflictos, priorización y validación de requerimientos se propone incluir lo correspondiente a los requerimientos de usabilidad.

### 3.1 Perfilando al usuario

La primera etapa del ciclo de la IU es realizar un estudio de usuarios, el objetivo es determinar las características individuales de los usuarios, los roles o papeles desempeñados por los usuarios así como la variabilidad en las tareas, que son los factores que impactan directamente en la usabilidad. Áncora ya realiza un análisis de tareas mediante los guiones de la situación actual y de la propuesta computacional que obtiene el analista después de establecer contacto con los usuarios mediante observación, y aplicación de 4 tarjetas, para la Red Semántica Natural del Dominio de la aplicación, y de la encuesta de actitud. Aprovechando este momento, se incluyó una tarjeta adicional para que los usuarios, en forma anónima, listen la paquetería o sistemas que han empleado, jerarquizándolos en cuanto a gusto y experiencia en la paquetería o sistemas.

Posteriormente, en Áncora, se aplica la encuesta de actitud, cuyos objetivos son: detectar la actitud que los usuarios tienen hacia la eventualidad del desarrollo de un nuevo software y la detección de los principales problemas. Conjuntamente, para fines de usabilidad, se aplica un cuestionario para la definición de perfiles de usuarios que contiene, principalmente, preguntas sobre: edad, nivel de preparación, ocupación, gustos, motivación, conocimiento del dominio de aplicación y experiencia en el manejo de cómputo e Internet. Estos datos se resumen en un instrumento en Excel diseñado de acuerdo a [2], el cual permitirá determinar los requerimientos de usabilidad. Con los resultados del cuestionario, se procede al establecimiento de metas de usabilidad.

### 3.2 Establecimiento de metas de usabilidad

Una vez que se han determinado los requerimientos de usabilidad, éstos se relacionan con cada una de las tareas analizadas. Como ya se tiene realizado el estudio de usuarios y ya se conocen los perfiles de éstos, se pueden determinar fácilmente cuales serán los requerimientos de usabilidad que la aplicación deberá cumplir para determinada tarea (escena del guión o caso de uso).

Escena	Características de la interfaz	FA	E	FR	BE	S
2: Registrar Empresa	- Capacidad de captura rápida de datos - Validación de datos - Ayuda en línea - Soporte para interrupciones - Retroalimentación - Capacidad de deshacer	√	√	√	√	√
3: Registrar Consultor	- Capacidad de captura rápida de datos - Validación de datos - Soporte para interrupciones - Retroalimentación - Capacidad de deshacer	√	√	√	√	√

**Tabla 1. Requerimientos de usabilidad (FA = fácil de aprender, E = Eficiente, FR = Fácil de Recordar, BE = Bajo en incidencias de Errores, S = Satisfacción)**

Para cada una de las escenas del guión de la propuesta computacional se agregan seis columnas, la primera con las características que la interfaz deberá tener, las cuales se obtienen a partir de las actividades descritas por la escena correspondiente, cada una

de las cinco columnas restantes corresponde un requerimiento de usabilidad de los propuestos por [3] y se coloca una marca en el requerimientos si, de acuerdo a lo encontrado en el estudio de usuarios, dicho factor es importante para la escena. En la Tabla 1 se muestra un ejemplo del guión con las columnas de características y requerimientos de usabilidad por escena del guión de la Figura 1. Para cada una de las escenas y para cada uno de los requerimientos de usabilidad marcados, se definen las metas de usabilidad, “una meta de usabilidad significativamente establecida es aquella que es exitosa al comunicar con la menor ambigüedad posible, un intento del equipo de desarrollo del producto de entregar un sistema usable” [6]. Hay tres componentes en una meta de usabilidad: 1) Desempeño, ¿qué debe poder hacer el usuario?, 2) Circunstancias, ¿bajo que condiciones debe el usuario ser capaz de hacerlo?, 3) Criterios, ¿qué tan bien debe hacerse?

Muchas de las metas de usabilidad se obtienen del análisis de tareas propuestas en [1]. Para cada meta se construye una tabla (ver Tabla 2), la cual corresponde al requerimiento eficiencia de la Escena 2 del sistema de diagnóstico empresarial (Figura 1). Es útil definir los criterios de éxito en tres niveles, ya que determinan un mecanismo de medición para la meta, lo que a su vez permite planear las pruebas de usabilidad. De esta manera, el último renglón permite sentar las bases para la etapa de evaluación. Al definir los criterios, debe tenerse en mente: 1) al usuario no al producto [1] y 2) que no todos los criterios aplican igual a todas las metas ni los valores en cada nivel pueden ser los mismos para todas ellas.

<b>Registrar Empresa</b>	<b>Meta: Eficiencia</b>		
¿Quién?	Asistente Técnico		
¿Qué debe poder hacer el usuario?	Capturar toda la información de la empresa en el sistema		
¿Bajo qué condiciones debe poder realizarlo?	1) tenga todos los recursos a la mano, 2) haya recibido capacitación previa o haya leído previamente el manual de usuario		
¿Cuáles son los criterios de éxito?	Excelente	Aceptable	Inaceptable
	3 minutos	5 minutos	Más de 5 minutos
¿Qué método de evaluación se empleará?	Se les entregará un formato con información de la empresa para su captura y se tomará el tiempo de terminación exitosa de la tarea.		

**Tabla 2. Metas de usabilidad de la Escena 2**

#### **4. Elaboración del Prototipo y Evaluación de Usabilidad**

Una vez que están definidas las metas de usabilidad, se procede al diseño de la interfaz. Para ello, existen diversas maneras, entre ellas el diseño participativo [3] con la ayuda de prototipos. Así que, se aprovechó que dentro del ciclo de vida de Ancora existe la actividad de la construcción del prototipo en papel (el manual preliminar del usuario). Este método resultó de gran ayuda tanto para los desarrolladores como para los usuarios. Para estos últimos, al tener una interfaz en papel con la explicación completa de su funcionamiento, les es más fácil hacer sus observaciones o propuestas directamente en el papel y para el desarrollador porque puede detectar y corregir las observaciones hechas por los usuarios sin estar ellos presentes. Así, se puede iterar tantas veces como

se considere necesario hasta que el desarrollador tenga claras las características que la interfaz deba tener.

Por otra parte, en la construcción del prototipo Áncora define que, para sistemas de un solo guión, cada escena de la propuesta computacional será una opción en la barra de menús (o de primer nivel) y las acciones dentro de cada escena (quintetas) corresponden a menús Pop-up o de segundo nivel.

La primera de las evaluaciones de los requerimientos de usabilidad, bajo esta propuesta, se lleva a cabo en la Reunión de Reflexión y Diseño (RRD) [5] dentro del ciclo de vida de Áncora. La RRD ayuda a jerarquizar y validar los requerimientos del software, por lo que se agrega un punto adicional sobre la opinión de los usuarios acerca de la interfaz del prototipo en papel. Los aspectos que se consideran fueron: 1) opinión acerca de la disposición de las opciones de la interfaz, 2) despliegue de resultados, colores empleados y 3) secuencia lógica de operaciones. Esto ha permitido obtener buenas ideas para el diseño de la interfaz y determinar los puntos comunes entre todos los usuarios, lo que ha facilitado la aceptación del sistema.

## 5. Evaluación y Conclusiones

Como una primera evaluación sobre la introducción de elementos de usabilidad en la fase de requerimientos, se llevó a cabo una encuesta de usabilidad, basada en la propuesta [4], de los productos terminados. Se observó que la usabilidad en ellos aumentó, logrando mensajes más claros, mejorándolos estéticamente y manejando el lenguaje del usuario. Sin embargo, tienen calificación media en algunos puntos como son: las ayudas y la facilidad de aprendizaje.

Al haber trabajado elementos de usabilidad junto con la Ingeniería de Requerimientos mediante Áncora se logró detectar la necesidad de formalizar el proceso de usabilidad, ya que las adecuaciones realizadas de manera experimental arrojaron buenos resultados. Se presentó una primera formalización de dicho proceso. Se incluyeron en el ciclo de vida de Áncora las actividades correspondientes a: los perfiles de usuarios, establecimiento de requerimientos, metas de usabilidad y diseño y evaluación de la interfaz propuesta. Se seguirá trabajando para mejorar lo ya realizado e incursionar en los aspectos faltantes de las pruebas de usabilidad.

## Referencias

- [1] J.S. Dumas, J.C. Redish, J.C. "A practical Guide to Usability Testing", Intellect Ltd, 1999
- [2] D. J. Mayhew, "The usability engineering lifecycle: a practitioners approach for user interface design", Morgan Kaufmann Publishers, 1999
- [3] J. Nielsen, "Usability Engineering", New York Academic Press, 1993.
- [4] B. Shneiderman, "Designing the User-Interface", tercera edición, Addison-Wesley, 1997
- [5] M. A. Sumano, "Método para el análisis de requerimientos de software con enfoque conjunto psicológico, social y lingüístico conducente al reuso", Tesis doctoral, IPN, 2002.
- [6] XEROX Corporation. Usability Analysis & Design version 1.0.

# Desarrollo de un Marco de Trabajo para Juegos Matemáticos Multiusuarios en un Ambiente Distribuido

Adriana Miramontes Rivera, Gabriel López-Morteo, Gilberto López

Departamento de Ciencias de la Computación  
Centro de Investigación Científica y de Educación Superior de Ensenada, CICESE.  
e-mail: adrysm90@yahoo.com, galopez@cicese.mx, glopez@cicese.mx  
Km. 107 Carretera Tijuana - Ensenada, Código Postal 22860, Apdo. Postal 2732,  
Ensenada, B.C. México.

**Abstract:** *A research project, based on the way to learn mathematics through Interactive Instructors of Recreational Mathematics (IIRM), is being developed nowadays in the Computer Science Department of CICESE. The objective of this task is to make a group of individuals, in a virtual community, to support each other by means of recreational activities related to everyday situations. A Framework (FIIRM) is now being developed in order to set up a well defined technological platform, in which its first stage is focused on the construction of multiplayer mathematical games. We consider that the standardization in design and implementation of the multiplayer games is important in order to achieve our goal of the foment of mathematics learning through the use of the IIRM, in order to avoid user's confusion by abrupt changes in quality and features between math games within the system.*

**Resumen:** *En el Departamento de Ciencias de la Computación de CICESE, se desarrolla un proyecto de investigación sobre la manera de fomentar el aprendizaje de las matemáticas a través de Instructores Interactivos de Diversiones Matemáticas (IIDM), buscando que la colaboración entre individuos de una comunidad virtual se realice mediante actividades divertidas y relacionadas con situaciones cotidianas. Con el objetivo de establecer una plataforma tecnológica bien definida para el desarrollo de los IIDM, se inició la creación de un Framework (FIIDM) del cual, la primera etapa de su desarrollo está orientada en a la construcción de juegos matemáticos multiusuarios. Consideramos que el desarrollo del FIIDM es importante para la estandarización en el diseño e implementación de los juegos multiusuarios, con el propósito de ayudar al desarrollador para hacer uso de las diversas herramientas de la arquitectura de soporte y fomentar la reutilización de componentes dentro de los IIDM.*

**Palabras Clave:** Framework, Juegos matemáticos multiusuarios, software educativo

## 1. Introducción

Actualmente en CICESE se desarrolla un proyecto de investigación sobre la manera de fomentar el aprendizaje de las matemáticas a través del uso de aplicaciones orientadas a la educación, a través de situaciones cotidianas y divertidas.

Dentro de este esquema se ha creado el concepto de Instructores Interactivos de Diversiones Matemáticas (IIDM), definidos como componentes de software especializados en conceptos matemáticos representados a través de las matemáticas recreativas [2][3][4][5] y que tienen una estructura interna basada en XML.

En los IIDM se han explorado el uso de los juegos multiusuarios bajo la perspectiva de que éstos prometen convertirse en una de las aplicaciones más importantes dentro de los ambientes colaborativos [7], buscando que un concepto que en las aulas pudiera ser tedioso y por lo tanto difícil de aprender, se convierta en algo más sencillo y práctico a través de añadir elementos lúdicos y prácticos a la educación.

La construcción de los IIDM es compleja ya que requiere que el desarrollador haga uso de elementos variados para la programación, como por ejemplo los mecanismos de inicio y control de la sesión multiusuarios para la programación de elementos de interfaces modificadas para el trabajo distribuido. Cada instancia de IIDM puede variar en aspectos tales como las políticas de acceso, las reglas de creación de juego, las reglas de ejecución de juego, los roles que se deberán desempeñar dentro del juego, la cantidad de jugadores permitidos dentro de una aplicación, y las políticas de privacidad entre otras cosas. Esta información es recopilada a través de varias etapas, en las cuales el usuario va introduciendo la información necesaria para seleccionar y parametrizar su sesión de jugador, como por ejemplo si seleccionó jugar con uno o más usuarios, o se decidió por la iniciación de una instancia nueva. En cualquiera de éstos dos casos, el usuario tendrá que hacer uso también de mecanismos para la invitación a otros jugadores para participar en una sesión de juego, siendo éste proceso controlado por el sistema contenedor de los IIDM.

Debido a esta variedad de opciones, la arquitectura de soporte a los IIDM se vuelve compleja. Por ello y con el objetivo de establecer una plataforma tecnológica bien definida para el desarrollo de los IIDM, se ha iniciado la creación de un Framework para el Desarrollo de Instructores Interactivos de Diversiones Matemáticas (FIIDM) para la construcción de juegos matemáticos distribuidos y de uno o más jugadores.

La construcción del FIIDM está basada en componentes y hace uso de patrones de diseño para incrementar su potencial de reutilización de código. Con el propósito de reforzar la estandarización de mecanismos en la arquitectura, todas las comunicaciones intra-componentes se realizan a través de cadenas XML.

El FIIDM se apoya en el concepto de antesalas, en donde el usuario recibe toda la información necesaria para integrarse a una aplicación en ejecución o para crear una nueva de cualquiera de los juegos ofrecidos en el sistema. La información que se presenta al usuario para su elección está dosificada y es dependiente tanto del perfil del usuario como de las elecciones que este vaya haciendo durante el proceso.

Se espera que con el uso del FIIDM disminuya el tiempo de desarrollo de los IIDM

a la vez que disminuya la complejidad de su construcción. Consideramos que para lograr esta meta es importante la estandarización en el diseño e implementación de los juegos multiusuarios, con lo cual se habilita al desarrollador para hacer uso de las diversas herramientas de la arquitectura de soporte. Esto se apoya en la definición de un Framework por [6], quienes lo definen como un conjunto de clases que colaboran para ayudar a los desarrolladores a planificar el uso de dicho marco y que regularmente es empleado para la reutilización de código en nuevos proyectos de desarrollo.

Sin la implementación del FIIDM, la construcción de los IIDM mantiene una dependencia total de las decisiones de diseño e implementación del desarrollador, por lo que cada IIDM dentro del sistema estaría construido de manera diferente. Si a eso se agrega que el desarrollador tiene que construir la aplicación, las herramientas necesarias para su buen funcionamiento y que la reutilización que cada desarrollador puede hacer en la construcción de otro IIDM sería mínima, se corre el riesgo de que el trabajar de manera independiente al FIIDM pudiera resultar en el desarrollo de aplicaciones bajo un panorama poco efectivo y con un nivel de calidad variable.

En este escrito, se presenta el trabajo en proceso del desarrollo del FIIDM; en particular en lo que se refiere a la especificación de sus requerimientos y la arquitectura resultante. Además, se muestra el primer prototipo de IIDM desarrollado con el Framework.

## **2. Metodología**

Para el desarrollo del Framework, se siguieron los lineamientos para diseño basado en componentes, propuesto por [6] Esta metodología se basa en el empleo de UML para la realización de la definición de los componentes que conforman la arquitectura. Esta claro que la definición de los componentes nos lleva a tener un sistema con un débil acoplamiento, que en entre otras cosas es una de las ventajas del diseño basado en componentes y que facilita su reutilización. Esta definición la logramos basándonos en la modelación inicial de los estados de cada uno de los elementos de los componentes conceptuales de la arquitectura del framework en un escenario de uso en particular y seguimos el modelo de desarrollo iterativo para el proceso de análisis y diseño [1]. Fue en este proceso que se identificaron escenarios y procesos repetitivos dentro de la arquitectura, es por esto que optamos por el empleo de patrones de diseño específicos para la resolución de las necesidades que debíamos solventar una y otra vez. Posteriormente se definieron los candidatos a clases haciendo uso de las tarjetas CRC. En este punto, encontramos que la mejor manera de definir claramente a los candidatos a componentes, fue la modelación de los escenarios de uso mediante diagramas de secuencia. Posteriormente, se definieron los archivos de definición de datos (DTD) para la generación de las cadenas XML que son empleadas en las comunicaciones internas entre los elementos del Framework. Finalmente, se integraron los elementos del FIIDM junto



con los servicios de la arquitectura de soporte.

### **3. Resultados.**

El desarrollo del FIIDM se está ejecutando en varias etapas sucesivas. Actualmente se cuenta con la totalidad del análisis y diseño del Framework, y se están codificando las clases empleando el lenguaje de programación Java. La arquitectura de soporte se encuentra operando y disponible a través de un proyecto previo llamado "Los Supersabios. Un Ambiente de Aprendizaje basado en IIDM". A continuación se presentan los elementos más relevantes del Framework organizados a través de escenarios de uso.

#### **3.1 Arquitectura del Framework.**

El framework cuenta con una arquitectura basada en componentes que se comunican entre ellos a través de cadenas XML. Cada uno de ellos provee servicios especiales para la ejecución de un juego distribuido multiusuarios. Los componentes principales que conforman la arquitectura son:

Administrador de Antesalas: Crea una instancia de antesala por cada petición recibida y posteriormente obtiene de ella la copia de la tabla de los datos y envía los datos a la interfaz gráfica para el despliegue.

Administrador de Despachadores de Mensajes: Crea un despachador por cada mensaje recibido.

Administrador de Mensajes: Crea un objeto de tipo mensaje (protocolo mensaje) por cada mensaje que recibe del servidor de mensajería. Envía una notificación al Administrador de Despachadores de Mensajes y posteriormente le envía los parámetros al receptor.

Administrador de Conexiones de Bases de Datos: Mantiene y administra las conexiones al manejador de bases de datos.

Administrador de Roles: Asigna un rol al usuario que hizo la petición de acuerdo a la creación de la instancia, esto es, el rol de administrador si se trata de el creador de la instancia o el rol de jugador si no lo es.

Administrador de Estructura de Juegos: Instancia un objeto del tipo Estructura por cada petición de juego nuevo que se recibe. Envía el objeto Estructura al evaluador y finalmente la envía al juego para que se realice la instanciación con los parámetros ya definidos.

Administrador de Peticiones: Instancia un objeto Petición por cada solicitud de instanciación de una sesión de juego que recibe del servidor. Envía la notificación de que la solicitud o el protocolo de solicitud esta armado y después envía todos sus parámetros a la estructura, cuando ésta

lo solicita.

### 3.2 La elección e inicio del juego.

El FIIDM define como Antesala al espacio en donde el usuario obtiene toda la información necesaria para hacer la elección del juego (Fig. 1). Es aquí donde el usuario decide si se incorporará a una instancia ya activa o si se será requerida la creación de una instancia nueva. Se deciden también las políticas de privacidad del juego, la creación o ingreso a torneos y las reglas de creación de juego. Esta información es necesaria para a parametrización inicial de una instancia. Toda esta información se actualiza constantemente para evitar que se haga una parametrización de una instancia con datos erróneos. La parametrización de la instancia de un juego, es totalmente dependiente del diseño de la construcción del IIDM, ya que ésta puede tener lugar antes de la instanciación o bien, después de que el juego ha sido iniciado con valores estandarizados.

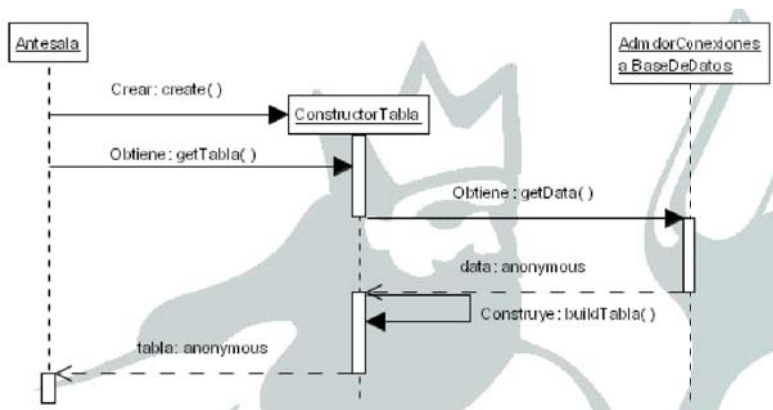


Figura 1. Diagrama de Secuencia que muestra el funcionamiento de la antesala.

### 3.3 El manejo del inicio de sesión.

Una vez definidos los parámetros para la instanciación, un escenario de uso típico es aquel en el cual un usuario haya elegido la creación de una instancia nueva y las reglas del juego que hagan necesaria la interacción con otros jugadores. En ese caso será necesario hacer invitaciones para que otros miembros de la comunidad virtual puedan incorporarse a una sesión de juego. El FIIDM cuenta con una herramienta que brinda soporte para la realización de invitaciones a usuarios que se encuentren en línea al momento de la instanciación de un juego. Esta invitación puede ser aceptada o declinada por el o los usuarios invitados a incorporarse a la sesión de

juego. Estas invitaciones se realizan con el mismo mecanismo que la comunicación interna de los jugadores, la cual se describe a detalle más adelante.

### **3.4 El soporte para usuario de incorporación tardía.**

El FIIDM también ofrece soporte para los usuarios que deseen incorporarse tardíamente a una instancia en ejecución. Siempre que las políticas de privacidad y las reglas del juego lo permitan, un usuario puede incorporarse a un juego previamente iniciado. Para ello, el historial del juego es almacenado en cada una de las copias del juego pudiendo ser enviada hacia una base de datos centralizada o bien a una nueva instancia que se incorpora a la sesión como un usuario que se incorpora tardíamente. Al igual que algunas otras herramientas de la arquitectura, ésta puede ser aplicada en conjunto con otros elementos para crear un esquema flexible de juegos que puedan soportar este esquema, por lo que un usuario podría integrarse a una sesión de juego en calidad de espectador o de un miembro activo, pero siempre con completa conciencia de lo que se ha venido desarrollando durante la ejecución del juego antes de que ingresara a la sesión.

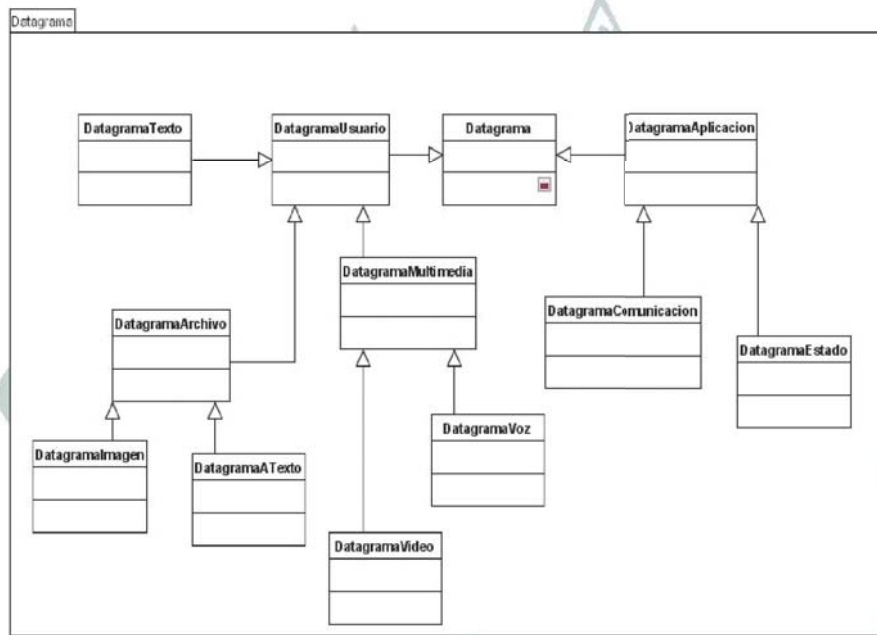
### **3.5 El manejo de mensajes y eventos.**

Como ya se ha mencionado, la colaboración forma parte fundamental para el desarrollo del aprendizaje de las matemáticas que los IIDM buscan enseñar a través de estas aplicaciones de tipo lúdico. Es por ello que la comunicación entre usuarios adquiere una importancia especial en nuestro proyecto. Siguiendo con la flexibilidad que cada una de las herramientas del FIIDM ofrece, se ha incorporado a la arquitectura una herramienta para el intercambio de mensajes entre jugadores mientras se lleva a cabo un juego. Todo esto con la finalidad de que la colaboración entre individuos esté sustentada por el Framework. El envío de mensajes se realiza a través del servidor de mensajería Jabber, el cual define una familia de protocolos de comunicación basados en mensajes empleando XML.

Esta comunicación puede darse en varios niveles, ya sea que se establezca entre dos individuos o que se de entre un grupo más grande de personas. Es el diseñador el encargado de definir si la comunicación se puede dar solamente entre miembros de un mismo equipo o no.

Para la construcción de un IIDM es necesaria la comunicación entre instancias que interactúan en conjunto. Esto fue posible siguiendo el mismo esquema de paso de mensajes que se usa para la comunicación entre jugadores. Esta comunicación se realizó mediante la construcción de pequeños datagramas (*Fig. 2*) de uso interno contruidos en XML, los cuales pueden ser fácilmente analizados a través de un parser. Entre otras cosas esta herramienta nos es útil para programar el nivel de conciencia de colaboración que deseamos tener dentro de nuestras aplicaciones. Empleamos el mismo servicio de mensajería Jabber para el envío de datos y eventos entre aplicaciones. Se incluyó en cada aplicación un analizador de protocolo de

eventos específicos para los IIDM. Gracias a este mecanismo se simplificó la comunicación entre las instancias distribuidas de un juego.



**Fig 2. Diagrama de clases jerarquía de Datagramas para Comunicación interna**

### 3.6 El manejo de puntuación en los juegos

En un momento dado, es posible tener activos una variedad de esquemas de juegos cada uno empleando diferentes esquemas de puntuación. Incluso, un mismo tipo de juego puede implementar varios esquemas de puntuación dependiendo del contexto en el que se ejecute, como por ejemplo en el caso en que éste forme parte de una serie o torneo, que el juego se ejecute por un grupo de usuarios o bien que un solo usuario interactúe con la computadora. Por ello el FIIDM cuenta con un controlador de esquemas de puntuación, el cual se encarga de definir distintos esquemas del manejo de la puntuación como en el caso en que los puntos se incrementen por una o varias unidades, o bien que los puntos se asignen a un individuo o a un equipo sin que se actualice la el tablero gráfico.

### **3.7 El manejo de turnos en los juegos.**

Al igual que en el caso de manejo de puntos, la necesidad de un controlador de turnos se vuelve inminente para lograr el control de los diversos esquemas que se pueden presentar en las aplicaciones construidas sobre el FIIDM. Esto es, para las diferentes etapas de un mismo juego se pueden manejar esquemas de turno distintos y a su vez un IIDM puede manejar un esquema totalmente distinto a otro.

Supongamos el caso en el que un juego varía su esquema dependiendo del nivel de complejidad de sus etapas y que en lugar de que los turnos sean secuenciales, en una de las etapas posteriores, el turno corresponda siempre a aquel jugador que cumpla con toda una serie de características previamente definidas en las reglas del juego.

### **3.8 El manejo de turnos en los juegos.**

El soporte para el almacenamiento del historial del juego, tiene un papel muy importante ya que se constituye como un complemento del soporte para usuarios de incorporación tardía a una sesión de juego: permite que un juego sea retomado por un usuario en una sesión posterior o que un usuario que quiere incorporarse a un juego pueda recibir una copia de la instancia sin perderse de la actividad desarrollada con anterioridad. Esta copia del juego es guardada en una base de datos centralizada de donde es recuperada posteriormente e incorporada al historial activo del juego, el cual está presente en tiempo de ejecución (run-time). Es también un elemento útil para la continuidad de series o torneos, pues es aquí donde se guarda la información referente a partidas anteriores. Un elemento importante, especialmente si se considera que los torneos pueden ser muy largos y definirse en periodos de tiempos prolongados.

## **4.- Conclusiones**

El uso del FIIDM tiene un prometedor potencial para crear IIDM de calidad. El diseño basado en componentes del FIIDM está fundamentado en los distintos administradores que forman su estructura principal: el administrador de antesalas, el administrador de despachadores de mensajes, el administrador de mensajes, el administrador de conexiones de bases de datos, el administrador de roles, el administrador de estructura de los juegos y el administrador de peticiones. En su conjunto, estos componentes ofrecen todas las características necesarias para crear juegos matemáticos multiusuarios cuyas características den soporte al trabajo colaborativo orientado al fomento para el aprendizaje de las matemáticas. El FIIDM no solo beneficia al desarrollador sino también al usuario, puesto que se logra la estandarización en las IIDM en cuanto a sus funciones y la calidad de sus servicios. Por otra parte el uso del FIIDM mantiene los lineamientos establecidos en el

proyecto, en lo que se refiere a la reutilización de código y la incorporación a los IIDM de los servicios que ofrece la infraestructura de soporte. Con el FIIDM, el desarrollador puede construir IIDM de una manera más sencilla a la vez que fomenta la reutilización de código mediante la adopción de interfaces y estructuras heredadas por clases definidas en el Framework. Actualmente continuamos con la codificación del FIIDM y se están realizando pruebas de implementación del Framework en diferentes juegos matemáticos recreativos.

## 5.- Referencias

- [1] Avison D. y Fitzgerald G. 1997. Information Systems Development Methodologies, Techniques and Tools. 2 ed. London: McGraw-Hill, 1997.
  
- [2] López-Morteo G. and López Mariscal G. 2003. An Electronic Ludic Learning Environment for Mathematics based on Learning Objects . Proceedings of ED-MEDIA 2003. World Conference on Educational Multimedia, Hypermedia and Telecommunications. Association for the Advancement of Computing in Education. 23 al 28 de junio del 2003, Honolulu, Hawaii.
  
- [3] López-Morteo G., López Mariscal G e Ibarra Esquer J. 2002. Enticing learners to mathematics through a ludic collaborative learning environment. Memorias del Segundo Congreso Internacional de Educación Abierta y a Distancia. 6 al 8 de junio del 2002. Colección Innovación Educativa y Nuevas Tecnologías, ANUIES
  
- [4] López-Morteo G., López Mariscal G. e Ibarra Esquer J.. 2001. Ambiente de Aprendizaje Basado en Instructores Interactivos de Diversiones Matemáticas . International Conference on New Technologies in Science Education CINTEC 2000, 4 al 6 de julio, Aveiro, Portugal
  
- [5] López-Morteo G. y López Mariscal G. 2000. Una concepción lúdica del software educativo para las matemáticas. Educación Abierta y a Distancia. Experiencias y Perspectivas. Primer Congreso Regional de Educación Abierta y a Distancia. 8 al 10 de junio 2000, Ensenada, Baja California, México. Coordinadores Carmen Perez Fragoso, Jesus Favela Vara y Gilberto Lopez Mariscal. 127-140.
  
- [6] Stevens P. y Pooley B., 2002. Utilización de UML en Ingeniería del Software con Objetos y Componentes Addison-Wisley
  
- [7] Zhang Y. y Weiss M., 2003. Virtual communities and team formation. Crossroads, 10(1), ACM Press, 13-21.

## Diagnóstico Automático de Usabilidad de Sitios Web

Fernando Sánchez Martínez, Oscar Mayora Ibarra

00379098@academ01.mor.itesm.mx, omayora@itesm.mx

ITESM, Campus Cuernavaca

Departamento de Ciencias de la Computación

Av. Paseo de la Reforma 182-A

Col. Lomas de Cuernavaca C.P. 62589

Temixco Morelos, MEXICO

**Abstrac.** Nowadays is recognized the importance that the usability has and its effect in the design of Web Sites, nevertheless, is those who do not consider, during the development of the site, to include a stage to evaluate the usability. On the other hand, previous studies have demonstrated that to improve a Web Site, undoubtedly, the best investment of time and performance consists of introducing improvements in its usability. This involves an increase in the time of work and therefore in costs, in general the use of technologies of usability has demonstrated an improvement in the final product. The present article proposes the development of a tool of automatic diagnosis of usability of Web sites, which it will serve to reducing the time of evaluation and consistently the costs.

**Resumen.** Actualmente se reconoce la importancia que tiene la usabilidad y su efecto en el diseño de Sitios Web, sin embargo, hay quienes no consideran, durante el desarrollo del sitio, incluir una etapa para evaluar la usabilidad. Por otra parte, estudios anteriores han demostrado que para mejorar un Sitio Web, sin duda, la mejor inversión de tiempo y rendimiento consiste en introducir mejoras en su usabilidad. Esto implica un incremento en el tiempo de trabajo y por lo tanto en costos, en general el uso de técnicas de usabilidad ha demostrado una mejoría en el producto final. El presente artículo propone el desarrollo de una herramienta de diagnóstico automático de usabilidad de Sitios Web, que servirá a reducir el tiempo de evaluación y consecuentemente los costos.

**Palabras clave.** Usabilidad, diagnóstico, diseño de sitios web.

### 1. Introducción

La usabilidad se refiere a la eficiencia y satisfacción con la que usuarios pueden alcanzar metas específicas en un sistema computacional determinado. Se define la ingeniería de usabilidad como el proceso para definir, medir y sucesivamente mejorar, el uso de los productos [1]. Estas técnicas se han desarrollado debido a la necesidad de trasladar la usabilidad desde el dominio de la opinión personal a un atributo que sea cuantificable como otros atributos de la ingeniería. Debido al gran auge que ha tenido en el ámbito mundial el uso y navegación por Internet y la economía electrónica, los diseñadores de Sitios Web y quienes los contratan para la construcción de sus portales, se han preocupado por las razones que implican que un usuario regrese o no al sitio; pues en el comercio electrónico, el usuario primero se enfrenta a la usabilidad y después hace el pago. La experiencia ha enseñado a los diseñadores que sitios prometedores de gran éxito por el servicio que ofrecen, se han visto opacados debido a la impresión que causan a los usuarios, ya que la calidad es afectada o fortalecida directamente por el diseño y

usabilidad del Sitio Web. La experiencia del usuario durante la navegación, la facilidad de entender la distribución de las páginas, el tiempo de descarga, la cantidad de imágenes o texto, los multimedia e incluso la disponibilidad de ciertos links, influye en la disposición o deseo de los usuarios a regresar a un determinado sitio.

En los últimos diez años se han realizado una gran cantidad de estudios de usabilidad del Web y durante este tiempo muchas cosas han cambiado. Donde Nielsen entre otros, se ha dado a la tarea de reexaminar el estado del Web distinguiendo varios problemas comunes de diseño [2]. Para evitar los problemas de usabilidad en el diseño de interfaces se considera llevar a cabo una evaluación como una etapa más en el desarrollo de una aplicación o un Sitio Web. La evaluación o diagnóstico de usabilidad consiste en las metodologías que miden los aspectos de usabilidad de una interfaz utilizada por un sistema e identificar problemas específicos; siendo una parte importante del proceso total del diseño de la interfaz de usuario, que consiste en ciclos iterativos de diseñar, de prototipos y de la evaluación.

El hecho de asegurar la usabilidad como parte del proceso de diseño, es una excelente práctica, ya que en cuando más rápido es detectado un problema es más fácil su solución y cuesta menos resolverlo que en etapas posteriores. Por ello, contar con una herramienta que de manera automática haga un diagnóstico preliminar y le informe al equipo de trabajo cuáles son las páginas que no presentan problemas, cuales otras presentan problemas de sobrecarga de texto, animaciones o imágenes, entre otro tipo de información; es un ahorro considerable en tiempo, esfuerzo y costos.

## **2. Estado del Arte**

Se ha propuesto una amplia gama de técnicas de evaluación y un subconjunto de estas son actualmente de uso común. Algunas técnicas de evaluación, como las pruebas formales de usuarios, sólo pueden ser aplicadas después de que el diseño de la interfaz se ha desarrollado. Otras, como la evaluación heurística, puede ser aplicada en las primeras etapas de diseño. Cada técnica tiene sus propias exigencias, y generalmente diversas técnicas presentan diferentes problemas de usabilidad. Cuando se hace una evaluación de usabilidad en una interfaz, las conclusiones de diferentes evaluadores pueden variar extensamente, incluso si usan la misma técnica de evaluación. Por lo que los expertos recomiendan que a la misma interfaz se le apliquen varias técnicas de usabilidad y así detectar todas las posibles situaciones críticas. Muchos métodos de evaluación de usabilidad requieren de la grabación de las acciones que un usuario ejecuta mientras interactúa con una interfaz. Esto puede ser hecho por un evaluador que toma apuntes mientras el participante usa el sistema en vivo o viendo repetidamente una cinta de vídeo de la sesión: ambas son actividades que llevan mucho tiempo. Como las técnicas de captura alternativas, automatizadas pueden registrar la actividad del usuario automáticamente. Una distinción importante puede ser hecha entre la información que es fácil para registrar, pero difícil de interpretar (p.ej., keystrokes) y la información que es significativa, pero difícil de etiquetar automáticamente, como la terminación de una tarea.

En el ámbito de la evaluación de usabilidad en Sitios Web se han registrado muy ricas aportaciones que han permitido a los diseñadores obtener mejores resultados. Uno de los investigadores que mayores avances a logrado en el campo de la usabilidad es Jakob Nielsen [3], quien ha publicado una gran cantidad de documentos relacionados con el tema enfocándose no solo a los Sitios Web sino también a las aplicaciones



computacionales en general. Para realizar la propuesta de este proyecto de investigación se han analizado las diferentes herramientas de diagnóstico de automático de usabilidad de Sitios Web que están disponibles en el mercado, entre los que destacaron:

- WebSat [4]
- UsableNet's Lift Online y Lift Onsite [5]
- WebCriteria [6]
- NetRaker Suite [7]
- Ergosoft – ErgoBrowser [8]
- A – Prompt [9]
- Bobby [10]
- Doctor Html [11]
- LinkBot [12]
- NetMechanich [13]
- WebGarage [14]

En 1999, el National Institute of Standards and Technology (NIST) anunció el lanzamiento de *WebSat* (Web Static Analyzer Tool) que tiene como tarea principal identificar problemas potenciales de usabilidad, comprobando el código HTML. Es un software gratuito y que da a los usuarios la posibilidad de utilizarlo en línea o como aplicación. Evalúa si la página hace uso de etiquetas para usuarios visualmente discapacitados. Checa la presencia de botones Submit and Reset. Analiza el tamaño y codificación de los gráficos en relación con la velocidad de descarga de la página. Comprueba los tags y el código que haría más fácil migrar la página a otro sitio de navegación. Hace una comprobación básica de cómo los links son codificados. Su mayor limitación es que no desarrolla la evaluación en el sitio completo sino solo por página. Posteriormente, apareció *Lift Online* y *Lift Onsite*, teniendo como valor agregado a la evaluación una serie de recomendaciones de cambios específicos en el código. Su limitante es también que el diagnóstico se desarrolla por página.

*WebCriteria* utiliza un agente inteligente que rastrea el sitio, simulando la experiencia de un usuario, analizando la accesabilidad, el tiempo de descarga y el contenido. En accesabilidad evalúa la velocidad con la que un usuario puede encontrar algo sobre el sitio. El primer factor de accesabilidad es el tiempo de exploración del link que se basa en la premisa de que los links colocados más arriba son los más accesibles. El segundo factor es el browse path, que hace un promedio del menor número de clicks entre la página de inicio a cualquier página en particular. La aplicación resume el porcentaje de diferentes medios (texto, gráficos, multimedios) y tecnologías como PDF, flash, etc.

*NetRaker Suite* es una herramienta de suscripción mensual que proporciona al usuario una versión limitada de una prueba de usabilidad. Su ventaja principal es que provee la retroalimentación de una tarea específica en oposición a las comprobaciones de código.

Ergosoft desarrolló un navegador Web llamado *ergoBrowser* destinado a facilitar el trabajo a los profesionales de usabilidad, ya que facilita la evaluación al grabar el tiempo ocupado en cada página y elaborando un registro completo de teclazos, el número de clicks en las páginas y scrollbars. Su limitante es que no trabaja en redes locales ni con páginas que utilicen frames. Otro desarrollo realizado en la Universidad de Toronto produjo *A-Prompt*, que es una herramienta que permite desarrollar páginas Web accesibles a minusválidos. Localiza barreras de accesabilidad, realiza reparaciones cuando es posible y asiste al evaluador en las reparaciones manuales.

*Bobby* es un servicio online o software offline que analiza páginas Web y ayuda a reparar las barreras de accesabilidad, promoviendo el uso de directrices.

Por su parte, *Doctor HTML*, es una de las herramientas más completas disponibles en el mercado para el diagnóstico de usabilidad. Fue lanzado en 1997 y permite hacer un diagnóstico del sitio completo desde Internet o en la Intranet. Cada uno de los errores identificados por la aplicación son acompañados de un comentario descriptivo. Funciona con código HTML abierto, verifica la presencia de los tags, localiza ligas muertas, la sintaxis de las imágenes, valida la estructura de los forms, da formato al código HTML, verifica la ortografía del contenido. Además, elimina los espacios innecesarios, verifica frames, evalúa la compatibilidad con los navegadores más comunes, el uso de fuentes estándares. Muestra la página que está siendo evaluada y si se requiere de un nombre de usuario y password puedes ser agregados.

*LinkBot*, evalúa el sitio, generando el árbol de navegación y muestra una lista de los links rotos, archivos huérfanos, páginas lentas y páginas que omiten los títulos o ALT.

*NetMechanic* es una aplicación que analiza el código HTML para encontrar errores y puede corregirlos automáticamente.

Finalmente, *WebGarage* que valida los links, el tiempo de descarga, la compatibilidad con los navegadores y valida el código HTML.

Para desarrollar el análisis de las diversas herramientas de diagnóstico se definió una taxonomía en base a los elementos que se verifican en cada una de ellas y posteriormente se presenta una tabla con los resultados. Los elementos de evaluación son los siguientes:

- Identificación de errores. Consiste en la localización e identificación de instrucciones en HTML que se expresan de manera incorrecta en alguna página del sitio web.
- Identificación de etiquetas para discapacitados. Localización e identificación de etiquetas útiles para discapacitados visuales.
- Tamaño de gráficos. Contenido en la página de gráficos. Expresado en bytes.
- Tamaño de texto. Cantidad de texto incluido en la página.
- Tamaño de multimedios. Memoria ocupada en multimedios.
- Evaluación de migración. Se verifica si el sitio puede migrar de plataforma.
- Comprobación de links. Se verifica si no existen ligas muertas.
- Recomendaciones de cambios en el código. Si la herramienta presenta recomendaciones al diseñador acerca de cómo mejorar el código html.
- Tiempo de descarga. Si se incluye un análisis del tiempo de descarga de cada página.
- Corrección de errores en el código html automáticamente.
- Asistencia en reparación manuales del código.
- Compatibilidad de navegadores. Análisis del sitio con respecto a diferentes navegadores de Internet.
- Ortografía del contenido. Si la herramienta evalúa la ortografía del texto contenido en el sitio.
- Generar árbol de navegación. Si se muestra al usuario el árbol de navegación del sitio completo.
- Desarrollo de tareas en el sitio. Genera un reporte que indica el tiempo que tarda un usuario en realizar o llegar a un punto específico dentro del sitio; incluye clicks, teclazos, scroll bar, etc.

- Niveles de profundidad. Se evalúa el número de niveles de profundidad horizontal y vertical del sitio.
- Servicio fuera de línea. Si la herramienta puede usarse como aplicación local.
- Servicio en línea. Si esta disponible vía web.
- Evaluación de una sola página. Si la herramienta evalúa solo por página
- Evaluación del Sitio completo. Caso contrario al anterior.
- Costo. Gratuito, pago por páginas evaluadas, por mensualidad o por licencias.

**Tabla 1. Características que identifican a las herramientas de diagnóstico**

Características	WS	ULO	WC	NS	E	AP	B	DH	LB	NM	WG
Identificación de errores HTML	X	X					X	X		X	X
Identificación de etiquetas para discapacitados	X					X					
Tamaño de gráficos	X		X					X			
Tamaño de texto			X					X			
Tamaño de multimedios			X					X			
Evalúa la migración	X										
Comprobación de links (ligas muertas)	X	X	X				X	X	X		X
Recomendaciones de cambios en el código		X					X	X			
Tiempo de descarga		X	X					X	X		X
Corrección de errores en el código HTML automáticamente						X				X	
Asistencia en reparaciones						X					
Compatibilidad con navegadores								X			X
Ortografía del contenido								X			
Genera árbol de navegación									X		
Desarrollo de tareas en el sitio (Tiempo de exploración)			X	X	X						
Inconsistencias en etiquetas								X			
Servicio fuera de línea	X						X	X			
Servicio en línea	X			X	X		X	X			X
Evaluación de una sola página	X	X					X				
Evaluación del Sitio completo					X			X	X		X
Costo	Gratis	Páginas		Mensual	Licencias	Gratis	Gratis	Licencias	Licencias	Páginas	Limitado

WS	WebSat	E	Ergosoft–	LB	LinkBot
ULO	UsableNet’s Lift Online y Lift Onsite	AP	ErgoBrowser A – Prompt	NM	NetMechanich
WC	WebCriteria	B	Bobby	WG	WebGarage
NS	NetRaker Suite	DH	Doctor Html		

En el análisis de las ventajas y limitaciones de cada una de las herramientas de diagnóstico de usabilidad que están disponibles en el mercado se puede identificar que en su mayoría están dedicadas al análisis del código HTML y a identificar errores de codificación, únicamente; en este caso se encuentran WebSAT, Lift Online y Lift Onsite, NetRaker Suite, Bobby, Doctor HTML, NetMechanic y WebGarage. Estas herramientas son de un gran apoyo para aquellos que están codificando las páginas debido al enfoque que tienen; siendo de estas, solo Doctor HTML y WebGarage los que evalúan un sitio completo.

Por otra parte, están la herramientas que facilitan las evaluaciones de usabilidad, contando el número de teclazos o clicks en los links del sitio, simulando la experiencia de los usuarios y midiendo el tiempo de descarga, tal como WebCriteria, ergoBrowser y LinkBot. En esta otra categoría, las tres herramientas se enfocan al aspectos funcionalidad del sitio, evaluando y grabando los movimientos de los usuarios dentro del sitio web cuando se les solicita realizar una tarea específicamente. No se consideran aspectos de codificación, ni el espacio ocupado en texto, imágenes o multimedios. Estas herramientas están destinadas al proceso de evaluación del sitio ante los usuarios, y no en el tiempo de codificación o diseño.

Finalmente A-Prompt que es una herramienta que guía en el desarrollo de páginas Web, identificando y corrigiendo errores automáticamente o asistiendo al desarrollador en la reparación manual. Es una herramienta dirigida a los codificadores de páginas y no a los diseñadores directamente, pues no evalúa aspectos de espacio, tiempo de descarga, ni la experiencia de los usuarios ante el sitio.

### 3. DAU: Herramienta de Diagnóstico Automático de Usabilidad

Desarrollar la evaluación de usabilidad de un Sitio en busca de fuentes potenciales de conflicto requiere de muchas horas de trabajo para poder recorrerlo completamente e ir evaluando las diferentes directrices, página por página dentro del mismo dominio. Si a esto sumamos que puede tratarse de un Sitio Web de gran tamaño — empresas internacionales, instituciones gubernamentales, instituciones educativas, comercios electrónicos, etc.— es un trabajo que requiere de un gran equipo de trabajo, que deben dedicar muchas horas en revisar el sitio e ir haciendo anotaciones con respecto a las variables de usabilidad.

La automatización de la evaluación de usabilidad tiene varias ventajas potenciales sobre la evaluación no automatizada, entre las que están:

- Reducir el costo de evaluación de usabilidad.
- Consistencia de los errores localizados.
- Detectar errores en las primeras etapas de trabajo.
- Reducir la necesidad de experiencia entre evaluadores individuales.
- Aumento de la cobertura de parámetros evaluados.
- La incorporación de la evaluación dentro de la fase de diseño de desarrollo, a diferencia de la aplicación después implantado.

Después del análisis desarrollado a las diversas herramientas que existen en el mercado se puede determinar que cada una de ellas se enfoca únicamente a una o dos de las etapas de desarrollo del sitio web, es decir, en el momento de codificación o ya cuando está implementado, en la experiencia de los usuarios. Es de suma importancia, no solo considerar estas etapas sino también los problemas que son difíciles de detectar, como son las ligas muertas, la detección de una navegación usable, los niveles de profundidad y los tiempos de navegación debido al tamaño del contenido de cada página del sitio.

Es necesaria una herramienta que permita a los diseñadores no solo evaluar código y accesibilidad sino también que haga un diagnóstico preliminar que informe al equipo de trabajo cuáles son las páginas con

problemas de sobrecarga, ligas no existentes, páginas huérfanas, información duplicada, niveles de profundidad, y cuales están dentro de los parámetros de usabilidad.

La experiencia en el desarrollo de aplicaciones nos ha llevado a aprender que entre más tiempo pase en detectarse un error, más costoso puede ser solucionarlo. Si se habla de desarrollo de sitios web, detectar páginas con demasiada carga de información puede poner en riesgo el sitio al enfrentarse al aburrimiento o enfado de los usuarios. Muchos de los errores de este tipo ni siquiera son detectados. Es por ello que llevar un diagnóstico con estas características a las primeras etapas de desarrollo permitirá a cualquier equipo de trabajo detectar errores o defectos en el diseño o distribución de la información a tiempo; reduciendo de esta manera los posibles incrementos en tiempo de desarrollo y reconstrucción del sitio. Para la evaluación de este tipo de parámetros de usabilidad se propone el desarrollo de DAU, una herramienta de diagnóstico dirigida a los diseñadores de sitios web. DAU es una herramienta en proceso de construcción que de manera automática hará un diagnóstico preliminar y le informará al equipo de trabajo cuáles son las páginas con problemas y cuales están dentro de los parámetros de usabilidad. En la tabla 2 se podrán observar las características que presentará la aplicación y que podrá comparar con la taxonomía evaluada en las otras herramientas disponibles en el mercado.

**Tabla 2. Características que identificarán a DAU.**

Características	DAU
Identificación de errores HTML	X
Identificación de etiquetas para discapacitados	
Tamaño de gráficos	X
Tamaño de texto	X
Tamaño de multimedios	X
Evalúa la migración	
Comprobación de links (ligas muertas)	X
Recomendaciones de cambios en el código	
Tiempo de descarga	X
Corrección de errores en el código HTML automáticamente	
Asistencia en reparaciones manuales	
Compatibilidad con navegadores	
Ortografía del contenido	
Genera árbol de navegación	X
Desarrollo de tareas en el sitio (Tiempo de exploración)	
Información duplicada	X
Inconsistencias en etiquetas	
Niveles de profundidad	X
Servicio fuera de línea	X
Servicio en línea	X
Evaluación de una sola página	
Evaluación del Sitio completo	X
Costo	Gratis

El sistema DAU contará con las siguientes funcionalidades:

- Generar el árbol de navegación del Sitio Web dentro basado en el dominio. Este árbol se presentará gráficamente en la aplicación y el usuario tendrá la posibilidad de cerrar o abrir los niveles de profundidad del sitio web y ver los resultados de evaluación por nivel de profundidad o por página, lo que le permitirá discriminar los reportes de acuerdo a lo que le interese en ese momento.
- Con respecto a los parámetros de evaluación, DAU dará la posibilidad a los diseñadores seleccionar únicamente los parámetros que sean de su interés. Los parámetros son:
  - o Ligas muertas
  - o Información duplicada
  - o Posibles inconsistencias de etiquetas
  - o Detección de navegación usable
  - o Detección de riesgo de exceso de texto en una página determinada
  - o Detección de riesgo de exceso de gráficos y animaciones en una página determinada

- Detección de contenidos y excesos
  - Detección de exceso de niveles de profundidad horizontal y vertical
  - Identificación de tiempos óptimos de navegación en los distintos niveles del sitio
  - Detección de problemas en el código HTML
  - Páginas correctas en usabilidad
- Emitir reportes de usabilidad con cierto grado de flexibilidad. La flexibilidad se especificará en la selección de los parámetros evaluados y en el nivel del árbol que se desee analizar. Estos reportes podrán verse en pantalla o si el usuario lo desea imprimirlos.

Otra característica de DAU, será la de evitar que el diseñador deba recorrer el sitio completo de manera manual e ir haciendo la evaluación midiendo el tiempo de descarga, el contenido de las páginas, etc. Al contar con la herramienta, el tiempo de diagnóstico y la consistencia de criterios de análisis reducirá el tiempo de evaluación y a su vez ajustará el sitio web a ciertos estándares predeterminados. Es importante hacer notar que se considera a las técnicas automatizadas como un complemento útil a la adición de técnicas de evaluación estándar, como la evaluación heurística y pruebas de usuario, no un sustituto. Diferentes técnicas localizan diferentes tipos de problemas, y las medidas subjetivas, como la satisfacción de usuario no es tan confiable si se hace con métodos automatizados.

#### 4. Conclusión

Asegurar la usabilidad como parte del proceso de diseño, es una excelente práctica, ya que entre más pronto es detectado un problema es más fácil su solución y cuesta menos resolverlo que en etapas posteriores. SE detectaron varias herramientas que llevan a cabo un cierto tipo de evaluación de usabilidad, algunos orientados específicamente en la codificación, otras en la experiencia del usuario y una última como guía de desarrollo de sitios web; pero ninguna que haga un diagnóstico preliminar que considere la presentación de un árbol de navegación y la emisión de reportes flexibles a los parámetros de evaluación y con la posibilidad de ampliar sus horizontes de evaluación. En el presente artículo, se propone el desarrollo de una herramienta que de manera automática haga un evaluación preliminar y le informe al equipo de diseño cuáles son las páginas que no presentan problemas, cuales otras tienen sobrecarga de texto, animaciones o imágenes, entre otros defectos; lo que implica un ahorro considerable en tiempo, esfuerzo y costos.

#### 5. Referencias

- [1] Denis Wixon y Chauncey Wilson
- [2] Jakob Nielsen's Website <http://www.useit.com/>
- [3] Jakob Nielsen. Designing Web Usability : The Practice of Simplicity
- [4] WebSat. <http://zing.ncsl.gov/webmet/sat/websat-process.html>
- [5] UsableNet's Lift Online y Lift Onsite. <http://www.usablenet.com>
- [6] WebCriteria. <http://www.webcriteria.com>
- [7] NetRaker Suite. <http://www.netraker.com>
- [8] Ergosoft – ErgoBrowser. <http://www.ergolabs.com>
- [9] A – Prompt. <http://aprompt.snow.utoronto.ca>
- [10] Bobby. <http://www.watchfire.com>
- [11] Doctor Html. <http://www2.imagiware.com/RXHTML>
- [12] LinkBot. <http://www.webdevelopersjournal.com/software/linkbot.html>
- [13] NetMechanic. <http://www.netmechanic.com/toolbox/html-code.htm>
- [14] WebGarage. <http://www.stetson.edu/administration/it/mediaservices/webgarage.php>
- [15] R. Molich, N. Bevan, S. Butler, I. Curson, E. Kindlund, J. Kirakowski, and D. Miller. Comparative evaluation of usability tests. In Proceedings of UPA98, pages 189-200, June 1998.

- [16] R. Molich, A. D. Thomsen, B. Karyukina, L. Schmidt, M. Ede, W. van Oel, and M. Arcuri. Comparative evaluation of usability tests. In Proceedings of ACM CHI'99 Conference on Human Factors in Computing Systems, pages 83-86, May 1999.
- [17] Jakob Nielsen. *Usability Engineering*. Academic Press, Boston, MA, 1993.
- [18] NIELSEN, J. 1993. *Usability Engineering*. Boston, MA: Academic Press.
- [19] MELODY Y. IVORY, MARTI A. HEARST. The State of the Art in Automating Usability Evaluation of User Interfaces. University of California, Berkeley

# Estudio sobre el uso de gestos en una interfaz multimodal hombre-máquina simulada

Luis Villaseñor Pineda & Claudia Vázquez Valerdi

Laboratorio de Tecnologías del Lenguaje  
Instituto Nacional de Astrofísica, Óptica y Electrónica  
Sta. María Tonantzintla, Puebla, México  
villasen@inaoep.mx

**Resumen.** La nueva generación de interfaces hombre-máquina son las llamadas interfaces multimodales. En ellas la comunicación con el sistema automático se realiza gracias a más de un modo de entrada, principalmente la voz y el gesto. El presente trabajo presenta un estudio empírico sobre el tipo de expresiones verbales que son acompañadas con indicaciones gestuales en una interfaz multimodal simulada. Este estudio describe y propone una clasificación del tipo de expresiones que los usuarios acompañaron de un gesto.

## 1 Introducción

Entre los proyectos realizados en el Laboratorio de Tecnologías del Lenguaje del INAOE se estudian los medios para extender la expresividad en la comunicación hombre-máquina al incluir varios modos de entrada. Lo que se busca es permitir una comunicación oral apoyada por otros modos de entrada como el ratón o el teclado.

Con este propósito se estudiaron video-grabaciones de sesiones de trabajo entre dos personas. El corpus utilizado en este estudio es el corpus *DIME* [Villaseñor et al. 01]. Este corpus multimodal provee información empírica sobre el uso y la interrelación en los intercambios conversacionales, además de brindar elementos para el estudio del lenguaje hablado, incluyendo fenómenos comunes como la anáfora, la elipsis, las expresiones deícticas, etc. La recolección del corpus se realizó mediante el escenario del Mago de Oz [Dahlbäck et al. 93]. Este escenario consiste de una persona (el mago) que toma el papel del sistema computacional y de una persona (el sujeto) a la cual se le solicita la resolución de una tarea con ayuda del mago. La interacción se lleva a cabo a través de un sistema de audio y computadoras que permiten observar la tarea realizada por el otro sin contacto visual entre ellos. La gráfica 1 muestra la plataforma usada para la adquisición del corpus DIME. El corpus consiste de 31 grabaciones con una duración total de 7 horas 10 minutos (en promedio 14 minutos por grabación).

Los datos aportados por este corpus nos han permitido analizar diferentes fenómenos en este tipo de interacción [Taylor et al 89]. En particular este estudio se enfoca al lenguaje no verbal, el cual funcionó para regular, interpretar y retroalimentar el proceso de comunicación. Por ejemplo en:



mago: ¿quieres que borre **este objeto**?  
sujeto: sí

El mago emitió la expresión *este objeto* acompañada de la señalización directa sobre el objeto al que se refiere, este *gesto* ayuda al sujeto a ubicar el objeto eliminando cualquier ambigüedad, ya que en este caso la expresión *este objeto* no tienen solución posible. Es decir, el gesto fue indispensable para establecer la correcta ubicación del objeto, además de regular e interpretar la comunicación.

El presente trabajo describe y clasifica las expresiones verbales acompañadas de gestos. A continuación se presenta brevemente los aspectos fundamentales del lenguaje no verbal; enseguida, se describen los detalles de la transcripción usada así como los diferentes tipos de uso de gestos identificados. Por último, en la cuarta sección se discuten las perspectivas de este trabajo.

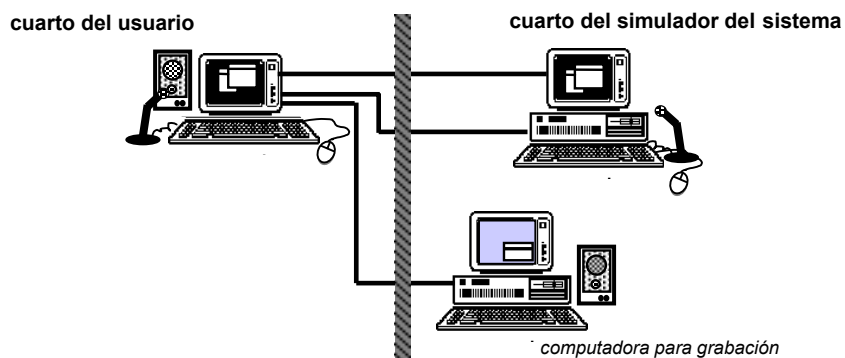


Fig. 1. Plataforma para la grabación del corpus DIME

## 2 El lenguaje no verbal

Diversos aspectos no verbales del comportamiento comunicativo son tan connaturales en las interacciones de la vida cotidiana que es difícil ser plenamente conscientes de sus funciones y significado. Entre las diferentes manifestaciones del lenguaje no verbal podemos encontrar los siguientes: el contacto corporal, la distancia interpersonal, la orientación, la postura, la mímica del rostro, el comportamiento visual, etc. (para una descripción completa véase [Vázquez 04]).

Entre las funciones discursivas del lenguaje no verbal tenemos:

1. Marca de manera consciente o inconsciente la sucesión gramatical de las palabras.
2. Resalta algún aspecto semántico del mensaje discursivo.
3. Señala algo en particular en el contexto referencial (gestos deícticos).
4. Muestra objetos, modos de actuar, estados de ánimo, reacciones, emociones, etc.

5. Tiene un gran contenido semántico en cuanto a la actitud e intención comunicativa de los sujetos en los proceso de interacción.
6. Puede oponerse al lenguaje verbal, ya que el lenguaje no verbal no puede mentir, es de difícil manipulación.

Es tan importante el lenguaje no verbal que no sólo ayuda a desmentir lo que se dice verbalmente sino además provee tanta información como para interpretar múltiples significados. En lo que respecta a este trabajo sólo nos enfocaremos a la función del lenguaje no verbal cuando este sirve para señalar algo en el contexto referencial. Esta función está íntimamente asociada con la *deixis*. En los párrafos siguientes se describe este fenómeno lingüístico.

## 2.1 La deixis

Como se mencionó anteriormente deseamos establecer las relaciones entre el lenguaje verbal y no verbal. De particular importancia para nosotros es lo que los lingüistas llaman *deixis*. Una expresión que es capaz de “mostrar” la situación de enunciación de un hablante en alguna medida es denominada frase o palabra deíctica. Es decir, los elementos deícticos tienen una función referencial, pero deben ser interpretados con respecto al hablante. Básicamente tenemos dos tipos de *deixis*<sup>1</sup>:

- *Deixis situacional*: cuyo punto de partida lo establece el hablante en el acto de la enunciación a partir de un conjunto de coordenadas actuales de persona, espacio y tiempo. Deixis mostrativa, con carácter egocéntrico de los pronombres personales, formas verbales, adverbios locativos.
- *Deixis contextual* (o *deixis “fórica”*: anáfora y catáfora): actúa en el interior del texto, refiriéndose a otras palabras que ya han aparecido o van a aparecer en él. Este tipo de *deixis* no tiene carácter egocéntrico, sino sólo aporta información acerca de la situación general de enunciación y del subjetivo punto de vista del hablante en ella.

En general, las expresiones que indican “persona”, “espacio” o “tiempo” no tienen más remedio que ser egocéntricas, pues incluso cuando la referencia respecto de la cual se toman no es el locutor, sino un punto externo, este punto no es “externo” respecto del locutor, que es el responsable del enunciado y el que por necesidad ha de expresar siempre en él un determinado punto de vista.

El caso de la *deixis* situacional podemos clasificarlos de la siguiente manera:

- De lugar (por ejemplo: aquí/acá, ahí, allí/allá, etc.)
- De tiempo (por ejemplo: ahora, mañana, etc.)

Los deícticos espaciales y temporales se decodifican, gracias a la fecha y lugar de enunciación dada. En nuestro estudio no tomaremos en cuenta los deícticos temporales, ya que en su mayoría son usados como muletillas, por ejemplo:

U: sí, **ahora**, el:: estante que está aquí arriba moverlo arriba del:: del refrigerador (ejemplo del diálogo 3, tarea1- turno 6).

Lo que es más, en el contexto del corpus DIME, el sistema computacional no está siendo programado para funciones futuras sino presentes.

---

<sup>1</sup> <http://www.ucm.es/info/circulo/no1/vigara.htm>

## 2.2 El pronombre como deíctico

Otro elemento lingüístico asociado al lenguaje no verbal es el pronombre. Esta relación es discutible, pues podemos ver un pronombre cumpliendo una función anafórica al reemplazar un término anterior y no un deíctico al no “mostrar” un objeto vinculado a la situación de enunciación. Sin embargo, en nuestro trabajo recuperamos muestras en las que se utilizan, por ejemplo pronombres demostrativos, sin vinculación anafórica o catafórica. Además, los pronombres implican un gesto que designa al objeto al mismo tiempo que se pronuncia la instancia del término. La relación entre los pronombres y los deícticos la establece también el Esbozo de la Real Academia de la Lengua Española (1978:202):

“Hay una propiedad común a todos los pronombres que no es morfológica ni propiamente sintáctica, aunque tenga consecuencias de orden sintáctico. Son nulos o escasos los contenidos semánticos del pronombre. No sabemos lo que significa el pronombre *esto*, por ejemplo, pero sí para lo que sirve: para **señalar** a algo que está ante nuestros ojos o algo que acabamos de pronunciar o escribir, sin mentar su concepto. A estas dos funciones aludía la Gramática griega al decir que todo pronombre es deíctico (señala inconceptualmente a lo que vemos o recordamos) o anafórico (remite a lo que se acaba de enunciar). El pronombre *yo* señala deícticamente al que está diciendo *yo*, el relativo *que* remite anafóricamente a su antecedente en el contexto.”

## 3 Transcripción y análisis de los diálogos

El primer paso para el estudio del lenguaje no verbal usado en la interacción hombre-máquina fue la transcripción de toda situación en la que un gesto fue usado. Esta anotación nos permitió, de manera explícita, establecer las relaciones entre el lenguaje verbal y no verbal con respecto a su función deíctica.

Con la transcripción empieza el análisis. En ella se detallan los silencios, tono de la voz y tartamudeos. Naturalmente la transcripción siempre se revisó y adecuó a las necesidades que se descubrieron en el proceso de descripción y análisis. Para el detalle de las convenciones de la transcripción véase [Tusón, 1997].

A partir de esta transcripción se identificaron tres grandes tipos de expresiones lingüísticas acompañadas de gesto. Las siguientes tablas presentan estos tres tipos y describen sus características. Del conjunto total de diálogos se tomaron únicamente 4 diálogos.

En la Tabla 1 encontramos el caso más común de uso coordinado de lenguaje verbal y no verbal: los *deícticos gramaticales* (éste, ése, aquel, ahí, allí, acá, etc.). Para describirlos se usaron diferentes características: (i) el deíctico gramatical usado por el sujeto (columna 1); (ii) la presencia de un gesto (1 o 2)<sup>2</sup> o no (0) (columna 2);

---

<sup>2</sup> En cuanto a la manifestación gestual, o selección, tomamos dos variantes: el (1) indica que el gesto acompaña al deíctico de forma directa con el cursor, el (2) hace referencia a que la selección es indirecta, es decir, que ya estaba seleccionada la ventana o área a la que se le hace referencia, o bien, el cursor ya estaba colocado en el objeto o lugar indicado, antes o durante la emisión de la expresión deíctica.

(iii) el referente asociado (columna 3); (iv) se indica si el deíctico se resuelve por texto o no(0), ya sea por anáfora (1) o catáfora (2) (columna 4); y (v) el tipo de deíctico, ya sea asociado a un objeto (1) o a un lugar (2). Por supuesto, también se tiene información para determinar en que diálogo y en que intervención aparece el deíctico.

Deíctico	Acompañada por un gesto	Referente	Posible resolución por texto	Tipo
ésta	1	ventana superior izquierda	1	1
ésta	1	ventana superior derecha	1	1
ésta	1	ventana inferior	1	1
éste	1	alacena	0	1
éste	1	campaña extractora	0	1
Éste	1	alacena	0	1
aquí	1	esquina inferior izquierda	0	2
Éste	1	gabinete de piso, blanco, doble	2	1

**Table 1.** Ejemplos de deícticos gramaticales presentes en el corpus DIME

Expresión referencial	Acompañada por gesto	Referente	Posible resolución por texto	Tipo
este mueblecito	1	campaña extractora blanca	2	1
este mueble	1	gabinete de piso blanco doble	2	1
este mueble	1	gabinete de piso blanco doble	1	1
esta esquina	1	esquina superior izquierda	0	2
este mueble	1	gabinete de piso blanco doble	1	2
esta estufa	1	estufa blanca	0	1
aquí encima	1	arriba del gabinete de piso blanco doble	0	2

**Tabla 2.** Ejemplos de expresiones referenciales presentes en el corpus DIME.

La Tabla 2 muestra expresiones que cumplen un papel de mostrar el referente (la tabla también está compuesta de cinco columnas, las mismas de la Tabla 1). Estas *expresiones referenciales* no son propiamente deícticos gramaticales. Pero consideramos que la información que se explicita por medio del lenguaje oral no es suficiente para la resolución de la orden, y como nos estamos enfocando a un ambiente computacional, es indispensable canalizar este tipo de expresiones y darles solución referencial. Por ejemplo, tenemos la expresión *este mueble*, la cual no es clara si no nos encontramos en el contexto; sabemos que la tarea es realizar un diseño de cocina en el cual están involucrados muchos muebles, y si escuchamos esta expresión, nos es imposible saber exactamente a qué mueble se refiere el usuario. Así como esta expresión tenemos otras similares, en cuanto a que sólo la expresión verbal no se basta a sí misma para explicar el referente. Igualmente, la computadora, no

sabr  a qu  mueble se refiere el usuario sin ayuda del contexto o gesto asociado. Tradicionalmente, este tipo de expresiones se resolver an como an foras (o cat foras), sin embargo, gracias a nuestro estudio podemos constatar que los usuarios usan de manera natural el lenguaje verbal y no verbal en estas condiciones. Aprovechando todos los medios a su alcance para asegurar la correcta interpretaci n de su intervenci n. Por supuesto, esto tiene una implicaci n pr ctica en la construcci n de una interfaz hombre-m quina, pues desde el punto de vista computacional, ser  m s f cil descifrar la orden aprovechando la informaci n gestual.

En la Tabla 3 se han tomado en cuenta expresiones que s  se resuelven por el texto: “gabinete de piso blanco doble”. Sin embargo, los usuarios acompa an la expresi n verbal con una indicaci n gestual. Siempre con la intenci n de hacer llegar su mensaje de la manera m s clara posible. Nuevamente, desde el punto de vista computacional, aprovechar este comportamiento por parte de los usuarios har  m s f cil de descifrar el mensaje. La tabla 3 describe estas *expresiones definidas* a trav s de la siguiente informaci n: (i) la expresi n “definida” que hacen referencia a un objeto o lugar (columna 1); (ii) indicamos si la expresi n est  acompa ada de gesto (1 o 2) o no (0); (iii) el tipo de referente asociado, ya sea un objeto o un lugar (columna 3).

Expresi�n <i>definida</i> de un objeto o lugar	Acompa�ada por gesto	Tipo
gabinete de piso blanco doble	2	1
este gabinete de piso, blanco, doble	2	1
la campana	1	1
la estufa	1	1
debajo de la ventana	0	2
mueblecito este, blanco o armario	1	1
al mismo nivel que la campana	0	2
los refrigeradores	0	1

**Tabla 3.** Ejemplos de expresiones definidas presentes en el corpus DIME.

En la Tabla 3 hay expresiones tan cortas como *el refrigerador*, consideramos que contextualmente, sint ctica y gramaticalmente el usuario emite este tipo de expresiones con el art culo definido masculino *el* para dirigirse a un s lo refrigerador que hay en el contexto, y no es necesario para el simulador del sistema que el usuario especifique a qu  refrigerador se refiere, ya que es el  nico. Tambi n podemos encontrar expresiones largas y con demostrativos (a n sin ser necesarios para el tipo de oraci n formulada) como *estos tres son los modelos de refrigeradores*, est n ubicadas en esta tabla porque son expresiones que se resuelven por s  mismas, adem s, entendemos el referente.

Despu s identificar y anotar estos tres tipos de expresiones ling sticas asociadas a gestos se obtuvieron los siguientes resultados. La Tabla 4 resume la informaci n para los deicticos gramaticales. Como puede observarse obtuvimos un total de 95 deicticos, de los cuales 92 de ellos tuvieron un gesto asociado, es decir el 97%, lo que representa un alto grado de dependencia del gesto al emitir este tipo de expresiones, lo

cual concuerda con lo esperado. Sin embargo, nos topamos con tres casos que no tienen un gesto asociado. Esto se debió a que el deíctico fue resuelto por el texto (anáfora) ya que se presenta su referente con anterioridad. A continuación se presenta el fragmento en que se da esta situación:

- T 62- U: ¿me puedes poner/ la mesa con cajones junto a la estufa del lado derecho?  
 T 63- S: ¿del lado derecho de la estufa?  
 T 64- U: sí, aquí  
 T 65- S: ok, (trabajando) el mueble no cabe **ahí**

El deíctico gramatical *aquí* sustituye al *lado derecho de la estufa*, dado que fue señalado el lugar pocos segundos antes, posiblemente el interlocutor consideró obvio el referente y no volvió a especificarlo con un gesto.

	Número de deícticos	Porcentajes con respecto al total de deícticos
Total de deícticos con gesto asociado	92	97%
-- cursor (directa)	87	
-- objeto seleccionado (indirecta)	5	
Total de deícticos sin gesto asociado	3	3%
Total de deícticos designando un objeto	49	52%
Total de deícticos designando un lugar	46	48%
Total de deícticos con posible resolución por texto	24	25%
-- con posible resolución por anáfora	19	20%
-- con posible resolución por catáfora	5	5%
<b>Total de deícticos gramaticales</b>	<b>95</b>	

**Tabla 4.** Resumen sobre los deícticos gramaticales encontrados en el corpus DIME.

La mayoría de los deícticos en esta tabla, 87, tuvieron el gesto asociado de manera *directa*, o sea, que en el momento de emitir la expresión, el usuario hizo al mismo tiempo la indicación con el cursor. Sólo 5 de ellos tuvieron el gesto asociado de manera *indirecta*: como ya vimos antes, el referente del objeto o lugar ya estaba seleccionado o el cursor se mantuvo en el referente durante la emisión de la expresión.

Es relevante notar que la utilización de los deícticos fueron apenas más abundantes para indicar un objeto (49 = 52%) que un lugar (46 = 48%). En cuanto a la posible resolución de los deícticos a través del discurso tenemos un 24 (25%), donde 19 de ellos se pueden resolver por anáfora (20%), y 5 por catáfora (5%). De los que se pueden resolver por anáfora 2 no tuvieron gesto asociado, 15 con gesto directo, y el resto (2) con indirecto.

La Tabla 5 muestra los resultados obtenidos para las expresiones referenciales. Éstas fueron un total de 121, 96 de ellas tuvieron un gesto asociado (79%), de las cuales 86 fueron con señalización directa y 10 indirecta. De estas expresiones podemos encontrar un total de 74 (61%) para designar un objeto y 47 (39%) para un lugar; caso contrastante con la Tabla 4 que mostraba una equivalencia en la

designación de cualquiera de estos dos elementos, en este caso se presenta una preferencia por las expresiones que indican un objeto. Muy probablemente por ser más fácil nombrar un objeto que un lugar. Sólo el 43 (36%) expresiones pueden tener una posible resolución por el texto, 30 de ellas por anáfora (25%) y 13 por catáfora (11%).

	Número de expresiones	Porcentajes con respecto al total de expresiones
Total de expresiones con gesto asociado	96	79%
-- cursor (directa)	86	
-- objeto seleccionado (indirecta)	10	
Total de expresiones sin gesto asociado	25	21%
Total de expresiones designando un objeto	74	61%
Total de expresiones designando un lugar	47	39%
Total de expresiones con posible resolución por texto	43	36%
- con posible resolución por anáfora	30	25%
- con posible resolución por catáfora	13	11%
<b>Total de expresiones referenciales</b>	121	

**Tabla 5.** Resumen sobre las expresiones referenciales encontradas en el corpus DIME.

Por último, en la Tabla 6 se presentan los resultados de las expresiones definidas. Se obtuvieron 247 expresiones, número mucho mayor en comparación con las otras dos tablas. 123 expresiones tienen un gesto que las acompaña lo que representa el 50%, 102 tienen el gesto asociado de manera directa (41%) y 22 indirecta (9%). Del total de expresiones 182 expresiones fueron empleadas para designar un objeto (73%) y 66 (27%) para designar un lugar.

	Número de expresiones	Porcentajes con respecto al total de expresiones
Total de expresiones designando un objeto	181	73%
Total de expresiones designando un lugar	66	27%
Total de expresiones con gesto	123	50%
Total de expresiones con gesto (directa)	102	41%
Total de expresiones con gesto (indirecta)	22	9%
<b>Total de expresiones definidas</b>	247	

**Tabla 6.** Resumen sobre las expresiones definidas encontradas en el corpus DIME.

## 4 Perspectivas

El presente estudio es parte del esfuerzo emprendido para determinar claramente cuales son los fenómenos propios de la comunicación oral humana que deben estar presentes en la comunicación oral hombre-máquina para que ésta sea realmente útil.

Este trabajo estudia las implicaciones del uso de gestos en la comunicación hombre-máquina. Es evidente que el uso del lenguaje no verbal como modo de entrada en la comunicación hombre-máquina es natural para el usuario y que mejora la eficacia de esta comunicación. Cabe recordar que el objetivo primordial de la investigación en las interfaces hombre-máquina es el desarrollo de *modos* de comunicación entre el usuario y la computadora que sean tolerantes a errores y de fácil aprendizaje.

El uso de gestos en la comunicación hombre-máquina es de gran valía por su enorme carga comunicacional. Si recordamos nuestro caso de estudio, las dos personas se comunicaron sin establecer contacto visual directo. Este contacto visual es de enorme importancia para confirmar la recepción (y entendimiento) del mensaje emitido. Bajo estas condiciones, el emisor del mensaje cambia su estrategia y trata, con todos los medios a su alcance, de ser lo más claro posible para que el mensaje llegue con éxito a su interlocutor. Esta es una probable explicación a la alta frecuencia de uso de gestos (50%) en las expresiones definidas. Expresiones que describen verbalmente un lugar u objeto sin ambigüedad alguna. Es decir, que aún “sin la necesidad” de reforzar la comunicación, los usuarios consideraron indispensable acompañarlas con un gesto seguramente para precisar el referente.

Los resultados de este estudio darán las bases para la creación de sistemas computacionales que integren y aprovechen el uso de gestos en la comunicación hombre-máquina. Los sistemas multimodales actuales fusionan la información verbal y no verbal a partir de sólo unos cuantos deícticos gramaticales<sup>3</sup>. Nuestro estudio no sólo enlista exhaustivamente los deícticos gramaticales usados, sino extiende este repertorio a expresiones que, a pesar de ser expresiones definidas o “completas”, son acompañadas por gestos. Desde el punto de vista del desarrollo de un sistema computacional este hecho es de gran importancia pues impactará en la eficiencia y eficacia al momento de determinar un referente. Decimos que será más eficiente pues no será necesario recurrir a procesos computacionalmente costosos dedicados a la inferir de que objeto (o lugar) se habla, sólo bastará tomar la posición del cursor para determinar el objeto (o lugar) involucrado. Por ejemplo, en el caso de una orden involucrando la expresión definida “la estufa roja” necesitamos un proceso computacional que revise cada una de las descripciones de los objetos en la escena, enfocándose en el tipo y color hasta determinar de qué objeto se trata. Este proceso de *búsqueda de referentes* puede llegar a ser mucho más costoso al imaginar expresiones como “la estufa al lado de la alacena”. Con ayuda de los gestos el proceso se simplifica mejorando su eficiencia y, lo que es más, su eficacia. Nuestro estudio comprueba la presencia de gestos y su uso, y determina claramente que éstos no sólo son usados como acompañamiento de los deícticos gramaticales.

---

<sup>3</sup> Cabe mencionar que dichos sistemas sólo existen para el idioma inglés y que hasta el momento este es el primer estudio para el español.



## Agradecimientos

Los autores agradecen el financiamiento parcial otorgado por el CONACYT, el SIN y el Laboratorio Franco-Mexicano de Informática (LAFMI) bajo el marco del proyecto “Interacción oral hombre-máquina”.

## Referencias

- [Dahlbäck et al, 93] Dahlbäck, N., Jönsson, A. & Ahrenberg, L. Wizard of Oz Studies - Why and How. *Knowledge-based Systems*, 6(4), (1993) 258-266.
- [Taylor et al 89] Taylor, M., Néel, F. & Bouwhuis, D. (Eds) *The Structure of Multimodal Dialogue*. Elsevier Science Publishers, 1989.
- [Tusón 97] Tusón De Valls, A. (1997). *Análisis de la conversación*, Barcelona, Ariel. 1997
- [Vázquez 04] Vázquez, C. *Análisis de deícticos en la conversación orientada por la tarea entre un usuario y un sistema computacional simulado*. Tesis de Licenciatura. Facultad de Filosofía y Letras. BUAP. 2004.
- [Villaseñor et al 01] Villaseñor, L., Massé A. Pineda, L. *The DIME corpus*. Memorias del 3er Encuentro Internacional de Ciencias de la Computación. pp 599-600, SMCC-INEGI, México. 2001.

# Diseño de Aplicaciones Interactivas utilizando el Paradigma de Patrones

Jaime Muñoz Arteaga<sup>1</sup>, Juan M. González Calleros<sup>2</sup> y Jorge Aguilar Cisneros<sup>3</sup>

<sup>1</sup>Departamento de Sistemas de Información de la Universidad Autónoma de Aguascalientes  
Av. Universidad # 940, Fraccionamiento Bosques, C.P. 20100, Aguascalientes, México  
[jmunozar@correo.uaa.mx](mailto:jmunozar@correo.uaa.mx)

<sup>2</sup>Coordinación de Ciencias Computacionales del (INAOE), Calle Luis Enrique Erro No. 1,  
Tonantzintla, Puebla, 72840, México,  
[jmgonzalez@ccc.inaoep.mx](mailto:jmgonzalez@ccc.inaoep.mx)

<sup>3</sup>Universidad Tecnológica de Puebla, Antiguo camino a la Resurrección #1002-A Zona  
Industrial, Puebla, , México, C.P 72300  
[jaguilar@utpuebla.edu.mx](mailto:jaguilar@utpuebla.edu.mx)

**Abstract.** An interactive application allows to guide and to make easy the user's task. However the design of this kind of application is known as a difficult task, one of the main troubles is to take into account the usability aspects in agreement with the functionality aspects. These two aspects in this paper are taken into account specifying an interactive application in terms of pattern approach. The patterns capture the best practices in the user-based design. The goal is to offer to designer an specification methodology for specify an interactive application independent of any programming language and this specification could facilitate the communication with the people involved in the software development.

**Resumen.** Una aplicación interactiva permite guiar y facilitar la tarea del usuario. Sin embargo el diseño de este tipo de aplicaciones se reconoce como una tarea difícil, una de las principales dificultades es integrar los requerimientos del usuario en el software de la aplicación. Estos dos aspectos en el presente trabajo son tomados en cuenta, especificando una aplicación interactiva en términos del paradigma de patrones. Los patrones pueden representar soluciones exitosas en el diseño centrado en el usuario. El objetivo es proveer al diseñador de una metodología para diseñar una aplicación interactiva independientemente de un lenguaje de programación y que el diseño facilite la comunicación del personal involucrado con el desarrollo de software.

## 1 Introducción

Cuando se diseña una aplicación interactiva implica especificar tanto la funcionalidad (o parte interna) como la interfaz de usuario (o parte externa) así como las relaciones entre estos dos módulos de software. Una aplicación interactiva tiene por objetivo facilitar y guiar al usuario en su tarea a través de la interfaz gráfica del usuario. Para esto el software no sólo debe de asegurar la funcionalidad misma de la aplicación,

sino también el software debe de asegurar la comunicación que se requieren entre el usuario y la aplicación. En particular es necesario adaptar y reutilizar el software de la interfaz de usuario dada la diversidad de perfiles de usuarios que pueden llegar a tener una aplicación interactiva.

El presente trabajo propone el diseño de aplicaciones interactivas en base al enfoque de patrones; donde la funcionalidad se especifica en términos de patrones de diseño [1] y la interfaz de usuario en términos de patrones de interacción [6]. La sección 2 de este trabajo describe a detalle la problemática que presentan el diseño de aplicaciones interactivas. En la sección 3 se propone una metodología como solución para diseñar este tipo de aplicaciones. La sección 4 presenta un caso de estudio y finalmente el enfoque aquí propuesto es puesto en comparación con otros trabajos relacionados.

## **2 Problemática**

La concepción de aplicaciones interactivas es un tema de investigación de gran interés particularmente en las áreas de la Interacción Humano-Computadora (IHC) y la Ingeniería de Software (IS). En el marco de las relaciones entre estos campos de estudio se abordan dos aspectos:

- La dificultad para construir software con alto nivel de utilidad y usabilidad
- Los problemas de comunicación entre los especialistas involucrados en el desarrollo de una aplicación.

En cuanto al primer aspecto la usabilidad es una cualidad del software que está estrechamente ligada a la interfaz de la aplicación. De una manera intuitiva, la usabilidad del software está determinada por el grado de satisfacción del usuario cuando utiliza el producto. Esta característica puede determinar el éxito o el fracaso de un producto. En cuanto al segundo aspecto el proceso de realización de una AI debe integrar un grupo de personas con diferentes especialidades (analista, diseñadores, programadores, ergónomos etc...) así como el mismo usuario con el fin de obtener un producto usable y de calidad.

La falta de integración entre la Ingeniería de Software y la Interacción Humano-Computadora se manifiesta de distintas formas. Por un lado, la ausencia del tratamiento de la construcción de interfaces en las metodologías más conocidas y usadas, en particular en el caso de las metodologías orientadas a objeto revela parte de las dificultades. En efecto, la mayoría de las metodologías adolecen de un tratamiento que integre el desarrollo de la interfaz y el desarrollo de la aplicación, hay un divorcio entre estos dos aspectos indisolublemente asociados de una aplicación, y si bien en ésta son inseparables, su construcción es realizada, en muchos casos, de manera independiente. Por otro lado, puede señalarse que esa separación en la metodología es un reflejo de las diferencias de los modelos mentales de ambos participantes, en lo concerniente al proceso, donde cada parte tiene su propia perspectiva y su propio conocimiento. Por ejemplo, para un ingeniero de software la funcionalidad de una aplicación puede ser la cualidad más importante, mientras que para un especialista en Interacción Humano-Computadora la usabilidad puede ser el objetivo. Tal separación tiene implicaciones también a nivel de la comunicación.

### 3 Metodología para Especificar Aplicación Interactivas

El presente trabajo propone especificar a nivel de análisis el poner en relación los requerimientos de la parte externa con los de parte interna tomando como base un prototipo de la interfaz de usuario de la aplicación interactiva a desarrollar. Para la etapa de diseño se propone aplicar el enfoque de patrones como un mecanismo que facilite la fluidez de la comunicación entre las partes (en particular el diseñador y el programador), donde cada una tiene su propia experiencia y conocimientos (fig. 1).

Por definición, un patrón es una solución a un problema recurrente dentro de un contexto específico. La especificación de un patrón permite comunicar la experiencia y el conocimiento en un área de interés [1]. La aplicación de la experiencia registrada por los patrones conlleva una serie de beneficios, por ejemplo los patrones de diseño permiten el reuso exitoso de diseños y arquitecturas, son un medio para compartir el conocimiento y nos proporcionan un lenguaje común entre diseñadores.

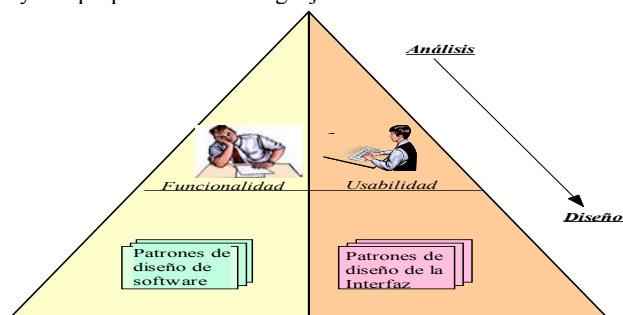


Fig. 1 Especificación de una aplicación interactiva utilizando el paradigma de patrones

Conforme a la figura 1, el presente trabajo preconiza diseñar la funcionalidad (parte interna) de una aplicación interactiva en términos de patrones de diseño [1] y la interfaz de usuario (parte externa) en términos de patrones de interacción [6]. Ambos tipos de patrones tienen un formato bien definido a saber: *nombre del patrón, problema que resuelve, solución propuesta, contexto, ejemplos etc.* Esto permitirá tanto al diseñador como al programador analizar a detalle aspectos de la funcionalidad y la usabilidad de una aplicación

### 4 Caso de estudio

El presente caso de estudio presenta el diseño de una aplicación interactiva dedicada a la administración (visualizar, editar y agregar información) de catálogo de autos. La figura 2 muestra el prototipo de la interfaz usuario en donde se identifican tres partes a saber: la zona de edición, la zona de comandos y la zona de despliegue de información.

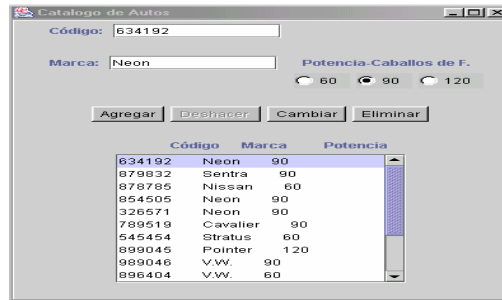


Fig. 2 Prototipo de la interfaz de la aplicación gestionaora de catálogo de autos

A través de los botones de la zona de comandos el usuario puede modificar, agregar y eliminar los registros sobre autos. El usuario puede editar y visualizar los contenidos de los registros que se encuentran en el catálogo utilizando respectivamente los componentes de entrada en la zona de edición y el componente de lista en la zona de despliegue de cada uno de los registros existente en el catálogo de autos.

#### 4.1 Análisis de requerimientos

A nivel de la etapa de análisis los requerimientos son plasmados en forma tabular la cual permite razonar sobre la relación de los elementos de la parte externa con los de la parte interna de la aplicación de nuestro caso de estudio.

Tabla 1 Análisis de requerimientos de la aplicación catálogo de autos.

Parte Externa (Interfaz Usuario)			Parte Interna (Funcionalidad)	
Interactor	Eventos	Servicios	Estado	Clases
Botón 1	Click	Agregar	Activación	GestionActivacion
Botón 2	Click	Deshacer	Activación	GestionActivacion
Botón 3	Click	Cambiar	Activación	GestionActivacion
Botón 4	Click	Eliminar	Activación	GestionActivacion
Radio botón 1	Click	Editar	Edición	GestionEdicion
Radio botón 2	Click	Editar	Edición	GestionEdicion
Radio botón 3	Click	Editar	Edición	GestionEdicion
Cuadro edición 1	AnyKey	Editar	Edición	GestionEdicion
Cuadro edición 2	AnyKey	Editar	Edición	GestionEdicion
Cuadro de lista	Click	Seleccionar	Selección	GestionCatalogo


Note que en la tabla, por ejemplo, más de un objeto de interacción puede estar asociado a un servicio que ofrece la aplicación al usuario, este es el caso del servicio *Editar*, el cual tiene 3 *Radio botones* y 2 *Cuadro de Edición* asociados. Otro relación importante es que por cada servicio se muestra su variable de estado y la clase

orientada a objetos que le corresponde, por ejemplo el servicio *Editar* está implementado por la clase *GestionEdicion* y dicho servicio estará disponible (o indisponible) en función del valor del estado *Edición* que tome durante el proceso de interacción.


La información de la presente tabla 1 es de utilidad para la etapa de diseño se presenta en la dos subsecciones siguientes.

#### 4.2 Especificación de la interfaz en términos de patrones de interacción

A partir de la información obtenida en la etapa de análisis y de las clasificaciones de patrones de interacción propuestos por Welie [6] y Muñoz [4] es posible identificar los patrones *Lista* [6] y *Activación proactiva* [4] para así diseñar la interfaz de usuario de la aplicación gestionaora de catálogos de autos.

<b>Nombre</b>	<b>Activación proactiva</b>
<b>Problema</b>	<b>El usuario necesita conocer a priori las acciones posibles con el fin de determinar su siguiente acción.</b>
<b>Principio usabilidad</b>	Predictibilidad y acciones explícitas
<b>Contexto</b>	El espacio de interacción debe ser visualizada continuamente al usuario.
<b>Fuerza</b>	Guiar al usuario en sus acciones ofreciendo los servicios disponibles en función del estado actual de una aplicación interactiva.
<b>Solución</b>	<b>Atenuar los widgets de la interfaz para evitarle al usuario activar servicios indisponibles en la interfaz del usuario. Cuando un servicio este disponible los widgets asociados debe estar resaltados.</b>
<b>Cosecuencia</b>	Este tipo de retroalimentación es bastante común a pesar que no ofrecer información suficiente sobre la razón del porque de la disponibilidad de los servicios del sistema.
<b>Ejemplo</b>	 <p>En la parte de interfaz anterior, el conjunto de servicios validos son aquellos botones con la etiqueta <i>Agregar</i>, <i>Cambiar</i> y <i>Eliminar</i>.</p>

<b>Nombre</b>	<b>Lista</b>
<b>Problema</b>	<b>¿Cómo mostrar el estado de una aplicación cuya información tiene una estructura plana?</b>
<b>Principio usabilidad</b>	Reducir la carga cognitiva y representación coherente.
<b>Contexto</b>	Una gran cantidad de información será secuencial. El usuario puede facilitarse el acceso de la información desplazándose a través de una de lista.
<b>Fuerza</b>	Una representación de lista es fácil de entender y de manipular. El usuario manipula y visualiza la relación de los elementos.
<b>Solución</b>	<b>Mostrar los datos en forma de lista donde exista un punto de base y</b>

	<b>el punto terminal, representan datos elementales.</b>
<b>Conse- cuencia</b>	La representación plana motiva al usuario a descubrir información que le puede ser útil para alcanzar su objetivo. Dicha representación también le permite tener una imagen mental correcta del acceso de la información que ofrece el sistema interactivo.
<b>Ejemplo</b>	 <p>A través de esta representación plana el usuario puede acceder y elegir sus datos que le son útiles.</p>

#### 4.3 Especificación de la interfaz en términos de patrones de software

A partir de la información obtenida en la etapa de análisis y de la clasificación de patrones de diseño propuesto por [1] es posible identificar los patrones *Comando* e *Iterador* para así diseñar el software con la funcionalidad de la aplicación gestionadora de catálogo de autos.

Nombre	Comando
<b>Propósito</b>	Encapsula una petición en un objeto, permitiendo así parametrizar a los clientes con diferentes peticiones hacer fila o llevar un registro de las peticiones, y poder deshacer las operaciones.
<b>Motivación</b>	A veces es necesario enviar peticiones a objetos sin saber acerca de la operación solicitada o de quien es el receptor de la petición. Por ejemplo los toolkits de interfaces de usuario incluyen botones y menús que realizan una petición en respuesta a una entrada de usuario.
<b>Aplicabilidad</b>	Usar el patrón comando cuando se quiera: <ul style="list-style-type: none"> <li>· Parametrizar objetos con una acción a realizar como ocurre con los objetos.</li> <li>· Especificar, poner en fila y ejecutar peticiones en diferentes instantes de tiempo</li> <li>· Permitir deshacer y registrar los cambios de manera que se puedan volver a aplicar en caso de una caída del sistema</li> </ul>

<b>Estructura</b>	
<b>Participantes</b>	<p><i>Orden (Gestión de Activación)</i> :  declara una interfaz para ejecutar una operación  <i>OrdenConcreta(Agregar,cambiar, etc.)</i>  – Define un enlace entre un objeto Receptor y una acción  Implementa el método de ejecución</p>
<b>Colaboraciones</b>	<p>El cliente crea un objeto OrdenConcreta y especifica su receptor.  Un objeto Invocador almacena el objeto OrdenConcreta.  El Invocador envía una petición llamando ejecutar sobre la orden..</p>
<b>Cosecuencia</b>	<p>El patrón comando tiene las siguientes consecuencias:  1. Las órdenes son objetos de primera clase. Pueden ser manipulados y extendidos como cualquier otro objeto,  2. Es fácil añadir nuevas ordenes, ya que no hay cambiar las clases existentes.</p>

<b>Nombre</b>	<b>Iterador</b>
<b>Propósito</b>	<p>El patrón Iterador, proporciona una manera de acceder secuencialmente a los elementos de un objeto agregado sin exponer su representación interna.</p>
<b>Motivación</b>	<p>Es conveniente usar este patrón cuando una clase necesita acceder el contenido de una colección, sin hacerse dependiente de la clase que implementa la colección. También es útil cuando una clase necesita acceder el contenido de múltiples colecciones, de manera uniforme.</p>
<b>Aplicabilidad</b>	<p>El patrón <i>Iterador</i> se emplea: Para acceder el contenido de un objeto agregado sin exponer su representación interna y permite varios recorridos sobre objetos agregados.</p>
<b>Estructura</b>	
<b>Participantes</b>	<p><i>Iterador</i>: Es una interfaz para recorrer los elementos y acceder a ellos.  <i>IteradorConcreto</i>: Implementa la interfaz de <i>Iterador</i>.  <i>GestorCatalogo</i>: Define una interfaz para crear un objeto <i>Iterador</i>.  <i>GestorCatalogoConcreto</i>: Implementa la interfaz de creación de <i>Iterador</i> retornando una instancia <i>IteradorConcreto</i> apropiado.</p>
<b>Colabora-</b>	<p>Una clase <i>IteradorConcreto</i> sabe cual es el objeto actual del agregado</p>



<b>ciones</b>	y puede calcular el objeto siguiente en el recorrido.
<b>Cosecuencia</b>	<ul style="list-style-type: none"> <li>– Permite variaciones en el recorrido de un agregado.</li> <li>– La interfaz de recorrido de Iterador elimina la necesidad de una interfaz similar en Agregado.</li> </ul> <p>Se puede hacer más de un recorrido a la vez sobre un agregado. Un Iterador mantiene su propio estado del recorrido.</p>

## Trabajos relacionados

En esta sección el presente trabajo es comparado con otros trabajos relacionados en el diseño de aplicaciones interactivas.

**Tabla 2.** Comparación de trabajos relacionados

	Parte Externa			Parte Interna		Patrones	
	Evento	Paradigma OO	Servicios	Iteractor	Estado	Diseño	Interacción
Explora[2]	<i>No</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>No</i>
WDPL[3]	<i>No</i>	<i>No</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>No</i>	<i>Si</i>
PIUI[5]	<i>No</i>	<i>No</i>	<i>Si</i>	<i>No</i>	<i>No</i>	<i>Si</i>	<i>No</i>
DiaMOD[6]	<i>No</i>	<i>No</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>No</i>
In&outSpec	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>	<i>Si</i>

Conforme a la tabla 2 el modelo DiaMODL[6] y el enfoque aquí presentado, titulado In&OutSpec (especificación de la parte interna y externa de una aplicación), muestran cubrir mejor las criterios de la tabla. Note también que el enfoque In&OutSpec es el único que especifica una aplicación interactiva aplicando patrones de diseño de software y de interfaces de usuario.

## 6. Conclusiones

El presente trabajo preconiza a nivel de análisis y diseño la reutilización y el fácil mantenimiento de la parte interna y externa de una aplicación interactiva. A nivel de análisis se propone poner en relación los requerimientos de la parte interna y externa de una aplicación. A nivel de diseño se propone una especificación más cerca al código en términos de patrones de diseño y una especificación más cercana a los requerimientos del usuario en términos de patrones de interacción. Uno de los objetivos alcanzados con los patrones es el de facilitar la fluidez de la comunicación entre las partes, donde cada una tiene su propia experiencia y conocimientos. Los patrones de diseño permiten especificar el código de la funcionalidad de una aplicación dando soluciones al diseñador para reutilizar y mantener fácilmente el código. Por su parte, los patrones de interacción propuestos aquí permiten especificar las experiencias exitosas en el diseño de la interfaz gráfica, tomando en cuenta las necesidades del usuario y los componentes de una interfaz gráfica. El modelo aquí propuesto puede ser extendido a aplicar los patrones de software para la generación

de código de una aplicación interactiva. Esto con el fin de llevar a cabo un rápido prototipaje y realizar así pruebas en conjunto con el usuario.

## 7 Referencias

- 1 Gamma E., Helm R., Johnson R., Vlissides J. and Booch G.: Design Patterns : Elements of Reusable Object-Oriented Software, Addison-Wesley Professional Computing (1995)
- 2 Molina, Pedro J., Torres, Ismael, and Pastor, Oscar: Patrones de interfaz de usuario para la exploración orientada a objetos. Congress IPO'2002 (Interacción Persona Ordenador). (2002) 27-36.
- 3 Montero, Francisco, Lozano, María Dolores, González, Pascual, and Ramos, Isidro: Designing Web Sites by Using Design Patterns. Second Latin American Conference on Pattern Languages of Programming: SugarLoafPLoP. Río de Janeiro, Brasil. (2002). 209-224.
- 4 Muñoz Arteaga J: A visual feedback for interactive systems. In Task Models and Diagrams for User Interface Design. International Workshop on Task models and Diagrams for user interface design Tamodia'2002. INFOREC Publishing House, Bucharest, Romania (2002).
- 5 Orantes Jiménez, Sandra Dinora: Patrones de Diseño para Interfaces de Usuario. Instituto Politécnico Nacional. México, D. F. (1999)
- 6 Welie, van Martijn and Trétteberg, Hallvard: Interaction Patterns in User Interfaces. 7<sup>th</sup> Pattern Languages of Programs Conference. Allerton Park Monticello, Illinois, USA. (2000)

# Impacto de la usabilidad en la implantación de un ambiente de tesis digitales

Lourdes Fernández Ramírez, J. Alfredo Sánchez, Alberto García

Bibliotecas, Universidad de las Américas, Puebla  
e-mail: {lulu, alfredo, alberto}@mail.udlap.mx  
Santa Catarina Mártir, San Andrés Cholula, Puebla, México 72820

**Abstract:** We present the results and implications of a usability study in the context of a collaborative environment for the construction and use of an institutional digital theses collection. The academic departments, which play an important role in the workflow, have been involved with different participation levels. For this reason, it is our interest to identify which factors influence how tasks are performed. The results show that the usability is good and that it does not influence task performance. On the other hand, the results show an important correlation between performance and direct assignment by a supervisor of the activities within the organization.

**Resumen:** Se presentan los resultados e implicaciones de un estudio de usabilidad realizado en el contexto de un ambiente de colaboración, el cual está orientado a la construcción y uso de una colección institucional de tesis digitales. Los departamentos académicos, que juegan un papel importante en el flujo de trabajo del sistema, se han involucrado con diferentes niveles de participación. De tal manera, nuestro interés ha sido determinar los factores que influyen sobre las tareas a desempeñar. Los resultados muestran que la usabilidad del ambiente es buena y que no influye en el desempeño de las tareas. En contraste, los resultados muestran una correlación importante entre el desempeño y la asignación directa de las actividades dentro de la organización.

**Palabras clave:** Tesis digitales, Flujo de trabajo, Usabilidad.

## 1. Introducción

La Universidad de las Américas, Puebla (UDLA), es la primera universidad en México en establecer la entrega de tesis en formato digital como requisito de graduación o titulación para todos sus programas académicos. En 1999, la Biblioteca inició un proyecto para crear un acervo digital de tesis que incluyera todas las tesis generadas por sus graduados [3]. A finales del 2002, la Universidad aprobó el requisito de tesis digitales para todos sus programas académicos (39 licenciaturas, 25 maestrías y 2 doctorados). Puesto que el volumen de tesis a manejar se incrementaría considerablemente se desarrolló un ambiente que permitieran a los involucrados en el

proyecto colaborar en la construcción de la colección de tesis digitales. A este ambiente de colaboración se le ha denominado genéricamente “Tales”.

Hoy en día, después de un año de establecer el requisito de tesis digital para los 26 departamentos de la universidad, el nivel de participación ha variado de un departamento a otro, aunque el requisito es general. El propósito de nuestra investigación es identificar los factores que influyen el nivel de participación de los participantes y si el sistema de flujo de trabajo es el más adecuado para realizar las tareas. Este trabajo se enfoca a la función del coordinador del departamento académico que es quien inicia el flujo de trabajo en la parte administrativa. La tarea del coordinador del departamento consiste en registrar en *Tales* [2], el comité de evaluación del examen profesional que corresponde a cada estudiante. Esta tarea es muy importante ya que sin ésta el director de la tesis no puede proceder a hacer la revisión y autorización de la misma. El área de Bibliotecas Digitales procesa y hace disponible la tesis únicamente hasta que es autorizada.

Nuestro interés es saber si son la falta de asignación directa, el tiempo de procesamiento o la usabilidad del ambiente Tales las causas que determinan que la participación en el registro de comité de evaluación de tesis varíe entre los diferentes departamentos académicos. Para responder a este problema se aplicaron encuestas a los profesores de tiempo completo y a asistentes de nuestra Universidad que desempeñan o podrían desempeñar este rol. Es importante mencionar que en este proyecto no se ha considerado el punto de vista de otro tipo de usuarios, como estudiantes o directores de tesis.

## 2. Creación colaborativa de una colección digital de tesis

Tales se desarrolla dentro de la iniciativa internacional de una Red de Bibliotecas Digitales de Tesis y Disertaciones (*Networked Digital Library of Theses and Dissertations, NDLTD*) [4]. La colección de tesis digitales se integra a la biblioteca tradicional y a las bibliotecas digitales, beneficiándose de los servicios que ambas ofrecen. Por ejemplo, de la biblioteca tradicional recibe la clasificación y catalogación, mientras que de la Biblioteca Digital puede recibir servicios de recomendación [6] y referencia virtual [7].

Con el propósito de integrar en un ambiente colaborativo todas las etapas del ciclo de vida de la colección de tesis digitales se diseñó e implementó *Tales* [2]. El sistema de flujo de trabajo apoya a los participantes en el proceso desde producción hasta publicación: autores, directores de tesis, coordinadores de los departamentos académicos, administradores y personal de biblioteca. El ciclo inicia con la entrada de los documentos a la colección para revisión, continúa con la anotación y publicación, y finaliza con las búsquedas y navegación sobre la colección.

Otras instituciones miembros de la organización NDLTD siguen un flujo de trabajo propuesto en el software de distribución libre sugerido por la organización [1]. Éste basa sus tareas en el autor (estudiante graduado) y biblioteca. Cabe mencionar que la

mayoría de las instituciones participantes sólo consideran programas de maestría y doctorado; esto reduce, la mayoría de las veces, el número de tesis a procesar por semestre.

### 3. Metodología

De acuerdo a la política publicada por Junta Académica a finales del 2002 son los departamentos académicos los encargados de hacer el registro de los comités de evaluación de tesis. Como ya se mencionó, es de nuestro interés comprobar si la falta de asignación directa, el tiempo de proceso y la usabilidad del ambiente diseñado son factores que influyen negativamente en la tarea de registrar comités a través de Tales. Esta investigación contiene 3 variables independientes y una dependiente. Las variables independientes son: asignación directa, tiempo de proceso (número de tesis registradas) y usabilidad del ambiente. Mientras que la variable dependiente es el número de comités registrados a través del ambiente.

*Asignación directa.* Se refiere a recibir una instrucción directa por parte de un superior para desempeñar una tarea y se calcula directamente con base en la pregunta 5 del instrumento: sí hubo asignación o no hubo asignación.

*Tiempo de proceso.* El tiempo de proceso en registro de comités de evaluación es directamente proporcional al número de estudiantes que registran su tesis vía Tales. Cada estudiante registrado corresponderá a una unidad de tiempo. La información de número de tesis registradas en el ambiente se obtuvo de la base de datos de Tesis digitales.

*Usabilidad de Tales.* La usabilidad es la medida de calidad en la experiencia de uso que tiene el usuario al interactuar con el sistema, sea éste un sitio de Web, una aplicación de software, alguna tecnología móvil o de algún otro dispositivo [5]. Para esta investigación el resultado de usabilidad se calcula con base en las respuestas de las preguntas 10 a la 19 que miden las características de un sistema usable. Se obtiene un promedio de de las respuestas en una escala del 1 a 5; donde 1 es calificación más baja y 5 la más alta. Cabe mencionar que no todos los encuestados respondieron estas preguntas porque sólo algunos estaban capacitados para hacerlo (quienes ya han utilizado Tales).

*Número de comités registrados a través del ambiente.* El registro de comité es la asignación, a través de una forma disponible vía Web, de cada uno de los evaluadores de la tesis: presidente, secretario, vocal, suplente (en algunos casos) y director. El número de comités registrados a estudiantes con ayuda del ambiente Tales se obtiene de la base de datos de Tesis digitales.

Esta investigación se realizó sin manipular deliberadamente las variables; las variables independientes ya habían ocurrido. Se realizaron mediciones simultáneas en donde cada escuela de la universidad corresponde a un grupo de enfoque.

## 4. Resultados

En la Tabla 1 se muestran los resultados obtenidos del promedio de las respuestas de usabilidad (los nombres de las escuelas se han sustituido por E1-E5 para proteger su anonimato). Se obtuvo una media de los reactivos concernientes a esta área en el ambiente (10 reactivos). Se muestra un resultado satisfactorio ya que 4 escuelas presentan una media mayor a 4 puntos (en una escala de 1 a 5).

Escuela:	E1	E2	E3	E4	E5	promedio
media	4.4	3.6	4.3	4.1	4.9	4.3
varianza	0.087	0.267	0.162	0.045	0.100	0.054
desviación estándar	0.295	0.516	0.403	0.212	0.316	0.233

**Tabla 1.** Resultados de la prueba de usabilidad

La Tabla 2 muestra los resultados de la prueba de usabilidad desglosado por tipo de respuesta. Aquí se identifica que más del 80% de los usuarios evaluaron el ambiente satisfactoriamente.

	E1	E2	E3	E4	E5	promedio
5. excelente	40%	0%	<b>50%</b>	22%	<b>90%</b>	40%
4. muy bien	<b>55%</b>	<b>60%</b>	32%	<b>70%</b>	10%	<b>45%</b>
3. regular	5%	40%	13%	8%	0%	13%
2. mal	0%	0%	5%	0%	0%	1%
1. muy mal	0%	0%	0%	0%	0%	0%

**Tabla 2.** Resultados de la prueba de usabilidad

La asignación directa se muestra en la Tabla 3. Se puede identificar en ella que hay una variación considerable entre las diferentes escuelas.

	E1	E2	E3	E4	E5	promedio
sí	50%	60%	80%	78%	40%	65%
no	50%	40%	20%	22%	60%	35%

**Tabla 3.** Resultados de asignación directa

En la Tabla 4 podemos ver cuál es la proporción de las tesis registradas que tienen el comité evaluador asignado a través del sistema.

	E1	E2	E3	E4	E5	promedio
	72%	69%	71%	63%	26%	59%

**Tabla 4.** Proporción de comités registrados a tesis registradas

## 5. Análisis de resultados

Al comparar cada una de las variables independientes con la variable dependiente pueden derivarse algunas observaciones importantes. Respecto a la usabilidad del ambiente, no parece influir directamente o negativamente, ya que la escuela con el mejor desempeño en cuanto a usabilidad (E5) tiene la menor proporción de comités registrados. De la misma manera, la escuela con la evaluación más baja (E2), tiene un porcentaje alto (comparado con la media) de comités registrados. La asignación directa si parece ser un factor que influye negativamente ya que la escuela que tiene menor porcentaje de asignación directa (E5), también tiene la menor proporción de comités registrados por tesis. En los demás casos los resultados son muy similares. El número de comités a registrar tampoco parece tener influencia directa ya que casi todas las escuelas (excepto E5) un porcentaje similar de comités registrados respecto de tesis recibidas.

## 6. Conclusiones y perspectivas

Como resultado del estudio realizado se tomaran diversas acciones, las cuales se comentan brevemente a continuación:

(1) Se simplificará la interfaz al asignar roles a miembros del jurado evaluador, con la opción de agregar o eliminar miembros e incluyendo opciones para agregar profesores no registrados. (2) Se implementará un servicio de alertas periódicas para que los responsables identifiquen fácilmente a estudiantes con el registro de comité pendiente. (3) Se generarán reportes con las tesis que están pendientes por registro de comité o para revisión, resaltando la fecha o tiempo de espera. (4) Los directores de tesis se podrán registrar automáticamente en los comités de sus asesorados.

El servicio de alertas ayudará al responsable a verificar sus registros pendientes y su prioridad. La identificación de un responsable permitirá que los autores y asesores los contacten directamente, lo que reducirá tiempos de espera. Los asesores no tendrán que esperar a que alguien más haga el registro porque podrán hacerlo ellos mismos. La intención del sistema desarrollado es ser un ambiente que ayude en las tareas de registro de comité, no incrementar su trabajo. Por ejemplo, algunos hacen el registro como estaban acostumbrados a hacerlo (ya sea papel y lápiz o de manera electrónica) y después lo hacen en el ambiente. De esta manera se duplica el trabajo y esto es lo que se quiere evitar. El ambiente debería cubrir lo que ya hace actualmente y más. Es importante que cada semestre se revise con cada departamento quién es el encargado de hacer este registro y si el ambiente cubre o no sus necesidades.

A pesar de que la evaluación de la usabilidad en general es positiva, la retroalimentación recibida durante el desarrollo de este proyecto permitirá hacer modificaciones al ambiente para mejorar su usabilidad y utilidad. Entre estas modificaciones sobresalen, la modificación de la interfaz para registro de comités y la

implementación de alertas periódicas para que los responsables identifiquen fácilmente qué estudiantes tienen pendiente el registro de comités.

Nuestras recomendaciones para interesados en construir colecciones o servicios similares son considerando que cada caso tendrá necesidades diferentes. Es importante que los involucrados se rijan sobre una norma y que ésta se de a conocer a cada uno de los participantes; que no haya ambigüedades en su nivel de participación. Puede existir un rechazo por la asignación de nuevas actividades pero se verá disminuido cuando los participantes reciban las instrucciones y el soporte adecuados; siempre tomando en consideración su opinión. El tamaño de la colección es un factor que puede determinar si es mejor un flujo de trabajo donde se involucren a varios participantes o si la opción de un servicio centralizado es la más adecuada.

## 7. Referencias

- [1] Douglas, K., Coles, B., Porter, G., Van de Velde, E. 2003. Taking the Plunge: Requiring the ETD. En memorias del Sexto Congreso Internacional de Tesis y Disertaciones Electrónicas (ETD 2003, Berlín, Alemania, Mayo).
- [2] Fernández Ramírez, L., Sánchez, J. A. 2003. Community Tales: An infrastructure for the collaborative construction of digital theses repositories. En memorias del Sexto Congreso Internacional de Tesis y Disertaciones Electrónicas (ETD 2003, Berlín, Alemania, Mayo).
- [3] Fernández Ramírez, L., Sánchez, J. A., García, A. 2002. Tales: Integración de tesis en una biblioteca digital avanzada. Scire: Representación y Organización del Conocimiento 8, 2 (Jul.-Dic.), 61-70. Zaragoza, España. (ISSN 1135-3761)
- [4] Fox, E., Eaton, J. L., McMillan, G., Kipp, N. A., Mather, P., Phanouriou, C. 2001. Networked Digital Library of Theses and Dissertations. NDLTD Project. Disponible en: <http://www.ndltd.org/>; fecha de última modificación: 26 marzo 2001; fecha de último acceso: julio 2001.
- [5] Nielsen, J. 1993. Usability engineering; Morgan Kaufmann publishers. USA
- [6] Ramírez Vargas, M. 2001. Servicios de recomendación en la biblioteca digital de la UDLA-P. Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. Mayo.
- [7] Sánchez, J. A., García, A. J., Proal, C., Fernández, L. 2001. Enabling the collaborative construction and reuse of knowledge through a virtual reference environment. Proceedings of the Seventh International Workshop on Groupware (Darmstadt, Germany, Sept. 6-8). IEEE Computer Society Press, Los Alamitos, Calif. 90-97.



# **Evaluación de Maestros: Una Herramienta en Línea**

Gerardo Barreto Ironelis Valdez Batista Omar Álvarez Sara Sandoval

Facultad de Telemática Universidad de Colima  
Av. Universidad 333 Col. Las Víboras Colima, Col. 28044 México  
gradbato@yahoo.com.mx

## **Abstract**

The teacher evaluation, a process that is periodically carried out, is considered tedious because of the quantity of calculation and consideration that should be evaluated with the process. Given that the teacher is considered as the principal agent of educational development, the catalyst of knowledge, ability, attitudes and values in student. Teacher evaluation forms the base for quality in the educational institution. In this chapter we have the proposal and solution to the complexity of this process in search of quality, through the implementation, test and functionality definition of a designed tool to be used on-line, which allows the students to evaluate their teachers. This tool was designed under the directions cited by Nielsen (2002).

## **Resumen**

La evaluación de maestros, es un proceso que se realiza periódicamente, se considera tedioso por la cantidad de cálculos y condiciones que se deben evaluar con éste. Dado que el maestro es considerado como el agente principal del desarrollo educativo, el detonador del conocimiento, habilidades, actitudes y valores en los estudiantes. Es la base de la calidad de la institución educativa. En este apartado se tiene la propuesta y solución a la complejidad de éste proceso en busca de calidad, mediante la implementación, prueba y definición de funcionalidad de una herramienta diseñada para ser usada en línea, la misma permite a los alumnos evaluar a los maestros. Esta herramienta se diseño basada en las directrices citadas por Nielsen (2002).

**Palabras claves:** HCI, Anáhua, EM, Web.

## **1. Introducción**

En los últimos tiempos se habla más que nunca de la necesidad de una educación y de universidades de calidad. El clima de desarrollo tecnológico que se vive en la actualidad, el creciente interés por la eficacia, la efectividad de los servicios y los

altos costes de la educación superior tiene mucho que ver con esa preocupación creciente por encontrar la calidad anhelada.

El docente (maestro) es considerado en las universidades como elemento indispensable, parte integrante del sistema que contribuye para lograr los objetivos finales y la calidad en la educación. No es uno más entre los que participan, sino que es, a modo personal, uno de los principales que influyen decisivamente en el producto final [1].

La evaluación de maestros por parte de los estudiantes es un proceso continuo, lo cual representa un gasto en las casas de estudios, debido a los insumos requeridos para este proceso como papel, lápiz, etc. Agregando la mano de obra directa e indirecta que en muchos casos deben ser tomados empleados de diferentes departamentos, lo que finalmente representa un gasto y atraso en la entrega de resultados. Otro aspecto que se debe considerar en el proceso de evaluación de maestros, son los errores que se tienen al manejar valores y el conteo de forma manual.

Por ésta razón se decide crear una herramienta disponible en Internet, dado que esta red se ha considerado como una gran base de datos de cobertura global. En este contexto, herramientas para la evaluación de maestros por parte de los estudiantes vía Web, ofrece un mayor soporte para la realización de este proceso, en ocasiones arduo y costoso para las instituciones educativas [2]. Con la herramienta de evaluación, se obtiene una recolección de datos confiables bajo perspectivas distintas, que aseguren la calidad y exactitud en la solución planteada.

## **2. Revisión de la Literatura**

El maestro es considerado como el agente principal del desarrollo educativo, el detonador del conocimiento, habilidades, actitudes y valores en los estudiantes. Es la base de la calidad de la institución educativa. Por estas y otras razones todas las universidades e instituciones educativas llevan a cabo el proceso de evaluación docente por parte de los estudiantes [3]. Los principales conductas que se persiguen evaluar son: asistencia, puntualidad, dominio de la materia, cumplimiento del programa, estructuración del curso, dinámica de la clase, manejo del grupo, motivación, evaluación, interés por enseñar, etc. La evaluación por parte de los estudiantes implica, definir el perfil docente ideal, generar y recopilar información, presentarla, analizarla e interpretarla para finalmente: diagnosticar, formular juicios y tomar decisiones.

Las decisiones que se deben tomar son:

- De índole formativa:
  - a) Identificar vías que realcen las destrezas profesionales del personal académico.
  - b) Instrumentar la profesionalización de la enseñanza.
  - c) Modernizar y actualizar al maestro en el uso de nuevas tecnologías y recursos didácticos diversos,

- d) Planificar el perfeccionamiento y el desarrollo profesional.
- e) Identificar intervenciones de capacitación y mejora.
- De índole administrativa:
  - a) Promover cambios en la estructura académica.
  - b) Remoción del personal no adecuado.

La universidad Anáhuac [4], define la evaluación docente como un proceso que está inmerso en una filosofía de la educación, va más allá de la generación descriptiva o comparativa, de indicadores y estadísticas. Es una evaluación orientada al desarrollo que se identifica con la evaluación formativa, cuyos objetivos son señalar aspectos que pueden ser mejorados y guiar las decisiones en torno a como mejorarlos. La información está dirigida a los propios sujetos evaluados, quienes participan activamente en el proceso. Dicha universidad inicia un proceso formal de la evaluación docente a partir del año de 1988 y hoy en día reconoce que su calidad a nivel docente ha crecido notablemente.

Dado que el proceso de evaluación tiene que ser realizado en las instituciones públicas y privadas, en el 2003 el Ministerio de Educación Nacional de Colombia promulgó el Decreto 2582. El objetivo de este decreto es establecer las reglas y mecanismos generales para la evaluación del desempeño de los docentes y directivos docentes que laboran en los establecimientos educativos estatales [4]. El decreto define los principios y el contenido de la evaluación, asigna las responsabilidades del Ministerio de Educación Nacional, las Secretarías de Educación, los rectores y directores; detalla el proceso general de evaluación, establece el período en el que se aplicará y precisa el uso de sus resultados.

La evaluación del desempeño de los docentes es una parte importante de la evaluación institucional y se realizará al terminar cada año lectivo. Ésta evaluará la actuación del docente como profesional que participa en el desarrollo del proyecto educativo institucional, planea su labor, aplica estrategias pedagógicas apropiadas, evalúa los resultados del aprendizaje y se integra con la comunidad educativa.

La evaluación del docente en una institución de educación superior, surge a raíz de considerar que uno de los principales problemas y dificultades que han impedido la aceptación y consolidación de los procesos de evaluación del profesorado universitario, han tenido su origen en que éstos se inician sin el adecuado consenso de las partes implicadas, lo que indudablemente genera actitudes de desconfianza, inhibición y rechazo. Por ello se consideró que antes de intentar hacer un planteamiento de evaluación, se debe considerar, tanto las opiniones de los profesores como de aquellos otros miembros de la comunidad implicados en esta evaluación, es decir: saber qué piensan, sobre quién, cómo y a través de qué procedimientos debe efectuarse el proceso de control de la actividad del profesorado al servicio de la Institución. Especialmente interesa conocer sus opiniones con relación a la utilidad o usos que puedan tener estos procesos de evaluación del profesor, con la finalidad de diseñar un modelo de evaluación docente acorde a sus características contextuales [5].

La puesta en marcha del sistema de evaluación docente es un paso en la dirección correcta para conseguir mejoras tangibles en la educación. Pero se trata sólo del primer paso. Lo que se requiere es aprovechar la radiografía para corregir las malas prácticas docentes, mejorar el sistema de apoyo técnico a los maestros, potenciar las buenas experiencias. Por ejemplo, sería provechoso que aquellos maestros destacados pudieran convertirse en verdaderos monitores del perfeccionamiento de sus colegas. La calidad de la enseñanza se define en la sala de clases. Si los profesores no tienen un buen desempeño, no sirven de mucho los grandes lineamientos sobre cómo mejorar la educación [6].

### **3. Solución propuesta**

El sistema de evaluación de maestros en línea, fue creado con el propósito de obtener una herramienta que permita recolectar los datos que sirven para evaluar a los maestros, permitiendo a los estudiantes de una manera fácil, clara, y eficaz responder las preguntas necesarias, las cuales arrojarán los datos que finalmente serán tabulados y presentados en estadísticas. Pero, ¿Qué es la herramienta de Evaluación de Maestros (EM)? Es una herramienta en línea que permite a los alumnos evaluar a un maestro a manera de puntaje y también tiene la posibilidad de emitir comentarios a un determinado maestro si así lo desea.

La manera en que la herramienta de “Evaluación de Maestros” guarda, procesa y presenta la información la convierte en un elemento para definirla como indispensable, para obtener resultados eficientes y confiables en el proceso de evaluación de maestros.

Las ventajas que brinda esta herramienta son muchas, pero para este apartado definiremos solo algunas:

- Generación de reportes ordenados por grupo de clase.
- Generación de estadísticas general o individual.
- Facilidad para la obtención de datos referente a los maestros y los grupos, ya que los mismos pueden ser cargados directamente desde archivo de texto.
- Guarda la información existente en la base de datos en un archivo externo (archivo de texto simple), archivo que puede ser cargado nuevamente si el usuario lo desea; esta opción permite mantener la información ordenada por periodo.
- Seguridad en la información y acceso.
- Control de alumnos, mediante una lista que indica quienes han evaluado y quienes no.

Su implementación se elaboró en plataformas de software libre, los cuales son ampliamente difundidos en el área del desarrollo de sistemas con la madurez suficiente para ser aplicados acertadamente en este proyecto (Tabla 1).

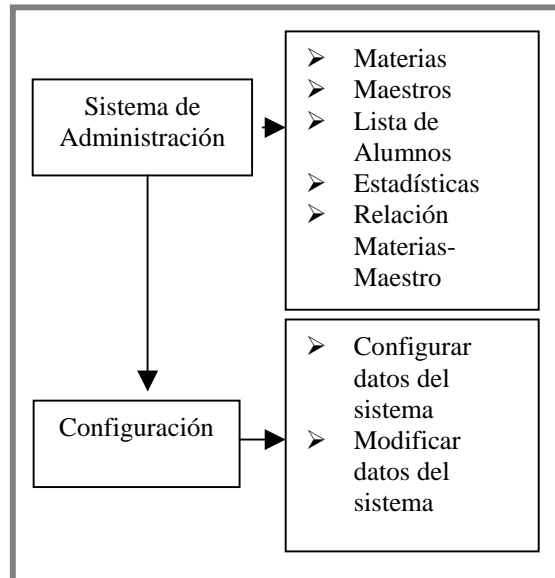
Tipo	Nombre	Características
Servidor Web	Apache	Servidor web más empleado en el mundo, flexibilidad de configuración.
Base de datos	MySQL	Motor de base de datos relacional, compacto y eficiente.
Interfaz	HTML	Lenguaje de definición de formato y contenido.
Interfaz	JavaScript	Lenguaje de programación.
Interfaz	PHP	Lenguaje de programación.

**Tabla1. Software utilizados para el desarrollo de la herramienta.**

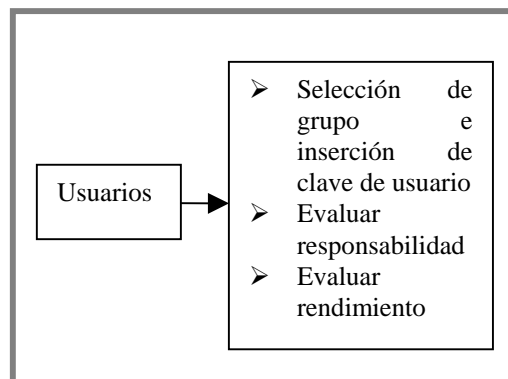
Es una herramienta creada bajo el entorno de software libre, con mira a dotarla de algunas de las ventajas del software libre, de las cuales se pueden citar algunas:

- **Bajo coste de propiedad:**  
Permite un muy sensible ahorro de costes en infraestructura y mantenimiento para proyectos a medida, en función de las necesidades del usuario.
- **Garantías de seguridad:**  
El hecho de disponer del código fuente de sus aplicaciones garantiza la seguridad, además de la fiabilidad del producto.
- **Cualquier desarrollo puede ser personalizado** para un cliente, debido a que disponemos del código fuente. Permite adaptar soluciones ya creadas para proyectos a medida.
- Permite un mayor avance de la tecnología debido a que el conocimiento se comparte. La ciencia avanza porque los científicos comparten los descubrimientos de manera libre.

Tomando en cuenta las ventajas de software libre y las ventajas observadas en la propuesta para la herramienta de evaluación de maestros, se va implementando de manera segmentada o en Frames, para fines de ir evaluando el alcance de cada una de las características definidas en el análisis y diseño de la herramienta. De dichos Frames, se obtienen los módulos que forman el Sistema de Evaluación de Maestros y la información y opciones que los componen a cada uno de ellos (Figuras 1 y 2).



**Figura 1. Módulo de administración y configuración.**



**Figura 2. Módulo de evaluación para usuarios.**

Según el funcionamiento de la herramienta, se puede decir que la misma entra en la definición de HCI (Interacción Humano Computadora). HCI es una disciplina que se refiere al diseño, evaluación y puesta en marcha de los sistemas de cálculo y operaciones interactivos de humano-maquina [4]. Una de las características de HCI, es la manipulación de información, que es precisamente lo que hace referenciar la herramienta de evaluación con el HCI, dado que EM es una herramienta para coleccionar, manipular y procesar datos referente a la evaluación de maestros, haciendo uso de las disponibilidades de la web.

Para el diseño de la interfaz se toman en cuenta las directrices citadas por Nielsen (2002), quien recomienda en primer lugar determinar el contenido de la página de inicio según las necesidades de las empresas, en ésta se usan colores en distintos tonos, basados siempre en las recomendaciones de Nielsen, que recomienda evitar colores fuertes a la vista del usuario. En esta página de inicio se encuentran también el logotipo de la herramienta en una tamaño razonable; el eslogan, para definir explícitamente las actividades del sitio; objetos de orientación para el usuario [7].

A continuación se describe de forma general el funcionamiento de esta herramienta.

Al iniciar, la herramienta detecta que no se ha configurado el sistema y le avisa al usuario, en seguida tiene que consultar al sistema de administración el cual lo redireccionará al módulo de configuración, ya que si no se configura el sistema no podrá seguir adelante, los datos necesarios para la configuración del sistema se presentan de modo claro, ordenados y precisos (Figura 3).



**Figura 3. Sistema de Configuración.**

Una vez proporcionados los datos podrá dirigirse al módulo de administración (Ver Figura 4).



**Figura 4. Sistema de Administración.**

En cuanto a las estadísticas, como ya se había mencionado, se pueden presentar por cada grupo), de manera general por grupo (Figura 5), y para establecer mediante estadísticas generadas por la herramienta el mejor maestro (Figura 6). En esta los puntos van de 0 a 100, siendo 100 el puntaje más alto. Cuando se visualizan las estadísticas se presentan junto con los puntos obtenidos de cada maestro los alumnos que evaluaron a dicho maestro y los comentarios obtenidos, y al final de las estadísticas por cada grupo o de manera general hay un enlace para las estadísticas y otro para los comentarios los cuales mandan llamar una nueva ventana con los datos correspondientes con la posibilidad de imprimirlos.



Figura 5. Estadísticas por cada grupo.

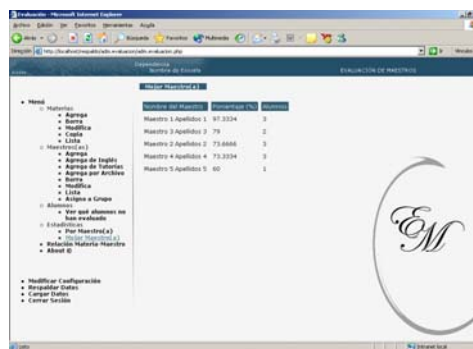


Figura 6. Estadísticas visualizando el mejor maestro.

Una de las metas esperadas con el desarrollo del sistema de Evaluación de Maestros configurable totalmente, en cuanto a puntos de evaluación y el ordenamiento de materias, maestros, y alumnos, de acuerdo a las necesidades y planes de estudios de las instituciones. En esta versión solo se pueden evaluar 10 puntos para responsabilidad y otros 10 puntos para rendimiento ( Figura 7).



La funcionalidad de la herramienta por su facilidad de uso ha sido aceptada por los estudiantes de manera muy positiva, ya que los mismo han contribuido con sus recomendaciones y comentarios, a las modificaciones pertinentes, para obtener una herramienta amigable, fácil de usar, fiable y abierta a ser adaptada a las necesidades de cada institución que decida usarla como herramienta de evaluación a los maestros.

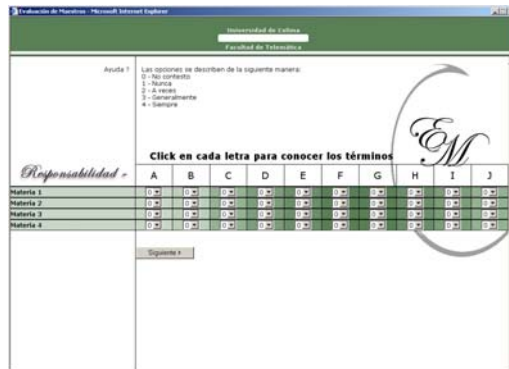


Figura 7. Ventana de evaluación.

#### 4. Conclusiones

La automatización de la información de recursos siempre ha sido un factor importante en cualquier organización ya sea por los costos que implica cierta actividad o por la duración que pueda implicar el realizarla. Anteriormente se ha observado el trabajo y tiempo que se lleva el planear e implementar una evaluación de maestros con el método tradicional, en el que se utilizan lápiz y papel como artefactos principales para la recolección de los datos, la tabulación de esos datos; todos estos conllevan una serie de pasos que al realizarlos manualmente se incurre en errores que pueden llegar a pasar desapercibidos en la organización, pero igual pueden generar una serie de problemas, riesgos y conflictos de soluciones. Ésta herramienta ha sido implementada y probada en dos facultades (escuelas) de la Universidad de Colima, y se observó que la evaluación a los maestros por parte de los estudiantes se realizó en menor tiempo, con igual o mejor precisión de cálculos, y que al final de ésta ya se contaban con los resultados a modo de reporte, los cuales, solo era cuestión de imprimirlos, tal como se pudo observar en Figura 5. Al comparar los resultados obtenidos con el reporte de la herramienta el error porcentual fue de un dos a tres por ciento, lo que significa que el rango de confiabilidad de los resultados es de 97%.

Uno de los problemas que se presentaron al inicio, en la prueba de la herramienta por parte de las facultades de la universidad que se pusieron en la disposición, fue que la herramienta no ofrecía la flexibilidad para que los alumnos de un mismo grupo tuvieran diferentes maestros, por ejemplo el caso de Inglés, para esta materia las distintas facultades (escuelas) hacen un examen de nivel y según las calificaciones obtenidas, son colocados en distintos grupos, y es en el caso en que los estudiantes de

un grupo, pueden en una materia como la citada previamente pueden estar en más de un grupo en un mismo semestre. Una vez desarrollada esta parte, la herramienta fue probada nuevamente y los resultados fueron tal y como se planeo al momento de la revisión. Una característica más que presenta la herramienta es que brinda confianza y seguridad en los cálculos de la información.

La evaluación de maestros es un proceso de mucha importancia, ya que con este se encaminan las instituciones educativas, a ser instituciones con calidad, respeto y confianza. Por tal razón el crear una herramienta para realizar este proceso vía Web, permite al estudiante hacerlo sin presión, con más conciencia, seguridad y facilidad. Por lo que a futuro se espera que esta herramienta sea totalmente configurable al gusto y necesidad, según los planes de estudio de las casas de estudios, tanto para las preguntas de evaluación que aplican, el peso o valor de cada una y a forma que deseen manejar los resultados.

## 5. Agradecimientos

Al desarrollar una herramienta como ésta, siempre es indispensable contar con el apoyo de alguna dependencia, la cual tenga la necesidad y la problemática definida a la hora de presentar una propuesta de solución. Por esta razón se agradece a las facultades de Telemática y Trabajo Social, ambas, de la Universidad de Colima quienes apoyaron desde la implementación hasta la prueba, siempre con la disposición de sugerencias para que la funcionalidad de la herramienta para evaluación de maestros sea la más próxima a los objetivos de desarrollo.

## 6. Referencias

- [1] F. A. a. R. González, "La calidad de la enseñanza superior y otros temas universitarios," *Universidad Politécnica., Madrid*, 1994.
- [2] J. S. d. I. Heras, ""Nuevos servicios a comunidades virtuales de usuarios (CVUs) de RedIRIS,"" *Boletín de la RedIris*, 1999.
- [3] U. d. Anáhuac, "La evaluación docente,"., pp. pp. 30, 2002.
- [4] Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, S. and, and Verplank, "ACM SIGCHI Curricula for Human-Computer Interaction," *ACM SIGCHI*, pp. 5, 2002.
- [5] J. C. a. A. Porto, "Evaluación de la docencia del profesorado universitario por los alumnos: En PÉREZ JUSTE, R.; CRISTÓBAL MARTÍN, MA. Y MARTÍNEZ AZNAR, BERTHA (Coords.)," in *Evaluación de Profesores y Reformas Educativas*, U. T. S.A, Ed. Madrid, 1995, pp. 45-89.
- [6] E. P. L. Nación, "La evaluación de los profesores," in *La Nación*, Digital ed. Chile, 2004.
- [7] J. Nielsen, "Usability Engineering," *AP Professional*, 1993.

# Interfaz con reconocimiento de voz para la manipulación robótica

Carlos Miranda-Palma<sup>1</sup>, Oscar Mayora-Ibarra<sup>2</sup> y Esmeralda Uruga-Serratos<sup>3</sup>

<sup>1</sup> UADY, Calle 48 A s/n x 31, 97700, Tizimín, Yucatán, México

<sup>2</sup> ITESM, Av. Paseo de la Reforma 182-A, 62589, Lomas de Cuernavaca, Temixco, Morelos, México

<sup>3</sup> IIMAS- UNAM, Circuito Escolar s/n, Ciudad Universitaria, 04510, D.F., México

cmiranda@tunku.uady.mx, omayora@itesm.mx, euraga@leibniz.iimas.unam.mx

**Abstract.** In this paper we describe an interface with speech recognition for the manipulation of a mobile Nomad Scout II robot. We describe the speech recognition system developed and we present the results obtained during the training of the acoustic models and during the recognition of the commands for the manipulation of the mobile robot. For the creation of the acoustic models, we used the HTK software of speech recognition.

**Resumen.** En este trabajo describimos una interfaz con reconocimiento de voz para la manipulación de un robot móvil Nomad Scout II. Se describe el sistema de reconocimiento de voz desarrollado y se presentan los resultados obtenidos durante el entrenamiento de los modelos acústicos y durante el reconocimiento de los comandos para la manipulación del robot móvil. Para la creación de los modelos acústicos, se utilizó el software de reconocimiento de voz HTK.

**Palabras clave:** Interfaz, reconocimiento de voz, corpus de voz.

## 1. Introducción

El uso de robots móviles para diversas áreas de la investigación se ha incrementado en estos últimos tiempos desde una simple manipulación hasta tareas complejas y de aprendizaje. Recientemente se han desarrollado diversos mecanismos para controlar sistemas de navegación robótica: a través de una lista de comandos secuenciales introducidos en modo de texto a una página Web [1], con gestos dinámicos realizados con la mano [2], y recientemente el uso del reconocimiento de voz [3]. Esto nos permite diversificar y ampliar las opciones de interacción en sistemas de navegación de un robot móvil. También nos brinda la oportunidad de realizar comparaciones de resultados al utilizar estas diversas opciones de navegación, y de realizar estudios de usabilidad de las interfaces desarrolladas para la manipulación de un robot móvil [4].

En el IIMAS de la UNAM también se han venido desarrollando diversos proyectos que incluyen la comunicación hablada como forma de interacción [5][6]. En [7] se presenta un sistema de reconocimiento del habla que también fue integrado al sistema de navegación de un robot móvil. Algunos de estos sistemas y herramientas desarrolladas

en estos proyectos fueron utilizadas y/o adaptadas para crear la interfaz con reconocimiento de voz presentada en este trabajo [8][9].

En este trabajo presentamos una interfaz para la manipulación de un robot móvil empleando la forma más natural de comunicación humana: el habla. Se presenta el desarrollo de un sistema de reconocimiento de voz continua en español para la navegación robótica por medio de comandos vocales. Los comandos hablados que se implementaron para su reconocimiento son 6: *Avanza*, *Gira\_Derecha*, *Gira\_Izquierda*, *Retrocede*, *Inicio* y *Fin*.

Para la creación del sistema de reconocimiento de voz continua en español se desarrolló un corpus de voz del español hablado en México. En este trabajo presentamos las características de los hablantes, el diseño del corpus y los resultados obtenidos en la evaluación del sistema de reconocimiento de voz.

## **2. Corpus de Voz**

Se han desarrollado algunos corpus de voz en español castellano y sudamericano, pero para el español hablado en México, que utiliza un sistema fonético y prosódico diferente a estos corpus de voz ya creados, muy pocos. En [10] y [11] se presentan algunos trabajos sobre la creación de corpus de voz para el español hablado en México.

En este trabajo el corpus fue diseñado para un propósito general. Para la navegación del robot móvil sólo se necesitaba un corpus que permitiera modelar básicamente 6 comandos vocales, sin embargo se diseñó un corpus cuyo contenido fonológico permitiera la construcción de un modelo acústico para que no sólo pueda reconocer estos comandos sino que esté diseñado para aumentar el número de comandos a reconocer o que pueda reconocer otro tipo de palabras o frases de tal manera que posteriormente pueda ser usado por otra aplicación.

Como el propósito del corpus es general se diseñó el contenido lingüístico del corpus basado en oraciones formadas con frases que contengan a la mayoría de los fonemas para el español hablado en México y que nos permitan obtener un número de muestras suficientes para crear sus modelos acústicos [11]. Las oraciones se diseñaron con significado lingüístico.

La grabación de los datos de voz consistió en que los hablantes leyeran de manera natural el conjunto de frases diseñadas. Cada uno de los hablantes grabó un conjunto de 54 frases. Las grabaciones se realizaron con un grupo formado por 40 personas cuya lengua materna es el español hablado en México. Estos 40 hablantes, (23 hombres y 17 mujeres), son provenientes de diferentes estados de la República Mexicana (9 de Morelos, 8 de Yucatán, 5 de Tabasco, 5 de Guerrero, 3 de Veracruz, 2 de Tamaulipas, y 1 de los siguientes Estados: Campeche, San Luis Potosí, Michoacán, Chiapas, Sinaloa, Chihuahua, Jalisco y Distrito Federal).

### 3. El sistema de reconocimiento de voz

El sistema de reconocimiento de voz utiliza 3 modelos principales:

- *Un modelo del lenguaje* que está basado en una gramática de estados finitos. En esta gramática se definen las posibles frases que serán utilizadas para interactuar con el sistema de navegación robótica y que se describen la sección 4. Con esta gramática se reconocen los 6 comandos vocales de navegación: *Avanza, Gira\_Derecha, Gira\_Izquierda, Retrocede, Inicio y Fin*.
- *Un modelo de pronunciación* formado por el vocabulario necesario para definir las frases de la gramática, donde la pronunciación se representa como la secuencia de fonemas correspondiente a cada palabra. El vocabulario de este sistema cuenta con 34 palabras en el dominio de navegación robótica.
- *Un modelo acústico* basado en modelos ocultos de Markov que fue entrenado con el corpus de voz descrito en la sección 2. Se modelaron unidades independientes del contexto para identificar los fonemas del Español hablado en México. Para el entrenamiento se usaron las librerías del Hidden Markov Tool Kit, HTK [12].

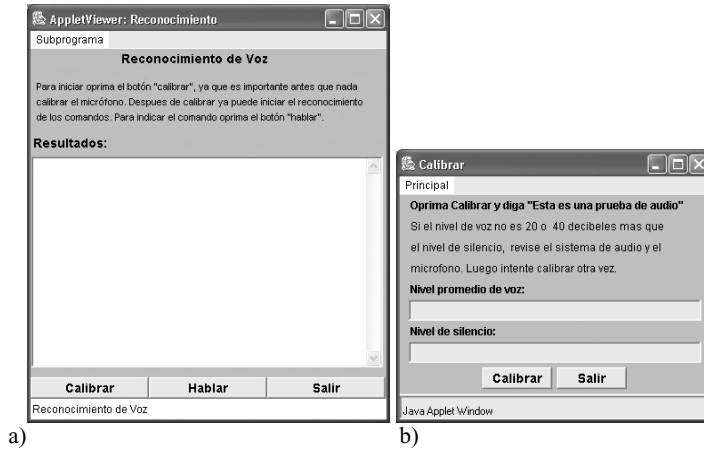
Aunque la gramática que se utiliza actualmente está limitada a reconocer frases para 6 comandos de navegación, el sistema tiene la flexibilidad para aumentar el número y complejidad de las frases a reconocer, y en caso necesario, poder agregar más comandos al sistema. Para poder utilizar un vocabulario más amplio y un lenguaje más flexible en la interacción con el sistema será necesario extender el modelo de pronunciación y el modelo del lenguaje, respectivamente.

Para evaluar que el modelo acústico del sistema de reconocimiento de voz es general e independiente del hablante, este modelo se integró en un sistema de diálogo capaz de reconocer poco más de 1000 palabras en Español [5]. El corpus de evaluación consiste de 611 oraciones con 2651 palabras y fue grabado con 8 hablantes que participaron en 4 diálogos (dos personas por diálogo) para diseñar una cocina [9]. El desempeño de este sistema de reconocimiento de voz se evaluó a nivel de oraciones y a nivel de palabras.

### 4. Interfaz con reconocimiento de voz

En la interfaz para la manipulación del robot a través de comandos hablados se utilizaron las librerías de graphVite para construir sistemas reconocimiento del habla.

Los comandos son detectados a través de un micrófono al interactuar con la interfaz de la figura 1 a). Dicha interfaz permite introducir los comandos de manera hablada y se ejecuta como un applet en un cliente de Web. El usuario al dar clic en el botón “Hablar” introduce la instrucción de navegación del robot.



**Figura 1:** a) Interfaz con reconocimiento de voz, b) Interfaz para la calibración del micrófono.

El botón “Calibrar” permite verificar el nivel de sonido entre voz y silencio para la calibración del micrófono. Al oprimir este botón aparece la interfaz de la figura 1 b). Al terminar el proceso de calibración del micrófono se oprime el botón “Salir” y regresamos a la interfaz principal.

Como se mencionó anteriormente, el botón “Hablar” es usado para introducir la secuencia de comandos para su ejecución en el robot. La secuencia de comandos reconocidos por la interfaz se pueden observar en la ventana de “Resultados”. Finalmente, el botón “Salir” nos permite terminar con la interacción con la interfaz.

La interfaz reconoce los comandos en forma de texto y estos son traducidos a su respectivo comando y parámetro para enviarlo al sistema de manipulación del robot. En la tabla 1, se presentan algunos ejemplos de esta correspondencia entre el texto reconocido y el comando y su parámetro correspondiente.

Texto	Comando	Parámetro
Avanza cuatro unidades	Avanza	4
Gira a la derecha cuarenta y cinco grados	Gira_Derecha	45
Gira a la izquierda ciento ochenta grados	Gira_Izquierda	180
Avanza hacia atrás	Retrocede	
Detente	Fin	

**Tabla 1:** Comando y su parámetro obtenido del texto reconocido

Esta interfaz fue utilizada para evaluar en vivo el desempeño del sistema de reconocimiento de voz. Cuando la interfaz con reconocimiento de voz ha reconocido un comando y su parámetro, éstos son validados por el sistema y son agregados a un arreglo. Este arreglo es enviado por medio de un socket Cliente implementado en la

interfaz a un socket Servidor desarrollado en C ANSI, este socket Servidor es el encargado de ejecutar el comando en el robot móvil.

Cabe mencionar que la interfaz sólo reconoce como comandos válidos aquellos comandos que se pronuncien con la gramática definida en el modelo del lenguaje. Es decir, si el usuario del sistema pronuncia un comando de la siguiente forma: “Gira sesenta grados a la izquierda” o “A la izquierda gira sesenta grados”, el sistema lo reconocerá como comando no válido, se tendría que haber pronunciado: “Gira a la izquierda sesenta grados” para que el sistema lo reconozca como un comando válido.

<b>Comando Pronunciado</b>	<b>%Palabras</b>
Avanza una unidad	87.79
Avanza tres unidades	96.67
<b>Avanza cuatro unidades</b>	<b>100.00</b>
Avanza cinco unidades	98.89
Avanza seis unidades	97.78
<b>Avanza siete unidades</b>	<b>100.00</b>
Avanza ocho unidades	88.89
Avanza nueve unidades	82.23
<b>Gira derecha treinta grados</b>	<b>100.00</b>
Gira derecha cuarenta y cinco grados	95.56
<b>Gira derecha noventa grados</b>	<b>100.00</b>
Gira a la derecha treinta grados	86.67
Gira a la derecha sesenta grados	91.12
Gira a la derecha ciento ochenta grados	90.48
Gira izquierda treinta grados	98.34
<b>Gira izquierda cuarenta y cinco grados</b>	<b>100.00</b>
Gira izquierda sesenta grados	95.01
Gira izquierda ciento ochenta grados	94.68
Gira a la izquierda cuarenta y cinco grados	95.84
Gira a la izquierda noventa grados	88.89
<b>Retrocede</b>	<b>100.00</b>
Avanza hacia atrás (Retrocede)	73.34
Inicio	86.67
<b>Detente (Fin)</b>	<b>100.00</b>

**Tabla 2:** Resultados al evaluar el reconocedor de voz con 5 hablantes

## 5. Resultados

La metodología utilizada para evaluar el desempeño de la interfaz con reconocedor de voz se basó en el criterio de evaluación del porcentaje de reconocimiento correcto. Este porcentaje se determinó en el desempeño del reconocimiento correcto a nivel de palabras.

Para la evaluación se realizaron 30 iteraciones para cada uno de los comandos. Esta evaluación se realizó con 5 hablantes, 3 hombres y 2 mujeres.

En general, los resultados obtenidos con el reconocedor fueron satisfactorios, ya que los niveles de reconocimiento estuvieron en el rango del 80 al 100 por ciento. En la tabla 2, se presentan algunos de estos resultados. La primera columna es el comando pronunciado y la segunda columna es el porcentaje de palabras reconocidas correctamente. El mejor desempeño del reconocedor se obtuvo para un grupo de 7 comandos cuyo porcentaje de reconocimiento de palabras fue del 100 por ciento.

## 6. Conclusiones

En este trabajo se describió el proceso realizado para la creación de un corpus de voz continua en español con las siguientes características: datos de voz en español, estos datos fueron obtenidos a partir de 1,155 frases leídas con naturalidad por 40 personas, lo cual fue suficiente para construir un reconocedor fonético independiente del locutor y de propósito general.

Los fonemas del Español mexicano se modelaron acústicamente utilizando el tipo de modelado independiente del contexto y el entrenamiento de los mismos se realizó utilizando los modelos ocultos de Markov. Se construyó un sistema de reconocimiento de voz continua en español con los modelos fonéticos entrenados. Este sistema es multilocutor y maneja un vocabulario de 34 palabras en el dominio de comandos de navegación robótica.

Se presentó una interfaz gráfica con reconocimiento de voz. El sistema de reconocimiento de voz mostró un desempeño del 80 al 100 por ciento de reconocimiento correcto de palabras del sistema de navegación robótica.

El sistema permite que no sólo se pueda percibir en la interfaz el comando reconocido a través del habla sino que a la vez se pueda observar que ese comando se ejecute en el robot móvil.

## Referencias

- [1] L. Enrique Sucar, M. A. López Trinidad, Laboratorio Virtual de Robótica Móvil. *Segundo Congreso Nacional de Robótica*, Toluca (México), 2000.
- [2] Avilés H. and Sucar E., Dynamic Bayesian Networks for Visual Recognition of Dynamic Gestures. *Journal of Intelligent & Fuzzy Systems*, Vol. 12 pp. 243-250, 2002.
- [3] Miranda-Palma C. and Mayora-Ibarra O. Robotic Remote Navigation by Speech Commands with Automatic Obstacles Detection. *Proceedings of RA 2003*, Salzburg, Austria, June 2003.



- [4] Mayora-Ibarra O., Sucar-Succar E., Aviles H. and Miranda-Palma C. From HCI to HRI: Usability Inspection in Multimodal Human – Robot Interactions. Proceedings of ROMAN'03, San Francisco, USA, November 2003.
- [5] Pineda, L.A., Massé, A., Meza, I., Salas, M., Schwarz, E., Uruga, E. y Villaseñor, L. *The DIME project*. MICAI 2002: Mexican International Conference on Artificial Intelligence. Lecture Notes in Artificial Intelligence. Vol. 2313. pp 166-175. 2002
- [6] Romero, L., *Impulsa el IIMAS la investigación en Cómputo*. Publicado el 2 de septiembre de 2002 en Gaceta UNAM. No. 3569. ISBN 0188-5138. <http://www.unam.mx/gaceta>
- [7] Ochoa, P. *Integración de un sistema de reconocimiento de voz al sistema de navegación de un robot móvil*. Tesis de Licenciatura en Ingeniería en Computación, UNAM, Diciembre 2002.
- [8] Uruga, E., Pineda, L. A., *Automatic Generation of Pronunciation Lexicons for Spanish*, Memorias del "Third International CICLing 2002: Computational Linguistics and Intelligent Text Processing". Lecture Notes in Computer Science, Vol. 2276, Editado por: Gelbukh, A. Editorial Springer-Verlag. pp 330-338. 2002.
- [9] Gamboa, C., *Un sistema de reconocimiento de voz para el español*. Tesis de Licenciatura en Informática, UNAM. Noviembre 2002.
- [10] Uruga, E., (1999), *Modelado Fonético para un Sistema de Reconocimiento de Voz Continua en Español*, Tesis de Maestría, ITESM Campus Cuernavaca, Maestría en Ciencias Computacionales.
- [11] Uruga, E., Gamboa, C., VOXMEX speech database: Design of a phonetically balanced corpus. International Conference on Spoken Language Resources and Evaluation, LREC 2004. Portugal, Mayo 2004.
- [12] Young, S. *The HTK Book*. Cambridge University Engineering Department, December 2002.

# Towards a Methodology to Envision and Evaluate Ubiquitous Computing

Víctor M. González<sup>1</sup>, Jesus Favela<sup>2</sup>, and Marcela Rodríguez<sup>2</sup>

<sup>1</sup>Department of Informatics, University of California, Irvine, CA, USA  
vmgyg@ics.uci.edu

<sup>2</sup>Departamento de Ciencias de la Computación, CICESE, Ensenada, México  
{favela, marcerod}@cicese.mx

**Abstract:** The design of ubiquitous computing solutions is intrinsically complex and should be grounded on a deep understanding of the ways that activities aimed to be supported are enacted in practice. From our experience in the development of an ubicomp medical application we have sketched a development methodology that emphasizes workplace studies to enrich requirement gathering and early, low-cost evaluation before the system is actually deployed. We plan to continue shaping this methodology with the development of new ubiquitous computing solutions.

## 1. Introduction

The inherent complexity in the development of useful and usable computer solutions, more than any other area, becomes critically accentuated in the area of ubiquitous computing (ubicomp). Proposed by Weiser in the early nineties, the ubiquitous computing vision favors a new conceptualization of the role of computers which takes them beyond particular locations (e.g. the desktop) and embeds them on the whole physical environment [10]. Ubicomp envisions a collection of computers of different capacities ranging from portable devices to digital whiteboards providing a wide spectrum of functionality. Consequently designs for ubiquitous computing must consider issues that traditional applications take for granted such as mobility of devices, synchronization of data, energy consumption, access mechanisms, information adaptation, etc. The complexity to create these systems and make them both useful and usable demands efforts at many different levels.

Among those efforts, the evaluation of ubiquitous computing environments in particular, has become an issue of considerable attention within the HCI community in recent years. In ubicomp computing becomes secondary or peripheral to the task at hand rather than the user's focus of attention, consequently evaluations has to expand their scope to consider the interactions of the individual with their whole environments and not just with single devices. Furthermore the evaluation has to be centered around whole activities rather than on individual tasks supported by some devices. Thus, traditional evaluation techniques used in HCI, based on controlled experiments where the analysis centers on constrained and isolated tasks have limited

applicability. In other hand, full deployment and testing of ubicomp technology which would make possible to have comprehensive evaluations is very complex for some kinds of applications, demands a considerable investment in infrastructure, and the definition of requirements too early in the design process. A better alternative would be to have a middle ground methodology that allow designers to evaluate and envision the opportunities of ubicomp within the context of use but does not demand a full deployment.

In this paper we present a methodological approach to design, develop, and evaluate ubiquitous computing applications. Although we focus on the healthcare domain, we believe that the methodology can be extended to other areas of application of ubicomp. We aim to define a methodology that improves the chances of the technology being adopted. Design and prototype development are based on a grounded understanding of user needs and the identification of validated scenarios of use. The evaluation phase is performed in two stages, the first one being a preliminary estimation of user intention of adoption and a second based on a contextual study with the technology in actual use. This is particularly important in ubicomp healthcare scenarios where the costs and risks associated with the implementation of new, complex technology are considerable.

## **2. Closing the gap: from envisioned designs to practical solutions**

Central to the efforts to develop ubiquitous computing solutions is the challenge of closing the gap between envisioned designs and practical solutions. As innovative as a solution might appear to designers it is within the framework of user's work practices that benefits are established. Aiming to evaluate technologies *in situ*, the concept of "living laboratories" in their study of home and classroom applications has been explored [1, 7]. In those settings real users can get in touch with technologies in environments very similar to the context for which the designs are aimed for. Without neglecting the importance of those efforts, we argue that many other situations and solutions for ubiquitous computing are less likely to be explored through those methodologies. Areas where the failure of applications will derive on financial repercussions, or threaten the life of people are upfront discarded to be explored with living laboratories. However, we would also argue that in general evaluating ubiquitous computing applications where work productivity can be affected will result on users being less willing to enact real work on living laboratories.

Even considering the complexity of evaluations, perhaps a more fundamental aspect to close the gap relates to mechanisms to envision the solution itself. There is always a danger that the envisioned solution might be too far removed from the work practices to be supported. Therefore designers, without limiting their innovation, have to be conscious of grounding their visions on the essence of what people actually do and the way they like to do it. Proponents of participatory design have for many years advocated to include users early in the projects in such a way that solutions can be developed with the constant feedback of those that will benefit from them [5]. Such scheme of feedback can result essential on the design of ubiquitous computing given the multiple effects that the paradigm would inflict on users. In contrast with desktop

computing, envisioning ubiquitous computing involves an analysis of working practices from an ecological perspective which considers how users and their environment interact and have influence over each other [8].

The direct participation of real users in envisioning and designing ubiquitous computing has started to be used in some projects. In particular the development of ubiquitous computing applications for healthcare support has been fruitfully explored with this scheme [2, 9]. Using dynamic and static scenarios, surveys, and other methods users are invited to evaluate technologies during one-day workshops or shorter evaluation meetings. There, people have the opportunity to see prototypes of emerging technologies and provide direct feedback to the designers. Those efforts in general are preceded and complemented with the use of ethnographic techniques that help consolidate initial understandings about the practices of individuals. The methodology that we present here is inspired in those previous efforts but oriented towards a more systematic consolidation of those processes.

### 3. A Methodology for the Development of Healthcare Ubicomp

Figure 1 depicts the process we followed in the implementation of an ubicomp healthcare application described in [9]. The application is a context-aware collaborative system that delivers information based on the location and role of people, the timing and the artifact of relevant artifacts, including devices and documents. The system can also be used by the hospital staff to exchange messages that depend, for their delivery, on the status of people, resources and/or devices.

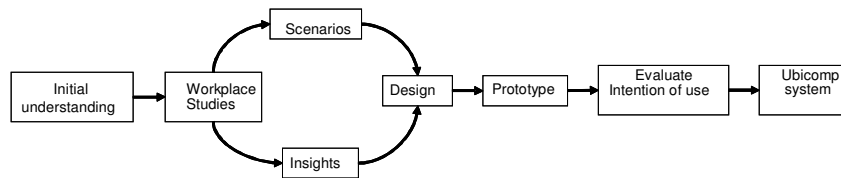


Fig. 1. The user interface of the System's Context-aware Client

We next describe the methodology illustrating it with our experience in the development of the above-mentioned system.

#### 3.1 Initial Understanding

Our efforts began with an initial understanding of the practices we wanted to support, these came from previous experience in healthcare process re-engineering and a preliminary vision of the nature of the technology we planned to develop: context-aware support for hospital processes.

### 3.2 Workplace Study

With the aim of acquiring a robust understanding of the essential contextual elements that support the management of information in a hospital setting we conducted a workplace study. We used qualitative methodologies to gain understandings beyond requirements gathering of how routine and non-routine work is performed in a daily basis. For a period of three months using participant observation and interviews we studied three areas of the hospital. Every day, an average of two hours was spent following and observing nurses, physicians, and lab staff, while doing their work. We moved around the hospital by ourselves or following a physician or a nurse. In this way we conducted many informal interviews that let us to become familiar with the setting, people and their practices.

Together, observations and interviews were used to construct an understanding of the effects of contextual elements on managing information in the hospital. We analyzed the transcripts of the interviews and notes and came out with a grounded representation of the work that let us to synthesize the particular characteristics that context-aware technologies ought to support. The workplace study provided us with two important inputs to our design: scenarios of use and insights into work practices.

### 3.3 Scenarios of Use

We decided to use scenarios to frame our understanding of hospital's work practices and also to project our vision of how their work could be augmented with context-aware tools. We used a scenario-based approach [3] to frame our findings from the study into specific *vignettes* that offer a new vision of how context aware tools can operate into the current work practices. Scenarios are a convenient approach to evaluate novel technologies [6]. One of these scenarios is the following:

*Rita, a doctor in a local hospital, makes her final round and notices that a patient is not responding well to her medication. Rita wishes to leave a note to the doctor who will be reviewing the patient in the afternoon shift. She doesn't know who that will be, so she writes a message to the first doctor to check the patient after her.*

### 3.4 Insights and Study Findings

From the study we identified characteristics that context-aware technologies should support. In particular, we identified that collaboration in the hospital is highly based on a set of contextual elements such as: (1) the location of people and devices, (2) the timing of messages to be delivered, (3) the role-oriented nature of the work and (4) the artifact-mediate nature of information gathering. For instance, access to a patient's medical records is most relevant when the doctor or nurse is with that patient. Likewise, a doctor might want to leave a message that describes recommendations for treatment to the nurse on the next shift, whom he might not know a-priori who that will be or a doctor might want to display the patient's lab analysis results in the closest public display as soon as they become available.

### 3.5 Design and Implementation of a Prototype

Scenarios and findings guided the design that led into a working prototype. We extended the Instant Messaging paradigm to support what we refer to as context-aware communication, which allows users to specify a set of circumstances that need to apply for a message to be delivered. For example, the sender can indicate that the message will be sent when the recipient enters the emergency room; or for a message to be sent to the last person to leave the laboratory when the air conditioning is on.

To illustrate the use of the context-aware messaging system we revisit the scenario presented before. As Rita, the physician in turn checks her last patient; she decides to send a message to the doctor who will be reviewing the patient in the afternoon shift. She turns to her PDA showing the Context-aware client, which lists the staff and devices available in the hospital, to send a message to the first doctor to check the patient during the next shift. As illustrated in Figure 2a, this client is able not only to notify the status of other users, but also to show resources available in the vicinity such as a printer, and their status as well as the services they can provide to the user. In addition to the information provided by an instant messaging system, the Context-aware Client also shows the location of users and devices if known. This information is shown parenthesis after the user's name as shown in Figure 2a.

When a context-aware message is created the sender specifies the criteria that need to be met for the message to be delivered. Figure 2b shows the form used by Rita to write the message and specify the context for its delivery. Rita writes the message and specifies that it should be sent to *any doctor* to be in *Room 226, after 2pm, today*.



Fig. 2. The Context-aware System provides users with awareness information of other people and devices, and lets them exchange messages that depend on context for their delivery.

Additional details on the system's functionality can be consulted in [9].

### 3.6 Preliminary Evaluation

A preliminary evaluation of ubicomp technology is important before actual deployment and investing on the required infrastructure, to evaluate the system's core

characteristics, the staff's intention to use the system, and their perception of system utility and ease of use. According to the Technology Acceptance Model [4], these aspects are fundamental determinants of system use.

The participants (28 medical staff) validated both scenarios and provided us with additional insights and opportunities for applying our technology. The results of the TAM questionnaire show that 91 percent of the participants would use the system. Additionally, 84 percent believe that using the system would enhance their job performance -a high degree of perceived usefulness- and 78 percent perceived the system to be easy to use.

### 3.7 System's Refinement and in-situ evaluation

From the results of the preliminary evaluation we have modified the system and improved its robustness. The evaluation also motivated the hospital's administration to support conducting a pilot study. We plan to deploy the system within the next few months and conduct a second field study to evaluate its benefits.

## 4. Conclusions

We present the methodology we have followed for the design of an ubicomp healthcare application. The methodology has been applied in the development of a context-aware system that supports communication and coordination of hospital staff. More recently, we have used a variant of this methodology to explore the uses of large public displays in a hospital and its integration with ubiquitous system we have already developed. The modifications consist on performing the preliminary evaluation earlier in the design phase and without a working prototype, but mockups of the technology. We are in the process of conducting a second evaluation with a working prototype in which users have the opportunity to enact scenarios with it.

## References

1. Abowd, G.D., Classroom 2000: An experiment with the instruction of a living educational environment. *IBM Systems Journal*, 1999. **38**(4): p. 508-530.
2. Bardram, J.E., Scenario-Based Design of Cooperative Systems – Re-designing an Hospital Information System in Denmark. *Group Decision and Negotiation*, 2000. **9**: p. 237-250.
3. Carroll, J., *Making Use*. 2000, Cambridge, Massachusetts: MIT Press.
4. Davis, F., Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 1989. **13**(3): p. 319-340.
5. Greenbaum, J. and M. Kyng, *Design at Work*. 1991: Lawrence Erlbaum Assoc.
6. Ikonen, V. and K. Rentto. Scenario Evaluation for Ubiquitous Computing - Stories Come True? in *Ubicomp 2002 - Workshop on User-Centered Evaluation of Ubiquitous Computing Application*. 2002. Goteborg, Sweden.
7. Kidd, C.D., R. Orr, G.D. Abowd, C.G. Atkeson, I.A. Essa, B. MacIntyre, E.D. Mynatt, T.E. Starner, and W. Newstetter. The aware home: A living laboratory for ubiquitous computing

- research. in *Proceedings of the International Workshop on Cooperative Buildings*. 1999. Pittsburgh, Pennsylvania: Spring Verlag. 191-198.
8. Kirsh, D., The Context of Work. *Human Computer Interaction*, 2001. **16**(2): p. 306-322.
  9. Munoz, M.A., M. Rodriguez, J. Favela, A.I. Martinez-Garcia, and V.M. Gonzalez, Context-aware mobile communication in hospitals. *Computer*, 2003. **36**(9): p. 38-46.
  10. Weiser, M., The Computer of the 21st Century. *Scientific American*, 1991. **265**(3): p. 94-104.



# Importancia de Modelos del Proceso Creativo en la IHM

Rafael Pérez y Pérez, Fernando Gamboa, UNAM, rpyp@servidor.unam.mx

*El presente documento pone de manifiesto la necesidad de incorporar algoritmos del proceso creativo en la IHM. Esta necesidad se vuelve particularmente relevante en sistemas que se adaptan al usuario: un sistema que es capaz de responder en formas no establecidas previamente por los diseñadores del mismo, brinda un nuevo enfoque que puede permitir resolver el problema de una adaptación eficiente. Para ello proponemos un modelo en computadora del proceso creativo y describimos un sistema de orientación vocacional que estamos desarrollando en nuestro laboratorio y cuya principal característica es que se adapta en forma creativa al usuario.*

## Abstract

*This paper points out the relevance of incorporating creative algorithms in HCI. Creativity is particularly important for adaptive computer systems: a system able to behave in ways not explicitly predefined by the designers opens new possibilities to develop more efficient adaptive systems. In this paper we introduce a computer model of creativity and describe a computer system under development to support vocational guidance and whose main characteristic is its capacity to adapt in a creative ways to the user.*

## 1. Introducción

A riesgo de establecer lo obvio, podemos afirmar que el objetivo principal del área de Interacción Humano-Computadora es el desarrollo de interfaces capaces de adaptarse al usuario y sus necesidades. Existen varios enfoques que buscan lograr este cometido: análisis del usuario y su tarea, de modo a diseñar sistemas que contemplen sus objetivos primordiales [1-6]; técnicas multimodales, capaces de “observar” al usuario y comunicarse con él (ella) utilizando un lenguaje multimedia intensivo [7-10]; interfaces “inteligentes”, que buscan modelar al usuario y su experiencia, de tal manera que puedan guiarlo de manera personalizada [1, 11, 12]; etc. A estos enfoques, que aquí denominaremos “clásicos”, se les adjunta uno que recientemente ha sido desarrollado por la comunidad de Inteligencia Artificial, de su vertiente más apegada a la Psicología Cognitiva: los modelos del proceso creativo [13-15].

Bajo este nuevo enfoque se establece que, en un sistema adaptable, el comportamiento del programa está en función de las características del usuario en turno: si todas las posibles respuestas que dicho sistema puede dar al usuario están previamente definidas, se dice que tenemos un sistema adaptable estático; en caso contrario, tenemos un sistema adaptable dinámico.

Por supuesto no se implica que un sistema dinámico sea mejor que uno estático; la elección depende de las características del problema a resolver. Por ejemplo, en un

cajero automático es posible imaginar la casi totalidad de posibles situaciones a las cuales se enfrentará un usuario. Sin embargo, existen otros contextos en lo que esto no es tan claro. Un ejemplo es la consultoría: cada caso es único y diferente a los demás. Es precisamente en estas circunstancias cuando se requieren modelos del proceso creativo que permitan dar respuestas no definidas con anterioridad. De esta manera, para los seguidores de esta nueva vertiente, un sistema adaptable dinámico deber ser capaz de dar respuestas creativas, es decir, respuestas no definidas previamente en forma explícita por los diseñadores del sistema [para una descripción de las características que debe poseer un programa de cómputo que modela el proceso creativo ver 16].

En el Laboratorio de Interacción Humano-Máquina y Multimedia del CCADET estamos desarrollando un programa en computadora para el apoyo a la orientación vocacional en la UNAM. Como característica principal, nuestro sistema puede adaptarse en forma creativa a los diferentes requerimientos de los usuarios del mismo. El sistema se basa en un modelo del proceso creativo denominado *Modelo E-R* el cual ha sido probado en otros dominios.

El presente documento está organizado de la siguiente manera: la sección 2 describe las características principales del *Modelo E-R*; la sección 3 describe las características principales del programa creativo de orientación vocacional que estamos desarrollando; la sección 4 discute las consecuencias e importancia de nuestro sistema en la IHM.

## **2. El Modelo E-R**

Pérez y Pérez [17] desarrolló un modelo para la generación de programas creativos en computadora denominado Modelo E-R. Está basado en la descripción de Sharples [13] sobre el proceso que seguimos los seres humanos al momento escribir. El Modelo E-R ha sido implementado exitosamente en un programa de cómputo para la generación de cuentos [18]. En nuestro Laboratorio actualmente se desarrollan cuatro sistemas creativos basados en el Modelo E-R: uno para la improvisación musical, otro para la solución de problemas en geometría, uno para la generación de cuentos y uno para la orientación vocacional. Los tres primeros se encuentran en una etapa avanzada de desarrollo y pronto publicaremos los resultados obtenidos. Como el lector puede observar, el modelo puede ser empleado en diferentes dominios. El Modelo E-R contempla tres tareas básicas:

- Definición de experiencias.
- Transformación de dichas experiencias en estructuras del conocimiento.
- Generación de material novedoso e interesante a través de ciclos E-R.

### **2.1 Definición de experiencias**

El Modelo E-R requiere que se definan por parte del usuario una serie de experiencias (o casos previos) exitosos. Por ejemplo, para la generación de cuentos dichas experiencias están formadas por secuencias de acciones que ejemplifican historias construidas adecuadamente; para la solución de problemas, los casos previos los

constituyen secuencias de acciones utilizadas para resolver exitosamente diversos problemas de geometría; etc. En el Modelo E-R, a cada una de las acciones que forman el conjunto de experiencias o casos previos se les denominan Acciones Primarias; cada una de ellas tiene asociadas precondiciones y poscondiciones.

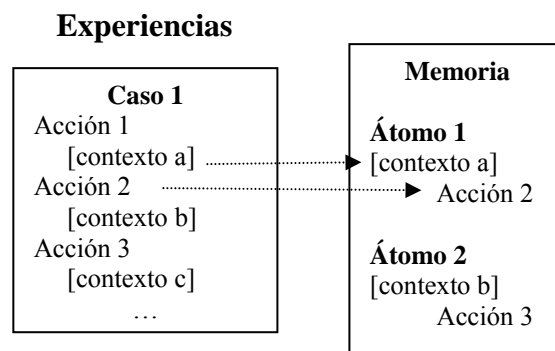
Precondiciones → Acción Primaria → Poscondiciones

Las precondiciones constituyen los requerimientos que deben ser satisfechos para que una acción se pueda ejecutar. Las poscondiciones constituyen las consecuencias de ejecutar una acción; existe una estructura conocida como *contexto*, la cual registra todas las poscondiciones que se activan cada vez que se ejecuta una acción primaria. De esta manera, el contexto representa el estado actual del mundo virtual creado por el sistema de cómputo. Normalmente, las precondiciones y poscondiciones están formadas por relaciones entre los diversos actores de un dominio dado: personajes en un cuento; puntos, círculos, rectas en la solución de problemas de geometría; notas en la improvisación musical; etc.

## 2.2 Estructuras del Conocimiento

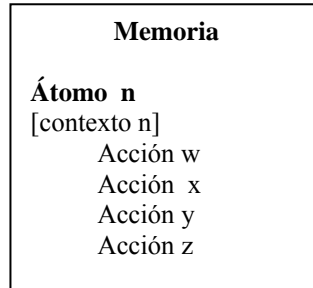
Para crear las estructuras del conocimiento en memoria es necesario procesar caso por caso y acción por acción el conjunto de experiencias. Para ello se siguen los siguientes pasos (ver figura 1):

1. Se obtiene una acción del conjunto de experiencias y se ejecuta; las consecuencias o poscondiciones de dicha acción generan un nuevo contexto.
  2. Dicho contexto se copia a memoria en una nueva estructura denominada átomo.
  3. La siguiente acción en las experiencias se asocia a dicho átomo.
  4. Se regresa al paso 1 hasta que las acciones que forman el caso terminen.
- Este proceso se repite para cada uno de los casos que conforman las experiencias o casos previos.



**Figura 1. Cómo se crean los átomos en memoria.**

Cuando el número de experiencias o casos previos son los suficientemente representativos, es posible tener en memoria grupos de varias acciones asociados a cada átomo (ver figura 2).



**Figura 2. Representación de un átomo con su contexto y varias acciones asociadas.**

### 2.3 Ciclos E-R

Una vez que se han creado las estructuras del conocimiento en memoria, el sistema ejecuta un proceso cíclico entre dos modos de operación conocidos como estado-E y estado-R.

Estado-E. Cuando el sistema trabaja en el estado-E emplea todos sus recursos para recuperar de memoria acciones primarias que le permitan llevar a cabo la tarea asignada. Funciona de la siguiente manera:

- se escoge una acción.
- el sistema la ejecuta y sus poscondiciones generan un nuevo contexto.
- el sistema busca en memoria un átomo igual al nuevo contexto.
- recupera todas las acciones asociadas a ese átomo, escoge una y vuelve a comenzar el ciclo.

Durante este proceso el sistema ignora por completo las precondiciones; sólo emplea el contexto producido por cada acción como guía para encontrar un átomo y poder así recuperar de memoria más acciones para continuar su tarea. Si el sistema no puede localizar en memoria un átomo igual al contexto actual, emplea diferentes estrategias para encontrar al menos uno similar. La primera consiste en buscar un átomo que contenga como parte de su estructura el contexto; la única restricción es que el átomo y el contexto sean iguales en al menos un porcentaje previamente determinado. A esta estrategia se le conoce como Contexto Dentro del Átomo (CODA). La segunda consiste en buscar un átomo cuya estructura se encuentre dentro del contexto; la única restricción es que el átomo y el contexto sean iguales en al menos un porcentaje previamente determinado. A esta estrategia se le conoce como Átomo Dentro del Contexto (ADCO). La tercera consiste en Modificar el Contexto (MOCO) agregando, eliminando o alterando elementos del mismo. Esta estrategia requiere que las características esenciales del contexto se conserven después de cada transformación. De otra manera, las acciones recuperadas de memoria tendrán muy poca conexión con el estado actual del mundo virtual creado por el sistema.

Una vez que se han recuperado de memoria un conjunto de posibles acciones a ejecutar, es necesario eliminar todas aquellas que no satisfagan cabalmente los requerimientos de la tarea. De ello se ocupan una serie de rutinas llamadas filtros. Existen dos tipos de constricciones asociadas a los filtros: 1) las constricciones estáticas, que están definidas dentro del código del programa (por ejemplo, en el desarrollo de un cuento está prohibido que un personaje ejecute dos veces seguidas la misma acción. Las constricciones estáticas le indican a los filtros que este tipo de situaciones se deben evitar); 2) las constricciones dinámicas, cuyo valor se actualiza constantemente mientras el programa se está ejecutando (por ejemplo, si durante el desarrollo de un cuento el sistema detecta que el mismo está resultando aburrido, las constricciones dinámicas le indican a los filtros que deben eliminar todas aquellas acciones que no contribuyan a hacer una historia más interesante). De esta manera, si alguna de las acciones recuperadas de memoria no satisface todas las constricciones estáticas y dinámicas, los filtros se ocupan de eliminarla. El funcionamiento del estado-E se puede resumir en los siguientes pasos:

1. Se ejecuta una acción la cual produce un nuevo contexto.
  2. El sistema busca en memoria un átomo que sea igual al nuevo contexto.
  3. Si la búsqueda es exitosa
    - Recupera de memoria el conjunto de posibles acciones a ejecutar.
    - Elimina por medio de los filtros todas aquellas acciones no útiles.
    - Entre las acciones restantes selecciona una y regresa al paso 1.
- de lo contrario
- Intenta la estrategia CODA; si la búsqueda es exitosa regresa al paso 3.
  - Intenta la estrategia ADCO; si la búsqueda es exitosa regresa al paso 3.
  - Intenta la estrategia MOCO; si la búsqueda es exitosa regresa al paso 3.
  - El sistema se declara bloqueado (no puede recuperar acciones para continuar su tarea).

Estado-R. Cuando el sistema trabaja en el estado-R:

- Verifica que todas las precondiciones de las acciones primarias empleadas hasta ese momento estén satisfechas. En caso necesario inserta acciones para cumplir con este requerimiento.
- Evalúa el material producido durante el estado-E para revisar que cumpla con los requerimientos de la tarea asignada. Si la evaluación resulta negativa, el sistema genera una serie de lineamientos (llamados constricciones dinámicas) que los filtros emplean durante la generación de material en el estado-E para eliminar aquellas acciones que no satisfagan los requerimientos de la tarea.
- En caso de que se declare un bloqueo durante el estado-E, se activan una serie de rutinas para intentar romper dicho bloqueo.

El desarrollo de cuentos novedosos e interesantes, la solución de problemas de geometría, la producción de improvisaciones musicales, etc., son el resultado de una estrecha interacción entre el estado-E y el estado-R.

### **3. Programa Creativo de Orientación Vocacional**

El Programa Creativo de Orientación Vocacional (PROCOV) es un sistema capaz de dar respuestas creativas —es decir, respuestas no definidas previamente en forma explícita por los diseñadores del sistema— a un usuario. El objetivo de PROCOV es apoyar a estudiantes de bachillerato en el proceso de toma de decisión de la carrera a estudiar en la universidad. El sistema actúa como un orientador que, por medio de una serie de preguntas, se va aproximando a el perfil del usuario. Si es necesario, PROCOV puede solicitar desarrollar alguna actividad previamente definida con el fin de obtener más datos del usuario. El sistema inclusive puede sugerir contestar alguna prueba psicológica para completar sus datos. Una vez que el sistema obtiene el perfil del usuario genera una recomendación vocacional. El conjunto de preguntas, actividades y pruebas psicológicas que lleva a cabo el sistema constituyen la *estrategia de orientación* empleada por PROCOV para asesorar al estudiante. El sistema cuenta con un conjunto de experiencias o casos previos, proporcionados por expertos orientadores de la UNAM, que funcionan como base para llevar a cabo la asesoría. Como característica principal, cuando PROCOV se enfrenta a una situación no contemplada en dichas experiencias, es capaz de generar una estrategia de orientación nueva que le permita afrontar adecuadamente la situación.

La recomendación vocacional se presenta al usuario en forma de secuencias de imágenes y textos. PROCOV tiene como objetivo que dicha presentación sea original e interesante para que el usuario se identifique con ella. Por original nos referimos a que una persona no vea la misma presentación dos veces; por interesante nos referimos a que la presentación capture la atención del usuario.

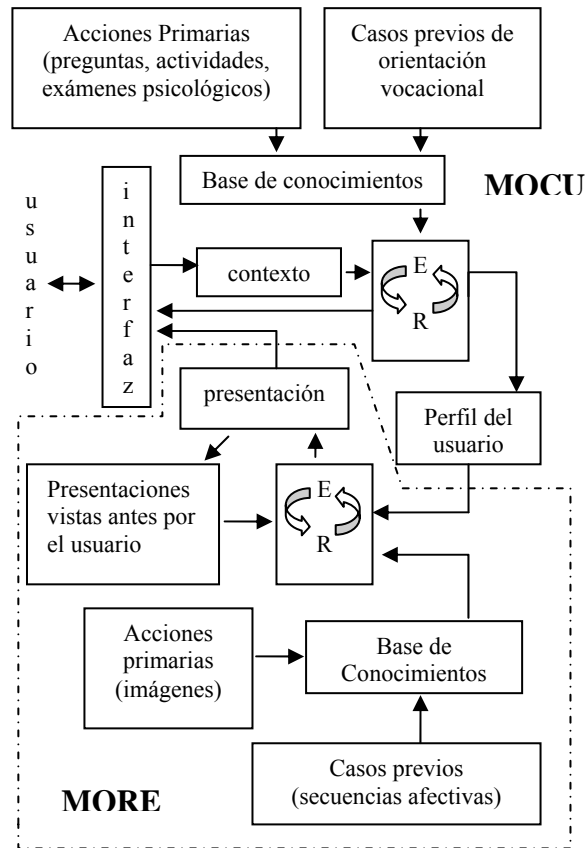
De esta manera, PROCOV constituye un programa capaz de adaptarse en forma creativa a los requerimientos de un usuario. PROCOV está dividido en dos módulos principales: Conocimiento del usuario y Recomendaciones.

#### **3.1 Módulo de Conocimiento del Usuario**

En el módulo de conocimiento del usuario (MOCU) es donde el sistema genera su estrategia de orientación para así poder establecer un perfil del usuario (ver figura 3). Con base en el modelo E-R, el MOCU requiere de un conjunto de acciones primarias y de un grupo de experiencias o casos previos. En el MOCU las acciones primarias están divididas en tres grupos:

- Preguntas.
- Actividades diversas.
- Aplicación de exámenes psicológicos.

Cada acción tiene asignada sus precondiciones y poscondiciones; cada vez que el sistema ejecuta una de estas acciones, la estructura representando el contexto se actualiza.



**Figura 3. Diagrama de PROCOV: la parte superior corresponde al Módulo de Conocimiento del Usuario (MOCU); la parte inferior corresponde al Módulo de Recomendaciones (MORE).**

Las preguntas pueden ser de varios tipos. Con base en el libro *Eligiendo mi carrera. Un proyecto de vida* publicado por la Dirección General de Orientación y Servicios Educativos de la UNAM [19], los temas que exploran las preguntas están relacionados con: desarrollo de un proyecto de vida, elementos que intervienen en la elección de carrera, conocimiento sobre diversas áreas académicas, conocimiento de sí mismo, valores personales y profesionales, intereses y aptitudes, perfil profesional y campos ocupacionales. Ejemplos de preguntas son: ¿cuánta información necesitas para decidir qué carrera elegir?, ¿cuáles consideras los factores más relevantes que se deben tomar en cuenta al momento de elegir una carrera?, etc. Por medio de estas preguntas el sistema genera un perfil del usuario. El tipo de respuestas puede llevar al sistema a realizar más preguntas, sugerir realizar actividades diversas o exámenes psicológicos. Las actividades diversas pueden consistir en juegos, actividades de reflexión, de investigación, etc. Por ejemplo, realizar una gráfica de

intereses y aptitudes, hacer una lista de las materias que le agradan al estudiante y las calificaciones que ha obtenido, hacer una lista de las materias que no le agradan al estudiante y las calificaciones que ha obtenido, comparar las listas, etc. A través de los exámenes psicológicos se obtiene información del usuario referente a sus habilidades, personalidad, intereses y hábitos de estudios. De esta manera, la estrategia de orientación se va creando al mismo tiempo que el sistema interactúa con el usuario.

La experiencia o casos previos son proporcionados por orientadores expertos de la UNAM. Esta experiencia consiste en ejemplos, paso a paso, de un proceso de orientación vocacional. De esta manera el sistema aprende cuándo realizar determinadas preguntas, cuándo proponer alguna actividad o cuándo aplicar un examen psicológico.

### **3.2 Módulo de Recomendaciones**

Una vez establecido el perfil del estudiante, el módulo de recomendaciones (MORE) genera una sugerencia de tres carreras a estudiar (ver figura 3). Dicha sugerencia toma la forma de una secuencia de imágenes, principalmente fotografías, y textos que se presentan al usuario. La presentación tiene tres objetivos básicos:

- Proporcionar información básica sobre la carrera que se está recomendando.
- Captar la atención del usuario. Para ello, se emplean secuencias de fotografías y textos que generen en el usuario una respuesta afectiva. Para lograr este objetivo primero trabajamos con dos artistas en diseño y comunicación visual y un experto en psicología, con quienes se establece los requerimientos necesarios para generar una respuesta afectiva por medio de secuencias fotográficas y textos. A continuación realizamos experimentos para comprobar que dichas secuencias generan los efectos deseados.
- Generar presentaciones creativas. Para mantener el interés del usuario en el sistema es necesario que las presentaciones sean novedosas. Es decir, una misma persona no debe ver la misma presentación dos veces, tampoco debe ver dos presentaciones muy parecidas, etc. Por tal razón, el MORE también está diseñado siguiendo el modelo E-R: cada imagen representa una acción primaria y tiene asociada sus respectivas precondiciones y poscondiciones; las experiencias están formadas por ejemplos de secuencias que generan respuestas afectivas en el usuario.

### **4. Discusión**

En este artículo hemos expuesto la posibilidad de integrar nuevos modelos a los mecanismos de interacción tradicionalmente utilizados por la comunidad de IHM. En particular se ha expuesto la pertinencia de incluir modelos del proceso creativo como un medio para potenciar la capacidad de adaptación de un sistema ante un nuevo usuario. En efecto, los modelos de procesos creativos permiten que el sistema tome decisiones, de manera dinámica, acerca de la forma idónea de presentar información al usuario. Se presentaron además los avances que se tienen con un sistema de consultoría vocacional, que integran estas técnicas. Los resultados obtenidos con el uso de dicho sistema en el terreno serán reportado en publicaciones posteriores.



## 5. Referencias

1. Balbo, S. and C. Lindley. *Adaptation of a task analysis methodology to the design of a decision support system*. in *INTERACT'97*. 1997. Sydney, Australia: Chapman and Hall, IFIP publishers.
2. Alonso, B., *Analyse des tâches mono et multi-opérateurs du contrôle aérien par le formalisme MAD pour l'espécification ergonomique de l'interface*, in *Ergonomie*. 1996, Paris 5: Paris. p. 325.
3. Gamboa, F. and D.L. Scapin, *Editing MAD\* task descriptions for specifying user interfaces at both, semantic and presentation levels*, in *Design, Specification and Verification of Interactive Systems '97*, M.D. Harrison and J.C. Torres, Editors. 1997, Springer-Verlag/Wien: Austria. p. 193-208.
4. Tarby, J.C. and M.F. Barthelet. *From Tasks to Objects: an example with Diane+*. in *CHI' 98*. 1998. Los Angeles.
5. Dominguez, G., *Herramienta para la vinculación del Análisis de la Tarea del Usuario con el Proceso Unificado*, in *Posgrado en Ciencia e Ingeniería de la Computación*. 2003, UNAM: México D.F.
6. Hackos, J.T. and J.C. Redish, *User and Task Analysis for Interface Design*. 1st ed. 1998, New York, Chichester, Weinheim, Brisbane, Singapore, Toronto: John Wiley & Sons, Inc. 488.
7. Lansdale, M.W. and T.C. Ormerod, *Understanding interfaces. A handbook of Human - Computer Dialog*. 1st ed. 1994, London San Diego New York: Academic Press. 289.
8. Andleigh, P.K. and K. Thakrar, *Multimedia Systems Design*. 1st ed. 1996, Upper Saddle River, NJ 07458: Prentice Hall PTR. 654.
9. Brice, R., *Multimedia and Virtual Reality Engineering*. 1st ed. 1997, Oxford: ewnes. 307.
10. Norman, D.A., *The Invisible computer*. 1998, Cambridge, Massachusetts; London, England: The MIT Press. 302.
11. Opperman, R., ed. *Adaptive User Support. Ergonomic Design of manually and Automatically Adaptable Software*. 1st ed. Computers, Cognition, and Work, ed. G.M. Olson, J.S. Olson, and B. Curtis. 1994, Lawrence Erlbaum Associates, publishers: New Jersey, NK. 253.
12. Salvendy, G. and M. Smith, eds. *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*. 1st ed. Advances in Human Factors / Ergonomics, ed. G. Salvendy. 1989, Elsevier: Amsterdam-oxford.New york. 900.
13. Sharples, M., *How we Write: Writing as Creative Design*. 1999, London: Routledge.
14. Dartnall, T., *Creativity, Cognition, and Knowledge*. 2002, Westport, Connecticut, London: PRAEGER.
15. Dartnall, T., *Artificial Intelligence and creativity: An interdisciplinary approach*. 1994, Dordrecht, Netherlands: Kluwer Academic Publishers.
16. Pérez y Pérez, R. and M. Sharples, *Three Computer-Based Models of Storytelling: BRUTUS, MINSTREL and MEXICA*. Knowledge Based Systems Journal, 2004. **17**(1): p. 15-29.
17. Pérez y Pérez, R., *MEXICA: a Computer Model of Creativity in Writing*, PhD. dissertation. 1999, University of Sussex, England.

18. Pérez y Pérez, R. and M. Sharples, *MEXICA: A Computer Model of a Cognitive Account of Creative Writing*. Journal of Experimental and Theoretical Artificial Intelligence, 2001, **13**(2): p. 119-139.
19. Dirección General de Orientación y Servicios Educativos, *Eligiendo mi carrera. Un proyecto de vida*. 2002, México: Universidad Nacional Autónoma de México.

## **Diseño e Implementación de una Interfaz de Comunicación en Internet, Orientada a Personas con Discapacidad Visual**

<b>Noemí Guerra</b> <i>Estudiante de la Facultad de Telemática</i>	<b>Román Gallardo</b> <i>Profesor de la Facultad de Telemática</i>	<b>María Andrade</b> <i>Profesora de la Facultad de Telemática</i>	<b>Ricardo Acosta</b> <i>Profesora de la Facultad de Telemática</i>
---	---	---	--

*Universidad de Colima, Colima, Col., México*

<i>nlg22@ucol.mx</i>	<i>aroman@ucol.mx</i>	<i>mandrad@ucol.mx</i>	<i>acosta@ucol.mx</i>
----------------------	-----------------------	------------------------	-----------------------

### **Resumen**

La carencia de alguna de las capacidades físicas de un individuo impide el pleno desarrollo de las diversas actividades que el mismo realiza, esto se traduce en una ***desventaja para la persona***. Debido a esta situación, el presente trabajo es un estudio que expone la problemática que existe en la sociedad invidente de la Organización **CREE\*** del estado de Colima, México, en cuanto a las necesidades de comunicación de mensajería instantánea a través de la red Internet, que tienen dichas personas; ya que actualmente las aplicaciones para personas que presentan discapacidades visuales son de tipo general y es muy limitada su distribución.

Por lo tanto, al implementar los conocimientos de las telecomunicaciones aunadas a la información (Telemática) y un buen diseño de la interfaz de comunicación especializada apoyada en las nuevas tecnologías de *interacción humano-computadora*, permitirá que estas personas desarrollen mejor sus habilidades de comunicación e integración en la sociedad actual.

### **Abstract**

The deficiency of some of the physical capacities of an individual prevents the total development of the diverse activities that the same one makes, this is translated in a disadvantage for the person. Due to this situation, the present work is a study that exposes the problematic one that exists in the blind society of the **CREE** Organization in the Colima State, México, as far as the necessities of instant message communication through network Internet, which they have these people; since at the moment the applications for visually impaired people are of general type and is very limited their distribution. Therefore, when implementing the knowledge of the telecommunications combined to the information (Telematic) and a good design of the interface of communication specialized supported in the new technologies of

---

\* CREE (Centro de Rehabilitación y Educación Especial).

interaction human-computer, will allow that these people develop better to their abilities of communication and integration in the present society.

**Palabras clave:** Interfaz, Web, HCI, TTS, Telemática.

## 1. Introducción

Gran porcentaje de la información que manejamos llega a través de la vista, por ello las personas con ceguera y/o deficiencia visual han tenido que buscar los medios y modos alternativos de acceso a esta, para que la falta de visión no represente más limitaciones de las estrictamente necesarias. Cuando la ceguera es total, el oído y el tacto pasan a ser los principales canales en la recepción de la información, mientras que para las personas con deficiencia visual el resto de visión que poseen es un recurso más a utilizar [1]. Por lo tanto, el lenguaje es una forma convencional utilizada para la comunicación de pensamientos y sentimientos entre los seres humanos, el cual es representado por símbolos ya sean escritos o hablados. La voz es parte integral de nuestras vidas y las personas con capacidades físicas y mentales normales e incluso quienes presenten algún tipo de discapacidad, en este caso visual, utilizan la voz como el medio principal de comunicación y una de sus grandes diferencias con respecto al lenguaje escrito, es que el habla permite comunicar necesidades inmediatas. Analizando lo anterior, vemos las dificultades que se pueden encontrar en la red Internet ya que este es un medio inestimable de acceso a la información para personas con discapacidad visual, aunque para que ellas obtengan los beneficios y la accesibilidad que brinda la misma, es necesario lograr la estandarización de los sitios web. El enfoque de estandarización se refiere a que por lo general todas las páginas en Internet están actualizando sus formatos y su información cada cierto tiempo; por ejemplo los sitios web de: periódicos, canales de noticias, finanzas, clima, entre otros; tienen que actualizar su información diariamente y de hecho, existen sitios en Internet que cambian su información cada mes y algunas nunca las vuelven a actualizar desde su creación [2]. Aunque en la red existen guías que garantizan la accesibilidad a los sitios web construidos según sus directrices y se han desarrollado herramientas de software destinadas a proporcionar una mayor y mejor accesibilidad a estas personas, la mayoría trata el problema desde su propio punto de vista, en lugar de apoyarse en el conocimiento de expertos en el diseño de interfaces para usuarios discapacitados. A partir de esta problemática surgió la idea por llevar a cabo un proyecto que resolviera las necesidades de comunicación de mensajería instantánea de la población con discapacidad visual. La aplicación propuesta es una herramienta que le permite a usuarios invidentes entablar una comunicación en línea (mensajería instantánea) con la funcionalidad de conversión TTS (Text To Speech) en idioma español. La misma ofrece una ayuda guiada a través de voz, sonido, manejo de las teclas de función, las cual le permiten al individuo poder interactuar de forma sencilla con la misma y se ha sido denominado **sistema de comunicación para invidentes (SCI)**. El aspecto más característico de este proyecto consiste en procesar la información de entrada, ya que los mensajes que son enviados por el emisor se encuentran en formato texto y son transformados y enviados en voz al canal de recepción. Además de esta funcionalidad de conversión text to speech, los

mensajes también aparecen de manera escrita en la ventana de conversación del usuario final, como método de apoyo a la persona que presente un grado de ceguera parcial.

## 2. Revisión de la Literatura

En este nuevo siglo, las computadoras y su uso cotidiano son algo que no puede pasar inadvertido. Debido a esto, cualquier tecnología que se desarrolle en este ámbito tendrá una importancia relevante en nuestra sociedad, ya que día a día surgen nuevas herramientas, programas y equipos, que nos permiten desarrollar e implementar sistemas, que hagan frente a la problemática y necesidades de todo tipo de usuario. Actualmente, existen en el mercado muchos dispositivos que apoyan a los discapacitados visuales en el uso y manejo de la computadora como herramienta de aprendizaje, pero su desventaja es que la mayoría de ellos son muy caros; entre estos productos podemos mencionar: máquinas lectoras para ciegos, enseñanza asistida por ordenador, terminales parlantes, teclados tipo Braille, ratón o mouse para invidentes, por mencionar algunos [3]. Además, se han creado software de diversos tipos para personas que presenten estas características; por mencionar algunos existen programas de: entretenimiento, videojuegos, lectura de pantallas (Screen Readers) [9], reconocimiento de voz y texto, sitios web de bibliotecas digitales [10]. Pero aún con estas innovaciones se considera que no se ha aprovechado al máximo los recursos que ofrece la red internet para apoyar más a los invidentes; ya que se considera que es un descuido de parte de los webmasters la no creación de sitios más *estándares* porque el hecho de que una persona presente esta discapacidad no le impide beneficiarse de la información y las nuevas tecnologías, a la cual todos tenemos derecho. Por otro lado, ya que internet actualmente es considerada una de las herramientas más importantes de comunicación entre las personas a nivel mundial, es preciso que estas personas tengan a su disposición aplicaciones de comunicación especializada; pero la realidad es que éstos son proyectos que aún se encuentran en vías de desarrollo y las que existen no están disponibles para los usuarios generales. En la actualidad, los conversores texto a voz están alcanzando un nivel de calidad que posibilita su uso en aplicaciones tanto personales como dirigidas al público en general. Esta tecnología va aunada al reconocimiento de voz el cual ofrece ventajas tales como: una manera más rápida de introducir datos, comodidad al no tener que estar utilizando el teclado o el ratón (mouse) para la captura de información, no es necesario utilizar el sentido visual, porque no se requiere estar de manera permanente delante de un monitor y además permite la realización de otras actividades manuales o visuales mientras se utiliza la computadora [4]. Además, para la elaboración de este proyecto se hizo un estudio acerca de las metodologías y conceptos de la *interacción humano computadora*, disciplina que contiene todos los parámetros requeridos para diseñar sistemas *eficientes, usables y funcionales*. La interacción humano-computadora (HCI) maneja un concepto clave denominado *usabilidad*, el cual consiste en hacer sistemas fáciles de utilizar y de aprender [5]. Para obtener una aplicación que cumpla con estos aspectos, se deben entender, analizar y evaluar los factores que determinan cómo la gente opera las tecnologías, para entonces proceder a desarrollar herramientas y técnicas que garanticen que el sistema va a realizar las actividades para las que ha

sido diseñada [5], y que a la vez proporcione una interacción eficiente, efectiva y segura para aquellos usuarios que presenten algún tipo de discapacidad. Es por esto que con el avance de la tecnología se ha hecho necesario desarrollar software que permita a personas con discapacidades tener acceso a la información a través de los sistemas de cómputo, ya que existe mucha información en la red, pero lamentablemente no se encuentra al alcance de todas las personas, por lo que este caso de estudio está orientado al estudio e implementación de una interfaz de comunicación en línea para personas invidentes; por lo tanto se puede decir que la necesidad por entablar dicha forma de comunicación se está incrementando [2].

### 3. Solución propuesta

El sistema de comunicación para invidentes aborda el área de la comunicación a través de la conversión *texto a voz* en idioma español, implementado en una interfaz que consiste en un mensajero instantáneo. Esta herramienta ha sido desarrollada para proporcionar a la población con discapacidad visual y a las personas que lo requieran, la oportunidad de comunicarse en línea con otras que presenten las mismas dificultades o con cualquier persona que se encuentre en la red. Dicho proyecto tomó como caso de estudio y aplicación a la organización **CREE** del estado de Colima, institución que fue creada con la misión de prestar atención y servicios de incorporación integral médica, educativa y profesional a la población con discapacidad. Los objetivos específicos desarrollados en este trabajo consistieron en documentar y evaluar los software de programación que actualmente existen para desarrollar aplicaciones de conversión de texto-voz en idioma español, los cuales representan la herramienta básica para llevar a cabo este proyecto. Además se analizaron las metodologías adoptadas por las teorías de Interacción Humano-Computadora (HCI) y se aplicaron las directrices que describen como realizar el desarrollo de aplicaciones para personas que presentan algún tipo de discapacidad [5]; después se procedió a analizar, diseñar y crear una interfaz de comunicación accesible, eficiente, y amigable para dichos usuarios, el cual fue evaluado mediante una prueba piloto aplicada a una muestra de la población con discapacidad visual del *Centro de Tecnología Adaptada* de la Organización **CREE**, quienes son los usuarios directos del mismo [6]. Dicha área ha sido destinada a brindar a la población con discapacidad y/o necesidades educativas especiales, las herramientas de cómputo necesarias que les permitan una rehabilitación más rápida a través del uso clínico o pedagógico de éstas. La población con discapacidad visual que se atiende en el CREE asciende a 43 personas, las cuales están afiliadas a la Organización de Ciegos Colimenses y asisten a dicho centro de rehabilitación porque el mismo les ofrece la infraestructura y tecnologías necesarias para su desarrollo [14]. En otros términos, la problemática que afronta el CREE ante los invidentes gira entorno a necesidades educativas, de capacitación y de orientación en el plano social y laboral (ver Tabla 1).

Necesidades de los Invidentes	Metas del Centro de Tecnología Adaptada del CREE	Porcentaje (%) de invidentes que reciben estos servicios
Educación especial	Respaldar el acceso a niveles superiores de preparación académica utilizando la computadora como herramienta principal de apoyo y desarrollar sus habilidades con el manejo de diversos software diseñados y donados al centro por la IBM.	51.16
Capacitación	Con la ayuda de un especialista y el manejo de los programas recomendados, el paciente o alumno tendrá la posibilidad de lograr con mayor eficacia su capacitación y/o rehabilitación.	18.61
Orientación (social y laboral)	A través de la comunicación obtenida mediante métodos especiales, se debe mejorar la interacción del paciente con las personas y actividades que lo rodean, para que el mismo pueda integrarse plenamente al ámbito social y laboral.	30.23

**Tabla 1. Necesidades específicas de los invidentes del CREE.**

El hombre necesita comunicarse por naturaleza, y esto es lo que le ha llevado a buscar nuevas formas de comunicación durante toda su existencia, desde los sistemas Morse pasando por la radio y el teléfono, los cuales en síntesis constituyen el origen de lo que actualmente se conoce como las **“autopistas de la información”**. Esta esencia comunicativa, junto al constante desarrollo tecnológico, abaratamiento y variedad existente en los equipos informáticos, han aportado grandes ventajas a la humanidad, ya que la computadora, dispositivos periféricos, software o aplicaciones y las redes telemáticas han llegado a formar parte esencial en casi todos los aspectos de la vida diaria. Esta nueva generación al acceso, manejo, emisión y recepción de todo tipo de información se encuentra inmersa bajo el concepto de HCI. La tecnología actual, intenta que dicha interacción del hombre con las máquinas, sea lo más natural posible en el sentido de posibilitar una comunicación oral con las mismas, y así obtener un mayor acercamiento por parte de ambos elementos; lo que no es una tarea demasiado sencilla, cuando nos referimos a personas que no están habituadas a la informática o que ven a esta rama de estudios como algo imposible por su dificultad en el entorno visual. En consecuencia, el **SCI** se basa en las teorías HCI y en las modernas técnicas de conversión texto-voz, con el objetivo de obtener una comunicación e interacción con diversas personas de manera ordenada, sofisticada y eficiente. Para los discapacitados visuales resulta imposible la utilización de ratones y dispositivos de señalamiento, por esta razón el teclado representa el único medio disponible para dar instrucciones a la computadora; y su uso cotidiano permite adaptarse fácil y rápidamente a su entorno. Además, los teclados de computadora cuentan con un relieve en las teclas F, J y 5; lo cual le permite a estas personas tener un marco de referencia para así incrementar las destrezas en su manejo [13]. Basado en estos análisis se decidió utilizar las teclas de función como patrón de diseño para *acceder más rápidamente a las principales funciones del sistema de síntesis de voz*, ya que de esta manera se ofrece al invidente un camino directo y específico sobre el cual se desenvolverá de forma amigable e independiente en el mismo. Por lo tanto, una delimitación de este proyecto radica en que los usuarios (discapacitados visuales)

que utilicen esta interfaz deben tener conocimientos básicos de informática y saber manejar el teclado, para que puedan acceder sin mayor dificultad al sistema operativo de la computadora y por consiguiente, directamente al sistema. Dicha interfaz guía al usuario mediante una serie de instrucciones habladas que facilitan al mismo su interacción con la herramienta, ya que ha sido desarrollada para que sea una aplicación sencilla y amigable para estas personas. El análisis y diseño del Sistema de Comunicación para Invidentes se basó en una metodología de diseño orientada a objetos conocida como UML (Unified Modeling Language), el cual es un lenguaje con una notación estándar internacional formado de una serie de reglas y múltiples elementos gráficos que se combinan para conformar notaciones y diagramas estándares, permitiendo construir modelos precisos y completos; además de describir la semántica esencial de lo que éstas representaciones y símbolos significan, por lo que es considerado el proveedor de los *planos del software* [7].

Básicamente, las aportaciones y beneficios que el sistema propuesto ofrece a los invidentes con respecto a otros esfuerzos son notables (ver Tabla 2).

Sistemas de Comunicación utilizados por Invidentes	Costo de adquisición	Manejo de la interfaz (a través de:)	Costo estándar de los sistemas de acceso a la interfaz
MSN Messenger [11]	gratuito	Jaws Screen Reader	796.87 USD [9]
Mirc [12]	20 USD [12]	Window Eyes Screen Reader	517.63 USD [9]
Freetel [12]	gratuito	Look Out Screen Reader	115.71 USD [9]
Netscape Messenger [12]	gratuito	Hal Screen Reader	302.88 USD [9]
SCI	gratuito	Sistema de Síntesis de Voz (TTS)	Libre de costo (Servicio adaptado al SCI) [8]

Tabla 2. Ventajas del SCI con respecto a otros sistemas de mensajería instantánea y chat.

El desarrollo de la herramienta de comunicación **SCI**, se basa en los parámetros de la conversión texto-voz, pero apoyado en los nuevos kits o paquetes de desarrollo que ofrece Microsoft, para la implementación de aplicaciones TTS. Al trabajar con esta tecnología la voz sintetizada que se obtiene en el canal de recepción presenta mayor calidad, ya que estas herramientas de programación han sido sofisticadas entorno a aspectos tales como la entonación, expresión de emociones y otros parámetros que proporcionan naturalidad a la voz obtenida. El desarrollo de esta aplicación se ha basado en las siguientes herramientas:

- ❖ Microsoft Agent Ring”, agentes de sonido [8].
- ❖ Microsoft ActiveX Controls, Text To Speech.

los cuales trabajan bajo el entorno de la herramienta de programación Microsoft Visual Basic 6.0. En consecuencia, el modo de operación del sistema consiste en lo siguiente: cuando el usuario va a ingresar a la aplicación, automáticamente la interfaz le brinda de forma oral todas las indicaciones de cómo acceder a cada una de las funciones que realiza la misma, entre las cuales se encuentra las opciones de:

- conectarse a la aplicación.



- establecer comunicación con un usuario mediante la selección o búsqueda del contacto.
- conocer la lista de usuarios que se encuentran activos en el momento.
- interactuar de forma sencilla con nueve personas a la vez.
- cancelar o terminar la conversación establecida en cada una de las ventanas de la mensajería instantánea, y
- salir por completo de la aplicación.

Por lo tanto, ahora se procederá a describir de forma general el funcionamiento de esta herramienta. Para empezar el usuario debe darse de alta en el sistema (ver Figura 1), lo cual se ejecuta de manera sencilla ya que como se dijo anteriormente toda la interfaz cuenta con una guía hablada; de inmediato el sistema le comunicará al usuario que se ha conectado al mismo y a la vez le brinda las indicaciones para que pueda iniciar la comunicación con las personas que desee (ver Figura 2).

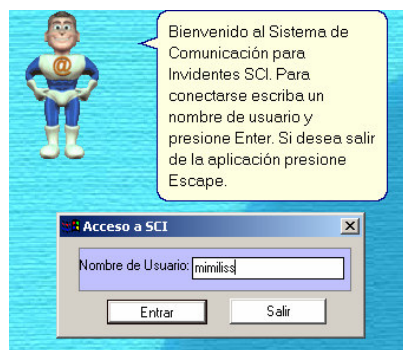


Figura 1. Interfaz principal del SCI.

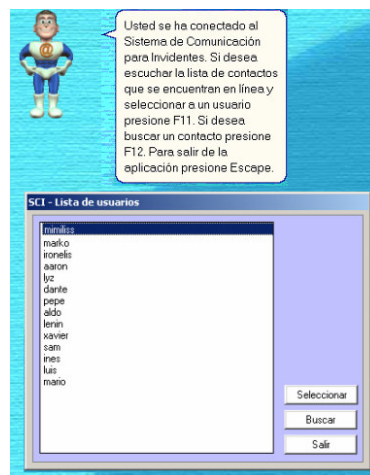


Figura 2. Interfaz de selección y búsqueda de contactos.

El usuario puede **seleccionar** a la(s) persona(s) con la que desea entablar la comunicación texto-voz presionando la tecla de función F11 y en cuanto escuche el nombre de usuario de la persona en cuestión, debe presionar enter; recordando que sólo podrá interactuar con nueve usuarios a la vez, como método de control inicial en el desarrollo y manejo de esta herramienta. Después de que el emisor selecciona a la persona con la que se desea comunicarse, automáticamente se le activa una ventana en la cual él puede escribir los mensajes que desea enviarle. Dichos mensajes van a llegar de manera **escrita y transformados en voz** al canal de recepción de la comunicación, y lo mismo sucede con los mensajes que le llegan a la persona que inicialmente era el emisor, el cual en este caso pasa a ser receptor. Otra de las funciones más relevantes de este sistema consiste en la **búsqueda personalizada de usuarios**, la cual se hace activando la tecla de función F12 y la interfaz le despliega una pequeña ventana de forma automática en la cual puede escribir el nombre de usuario de la persona que

desea contactar. Si la persona se encuentra en línea, de inmediato se activará la ventana de conversación con el mismo; de otro modo el sistema le comunicará a través de voz y también de forma escrita que dicho usuario no se encuentra en línea en ese momento. Por último, el usuario puede culminar la conversación con sus contactos o salir por completo de la aplicación tan sólo presionando la tecla *escape*. Para la evaluación del sistema se utilizó el cuestionario SUS (System Usability Scale), el cual se ajustó al caso de estudio en cuestión. SUS fue desarrollado en el año de 1986 como parte de la introducción de la ingeniería de usabilidad y ha demostrado ser una herramienta valiosa para la evaluación, ya que es un sistema robusto y confiable, además de que correlaciona bien sus datos con otros sistemas de medida [15]. Según los datos obtenidos de acuerdo al análisis efectuado durante la prueba piloto, el SCI obtuvo una puntuación SUS de 63.75, lo cual lo sitúa en un nivel de usabilidad aceptable pero que a la vez se puede mejorar. Este proyecto brinda tanto a la sociedad con discapacidad visual del CREE, como de otras localidades la oportunidad de poder interactuar con otras personas que presenten las mismas características o no; con el objetivo de que compartan experiencias, conocimientos, diversión, entretenimiento y lo más importante radical en que su mundo social se expanda y no se quede tan limitado como hasta el momento ha sido para estas personas especiales. La primera fase de este proyecto está concluida pero el SCI puede ser complementado con múltiples funciones entre ellas: envío de archivos, interacción con más de nueve personas a la vez, adaptación a otros idiomas, historial de mensajes por cada usuario, adaptación de código fuente a PHP ó ASP y plataforma Linux; entre otros aspectos.

#### 4. Conclusiones

La herramienta **SCI** es una interfaz de comunicación orientada a usuarios con discapacidad visual, la cual se basa en la transmisión de información en línea (mensajería instantánea), en donde los datos que el emisor envía de manera escrita a través del canal de comunicación llegan al receptor transformados en voz. Su importancia radica en que es un sistema especializado, ya que el análisis, diseño y construcción de esta herramienta se ha fundamentado en las técnicas del UML, los conceptos básicos propuestos por las metodologías HCI, entre otros aspectos de igual relevancia; lo que ha permitido obtener una aplicación que ofrece la interacción de estos usuarios con la misma y por lo tanto con otras personas que presenten o no las mismas debilidades, ya que también puede ser utilizada por un individuo que no sea invidente, porque la interfaz ofrece tanto una guía hablada como escrita de cada una de las funciones que realiza. Para lograr este objetivo se procedió a analizar y evaluar ciertos parámetros que se consideraron relevantes al momento de implementar dicha herramienta, tales como los niveles de aceptación, usabilidad y utilidad, los cuales obtuvieron un porcentaje altamente satisfactorio en los resultados de la prueba piloto que se aplicó a los participantes con discapacidad visual de dicho estudio. Esta aplicación es de adquisición gratuita y se encontrará disponible en la red Internet para que la persona que la considere, útil la descargue en su *computadora personal* (PC) y logre una eficiente y fácil comunicación en línea.

## 5. Agradecimientos

Por su invaluable ayuda en el desarrollo de este proyecto de tesis, mil gracias a mis grandes amigos: Ironelis, Lucy, Marko, Lyz, Miguel y especialmente a mi familia.

## 6. Referencias

- [1] J. Saramago, *Ensayo sobre la ceguera*, México: Aguilar, Altea, Taurus, Alfaguara, S.A, 2001.
- [2] A. Molteni, (2003, Mayo 17), "Internet y los discapacitados", [Online]. [http://www.htmlpoint.com/accessibilita/accessibilita\\_01.htm](http://www.htmlpoint.com/accessibilita/accessibilita_01.htm)
- [3] I. Galo, (), "Kit de acceso a Internet para invidentes", in Diario de Noticias, Digital ed. Pamplona, 2001.
- [4] F. Castejón, J. Escalada, L. Monzón, M. Rodríguez, P. Sanz Velasco, (2003, Abril 20), "Un conversor texto-voz para español", [Online]. <http://www.tid.es/presencia/publicaciones/comsid/esp/articulos/vol52/artic8/8.html>.
- [5] B. Schneiderman, *Designing the user interface strategies for effective Human Computer Interaction*, Third edition, Addison Wesley Iberoamericana.1998.
- [6] M. Casian, *Centro de Tecnología Adaptada DIF*, Colima: CREE. 1998.
- [7] G. Booch, J. Rumbaugh, I. Jacobson, *El Lenguaje Unificado De Modelado*, Madrid: Addison Wesley Iberoamericana. 1999.
- [8] G. Scott Bell, (2003, Mayo 10), "Microsoft Agent Ring", [Online]. <http://www.msagentring.org/>.
- [9] RNIB, (2003, Noviembre 19), "Speech Output Systems", [Online]. [http://www.rnib.org.uk/xpedio/groups/public/documents/PublicWebsite/public\\_techspeechword.doc](http://www.rnib.org.uk/xpedio/groups/public/documents/PublicWebsite/public_techspeechword.doc)
- [10] A. Bianchi, "Recursos en Internet: Integración de los invidentes al mundo informatizado", Boletín informativo Paso a Paso, Universidad de Carabobo, Caracas, (2002, Diciembre).
- [11] V. Meaux, (2003, Junio 7), "Guía de MSN Messenger 4.6 con JAWS 3.7 y 4.0", [Online]. <http://www.funcaragol.org/html/gjaws.htm>
- [12] A. Molteni, (2003, Mayo 17), "Los softwares más utilizados por los invidentes", [Online]. [http://www.htmlpoint.com/accessibilita/accessibilita\\_26.htm](http://www.htmlpoint.com/accessibilita/accessibilita_26.htm)
- [13] C. Arriola, (2003, Julio 13), "La informática al servicio de los invidentes. Un caso de la vida real", [Online]. [www.ciberhabitat.com/hospital/testimonio/](http://www.ciberhabitat.com/hospital/testimonio/)
- [14] M. Casian, *Centro de Tecnología Adaptada DIF-Colima*, Colima: CREE. 1998.
- [15] Jordan et al., *Usability evaluation in industry*, London, UK: Taylor & Francis. 1996.

# Hacia un Sistema de Diagnóstico y Compensación de Dislexia para Niños de Nivel Primaria

Mario A Moreno Rocha<sup>1</sup>, Rosario Peralta Calvo<sup>2</sup>, Ricardo Ruiz Rodríguez<sup>3</sup>

Cuerpo Académico de Ingeniería de Software, Universidad Tecnológica de la Mixteca,  
Km. 2.5 Carretera a Acatlima, Huajuapán de León, Oaxaca 69000, México  
Tel. +52 (953) 53 20399 ext 200  
{mmoreno<sup>1</sup>, mcalvo<sup>2</sup>, rruiz<sup>3</sup>}@mixteco.utm.mx

**Abstract.** This article will present development being done for a computer system aimed to primary school level children, which diagnoses and compensates dyslexia, explaining how Human-Computer Interaction and the utilisation of new technologies are essential in this development. It will also discuss how this system could support children with a learning disability, dyslexia in this case, in their scholar achievement.

**Resumen.** En el presente artículo se dará a conocer el desarrollo que se está siguiendo para la conformación de un sistema de cómputo para el diagnóstico y compensación de niños con dislexia de nivel primaria, explicando cómo la Interacción Humano-Computadora y la aplicación de nuevas tecnologías son parte esencial en este desarrollo. También se discutirá cómo este sistema pudiera apoyar a niños que presenten alguna discapacidad de aprendizaje, en este caso dislexia, en su desempeño escolar.

**Palabras clave:** Discapacidad de aprendizaje, dislexia, usabilidad, HCI.

## 1. Introducción

Como es de todos conocido, la computadora ayuda a la automatización de innumerables procesos. Su uso se ha ampliado a otras áreas, como en el caso de la Educación Especial. A través de diferentes programas de aplicación o software especializado, se han apoyado, a través de los años, a personas con necesidades educativas especiales. [7]. Los estudios muestran que los niños con problemas de este tipo constituyen aproximadamente entre el 3% y 6% de la población estudiantil estadounidense. El Departamento de Educación de los Estados Unidos afirma que aproximadamente un cinco por ciento de la población de una escuela puede tener problemas de aprendizaje [5]. En México, a saber, no existen estudios tan completos como éstos, pero entre las discapacidades de aprendizaje más importantes entre niños de nivel primaria, podemos encontrar el Déficit de Atención (ADD), discalculia y dislexia, entre otros.

La dislexia es causada por dificultades en el sistema nervioso central que afecta la habilidad de escritura. Más específicamente, el individuo puede presentar problemas al recordar, reconocer, invertir letras, o tener una pobre escritura [4], todas ellas habilidades básicas en esta etapa temprana de aprendizaje. La dislexia, a través de

terapias especiales, puede compensarse para ayudar al niño a que continúe su educación básica. El diagnóstico y tratamiento de dislexia, un proceso largo y difícil, es realizado por maestros en Educación Especial.

En la Cd. de Huajuapán de León, Oaxaca, existen pocas instituciones que realizan esta labor. En un mismo y concurrido salón de clases, las maestras tratan un sinnúmero de discapacidades de aprendizaje simultáneamente, disminuyendo al mismo tiempo, la capacidad de su experiencia y tiempo dedicado a cada uno de los niños. Por esta razón, se propone el desarrollo de un software de apoyo para el docente de educación especial, en donde se automatizarán algunos de los procesos que realiza en el salón de clases, para diagnosticar la dislexia que presente el niño y compensarla en algunos aspectos. También se propone utilizar las disciplinas de la Interacción Humano-Computadora (UCD, estudios de usabilidad, desarrollos de interfaces) y la aplicación de nuevas tecnologías para la solución del problema planteado.

## **2. Antecedentes al Sistema y la Labor del CASI**

Como antecedente a este trabajo, se tiene el desarrollado por Alan Newell y Peter Gregor, quienes describen la forma en que un procesador de textos puede ayudar a aliviar algunas dificultades presentadas por las personas con dislexia [3]. El presente sistema en aún en desarrollo, a diferencia de aquél del Prof. Newell, considera el proceso de diagnóstico que en un momento dado se le realizaría al niño que se presume tiene dislexia, pero únicamente tratando el área escolar, y entre edades de seis y siete años, considerando que es un apoyo para el docente de educación especial.

Por otro lado, dentro del Instituto de Electrónica y Computación (IEC) de la Universidad Tecnológica de la Mixteca, el Cuerpo Académico de Ingeniería de Software (CASI), realiza investigación y desarrollo dentro del área de Interacción Humano-Computadora (HCI), y es de especial interés, el desarrollo para la compensación de discapacidades en general, y las cognitivas en particular.

Para este proyecto, las pruebas se realizaron en el Laboratorio de Usabilidad (UsaLab) del IEC. El proyecto de dislexia está siendo desarrollado como un proyecto de tesis de licenciatura, con la supervisión del líder de la línea de investigación en HCI del CASI, así como de la invaluable asesoría de una maestra en educación especial.

## **3. Metodologías de Desarrollo**

La complejidad inherente en el proyecto de diagnóstico y compensación de dislexia requiere de un enfoque multidisciplinario, en la que las ciencias computacionales, así como la psicología, participan con metodologías bien definidas: Diseño Centrado al Usuario (UCD) y Metodologías de Evaluación en Psicología [7]. La metodología

seleccionada para el diagnóstico se basa en la realización de pruebas de escritura y lectura a los niños que el docente sospeche padezcan de dislexia.

#### 4. Desarrollo del Sistema

El desarrollo del software siguió la metodología descrita por Preece [6]. El proceso iterativo está ilustrado en la Figura 1, y sus detalles abajo.

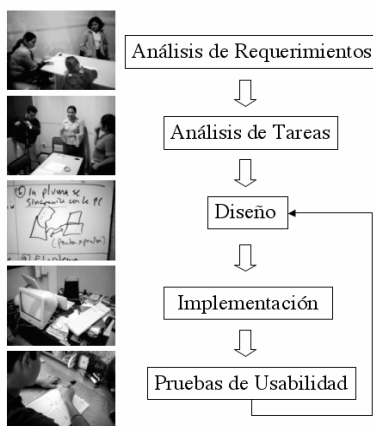


Fig 1. Proceso de desarrollo del software de acuerdo al Diseño Centrado al Usuario

- **Análisis de requerimientos:** se hicieron grabaciones de video de la maestra en educación especial desarrollando las pruebas de diagnóstico en el mismo salón de clases. Después, se reprodujeron las pruebas en el UsaLab, para su posterior análisis junto con cuestionarios y entrevistas, así como las observaciones realizadas en el segundo video.
- **Análisis de tareas:** se eligieron y determinaron las tareas, metas, procedimientos, factores de éxito, tiempos y escenarios a automatizar (GOMS). De esta manera se generaron las áreas del sistema.
- **Diseño:** el diseño de pantallas por prototipos de baja fidelidad, y *card sort*. Uso de guías de desarrollo de interfaces y experiencia previa en sistemas similares para niños.
- **Implementación:** se desarrolla el sistema bajo MS Windows XP y con el ambiente de desarrollo Delphi ver. 5.0.
- **Pruebas de usabilidad:** para la maestra, se desarrollan pruebas de usabilidad basadas en *thinking aloud protocols* y *cognitive walkthrough*. Para el desarrollo de pruebas con los niños, seguimos las pautas discutidas en [1], siendo útiles las mismas técnicas empleadas con las maestras.

## 5. Nuevas Tecnologías Empleadas

Otro factor interesante en la automatización de las actividades del diagnóstico fue la incorporación de nuevas tecnologías. El análisis de tareas mostró la importancia de la escritura del niño ante el dictado de las palabras seleccionadas para realizar dicho diagnóstico, para ser evaluada en el momento, o después, por el profesor de educación especial.

Debido a los requerimientos del diagnóstico, se consideran dos opciones para el registro de la escritura, la primera de ellas es el uso de papel y pluma normales, y a través del uso de un scanner, obtener la imagen para su diagnóstico; la segunda es el uso de una pluma digital, que permite escribir como una pluma normal, pero almacena permanentemente la escritura, dibujos y notas en una PC

Al final, fue seleccionada la pluma Logitech io, para contar con otra alternativa aparte del scanner, y sobre todo, darle una aplicación en un área que no fue considerada al momento en que fue desarrollada.

## 6. Sistema Implementado

La implementación de nuevas tecnologías no estuvo exenta de errores. La más significativa fue la imposibilidad de traducir a texto la imagen obtenida por la pluma digital, para su posterior análisis por el sistema. El porcentaje de efectividad no fue el adecuado, además que las pruebas mostraron que el idioma Castellano es especialmente difícil de convertir.

Se logró obtener mejores resultados tomando la imagen de la escritura del niño, ya sea con la pluma digital o un scanner, y haciendo el análisis manual a través de la herramienta desarrollada para el sistema. El sistema contabiliza automáticamente el número de errores, de tal forma que se puedan generar los resultados para ubicar en un nivel de dislexia al niño. La Fig. 2, muestra dicha herramienta.



Fig 2. Marcación de errores en la escritura del niño.

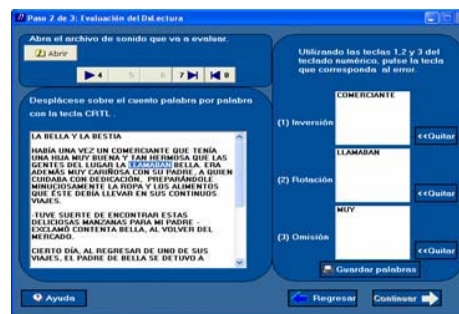


Fig 3. Marcación de errores en la lectura del niño.

## 6.1 ViaVoice para el Diagnóstico de la Lectura

Para realizar el diagnóstico de la lectura del niño, el docente debe grabar al niño haciendo lecturas de cuentos infantiles, divididos en niveles de complejidad, para determinar el grado de dislexia existente. Se consideró el software de IBM ViaVoice, para trabajar de la misma manera que el diagnóstico de la escritura: grabando la lectura del niño, luego convirtiéndolo a texto y haciendo una comparación contra el documento original.

De nuevo, el grado de efectividad prometido por ViaVoice (del 90 al 95%) se convirtió en un 50 al 60% en la práctica. Entre las causas identificadas fue que el niño debía configurar el software para reconocer su lectura, misión imposible con niños con dislexia. ViaVoice fue entonces descartado.

La implementación del diagnóstico de la lectura fue realizada por una herramienta desarrollada especialmente, en que el docente tiene total control de la reproducción de la grabación. Ver Fig. 3.

## 6.2 Compensación de la dislexia

La compensación consiste en proporcionar ejercicios al niño que le permitirán sobrellevar, y con el tiempo, corregir, parte del problema de dislexia en la escritura y lectura. Para la selección de los ejercicios, se recurrió a la bibliografía que los maestros en educación especial llevan a cabo para realizar dicha compensación en el aula de clase [2]. De igual manera se dio con el contenido de los ejercicios, cuya característica principal fue el grado de complejidad y el volumen de la información.



Fig 4. Ejercicio de “Relacionar Palabras”.



Fig 5. Ejercicio de “Armar Palabras con Sílabas”.

### 6.2.1 Aplicación de los ejercicios

La maestra de educación especial selecciona el nivel, el tipo y el bloque de ejercicio que el niño va a realizar, dejándole después el control del ejercicio al niño. La Fig. 4 muestra el ejercicio “Relacionar Palabras”, que consiste en unir la palabra del lado derecho con su igual de la izquierda. La Fig. 5 muestra el ejercicio “Armar Palabras



con Sílabas”, la cual contiene cinco globos de colores, en donde cada globo tiene las sílabas de una palabra y el niño, con ayuda del ratón, arrastra estas sílabas al cuadro correspondiente y las pondrá en el orden que cree es el correcto.

## 7. Conclusiones y Trabajo Futuro

El sistema se encuentra actualmente en pruebas con maestras de educación especial en el Diagnóstico de Escritura y Lectura. Hasta ahora, los resultados mostrados son muy halagadores, ya que cumplen los requerimientos iniciales y ha sido de gran ayuda a las maestras en tan difícil labor. La incorporación de HCI y de nuevas tecnologías, han mostrado un nuevo panorama a la forma en que las ciencias computacionales pueden ayudar a otras áreas como la educación especial. En este sentido es importante decir que se prepara un estudio para la determinación del grado de utilidad de la herramienta como apoyo para el docente de educación especial en el proceso de diagnóstico, así como para la compensación en el alumno a través del uso del software.

Por último, vale la pena mencionar que estamos interesados en compartir nuestra experiencia y obtener retroalimentación que nos ayude a mejorar, y de ahí la importancia y nuestro interés en poder participar en este tipo de foros. Así mismo, esperamos poder compartir más nuestras experiencias en este aspecto, dado que en este artículo sólo se menciona lo más sobresaliente debido a limitaciones de espacio.

## 9. Referencias

- [1] Hanna Libby, Risdén Kirsten, Alexander Kirsten, “Guidelines for usability testing with children”, *interactions archive*, Volume 4, Issue 5 Sept. / Oct. 1997 Pp. 9 – 14, 1997, ISSN:1072-5520, ACM Press New York, NY, USA
- [2] Narvarte, Mariana E. (2001). *Trastornos Escolares: Detección – Diagnóstico y Tratamiento*, GIL, SA de CV, México.
- [3] Newell, Allan F & Gregor, Peter, (2000). “An empirical investigation of ways in which some of the problems encountered by some dyslexics may be alleviated using computer techniques”.
- [Electronic Version]. *Proceedings of the fourth international ACM conference on Assistive technologies*, 2000, Pp. 85-91.
- [4] Our Special Kids. Org <http://www.ourspecialkids.org/definitions.html>
- [5] Pierangelo, Roger & Jacoby, Robert, (1998). *Guía de Educación Especial para Niños Discapacitados*, Prentice Hall, México
- [6] Preece, J. *A guide to usability, human factors in computing*. Addison-Wesley Publishing Company, 1993, ISBN: 0-201-62768-X
- [7] SEP (2003). *Antología: La integración educativa en la escuela primaria*, IEEPO, México.

# Creación de un diccionario electrónico de sexualidad

Gerardo Sierra, Marlene Moreno, Eduardo Solares, Alfonso Medina  
Grupo de Ingeniería Lingüística  
Instituto de Ingeniería, UNAM.

e-mail: {gsierram, mmorenop, esolaesm, amedinau}@iingen.unam.mx  
3er cubículo de la Torre de Ingeniería, Circuito Interior, Ciudad Universitaria,  
Del. Coyoacán, Apartado Postal 70-472, CP 04510, Distrito Federal, México.

**Abstract:** Although several glossaries and vocabularies exist about sexuality, a structured collecting with homogenous criteria of basic information about the exercise of sexuality in Mexico is needed. The ever increasing use of electronic media as sources of information and the advance of technologies make it possible to develop a dictionary to be available in CD. This paper describes a design alternative for a basic electronic dictionary based on the analysis of requirements of potential users.

**Resumen:** Si bien existen glosarios y vocabularios sobre sexualidad, hace falta una recopilación estructurada con criterios homogéneos de la información básica sobre el ejercicio de la sexualidad en México. El creciente uso de medios electrónicos como fuentes de información y el avance en las tecnologías hace posible la construcción de un diccionario que se haga disponible en CD. Este trabajo presenta una alternativa del diseño de un diccionario electrónico básico basado en el análisis de requisitos de usuarios potenciales.

**Palabras Clave:** Lexicografía computacional, diseño interfaz-usuario, sexualidad.

## 1. Introducción

No es ningún secreto que existe una gran necesidad de información veraz sobre temas sexuales en México y en el mundo [1]; los embarazos inesperados, las enfermedades de transmisión sexual, la violencia, la discriminación, la homofobia, etc., están relacionadas con una incompleta o no comprensiva y represiva educación sexual.

Para comprender los temas relacionados con la sexualidad es necesario tener conocimiento sobre el significado de las palabras que se utilizan; un diccionario especializado es la opción para proporcionar esa información de manera clara, sucinta y explícita, sirviendo para los propósitos principales de representación y transmisión del conocimiento especializado.

El Grupo de Ingeniería Lingüística (GIL) de la UNAM, junto con el Colegio de México y sexólogos calificados, diseñan un diccionario electrónico de sexualidad, en el que se ofrecerán además de respuestas a las dudas de los usuarios, información

lexicográfica especializada. En este documento nos enfocaremos a los problemas de identificación de usuarios, análisis de requerimientos, diseño y evaluación del sistema.

## **2. Análisis**

Entre los usuarios potenciales del sistema encontramos a: sexólogos, profesionales de la educación, personal médico y paramédico, psicólogos, trabajadores sociales, terapeutas en general, organizaciones de la sociedad civil con trabajos relacionados a los temas; escuelas y estudiantes de secundaria, preparatoria y universidad; niños, padres y madres de familia.

Tomando como base los trabajos realizados por Kharma [2], Hartmann [3], Kipfer [4], Béjoint [5], Quirk [6] y Tomaszczyk [7], entre los porcentajes de importancia de algunas de las necesidades básicas de los usuarios en el uso de los diccionarios se encuentran: contenido en significados (90%), gramática adecuada (60%), sinónimos (55%), antónimos (5%), historia de la palabra (3%) y juegos (5%).

El hecho de que las encuestas mencionen una diversidad tan amplia de las necesidades básicas en la consulta de diccionarios, como sería para el uso de juegos, no quiere decir que los diccionarios están pensados para tal fin o que los usuarios tengan ese objetivo para utilizarlos, pues ejemplifica los distintos usos de los diccionarios, entre los que están los juegos.

Nuestro diccionario contemplará, por tanto, estas necesidades aunadas a las de usuarios que utilizan sistemas de información electrónica, como son; diferentes formatos de presentación, opciones de búsqueda, portabilidad, interactividad, etc.

## **3. Diseño del sistema**

Para asegurar una perspectiva científica en el diseño del sistema, contaremos con la participación de los expertos en los campos siguientes: terminólogos, lexicógrafos, ingenieros de la lengua, sexólogos, programadores y analistas, entre otros; todos ellos forman un equipo interdisciplinario de trabajo con tareas específicas dentro del proyecto.

Para asegurar que el diccionario presente información adecuada a las distintas edades de los usuarios, el sistema contara con una etapa de validación con series de preguntas que permitirán identificar la edad del usuario. Con ello, el sistema delimitará el contenido de la información que el usuario visualizará.

### 3.1 Opciones de búsqueda

El usuario tendrá la posibilidad de realizar cuatro búsquedas básicas para obtener información en el diccionario:

- Búsqueda por términos, también conocida como búsqueda semasiológica, permite el ingreso de términos referentes a sexualidad, mostrando como respuesta su definición, temas y términos relacionados, a los que puede acceder, si así lo desea, mediante hipervínculos en pantalla.
- Búsqueda por temas ya definidos por los expertos, que presentará además de otros temas relacionados con el área, hipervínculos de términos inmersos en el tema.
- Búsqueda por áreas de uso, mostrará términos y temas contenidos en el área temática elegida, para que el usuario obtenga una visión más amplia de los conceptos buscados.
- Búsqueda inversa, también conocida como búsqueda onomasiológica, en la cual se permitirá buscar términos a partir de la descripción del concepto en lenguaje natural. Esta característica es una aportación innovadora en el área de diccionarios electrónicos desarrollada por el Grupo de Ingeniería Lingüística [8]. Con base en esta búsqueda, el usuario escribirá cualquier idea que tenga acerca de un término y obtendrá un listado de los términos con sus respectivas definiciones formales, que mejor cumplen con esta descripción.

Las búsquedas podrán ser refinadas por medio de filtros elegidos por el usuario durante sus consultas. Los formatos de respuesta serán presentados en texto (definiciones, términos, y temas relacionados), imagen y audio (voces pregrabadas referentes a los contextos de uso de la terminología). Entre los filtros se encuentra, por ejemplo, el relativo a las fuentes y/o las diferentes escuelas de sexualidad.

Es importante señalar que el diccionario contará con algunas herramientas de búsqueda inteligente que utilizan técnicas específicas del procesamiento de lenguaje natural. Una de ellas es un proceso de corrección ortográfica de carácter por carácter en el proceso de la búsqueda por términos, en la que, conforme el usuario vaya introduciendo el término, el sistema irá presentando a aquellos que correspondan al término escrito correctamente o al que tenga mayor similitud. Asimismo, existirá la corrección ortográfica de términos completos introducidos por el usuario; un ejemplo de ello sería al utilizar números y palabras para expresar términos como 6 + nueve, 69, 6 y 9, 6+9, seis nueve, 96, etc., en donde el diccionario comprenderá la petición del usuario y lo llevará al término “sesenta y nueve”.

Además de la utilización del sistema con fines lexicográficos el sistema contará con un apartado dedicado a juegos, lo que garantizará un enriquecimiento conceptual sobre la sexualidad. Entre los juegos se encuentran: sopa de letras y crucigramas, los cuales provocaran interés para jugar y aprender.

### **3.2 Interfaz de usuario**

Un factor importante a considerar dentro de la interacción entre el humano y la computadora es el diálogo requerido entre ambas partes; en general, el sistema debe comunicar al usuario los procesos que está llevando a cabo, los que desarrolló y los resultados obtenidos, mientras que el usuario debe proporcionar las instrucciones, datos e información al sistema para desarrollar la tarea deseada. La forma en que se comunican ambas partes es conocida como la “interfaz de usuario”. Ésta no es únicamente lo que el usuario puede ver, tocar y oír, ya que también involucra los conceptos que el usuario debe conocer acerca del sistema de cómputo y cómo puede utilizarlos para llevar a cabo las diferentes tareas de forma eficiente.

En este sentido, debemos tomar en cuenta el concepto de usabilidad, el cual hace referencia a la rapidez y facilidad con que los usuarios llevan a cabo sus tareas a través del uso de una interfaz. En nuestro caso, mencionaremos a grandes rasgos tres vertientes:

- Una aproximación al usuario. Para desarrollar la interfaz y que sea usable, se tiene que conocer, entender y trabajar con las personas que representan a los usuarios finales o potenciales del diccionario. Entre otras cosas, se contempla la participación de grupos focales de usuarios potenciales para evaluar el diccionario tanto en el diseño de las interfaces como en la información contenida y en la claridad de los conceptos lexicográficos.
- Un amplio conocimiento del contexto de uso. Para desarrollar cualquier producto usable debemos tener muy claro los objetivos del usuario, así como los trabajos y tareas del usuario que el sistema automatiza, modifica o embellece. En esta vertiente, resulta vital la participación de los ingenieros lingüistas que desarrollarán las herramientas de búsqueda inteligente con las técnicas de procesamiento de lenguaje natural.
- Satisfacción de las necesidades del usuario. Los usuarios son gente ocupada intentando llevar a cabo una tarea, por lo que el sistema debe relacionar usabilidad con productividad y calidad. Se buscará que el diccionario satisfaga al usuario desde el punto de vista del uso del sistema, a través de una interfaz confiable, amigable, ergonómica y robusta para el manejo de información lexicográfica, que le permita al usuario realizar múltiples tipos de consulta, a la vez que contenga valores agregados que hagan atractivo el uso del diccionario.

### **3.3 Plataforma de desarrollo**

Nuestro diccionario estará disponible en CD-ROM; las ventajas de hacerlo de esta forma son que en principio el usuario podrá trasladar el sistema a cualquier lugar y tener acceso al diccionario instalándolo en su PC o simplemente ejecutándolo desde la unidad de CD.

Como plataforma de desarrollo utilizaremos Java, ya que su diseño es orientado a objetos y la tendencia del futuro, a la que Java se suma, apunta hacia este tipo de programación, especialmente en entornos cada vez más complejos y basados en red.

#### **4. Evaluación**

La evaluación del sistema estará basada en criterios de diseño de interfaces usuario y pruebas de usabilidad. Contaremos con una serie de métodos y herramientas que incorporarán estas consideraciones [9].

Es totalmente necesario conocer el punto de vista del usuario sobre lo que le permite y no le permite hacer la interfaz. El usuario final es el que tiene la última palabra sobre el desempeño del mismo, y partiendo de las pruebas que se realicen con él, se pueden llevar a cabo modificaciones que permitan un mejor uso del sistema. Para ello, se tendrán los grupos focales, que serán usuarios potenciales con distintas necesidades en el uso del diccionario.

#### **5. Conclusiones**

La finalidad del diccionario básico de sexualidad será servir como medio didáctico e informativo a todas las personas que hagan uso de éste. Encontrarán información veraz y concreta acerca de los términos usados en el área de sexualidad, así como una gran variedad de imágenes y audio, y se navegará entre los diferentes formatos sin complicación alguna.

Si bien existen varios diccionarios en CD, por los rasgos propios de la temática propuesta, este diccionario tendrá características que lo hacen único en su género desde el punto de vista de interacción humano-computadora, tales como:

1. Buscadores inteligentes.
2. Inclusión de imágenes, sonidos, y juegos.
3. Variedad de búsquedas, como la del tipo onomasiológica.

Este proyecto será el comienzo de una serie de diccionarios electrónicos especializados que estarán a la vanguardia entre los sistemas computacionales existentes en el mercado, y que presentan la información de una forma agradable, y eficiente, motivando al usuario a usar estos sistemas cotidianamente.

#### **Referencias**

[1] Promotion of Sexual Health, Recommendations for Action PAHO, WHO, WASS, Antigua, Guatemala, 2002.

- [2] N.N. Kharma, "Wanted: a brand-new type of learner's dictionary", *Multilingua* 4 (2), 1985, pp. 85-90.
- [3] Hartmann, R.R.K. "Four perspectives on dictionary use: A critical review of research methods". In *Lexicographica: The dictionary and the language learner*, A. Cowie (ed), Max Niemeyer Verlag, Tübingen, 1987.
- [4] B.A. Kipfer, "Dictionaries and the intermediate student: communicative needs and the development of user reference skills", *Lexicographica* 17, 1987, pp. 44-54.
- [5] H.Béjoint, "The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills", *Applied Linguistics* 2 (3), 1981, pp. 207-222.
- [6] Quirk, R., "The social impact of dictionaries in the U.K." In *Lexicography in English*, R.I. McDavid and A.R. Duckert (eds.), New York Academic of Sciences, New York, 1973.
- [7] Tomaszczyk, J. "Dictionaries: users and uses", *Glottodidactica* 12, 1979, pp. 103-119.
- [8] Sierra G., *Design of a concept-oriented tool for terminology*, PhD Thesis, UMIST, Manchester, 1999.
- [9] F. Gamboa, E. Martínez, R. Juárez, J.L. Pérez, "Sistema de evaluación para un tutorial de física", *MXLII Congreso Nacional de Física*, Villahermosa Tabasco, 1999.

# **Interacciones en la realidad virtual y las simulaciones físicas**

**Virtual Reality Interaction and Physical Simulation**

Cesar Mendoza  
Fabio Ganovelli  
(Editores)





## Introduction

In recent years, much research has been done in virtual reality technology to advance other fields such as medicine, engineering, training, education and entertainment. Designing usable and effective interactive virtual worlds involves many areas of research and the new advances in computer science and the use of more powerful machines allows the development of new applications that were unimaginable in the past.

However, most of this research is done in USA, Europe, or Japan. Although in Mexico this domain has started to grow, there is an important need to spread out the works in the area. Hence, in its first version, the workshop focuses on describing new methods and on-going research that give a wide hint of the research in virtual reality and computer graphics.

The Virtual Reality Interactions and Physical Simulations (VRIPHYS 04) workshop aims to identify and put together the Mexican scientific community doing research in virtual reality interactions and physical simulations in computer graphics.

Moreover, since the workshop is open to international participation, we seek to create an environment able to motivate any kind of cooperation between the Mexican and international communities.

The sessions of the workshop were created after the paper selection process rather than based on a predetermined set of topics. This process leads to creative session titles and fits better the expectations of the attendees. The selected papers come from Mexico, Italy, Spain, Japan and Czech Republic and touch different subjects of computer graphics and virtual reality.

The workshop includes an invited talk from one of the most representative researchers in computer graphics in Mexico, Isaac Rudomin.

Besides the technical communications, there is a panel session, which gave attending researchers the opportunity to express and exchange ideas, to give feedback to the workshop for next editions and to give proposals to establish tools that help the development of the computer graphics and virtual reality domains in Mexico.

As part of the activities of the workshop, it includes a visit to the Virtual Reality Laboratory in the University of Colima.

Considering virtual reality as a multidisciplinary domain, the workshop seeks to present the potential advantages of using virtual environments in other research fields. We expect to attract application developers and show them the envelope of what can be done in virtual reality and computer graphics. We seek to encourage developers to work together with researchers and provide products that are truly usable, rather than just a demo.

We would like to thank the authors of the paper submissions, the ENC'2004 organizing committee, the workshop coordinators, and the scientific committee itself and the tertiary reviewers who took the time to write careful, constructive, and detailed reviews.

## Scientific Committee

Co-Presidents:

Cesar Mendoza, CNR Italy & Trinity College, Ireland  
Fabio Ganovelli, CNR, Italy

Carol O'Sullivan, Trinity College, Ireland  
Isaac Rudomin, ITESM, CEM, Mexico  
Bedrich Benes, ITESM CCM, Mexico  
Miguel Garcia, Universidad de Colima, Mexico  
Roberto Scopigno, CNR Italy  
Philippe Meseure, SIC France  
Stephane Cotin, CIMIT USA  
Christian Laugier, INRIA Rhone-Alpes France  
Genevieve Lucet, Ixtli UNAM Mexico  
Christophe Chaillou, INRIA Futurs France  
Riccardo Scateni, Università di Cagliari, Italy  
Isabel Navazo, Universitat Politècnica de Catalunya, Spain  
Gerardo Ayala, UDLA, Mexico  
Victor Ng-Thow-Hing, Honda Research Institute, USA  
Francois Faure, INRIA Rhone-Alpes France

**Acknowledgements** The workshop was organized in cooperation with Eurographics.



# Enhanced Hair Paint

Isaac Rudomín, Benjamín Hernández

ITESM-CEM, Atizapan México  
{rudomin, hbenjamin}@itesm.mx

**Abstract.** Hair Paint is a modeling and animating tool for hair and furry objects. In this paper some improvements are done to Hair Paint to add some new features like adding hair gel to modify the physical behavior of hair or fur, addition of pony tails and hair growth direction to make more complex hair styles. We also implemented mechanisms that avoid the hairstyle mess up due to wind forces.

**Keywords.** Physical restrictions, Hair Animation, Hair Paint.

## 1 Introduction

Human hair modeling, animation and rendering are active research areas. The initial research in human hair simulation has so far focused in one of the three areas: hair modeling allows us to represent distribution, form and styling; hair animation allows us to reproduce movement and the interactions due to this movement; and hair rendering allows us to represent light interactions that cause reflections and shadows so the lack of one of these characteristic produces an unrealistic hair appearance.

Hairstyling can be an awkward task for the user. There are some alternatives that have been published such as designing an interactive 3D user interface [1][2] so the user can style hair on any section of the 3D model. Another alternative [3] is seeding hair randomly based on the triangular faces of the 3D model and barycentric coordinates. This approach simulates the random positions of the hair roots but the user cannot easily control the distribution of hairs over the 3D model. Another technique [4] allows the user to manipulate a kind of electronic comb around a (real) mannequin's head, while the results of the hairstyle are shown on a computer monitor. This is an interesting method but requires hardware not commonly available. A new method [5], for obtain hair geometry and implicitly get a hairstyling, analyses a set of real hair color images with different light positions to determine the hair direction due its anisotropic behavior.

There are some commercial alternatives such as: Shave and Haircut for Maya, from Joseph Alter Inc., Maya Fur from Alias Wavefront, Poser from Curious Labs and Sasquash from Warley Laboratories: all these systems allow hair styling. Common users need to have some knowledge of these tools to control this interfaces since every tool has a different one, and know all the commands. In some cases these commands

are difficult to understand, and sometimes they require refinements in a 3D interface that is complicated to use for common users.

Hair simulation systems can be classified according to the way they model hair. Some systems use alpha-mapped textured surfaces [7] to simplify the lighting model, making it independent of the number of hairs but produce a flat hair model. In Daldegan et al [8] a method is proposed to define hair characteristics strand by strand such as density, distribution and orientation to complete the hairstyle. Other methods model hair strand by strand [9][10][11], in them the animation is usually physically based. Other techniques exploit hair's tendency to form clusters due to static electricity, so the hair modeling is carried out using clusters [12][13][14]. In these methods clustering greatly simplifies interaction processing. Animation is computed on these clusters defined by a skeleton curve. These skeleton curves are then rendered using more curves to complete the hair model. Still another approach [15] consists in modeling hair in a hierarchical way so there is a general hairstyle first, and then refinements can be made locally to increment the hairstyle complexity. The method has been tried for non real time animation [16]. Bando et al [17] shows a different way of animating hair using particles. They use sets of particles that serve as sampling points for the volume of the hair and are scattered along the hair model. These sets serve to track the motion of the volume of the hair, and the dynamics of the hair, including hair-hair interactions, are simulated using the interacting particles. For hair rendering, the Kajiya and Kay model [18] is normally used to simulate light scattering from hair. Marschner et al [19] extends this model by adding new measurements since a hair strand is modeled as a transparent elliptical cylinder so transmission and internal reflection are accounted for.

In a previous paper [6], we discussed a system, called Hair Paint, that based on a 2D paint program interface and color scale images to specify hair characteristics, allows the user to model hair and animate it in a fast way. In this paper we describe the extensions we have done to this tool. The extensions added allow the user to modify the hair physics and allows for a greater variety in the hairstyles the user can model.

## 2 Hair Paint Reviewed

Hair paint is an easy-to-use tool for modeling hair based on a paint application and a 3D view of the hairstyle [6]. Hair paint adds two important characteristics for any hair styling tool: interactivity and ease of use. Hair Paint allows users interactively see and control their hairstyle in real time and modify it if something goes wrong. Since Hair Paint is an easy-to-use and intuitive modeling tool the user can design hairstyles in a fast way.

For hair modeling Hair Paint uses an explicit representation of hair and allows users to specify geometrical characteristics (density/root position and length) and style characteristics (curliness, braids and color). These characteristics are specified via color scale images where darker values indicate minimum values and brighter values indicates maximum values, i.e. for curliness, we get straight hair for darker values and different degrees of curliness for brighter colors. [6, 20] describe in detail the methods used to model these characteristics.

Since the user specifies hair characteristics via color scale images, Hair Paint has a paint area where users create their artwork and a 3D view area where they can see their hairstyle (Figure 1).

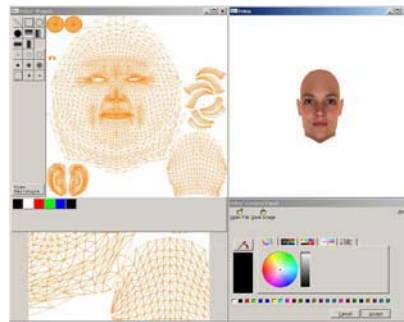


Figure 1. Hair Paint Interface

A detailed description of the modeling tools provided by Hair Paint is included in [6]. In figure 2 we can see some examples.

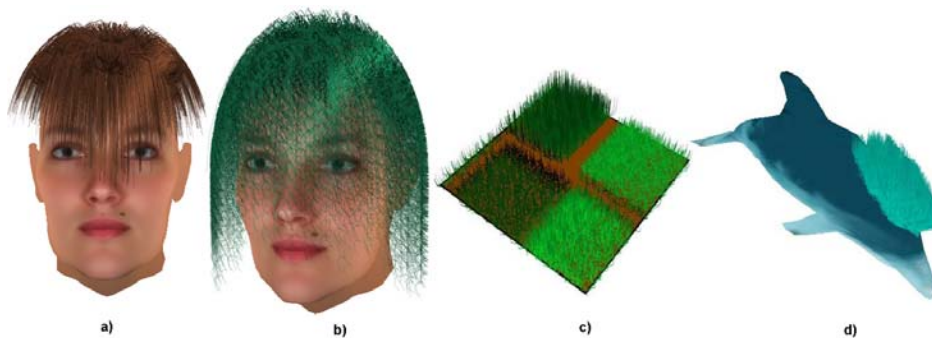


Figure 2. a) Different length are shown b) Hairstyle with curly hair c) A prairie with different grass lengths, color and density d) The Fringe Dolphin

Hair Paint allows users to animate hair in an automated way so they don't have to define in detail what or how hair must move. A detailed description of the animation method used in Hair Paint is included in [20]; in general terms a simplified physically based method that automatically computes physics only on some hairs (basis hairs) is used, calculating other dependent hairs from them. Although hair modeling and hair animation is the most important part for Hair Paint, hair rendering is also considered. To capture the anisotropic behavior of hair and reduce aliasing effects we use hardware acceleration, available in most common graphics cards.

### 3 Enhancing Hair Paint

Expanding the capabilities of Hair Paint is necessary to make possible to users model more complex hairstyles and allow them to modify the physical behavior of hair. To

achieve these capabilities we propose the addition of three new tools that allow users to specify style characteristics such as: gel to modify the hair's physical behavior, growth of direction to model, for example, fringes and finally a tool for modeling pony tails. All of these characteristics are specified using color scale images, just as was the case for the original tools.

### 3.1 Gel

As mentioned in section 2, hair animation is a semiautomatic process. This, in some way, is practical for users since it simplifies the specification of animation process, on the other hand, there are different kinds of hair strands, for example curly or straight, that have different physical behavior due to length, form, thickness, etc. Another aspect that needs to be considered is that a hairstyle could have hair areas where strands move freely and other areas where strand movement is restricted. For these reasons and to provide users with more control over hair animation the gel tool was added. There are other efforts [22] to specify the gel effect over hair, here we propose our alternative.

Basically gel effects can be summarized as follows: the first effect is the modification of the amount of movement of the hair strands, i.e. the hair strands can be affected or not by wind forces. The second effect is the clumping effect, i.e. the gel can cause hair to stick together. There is another effect, although it is left for future work, related to hair appearance: gel can cause hair to look wet and bright.

Implementing the first effect can be done by modifying the hair strand's mass so it results in different motion from one strand to another but it could be problematic since minimum mass values close to zero could destabilize the system. To avoid this potential problem, we decided to modify the forces due to gravity and wind that interact with the hair strands. This force modification is done by multiplying it by a factor  $n \in [0, 1]$  obtained from the color scale image as described below.

For the second effect, we used the capability of our animation algorithm, first presented in [20], to modify the dependent hair strand's position. From Algorithm 1, we can obtain clumpy hair if we substitute the equation  $Position(x,y,z) += \Delta M$  by  $Position(x,y,z) += \Delta M - clumpiness$ , this equation, allows us to control the distance between basis hair strands and dependent hair strands to form clumps.

```

- Seed hairs on the 3D model's scalp (or wherever hair is
desired)
- For each face of the 3D model, pick randomly some hair
strands as basis hairs and associate the closest non
basis hairs as dependent hairs
- While animating hair
  - For each basis hair strand
    - Calculate equations of movement
    and keep the results ( $r_{old}$   $r_{new}$ )
    - Calculate  $\Delta M = |r_{new} - r_{old}|$ 
    - For each associated non basis hair
    strand change the position of all its
    nodes by  $Position(x, y, z) += \Delta M$ 

```

Algorithm 1. Basis – Dependent Hair Strands

The use of the color scale image is based on the following method. To specify gel, yellow scale images are used where darker color values represent maximum values for  $n$  and brighter color values represent minimum values for  $n$ . Another aspect that has been taken into consideration is that hair moves faster at the tip than in parts near the root. For that reason the factor value obtained from the yellow scale image is modified along the strand nodes to obtain this behavior.

In figure 3, we can see the effect of gel tool and the associated gel map.



Figure 3. Gel effects. a) A gel map used for the hairstyle on the right, notice that gel is just applied in the left side of the hairstyle. b) The resultant hairstyle, notice that from the center to the left side of the hairstyle, the hair doesn't look fluffy (like the opposite hairstyle) since clumps are created.

### 3.2 Growth direction

By specifying hair growth direction we can simulate the effect of combing hair, allowing users to model a wider variety of hairstyles than they can model with normal Hair Paint. In Hair Paint hair grows in direction of the polygonal face normal, i.e. strands form an angle of  $90^\circ$  between face this results in punk or military hairstyles if animation isn't applied.

To modify strand growth direction using color scale images it is necessary to use the three color components Red, Green, and Blue. Each component will modify the angle between polygonal face and hair strand in  $x$ ,  $y$  and  $z$  direction respectively. Since the RGB scale we used is from 0 to 255, we define angles from 0 to 255, where 0 specifies an angle of  $0^\circ$ , 128 an angle of  $90^\circ$  and 255 an angle of  $180^\circ$ . This allows us to have an absolute range from 0 to  $180^\circ$  for each  $x$ ,  $y$  and  $z$  component. It is necessary to mention that this angle range is used at the root of strands, if the same angle at the root would be used along the strand we could obtain straight strands from root to tip (like punk or military hairstyles), but with different initial angle. To avoid this and achieve the effect of a curved hair strand from the initial angle obtained at the root we modify this initial angle along all the strand nodes by multiplying it by a factor  $f$ , where  $f \in (0, 1]$ , we vary this factor  $f$  according the particle hair strand position, i.e. at the root,  $f$  equals to one and is decremented along the strand until the tip is reached.

For the modification of the hair strand angle we use polar coordinates to modify the growth direction. Figure 4 shows an example of the use of polar coordinates to modify the growth direction for the  $xy$  plane, for  $yz$  and  $xz$  plane the process is similar.



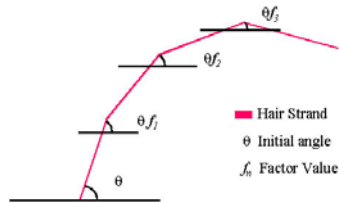


Figure 4. Modifying the initial growth angle

Figure 5 shows a hairstyle using the hair growth direction tool.

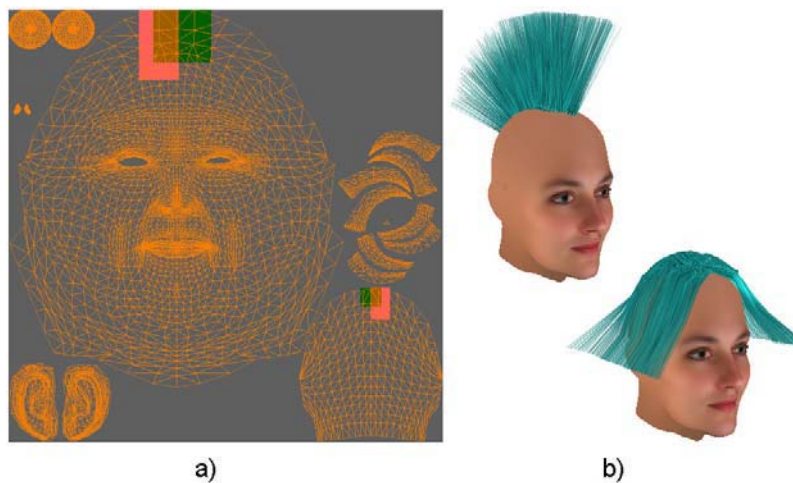


Figure 5. a) Growth Direction Map. b) Left, hairstyle without specified growth direction map. Right, hairstyle using the growth direction map in a).

### 3.3 Pony Tails

The last enhancement added to Hair Paint is the ability to model pony tails. These are modeled using GC (generalized cylinder [21]) rendered with OpenGL bezier curves to obtain smooth appearance and reduce discontinuities along the GC.

As mentioned in [21], a GC is composed by a skeleton curve that describes a path followed by Frenet frames. We take this skeleton curve and we apply physically based animation over the nodes of this curve using Verlet integration [20]. We change the radius of the Frenet frames along the GC to model the characteristic appearance of a pony tail where at the tip seems to be more strand relaxation than at the root, also we can use different radius to model different pony tail types (figure 6).

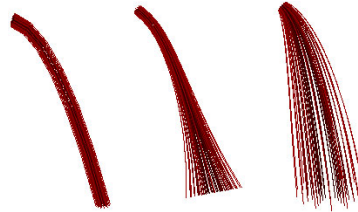


Figure 6. Pony Tails

Specification of pony tails in our modeling tool is done by violet scale image to control position and length. The length of the pony tail depends of the intensity of the image so for darker values the pony tail is shorter than for brighter values. Figure 7 shows a hairstyle with pony tail.



Figure 7. Pony Tail Hairstyle

## 4 Conclusions

The enhancements presented in this paper extend the modeling and animation capabilities of Hair Paint. For modeling hairstyles the addition of the tools to specify growth direction and pony tails allows users to define hair characteristics allowing them to model more complex hairstyles. The gel tool allows users to modify the physical behavior of hair strands. Users can also use it to restrict the movement of hair and avoid the excessive deformation of the hair that would cause the modeled hairstyle to be lost when animation is applied by reducing the wind force effects (figure 3).

## References

1. Chen, Saeyor, Dohi, And Ishizuka. A System of 3D Hairstyle Synthesis Based on the Wisp Model. *The Visual Computer*, 15(4), 159-170.
2. Kim And Neumann. Interactive Multiresolution Hair Modeling and Editing. *ACM Transactions on Graphics* 2002, 21, 3, 620-629.
3. Van Gelder And Wilhelms. An Interactive Fur Modeling Technique. *Proceedings of the conference on Graphics interface 1997*, p.181-188, May 1997, Kelowna, British Columbia, Canada.

4. Lee, Chen, Leu And Ouhyoung. A Rotor Platform Assisted System for 3D Hairstyles. *Journal of WSCG*, 10 (1-3), 271-278.
5. Sylvain Paris, Hector M. Briceño, Francois X. Sillion. Capture of Hair Geometry from Multiple Images. To be presented in Siggraph 2004.
6. Isaac Rudomin, Benjamin Hernandez. Hair Paint. To be Presented in Computer Graphics International 2004 (CGI), Creta, Grecia, Junio 16 – 19, 2004
7. Koon And Zhiyong. A Simple Physics Model to Animate Human Hair Modeled in 2D Strips in Real Time. *Proceedings of Eurographics Workshop 2002*, 127-138.
8. Daldegan, Thalmann, Kurihara And Thalmann. An integrated system for modeling, animating and rendering hair. *Computer Graphics Forum (Eurographics '93)* 12, 3 (1993), 211–221.
9. Anjyo, Usami, And Kurihara. A Simple Method for Extracting the Natural Beauty of Hair. *Proceedings of ACM SIGGRAPH 92*, 26(4), ACM, 111-120.
10. Kurihara, Anjyo And Thalmann, 1993. Hair Animation with Collision Detection. Computer Graphics Laboratory, Swiss Federal Institute of Technology. *Computer Graphics International*, 1996.
11. Chang, Jin, And Yu. A Practical Model for Hair Mutual Interactions. In *Symposium on Computer Animation 2002*, ACM SIGGRAPH, p. 73-80.
12. Kim And Neumann. A Thin Shell Volume for Modeling Human Hair. In *Computer Animation 2000*, Philadelphia, IEEE Computer Society, 121-128.
13. Plante, Cani, And Poulin. A Layered Wisp Model for Simulating Interactions Inside Long Hair. *Eurographics Workshop on Computer Animation and Simulation 2001*, 139–148.
14. Chen, Saeyor, Dohi, And Ishizuka. A System of 3D Hairstyle Synthesis Based on the Wisp Model. *The Visual Computer*, 15(4), 159-170.
15. Kim And Neumann. Interactive Multiresolution Hair Modeling and Editing. *ACM Transactions on Graphics* 2002, 21, 3, 620-629.
16. Bertails, F. Kim. Cani, M. Neumann, U. AdaptiveWisp Tree – a multiresolution control structure for simulating dynamic clustering in hair motion. *Symposium on Computer Animation'03*, July 2003
17. Bando, Chen, Nishita. Animating Hair with Loosely Connected Particles. *Computer Graphics Forum* vol 22, n° 3 (2002)
18. Kajiya And Kay. Rendering Fur with Three Dimensional Textures. In *Computer Graphics (Proceedings of ACM SIGGRAPH 89)*, 23(4), ACM, 271-280.
19. Marschner, Jensen, Cammarano. Light Scattering from Human Hair Fibers. *Proceedings of ACM SIGGRAPH 2003*.
20. Benjamin Hernandez, Isaac Rudomin. Styling by painting and real time animation of hair using basis-dependent hair strands. *WSCG 2004 POSTER proceedings ISBN 80-903100-6-0*
21. Bloomenthal, J. 1990. Calculation of Reference Frames Along A Space Curve. In A. Glassner, editor, *Graphics Gems*, Academic Press.
22. K. Ward, N. Galoppo, M. Lin. Modeling hair influenced by water and styling products. *Proceedings of Computer Animation and Social Agents*, 2004.

# Volume Preserved Deformation of Human Face with Hand

Yohei Iwasaki, Isamu Mizuno, and Toyohisa Kaneko

Toyohashi University of Technology,  
Department of Information and Computer Sciences  
kaneko@ics.tut.ac.jp

**Abstract.** This paper is concerned with volume preserved deformation of a human face pressed or lifted at a point on the face with a human hand. The 3D models of the human face and hand are made from 3D CT data. Elasticity is realized with a spring-mass network. Although most past works treated volume preservation by controlling the shape of surface, this paper proposes a new method that is based on the volume of each constituent tetrahedron.

CT data provides polygonal data of soft tissue and bones with the marching cubes algorithm. Then nodes are allocated using a bubble mesh algorithm, which is capable of placing nodes approximately in an equal distance.

Based upon the network models, the newly proposed method was applied to two examples of soft object interactions in order to demonstrate its applicability and usefulness. The first was a collision of a hand to a part of the cheek and the second was lifting a part of the lower lip with a hand. Good agreement between the actual and simulated scene was observed.

Keywords: Volume preserved deformation, facial simulation, CT data, deformation with external force

## 1 Introduction

In virtual reality, a real-time operation is carried out for various objects including soft objects such as human face, organs etc. This requires modeling of soft objects and mutual interaction between soft and hard objects. This paper deals with the interaction between human organs, especially between a face and a hand.

To create soft objects, CT(computing tomography) data generally available at hospitals are exploited in this paper. The marching cubes algorithm is applied to CT data which are three-dimensional, providing the polygonal surfaces of skin and also bones.

The property of maintaining constant volume during mutual interaction of two soft objects is very important. Any deformation on the skin is thought to be carried out keeping the total volume in a local region, while bones are treated as rigid bodies. If a part on a soft object is pressed down, there must be a bulge in its neighboring area. This can be accomplished under the assumption of volume preserved deformation.

### 1.1 Related Works

The work of Terzopoulos et al. [1] is most well known for modeling soft objects in the field of computer graphics. It introduced a spring-mass network model to simulate

deformation of such primarily two-dimensional objects as a cloth and a flag. Then it was extended to a 3D surface that is represented with second-order polynomials [2]. These works were applied to make animations of deforming objects such as teapots and cars [3, 4].

Most of the above works are concerned with deformation of surfaces [4] alone, where the number of nodes is substantially smaller than that of a true 3D model containing internal nodes. In order to reduce the computation time, James and Pai [5] proposed a BEM (boundary element method) based upon a pre-computed matrix describing the relationship between external force and displacement. Although a FEM (finite element method) is known to be more accurate and more time-consuming, Capell et al. [6] proposed a fast computing FEM method based on a simple structure for animating animal movement.

There are works for soft tissue deformation in the area of medical simulations [7–10], where a part of organs is moved or cut with a knife usually in real-time.

There are a number of works for human expressions (for instance [11–13]). This paper does not address the problem of generating human expressions with controlling facial muscles etc. but the problem of making expressions with external force, for instance with a hand.

## 1.2 Node Allocation with the Close-packed Structure

A spring-mass network is used here to represent a soft or elastic object. Initial allocation of nodes is carried out based upon the close-packed structure. This is analogous to cutting or sculpturing a close-packed crystal object into a 3D shape that approximates the original object.

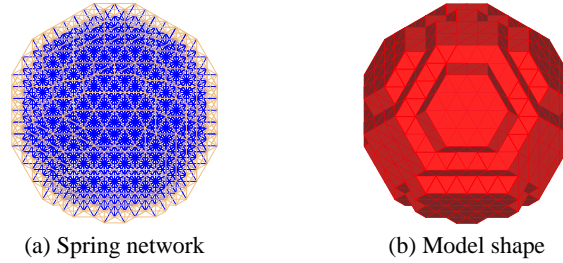
It is known that there are two types of close-packed structure: cubic and hexagonal close-packed structure [14]. In either case, the resulting edge length will be either 1 or  $\sqrt{2}$ . (The occurrence percentage of  $\sqrt{2}$  is about 14%.) In our paper, the cubic close-packed (ccp) structure is employed. In this case there are 12 neighboring nodes per node.

## 1.3 3D modeling with Close-packed Structure

The marching cubes algorithm is applied to given CT data with two thresholds (one is between the air and soft tissue and the other between the soft tissue and the bone), providing the surface skin and bone polygons. Thus the facial object is represented as the region of soft tissue between the skin and bone.

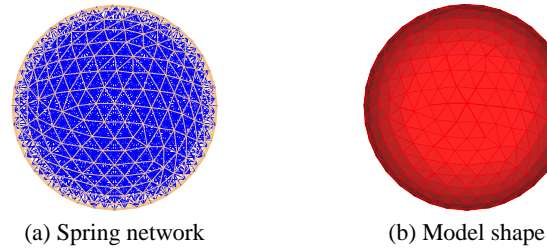
Then this region is covered by a 3D cubic close-packed structure which is slightly larger than the original region. Let  $r$  be the desired average edge distance of the network model to be constructed. Consider each node of the surface nodes of ccp structure. If its distance to the object surface is larger than  $.5r$ , the node is judged to be external and is removed. The node whose distance is smaller than  $.5r$  remains. For all the remaining nodes, their distance to the surface are again computed. Those nodes whose distances are larger than  $.5r$  are then labeled as internal nodes. Those smaller than  $.5r$  are as surface nodes.

Then the surface nodes of this resulting ccp model are moved onto the surface of the given object with the bubble mesh algorithm [15]. After the surface nodes are adjusted, this process is carried out for the internal nodes.



**Fig. 1.** Initial spring network model with close-packed structure

As an example, consider a sphere of radius 5cm. Let 2.5mm be the average distance. The initial network and shape are shown in Figure 1. Then the bubble mesh algorithm is carried out. The resulting network model is shown in Figure 2 and the statistics of edge distances is tabulated in Table 1. It is seen that the resulting model has fairly uniform edge lengths. The shape is very close to a sphere while the original shape is not smooth. It should be noted that node-to-node connection is set exactly as that of the initial ccp structure even if each node is slightly moved after the application of the bubble mesh algorithm.



**Fig. 2.** Spring network model for a sphere

**Table 1.** statistics of edge lengths[mm]

	No. of edges	Min	Max	Avg.	Sigma
before	5196	2.61	7.43	4.99	0.32
after	5196	3.37	6.71	5.00	0.24

## 1.4 Spring constant

One important parameter of a spring network is a spring constant assigned to each edge. In our case, this value is actually measured using a specially designed system [16] with a force sensor (Nitta UFS2012A) and a laser distance measurement tool (Keence). Local spring constants are set so that displacement value on the cheek with pressed force is the same between the virtual system realized with this network model and the real system. We measured nine points on the face. Spring constants other than these measured points are interpolated.

## 2 Volume Preserved Deformation

### 2.1 Kinematics

Deformation in a network is carried out based upon Newton's kinematics and Hook's law applicable to a spring. The kinematics equations [17, 18] are given as:

$$\begin{aligned} \mathbf{v}_i(t) &= \mathbf{v}_i(t - \tau) \exp\left(-\frac{c}{m}\tau\right) \\ &\quad + \frac{\mathbf{f}_i}{c} \{1 - \exp\left(-\frac{c}{m}\tau\right)\} \\ \mathbf{u}_i(t) &= \mathbf{u}_i(t - \tau) + \left(\frac{m}{c}\mathbf{v}_i(t - \tau) - \frac{m}{c^2}\mathbf{f}_i\right) \\ &\quad \times \{1 - \exp\left(-\frac{c}{m}\tau\right)\} + \frac{\mathbf{f}_i}{c}\tau \end{aligned} \quad (1)$$

Here  $\tau$  is an iteration time interval (set to be 0.001 sec. in our experiments).  $\mathbf{v}_i(t)$  and  $\mathbf{u}_i(t)$  are velocity and position of node  $i$ .  $\mathbf{f}_i$  is the sum of force at node  $i$ .  $m$  is the mass which is computed based upon the CT using the density of water.  $c$  represents viscosity and was set using Equation (2) [19].

$$c = 2\sqrt{k \cdot m} \times \alpha \quad (2)$$

This equation provides the critical viscosity  $c$  for  $\alpha=1$ , mass  $m$ , spring constant  $k$ .

### 2.2 Volume Preserved Deformation

The human soft tissue contains water more than 70% so that volume is thought to be preserved during deformation. Terzopoulos et al.[4] proposed a method of placing additional (virtual) force along the normal direction of a point on the surface for volume preservation.

In this paper we propose a new method based upon constituent tetrahedrons. The soft object, human face, is already represented as a spring-mass network. This network as a 3D object can be broken into a set of tetrahedrons. Volume preservation is carried out in a subset of these tetrahedrons, namely a group of tetrahedrons included in a local region.

Volume preserved deformation is carried out as follows:

Step 1 (Computation of total volume)

Compute the volume of each of tetrahedrons and obtain the total volume  $T(t)$  at time  $t$  and compare it to the initial total volume  $T(0)$ . If  $T(t) - T(0) > 0$ , all the tetrahedrons are pressed with force from external to internal direction. If  $T(t) - T(0) < 0$  or  $0$ , then force is from the internal to the external. The force direction is along the normal direction of each of four planes of a tetrahedron.

Step 2 (Volume preserved force):

The force of the  $i$ -th tetrahedron for preserved volume is given as:

$$F_{V_j}(t) = p(V_j(t) - V_j(0)) \quad (3)$$

where  $V_j(0)$  and  $V_j(t)$  are volume at time 0 and time  $t$ , respectively.  $p$  is a constant determined experimentally and is set to be  $5.0 \times 10^7$ . (This value is related to the volume metric compression ratio showing force divided by displaced or compressed volume.) If  $V_j(t) - V_j(0) > 0$ , force is from internal to external. Otherwise, the opposite direction. If this direction is the opposite of the global direction at Step 1, then no force is given.

The force is distributed in proportion to each area of four planes per tetrahedron. The distribution ratio to the plane  $j$  is  $q_j = S_j / (S_1 + S_2 + S_3 + S_4)$  where  $S_j$  is the area of plane  $j$ .

Step 3 :

Equation (3) gives total volume preserved force. It is distributed to four planes to the normal directions ( $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \mathbf{n}_4$ ). At node  $i$ , the total force  $\mathbf{f}_i$  is given as

$$\mathbf{f}_i = F_{V_j(t)}(\mathbf{n}_1 q_1 + \mathbf{n}_2 q_2 + \mathbf{n}_3 q_3) \quad (4)$$

where the assumed normal directions are  $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ . Furthermore, the total force at a node is obtainable by summing up the forces from all the tetrahedrons sharing the node.

Step 4 (Deformation):

Hook's force and volume preserved force are added, providing the edge lengths and node coordinates are updated.

Step 5 (Termination):

The iteration is terminated if the following two conditions are met.

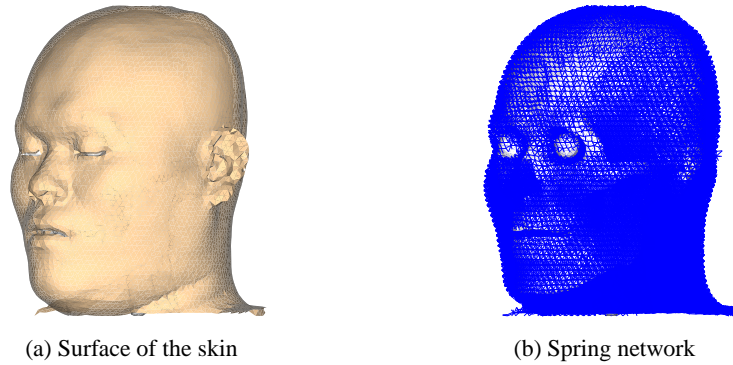
- 1)  $|T(t) - T(0)|$  is below a threshold  $\epsilon T(0)$  where  $\epsilon = .01$ .
- 2) Change per iteration in all the node coordinates is below a threshold  $\epsilon' 2r$ , where  $r$  is the average edge length and  $\epsilon'$  is set to be  $.01$ .



### 3 Deformation of Face with Hand

#### 3.1 Spring network of face

A spring-mass network model of a face was constructed with the average edge length of 5mm (i.e.  $r=2.5\text{mm}$ ). The resulting statistics is shown in Table 2 (see Figure 3).



**Fig. 3.** Spring network model of a human face

**Table 2.** Statistics of edge length[mm]

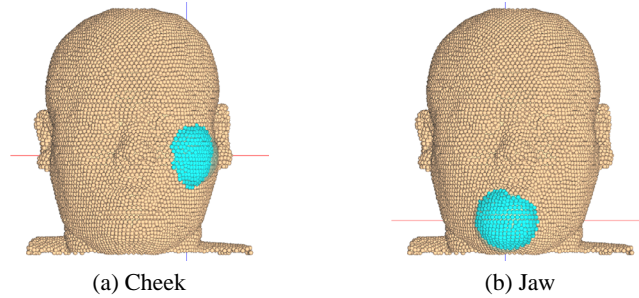
No. of edges	Min	Max	Avg.	Sigma
409791	2.042	11.199	5.069	0.257

**Table 3.** Statistics of node lengths[mm]

	No. of edges	Min	Max	Avg.	Sigma
Cheek	4880	2.885	7.755	5.122	0.175
Jaw	3359	2.896	8.374	5.004	0.385

#### 3.2 Simulation

Although details are omitted here, a similar process was carried out to obtain a spring-mass network model based upon CT data of a human hand with the same average length  $r$  (i.e. 5mm). Bones included in either model are treated as rigid bodies, namely objects that do not change their shape.

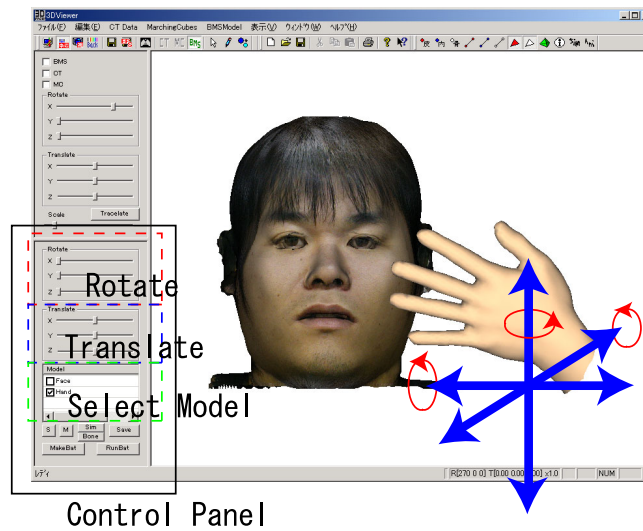


**Fig. 4.** Locations for evaluation

The hand model was moved towards the face model. Two types of deformative simulation were performed: (1) press on the cheek with a hand, and (2) lift-up of the lower lip with a finger.

### 3.3 GUI for Hand Motion

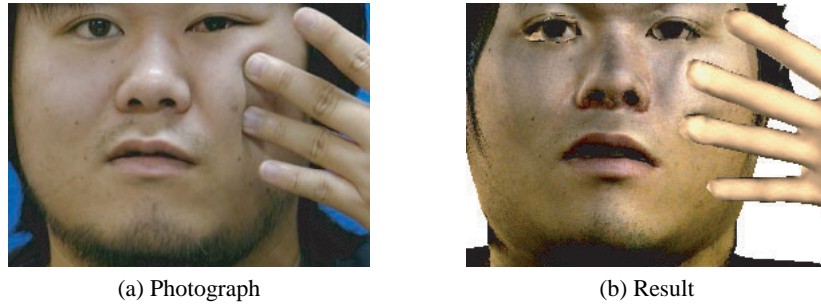
A graphical user interface GUI as shown in Figure Figure 5 was constructed in order to specify the motion of a hand. There are two motions: straight motion and rotational motion.



**Fig. 5.** User Interface

### 3.4 Pressing on cheek with hand

The hand model was moved towards the cheek by a increment of  $1.25\text{mm}$ . Figure 6(a) shows a frontal photograph and Figure 6(b) shows a frame of the simulated results.



(a) Photograph

(b) Result

**Fig. 6.** A part of cheek pushed with fingers



**Fig. 7.** A part of textured cheek pushed with fingers

The hand was moved from the left side (from the model, right from the observer) to the right side. The simulation process is carried out for each motion of  $1.25\text{mm}$ . Table 4 tabulates the time required for the case with and the case without volume preservation. The computation was carried out on a PC with 1.8GHz and 1GB memory. The time spent was 1.0 Hours and 1.5 Hours without and with volume preservation, respectively. Figure 7 shows the same picture observed from a view position with 45 degree rotation. The amount of bulge on the cheek as a result of hand motion is observed to be about 3mm, which is about the same of the bulge with simulation. We also checked changes in total volume and observed a preservation rate of 99.99% for face and that of 99.98% for hand.

**Table 4.** Computation time

		Vol.Pres.				No Vol.Pres.	
Frame No	Motion [mm]	exec time/frame	exec time/frame	Frame No	Motion [mm]	exec time/frame	exec time/frame
1	0.125	00:04:08	00:03:38	11	1.375	00:05:12	00:03:19
2	0.250	00:04:59	00:04:03	12	1.500	00:05:05	00:03:19
3	0.375	00:02:41	00:03:37	13	1.625	00:05:32	00:03:19
4	0.500	00:04:39	00:03:18	14	1.750	00:05:33	00:03:19
5	0.625	00:04:46	00:03:19	15	1.875	00:05:19	00:03:20
6	0.750	00:04:44	00:03:19	16	2.000	00:05:17	00:03:19
7	0.875	00:05:07	00:03:19	17	2.125	00:05:58	00:03:19
8	1.000	00:05:07	00:03:20	18	2.250	00:05:51	00:03:20
9	1.125	00:05:01	00:03:19	total		01:30:35	1:01:05
10	1.250	00:05:36	00:03:19	average		00:05:02	00:03:20

### 3.5 Lifting a part of the lower lip

The next simulation was to lift up a right part of the lower lip with a hand. Figure 8(a) shows a photograph while Figure 8(b) shows a simulated result. The agreement is reasonable. The total simulation time was about 1.0 Hours on the PC mentioned above.



(a) Photograph



(b) Result

**Fig. 8.** The lip lifted upwards with fingers

## 4 Conclusion and future works

We showed volume preserved deformation during the interaction of two soft objects: human face and hand. The following are the main contributions of this paper.

- 1) a method for constructing a spring-mass network model based upon CT data.
- 2) a new method of volume preservation based upon constituent tetrahedrons.
- 3) verification of the proposed methods with two examples.

Future work is to devise a method for speeding up simulation.

## References

1. Terzopoulos, D., Platt, J., Barr, A., Fleischer, K.: Elastically deformable models. In: Proc. of SIGGRAPH '87. (1987) 205–214
2. Terzopoulos, D., Metaxas, D.: Dynamic 3d models with local and global deformations: Deformable superquadrics. Volume 13. (1991) 703–714
3. Faloutsos, P., van de Panne, M., Terzopoulos, D.: Dynamic free-form deformations for animating synthesis. Volume 3. (1997) 201–214
4. Terzopoulos, D., Platt, J., Barr, A., other: Elastically deformable models. *Computer Graphics* **21** (1997) 205–214
5. James, D., Pai, D.K.: Artdefo: Accurate real time deformable objects. In: Proc. of SIGGRAPH '99. (1999) 66–72
6. Capell, S., Green, S., Curless, B., Duchamp, T., Popovic, Z.: Interactive skeleton-driven dynamic deformation. In: Proc. of SIGGRAPH 2002. (2002) 586–593
7. Cotin, S., Delingette, H., Ayache, N.: Real-time elastic deformations of soft tissues for surgery simulation. Volume 5. (1999) 62–73
8. Kaehler, K., Haber, J., Seidel, H.P.: Geometry-based muscle modeling for facial animation. In: Proc. of Graphics Interface 2001. (2001) 37–46
9. Bruyns, C., Ottensmeyer, M.: Measurement of soft-tissue mechanical properties to support development of a physically based virtual model. In: Proc. of MICCAI 2002. (2002) 282–289
10. Hauth, M., Gross, J., Strasse, W., Buess, G.: Soft tissue simulation based on measured data. In: Proc. of MICCAI 2003. (2003) 262–270
11. Lee, Y., Terzopoulos, D., Waters, K.: Realistic modeling for facial animation. (1995) 55–62
12. Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., D.H.Salesin: Synthesizing realistic facial expressions from photographs. In: Proc. of SIGGRAPH '98. (1998) 75–84
13. Noh, J.Y., Neumann, U.: Expression cloning. In: Proc. of SIGGRAPH 2001. (2001) 277–288
14. Hales, T.: Sphere packings i. In: *Discrete and Computational Geometry*. Number 17 (1997) 1–51
15. Shimada, K.: Physically-based mesh generation: Automated triangulation of surfaces and volumes via bubble packing. Ph.D. thesis, Massachusetts Institute of Technology (1993)
16. Mizobuchi, T., Hirota, K., Kaneko, T.: Deformation model of a human face based on measurements (in japanese). *Trans. of IEICE* **J85-D-II** (2002) 466–474
17. Hirota, K., Kaneko, T.: Haptic representation of elastic objects. *Presence* **10** (2001) 525–536
18. Hirota, K., Kaneko, T.: Implementation of elastic object in virtual environment, advances in human factors / ergonomics. *Design of Computing Systems* **21B** (1997) 969–972
19. Hirota, K., Tanoue, Y., Kaneko, T.: Simulation of three-dimensional cracks. *Visual Computing* **16** (2000) 371–378

# Point Based Rendering and Displaced Subdivision for Interactive Animation of Crowds of Clothed Characters

Isaac Rudomín and Erik Millán

Computer Science Department ITESM-CEM  
Km 3.5 Carretera al lago de Guadalupe  
Atizapán de Zaragoza, Estado de México, Mexico, 52926  
Phone: +52 55 5864 5659  
rudomin@itesm.mx

**Abstract.** This paper presents a set of techniques that can be used for managing different levels of detail in an interactive animation. Polygon subdivision schemes and displacement mapping are used to obtain higher resolution meshes, while point-based rendering is used to reduce the detail in the model when viewed at a distance. Data from subdivision is used in point-based rendering for a better distribution of samples. An application with crowds of characters and physically-based clothing is used to show the advantages of this approach.

**Keywords.** Level-of-detail, subdivision surfaces, displacement maps, point-based rendering, crowds.

## 1 Introduction

A common problem in computer graphics is to achieve a balance between a nice and smooth display of models and a fast and constant frame rate. Hence, level-of-detail techniques are commonly used to display more polygons when objects have a larger display area, and to show less when the area gets smaller. Many algorithms have been used to increase the detail in a model. One of the most widely used are subdivision surfaces, designed to work with polygonal meshes, which makes their use efficient in commodity hardware. There are other methods to achieve higher detail, such as bump mapping and vertex displacement, which used in conjunction with subdivision, produce better appearance while keeping a reduced number of primitives and simplifying the edition process.

The animation of clothed humans represents a complex problem for multiresolution techniques, as every vertex changes its position frame after frame due to external forces. The use of a multiresolution technique to generate additional level of detail can produce nice results while maintaining a small number of primitives. There are many approaches to reducing detail in complex models. Most include a preprocessing phase which will, either create less detailed models, or sort vertices for their later display. In this paper, we propose the use of a particular version of point-based rendering, adapted for animated clothed characters, in combination with a displaced subdivision scheme.

## 2 Previous work

### 2.1 Mesh refinement

The publication of the papers by Catmull and Clark [1] and Doo and Sabin [2] marked the beginning of the use subdivision techniques for surface modeling. Subdivision elegantly addresses many issues in computer graphics, such as arbitrary topology, scalability, uniformity of representation, numerical stability and code simplicity. There is a straightforward way to classify most of the subdivision schemas based on four criteria: the type of refinement rule (face or vertex split); the type of generated mesh (triangular or quadrilateral); whether the new vertices are approximated or interpolated; and the smoothness of the limit surfaces.

For our purposes, we analyze Loop and Butterfly schemes, which are face-split schemas for triangular meshes. The Loop scheme is a simple approximating face-split scheme for triangular meshes proposed in [3]. The Butterfly scheme was first proposed in [4] and defined on arbitrary triangular meshes. In [5] a modification of the Butterfly scheme was proposed, which guarantees the generation of C1-continuous surfaces.

There are more techniques to increase detail in a mesh. Bump mapping, as proposed in [6], is a technique to change the appearance of a surface by altering its normal vectors. This gives the impression of a higher resolution without the use of additional geometry. In contrast, displacement mapping is a method to present surface detail by defining an offset from a base surface. Its main difference from bump mapping is that the position of each vertex is modified. This method results in a more realistic rendering, where displaced geometry can be seen in the silhouettes.

An efficient implementation of displacement mapping is the use of displaced subdivision surfaces, introduced by Lee in [7]. These surfaces are represented as a scalar-valued displacement over a smooth domain surface. Here, surface detail can be represented using a simple mesh and a displacement map. In [8], displacement mapping is implemented in hardware to produce efficiently a more detailed geometry.

### 2.2 Mesh reduction

One of the first approaches for polygonal mesh reduction was sampling a surface with fewer vertices, then joining these new vertices to create a new polygonal structure. Turk [9] proposes randomly sampling new points within the triangles of a mesh, joining them to the original vertices, and removing the old vertices progressively. A similar approach is presented in [10], where vertices are removed one by one through a preprocessing phase, creating a hierarchical structure that can be rendered with any desired level of detail.

It is also possible to construct impostors as proposed in [11] and [12]. However, while some animation may be based in keyframes, cloth simulation is recalculated dynamically every frame, complicating the use of fixed animations.

A different approach for reducing the detail in a model is *point-based rendering*. Point-based rendering, as proposed by Grossman [13], uses points as primitives rather than polygons. In order to cover a larger area, points may be rendered as splats, flat shapes that cover many pixels. Rusinkiewicz [14] uses a technique called QSplat, based on a hierarchy of bounding spheres to sort and group the pixels. Similarly, Pfister [15] introduces Surfels, point primitives which are projected to the faces of cubes in a hierarchical octree. In these approaches, there is a preprocessing phase in which the points are classified and representations for lower detail are obtained. However, in an animation, this preprocessing would be necessary for every change in the position of the vertices, which, in this case, dramatically increases the complexity.

### 3 Multiresolution animation of virtual clothed humans

There has been a great amount of work published on the display and render of crowds of characters with different levels of detail. In [16], subdivision is proposed as a technique to increase detail in crowds, but no details on the implementation or its efficiency are shown. Aubel [17] presents a multiresolution approach to create the surface of a human; however, the mesh generation occurs once before the animation, so no interactive level of detail is achieved. Both [17] and [12] use animated impostors to represent a huge amount of characters in a scene. The disadvantage of this approach is that dynamic calculation of cloth subject to forces produce different animations that cannot always be predicted.

Many researchers have attempted to model the dynamic behavior of cloth and garments. There are two major tasks that a clothed character simulation system performs, which are modeling the behavior of cloth and detecting collisions between cloth and the character. Methods for modeling the behavior of cloth can be classified into geometric, physically based, and hybrid [18][19]. Geometric approaches try to mimic the shape of cloth, physically based methods simulate actual behavior by using physics, and hybrid methods combine the advantages of both approaches. Accurate physically-based cloth simulation systems, such as [20], define an implicit Euler integration scheme that allows large time steps in the simulation, even for stiff springs, as required to model the behavior of cloth. However, each step involves significant computation.

In previous work [21] we have implemented a hybrid geometric-physically based approach, using different algorithms for garments that behave differently. For those that stick to the body, the algorithm is basically geometric, a variation of skinning algorithms but using an ellipsoidal approximation to the body to simplify the calculation of clothing-body collision detection. The vertices in the cloth-mesh are associated with a body part and move with it. The ellipsoids generate a field used for collision detection. For free flowing garments, we augment this algorithm by adding physically-based simulation. The simulation shown in this paper is based on this work.

The animation of meshes in cloth simulation implies the calculation of the position of every vertex for each animation frame. However, more detailed meshes

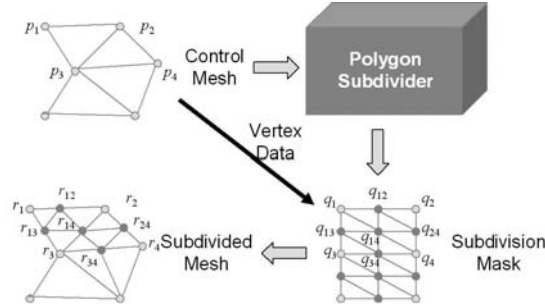


would demand a higher processing time. Furthermore, when a mesh has a small display area, some of the detail will not be appreciated by the user. Hence the convenience of using a coarser model helped by multiresolution techniques. For cloth simulation, we propose the use of subdivision surfaces and displacement mapping to increase the detail in a mesh, and point-based rendering to reduce its detail when required.

### 3.1 Adaptive subdivision surfaces for animation

The most common representation for models used in interactive applications are polygonal meshes, and particularly triangular meshes, since they are best displayed by graphics hardware. Hence, triangular subdivision schemes are the most eligible option for this application. It is also desirable for fabric garments that the points in the control mesh are also part of the subdivided mesh, since collisions with the underlying body are calculated for the original points. According to these considerations, the Butterfly subdivision scheme was selected, though similar performance might be achieved through different subdivision schemes.

Given a control mesh, we create an object called *polygon subdivider*. A polygon subdivider will analyze the topology of the mesh and add new triangles up to a certain subdivision level.



**Fig. 1.** Overview of the subdivision process

The position of the vertices in an animated cloth is likely to change in every frame, making necessary to calculate subdivision constantly. Nevertheless, it is common in models used for cloth animation, and for most types of animation, that the topology of the model remains constant. For this reason, some of the data can be calculated in advance. Consider we have a mesh with vertices  $p_0, p_1, \dots, p_i$ . A subdivision mask produced will produce, for each level, a set of vertices  $r_j$  whose position will be obtained in the following manner:

$$r_j = \sum_{k=0}^i w_k p_k$$

where  $w_k$  is the weight associated to each vertex in the control mesh to obtain the subdivided vertex. The values for each  $w_k$  are produced by the polygon subdivider and depend on the subdivision schema selected. These values are stored on a *subdivision mask*, and the actual value of the vertex depends on the position of the control mesh in a given moment. The polygon subdivider will produce a set of subdivision masks, one for each level of subdivision desired. In this way, a weighted sum is the only computation required to obtain the position for a vertex at a given level of subdivision.

Subdivision masks can also handle the calculation of normal vectors and texture coordinates, requiring only a weighted sum to obtain their value. Texture coordinates are not likely to change through the animation, so they can be computed in advance. In addition, subdivision masks can be extended to include new information, as it will be described later on this paper.

Once we have a set of masks, we can display a model at different levels of detail using adaptive subdivision. The subdivision level for a given triangle is selected according to the distance from a point of view. In our work, we used a threshold based on the area of a polygon and on the distance to the camera to select the detail. An overview of the entire subdivision process can be appreciated in figure 1.



**Fig. 2.** a) Masks used for polygon subdivision. b) Sample subdivided blouse with additional detail on the silhouette

Due to the adaptive subdivision, contiguous triangles can have different detail. To avoid the development of holes, triangles surrounded by more detailed neighbors are retriangulated according to the masks in Figure 2.a.

The usage of different levels of detail based on the distance of the model to the camera produces good results; however, lack of detail is highly noticeable by a user in silhouettes. This is why faces in the silhouette of the model are displayed with a higher level of detail. An example of this adaptive subdivision is exemplified in Figure 2.b. The image on the left has a higher detail on its contours, which can be better appreciated on the right image that has been artificially rotated to help observation.

### 3.2 Displacement mapping

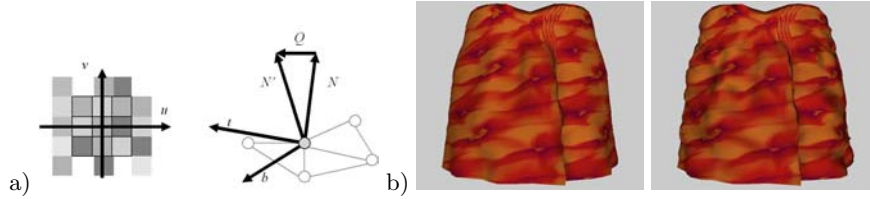
Additional level of detail can be added to a mesh through displacement mapping. A *displacement map* is designed as a grayscale image mapped to a polygonal mesh through texture coordinates; the intensity in the texture will indicate the magnitude of the displacement. Texture coordinates can be used for displacement to simplify the creation of displacement maps in previously textured meshes.

The original design of subdivision masks is extended to support displacement mapping. A displaced vertex  $d_j$  is obtained from a vertex  $r_j$  through

$$s_j = r_j + dN_j$$

where  $N_j$  is the normal vector of the vertex and  $d$  is the displacement value obtained from the map. This requires little additional effort, as these values have already been precalculated in the subdivision mask.

However, when a vertex is displaced, its normal vector will also change. This modified normal vector is fundamental to appreciate correct shading on the displaced model. However, due to the simplifications in subdivision masks, it is impossible to know the displaced positions of neighboring points at a given level of detail, crucial in normal calculation for subdivision surfaces.



**Fig. 3.** a) Normal perturbation for displacement mapping. b) Left: Skirt rendered without displacement mapping. Right: Skirt using displacement mapping

Nevertheless, we can use the subdivision masks for tangent vectors  $t_j$  and  $b_j$ , used to obtain the normal vector  $N_j$ . Hence, can obtain a perturbation vector  $Q_j$  as proposed in [6].

$$Q_j = u_j(N_j \times t_j) - v_j(N_j \times b_j)$$

Here,  $u_j$  and  $v_j$  are the partial derivatives of the displacement map, and can be obtained by simply subtracting neighboring values on this map. By adding this vector to the original normal we obtain the displaced normal. This is shown in figure 3.a. An example of a skirt rendered with and without displacement mapping is shown in Figure 3.b.

### 3.3 Point-based rendering

The use of subdivision surfaces to increase the level of detail of a mesh produces a large number of polygons and vertices. This extended set of points provides additional samples for point-based rendering, and consequently a better sampling of the space occupied by a model. In [22], samples are taken from the surface of a polygonal mesh. While this is useful, this paper does not propose a method to increase the level of detail in a mesh. Besides, the use of samples from a subdivided mesh provides a better representation for coarser polygonal meshes, as sampled geometry contains a more natural shape. However, the sampling and stratification techniques from [22] method can be easily integrated with our approach.

An octree hierarchy is used to classify the points in a mesh. This octree is traversed in a breadth-first manner to obtain a uniform sample distribution, providing a list of hierarchically-ordered vertices similar to [23].

For each cell, all subcells are visited with a random order. Once all of them are visited, they are visited again until all vertices within have been visited or until the number of points to render is reached. This random selection is used at every octree level, so the sampling distribution is always regular. The selected samples are stored in a list. This random list may be generated constantly, which produces some noise in display and improves the appearance of distant objects.

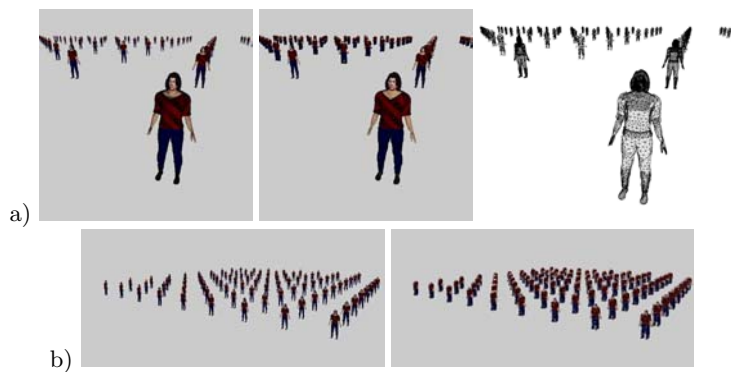


**Fig. 4.** Sample point-based renders of a blouse. Point size has been reduced to appreciate samples.

The use of an octree hierarchy produces good results, but it can be replaced to better adapt to a specific domain. In our clothing application, vertices in cloth models are mapped to ellipsoids within the body of the human, as in [21]. Based on this, in this paper rather than straight octrees we first classify vertices according to the closest ellipsoid and then, within each ellipsoidal category, we apply the octree algorithm. In this way, it is guaranteed that all parts of the body are visible from a distant point of view.

To display a point-based model, a number of samples is selected based on the distance from the model to the camera. These samples are picked from the list generated by the hierarchy, and calculated according to the subdivision mask

which gives us a way to obtain additional samples where needed. An example of this render is displayed in Figure 4.



**Fig. 5.** a) Left: Crowd using traditional render. Middle: Crowd using different levels of detail. Right: Wireframe of the crowd. Point size has been reduced to contrast points from polygons. b) Left: Traditional render of a crowd. Right: Crowd rendered using our approach.

In order to cull faces when displaying the model, most of the point-based algorithms, such as [13], obtain normal cones. Normal cones select an area where a set of points is not visible. However, in animated meshes, these areas would have to be constantly recalculated. To implement backface culling, we obtain, for each vertex, the dot product between the normal for each point and the direction to the camera. In addition, we modify the alpha value in contours to generate an antialiasing effect. To improve performance, it is possible to obtain these normals for blocks of points, however, this is left as future work.

## 4 Results

In order to test our approach, a sample application was programmed displaying 100 cloned models. A culling algorithm was used to only draw the visible objects. The animation for these models was calculated once and used for all of them. However, their render was performed adaptively according to their position and distance to the point of view.

Figure 5.a shows a scene with multiple levels of detail. According to the distance to the camera, models are rendered either by adding detail through subdivision or by reducing detail through point based render. In every model, a different number of vertices is selected, and the position and normal vectors for these vertices are calculated interactively. Figure 5.b show a comparison between a crowd rendered using point-based rendering and a crowd using regular rendering. In this case, point-based render significantly increases the frame rate.

**Table 1.** Frames per second for different renderings of the crowd.

<b>Figure</b>	<b>5.a middle</b>	<b>5.a left</b>	<b>5.b right</b>	<b>5.b left</b>
<b>LOD</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>No</b>
<b>Characters</b>	62	62	100	100
<b>Vertices</b>	359,058	1,263,126	366,777	2,037,300
<b>Triangles</b>	45,504	421,042	0	679,100
<b>Points</b>	222,546	0	366,777	0
<b>FPS</b>	5fps	5fps	33fps	3fps

Rendering details from different points of view are shown in Table 1. In close-ups, rendering time increases, but a higher detail is generated for closest models, improving their appearance. We are still working in implementing several optimization techniques that we expect can still yield significant increases in the frame rate.

## 5 Conclusions and future work

The use of multiresolution approaches to both increase and decrease the level of detail of models can produce nice looking results at interactive frame rates. The combination of these two methods proves useful to obtain good transitions between levels of detail for interpolating subdivision schemes in animation, but should also produce nice transitions in the case of approximating subdivision schemes.

In addition, the use of subdivision data for point-based rendering provides a better quality in the sampling used to pick rendered points. The partial pre-calculation of subdivision masks provides a way to incorporate dynamic level of detail interactively.

There are many optimizations possible for these framework. Corresponding information on different characters can be cached temporarily to be used by these different instances. Physical calculations for animations can be calculated only up to a certain level of detail, replacing them with less expensive calculations. Finally, this approach could be implemented in hardware through vertex and pixel shaders, removing the additional load from the main processor.

## References

1. Catmull, E.E., Clark, J.H.: Recursively generated b-spline surfaces on topological meshes. *Computer Aided Design* **10** (1978) 350–355
2. Doo, D., Sabin, M.: Analysis of the behaviour of recursive division surfaces near extraordinary points. *Computer Aided Design* **10** (1978) 356–360
3. Loop, D.: Smooth subdivision surfaces based on triangles. Master’s thesis, University of Utah, Department of Mathematics (1987)
4. Dyn, N., Levine, D., Gregory, J.A.: A butterfly subdivision scheme for surface interpolation with tension control. *ACM Trans. Graph.* **9** (1990) 160–169

5. Zorin, D., Schröder, P., Sweldens, W.: Interpolating subdivision for meshes with arbitrary topology. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM Press (1996) 189–192
6. Blinn, J.F.: Simulation of wrinkled surfaces. In: Proceedings of the 5th annual conference on Computer graphics and interactive techniques, ACM Press (1978) 286–292
7. Lee, A., Moreton, H., Hoppe, H.: Displaced subdivision surfaces. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co. (2000) 85–94
8. Gumhold, S., Hüttner, T.: Multiresolution rendering with displacement mapping. In: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware, ACM Press (1999) 55–66
9. Turk, G.: Retiling polygonal surfaces. *Computer Graphics* **26** (1992) 55–64
10. Hoppe, H.: Progressive meshes. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM Press (1996) 99–108
11. Tecchia, F., Chrysanthou, Y.: Real-time rendering of densely populated urban environments. In: Proceedings of the Eurographics Workshop on Rendering Techniques 2000, Springer-Verlag (2000) 83–88
12. Tecchia, F., Loscos, C., Chrysanthou, Y.: Image-based crowd rendering. *IEEE Comput. Graph. Appl.* **22** (2002) 36–43
13. Grossman, J.P., Dally, W.J.: Point sample rendering. In: *Rendering Techniques '98*, Springer (1998) 181–192
14. Rusinkiewicz, S., Levoy, M.: Qsplat: a multiresolution point rendering system for large meshes. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co. (2000) 343–352
15. Pfister, H., Zwicker, M., van Baar, J., Gross, M.: Surfels: surface elements as rendering primitives. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co. (2000) 335–342
16. O’Sullivan, C., Cassell, J., Vilhjálmsón, H., Dobbyn, S., Peters, C., Leeson, W., Giang, T., Dingliana, J.: Crowd and group simulation with levels of detail for geometry, motion and conversational behaviour. In: *Eurographics Irish Chapter Workshop Proceedings 2002*. (2002) 15–20
17. Aubel, A., Boulic, R., Thalmann, D.: Real-time display of virtual humans: Levels of details and impostors. *IEEE Trans. Circuits and Systems for Video Technology* **10** (2000) 207–217
18. Volino, P., Magnenat-Thalmann, N.: *Virtual Clothing*. Springer Verlag (2000)
19. House, D.H., Breen, D.E.: *Cloth modeling and animation*. A. K. Peters, Ltd. (2000)
20. Baraff, D., Witkin, A.: Large steps in cloth simulation. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques, ACM Press (1998) 43–54
21. Rudomin, I., Castillo, J.L.: Distance fields applied to character animation. In: *Proceedings of Eurographics*, Eurographics Association (2002)
22. Wand, M., Straßer, W.: Multi-resolution rendering of complex animated scenes. *Computer Graphics Forum* **21** (2002) Eurographics 2002.
23. Dachsbacher, C., Vogelgsang, C., Stamminger, M.: Sequential point trees. *ACM Trans. Graph.* **22** (2003) 657–662

# Haptic Rendering using Velocity Driven Level of Detail

Pavel Kolčárek and Jiří Sochor

Masaryk university  
HCI Laboratory  
Botanická 68a, Brno, 602 00, Czech republic  
{kolcarek, sochor}@fi.muni.cz

**Abstract.** This paper presents a new concept of using Level Of Detail (LOD) in haptics. LOD and model simplification methods are widely used in computer graphics. In recent years they have started to appear in haptics as well. All these techniques pursue a similar goal as in graphics — they try to simplify the original model as much as possible without incurring the loss of significant details. We take a different approach. A very complex scene can be confusing to the users, especially to visually impaired people, because they cannot adjust their movements according to the image on the screen. We have proposed a method, that will simplify the model in agreement with the speed of the user’s fingertip. When users explore the surface quickly, only a rough shape is presented to them, whereas slow movements result in highly detailed surface. This paper presents the current state of our research in this field.

**Keywords:** level of detail, haptics, velocity driven exploration, multiresolution mesh

## 1 Introduction

Virtual environments are becoming very popular in many areas of our lives. We can come across them in entertainment, education, medicine and virtual design. The important component of these environments is haptic display with its force and tactile feedback, that gives the user a greater sense of immersion in them. Force (or kinesthetic) feedback refers to the sensation from the application of physical force in response to user input. Tactile feedback refers to the sensation caused by stimulating the nerves of the skin such as by pressure and vibration.

The basic building block of the virtual environment is a 3D model. A wide palette of methods for describing a 3D model has been developed. The easiest solution, and probably the most commonly used, is a *polygonal mesh*. For its simplicity it is frequently used in the 3D computer graphics as well as in haptic rendering. Highly detailed meshes (obtained for example using range scanning systems [2]) need to be *simplified* [4, 6] to meet the requirements of a rendering device. To further improve rendering performance, several versions of a model at various levels of detail can be defined. We obtain a *Level-Of-Detail (LOD) approximation*. When the object is close to the viewer, a detailed mesh is used, whereas for the object far away from the viewer, a coarse representation is rendered.



As has been mentioned, LOD and simplification methods play an important role in 3D computer graphics, but in recent years they have entered the field of haptics [1, 3, 11]. We will focus on the main contributions in the next section, but prior to this we will concentrate on differences between using LOD in graphics and haptics and briefly outline our approach. In Section 3 we will discuss it in greater detail. Finally, in Section 4 we will present our results and outline our future research directions.

## 1.1 LOD: Graphics versus Haptics

LOD and multiresolution techniques have been developed to accelerate rendering and handling 3D models in graphics. If we want to use them in a different area, in this case in haptics, we need to modify the criteria of choosing a relevant LOD representation which will meet the different needs of haptic rendering. First of all, when the surface of a model is explored, the user is always in contact with it. Therefore, we cannot use typical selection criteria known from graphics such as the distance of the object from the viewer or size of the model on the computer screen. In our approach we are using the speed of the user's fingertip, and another possible criterion could be the pressure of the fingertip on the surface.

Next, we must take into consideration the sensitivity of the human sense of touch. We cannot use a discrete LOD consisting of several precomputed levels and switch between them, while the program is running, as in graphics. Abrupt surface changes are unnatural and disturbing. Finally, we must take the relation between surface curvature and the radius of the fingertip into consideration so that we don't miss important features of the surface [10].

Recently published methods using LOD in haptics [1, 3, 11] follow *the perceptually driven approach*—the model is simplified so that the haptic device can render it and for rendering the most faithful representation as possible is chosen. This technique is similar to that used in graphics, but the sense of touch, while allowing for active exploration, is spatially focused and has a far lower bandwidth than the visual sense [9]. Thus, if a visual representation is not available or cannot be used (as in the case with blind or visually impaired people), haptic representations of large data sets can be confusing to the user. Additionally, the experiments demonstrated that the greater the curvature of the surface, the lower the speed of the user's fingertip [13] and even a brief contact may suffice to extract a number of features that are diagnostic of the object's identity and also the location of contact [8].

With these results in mind we have proposed a new concept of LOD—*velocity driven LOD*. It follows the hierarchical mode of surface exploration and facilitates the orientation in the scene. The users fly quickly through the scene and get a rough overview of it. When they slow down, they get more and more details about the scene. We will describe our method in more detail in Section 3.

## 2 Related Work

LOD techniques are not an independent field of research, but are linked to other areas. Many approaches to using LOD have been developed over the past two decades. Firstly,

we briefly mention the LOD techniques used in computer graphics that are related to our work. Then we focus on their application in haptics.

## 2.1 LOD and Simplification in Haptics

The simplest solution is a discrete LOD. The levels are created during offline preprocessing, and the appropriate LOD is chosen to represent the object. For a continuous LOD, the simplification algorithm creates a data structure encoding a continuous spectrum of detail. Hoppe [6] proposed such approach in progressive meshes.

The continuous LOD can be extended to a view-dependent LOD. Besides terrain modeling it can be used for local mesh refinement for lighting or visibility, as Xia and Varshney [14] suggested. Another application of the view-dependent LOD is a gaze-controlled LOD. This approach was published by Zha *et al.* [15]. They refine a small area of the model which the user's gaze is focused on, while the rest of the mesh remains in a low detail.

The application of LOD in haptics is supported by the study undertaken by Zhang *et al.* [16]. The results of their experiments suggest that users might not be able to distinguish two haptic objects if they are beyond some refinement level. Okamura and Cutkosky [10] defined surface features based on local curvature. They give examples of how a large fingertip can miss some small surface details.

Several multiresolution methods for use in haptics have been proposed. El-Sana and Varshney [3] have proposed a continuously-adaptive rendering method. They construct a multiresolution hierarchy during the pre-processing stage and at run time they use high-detail representation for the regions around the probe pointer by applying a bell-shaped filter centered at the position of the probe. Asghar and Barner [1] developed a nonlinear multiresolution technique for scientific visualization. This technique preserves edges and sharp features. It is designed for visualizing a regularly sampled dataset, such as terrains.

Kim *et al.* [7] introduced an algorithm using localized contact computations. They decompose the surface of each polyhedron into convex pieces and construct bounding volume hierarchies for fast proximity queries. Multiple contacts are detected between each pair of decomposed convex pieces. Otaduy and Lin [11] develop this algorithm for their sensation preserving simplification algorithm. They construct a multiresolution hierarchy, which is then used to compute a fast contact response of the object-object interaction.

## 3 Velocity Driven LOD

The velocity driven LOD approach simulates the human process of object recognition. At the beginning of this process, a person gets an advance hypothesis about the object from several brief contacts with it, which Klatzky and Lederman [8] call a "haptic glance". Having this hypothesis in mind, the user tries to validate or to disprove it by further exploration of the object.

This process is much more difficult, if we are dependent solely on the kinesthetic and tactile information without visual feedback from the screen. The user can rapidly

become lost in complex or highly detailed scenes. Therefore, when the user quickly scans a virtual object, only a coarse shape is displayed. Slowing down the user’s fingertip involves a refinement of the mesh and hence the presentation of the surface in more detail.

Our method is based on the dynamic gaze-controlled LOD by Zha *et al.* [15], which has been designed for computer graphics. They proposed a method that displays the part of the scene at which the user’s gaze is focused on at a great detail, while the rest of the scene remains in low on detail.

We have achieved a selective mesh simplification which employs the general data structure, the vertex tree, and represents the mesh hierarchies. This structure, called a *merge tree* (MT), was introduced in [14, 15]. We construct the MT using the mesh simplification proposed by Hoppe [6], but unlike the other approaches we use it only for collapsing the edges.

### 3.1 Algorithm Overview

For haptic visualization, we are using the force-feedback device PHANToM and in our current implementation we are using the modified software architecture proposed by Ho *et al.* [5]. It is composed of three independent threads. A visual thread displays a model on the screen and handles user events, a haptic thread communicates with a PHANToM device and computes the final force exerted on the user. There is a communication thread between these threads. It processes the tasks presented in the run-time loop of Algorithm 1, except for the computation of final force. For a smooth run of the application, we are using two sets of data — active and non-active. From the former is computed the final force while the latter is used for a data transfer. When the non-active set is filled with a new data, both sets are switched.

Whole algorithm is divided into two main parts — preprocessing and run-time loop:

---

**Algorithm 1** Velocity driven LOD

---

- 1: {*Preprocessing part*}
  - 2: initialize and fill all data structures
  - 3: construct the merge tree
  - 4: {*Run-time part*}
  - 5: **loop**
  - 6:   estimate the velocity of the user’s fingertip
  - 7:   determine the appropriate LOD that will be used for displaying
  - 8:   detect collision of the probe with a scene
  - 9:     – find the bounding boxes
  - 10:    – select the relevant subset of triangles from a surface mesh
  - 11:   simplify the subset to the requested LOD
  - 12:   compute the force exerted on fingertip from the simplified triangles
  - 13: **end loop**
- 

Before construction of a merge tree, we initialize all data structures of the triangle mesh, add corresponding connectivity and create a simple bounding volume hierarchy

(BVH) for collision detections. For BVH, we use k-DOP bounding volumes, namely 14-DOPs. The construction of the merge tree is described in Section 3.2.

The algorithm proceeds through several steps during the main loop of the application. First of all, the current speed of the user’s fingertip is detected from the PHANTOM driver and a requested LOD is set. Only a simple linear transformation between the minimal and maximal achievable velocity is currently implemented.

After this, the algorithm detects a collision of the probe of the haptic device with a model. The probe is bounded by a sphere of a given radius and if the collision with BVH occurs, impacted triangles are selected for consequent simplification. This is one of the principal ideas of our method — we take only the local neighborhood of an explored surface into consideration. The rest is not important, because the human fingertip perceives the surface in a similar manner.

In the next step, the algorithm simplifies the obtained data. We discuss this procedure in more detail in Section 3.3. Finally, the final force is computed from the simplified set of triangles. For this step we use well known methods such as god-object or proxy position — for further details see the overview presented in [12].

### 3.2 Merge Tree Construction

The mesh simplification is based on two basic operators — *edge collapse* and its opposite, *vertex split*, as shown in Figure 1. With reference to the software framework used in haptics, we use only one operator, a half-edge collapse. During this operation, two vertices  $u$  and  $v$  are merged into one and triangles adjacent to these vertices are modified. Let us define the *affected triangles* of the vertex as the union of triangles that are adjacent to it. The *affected edges* of the vertex form a set of edges leading from the vertex.

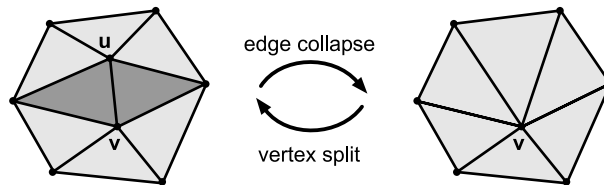
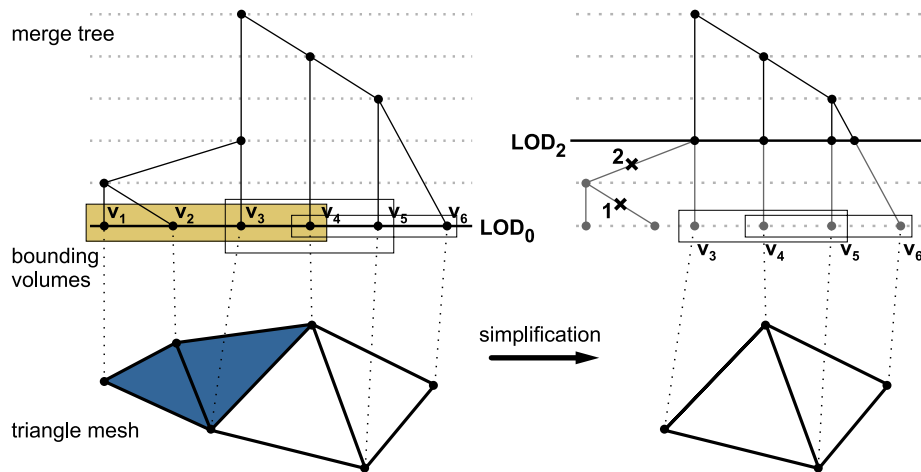


Fig. 1. Edge collapse and vertex split [6]

We construct the merge tree as follows: edge collapses are implemented by using a priority queue. In the beginning, we fill the queue with all edges of the mesh sorted by the length of edge. The shorter edges are placed further ahead in the queue. Then we iteratively remove the shortest edge from the top of the queue and collapse it until we reach the requested LOD. For each edge collapse, we locate the affected edges for both endpoints of the edge, update their length and position in the queue, register LOD number and endpoints to the merge tree and finally remove the collapsed edge from the priority queue and increase the LOD number for the next iteration. Because the two

successive levels of detail differ only in one edge collapse, the change between two LODs is continuous. This is very important for a smooth rendering, since the bigger changes between two LODs that are used in graphics cause an unnatural perception of explored surfaces.



**Fig. 2.** An original mesh and its merge tree (left) and the mesh after two edge collapses (right)

An example of a simple merge tree is shown in Figure 2. There is the original triangle mesh with a corresponding merge tree on the left side. Note that in practice there is no need to simplify a whole mesh into a single vertex, so we usually get a forest of trees. The horizontal lines indicate the LOD number. At the basic level,  $LOD_0$ , there are also simple bounding volumes outlined for this mesh. By applying one iteration of the Algorithm 1, the highlighted bounding volume and its two triangles are selected. The mesh is then simplified using the merge tree (see Algorithm 2 in the next section): after a first edge collapse the vertex  $v_2$  collapses into vertex  $v_1$  and after the second edge collapse the vertex  $v_1$  collapses into  $v_3$ . The resulting  $LOD_2$  mesh along with marked edge collapses is on the right side of the Fig. 2.

### 3.3 The Simplification of a Local Surface to a Requested LOD

When the application detects a collision with the surface, it selects a subset of triangles (see Algorithm 1). This subset is then simplified using the following algorithm.

This algorithm iteratively collapses the edges until it reaches the requested level of detail. With each iteration, it locates the endpoints of collapsed edge  $\langle u, v \rangle$  and removes and updates the affected triangles of the edge. The general case of collapsed edge is shown in Fig. 1. Note the condition of adding triangles on lines 6–8. The simplification process starts with several triangles, but it could happen that the triangle directly

---

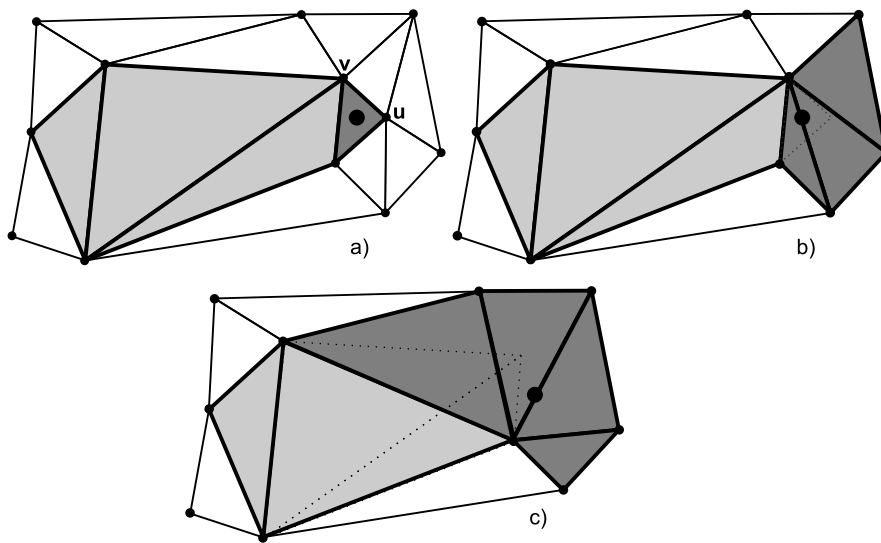
**Algorithm 2** Simplification of subset of triangles according to the merge tree

---

**Input:** subset of triangles of the original mesh, corresponding part of the merge tree

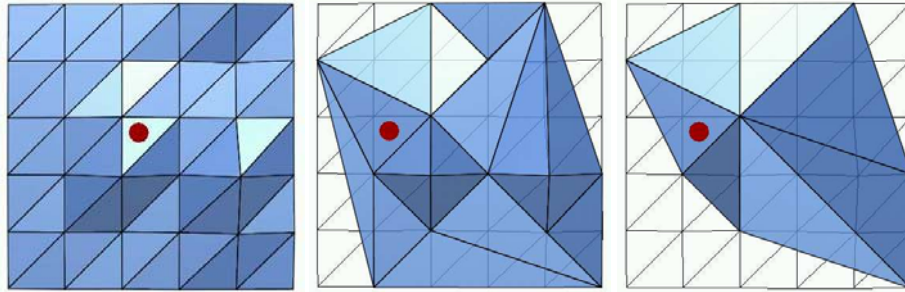
- 1: sort all vertices of the subset by their LOD number, the smallest numbers first
  - 2: **repeat** {*edge collapse*  $\langle u, v \rangle$ }
  - 3: get the vertex  $u$  with smallest LOD from the sorted list
  - 4: from the MT determine the edge to be collapsed –  $\langle u, v \rangle$
  - 5: collect the affected triangles of the vertex  $u$
  - 6: **if** some affected triangles are not in subset **then**
  - 7:     add them to it
  - 8: **end if**
  - 9: delete the edge  $\langle u, v \rangle$  and affected triangles from the subset
  - 10: re-triangulate the resulting hole
  - 11: **until** requested LOD is reached
- 

under the fingertip is removed. This would result in the user's finger falling through the surface.



**Fig. 3.** Progress of the set of triangles when the triangle under the fingertip is removed

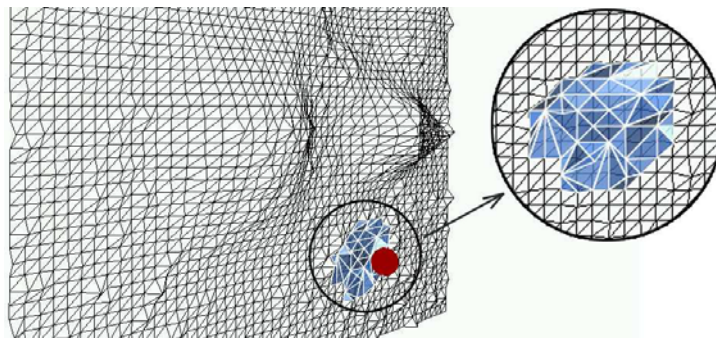
This situation is depicted in Fig. 3. In case a) is the original mesh with a selected subset of triangles. The fingertip position is marked with a dot. The highlighted triangle is removed after the the edge  $\langle u, v \rangle$  collapses, the final force is evaluated as zero and the user falls through the surface. To prevent this, the algorithm checks whether there are some affected triangles, that are not in the selected subset. If so, it adds them to the subset. This situation is shown in Fig. 3 b) and after one more iteration in 3 c).



**Fig. 4.** Original mesh and an example of its two simplifications

## 4 Results and Future Work

Because we are still working on this project, we have no final results or comparative tests with other methods yet. The method is implemented and being tested and the results are shown in Fig. 4. Below we discuss the issues we plan to address.

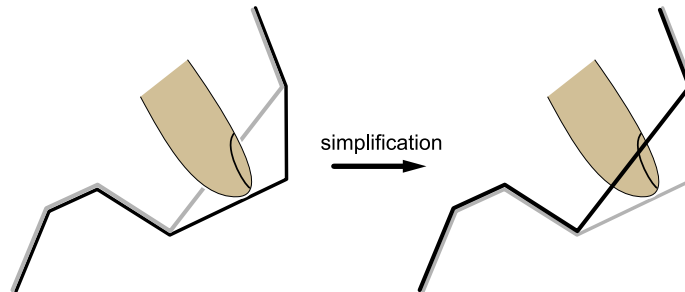


**Fig. 5.** More complex surface with magnified simplification area

While our approach is designed for 3D models having a 2-manifold surface with boundaries, we present our results for illustration on a simple regular mesh only. On the far left of Fig. 4 is the original mesh, and simplified meshes are on the right side. For the sake of clarity, the original mesh is drawn below the simplified set and the small sphere represents the position of the user. Figure 5 represents a more complex surface, with subset of triangles being simplified and magnified on the right.

### Future Work:

- *Interpolation between two successive LODs* — The method we have proposed uses a continuous LOD. Two successive levels differ only in a single edge collapse. Despite this, the change between two LODs is sometimes disturbing and after the LOD



**Fig. 6.** The change of LOD causes that the fingertip finds itself below the current surface

change, the user can find himself above or below the current surface. This situation is shown in Figure 6. To solve this problem, we want to develop an interpolation method between two successive LOD changes when necessary.

- *Determination of requested LOD* — We also plan to elaborate the computation of the requested LOD and replace the currently used linear transformation with a technique that would correspond to individual needs of each user. The possible solution can be some kind of learning algorithm.

**Acknowledgements:** This work has been supported by research grant from the Ministry of Education of the Czech Republic, Contract No. MSM143300003.

## References

1. Mohammad Waqas Asghar and Kenneth E. Barner, *Nonlinear multiresolution techniques with applications to scientific visualisation in a haptic environment*, IEEE Transactions on Visualisation and Computer Graphics **7** (2001), no. 1.
2. Brian Curless and Marc Levoy, *A volumetric method for building complex models from range images*, Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM Press, 1996, pp. 303–312.
3. Jihad El-Sana and Amitabh Varshney, *Continuously-adaptive haptic rendering*, Virtual Environments 2000 (2000), 135–144.
4. Michael Garland and Paul Heckbert, *Surface simplification using Quadric Error Metric*, Computer Graphics (SIGGRAPH '97 Proceedings) (1997), 209–216.
5. Chih-Hao Ho, Cagatay Basdogan, and Mandayam A. Srinivasan, *Efficient point-based rendering techniques for haptic display of virtual objects*, Presence **8** (1999), no. 5, 477–491.
6. Hugues Hoppe, *Progressive meshes*, SIGGRAPH '96 Proc., August 1996, pp. 99–108.
7. Young J. Kim, Miguel A. Otaduy, Ming C. Lin, and Dinesh Manocha, *Six-degree-of-freedom haptic display using localized contact computations*, Tenth Symposium on Haptic Interfaces For Virtual Environment and Teleoperator Systems (2002).
8. Roberta L. Klatzky and Susan J. Lederman, *Identifying objects from a haptic glance*, Perception and Psychophysics **57** (1995), no. 8, 1111–1123.
9. K. J. Kokjer, *The information capacity of the human fingertip*, IEEE Trans. Systems, Man, and Cybernetics **17** (1987), no. 1.



10. Allison M. Okamura and Mark R. Cutkosky, *Haptic exploration of fine surface features*, ICRA 1999 (1999), 2930–2936.
11. Miguel A. Otaduy and Ming C. Lin, *Sensation preserving simplification for haptic rendering*, ACM Trans. Graph. **22** (2003), no. 3, 543–553.
12. Diego C. Ruspini, Krasimir Kolarov, and Oussama Khatib, *The haptic display of complex graphical environments*, Proceedings of the 24th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., 1997, pp. 345–352.
13. Alan Wing, Chris Christou, and Adam Waller, *Touching trajectories: the relation between speed and curvature in exploring shape*, Proc. Eurohaptics 2001, July 2001, pp. 76–79.
14. Julie C. Xia and Amitabh Varshney, *Dynamic view-dependent simplification for polygonal models*, Proceedings of the IEEE Visualisation, 1996.
15. Hongbin Zha, Yoshinobu Makimoto, and Tsutomu Hasegawa, *Dynamic gaze-controlled levels of detail of polygonal objects in 3-D environment modeling*, Second International Conference on 3-D Imaging and Modeling, 10 1999, pp. 321–331.
16. Jian Zhang, Shahram Payandeh, and John Dill, *Levels of detail in reducing cost of haptic rendering: A preliminary user study*, HAPTICS '03 (2003), 205.

# ALICE: A Collaborative Virtual Reality Navigator

Víctor Theoktisto, Marta Fairén, and Isabel Navazo

Departament de Llenguatges i Sistemes Informàtics, UPC, Spain,  
{vtheok,mfairen,isabel}@lsi.upc.es

**Abstract.** This paper presents a Collaborative Virtual Reality Environment, ALICE, which implements remote collaboration by combining sophisticated computer graphics algorithms, advanced user interaction and state-of-the-art shared session management across the network. The architecture implemented to enable collaboration is based on multi-threaded software components under scalable P2P topology. ALICE has been already used on VR applications among diverse disciplines such as industrial engineering, medicine and architecture.

## 1 Introduction

Virtual Reality tools have been applied in many different fields such as engineering, to avoid the use of physical prototypes or interpret real or simulated results; medicine, to help interpretation of scanned data or surgery planning; or architecture, to enable visiting and stress-testing upcoming facilities. In all these environments, users inspect 3D scenes, navigate inside models and manipulate objects and properties.

Distributed environments have been around since the introduction of the first networks. Scope and complexity have kept pace with distributed systems evolution, migrating towards distributed processing, data sharing, multiple execution threads and sophisticated display technology. Computer Support for Cooperative Work (CSCW) [1] is an umbrella term for distributed applications in which multiple users collaborate toward common goals, under a high level event notification and message passing architecture. When combined with several degrees of information sharing, 3D data visualization and real world user-interaction metaphors they become Collaborative Virtual Reality Environments or CVRE's. Remote participants using visual identities (called *avatars*) may navigate inside the virtual space, interact with other remote avatars, and propagate changes to objects in the environment.

ALICE is a VR software platform for the real time inspection and navigation of very complex virtual models. Its collaborative features allow users to share their virtual experiences and participate actively in the environment. These collaborative features include a collaborative user interface model, the possibility of different session roles for the users, the use of avatars and 3D marker highlighting and the exclusion control for the collaborative manipulation of objects in the environment.

In section 2 we briefly present some related work on collaborative systems. In section 3 the ALICE inspector and navigator features, followed by its collaborative capabilities in section 4. Next, section 5 presents the callback implementation and also how these callbacks are used to allow collaborative manipulation of objects (section 6). Finally, we present some results and future work in section 7.

## 2 Related work

The first developed major distributed virtual environments were DIVE [2] and MASSIVE [3]. Different network technologies have been used to enable distribution on a CVRE (BSD-sockets, RPC, Java RMI, DCOM, CORBA, etc.). Systems such as DIVE, NPSNET-V [4], and Distributed Open Inventor [5] apply different solutions for collaboration awareness, using multiple execution threads where each one has an image of the other participants in the interaction.

The most sophisticated approaches delegate clients' network management to components outside of the CVRE ([6–8]).

A more extended sample of existing CVREs were profiled as part of this research in [9]. It also presents a characterization allowing the designer to specify the most suitable feature set for creating a visual sense of presence within a collaboration framework. Features described in the characterization include: *network substrate* to decide which transmission policy is best suited to the expected message flow in the environment; *scalable topology* to choose the scheme for information sharing and communications; *object complexity* to weigh in the network performance cost of broadcasting object changes; *environment persistence* to decide whether users' interactions with the environment have temporal or permanent effects in the CVRE system; and *user interaction* to include the user interface features desirable for a CVRE.

## 3 ALICE as a stand-alone virtual reality navigator

ALICE VR Real Time Inspector and Navigator is a stand-alone VR software platform for visualizing very complex virtual models. It has been used in a number of applications such as navigation in urban environments or interior ship design among others.

In order to allow the users of these applications to be able to navigate and inspect complex 3D models in several VR systems, ALICE offers the following features:

- *Stereoscopic visualization*: being able to work either with active stereo, for systems like Head Mounted Displays [10], or passive stereo, for other low cost VR systems [11, 12].
- *User position and orientation tracking*: allowing user's implicit interaction by following his movements and making him feel he is inspecting a real object instead of a virtual one.

- *Use of varied interaction devices:* being able to follow orders from a mouse, a joystick or a VR glove, for example.
- *Different VR modes of execution:* able to work over different VR display systems like stereoscopic tables, a CAVE, a Head Mounted Display, etc.

Apart from these external features, ALICE implements internally an extensible callback system and many advanced algorithms in computer graphics, in order to be able to work interactively with very complex scenes. It uses internally a hierarchical structure of the objects in the scene, keeping also for each one other non-geometrical information, allowing, for example, the use of textures. Among these advanced algorithms are the following:

**Simplification techniques:** ALICE maintains different levels of detail for the objects in the scene, allowing a faster visualization of complex models by choosing the right level of detail depending on the distance between the object and the observer [13] (further objects can be visualized with less detail without losing image quality).

**Visibility culling:** This technique eliminates from the visualization process those parts of the geometric model that will not be visible from the current observer (camera) position [14]. This technique is also combined in ALICE with the multiple levels of detail [15].

**Collision detection:** The collision detection is a key component for any VR system, because it is the base for the selection and manipulation of objects, robotics, vehicles simulators, etc. Collision detection allows, for example, to select objects with a VR glove, which is used to select them by virtually “touching” the objects [16].

## 4 Adding Collaborative VR capabilities to ALICE

Most science disciplines (and the entertainment industry) use VR techniques to enhance user experiences. As research shows, users *always* desire to share these virtual experiences. The object-oriented nature of current standalone VR applications, usually having rendering and user-interface components, facilitate their transformation into complete CVRE’s, by allowing the seamless attachment of a network-based component to enable collaboration.

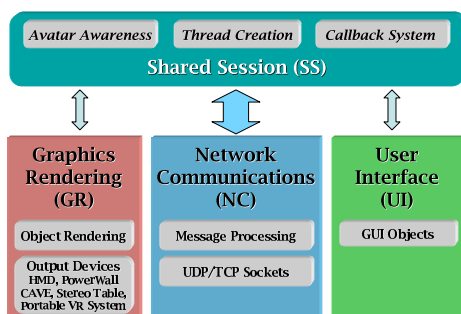
The ALICE collaborative features have been designed to be as much independent as possible of ALICE itself, so they can be applied easily to other VR applications. These features involve the implementation of a *multithreaded software components* architecture, a scalable *P2P sharing topology*, a flexible *crossplatform library for network transmission*, and a layer implementing *session awareness capabilities*.

### 4.1 Multithreaded Software Components

A standard software engineering practice in Computer Graphics is the factoring of application objects into at least two weakly cohesive software functional

components, graphics rendering and user interface. We decouple the Graphics Rendering (GR) and User Interface (UI) parts and instantiate them in separate threads. The same approach is taken with the new network communications component (NC), launched in its own separate concurrent thread. In this way, advantage is taken of the underlying operating system’s context switching, loading the new software components without altering functioning code. This extensible approach allows the addition of more component threads, such as one dedicated to tracker data acquisition or interaction with haptic devices.

A snapshot of a working framework model is shown in Figure 1, detailing each software component. The NC component thread handles communications and message parsing; the top Shared Session (SS) management layer launches all concurrent threads, tracks users’ avatars, propagate state changes to the UI and GR components using callbacks, and is in general responsible for the emerging collaborative behavior.



**Fig. 1.** Collaboration-enabling Threaded Processes. The framework includes the VR application components (GR, UI), a session layer (SS) and a network component (NC)

## 4.2 P2P Sharing Topology

Fitting any of the client/server topologies would have implied the creation of at least one central server from scratch and compromised the applications’ standalone behavior. We chose instead a peer-to-peer scalable topology, the most adequate for equal participants with separate access to their models. There are two possible topologies available in the framework: P2Pr [*Peer-to-Peer with scene replication*] and P2Ps [*Peer-to-Peer with scene sharing*]. ALICE uses the P2Pr topology.

Having no central server, both approaches require a third party application providing locating services for clients willing to enter in a session.

### Thin broker for session administration

In our proposal, this third party is called a *message broker*, tracking session participation and interaction. Shared state information is kept through the following services:

- A name service for location and client registration.

- A session management service.
- A session and client state report and mirroring service.

Given that the broker is not a bridge, client messages go directly to their destiny. Each client keeps track of other participants, and periodically may send its current state to the broker for shared session recording purposes.

### 4.3 Crossplatform Network Transmission

Since communication is what enables collaboration, the corresponding software component handles network communication capabilities. This is done by a cross-platform networking class that allows either datagram-oriented (UDP) or connection-oriented (TCP) communications under IPv4 and IPv6 networks. The NC thread, under a common message protocol, implements the following basic kind of services, each one running on its own separate listening socket:

- Shared event pipeline for sending environment state changes and callback messages
- Continuous streaming of some client properties, such as camera position and orientation
- A notifying service for the *Broker*.

When a client reports to the broker, it posts its network address and listening ports. The message protocol is short and simple. There are three kinds of messages: *session*, *location* and *manipulation*. Session messages are the ones exchanged between the broker and the clients: connecting and disconnecting, reporting internet addresses and ports, number of active cameras, avatar appearance, global scene file, and other relevant data. Location messages are mostly for *avatar* properties being broadcasted among all participants (see next subsection). Manipulation messages (such as a local client touching, grabbing or modifying an object) are sent to remote users by the callback system to maintain scene coherence among all participants (the callback mechanism is explained in section 5).

Since all of this happens in the NC thread, mutexes are activated when this thread is modifying data such as clients' cameras. There is a corresponding set of mutexes placed just before rendering to avoid race conditions so common to concurrent programming.

### 4.4 Session Awareness Management

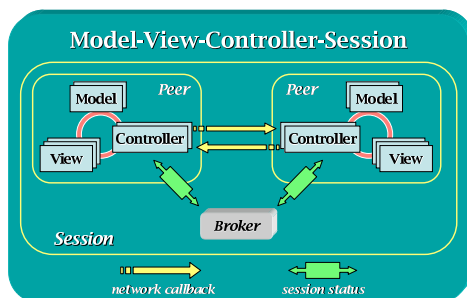
The ALICE collaboration feature set includes the following:

- Collaborative user interface model.
- Session administration with differentiated user roles.
- Client awareness using *avatars*.
- Shared annotation and 3D marker highlighting.

For a client in this scenario, there must be perceptual evidence that other entities (human or otherwise) are participating, so 3D client embodiments within the environment (avatars) are used to dynamically reflect their position and state in the scene. Clients may want to call others to attention by placing special 3D signals, leaving trails in the scene or modifying the environment.

### Collaborative user interface model

The problem to solve when recasting existing VR navigators as CVREs is how to implement the maximal collaborative feature set with the least possible implementation cost, and without affecting the original standalone behavior. We choose a hybrid Model-View-Controller-Session (MVCS) approach, where controllers operate using a callback mechanism, and the session layer coherence is maintained by existing network-aware controllers at each node, who also notify the broker (see figure 2).



**Fig. 2.** Model-View-Controller-Session (MVCS) Objects showing an external broker maintaining session states

### Session administration with differentiated session roles

We have identified five different collaborative user behaviors: *standalone*, *peer*, *incognito*, *slave*, and *master*. A *standalone* client is not aware of other clients. It defaults to the original isolated behavior of the application. *Peers* are clients that communicate among themselves using the common message protocol. Users traveling *incognito* may observe scene interaction in “voyeur” mode without other clients knowing it. A *slave* is a peer that is bound to another, correspondingly called a *master*, in the sense that the *master*’s current state is continuously replicated by the remote *slave(s)*. The *slave* shuts down its own cameras and reflects the *master*’s camera viewpoint and actions. This feature may also be used to “teleport” a participant to the position of another, which is very useful to avoid losing virtual eye-contact among peers. These client roles are voluntary and changeable during a session.

### Client awareness using avatars

Each client has its own 3D representation traversing the environment, having several active camera perspectives at any time. Avatars broadcast a number of

state attributes, such as *position*, *orientation* and *velocity* camera vectors for dead reckoning calculations.

### Shared annotation and 3D marker highlighting

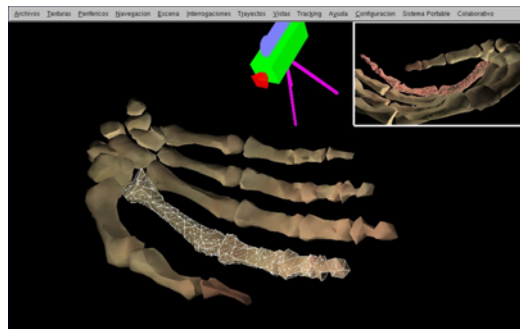
Users must not only be aware of each other, they must be able to call the attention of remote participants to some feature or object in the environment. This is accomplished by temporal 3D markers such as arrows, billboards or banners, objects that a “guide” pins at some interesting locations.

The broker must be active for a session to be initiated by at least two subscribing participants. Each client may choose a session role (usually as a *peer*) and an avatar representation (from a menu), while keeping a list of the current active interactions with other users. As they navigate, clients can chat to each other, or place 3D markers on the scene to call attention to some feature.

## 4.5 Remote user window

Another feature added to the collaborative capabilities of ALICE allows the user to open an extra small window where the rendering of a remote user can be visualized, while its own rendering is still available on the main window of the application. In figure 3 the main window is seeing the avatar of the remote user whose rendering is being visualized in the small window.

Since the purpose of this feature is to know how the remote user is seeing the scene at any time, some information about the local rendering of that user is also required to pass through the network (apart from the information of position and orientation of his camera). This information includes, for example, the position of far and near clipping planes.



**Fig. 3.** A main window with a remote client visualization on the upper-right corner.

## 5 The callback mechanism

When a user interacts with ALICE, he can cause some changes in the scene or in his representation in the collaborative session (changing the avatar or becoming



stand-alone, for example). These changes should be broadcasted to the rest of the participants on the session, so they can be aware of what the changes were.

An ALICE callback is an entity hooked to a certain task that has to be done in the application. Callbacks that imply a network communication, causing the replication of its task in all other participants in the collaborative session, require the implementation of methods for serializing, sending, receiving and deserializing the information associated with it. Since each callback contains different number of parameters to be sent over the network, its serialization and deserialization is performed by the callback itself. This also has an advantage since all the parameters can be stored and sent as one stream of data, without sending additional overhead information of what is actually sent (metadata). The only additional information that has to be sent with the callback network message is the identity of the callback which should deserialize the message to obtain the correct data.

Examples of callbacks included in the collaborative ALICE application are the following:

- *change avatar representation*, when the user selects a different representation for the avatar representing him;
- *add a new light* in the scene;
- *select/deselect an object*;
- *change some property* of a previously selected object;
- *eliminate an object* that had been selected;
- *move an object* that had been selected;

The last four callbacks listed are directly related with the collaborative manipulation of objects explained in the next section.

## 6 Collaborative manipulation of objects

ALICE, as a navigator and inspector, has limited modeling options in the scene, but it allows the user to move and delete objects and also to change several object attributes like color, for example.

Since the ALICE collaborative architecture has a component, the broker, that controls all the involved in the collaborative session, we use it to implement an object blocking mechanism for the selection of objects. When a user wants to manipulate an object he has to select it first by asking the broker whether the object is already selected by another user or not. The requests for selecting an object arriving to the broker are served sequentially, so the first request arriving is the one who achieves the selection. When the broker accepts a selection coming from a user, this selection is broadcasted (by using callbacks) to the rest of the users in the collaborative session.

In order to allow the user to be aware of the objects selected by remote users, these objects are labeled as remotely selected in the local copy of the scene, and the visualization routine draws a red wire frame around the object polygons. This information of remote selections is also kept in the list of remote users in



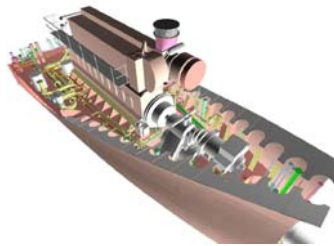
**Fig. 4.** Different VR systems where ALICE works and has been tested

order to be able to recognize the user who selected a certain object, which is useful when a second rendering of the user who selected the object has to be done in a remote user window (for example, in figure 3 the finger is selected by the local user and the small window is visualizing it as remotely selected).

## 7 Results and future work

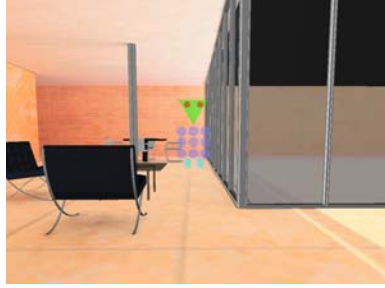
ALICE has been successfully used as a collaborative platform over different VR systems, such as HMDs, Stereoscopic tables [11] or the portable system [12] (see figure 4). Some of the applications are the following:

- *Inspection and navigation in the interior of a ship.* Ship designers cannot use real prototypes to evaluate their designs, so they need virtual prototypes to be evaluated (see figure 5). Different users can discuss about the resulting design without needing to be at the same place at the same time.

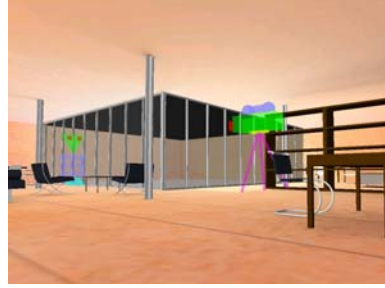


**Fig. 5.** Ship virtual prototype

- *Inspection and modification of architectural design.* The virtual design made by the architects can be shown to potential clients in order to know their opinion (see figure 6). A client, by collaborating with the architect, can also make small changes such as moving furniture or walls in order to propose other possibilities not far away of the initial design.



(a) "camera" avatar's viewscene



(b) An *incognito* viewscene, showing two avatars interacting

**Fig. 6.** Two users observe each other while navigating the architectural environment

- *Training in medicine.* A surgeon can show the students, for example, how the scalpel should be used for a certain surgery incision.

We have also tested ALICE's remote collaboration and navigation services in sessions with the Girona University (located 100 Km. from Barcelona) through a 10Mb wide area network connection. Network traffic is generated only when an avatar changes position or orientation, or when a callback is produced (for example when selecting an object), so no unnecessary traffic is produced.

The results obtained from our tests can be seen in the following table. The scene used on these tests (the interior of a ship) contains 20.000 polygons, but on purpose does not have complex textures because it could hinder or slow graphics performance. The table shows the results obtained in the communication of 2, 4 and 8 workstations from both labs (Girona's and our's).

Participants	1	2	4	8
Av. Number of messages	–	2539	8067	14331
Av. Total Net. Time (msec)	–	35	31	46
Av. Roundtrip Time (msec)	–	13	46	57
Framerate	47.3	45.2	44.2	42.7

In the table we observe the average total number of messages that were sent through the network in a 4 minutes navigation test. The total network time (in milliseconds) gives information about how much time Alice spent in the transmission of messages during these 4 minutes trials (this means that around 0.1-0.2% of total time is spent in network communications). The roundtrip time is also indicated in milliseconds. Since our framework uses unicast addresses, roundtrip time increases as more peers participate in the session. Finally, the table shows the average rendering framerate achieved for each case, which indicates that increasing the number of nodes affects graphics performance very slightly compared to the stand-alone performance.

As for future work, we are working on extending the collaborative breath of ALICE by including in the session layer a fourth thread for handling haptic devices, adding high frequency force-feedback events to the interactive session repertoire.

## References

1. C., S., Schummer, J., Seitz, P.: Modeling collaboration using shared objects. In: Proceedings of the international ACM SIGGROUP conference on Supporting group work, Phoenix, Arizona, United States (1999)
2. Carlsson, C., Hagsand, O.: Dive: A platform for multi-user virtual environments. *Computers and Graphics* **17** (1993) 663–669 <http://www.sics.se/dce/dive.html>.
3. Greenhalgh, C., Benford, S.: Massive, a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer, Human Interaction*. **2** (1994) 239–261
4. Capps, M., McGregor, D., Brutzman, D., Zyda, M.: Npsnet-v: A new beginning for dynamically extensible virtual environments. *IEEE Computer Graphics and Applications* **20** (2000) 12–15
5. Hesina, G., Schmalstieg, D., Fuhrmann, A., Purgathofer, W.: Distributed open inventor: A practical approach to distributed 3d graphics. In: Proceedings of ACM VRST'99. (1999) 74–80
6. Harting, P., Just, C., Cruz-Neira, C.: Distributed virtual reality using octopus. In: Proceedings of IEEE Virtual Reality 2001. (2001) 53–62
7. Leigh, J., Johnson, A.E., DeFanti, T.: Cavern: A distributed architecture for supporting scalable persistence and interoperability in collaborative virtual environments. *Journal of Virtual Reality Research, Development and Applications* **2** (1996) 217–237
8. Deriggi, F., Kubo, M., Sementille, A., Ferreira, J., dos Santos, S., Kirner., C.: Corba platform as support for distributed virtual environments. In: Proceedings of IEEE, Virtual Reality'99, Houston, Texas (1999)
9. Theoktisto, V., Fairén, M., Navazo, I.: Enabling collaboration in virtual reality navigators. Technical Report LSI-04-13-R, Universitat Politècnica de Catalunya (2004)
10. Sutherland, I.: A head-mounted three-dimensional display. In: AFIPS Conference Proceedings. Volume 33. (1968) 757–764
11. Andújar, C., Fairén, M., Brunet, P.: Affordable projection system for 3d interaction. In: 1st Ibero-American Symposium in Computer Graphics, University of Minho, Portugal (2002)
12. Fairén, M., Brunet, P., Techmann, T.: Minivr: A portable virtual reality system. *Computers & Graphics (to appear)* **28** (2004)
13. Andújar, C., Brunet, P., Ayala, D.: Topology-reducing surface simplification using a discrete solid representation. *ACM Transactions on Graphics* **20** (2002)
14. Andújar, C., Saona-Vázquez, C., Navazo, I.: LOD visibility culling and occluder synthesis. *Computer Aided Design* **32** (2000)
15. Andújar, C., Saona-Vázquez, C., Navazo, I., Brunet, P.: Integrating occlusion culling and levels of detail through hardly-visible sets. *Computer Graphics Forum* **19** (2000)
16. Franquesa-Niubo, M., Brunet, P.: Collision detection using multiresolution kd-trees. Technical Report LSI-01-36-R, UPC (2001)

# Designing head tracking devices for your VR needs

Isaac Rudomin<sup>1</sup>, Marissa Diaz<sup>1</sup>

<sup>1</sup> ITESM-CEM Instituto Tecnológico y de Estudios Superiores de Monterrey,  
Computer Science Department,  
Carretera Lago de Guadalupe Km. 3.5,  
Margarita Maza de Juarez,  
Atizapan de Zaragoza,  
México, Estado de México,  
rudomi@itesm.mx, dmarissa@itesm.mx

**Abstract.** This paper describes a head-tracking solution that we developed using custom-designed hardware and that is suitable even on a small budget. We examine the problems that this kind of design implies, how we solved them, and give alternative solutions that may fit in different kind of applications.

## 1 Introduction

In large-scale Virtual Reality rooms the cost of standard head tracking devices and motion sensors can be insignificant, but even in this case when the system requires specific information to attain interaction and the cost of changing the manufactured sensors is not included in the original funds, picking suitable technology so that one can custom design and build a sensor for the application is of great importance.

To develop useful Virtual Reality equipment it is necessary to realize that the main problem with representing objects in a virtual world is the perception that a user has of the virtual objects. Hinckley<sup>[13]</sup> supports the idea that most people don't have a notion of how a 3D virtual world should be, but have a good idea of the experience of the real world. When one tries to implement an interactive and immersive application one should keep this in mind.

In order to accomplish kinesthetic feedback and return references that are natural to the user, there should be a relationship between head movement and the virtual environment. The user normally expects that some actions in the real world to cause changes in the virtual world and therefore the need for interactive devices is revealed. For comfort, equipment attached to the user should be lightweight, small and wireless. To improve the 3D virtual world it is also necessary to reduce computational complexity and costs in design and hardware. The solution we explore follows these guidelines.

## 1.2 Head Tracking Techniques

To obtain the head's position in space, we can decompose the problem as follows: first we must determine the angle with respect to the axis of symmetry of the body, in other words, the movement caused by panning the same plane of view (roll); in second place we must determine the inclination of the head with respect to the horizontal axis (pitch); and lastly the rotation with respect to the z-axis (yaw).

Something to remember and that can make our life a little easier is to take into account that there are some unreachable positions, in other words the user won't twist his or her head in awkward ways and that those configurations can be excluded from the design.

In the following we briefly review the methods most commonly used to obtain this head position information<sup>[3]</sup>:

- Gyroscopic systems. - Gyroscopes allow the reading of roll, pitch and yaw but they have a drift caused by the acceleration.
- Cord Systems. – These rather old-fashioned methods consist in literally tying the user by attaching cords to his/her head and determine the movement by reading the tension in the rope system.
- Arm System. - Using a mechanical arm that transmits its rotation while moving potentiometers that are attached at the base. This one is relatively uncomfortable and can't give data about the pitch and yaw.
- Ultrasound. - Triangulation of the position based in receptors and transmitters of ultrasound (getting the bouncing of the ultrasonic signal to triangulate, typically there are three receptors).
- Cameras. - By using the data obtained from images to determine the position of the head, by using vision-based methods that have a high computational cost. Some methods use markings attached to the user instead of edge detection, since it is less complex computationally talking but less elegant.
- Magnetic Field Systems. – In the past these were developed with coils and other magnet systems, but at the present time it is possible to use ultra sensitive Hall effect sensors and giant magneto resistance devices.

We decided to work with magnetic sensors due to their flexibility and because they allow comfortable sensing. The following section will explain with some detail how this kind of system works and how we can implement economic and reliable applications that use them.

## 2. System design

The first option we reviewed was the Dinsmore 1490 digital compass. This device, as the manufacturer describes it, is a compass that gives a discontinuous measurement of the position using several 1525 Hall effect switches.

The 1490 has a magnet of good intensity that freely rotates while it is orienting with respect to the earth's magnetic field. It has four switches in positions scattered at  $90^\circ$  that gives information of the angle with a  $45^\circ$  resolution when they are activated individually or in pairs by the magnetic field, returning a signal in an open collector transistor output pin.

The 1490 can be used in applications that require an easy and fast development solution that gives accurate information with low resolution; if the application requires a little more resolution one can interpolate with software the angle between the switches just by taking on count the last measurement and the time involved in the motion. This is not as accurate as one may wish, but it works.

The second option we explored us the use of giant magneto resistor (GMR) devices that allow continuous measurement of the angle with respect to the North-South line and gives a resolution of  $2^\circ$ . There are IC's available that ensure precise angle readings for both  $180^\circ$  and  $360^\circ$ .

The advantages of GMR technology are obvious, the only thing one has to keep in mind while using these devices is that the voltage signal in the output is hard to transmit wirelessly without preprocessing to obtain the angle, but in order to do that it is one must apply a complex mathematical equation or map the value in a table which is hard to accomplish if only a small space can be used for circuits.

GMR devices can be used in applications that can afford to expend space in the transmitter and that need high resolution; it can be put in a surface mount package so one need not waste too much space in the design.

The third option are the 5617 and 5618 Hall effect sensors from Allegro, which are analog devices which allow a precise reading of the angle using as a reference the magnetic fields: in the case of the 5618 a strong reference is needed (something like two very big magnets or coils), but with the 5617 the measurement can be performed by referencing directly to the earth's magnetic field.

The Hall sensors from Allegro are fast, accurate and small, and besides they return an output that is suitable to be converted into a pulse with a width or frequency proportional to the angle. The main problem is that it is required to use two of these sensors to attain a  $360^\circ$  range reading and with this addition it is harder to transmit wirelessly the information. However, even with this problem it is easier than using the other technologies reviewed before.

One concept that must be clear is how the Hall effect works, this effect, by definition, states the relationship between the current in a conductor and the magnetic field, the way this current is affected causes a potential difference to occur in a plane perpendicular to the current that is proportional to the vector product of the current and the magnetic field so that if the direction and magnitude of the current are known the magnitude and direction of the magnetic field can be easily calculated since the variation of the output voltage can be represented as a vector containing the magnitude and direction of the magnetic field.

Giant Magneto Resistance devices are similar to the Hall effect in essence but in them the current influenced by the magnetic fields is crossing a bridge of resistances from which voltage results in a sinusoidal function that is related directly by mathematical equations to the angle of rotation and direction of the magnetic field.

The big asset of GMR is that the information it returns is related to the direction of the magnetic field and not to its intensity. The main problem is sensitivity and in some cases it needs a relay device like a compass or a free moving magnet to have a magnetic field reference to take the measurement.

An important point is that in our implementation we used a radio transmitter manufactured by Abacomm Technologies Company that has a patented laser calibration technology for the frequency that makes the transmitter small and perfect for this kind of application. If one is an expert in radio communication or wants to use another wireless option one must remember that keeping the design small is very important and that infrared requires that the user doesn't block the emitter and the receivers must be properly placed.

There is a certain fear about the usage of magnetic sensors because of the noise that can be added to the system if the magnetic conditions change but we developed a prototype, which can measure referenced to a direction in a specific room so that the variations in the magnetic field are eliminated.

## **2.1 Switch sensing**

Once we have the orientation of the user in the VR space we would like to know if the user moves beyond certain points of interest. For this purpose we can design a grid of sensors that gives us information about the position of the user on this grid. We would like to use this kind of device as input for our applications

Some times we want to get binary information of the actions that the user does in order to change the 3D virtual world. Something that is easily built is a light-sensing device but it implies that one must have a well illuminated room and can give wrong information if it is not well adjusted. Another approach is to place a beam across the



room so when the receiver doesn't have a response one can activate the switch but the main problem is the distance, power and special placement of these sensing devices.

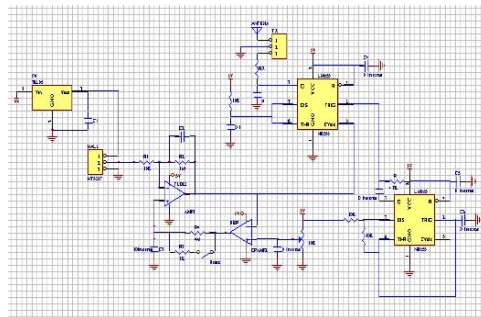
In our design we used a light sensor that initially returns a continuous signal, which is used to determine the binary value but that gives us an opportunity to avoid mistakes due to varying room lightning. In Figure 1 we see this device.



**Fig. 1.** Binary output sensing device for VR interaction and position detection.

### 3 Conclusions

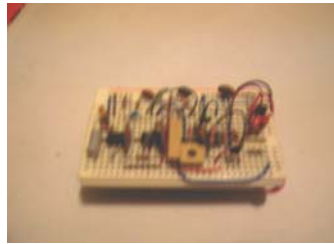
Our application needed a Head tracker with good accuracy in a 90° range; it has to be small and light-weight as well as economical. The Electronic diagram shown (figure 2) gives an idea of the design process we followed in order to obtain the unit described in figures 3 and 4.



**Fig. 2.** Electronic diagram for the head tracking device.

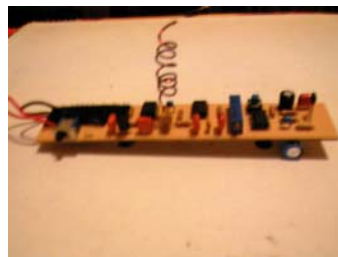
The important parts in this design are: The amplification phase on which we took the signal from the Hall sensor and used it as the input for an operational amplifier in an inversion configuration, the capacitor in this section is small to filter the noise, the

other operational amplifier is used to set the initial voltage and reset the system to the magnetic conditions in the room; The signal treatment phase is a 555 that measures the time and sends a pulse to start the pulse generated in the second 555 (configured as a mono-stable) and that returns a pulse which is proportional to the voltage in the input; And the transmission phase that is basically the antenna and the Abacomm transmitter.



**Fig. 3.** First Prototype.

For the receiver we used a basic stamp II micro controller to receive and send (to the PC) the information of the sensor, to program it the main problem was to adjust the reception frequency to the 1ms time of the transmitted information signal.



**Fig. 4.** Final stages of prototype, this unit can be placed atop VR lenses to transmit the head orientation data.

In the PC the important part was to receive and retransmit the information to the equipment that will render of the application. The option of high sensibility Hall sensors was the one that responded best in our case but for other applications one must evaluate the other options for sensing and transmitting data.

## 4 Results

The head tracker we built has decent accuracy in a 90° range; is small even as a prototype and is lightweight as well as economical. We hope that this design to be useful

to people who want to develop virtual reality applications that are reliable and economic. Compared to the normal head trackers on the market this device is substantially cheaper (\$55USD), for example Intertrax2 that is a popular one costs \$995USD. Our design beats even the affordable ones like the virtual realities "virtua track" that costs \$499 USD, it gives 3 degrees of freedom but it is based also in magnetic sensors and is almost 10 times the price. Some systems are quite complex and deliver more information than the needed like the 3<sup>rd</sup> tech HiBall-3100 that surely gives accurate position of the head and the body in a confined area but are expensive and sophisticated.

The accuracy of our system is good, it gives an error of  $\pm 2^\circ$  on each degree of freedom implemented, it can be upgraded from one degree of freedom to 3 degrees of freedom by adding extra sensors with a total cost of 10USD for each for a total of \$75USD.

## References

1. Baratoff, G., and Blanksteen, S. Tracking Devices. Encyclopedia of Virtual Environments. World Wide Web URL: <http://www.hitl.washington.edu/scivw/EVE/1.D.1.b.TrackingDevices.html>. 1993
2. BURDEA, GRIGORE; PHILIPPE COIFFET. Virtual Reality Technology. Printed in 1994 by John Wiley & Sons USA.
3. CARSTENS, JAMES. Electrical Sensors and Transducers. Prentice Hall USA 1993.
4. DORF RICHARD. Electric Circuits. Fifth edition . Printed by John Wiley & Sons at USA in 2001.
5. GRADECKI, JOE. The virtual Reality programmer's kit. Printed by John Wiley & Sons USA. 1994.
6. HERKE JAN NOORDMANS; HANS T.M. VAN DER VOORT; ARNOLD W.M. SMEULDERS. Spectral Volume Rendering. IEEE transactions on visualization computer graphics, vol. 6, no. 3, July-September 2000. Number 1077-2626/00.
7. JONES KAREN C.; MARC W. CYGNUS. Virtual reality for manufacturing simulation. Proceedings winter simulation conference 1993.
8. LANE, C.. Display Technologies. Encyclopedia of Virtual Environments. World Wide Web URL: <http://www.hitl.washington.edu/scivw/EVE/1.A.1.Displays.html> 1993
9. MINER, GAYLE. Lines and Electromagnetic Fields for Engineers. Printed by Oxford University Press at USA in 1996.
10. NATONEK E.; ZIMMERMAN L.; FLÜCKIGER L. Model Based Vision as feedback for Virtual Reality Robotics Environments. IEEE : 0-8186-7084-3/95
11. NEWMAN, WILLIAM; MICHAEL LAMMING. Interactive system design.. Printed at 1995 by Addison

Wesley in England

12.RANDY PAUSCH; JOHN C. KEN HINCKLEY , "A Survey of Design Issues in Spatial Input". University of Virginia. Departments of Neurosurgery and Computer Science . UIST 94 Noviembre 2-4. páginas 213-222  
Número de la ACM:1994 ACM 0-89791 -657-3/94/0011.

13.WEBB, JOHN, RONALD REIS. Programmable Logic Controllers. 4 edition by Prentice Hall. in 1999.

# Advanced Arthroscopy Training Simulator

## *insight*<sub>MIST</sub>

Jose M. Espadero, Sofía Bayona, Jose M. Fernández, Marcos García

### Abstract

Minimally Invasive Surgery (MIS) is the latest trend in surgery and it is performed with the use of small incisions, customized instruments, innovative imaging techniques and global navigation. The advantages are clear: less pain, trauma and blood loss, and minimal scarring. The result is a rapid recovery, high patient satisfaction and the best possible outcome. The main problem is that its performance requires the surgeon high skill and ability. Traditional learning methods are expensive and involve many problems.

This paper presents an adaptive Arthroscopy Training System that provides a realistic environment in which practitioners can develop hand-eye coordination dexterities, one of the key features when starting learning MIS techniques. Some of its characteristics will be: specifically arthroscopy adapted instrumental, dexterity evaluation system and library of anatomy models.

### KEY WORDS

Virtual Reality, Training Systems, Arthroscopy, Visualization and HCI, Minimal Invasive Surgery

## 1 Introduction

MIS refers to a growing number of surgical procedures in which the surgeon operates through small incisions, with customized specific instruments, using innovative imaging techniques and global navigation. The visual control of the organs and the surgical instruments is done through a camera with a cold light attached. A lot of benefits derive from MIS: the patient trauma is considerably decreased, allowing a much faster recovery. This reduces the necessity of having to intern the patient, resulting in large cost reductions. On the other hand, techniques for MIS are complicated, difficult to master and not at all intuitive, since the surgeon must perform the operation through indirect visual control, that is, by watching a screen instead of directly observing the patient body. This is an unusual skill, difficult to grasp. In addition, surgeons are nowadays performing MIS on more and more parts of the body, which implies a strong need for training procedures.

Traditional learning methods are expensive, involve many problems and are not as realistic as it would be desirable. Aspects like the interface that the surgeon will handle, faithful and realistic feedback, learning method, evaluation and cost must be considered when designing simulators. Since it is difficult to practice this type of procedures in corpses or animals, frequently the learning process is made from first contacts in high cost specialized courses and directly in the operating room under the supervision of an expert surgeon, with the consequent limitations and risks.

Whenever an operator has to be trained to perform tasks involving risk for human beings or critical economic costs, simulators are the best choice. Very common examples can be found in industry, transportation and military environments, with flight simulators as probably the most widely publicized case.

In medicine methods to acquire practical proficiency and reduce risks are also needed, because surgeons deal with human lives. Medical staff must receive a very special training to perform correctly their labour. For example, surgeons, after a period of theoretical education, to acquire practical knowledge observe a number of operations performed in real patients and they gradually incorporate, increasing the risks, to auxiliary tasks executed under strict expert supervision. Those practical sessions constitute a bottleneck in the learning process, given the restrictions in the number of trainees admitted to each operation and patients are frequently not pleased with being examined by non-experts. Other methods, such as using animals present legal and ethical problems. Phantoms or plastic models, are expensive to build, and have a limited lifespan because once an incision has been made during a training session, it cannot be removed.

All these factors lead to the need for artificial Surgery Training Systems which may be able to provide a non-degradable realistic environment, in which novices may learn and try as much as desired, with no additional cost [1] after the installation of the training system.

## 2 State of the art

Simulators have some type of physical representation with sensing instruments that inform the computer of the user's movements. Systems may include some sort of three dimensional (3D) imaging to make the environment more realistic and intuitive. The program computes user's movements and projects the correct response onto the screen[2]. If the model has haptic (touch) feedback, it will also create the illusion that the user is coming into physical contact with the model, and the user will *feel* the patient as well as see it[3][4][5].

This level of technology has come onto the market recently, so many products are still in the developmental stage.

However there are already tools for helping in teaching anatomy and training surgeons. For example, the University of California is building a networked educational system which links local knowledge from textbooks and anatomical atlases with electronic medical records and several on-line simulators [6].

There are also some projects on MIS procedures like ProMIS[7], SimSurgery[8], LapSim[9], LS500 Laparoscopy Simulation Platform[10],the AccuTouch Endoscopy

Simulator[11] for flexible bronchoscopy, upper and lower gastrointestinal flexible endoscopy or the AccuTouch Endovascular Simulator[12] for endovascular procedures such as PTCA, Stenting, and Cardiac Pacing. Endotower[13] and GI Mentor[14] are useful for endoscopic navigation.

Other endoscopic and laparoscopic training systems are: the Key Surgical Activities (KSA)[15], the Vascular Intervention System Trainer (VIST)[16], the Procedicus Virtual Arthroscopy (PVA)[17], the Lap Mentor[18] or the Minimally Invasive Surgical Trainer MIST[19].

It is usual when building training systems, to focus just on a procedure, because it simplifies the system requirements.

Regarding arthroscopy [20], there has been development specially in the knee pathologies[21, 22, 23].

### **3 Methodology**

In order to achieve an Advanced Arthroscopy Training Simulator an incremental development has been chosen. Among the different MIS modalities, arthroscopy was selected attending to several considerations, such as the distinctive problems associated to this kind of surgery, where the surgeon has reduced manoeuvrability; the potentially large demand for training, and the varied repertoire of interventions that could be covered by this kind of systems.

#### **3.1 Objectives**

Our multidisciplinary work group develops a simulator for the learning and improvement of arthroscopy surgery techniques. The following objectives were focused on during the whole stage of the system design:

- To build a multi-purposed adaptable training simulator
- To apply techniques of Virtual Reality to generate a simulation of great fidelity and realism
- To cover exhaustive and progressively all the different requirements of the apprentice learning process
- To provide the instructor with a didactic tool to facilitate procedure teaching and case explanations
- To offer a way of measuring dexterity and performance
- To make the practitioners used to work into stressful environments

## 3.2 Training Simulator phases

In order to achieve all these goals, this project has been subdivided in modules and structured in different phases.

The main aim of the first stage is to provide the possibility of performing virtual navigation, to teach spatial orientation and basic instrumental management. The apprentice will get used to handle the endoscope with the camera at its end while looking at the monitor where the image provided by this camera is shown. The task of localizing one determined zone of the anatomy and being able to see the tip of the other arthroscope while managing both instruments will be also accomplished in this phase.

The second step is to provide tactile feedback and tactile sensation. This will increase the system realism, because the surgeon through the instrumental will be able to touch and feel hard surfaces. To obtain tact, a haptic device will be used, so this sensation will be added to the visual feedback already provided, contributing to the immersion experience.

The last version will include the capacity of processing real data of patients who suffer from different pathologies. The library of models, common cases and unusual pathologies will be completed. All the characteristics will be added to polish environment realism and to reach greater accuracy, resulting in a simulation with a high degree of fidelity.

## 3.3 System Architecture

As shown in figure 1, a modular architecture has been constructed for the simulator, with the purpose of making an adaptable and scalable system.

Flexible *insight<sub>MIST</sub>* architecture allows its use in independent installations as well as in systems with multiple clients connected to a single data base with a central repository in the server.

**Tact and sound System** The specification of the second phase includes a force feedback system that provides great fidelity to the tact system, as well as real sounds with which to strengthen the simulation sensation of realism.

**Evaluation System** The dexterity evaluator collects information on trainee performance and provides dataset, graphical and statistical information that allow to evaluate the skill of the surgeon in formation. The evaluation system is extendable so that it can incorporate new algorithms of analysis and post-processing of simulation data.

**Ingestion of real data** Third stage specification includes the capacity to process real data that are superposed to the simulated ones, to reach the realism of the experience and the adaptation to particular cases. In this way it allows to be use like system of preparation and planning of real arthroscopy operations with real data.

**Training case Library** The library of cases includes different anatomical models, clinical reports and instrumental history. In this way the instructor will be able to



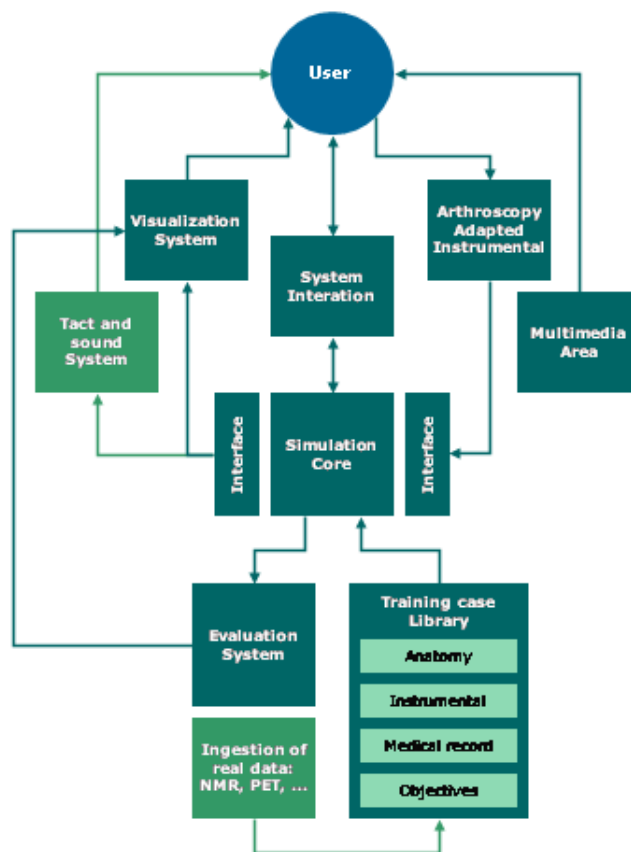


Figure 1: Functional Diagram

define new models of cases. The system is open, allowing the user to include and modify models, being possible to work on real data.

**Arthroscopy Adapted Instrumental** Specifically adapted arthroscopy instruments that resemble to the real ones are included in the Instrumental library in order to facilitate their use for the different procedures.

**Multimedia Area** It counts with multimedia courses to increase the educational capacities of the tool.

### 3.4 First Version System Description

This phase faces the first and main difficulty that arises while learning arthroscopy (and other MIS procedures): the initial disorientation novice practitioners experiment due to the fact that the surgeon and the endoscopic camera have different visual fields, which are also different from the operation work space. These fields are not only different; they also have different orientations and coordinate axis. In summary, the surgeon must handle the arthroscopes using the visual information provided by the endoscopic camera, watching a monitor instead of his own hands and the surgical instruments.

*insight<sub>MIST</sub>* counts with a control device with a pair of trocars whose displacements can be recorded by a computer, with the purpose of representing on the screen the content of the vision field of a simulated camera located at the end of one of the trocars. As noted above, for the first version it was not considered priority to provide tact feedback to the surgeon, since it is not essential to train the visual orientation (this aspect will be accomplished in the second version). The operator launches the program and specifies a file that contains data of the area to operate, in order to facilitate simulate surgical operations anywhere, and from diverse endoscopic entry points. The first dataset for simulating operations of right and left shoulders was extracted from the NPAC- Visible Human [24] and pre-processed afterwards.

Once the data file is loaded, the computer presents the trainee with an image of the area to operate, as acquired from a camera placed at the end of the left trocar (or right trocar, for left-handed practitioners), which acts as the arthroscope. It is also possible to see in the image, whenever the position of the arthroscope allows it, a representation of the tip of the instrumental operated with the right hand. The surgeon can make any camera movements (translation and rotations) and verify the results interactively.

There is also the possibility, introduced for didactic reasons, of providing a panoramic view of the whole scene, including both the position of the arthroscopes and the complete loaded anatomy. This allows novices to recover from situations of complete disorientation and to correct the instrument position during the first learning stages, helping them to plan the best entry point for the endoscopes in each particular operation.

Another interesting possibility consists of selecting the bevel angle between the endoscope axis and the camera optical axis. Having the camera aiming at a direction not coincident with the endoscope's axis allows surgeons to inspect a much larger area by rotating the endoscope and to peek through joints to areas of difficult access. This axis misalignment introduces an additional source of disorientation for novice practitioners.

The arthroscopy simulator was developed in an O2 Silicon Graphics station. Afterwards the whole system was ported to a standard PC with Linux.

### 3.5 Material used during the development

As control system, the *insight<sub>MIST</sub>* device from GMV company, is being developed. This device is a tool specifically designed for arthroscopic procedures simulations.

This device consists of two arms joined to a fixed base by a pair of joints. These arms will serve to simulate trocars used in the operation.

Optical sensors contained into the base and the joints allow to make a precise track of the each arm movements and to transmit them to a computer through a standard serial port RS-232. This type of sensors provides measures with angular and linear positions. In addition, this technology is immune to the sonorous, luminous noise as well as to electromagnetic sources, and has a very low latency which allows real time.

The device will be adapted to real instrumental, so that it will contain the trocar sheath for the telescope and the other instrument.

The game of joints that joins the arms to the base allows making movements with any degree of freedom, although always limited by the mechanics. It tracks the movements which allow the computer to reconstruct the position and state of each one, including the rotation over itself.

When the device is switched on, an auto-calibrated of the sensors is made. Next, the serial connection with the computer is made and the interface begins to respond to the requests for information about the instruments position and direction.

The application has been developed following programming criteria to assure its portability, and scalability, always having into account the next phases and assuring its flexibility

In order to facilitate the creation of new cases, as well as the modification of any of the available ones, to introduce new traumas for instance, or to modify the patient position during the operation, the format of the data used to model the anatomy of the different training cases is based on VRML, a language for modelling virtual reality employed by most of design 3D applications on the market at present.

## 4 Results

Virtual Reality techniques have been applied to build an advanced training simulator adaptable to different arthroscopic techniques with great fidelity. Realism is achieved through the gallery of real arthroscopic instruments models. The different modules and phases of the simulator will allow the practitioner to complete progressively all his learning process.

The trainee will be able first to get acquainted to the anatomy and to localize the different zones. Afterwards, it will be possible to learn to practice and recognize different pathologies, even those which are rare and will be found only once or twice by an expert surgeon in a lifetime.

Practitioners will get used to work into stressful environments when they will have to accomplish the different exercises in which time will be a critical factor and noises

and visual feedback will contribute to the stress component.

It offers a scale to measure the apprentice ability and precision as well as that of the experienced surgeon, thanks to its dexterity evaluation system.

The result will be a portable, compact equipment, easy to utilize which will be able to cover all the different stages of apprenticeship, from learning the optics in this first phase to being able in the last one to plan and practice a whole operation beforehand by loading real patient data.

## 5 Conclusions and perspectives

Nowadays, experienced surgeons perform arthroscopic interventions in the operation room, allowing a reduced number of students to watch and, later on, partially participate. This traditional training method entails limited practice and stress, since there is some risk for the patient. Another disadvantage is that there is a defined limit in the amount of time and the extent to which a novice can practice during real surgery, limiting therefore the amount of improvement obtained.

The Arthroscopy Training System is a very useful tool which can remove much of the constraints described above. It can be used extensively during practice sessions, without time limit, risks or real stress.

There are multiple surgery scenarios that can benefit from VR systems such as the Arthroscopy Training System presented here, with slight changes in the data loaded onto the simulator.

- First, the Surgery Training System may be used in several scenarios with data showing a normal healthy human anatomy, with the aim of familiarizing the apprentice with the instrumental involved in arthroscopy surgery and with what it is expected to be seen through the arthroscope.
- The system may be used then with data showing any pathology so that novices may learn how to recognise such pathologies.
- It can be adapted to be used as a completely objective method to test and evaluate a surgeon's dexterity, so that medical schools may generalise the use of such systems in their evaluation process.
- It is also possible to generate ad-hoc scenarios from real NMR data from a patient with some unusual or critical pathology before actually performing an operation. This would allow surgeons to train before actually performing the real surgery. This may permit surgeons to decide aspects such as which is the best position for the patient, or the optimal entry points to introduce the instrumental.

Being able to be loaded with different data, the Arthroscopy Training System can be used to practice different pathologies. This would permit students to learn to recognise and operate later on them. It can also be used for planning interventions, since it would allow the surgeon to essay and choose between the different options, studying different patient positions as well as which would be the best instrumental insertion points.

The evolving *insightMIST* supposes a technological challenge of great importance and it is intended to be an equipment of first level. The simulator displays in screen three-dimensional images with arthroscopic and panoramic visions, it includes simulation of arthroscopic instruments, and, in superior versions, it provides tactile sensations and it includes libraries of clinical cases. *insightMIST* tries to be a tool of support to the education for arthroscopic orientation and hand coordination with arthroscopic instruments. Also it pays special attention to aspects of involution and evaluation of the educational action by means of a suitable categorization of exercises, techniques of skill evaluation and tools multimedia.

## References

- [1] S. Bayona, L. Pastor, and J. Espadero. Entrenamiento quirúrgico mediante técnicas de realidad virtual. simulador para artroscopia. Technical report, Univ. Rey Juan Carlos, July 2002.
- [2] S. Bayona, J.M. Espadero, L. Pastor, and J.M. Fernández-Arroyo. A low-cost arthroscopy surgery training system. *IASTED03 VIIP 2003 proceedings*, 2:1005–1009, 2003.
- [3] C. Laugier, C. Mendoza, and K. Sundaraj. Towards a realistic medical simulator using virtual environments and force feedback. *Robotics Research, Springer Tracts in Advanced Robotics*, 6, 2003.
- [4] C.A. Mendoza. *Soft Tissue Interactive Simulations for Medical Applications Including 3D Cutting and Force Feedback*. PhD thesis, INPG, 2003.
- [5] K. Sundaraj. *Real-Time Dynamic Simulation and 3D Interaction of Biological Tissue: Application to Medical Simulators*. PhD thesis, INPG, 2004.
- [6] H.M. Hoffman. Virtual reality meets medical education. In K. Morgan, R.M. Satava, H.B. Sieberg, R. Mattheus, and J.P. Christensen, editors, *Interactive Technology and the New Paradigm for Healthcare*, pages 130–136. IOS Press, 1995.
- [7] Haptica. Promis. <http://www.haptica.com/id11.htm>.
- [8] Simsurgery. Simsurgery. <http://www.simsurgery.no/>.
- [9] Surgical Science Lapsim. <http://www.surgical-science.com/main/default/default.cfm>.
- [10] Surgical Science. Ls500 laparoscopy simulation platform. <http://www.xitact.com/products/lap/index.html>.
- [11] Immersion Medical. Accutouch endoscopy simulator. <http://www.immersion.com/medical/products/endoscopy/>.
- [12] Immersion Medical. Accutouch endovascular simulator. <http://www.immersion.com/medical/products/endovascular/>.

- [13] Verefi technologies. Endotower. <http://www.verefi.com/endotower.html>.
- [14] Simbionix. Gi mentor. [http://www.simbionix.com/GI\\_Mentor.html](http://www.simbionix.com/GI_Mentor.html).
- [15] Mentice. Ksa key surgical activities. <http://www.mentice.com/>.
- [16] Mentice. Vist vascular intervention system trainer. <http://www.mentice.com/>.
- [17] Mentice. Pva procedicus virtual arthroscopy. <http://www.mentice.com/>.
- [18] Simbionix. Lap mentor. [http://www.simbionix.com/LAP\\_Mentor.html](http://www.simbionix.com/LAP_Mentor.html).
- [19] Mentice. Mist minimally invasive surgical trainer. <http://www.mentice.com/>.
- [20] W. Muller and U. Bockholt. The virtual reality arthroscopy training simulator. *Medicine Meets Virtual Reality*, 6:13–19, 1998.
- [21] J.D. Mabrey, S.D. Gillogly, and J.R. Kasser et al. Virtual reality simulation of arthroscopy of the knee. *Arthroscopy*, July 2002.
- [22] R.A. Pedowitz, J. Esch, and S. Snyder. Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy*, July 2002.
- [23] VR in medicine and biology group of the Univ. of Sheffield. The sheffield knee arthroscopy training system (skats). <http://www.shef.ac.uk/vrmbg/skats.html>.
- [24] Northeast Parallel Architectures Center at Syracuse University. Npac-3d visible human. <http://www.npac.syr.edu/projects/3Dvisiblehuman>.

# Hardware-Accelerated Dynamic Volume Rendering for Real-Time Surgical Simulation

Marco Agus, Andrea Giachetti, Enrico Gobbetti  
Gianluigi Zanetti, and Antonio Zorcolo

CRS4

Center for Advanced Studies, Research and Development in Sardinia

POLARIS Edificio 1, 09010 Pula (CA), Italy

{magus, zarco, giach, gobbetti, zag}@crs4.it

**Abstract.** We developed a direct volume rendering technique, that supports low latency real time visual feedback in parallel with physical simulation on commodity graphics platforms. In our approach, a fast approximation of the diffuse shading equation is computed on the fly by the graphics pipe-line directly from the scalar data. We do this by exploiting the possibilities offered by multi-texturing with the register combiner OpenGL extension, that provides a configurable means to determine per-pixel fragment coloring. The effectiveness of our approach, that supports a full decoupling of simulation and rendering, is demonstrated in a training system for temporal bone surgery.

## 1 Introduction

The diffusion of minimally invasive procedures is bringing major improvements in the quality of the care provided to the patients. This is done, however, at the cost of the increase in the complexity of the surgical procedures performed, thus requiring an increase in the training needed by each specific intervention for both planning and procedural issues. At the same time, the shortage of cadavers for medical training and public concern with the inhuman treatment of animals is drastically limiting the traditional approaches to surgical training. Virtual reality simulators realistically mimicking a patient-specific operating environment would therefore significantly contribute to the improvement of surgical training. Developing high quality simulators is, however, extremely difficult, since the need to provide real-time feedback to users, while simulating physical effects, imposes stringent constraints on the simulation and visualization system. This paper focuses on the dynamic visualization problem, describing the volume rendering technique that we have developed for a temporal bone dissection simulator.

Direct volume rendering with shading, that works by integrating along selected projectors the value of a continuous emission/reflection/absorption volume function reconstructed from discrete sampling points [1], is the de-facto standard in the pre-operative analysis of medical data. By manipulating the mapping from values of the original volume data to emission, reflection, and absorption coefficients, various effects can be achieved, including isosurfaces and opaque objects. Using this data intensive technique on dynamic volumes under real-time constraints is, however, an open research problem.

This fact has limited simulators to employ surface based techniques, that have problems with semitransparent materials and rely on complex mesh structures that impose important synchronization overheads. A number of authors have proposed to exploit texture mapping and rasterization hardware to render scalar volumes at interactive speeds [2–5]. These techniques are based on uploading the scalar volume to texture memory prior to rendering object-aligned or view-direction-aligned textured volume slices. One of the major limitations of these methods is their inability to efficiently implement surface illumination models, since texture lookup is based only on data values and not on gradient information. The quality of the images is therefore insufficient for accurately perceiving surface shape. Various authors have thus proposed alternative techniques for supporting hardware-accelerated direct volume rendering with shading [4, 6–8]. While image quality is close to that of the best software solution, this comes at the expense of performance and texture memory overheads, since the proposed techniques require multiple passes through the rasterization hardware and/or precomputation of gradient volumes. This is unacceptable in surgical simulation, since the volume is continuously varying, and thus we cannot efficiently compute and reload gradient maps without strongly coupling the simulation and rendering tasks.

In this paper, we propose a texture-based volume rendering approach, that supports low latency real time visual feedback to occur in parallel with physical simulation, without requiring any synchronization among the threads. In our approach, a fast approximation of the diffuse shading equation is computed on the fly by the graphics pipe-line directly from the scalar data. We do this by exploiting the possibilities offered by multi-texturing with the register combiner OpenGL extension, that provides a configurable means to determine per-pixel fragment coloring. The rest of the paper describes our technique and illustrates its effectiveness in a virtual training system for temporal bone surgery [9].

## 2 Interactive volume rendering approach

Although volumetric data is defined over a continuous three-dimensional domain ( $R^3$ ), measurements and simulations provide volume data as 3D arrays, that can be easily used as scalar texture images, without pre-processing. We render this volume sampling the volume through texturing hardware using front-to-back slice composition. For each step and for all pixels, graphics hardware accesses the texture and extracts the scalar value. This sampled value is converted, through a transfer function, to a color triple and an opacity value that are saved inside combiner input registers. Combiners are then programmed to compute gradients and shading on-the-fly, and return the color used during the blending process. In the following, we provide more details about this process.

**Sampling through texture mapping.** Current consumer graphics hardware is based on an object-order rasterization approach, i.e. primitives (polygons, lines, points) are scan-converted and written pixel-per-pixel into the frame buffer. Since volume data do not consist of such primitives, a proxy geometry is defined for each individual slice through volume data. Each slice is textured with corresponding data from tvolume. The



volume is reconstructed during rasterization on slice polygons by applying a convolution of volume data with a filter kernel. The entire volume can be represented by a stack of such slices, if the number of slices satisfies restrictions imposed by Nyquist theorem. Our approach follows the technique proposed by Rezk-Salama and others [7], extending it with on-the-fly gradient and shading computation. In our case, since in a surgical setting viewpoint motion is constrained (limited to maximum 30 degrees), we are allowed to use object-aligned slices. At each frame, we traverse the array of slices, and we reload them as textures two at a time. After texture loading and reconstruction, the rasterization process derives, for each projected pixel, texture sampling position, and texturing hardware extracts the correspondent density, by bi-linear or tri-linear interpolation of closest texels; the resultant value is then mapped to an RGBA vector by the transfer function. In order to compute the surface gradient, four texture units are needed, that are used to sample the volume with offset  $dx$ ,  $dy$ ,  $dz$ , relatively to the central point.

This procedure is extremely efficient, since all the computation is performed in parallel in the graphics hardware and no particular synchronization is needed between the renderer and the process that is modifying the dataset. Only a single sweep through the volume is needed, and volume slices are sequentially loaded into texture memory on current standard PC graphics platform using AGP 8X transfers, which provides a peak bandwidth of 2108 MB/s.

**Transfer function.** The transfer function mapping, in our direct volume rendering approach, is obtained by exploiting the *glColorTable* primitive, that is commonly implemented in commodity graphics hardware. This function is used by compiling a look-up table, with transfer function values, and by installing it inside graphics memory. At the same time of sampling, texturing hardware performs transfer conversion of density values to RGBA colors contained inside the table. The color table contains associated colors instead of pure colors [10], in order to control the color interpolation error. The associated color employment has also beneficial effects to color accumulation process. The color look-up table lets users choose and calibrate the transfer function in real time; it can be computed and reloaded each time the user change some function parameters.

**Voxel color computation.** Lighting and shading process follows the standard lighting equation [10]:

$$\tilde{C}[n] = \alpha[n]C_a[n]l_a + \alpha[n]C_d[n] \|\bar{l}_d\| \max \left( \frac{\nabla f[n]}{\|\nabla f[n]\|} \cdot \frac{\bar{l}_d}{\|\bar{l}_d\|}, 0 \right) \quad (1)$$

where  $\tilde{C}[n]$  is the associated color,  $C_a[n]$ ,  $C_d[n]$  and  $\alpha[n]$  are the non-directional ambient reflective factor, the diffuse directional reflective factor and the voxel opacity, while  $l_a$  and  $\bar{l}_d$  are the ambient light intensity and the light intensity coming from the directional source. In order to highlight surface details, we employ the artifact proposed by [11] of weighting the opacity  $\alpha[n]$  with a surface *strength*, evaluated as a function of volume and his gradient:  $S = h(f(s), \nabla f(s))$  [12]. If we use the gradient modulus as *strength*, we have:

$$\tilde{C}[n] = l_a \|\nabla f[n]\| \alpha[n] C_a[n] + \|\bar{l}_d\| \alpha[n] C_d[n] \max \left( \nabla f[n] \cdot \frac{\bar{l}_d}{\|\bar{l}_d\|}, 0 \right). \quad (2)$$

Such a strength function, enables the visualization of boundary surface between tissues, and disables the visualization of parts with null gradient (like the internal parts of an object). Specifically, in our case, the shading components are supposed to be the combination of an ambient component and a directional component emitted by a source oriented along the volume z axis (slices normal). This way, the dot product between the light direction and the opacity gradient is the component  $\nabla_z f[n]$ , and equation 1 is simplified as follows:

$$\tilde{C}[n] = l_a \|\nabla f[n]\| \alpha[n] C_a[n] + l_d \alpha[n] C_d[n] \max (\nabla_z f[n], 0). \quad (3)$$



**Fig. 1.** Internal and external optical models.

But visual result of equation 3 is not fully satisfying: in fact only surface voxels contribute to pixel color, because strength becomes zero in tissue internal parts. This fact would be irrelevant if surfaces were consistent enough to completely mask the color of internal voxels. Anyway, low strength surfaces and small width walls let see the hollows produced by equation 3 (see fig. 1a). Hence, the optical model employed for internal volumes is different from that used exclusively for parts having non null gradient. The overall model is then defined by:

$$\tilde{C}[n] = \begin{cases} l_a \|\nabla f[n]\| \alpha[n] C_a[n] + l_d \alpha[n] C_d[n] \max (\nabla_z f[n], 0) & \text{if } \|\nabla f[n]\| > 0 \\ (l_a + l_d) \alpha[n] C_a[n] & \text{if } \|\nabla f[n]\| = 0 \end{cases}. \quad (4)$$

With this artifact, the image quality is greatly improved during rendering of low density tissues or in these case of tissue subtle layers with high density (for example bone), as shown in figure 1b.

**Opacity gradient computation.** In lighting equation 2, the surface normal is related to gradient  $\nabla f[n]$ , and the modulus is regarded as surface *strength*. If  $f[n]$  is used as opacity, instead of density, we can arbitrarily modify the surface appearance properties (opacity, width and consistence) by modifying the transfer function. Since OpenGL register combiners receive from texture hardware 4 opacity values  $\alpha(p)$ ,  $\alpha(p + dx)$ ,  $\alpha(p + dy)$ ,  $\alpha(p + dz)$ , they are able to approximate the opacity gradient with forward differences. Since combiners are SIMD arithmetic modules able to perform linear operations, it is relatively simple to derive forward differences and vector modules, but it is impossible to perform ratios and root extractions. Now the gradient norm computation involves a square root computation, that needs to be approximated with a polynomial function. Since the number of available combiners is limited and many of them are used to compute the gradient components as well as the lighting equation, we can only approximate the square root function with a quadratic function. The 2nd order polynomial is derived from a Taylor series evaluated in the neighborhood of an arbitrary point  $x_0$  of interval  $]0, 1]$ . The value of  $x_0$  has to be chosen in order to minimize the approximation error in the interval. According to equation 2, the gradient norm is used to weight voxel opacity and associated color contribution, so the best approximation is obtained when Taylor series is evaluated in  $x_0 = 1$ . The interpolation function is then  $\sqrt{x} \approx \frac{3}{8} + \frac{3}{4}x - \frac{1}{8}x^2$ .

**Performance enhancement.** Pixel fill-rate is the major limiting factor when using a texturing approach to volume visualization. In zoom rendering, an appropriately down-scaled image is rendered in the back buffer and then enlarged and copied to the front buffer [13, 14]. In this way, delays associated with buffer swap synchronization are avoided, and the number of pixels filled during volume rendering is reduced. In our implementation, the copy and zoom operations are implemented by copying the reduced size image in texture memory and then rendering a textured polygon in the front buffer. Hence, sophisticated texture interpolation algorithms can be used to reduce the artifacts caused by magnification.

### 3 Implementation and results

Our technique for direct volume rendering has been integrated in a prototype training system for mastoidectomy. The simulator system provides real-time visual and haptic feedback [9, 15] and it is modeled as a collection of loosely coupled concurrent components [16]. The overall system is divided in a "fast" subsystem, responsible for the high frequency tasks (surgical instrument tracking, force feedback computation, bone erosion), and a "slow" one, essentially dedicated to the production of data for visual feedback [17]. The system runs on two interconnected multiprocessor machines. Thanks to our volume rendering approach, the renderer is totally decoupled from the simulator

and the tracking system, and runs at his own frequency. The current configuration is the following: a single-processor PIV/1500 MHz with 256 MB PC133 RAM for the high-frequency tasks (haptics loop (1KHz) and interprocess communication loop); a dual-processor Intel Xeon 2.4 GHz with 2048 MB DDR PC400 RAM and a NVIDIA GeForce FX 5800 Ultra AGP 8X and running a 2.4 linux kernel, for the low frequency tasks (receiving loop, simulator evolution and visual rendering); a Phantom Desktop and a Phantom 1.0 haptic devices, that provide 6DOF tracking and 3DOF force feedback for the burr/irrigator and the sucker; a n-vision VB30 binocular display for presenting images to the user. We are currently using a volume of 256x256x128 cubical voxels (0.3 mm side) to represent the region where the operation takes place. We executed performance benchmarks on the system, which revealed that, using eight register combiners and 8 bit/texture volumes, peak texture transfer rate is about 400M texel/s, while peak fill rate is about 400M pixel/s per second. According to these results, the rendering system should theoretically be able to completely reload and render an entire 256X256X128 dataset in about 40 ms per frame. In the surgical simulator system, with this volume size, and using a window of about the same resolution (320X240 zoomed to 640X480), we obtain refresh timings of about 50 msec per frame, corresponding to a frame rate of 20 fps, which is close to the theoretical peak. The CPU overhead is negligible, and the simulation can run and update the volume in parallel in a totally decoupled manner. The performance of the prototype is thus sufficient to meet timing constraints, even though the computational and visualization platform is constructed from affordable and widely accessible components. The visual quality of the method is illustrated in figure 2, which shows snapshots captured during a virtual session of the surgical simulator. The principal steps of a basic mastoidectomy, performed by an Ear, Nose and Throat surgeon, are represented.

## **4 Conclusion and discussion**

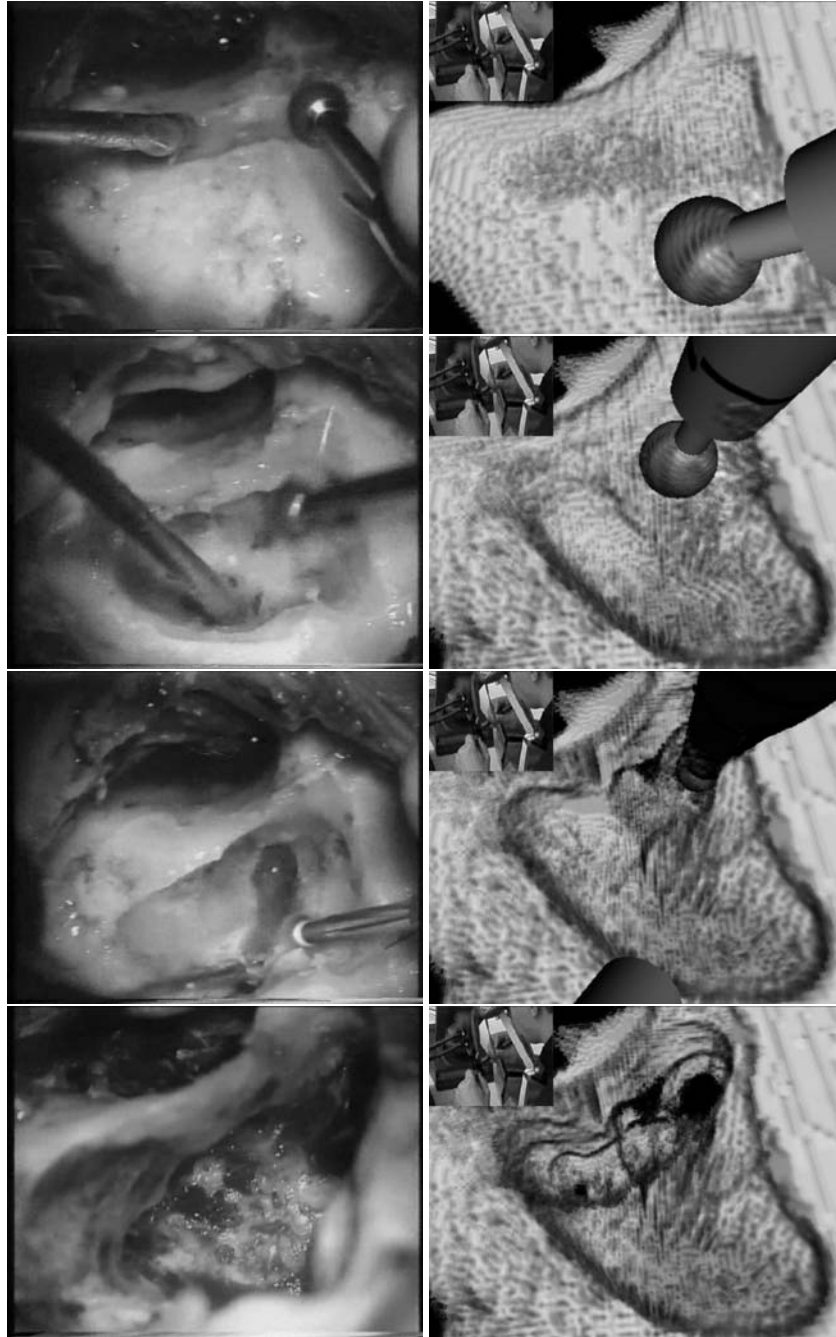
We presented a dynamic volume rendering technique which is well suited for the incorporation in surgical simulators. The technique supports low-latency and high frequency rendering of shaded semi-transparent materials. The method is extremely efficient, since all the computation is performed in parallel in the graphics hardware and no particular synchronization is needed between the renderer and the process that is modifying the dataset. Only a single sweep through the volume is needed, and volume slices are sequentially loaded into texture memory on current standard PC graphics platform using AGP transfers. The effectiveness of our approach is demonstrated in a training system for temporal bone surgery.

## **Acknowledgments**

We would like to thank Prof. Stefano Sellari Franceschini and his team, University of Pisa, for his collaboration in the design and testing of the system.

## References

1. Max, N.: Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics* **1** (1995) 99–108
2. Cabral, B., Cam, N., Foran, J.: Accelerated volume rendering and tomographic reconstruction using texture mapping hardware. In Kaufman, A., Krueger, W., eds.: 1994 Symposium on Volume Visualization, ACM SIGGRAPH (1994) 91–98 ISBN 0-89791-741-3.
3. Guan, S., Lipes, R.G.: Innovative volume rendering using 3D texture mapping. In: *Image Capture, Formatting and Display*. Volume 2164 of SPIE. SPIE (1994)
4. Van Gelder, A., Kim, K.: Direct volume rendering with shading via three-dimensional textures. In: 1996 Volume Visualization Symposium, IEEE (1996) 23–30 ISBN 0-89791-741-3.
5. Kulick, T.: Building an opengl volume renderer. *SGI Dev. News* (1996)
6. Westermann, R., Ertl, T.: Efficiently using graphics hardware in volume rendering applications. In Cohen, M., ed.: *SIGGRAPH 98 Conference Proceedings*. Annual Conference Series, ACM SIGGRAPH, Addison Wesley (1998) 169–178 ISBN 0-89791-999-8.
7. Rezk-Salama, C., Engel, K., Bauer, M., Greiner, G., Ertl, T.: Interactive volume rendering on standard PC graphics hardware using multi-textures and multi-stage rasterization. In Spencer, S.N., ed.: *Proceedings of the 2000 SIGGRAPH/EUROGRAPHICS Workshop on Graphics Hardware*, ACM Press (2000) 109–118
8. Engel, K., Kraus, M., Ertl, T.: High quality pre-integrated volume rendering using hardware-accelerated pixel shading. In: *EuroGraphics/SIGGRAPH Workshop on Graphics Hardware*. (2001)
9. Agus, M., Giachetti, A., Gobbetti, E., Zanetti, G., Zorcolo, A.: Adaptive techniques for real time haptic and visual simulation of bone dissection. In: *IEEE Virtual Reality Conference*, Conference held in Los Angeles, CA, USA, March 22–26 (2003) 102–109
10. Wittenbrink, C.M., Malzbender, T., Goss, M.E.: Opacity-weighted color interpolation for volume sampling. In: *IEEE Symposium on Volume Visualization*, IEEE, ACM SIGGRAPH (1998) 135–142
11. Levoy, M.: Display of surfaces from volume data. *IEEE Computer Graphics and Applications* **8** (1988) 29–37
12. Drebin, B., Carpenter, L., Hanrahan, P.: Volume rendering. In Wolfe, R., ed.: *Significant Seminal Papers of Computer Graphics: Pioneering Efforts that shaped the Field*, N.Y., ACM Press (1998) 363–372
13. Mazuryk, T., Schmalstieg, D., Gervautz, M.: Zoom rendering: Improving 3-D rendering performance with 2-D operations. Technical Report CG, Institute of Computer Graphics, Vienna University of Technology (1995)
14. Gobbetti, E., Pili, P., Zorcolo, A., Tuveri, M.: Interactive virtual angioscopy. In: *Proceedings IEEE Visualization*, Conference held in Research Triangle Park, NC, USA, IEEE Computer Society Press (1998) 435–438
15. Agus, M., Giachetti, A., Gobbetti, E., Zanetti, G., Zorcolo, A.: Real-time haptic and visual simulation of bone dissection. *Presence: Teleoperators and Virtual Environments* **12** (2003) 110–122
16. Agus, M., Giachetti, A., Gobbetti, E., Zanetti, G., Zorcolo, A.: A multiprocessor decoupled system for the simulation of temporal bone surgery. *Computing and Visualization in Science* **5** (2002)
17. Agus, M., Giachetti, A., Gobbetti, E., Zanetti, G., John, N.W., Stone, R.J.: Mastoidectomy simulation with combined visual and haptic feedback. In Westwood, J.D., Hoffmann, H.M., Mogel, G.T., Stredney, D., eds.: *Medicine Meets Virtual Reality 2002*, IOS Press (2002) 17–23



**Fig. 2. Comparison between real and virtual intervention:** the principal steps of a basic mastoidectomy, performed by a surgeon, are represented. Photos courtesy of Prof. Stefano Sellari Franceschini, University of Pisa.



# **Tecnologías del lenguaje humano**

Alexander Gelbukh  
Aurelio López López  
Manuel Montes y Gómez  
Luis Villaseñor Pineda  
(Editores)





## Introducción

**¿Qué son las Tecnologías del Lenguaje Humano?** El lenguaje humano ha sido objeto de estudio desde hace mucho tiempo, desde diversos puntos de vista y en diferentes disciplinas. El uso de la computadora no sólo permitió acelerar su estudio sino que abrió nuevos campos de investigación. Es justo alrededor de la computadora, que nacen las tecnologías del lenguaje humano (TLH). En breve, podemos definir las TLH como el conjunto de conocimientos y medios involucrados en el tratamiento automático de las lenguas humanas. Es importante remarcar que el lenguaje humano no puede reducirse únicamente a su forma oral o escrita. El lenguaje humano va más allá e incluye otros modos de comunicación. Por ejemplo, el habla se combina con ademanes indicativos y expresiones faciales; y los textos digitales presentan combinaciones con imágenes y sonidos. De esta manera, las tecnologías del lenguaje incluyen el procesamiento de la comunicación multimodal y de los documentos multimedia.

**Relevancia de las TLH para la comunidad hispano hablante** El idioma Español es hablado por aproximadamente 400 millones de personas, incluyendo todas sus variedades fonéticas. Sin embargo, la influencia hoy en día de nuestro idioma en la sociedad de la información es mínima. Desgraciadamente la ciencia y, más concretamente, la tecnología asociada al tratamiento del idioma Español no se ha desarrollado con el ímpetu necesario. Nos encontramos con un enorme rezago tecnológico que sólo podrá resolverse a través de una participación decidida. Para ello es indispensable crear herramientas propias capaces de tratar la información expresada en nuestro idioma. Es por ello indispensable el desarrollo de una infraestructura tecnológica para el tratamiento automático del lenguaje Español. Cabe hacer mención de la responsabilidad de México en la búsqueda de soluciones a esta problemática. México es el país hispano hablante más poblado con cerca de 100 millones de habitantes sin contar la enorme presencia de los emigrantes mexicanos en E. U. A., una comunidad de más de 20 millones de personas.

**Objetivo del taller de TLH** El presente taller tiene por objetivo principal reunir a todos aquellos grupos de investigación mexicanos interesados en las TLH. Este taller está abierto a todos aquellos esfuerzos relacionados con el procesamiento del lenguaje humano por medios computacionales. Como se menciona, esta área involucra aportaciones de diferentes disciplinas, entre ellas podemos mencionar: la Lingüística computacional, la Recuperación de información textual, la Categorización de textos, la Ingeniería lingüística, el Reconocimiento de voz, la Extracción de información, los Sistemas conversacionales, entre otras.

**Antecedentes inmediatos de este taller** Desde hace tiempo diferentes grupos de investigación en México han trabajado diversos aspectos del tratamiento automático del Español. Sin embargo, es apenas en años recientes que esta comunidad ha comenzado a reunirse con la finalidad de abordar esta enorme problemática de manera conjunta. Hasta ahora esta comunidad ha realizado dos reuniones regionales, una en la ciudad de Puebla organizada por la Benemérita Universidad Autónoma de Puebla y otra en la ciudad de México organizada por el Instituto de Ingeniería de la UNAM. El

objetivo de las reuniones se centró en la discusión de avances y la búsqueda de proyectos conjuntos. Por otro lado, como consecuencia inmediata de estas reuniones se creó la Asociación Mexicana de Procesamiento de Lenguaje Natural. Por supuesto, este taller será un medio idóneo para dar a conocer esta asociación e invitar a nuevos grupos a participar en ella.

## **Comité Científico**

Copresidentes:

Alexander Gelbukh, CIC-Instituto Politécnico Nacional

Aurelio López López, Instituto Nacional de Astrofísica, Óptica y Electrónica

Manuel Montes y Gómez, Instituto Nacional de Astrofísica, Óptica y Electrónica

Luis Villaseñor Pineda, Instituto Nacional de Astrofísica, Óptica y Electrónica

Javier Cuétara, Fac. de Filosofía y Letras-Univ. Nacional Autónoma de México

Sofía Galicia Haro, Fac. de Ciencias-Universidad Nacional Autónoma de México

Warren Greiff, MITRE Corporation

Héctor Jiménez, Benemérita Universidad Autónoma de Puebla

Ingrid Kirschning, Universidad de las Américas, Puebla

Carlos Alberto Reyes, Instituto Nacional de Astrofísica, Óptica y Electrónica

Luis Pineda Cortés, IIMAS-Universidad Nacional Autónoma de México

Paolo Rosso, Universidad Politécnica de Valencia

Grigori Sidorov, CIC-Instituto Politécnico Nacional

Gerardo Sierra, INI-Universidad Nacional Autónoma de México



# Identificación de Antónimos en Textos Planos

Cupertino Lucero, David Pinto y Héctor Jimenez-Salazar

Facultad de Ciencias de la Computación  
Benemérita Universidad Autónoma de Puebla  
14 sur y Av. San Claudio. Edif. 135. Ciudad Universitaria  
Puebla, Pue. 72570. México  
Tel. (+52222) 229 55 00 ext. 7212 Fax (+52222) 229 56 72  
QPr@hotmai.com, dpinto@cs.buap.mx, hjimenez@fcfm.buap.mx

**Abstract.** WordNet has been used in several applications of Natural Language Processing field, however, the fact that it is not a specialized lexical database (LDB) led us to construct LDB's for specific domains. Thereby it is important to develop automatic methods for the recognition of lexical relationships. In this paper we present a new method for identifying word pairs (extracted from a raw corpus) in oppositeness relationship. The method was applied to a set of related word pairs obtained from a corpus by using Grefenstette's method. Also, we used features extracted from word contexts. These features were evaluated by a word-distance between pairs that appears in the same context, lexical-syntactical patterns that matches regular expressions with word contexts, and by lexical co-occurrence networks built for each word of a pair. The method was tested with a set of word pairs taken from an economy corpus obtaining a 80-percent accuracy.

**Key Words:** Oppositeness relationship, lexical co-occurrence network.

**Resumen.** WordNet se ha usado en varias aplicaciones y en muchos campos del procesamiento del lenguaje natural, sin embargo, el hecho de que no sea una base de datos léxica (BDL) especializada ha conducido a la construcción de BDLs para dominios específicos. Por ello, es muy importante desarrollar métodos automáticos para el reconocimiento de relaciones léxicas. En este trabajo se presenta un nuevo método para identificar pares de palabras en relación de oposición provenientes de un corpus sin formato. El método fue aplicado a pares de palabras relacionadas que fueron obtenidas de un thesaurus creado por el método de Grefenstette. Para lograr esto, también se usan algunos rasgos extraídos de los contextos de las palabras. Estos rasgos fueron evaluados a partir de la distancia entre las palabras que aparecen en el mismo contexto, por patrones léxico-sintácticos usados para empatar expresiones regulares en los contextos de las palabras, y por una red de co-ocurrencia léxica construida para cada palabra relacionada. El método fue probado en un conjunto de pares de palabras tomadas de un thesaurus de economía, y se obtuvo un 80 por ciento de precisión, lo cual es muy alentador.

**Palabras Clave:** Relación de oposición, redes de co-ocurrencia léxica.

## 1. Introducción

WordNet, es una Base de Datos Léxica (BDL) que tiene una amplia cantidad de relaciones léxicas entre palabras para el idioma inglés. A través del tiempo, WordNet ha tenido varias aplicaciones en muchos campos del procesamiento del lenguaje natural, como es el caso de la recuperación de información, desambiguación del sentido de una palabra, y refinamiento de otras bases de datos léxicas, entre otras [4,8,9,11]. Sin embargo, la poca especialización de esta base de datos conduce a la necesidad de construir BDL para dominios específicos, tarea que necesita métodos automáticos para reducir el tiempo y el esfuerzo, y permitir así su aplicabilidad a otros dominios.

Los procedimientos automáticos que se han propuesto para identificar las relaciones léxicas subyacentes en el lenguaje natural siguen de cierta manera el procedimiento manual. En esencia, se trata de observar las concordancias de dos términos para proponer una relación entre ellos. Por ejemplo, P. Hindle [14] propuso un método basado en la asignación de rasgos sintácticos frecuentes de un término, para representar y determinar la similitud entre sustantivos. Un método semejante es el que propuso G. Grefenstette [1] para construir thesauri por "conocimiento pobre", partiendo de un texto grande en un dominio específico. Este investigador observó que dos términos son mutuamente vecinos, si uno es altamente frecuente en los contextos del otro, y viceversa. En esta misma dirección, un trabajo para el portugués de C. Veraschin [5] refinó el método, incluyendo mayor información sintáctica en la representación de los términos a relacionar usando frases preposicionales. Tal vez el trabajo más significativo sea el de M. Hearst [7], quien obtuvo relaciones de hiponimia a partir de patrones léxico sintácticos. Por otra parte, Sanderson y Croft [15] identifican relaciones de hiponimia con base en la idea de subsunción, sustentada en la contención del conjunto de documentos que contienen ambos términos en el conjunto de documentos que contiene a uno de ellos, es decir:  $x$  subsume a  $y$  si los textos que contienen a  $y$  tienen una alta probabilidad de contener a  $x$ . Mas recientemente, H. Jiménez [6] señaló que el término puede usarse con diferentes sentidos y, por tanto, se pueden identificar algunas relaciones de sinonimia e hiponimia para algunos sentidos de las palabras en cuestión, refiriendo a la subsunción mediante el agrupamiento de los rasgos que representan los sentidos de los términos. Un enfoque que decide si un par de palabras son antónimos usa vectores conceptuales [12]: descompone cada término del par, por medio de un MRD (Machine Readable Dictionary) y un thesaurus. Este método determina si el par está en relación de antonimia, realizando operaciones en el espacio vectorial, como la medida del coseno.

En este trabajo se presenta un método para identificar antónimos (o en relación complementaria, como "verdadero" y "falso", Cruse [13]). Este método está basado en la representación de pares de términos -que se han encontrado como relacionados pero sin conocer el tipo de relación- mediante tres rasgos provenientes de sus contextos: la puntuación inversamente proporcional aportada por la distancia que separa el par de palabras, algunos patrones léxico sintácticos y el grado de similitud determinado por una red de co-ocurrencia léxica

construida para cada palabra de los pares. El corpus del cual se extraen los pares relacionados también provee los contextos para determinar los rasgos. Se tomaron un conjunto de pares que fueron usados como ejemplos positivos. Los umbrales útiles en la clasificación fueron definidos con base en los valores obtenidos con los ejemplos positivos.

El determinar la relación de antonimia es un problema difícil de resolver debido a que los antónimos y los sinónimos presentan rasgos similares. Así, es necesario establecer un mecanismo que permita distinguir entre estos dos tipos de relaciones. De manera específica, D. Cruse [13] comenta sobre los opositivos lo siguiente:

“... in respect to all other features, they are identical, hence their semantic closeness; along the dimension of difference, they occupy opposing poles, hence the feeling of difference.”

esto también fue expresado por L. Wanner [10]: los antónimos  $x$  y  $y$  cumplen que  $x$  tenga como rasgos  $ABC$  y  $y$  tenga  $AB-C$ .

En este trabajo se recurre al uso de las redes de co-ocurrencia léxica, las cuales fueron usadas por Philip Edmonds [3] para seleccionar el sinónimo más adecuado en un contexto, y que sirven también como un mecanismo de filtrado en el proceso de determinación de antónimos.

Cada uno de los rasgos usados en la clasificación de relaciones de oposición se encuentran descritos en la sección 2 de este trabajo. La sección 3 indica cómo calcular los pesos asociados a los rasgos, y muestra una prueba del procedimiento de clasificación sobre un ejemplo de pares relacionados. Al final, se presenta una discusión de los resultados obtenidos.

## 2. Rasgos Utilizados en la Identificación de Antónimos

El método de clasificación de rasgos propuesto, está basado en una función de puntaje, compuesta por los pesos de los mismos rasgos. El método usa los umbrales determinados por ejemplos positivos. A continuación se describe cada uno de estos rasgos necesarios en el cálculo del puntaje, el cual permite clasificar pares de palabras.

**Distancia Inversamente Proporcional (DIP).** Este rasgo tiene su base en la observación de contextos que tienen palabras relacionadas. DIP representa qué tan cercanas están dos palabras. En contextos de palabras relacionadas, las palabras antónimas muy frecuentemente co-ocurren con una distancia pequeña. Esta observación está basada por el uso de antónimos con propósitos de contraste. Se define DIP de la siguiente manera: la distancia entre las palabras relacionadas (número de palabras que las separan) es complementada con respecto a la distancia máxima aportada por ejemplos positivos. Se toma el valor máximo de la distancia complementaria en los ejemplos positivos como:



$$\Delta_M = \max_{(x_1, y_1) \in Pos} \left\{ \max_{(x_2, y_2) \in \mathcal{B}} \bar{\Delta}(x_2, y_2) - \bar{\Delta}(x_1, y_1) \right\} \quad (1)$$

donde  $\bar{\Delta}(\mathbf{x}, \mathbf{y})$  es la distancia promedio entre las palabras  $x$  y  $y$  en sus contextos.

**Patrones Léxico Sintácticos (PLS).** Los PLSs han sido introducidos en el trabajo de P. Hearst [7] con buenos resultados en la identificación de hipónimos. Se identificaron diversos patrones en los contextos con relaciones de oposición guiados por palabras clave, signos de puntuación, y distancia entre palabras relacionadas. Los PLSs son representados como expresiones regulares. Ejemplos de estos patrones aparecen en la tabla 1. Los patrones contienen palabras clave como: *pero*, *desde*, *hasta*, *sino*, *y*, *o*.

Nr	Expresión Regular	Peso
1	Ant1 word*, <b>pero</b> word* Ant2	5
2	<b>desde</b> word* Ant1 <b>hasta</b> word* Ant2	4
3	Ant1 word* [,;] <b>sino</b> word* Ant2	5
4	Ant1 word{0,4}[y o] word{0,4} Ant2	1

Tabla 1. Expresiones regulares y sus pesos.

**Redes de Co-ocurrencia Léxica (RCL).** Debido a que los antónimos se comportan de manera similar a los sinónimos, se tuvo que discernir entre estas dos relaciones. Por lo tanto, se decidió manejar una representación para palabras que pudieran contener información usada en problemas con sinonimia. Las RCL fueron usadas en esta propuesta como un mecanismo de selección del sinónimo mas adecuado en un contexto [3]. El procedimiento usado para calcular una RCL para un término, llamado *raíz*, se describe a continuación:

1. Para formar el contexto de la raíz,  $x$ , se consideran las oraciones del corpus (usado en la construcción de nuestro thesaurus),  $\mathcal{C}$ , que la contienen:

$$A_1(x) = \{y | \mathbf{x}, y \text{ co-ocurren en una oración de } \mathcal{C}\} \quad (2)$$

2. El contexto de la raíz es filtrado, descartando todas las palabras cuya información mutua [2] es menor que 5; el resto de las palabras se les llama palabras asociadas de primer-orden:

$$A'_1(x) = \{y | y \in A_1(x) \wedge MI(\mathbf{x}, y) > 5\} \quad (3)$$

3. El proceso se repite para las palabras  $y \in A'_1(x)$  (palabras asociadas de segundo-orden). Esto depende del nivel deseado de la RCL. En general, las palabras asociadas de  $n$ -orden para  $x$  son determinadas de acuerdo con:

$$A'_n(x) = \bigcup_{y \in A'_{n-1}(x)} A'_1(y) \quad (4)$$

La necesidad de discernir entre relaciones de sinonimia y antonimia de un par  $(a_1$  y  $a_2)$ , requirió del uso de RCL para decidir si pueden ser consideradas sinónimos. Una manera de hacer esto es calcular la similitud relativa entre las palabras. La similitud, en términos de RCL, está definida en [3] con el propósito de calcular la similitud de una palabra  $w$  en un contexto, la cual está basada en la similitud entre las palabras  $w$  y  $x$ , donde  $x$  debería estar en la RCL de  $w$ . Esto permite suponer que cada arco de la RCL tenga un peso (este tópico será explicado en la sección 3), y considerando  $P = (w_0, w_1, \dots, w_n)$  como el camino de costo mínimo de  $w_0 = w$  a  $w_n = x$ , entonces, la similitud de  $w$  y  $x$  queda definida como sigue:

$$sig(\mathbf{w}; ) = \frac{1}{d^3} \sum_{w_i \in P} \frac{t(w_{i-1}, w_i)}{i}, \quad (5)$$

donde  $t(w_{i-1}, w_i)$  es el  $t$ -score definido en [2], y se calcula de la siguiente manera:

$$t(w_{i-1}, w_i) = \frac{P(w_{i-1}, w_i) - P(w_{i-1}) \cdot P(w_i)}{\sqrt{(\sigma^2(P(w_{i-1}, w_i)) + \sigma^2(P(w_{i-1}) \cdot P(w_i)))}} \quad (6)$$

$$\text{con } P(w_{i-1}, w_i) = \frac{fr(w_{i-1}, w_i)}{N} \text{ y } \sigma^2 P(w_{i-1}, w_i) \cong N \cdot P(w_{i-1}, w_i).$$

Dadas las palabras  $a_1$  y  $a_2$  con nodos en RCL  $L(a_1)$  y  $L(a_2)$ , respectivamente, se espera que ellas tengan una alta similitud si la suma de las similitudes de las palabras en  $L(a_1) \cap L(a_2)$  para ambas RCLs es alta. Por lo tanto, es necesario saber cuándo la similitud es alta. En este trabajo se hace referencia a la similitud total para calcular la similitud relativa. La similitud total,  $s_t$ , es calculada sumando todos los pesos de ambas  $L(a_1)$  y  $L(a_2)$ . En resumen, la similitud relativa entre las palabras  $a_1$  y  $a_2$  está definida como:

$$s_r(a_1, a_2) = \frac{1}{s_t} \sum_{w \in \{a_1, a_2\}, x \in L(a_1) \cap L(a_2)} sig(\mathbf{w}; ), \quad (7)$$

Se puede entonces decir que los rasgos ayudan a determinar el puntaje total de un par de términos relacionados, para determinar si estos son sinónimos.

### 3. Determinación del Puntaje Total

El puntaje total considera cada rasgo descrito anteriormente. De acuerdo con los valores observados en ejemplos positivos, un peso es asignado a cada rasgo. Los pesos más altos expresan que los valores de dichos rasgos fueron observados en ejemplos positivos. Por ejemplo, un peso bajo dado a una expresión regular indica que algunos ejemplos negativos empatan con tal patrón.

El puntaje total  $S_g(a_1, a_2)$  es la suma de todos los valores de los rasgos. Este valor se calcula como:

$$S_g(a_1, a_2) = W_{er}(a_1, a_2) + W_d(a_1, a_2) + W_{net}(a_1, a_2), \quad (8)$$

donde  $W_{er}(a_1, a_2)$  es el peso obtenido por las expresiones regulares que empatan los contextos que contienen tanto a  $a_1$  como a  $a_2$ , ( $W_d(a_1, a_2)$ ) es el peso aportado por la distancia inversamente proporcional y  $s_r(a_1, a_2)$  ( $W_{net}(a_1, a_2)$ ) un valor inversamente proporcional al valor de similitud entre  $a_1$  y  $a_2$ , obtenido a través de las RCLs. Cada peso está normalizado en el rango  $[0, 1]$ . Así,  $S_g$  es menor que 3. A continuación se describen cada uno de estos valores:

Dado  $E$ , el conjunto de todas las expresiones regulares que empatan los contextos de  $a_1$  y  $a_2$ ;  $weight(e)$ , el peso de  $e$  (algunos ejemplos se muestran en la tabla 1); y  $fr(e)$  la frecuencia relativa  $e$  que empata con los contextos, se define  $W_{er}(a_1, a_2) = \sum_{e \in E} weight(e) \cdot fr(e)$ .

$W_d(a_1, a_2)$  es considerado el puntaje máximo para ejemplos positivos. A partir de  $\Delta_M$  (ec. 1) se determina  $W_d(a_1, a_2)$  para lograr un valor normalizado entre  $[0, 1]$ :

$$W_d(a_1, a_2) = \frac{\Delta_M - \bar{\Delta}(a_1, a_2)}{\Delta_M} \quad (9)$$

Finalmente,  $W_{net}(a_1, a_2)$  sigue un cálculo similar a la ecuación 9:

$$W_{net}(a_1, a_2) = \frac{\max_{(x,y) \in \mathcal{B}} \{s_r(\mathcal{Y}, \ )\} - s_r(a_1, a_2)}{\max_{(x,y) \in \mathcal{B}} \{s_r(\mathcal{Y}, \ )\}} \quad (10)$$

En la prueba se usa un corpus compuesto por 26297 oraciones, 11575 términos (incluyendo nombre propios, después de eliminar las palabras cerradas y lematizar el resto). En la tabla 2 se muestra el conjunto de entrenamiento, el cual está compuesto por 15 pares, diez positivos y cinco negativos, que fueron usados para afinar los umbrales. El conjunto de prueba estuvo compuesto por 8 pares de antónimos y 10 pares que tienen algún otro tipo de relación semántica. La tabla 3 muestra algunos ejemplos de la aplicación del método, el veredicto indica si los pares de palabras, positivos y negativos, han sido identificados como antónimos o no antónimos respectivamente, de manera correcta o incorrecta. Es importante notar que solo cuatro pares fueron clasificados incorrectamente, y que el método también descarta hipónimos.

## 4. Discusión

Se considera que la identificación de relaciones de antonimia con el uso combinado de patrones léxico sintácticos y las redes de co-ocurrencia léxica es una línea interesante de investigación. Se ha observado que es conveniente usar una profundidad de tamaño 3 para una RCL, ya que a mayor profundidad, la cantidad de términos involucrados tiende a debilitar la asociación de términos del nodo raíz. Se destaca también la importancia que posee el cálculo de la información mutua como mecanismo de filtrado, ya que decrementa la posibilidad de que dos cualesquiera términos posean una relación de primer orden, y el puntaje  $t$  que define los pesos entre los términos.

El umbral tomado para la determinación de la relación entre un par de palabras relacionadas en este trabajo, está basado en el cálculo obtenido a

Pares de Palabras	Distancia	PLS	RCL	Veredicto
Absoluto-Relativo	0.64814815	0.79333333	0.97665313	Correcto
Compra-Venta	0.59259259	0.16666667	0.47302784	Correcto
Natural-Artificial	0.68518519	0.58	0.9724478	Correcto
Consumidor-Productor	0.62962963	0.56	0.47911833	Correcto
Escasez-Abundancia	1	1	0.36180394	Correcto
General-Particular	0	0.50666667	1	Correcto
Máximo-Mínimo	0.51851852	0.26666667	0.64022622	Correcto
Positivo-Negativo	0.44444444	0.80666667	0	Correcto
Verdad-Mentira	0.72222222	0.66666667	0.9650522	Correcto
Vida-Muerte	0.88888889	0.86666667	0.77392691	Correcto
Corrección-Ajuste	0.75925926	0.2	0.08164153	Correcto
Hombre-Humano	-0.38888889	0.10666667	0.97056265	Correcto
Mercancía-Producto	-0.05555556	0.18666667	0.83236659	Correcto
Moneda-Dinero	0.05555556	0.02	0.87558005	Correcto
Obrero-Trabajador	-0.48148148	0	0.86934455	Correcto

Tabla 2. Conjunto de entrenamiento: los primeros 10 pares son antónimos.

Pares de Palabras	Distancia	PLS	RCL	Veredicto
Bajo-Alto	0.53703704	0.2	0.62601508	Correcto
Activo-Pasivo	0.67037037	0.76666667	0.28741299	Correcto
Grande-Pequeño	0.47407407	0.53333333	0.97317285	Correcto
Oferta-Demanda	0.79259259	0.76666667	0.04350348	Correcto
Pregunta-Respuesta	0.53703704	0.02666667	0.1712587	Incorrecto
Público-Privado	0.48148148	0.53333333	0.75710557	Correcto
Social-Individual	0.27407407	0.63333333	1.0549594	Correcto
Interior-Exterior	0.50740741	0.73333333	0.62427494	Correcto
Confianza-Fe	-2.05555556	0	0.90762761	Correcto
Crédito-Préstamo	0.33333333	0.33333333	0.36107889	Correcto
Cultura-Democracia	0.7037037	0.26666667	0.67444896	Incorrecto
Harina-Trigo	0.6	0.43333333	0.05191415	Correcto
Inversión-Gasto	0.53148148	0.1	0.33990719	Correcto
Miembro-Comunidad	0.90740741	0	0.70562645	Correcto
Pobreza-Problema	0.56851852	0.03333333	1.03045244	Incorrecto
Productor-Benefactor	0.32222222	0.16666667	0.34353248	Correcto
Rasgo-Característica	0.85185185	0	0.72186775	Incorrecto
Semana-Día	0.30185185	0	0.50101508	Correcto

Tabla 3. Conjunto de prueba: los primeros 8 pares son antónimos.

través del conjunto de entrenamiento. Sin embargo, sería deseable obtener dicho umbral, por ejemplo, mediante una relación entre el tamaño del corpus y el número de contextos del par de palabras relacionadas.

Los resultados obtenidos son alentadores y motivan el estudio profundo del uso de la técnica propuesta como un mecanismo para la identificación de antónimos en un texto sin información adicional, como etiquetas o algunas estructuras.

Los resultados apuntan al uso de expresiones regulares como primer mecanismo de restricción para posteriormente filtrar a través del uso de redes de co-ocurrencia léxica. Por último, si existieran pares aún sin categorizar, entonces podría hacerse uso de la distancia promedio esperada para términos en relación de antonimia. Aún es necesario validar los resultados encontrados, en una muestra grande de pares de palabras y conocer si el método propuesto podría aplicarse a dominios diversos.

## Agradecimientos

Agradecemos los comentarios de los árbitros de este trabajo, así también el apoyo parcial recibido por parte del proyecto VIEP III 9-04/ING/G.

## Referencias

1. G. Grefenstette: "Explorations in Automatic Thesaurus Discovery", Kluwer Academic Publishers, Boston Hardbound, ISBN 0-7923-9468-2 July 1994.
2. Church, Kenneth Ward; Gale, William; Hanks, Patrick; Hindle, Donald; Moon, Rosamund: "Lexical Substitutability", In: Atkins, B. T. S.; Zampolli, Antonio (eds.): *Computational Approaches to the Lexicon*. Oxford University Press, pp. 153-180, 1994.
3. Edmonds P.: "Choosing the word most typical in context using a lexical co-occurrence network", *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, pp. 507-509, 1997.
4. Morato J., Marzal M.A., Lloréns J., Moreiro J.: "WordNet Applications", Petr Sojka, Karel Pala, Pavel Smrc, Christiane Fellbaum, Piek Vossen (Eds.): *Proceedings GWC 2004*, pp. 270-278, 2004.
5. Caroline Varaschin Gasperin Vera Lúcia Strube de Lima, "Experiments on Extracting Semantic Relations from Syntactic Relations", *CiCLing 2003*, LNCS 2588, pp. 314-324, 2003.
6. Jiménez-Salazar, H., "A Method of Automatic Detection of Lexical Relationships Using a Raw Corpus", *CiCLing 2003*, LNCS 2588, pp. 325-328, 2003.
7. Hearst, M.: "Automatic acquisition of hyponyms from large text corpora". *Proceedings of the 14th International Conference on Computational Linguistics*, Nantes, France, 1992.
8. Yorick Wilks, Roberta Catizone: "Lexical Tuning". *CiCLing 2002*, pp. 106-125, 2002.
9. Paolo Rosso, Francesco Masulli, Davide Buscaldi, Ferran Pla, Antonio Molina: "Automatic Noun Sense Disambiguation", *CiCLing 2003*: pp. 273-276

10. L. Wanner: "Lexical Functions in Lexicography and Natural Language Processing", John Benjamins Publishing Company, 1996.
11. Hearst, M.: "Automated Discovery of WordNet Relations", in *WordNet and Electronic Lexical Database*, C. Fellbaum (Ed.), The MIT Press, 1999, pp. 131-152.
12. Schwab, D., Lafourcade, M., Prince, V.: "Antonymy and Conceptual Vectors", in *the Proceedings of the 19th Conference on Computational Linguistics*, 2002, pp. 904-910.
13. Cruse, D.: "Lexical Semantics", Cambridge, Cambridge University Press, 1986.
14. Hindle, D.: "Noun Classification from Predicate-Argument Structures", in *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, 1990, pp. 268-275.
15. Sanderson M., Croft B.: "Deriving concept hierarchies from text", In *Proceedings of the 22 a Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 206 - 213, Berkeley, CA, August 1999.

# Some Tests in Text Categorization using Term Selection by DTP

Edgar Moyotl-Hernández, Héctor Jiménez-Salazar

Facultad de Ciencias de la Computación,  
B. Universidad Autónoma de Puebla,  
14 sur y Av. San Claudio. Edif. 135. Ciudad Universitaria,  
Puebla, Pue. 72570. México,  
Tel. (01222) 229 55 00 ext. 7212 Fax (01222) 229 56 72,  
emoyotl@mail.cs.buap.mx, hjimenez@fcfm.buap.mx

**Abstract.** Distance to Transition Point (DTP) has shown good performance in term selection for Text Categorization task. Previous experiment report that DTP behaves well as DF and CHI term selection techniques. In this paper we present the results of using DTP computed in a global and local fashion; considering the whole of categories of training set. The results confirm that performance of DTP globally computed is better than DTP locally computed. The test carried out took into account two classification methods:  $k$ -NN and Rocchio's algorithm; and three well known methods to select terms: DF, CHI and IG.

**Keywords:** distance to transition point, term selection, text categorization.

## 1 Introduction

We are interested in the automated assignment of texts into predefined categories, Text Categorization (TC), which is solved by supervised learning algorithms [10]. Specifically, our interest is in the selection of terms from texts in order to train supervised algorithms. This is an important problem because of the necessity to optimize time and memory space.

Some classification algorithms used in TC are: Rocchio's algorithm,  $k$ -Nearest Neighbor ( $k$ -NN), and Support Vector Machines [6][2][13]. Also there are several methods of term selection [7][12]. Most known term selection techniques assign weights to terms, and those which guess more importance are selected. Three very used criteria to term selection are [7][12]: Document Frequency (DF), is the number of documents in which term occurs;  $\chi^2$  statistic (CHI), measures the lack of independence between the term and the category; and Information Gain (IG), measures the number of bits of information by knowing the presence or absence of a term in a document.

Our approach to term selection is based in the occurrence frequency of a term in the whole of the training collection. It distinguishes a frequency called *Transition Point* (TP) which is used to compute the distance of each term frequency to TP (DTP). Previous experiments [5] have shown that DTP improves

in some aspects other term selection methods as DF and CHI; for example, this due to DTP calculation is an  $O(V)$  algorithm where  $V$  is the number of terms in the training set, and it improves slightly the performance of TC task. Later has motivated to continue the research on DTP.

In this paper we present an experiment varying the calculation of TP in two different ways: *global* TP, using the whole of the documents; and *local* TP, using the documents of each category. The experiment was accomplished on a small texts collection (6 categories with totally around 1,000 documents), using two classification methods:  $k$ -NN and Rocchio.

The following section explains some backgrounds about DTP. Sections 3, 4 and 5 describe the material and methods used in the test accomplished. At the end, we provide the conclusions reached in this work.

## 2 Distance to Transition Point

As we have said our method is based on *Transition Point* (TP). TP is the frequency of a term which splits terms into high frequency terms and low frequency terms. The present test is based on a previous experiment that shown that TP behaves well as a cut on the selected term by the classical method [4]. From the fact that TP indicates the frequency which around it there are key words of the text [11], was calculated a weight per term. Such weight, for term  $t_i$ , is calculated as the inverse of distance of  $t_i$  frequency to TP (DTP): the more closeness of frequency to TP, the more weight for that term. Firstly, let us see how to calculate TP; some details may be found in [5].

Let  $T$  be a text (or a set of texts), and let  $I_1$  be the number of words with frequency 1. TP is defined as [11]:

$$n = (\sqrt{1 + 8I_1} - 1)/2. \quad (1)$$

As we can see, TP calculation requires only scanning the full text in order to find  $I_1$ , which can be done in  $O(V)$ , where  $V$  is the number of terms.

Now, DTP is easily calculated making the difference between each term frequency and TP. We would hope that DTP calculated for each category ports more information and, therefore, TC were better. However, test accomplished shown that DTP on the basis of all categories has better performance. We will show this fact in the following sections.

## 3 Term Selection Methods

In this section, we give a brief introduction on three effective term selection (TS) techniques as they are presented in [12], including one unsupervised method (in the sense that it does not use category information) DF and two supervised method (it uses category information) CHI and IG. These methods assign a score to each individual term and then select the terms that score highest. In the following, let us denote with  $D$  the training documents set,  $N$  the number of



documents in  $D$ ,  $V$  is the number of terms in  $D$ , and let  $\{c_k\}_{k=1}^M$  the categories set. Selection methods used in the test were the following:

**Document Frequency (DF).** Document frequency is the number of documents in which a term  $t_i$  occurs. It is the simplest technique for term selection and easily scales to a large data set with a computation complexity approximately linear in the number  $N$ . It is a simple but effective term selection method for TC [12].

$\chi^2$  **statistic (CHI).** The  $\chi^2$  statistic measures the lack of independence between the term and the category. In the TC, given a two-way contingency table for each term  $t_i$  and category  $c_k$  as represented in Table 1, it is defined to be:

$$CHI(t_i, c_k) = \frac{N(ad - cb)^2}{(a + c)(b + d)(a + b)(c + d)}, \quad (2)$$

where,  $a$ ,  $b$ ,  $c$  and  $d$  indicate the number of documents for each cell in the contingency table (see table 1). We computed  $CHI$  for each category and each term in  $D$ , and then combined the category specific scores of each term into one score:  $CHI(t_i) = \max_{k=1}^M \{CHI(t_i, c_k)\}$ . The computation of  $CHI$  scores has a quadratic complexity, similar to IG [12].

**Information Gain (IG).** Information gain of a term measures the number of bits of information obtained by knowing the presence or absence of a term in a document. The information gain of term  $t_i$  is defined as

$$\begin{aligned} IG(t_i) = & - \sum_{k=1}^M P(c_k) \log P(c_k) \\ & + P(t_i) \sum_{k=1}^M P(c_k|t_i) \log P(c_k|t_i) \\ & + P(\bar{t}_i) \sum_{k=1}^M P(c_k|\bar{t}_i) \log P(c_k|\bar{t}_i) \end{aligned} \quad (3)$$

where, for example,  $P(c_k)$  is the number of documents belonging to the category  $c_k$  divided by the total number of documents in  $D$ ,  $P(t_i)$  is the number of documents without the term  $t_i$  divided by the total number of documents in  $D$ ,  $P(c_k|t_i)$  is the number of documents of category  $c_k$  with the term  $t_i$  divided by the number of documents with  $t_i$ , etc. The computation includes the estimation of the conditional probabilities of a category given a term, and the entropy computations in the definition. The probability estimation has a time complexity of  $O(N)$  and the entropy computations has a time complexity of  $O(M)$  [12].

**Distance to Transition Point (DTP).** DTP measures importance of a term  $t_i$  according to the distance of that term to TP:

$$DTP(t_i) = |TP - \text{frq}(t_i)|, \quad (4)$$

where  $\text{frq}(t_i)$  is the occurrence frequency of  $t_i$  in  $D$  (*global*) or  $D_k$  (*local*), TP is computed on  $D$  (*global*) or  $D_k$  (*local*), and  $D_k$  is the training documents set of  $c_k$ .

Category/Term	$t_i$	$\bar{t}_i$
$c_k$	a	b
$\bar{c}_k$	c	d

**Table 1.** Two-way contingency table.

## 4 Classification Methods

To assess the effectiveness of TS methods we used two classifiers:  $k$ -NN and Rocchio. Both classifiers treat documents as a feature vectors.  $k$ -NN is based on the categories assigned to the  $k$  nearest training documents to the new document. The categories of these neighbors are weighted using the similarity of each neighbor to the new document, where the similarity is measured by the cosine between the document vectors. If one category belongs to multiple neighbors then the sum of the similarity scores of these neighbors is the weight of the category. Rocchio is based on the relevance feedback algorithm originally proposed for information retrieval. It has been extensively used for TC. The basic idea is to construct a prototype vector for each category using training documents. Given a category, the vectors of documents belonging to this category are given a positive weight, and the vectors of remaining documents are given a negative weight. By summing up these positively and negatively weighted vectors, the prototype vector of this category is obtained. To classify a new document, the cosine between the new document and prototype vector is computed.

Both classifiers are *context sensitive* in the sense that no independence is assumed between either terms or categories [12].  $k$ -NN and Rocchio treat a document as a single point in a vector space, thus enabling a better observation on TS.

## 5 Test

The texts used in our experiments are the Spanish news downloaded from the Mexican newspaper *La Jornada* (year 2000). We preprocess the texts removing *stopwords*, punctuation and numbers, and stemming the remaining words by means of a Porter’s stemmer adapted to Spanish. We have used a total of 1,449 documents belonging to six different categories (Culture (C), Sports (S), Economy (E), World (W), Politics (P) and Society & Justice (J)) for training and two test sets (see Table 2). We only managed one label setting (i.e., each document was assigned in only one class).

	Categories	C	S	E	W	P	J
Training data	No. of documents	104	114	107	127	93	91
	No. of terms	7,131	4,686	3,807	5,860	4,796	4,412
Test data set1	No. of documents	58	57	69	78	89	56
	No. of terms	5,228	3,285	3,235	4,611	4,647	3,774
Test data set2	No. of documents	83	65	61	51	90	56
	No. of terms	6,349	3,799	2,793	3,611	4,879	3,778

**Table 2.** Training and testing data.

To evaluate the effectiveness of category assignments to documents by classifier, the standard precision, recall and  $F_1$  measure was used here. Precision is defined to be the number of categories correctly assigned divided by total number of categories assigned. Recall is the number of categories correctly assigned divided by the total number of categories that should be assigned. The  $F_1$  measure combines precision ( $P$ ) and recall ( $R$ ) with an equal weight in the following form  $F_1 = 2RP/R + P$ . These scores can be computed for the binary decisions on each individual category first and then be averaged over categories. Or they can be computed globally over all the binary decisions. The former way is called *macroaveraging* and the latter *microaveraging*. We have evaluated microaveraging  $F_1$ , since it is almost preferred to macroaveraging [10].

Percent of terms	Number of terms	$k$ -NN					Rocchio				
		DF	CHI	IG	$DTP_g$	$DTP_l$	DF	CHI	IG	$DTP_g$	$DTP_l$
1	142	0.610	0.702	0.710	0.680	0.669	0.616	0.705	0.718	0.690	0.659
3	426	0.696	0.748	0.764	0.763	0.743	0.702	0.738	0.754	0.750	0.764
5	710	0.750	0.776	0.781	0.765	0.744	0.750	0.756	0.761	0.775	0.768
10	1,419	0.791	0.801	0.793	0.791	0.801	0.777	0.781	0.776	0.808	0.783
15	2,129	0.795	0.797	0.802	0.807	0.792	0.777	0.782	0.783	0.812	0.787
20	2,838	0.799	0.798	0.806	0.807	0.802	0.782	0.786	0.790	0.820	0.793
25	3,548	0.801	0.799	0.807	0.809	0.799	0.788	0.795	0.791	0.819	0.792
50	7,095	0.799	0.786	0.803	0.809	0.802	0.795	0.798	0.803	0.822	0.798

**Table 3.**  $F_1$  average values for  $k$ -NN and Rocchio on test set1 and set2.

We have performed our TS experiments first with the standar  $k$ -NN classifier (with  $k = 30$ ), and subsequently with the Rocchio classifier (with  $\beta = 16$  y  $\alpha = 4$  as was suggested in [3]). In these experiments we have compared three baseline term selection functions, i.e. DF, CHI, IG and DTP (global denoted with  $DTP_g$  and local denoted with  $DTP_l$ ). Table 3 lists the  $F_1$  values for  $k$ -NN and Rocchio with different TS techniques at different percent of terms (the number of different terms in the training set is 14,190).

We calculate the correlation coefficients among the statistics with the formula:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (5)$$

$$(6)$$

where  $X, Y$  are random variables,  $Cov(X, Y)$  is their covariance, and  $\sigma_X \cdot \sigma_Y$  are their standard deviations respectively. The results are given in Table 4:

	DF	CHI	IG	$DTP_g$	$DTP_l$
DF	1	0.358	0.573	0.255	0.443
CHI	0.358	1	0.919	0.167	0.283
IG	0.573	0.919	1	0.215	0.361
$DTP_g$	0.255	0.215	0.167	1	0.1
$DTP_l$	0.443	0.361	0.283	0.1	1

**Table 4.** Correlation Coefficients of Statistics.

Yang and Pedersen [12] conducted a comparative study on several TS methods, and found: First, CHI and IG are most effective method to reduce the dimensionality of the term space. Second, DF performance, similarly, was shown scoring in favor of common terms over rare terms. Third, strong correlations between DF, CHI and IG values of a term are general rather than corpus-dependent. The strong correlations means that common terms are often informative, and reciprocally. Such correlation are displayed in table 4. Basically, these results seem to state that the most valuable terms for TC are those with medium frequency in the training set, i.e., those around of the TP. Another interesting conclusion in [12] is that using category information for TS does not seem to be crucial for excellent performance. DTP does not use category information present in the training set, but has a performance similar to CHI and IG.

## 6 Conclusions

We have showed some features of DTP behavior. First, DTP performance is similar to other term selection techniques, varying two classification methods, Rocchio and  $k$ -NN. Second, DTP calculation is in the same complexity class that DF, the better technique in execution time. Third, DTP is independent of category, it may be computed globally, and is better than DTP locally computed.

Although, the difference between DTP and other term selection techniques is not significative in TC, this result encourages to carry out further experiments to know how much our proposal enhances TC task with respect to reported results in the literature.

## References

1. Booth, A.: "A Law of Occurrences for Words of Low Frequency", *Information and control*, 10(4) pp 386-93, 1967.
2. Lam, W. & Ho, C.: "Using Generalized Instance Set for Automatic Texts Categorization", in *Proc. of 21th Ann. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp 81-89, 1998.
3. Lewis, D.D.: "Training Algorithms for Linear Text Classifiers", *SIGIR96*, 1996.
4. Moyotl, E. & Jiménez, H.: "An Analysis on Frequency of Terms for Text Categorization", to be published in *Proc. of SEPLN*, 2004.
5. Moyotl, E. & Jiménez, H.: "A Novel Method to Select Terms for Text Categorization", sent to *Iberamia 2004*.
6. Rocchio, J.J.: "Relevance Feedback in Information Retrieval". In G. Salton, (Ed.), *The SMART retrieval system: experiments in automatic document processing*, Prentice-Hall, 1971.
7. Rogati, M. & Yang, Y.: "High-Performing Feature Selection for Text Classification", *CIKM'02*, ACM, 2002.
8. Salton, G. & McGill, M.: *Introduction to Modern Information Retrieval*, 1983.
9. Salton, G., Wong, & Yang, C.S.: "A Vector Space Model for Automatic Indexing", *Information Retrieval and Language Processing*, pp 613-620, 1975.
10. Sebastiani, F.: "Machine Learning in Automated Text Categorization", *ACM Computing Surveys*, Vol. 34(1), pp 1-47, 2002.
11. Urbizagástegui-Alvarado, R.: "Las posibilidades de la ley de Zipf en la indización automática", *Reporte de la Universidad de California Riverside*, 1999.
12. Yang, Y. & Pedersen, P.: "A Comparative Study on Feature Selection in Text Categorization", in *Proc. of 14th Int. Conf. on Machine Learning*, pp 412-420, 1997.
13. Yang, Y. & Liu, X.: "A Re-examination of Text Categorization Methods", *ACM SIGIR*, pp 42-49, 1999.
14. Xue, D. & Sun, M: "A Study on Feature Weighting in Chinese Text Categorization", *Lecture Notes in Computer Science*, A. Gelbukh (Ed.), Vol. 2588, Springer, pp 592-601, 2003.
15. Zipf, G.K.: *Human Behaviour and the Principle of Least Effort*, Addison-Wesley, 1949.

# Herramientas de análisis para el Corpus Lingüístico en Ingeniería

Gabriel Garduño-Torres, Gerardo Sierra-Martínez,  
Alfonso Medina-Urrea

Grupo de Ingeniería Lingüística  
Instituto de Ingeniería, UNAM.

e-mail : {ggardunot, gsierram, amedinau}@iingen.unam.mx  
3<sup>er</sup> cubículo de la Torre de Ingeniería, Circuito Interior, Ciudad Universitaria,  
Del. Coyoacán, CP 04510, Distrito Federal, México. Apartado Postal 70-472

**Abstract:** This paper describes the analysis tools of what will be the first Corpus on Engineering in Spanish. The implementation of these will provide the means for engineers, linguists and linguistic engineers, among others, to explore the corpus according to their research needs. The results generated by the system proposed will have several applications; for instance, in terminology extraction and assistance for the writing and style of technical documents.

**Resumen:** En este documento se describen las herramientas de análisis del primer corpus en ingeniería en español. Con su implementación se proporcionará un instrumento de apoyo a los especialistas de las áreas de ingeniería, lingüística, ingeniería lingüística, entre otros, en sus labores de investigación. Los resultados generados por el sistema tendrán diversas aplicaciones; por ejemplo, en la extracción de terminología y en la asistencia para redactar documentos técnicos.

**Palabras Clave:** Corpus lingüísticos, ingeniería, concordancias, colocaciones, medidas de asociación.

## 1. Introducción.

Existen dos tipos de corpus lingüísticos: los generales y los de especialidad. Si entendemos corpus como un conjunto electrónico recopilado de textos hablados o escritos cuya finalidad es algún tipo de análisis lingüístico, los primeros se encargan de recoger todo tipo de géneros y son útiles para describir la lengua de una comunidad (para el español, por ejemplo, consúltese el CREA[6]). Los corpus especializados, por otro lado, recogen material lingüístico que puede aportar información para la descripción de un área de especialidad o tema en particular (véase, por ejemplo, el Corpus Técnico del IULA[7]).

El contar con un corpus de especialidad en un área determinada permite analizar y observar cuestiones relacionadas con el conocimiento de los expertos en esa especialidad; por ejemplo, podemos mencionar cómo se expresa el especialista y cómo escribe. También se pueden examinar rasgos textuales más refinados como el marcaje

tipográfico, que resalta los elementos temáticos más importantes, las variedades estilísticas, etc. Además, al tratarse de documentos lingüísticos los corpus sirven para analizar la estructura léxica, morfológica y gramatical de la lengua representada. Es decir, un corpus lingüístico en un área de especialidad sirve para muchas cosas, su uso se extiende no sólo a los investigadores en el área, sino también a lingüistas, ingenieros lingüistas y público en general. En el área de ingeniería no existe un corpus especializado, por lo menos no en el mundo de habla hispana. De aquí nace la idea de elaborar el primer *Corpus Lingüístico en Ingeniería*, en México y en español, el cual describimos a continuación.

## 2. El Corpus Lingüístico en Ingeniería.

Actualmente el Grupo de Ingeniería Lingüística (GIL) del Instituto de Ingeniería, UNAM, y gracias al patrocinio de CONACYT, desarrolla el primer corpus lingüístico en ingeniería. Los objetivos generales del proyecto son:

1. Elaborar, desarrollar y mantener un corpus lingüístico que contenga textos selectos en el área de ingeniería, los cuales deben estar debidamente codificados y organizados.
2. Diseñar herramientas de análisis adecuadas para manejar el corpus y explotar su utilidad en el desarrollo de diversas investigaciones en las áreas de ingeniería, ingeniería lingüística y lingüística.

De esta manera, se espera tener dos productos principales: una colección balanceada de textos representativos de la ingeniería en lengua española; y las herramientas computacionales necesarias para consultar y analizar dicha colección.

Por otra parte, el corpus será *multipropósito*, es decir, podrá satisfacer distintas necesidades dependiendo de la finalidad de su consulta. Por un lado, el corpus contribuirá a construir el consenso necesario para el establecimiento de terminologías robustas y no redundantes de las diversas áreas temáticas de la ingeniería; por otro lado, ayudará en la redacción de este tipo de documentos, libros o artículos proporcionando modelos de escritura o bien, ilustrando a través de un panorama estilístico diversos aspectos de la exposición de temas. Los ingenieros lingüistas podrán elaborar herramientas vinculadas con el procesamiento del lenguaje natural; y los lingüistas tendrán una herramienta de apoyo para investigaciones lexicográficas, terminográficas, gramaticales, pragmáticas y semánticas.

Asimismo, el corpus en ingeniería es uno de los proyectos eje del GIL porque redundará en distintas líneas de investigación, las cuales son:

1. Identificación de patrones recurrentes para la extracción automática de contextos definitorios en textos especializados [1].

2. Elaboración semiautomática de diccionarios especializados tanto de consulta onomasiológica[2] (del concepto al término) como semasiológica (del término al concepto).
3. Identificación de paradigmas semánticos y clusters[3].
4. Extracción terminológica a partir de documentos actuales y confiables en el área de ingeniería.
5. Elaboración de herramientas para facilitar la redacción de artículos e informes técnicos al ingeniero (generación automática de documentos).

Para garantizar la conservación de los detalles visuales de los documentos, se mantendrán los formatos tipográficos y de estilo, así como la estructura textual que los caracterice. Para esto, se marcarán los documentos con etiquetas estructurales XML.

Se ha observado que en la situación pragmática de introducir nuevos términos y definiciones en los textos de especialidad, los autores hacen uso de patrones recurrentes en los que se involucran marcas tipográficas (negritas, cursivas, subrayado, espaciado, etc.) con el fin de resaltar el nuevo término o la definición. Por eso, es importante conservar la tipografía marcada en el documento original, para mostrárselos al usuario, en el momento de su consulta y para utilizarlos en búsquedas automáticas de patrones. Por ejemplo, considérese el siguiente contexto definitorio:

Según G. Malagón (1996) un hospital se define como: **“una parte integrante de la organización médica, cuya función es la de proporcionar a la población...”**

Según la etiquetas XML que hemos definido, este contexto se anotaría de la siguiente manera:

```
<doc><p>Según <name>G. Malagón</name> (<num>1996</num>) un hospital se define como: “<b>una parte integrante de la organización médica, cuya función es la de proporcionar a la población...</b></p></doc>
```

Como podemos observar en el ejemplo, se tienen las etiquetas <doc>...</doc> que serán las “etiquetas madre” o “raíz” de la cuales dependerán todas las existentes dentro de los documentos. Las etiquetas <p>...</p> nos indican que todo lo que esté dentro de ellas es un párrafo; las etiquetas <name>...</name> contienen dentro de ellas algún nombre propio; las etiquetas <num>...</num> nos indican la existencia de un número; y finalmente las etiquetas <b>...</b> nos dicen que todo el texto ahí presente se encuentra en negritas.

También nos interesa marcar las categorías a las que pertenecen las palabras en un corpus, por ejemplo, las típicas categorías POS (*part of speech*). La marcación con estas es necesaria para determinar la estructura interna de las frases verbales, nominales, preposicionales y adverbiales, con las cuales se pueden construir los términos, las definiciones así como otras estructuras de interés lingüístico.

Con base en el conjunto de etiquetas EAGLES[8], que contiene códigos especiales para atributos asociados a cada categoría POS, podemos enfocarnos, por ejemplo, en el



primer y segundo código de dichas etiquetas (categoría gramatical y el tipo POS) para etiquetar el párrafo de arriba que quedaría de la siguiente manera:

```
<div><p>Según_SP <name>G. Malagón</name> (_FC <num>1996</num> )_FC un_TI  
hospital_NC se_PP define_VM como_CS :_FC “_FC <b>una_TI  
parte_NC...</b></p></div>
```

Naturalmente, para la aplicación de las herramientas de análisis que describiremos a continuación, es necesario contemplar una etapa de pre-procesamiento del corpus. Típicamente, en esta etapa se anotan las partes del texto que queremos que se distingan de las demás, como son los nombres propios, números, fechas, siglas, abreviaturas, etc. Los programas de explotación del corpus tratan de manera especial las etiquetas XML que marcan estas unidades textuales.

### 3. Herramientas de análisis

Una vez etiquetado el corpus se pueden aplicar las herramientas de análisis. Dada la variedad potencial de usuarios, dichas herramientas estarán disponibles en Internet. Según nuestros requerimientos, las características óptimas del ambiente de desarrollo del sistema son las siguientes:

1. Plataforma UNIX para asegurar el tiempo mínimo del cálculo de los procesos y el control de los accesos de los usuarios.
2. Lenguaje de programación Java, porque es portable, robusto y permite tomar las medidas de seguridad necesarias dentro de la implementación del software.
3. Lenguaje de programación Python y plataforma NLTK, porque el primero permite un extenso tratamiento de cadenas de texto y el segundo es un ambiente específico para el procesamiento de lenguaje natural.

Así, el empleo de las herramientas necesitará, por un lado, que el usuario cuente con un navegador de Web y, por el otro, que se haya registrado previamente mediante un formulario para que tenga acceso de forma gratuita al sistema.

Las herramientas le darán al usuario la opción de elegir una lista filtro (*stop list*) por defecto, o proporcionar una propia, para que los resultados generados por el sistema no muestren las palabras juzgadas irrelevantes en análisis particulares. Además, el usuario también podrá modificar la lista filtro para adecuarla a sus necesidades. A continuación describimos las herramientas mínimas de explotación del corpus.

#### 3.1 Estadísticas

Con el corpus etiquetado correctamente el programa podrá llevar a cabo estadísticas que generen resultados más confiables, por ejemplo, podrá contar cifras, en cuyo caso considerará los números como una sola unidad (275,376.25). Además, los nombres propios, siglas, etc. conservarán las letras mayúsculas que típicamente se utilizan para representarlos.

Se contarán las palabras por cada documento, por cada área temática y en el corpus completo. Asimismo se llevará un registro del número de documentos existentes en el corpus entero y en cada área temática.

El usuario podrá generar listas de palabras (Figura 1). Una lista de palabras es un listado con los *types* o vocablos del corpus (unigramas). A cada *type* le acompañará su frecuencia absoluta (número de veces que se repite dicha palabra), su frecuencia relativa (número de veces que ocurre la palabra en relación con el total de palabras en el corpus), la entropía (contenido de información) y sus etiquetas POS. Las frecuencias y las entropías serán calculadas y desplegadas para cada documento, para cada área temática y para el corpus en general. El usuario podrá ordenar los elementos de la lista por el rango de cualquiera de estos datos; por ejemplo, su frecuencia relativa en el área temática (ascendente o descendente) o por orden alfabético normal o inverso (se ordena a partir de la primera letra del lado derecho).

Word	Freq.	%
DE	22,720	7.69
LA	11,522	3.90
EN	7,819	2.65
Y	7,206	2.44

Fig 1. Lista de palabras

### 3.2 Concordancias

Las concordancias también son listas de palabras pero con sus contextos. Para que se extraigan las concordancias de una o más palabras, el usuario deberá introducir al sistema una palabra o expresión regular. Además, a partir de la lista de palabras mencionada arriba, el usuario podrá desplegar las concordancias de cualquier palabra con sólo seleccionarla en el listado. Junto con las concordancias vendrán datos tales como el código identificador del documento, los datos bibliográficos de la fuente, área temática y la posición de la palabra en el corpus.

A menos de que el usuario seleccione lo contrario, las concordancias traerán el contexto sin modificaciones; esto es, con mayúsculas y minúsculas y con los signos de puntuación. Para facilitar la visualización en pantalla, la palabra que se analiza vendrá en una columna centrada, y del lado izquierdo la parte del texto que le antecede, y del lado derecho la que le precede. Al momento de traer el texto a pantalla el usuario podrá elegir si desea que se extraiga con todo y etiquetas de marcaje. Esto dará la posibilidad de visualizar los rasgos tipográficos que contengan los textos.

El tamaño de la ventana (cantidad de texto que se trae acompañando a la palabra que se analiza) podrá ser modificado por el usuario según sus necesidades: por cantidad de palabras o caracteres (tanto a la izquierda como a la derecha de la palabra analizada) y

con opción de ajustarla a un párrafo. En la figura 2 se muestra un ejemplo de concordancia:

Los CFD se han convertido en una herramienta de diseño "madura" en lo relativo a los de  
tempo real, ha demostrado ser una herramienta ideal para el desarrollo de redes de obser  
nsiderarse actualmente como una herramienta más de diseño en las formas hidrodinami

**Fig 2. Concordancias de la palabra herramienta**

Las búsquedas permitirán la utilización de operadores booleanos (AND, OR y NOT), expresiones regulares y caracteres comodines (\*, + y ?). Finalmente, las concordancias podrán extraerse de un área temática específica o del corpus en general; y se podrán mostrar por orden de aparición o por orden alfabético de la palabra en cuestión.

### 3.3 Medidas de asociación

Además de las estadísticas de unigramas descritas en 3.1, el corpus contará con una herramienta para generar algunas estadísticas de bigramas[4] que nos permitirán determinar qué pares de palabras gráficas están estadísticamente más asociados entre sí. Esto es importante porque permitirá la detección de términos de más de una palabra gráfica y de las colocaciones típicas del corpus o de las áreas temáticas de la ingeniería por separado.

El conjunto mínimo de medidas de asociación entre pares de palabras gráficas que calcularemos contendrá las estadísticas conocidas como información mutua, razón de semejanza (*likelihood ratio*) y la prueba de independencia de  $\chi^2$ [4][5]. La herramienta generará la lista de pares de palabras más asociadas entre sí, según los valores del corpus, del área temática y del documento. Aparecerán ordenados por rango de asociación según la columna de la estadística seleccionada por el usuario. Naturalmente, aquí también se proporcionaría una lista filtro para eliminar las palabras gráficas con función gramatical. Obviamente, también aquí se le permitirá al usuario modificar esta lista o proporcionar la suya propia.

Si distinguimos entre unigramas sencillos y complejos (siendo éstos últimos secuencias de palabras gráficas), una característica interesante de esta herramienta será la posibilidad de generar listas de bigramas constituidos por pares de unigramas complejos (en efecto, bigramas, trigramas, etc.). En otras palabras, el usuario podrá seleccionar una serie de caracteres con espacios intermedios (cadena de palabras gráficas) para ser considerada como unigrama apto de asociarse a otro unigrama. De esta manera, las medidas de asociación entre éstos podrán calcularse y mostrarse en la lista de los bigramas más prominentes del corpus, del área temática y del documento.

### 3.4 Colocaciones

Con las concordancias, las estadísticas y las medidas de asociación los usuarios podrán determinar las colocaciones pertinentes a sus investigaciones en el corpus. No es ninguna novedad que en los textos las palabras tiendan a ocurrir acompañadas por otras o que normalmente ocurran en conjuntos. Una colocación se define como la ocurrencia de dos o más palabras que se encuentran cercanas en un texto y que tienden a ocurrir cercanas en ciertos contextos. En este sentido, se considera una colocación como la combinación frecuente de palabras, así como la combinación en la que una palabra requiere la presencia de otra para expresar un sentido dado. Con las colocaciones podremos evaluar las co-ocurrencias (y su frecuencia) de dos o más palabras en un contexto dado. Se podrá evaluar la cercanía de las partes de una colocación. Se puede tomar una de las palabras y asignarle distancia cero. Las palabras asociadas podrán detectarse al examinar las que la rodean, y sus distancias serán el número de palabras que se encuentran entre la primera y la segunda menos uno. Las distancias serán negativas cuando la segunda esté a la izquierda de la primera y positivas cuando a la derecha. Las colocaciones se presentarán de tal forma que en el centro se muestre la palabra analizada y a su lado izquierdo y derecho las palabras contiguas. Un ejemplo de cómo se visualizará una colocación se muestra a continuación:

Los correctores ortográficos son una **herramienta** computacional muy útil en la redacción de documentos

-5	-4	-3	-2	-1	nodo	+1	+2	+3	+4	+5
14	16	13	35	11	48	43	7	9	12	9

En este ejemplo se muestran en una concordancia las palabras gráficas contiguas al nodo “herramienta”. La distancia de cada palabra con respecto al nodo aparece en el siguiente renglón. Además las frecuencias de ocurrencia de cada palabra con respecto al nodo (en color azul) nos indican que muy probablemente la palabra “herramienta” esté directamente adherida a la palabra “computacional”, porque tienen una frecuencia muy alta y similar entre ellas. De aquí se podría inferir que “herramienta computacional” es un candidato a término del documento examinado.

## 4. Conclusiones

En este trabajo se describieron las herramientas de análisis que estarán disponibles en Internet para explorar el Corpus Lingüístico en Ingeniería. En esencia, se trata de instrumentos para generar concordancias y proporcionar datos estadísticos de unigramas y bigramas que permitan a los usuarios hacer inferencias sobre las estructuras lingüísticas y pragmáticas presentes en los documentos del corpus y determinar las colocaciones típicas de cada documento, cada área temática y del corpus entero.

Esperamos que el desarrollo de estas herramientas beneficie a la comunidad de investigadores en sus labores cotidianas y al público en general. Entre sus múltiples propósitos, permitirán la explotación del corpus para investigaciones terminográficas y servirá de apoyo para aquellos que encuentran dificultad en la redacción de sus

artículos ya que les proporcionará modelos de escritura y estilo. Actualmente hemos obtenido los siguientes resultados en la elaboración del corpus y sus herramientas:

1. Compilación de un corpus de textos electrónicos con más de 500,000 palabras, en las distintas áreas de ingeniería.
2. Desarrollo de una base de datos bibliográfica para llevar control de los documentos contenidos en el corpus y asegurar el balance de las áreas temáticas de la ingeniería.
3. Creación de manuales que describen los procesos de digitalización y etiquetado.
4. Resolución de problemas derivados del empleo del reconocimiento óptico de caracteres.
5. Implementación de una metodología eficaz para guardar y manejar documentos.
6. Desarrollo de una herramienta que ayudará en el marcaje semiautomático de los documentos, elaborada por medio de macros en Word.
7. Implementación de un programa en Python que ayudará en la transformación de los documentos que tienen un formato HTML a nuestro formato XML.
8. Diseño e implementación de las herramientas de análisis del corpus.

## Referencias

- [1] Alarcón Martínez, Rodrigo, “Análisis lingüístico de contextos definitorios en textos de especialidad”, Tesis de licenciatura, Ciudad Universitaria, UNAM, 2003
- [2] Castillo Hernández, Gabriel, “Algoritmo revisado para la extracción automática de agrupamientos semánticos”, Tesis de maestría, Ciudad Universitaria, UNAM, 2002
- [3] Sierra Martínez, Gerardo , “Design of a concept-oriented tool for terminology”, Tesis de doctorado, University of Manchester, August, 1999
- [4] Manning, Christopher D. and Schütze Hinrich, “Foundations of statistical natural language processing”, Fifth printing, The MIT Press, Massachusetts Institute of Technology, London 2002
- [5] Oakes, Michael P., “Statistics for Corpus Linguistics”, Edinburg University Press, Edinburg 1988
- [6] <http://www.rae.es/>
- [7] <http://brangaene.upf.es/bwananet/indexes.htm>
- [8] <http://www.lsi.upc.es/~nlp/tools/parole-sp.html>

# Uso del Contexto para la Búsqueda de Respuestas en Español

**Manuel Pérez-Coutiño, Manuel Montes-y-Gómez, Aurelio López-López**

Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)  
Luis Enrique Erro No. 1, Sta Ma Tonantzintla, 72840, Puebla, Pue., México.  
{mapco, mmontesg, allopez}@inaoep.mx

**Resumen.** Actualmente, la creciente cantidad de información disponible en medios electrónicos representa un reto para la evolución de los mecanismos de acceso y recuperación de la misma. Recientemente los sistemas de Búsqueda de Respuestas (BR) han reaparecido como una alternativa viable a este reto. A pesar su potencial el desarrollo de este tipo de sistemas para el idioma Español aún es escaso. En este documento se propone realizar la tarea de BR mediante el uso de la anotación predictiva del contexto léxico-sintáctico de las entidades nombradas que ocurren en cada documento de una colección fuente en español. La anotación de las entidades nombradas y su contexto representan cada documento como un conjunto de instancias de los conceptos en una ontología de nivel superior. El proceso de búsqueda de respuestas candidatas se realiza a partir de dicha representación, evaluando cada candidata contra las entidades nombradas y el contexto de la pregunta. La respuesta se selecciona a partir de dicha evaluación. Este artículo presenta la metodología de solución propuesta y los avances realizados durante el primer año de investigación.

**Abstract.** Nowadays, the growing amount of information available in electronic media represents a challenge in the evolution of machinery for information access and retrieval. Recently, Question Answering (QA) systems have reappeared as an alternative to such challenge. Despite of its potential, few systems of this kind are developed for Spanish language. This paper proposes to perform QA task using predictive annotation of lexical-syntactic context of named entities occurring in each document within a source collection in Spanish. Named entities annotation and their context represent each document as a set of instances from concepts in an upper level ontology. The process of searching for candidate answers start from such representation, evaluating each candidate against question's named entities and their context. The answer is selected based on the previous evaluation. The paper presents the solving methodology as well as the advances reached during the first year of research.

## 1 Introducción

Durante los últimos años hemos presenciado un crecimiento continuo y exponencial de la información en medios electrónicos tanto en Internet como en colecciones especializadas de recursos, muchos de los cuales se encuentran en forma textual.

Tradicionalmente los usuarios acceden a las fuentes textuales de información –no estructurada ó semiestructurada– mediante sistemas de recuperación de documentos que consisten en procesos capaces de identificar documentos relacionados a un

conjunto de términos o palabras clave dados por el usuario y que tienen por objetivo reflejar su necesidad de información. Sin embargo el uso de estos sistemas requiere de un gran esfuerzo adicional por parte del usuario para filtrar y analizar la lista de documentos obtenida, por lo que este tipo de sistemas o máquinas de búsqueda son incapaces de proporcionar una respuesta concisa a una necesidad específica de información [4].

La alternativa para responder concretamente a preguntas concisas son los sistemas de Búsqueda de Respuestas (BR), capaces de responder a preguntas realizadas por los usuarios en lenguaje natural. Podemos decir que la investigación en sistemas de búsqueda de respuestas se ha incrementado a partir de la introducción de un foro para su evaluación como parte de la Conferencia TREC<sup>1</sup> en 1999 (limitada al lenguaje inglés), y más recientemente en sistemas de Búsqueda de Respuestas Multilingüe [5], siendo en el año 2003 la primera ocasión que se incluyó la evaluación de sistemas de BR como parte del CLEF<sup>2</sup> y donde sólo se presentó un sistema de BR para tratar información en el lenguaje Español [10].

En este artículo se presenta una propuesta para la Búsqueda de Respuestas en Español con base en el uso de la anotación anticipada del contexto al nivel léxico y al nivel sintáctico. En los experimentos preliminares se ha probado el desempeño del uso de contextos léxicos en un banco de pruebas estándar.

## 2 Estado del Arte

Los sistemas actuales de BR afrontan su tarea desde la perspectiva del usuario casual. Es decir, se enfocan en responder preguntas simples sobre hechos concretos a partir de una colección de documentos donde la respuesta se encuentra en forma explícita en un sólo documento. Estas preguntas generalmente pueden responderse con palabras o frases que denotan el nombre de una persona, de un lugar, una fecha, etc. Sin embargo, los sistemas de BR del futuro permitirán resolver preguntas más complejas a partir de la fusión de la información contenida en varios documentos.

Los sistemas de BR típicamente consideran los siguientes procesos: (i) el análisis de la pregunta, (ii) la recuperación de documentos relacionados; (iii) la selección de pasajes relevantes, y (iv) la extracción de fragmentos respuesta. Los sistemas de BR existentes utilizan diferentes técnicas para el tratamiento tanto de las preguntas como de los documentos fuente utilizados para realizar dichos procesos. Uno de los aspectos que ha demostrado mayor efectividad [3,8,9,10] es el uso de reconocedores de entidades nombradas en diferentes niveles del proceso de BR. Una entidad nombrada (EN) es una palabra, o un sintagma que denota un objeto que puede caer en una de las siguientes categorías generales: persona, organización, lugar, fecha, cantidad.

Hablando de forma general, el uso de las EN en sistemas de BR comienza a partir del análisis de la pregunta, al asociar a la pregunta en turno la clase semántica esperada como respuesta. Es decir, dada una pregunta determinar si esta requiere como respuesta una EN de clase persona, fecha, etc. Entonces el proceso de extracción de fragmentos respuesta se realiza con base en la ocurrencia de EN de la

---

<sup>1</sup> TREC (Text Retrieval Conference), <http://trec.nist.gov/>

<sup>2</sup> CLEF (Cross Language Evaluation Forum), <http://clef-qa.itc.it/>

clase semántica esperada como respuesta dentro del fragmento de texto analizado [10]. Otras aproximaciones [3,9] utilizan la identificación de EN para establecer tripletas semánticas formadas por una entidad, el rol semántico que dicha entidad desempeña y el término con el que dicha entidad mantiene la relación.

En contraparte a la identificación de EN en tiempo de búsqueda, Prager [8] ha presentado una aproximación conocida como “Anotación Predictiva”. Dicha aproximación recae en tres componentes: Anotación predictiva, análisis de la pregunta y selección de la respuesta. La anotación predictiva consiste en analizar los documentos en la colección de entrada en busca de palabras que se cree puedan ser respuestas a posibles preguntas. Entonces el sistema les asigna etiquetas que indican el tipo de preguntas que pueden responder. Las etiquetas incluyen lugares, personas, duración, día y longitud. El análisis de la pregunta consiste en utilizar una variedad de tipos de pregunta estándar en los cuales se reemplazan ciertas palabras por las etiquetas adecuadas. Por ejemplo, la pregunta “*How tall is the Matterhorn*” será transformada a “*LENGTH\$ is Matterhorn*”, donde *LENGTH\$* es la etiqueta para denotar distancia. El sistema utiliza un algoritmo para asignar relevancia a los pasajes recuperados y así seleccionar la mejor respuesta.

### 3 Uso del Contexto

Este trabajo se apoya en las investigaciones de Prager [8] sin embargo, además de anotar las entidades nombradas que se encuentran en cada documento y la clase semántica correspondiente, dicha información se utiliza como base para la identificación del contexto léxico-sintáctico de cada entidad nombrada, extraer dicha información y generar la representación de cada documento de la colección documental para su posterior indexado. Dicho índice será entonces utilizado para realizar la extracción de respuestas candidatas a partir de: a) la comparación de la clase semántica esperada como respuesta del análisis de la pregunta, b) las entidades y el contexto de la pregunta, c) el uso de conocimiento externo. Finalmente se realizará la tarea de selección de la respuesta con base en la información recuperada, la similitud de las entidades y los contextos tanto de la pregunta como de las respuestas candidatas. La figura 1 muestra el esquema de la arquitectura esperada al término de esta investigación.

La versión preliminar del modelo de documento así como de la selección de respuestas utiliza un contexto definido a partir de características léxicas identificadas con un etiquetador de partes de la oración y, un reconocedor y clasificador de entidades nombradas. En los experimentos preliminares se utilizó la herramienta MACO [2] para obtener dicha información.

#### 3.1 Modelo de Documento

El objetivo de modelar y representar los documentos fuente para sistemas de BR es proveer un conjunto de recursos preprocesados que contengan información valiosa para facilitar las etapas de recuperación de respuestas candidatas, así como la de selección de la respuesta. Una característica importante del modelo propuesto es que provee un formato unificado para las fuentes de información, como se menciona en [1] “...también es necesario que las fuentes de información sean más heterogéneas y de mayor tamaño...”. Al desarrollar un modelo de documentos es posible que varias



fuentes puedan ser expresadas en un formato estandarizado, o al menos que la transformación y el mapeo entre fuentes de información equivalentes sea viable.

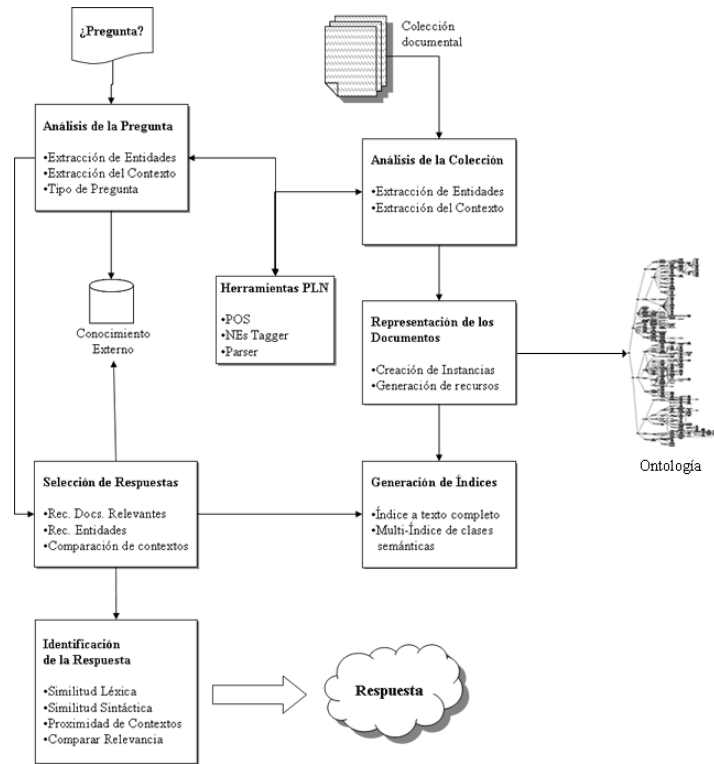


Figura 1. Esquema de la arquitectura esperada al término de esta investigación.

La figura 2 ilustra el modelo propuesto con base en los conceptos de SUMO [7]. El modelo de documento en su nivel conceptual considera un documento como un conjunto de objetos textuales cuyo contenido se refiere a diferentes entidades nombradas aún cuando cada documento se puede enfocar en uno o varios tópicos principales. El modelo supone que las entidades nombradas están fuertemente relacionadas a su contexto léxico, especialmente a sustantivos (que generalmente representan los tópicos más importantes) y verbos (las acciones asociadas a las entidades y a los tópicos del mismo contexto). De esta forma, un documento se puede ver como un conjunto de entidades y sus contextos. Más aún, cada entidad nombrada puede ser refinada por medio de ontologías [6] haciendo uso de la información en el contexto y/o de conocimiento externo como tesauros u otras ontologías. Dicho refinamiento idealmente debe realizarse siguiendo reglas o axiomas preestablecidos que permitan tanto la generalización como la inferencia a partir de una entidad (que es en sí una instancia de un concepto en la ontología), por lo cual el uso de una ontología de alto nivel ya existente resulta ser la opción más adecuada, en vez de desarrollar una desde cero.

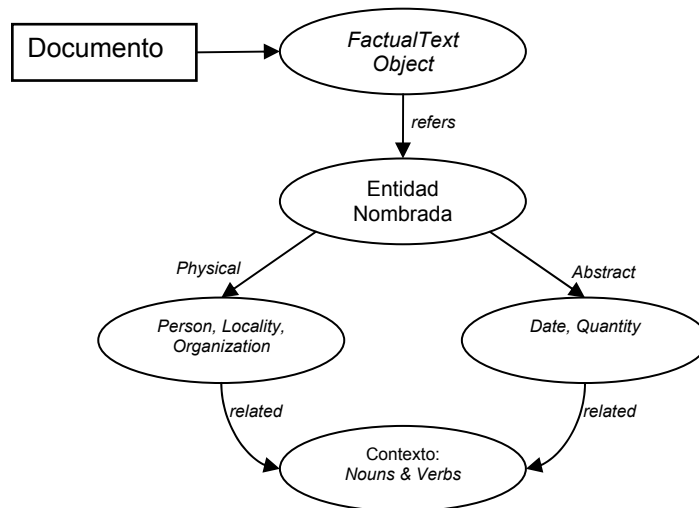


Figura 2. Modelo de documento propuesto, los conceptos y axiomas corresponden a los definidos en la ontología SUMO. Las etiquetas en cursivas se han conservado tal como se definen en SUMO.

### 3.2 Selección de Respuestas

El proceso de selección de respuestas está soportado por dos procesos previos, por un lado la representación de los documentos fuente, la cual se indexa en un proceso fuera de línea. Por otro lado, el procesamiento de la pregunta, mismo que no se detalla en este documento, y del que se extrae la siguiente información: La clase semántica de la entidad esperada como respuesta a la pregunta, así como las entidades nombradas y el contexto de la pregunta. El algoritmo de búsqueda de respuestas candidatas se describe a continuación.

- 1 Recuperar los documentos relevantes a las entidades nombradas de la pregunta.
- 2 Recuperar los contextos de los documentos obtenidos en 1.
  - 2.1 Analizar cada contexto a partir de su entidad nombrada asociada.
  - 2.2 Calcular la similitud entre el contexto en turno y el de la pregunta y sus entidades.
  - 2.3 Conservar como candidatas a contener la respuesta sólo aquellos cuya entidad nombrada asociada pertenece a la clase semántica esperada como respuesta.
- 3 Ordenar en forma decreciente los contextos de acuerdo a su similitud.
- 4 Obtener las entidades asociadas a los contextos candidatas.
- 5 Reportar las primeras tres respuestas candidatas como posibles respuestas.

A continuación se muestra un ejemplo de la aplicación del algoritmo propuesto para responder a la pregunta *¿Quién era el presidente del PNR en 1931?*:

1. Recuperar los documentos relevantes a las entidades nombradas de la pregunta: PNR (Organización), la cual se encuentra en los documentos {0,13,86}; y 1931 (Fecha), que se encuentra en los documentos {0}. Como consecuencia, el subconjunto de documentos que pueden responder a la pregunta es {0}
2. Recuperar los contextos de los documentos obtenidos en 1 (ver tabla 1).
  - a. Analizar cada contexto a partir de su entidad nombrada asociada.
  - b. Calcular la similitud entre el contexto en turno y el de la pregunta y sus entidades {ser, presidente, PNR, 1931}.
  - c. Conservar como candidatos a contener la respuesta sólo aquellos cuya entidad nombrada asociada pertenece a la clase semántica esperada como respuesta.
3. Ordenar en forma decreciente los contextos de acuerdo a su similitud.
4. Obtener las entidades asociadas a los contextos candidatos (columna “entidades nombradas de clase persona” en la tabla 1)
5. Reportar las primeras tres respuestas candidatas como posibles respuestas {Cárdenas, Cárdenas\_Presidente}.

Tabla 1. Entidades candidatas y su contexto.

Contexto	Entidades Nombradas de clase Persona	Sim.
{presidente, PNR, echar, mano}	Cárdenas	0.5
{prm, fundar, carácter, otorgar, ser, forma, permanecer, candidatura}	Cárdenas_Presidente	0.125
{creador, 30, año, plebiscito, añoranza, embellecer, muerte}	Portes_Gil	0
{arribar, 1964, pri, presidencia, Carlos, juventud, líder, camisa}	Madrazo	0
{Madrazo, arribar, 1964, pri, juventud, líder, traer}	Carlos	0
{pri, faltar, mano, gato}	Madrazo	0
{zurrar, Sinaloa, Madrazo, corto}	Polo_Sánchez_Celis	0
{Sinaloa, zurrar, Polo_Sánchez_Celis, corto, 11, mes, salir}	Madrazo	0
{pierna, cola, pri, salir, diputado, pelea, deber}	Polo_Martínez_Domínguez	0

En este ejemplo, “Cárdenas” es la respuesta correcta, el pasaje original se muestra en la tabla 2.

Tabla 2. Pasaje con la respuesta a la pregunta del ejemplo.

<p>“<b>Cárdenas</b> como presidente del PNR echó mano del Ersatz de democracia en 1931 por vez primera en la selección de candidatos a gobernadores”.</p>
---

## 4 Resultados Preliminares

Para la evaluación de este trabajo se siguió el criterio aplicado en la tarea de búsqueda de respuestas del CLEF en su edición del 2003 (QA@CLEF-2003) [5]. La colección de documentos usada fue provista por la agencia de noticias española EFE y contiene noticias de dominio abierto del año 1994 y 1995. La colección EFE1994 contiene 215,738 noticias (509 MB), mientras que la EFE1995 contiene 238,307 noticias (577

MB). El conjunto de preguntas de evaluación consta de 200 preguntas sobre hechos, de las cuales 20 no tienen respuesta. Es importante mencionar que el resultado alcanzado en el QA@CLEF-2003 para la tarea monolingüe en Español, del sistema presentado por la universidad de Alicante, España [10] fue del 35% de precisión en evaluación estricta y 38.5% en no-estricta, ambas sin usar recursos externos (como Internet).

La tabla 1 muestra los resultados obtenidos con la aproximación de nuestro trabajo con base en el uso de contextos léxicos y entidades nombradas. Estos alcanzan el 30% de precisión en evaluación estricta y 33.5% en no-estricta. Esto se debe en parte a la confusión del clasificador de entidades, misma que ha sido compensada parcialmente durante el cálculo de similitud (sección 3.2). Otro factor de impacto en la selección de respuestas se encuentra en el empate de respuestas candidatas, es decir, se encuentran diversas respuestas candidatas con el mismo valor de similitud. Para subsanar esto se calculan las frecuencias de ocurrencias de contextos idénticos en estructura, entidad asociada y similitud.

Tabla 1. Resultados preliminares de la aproximación propuesta (LabTL-Inaoe)

Sistema	Estricto		No Estricto	
	MRR	%	MRR	%
Alicante	0.2966	35.0	0.3175	38.5
LabTL-Inaoe	0.2349	30.0	0.2558	33.5

## 5 Trabajo en Proceso

A la fecha se realiza una evaluación detallada de los errores del método con objeto de determinar las limitaciones de esta aproximación tanto por los recursos de PLN necesarios como por el tipo de preguntas capaz de responder. También se realizan experimentos con contextos más amplios en longitud y partes de la oración a considerar. El siguiente paso es el uso de recursos externos como Euro-WordNet e Internet para incrementar la precisión del sistema. Una vez que se cuente con los resultados de estos experimentos se procederá a la siguiente etapa, el uso de contextos con base en información sintáctica.

**Agradecimientos.** Este trabajo se realizó con el apoyo parcial de CONACYT becas 166876 y U39957-Y, así como del laboratorio de Tecnologías del Lenguaje, INAOE. Los autores también agradecen al comité organizador del CLEF y a la agencia EFE por los recursos facilitados.

## Referencias

1. Burger, J. et al. *Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A)*. NIST 2001.
2. Carreras, X. and Padró, L. *A Flexible Distributed Architecture for Natural Language Analyzers*. In Proceedings of the LREC'02, Las Palmas de Gran Canaria, Spain, 2002.

3. Cowie J., et al., *Automatic Question Answering*, Proceedings of the International Conference on Multimedia Information Retrieval (RIAO 2000)., 2000.
4. Hirshman L. and Gaizauskas R. *Natural Language Question Answering: The View from Here*, Natural Language Engineering 7, 2001.
5. Magnini B., Romagnoli S., Vallin A., Herrera J., Peñas A., Peinado V., Verdejo F. and Rijke M. *The Multiple Language Question Answering Track at CLEF 2003*. CLEF 2003 Workshop, Springer-Verlag.
6. Mann, G.S. *Fine-Grained Proper Noun Ontologies for Question Answering*, SemaNet'02: Building and Using Semantic Networks, 2002.
7. Niles, I. and Pease A., *Toward a Standard Upper Ontology*, in Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), 2001.
8. Prager J., Radev D., Brown E., Coden A. and Samn V. *The Use of Predictive Annotation for Question Answering in TREC8*. NIST 1999.
9. Ravichandran D. and Hovy E. *Learning surface text patterns for a question answering system*. In ACL Conference, 2002.
10. Vicedo, J.L., Izquierdo R., Llopis F. and Muñoz R., *Question Answering in Spanish*. CLEF 2003 Workshop, Springer-Verlag.
11. Vicedo, J.L., Rodríguez, H., Peñas, A. and Massot, M. Los sistemas de Búsqueda de Respuestas desde una perspectiva actual. Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural, n.31, 2003.

# Proyecto de preparación del corpus paralelo alineado español-inglés

José Angel Vera Félix, Grigori Sidorov

Centro de Investigación en Computación, CIC.  
Av. Juan de Dios Batíz s/n, casi esq. Miguel Othón de Mendizabal,  
Unidad Profesional "Adolfo López Mateos", Edificio CIC.  
Col Nueva Industrial Vallejo, CP 07738, México D.F.  
javera@sagitario.cic.ipn.mx, sidorov@cic.ipn.mx

**Resumen.** Se presenta la implementación de una metodología enfocada a la alineación de textos paralelos, a nivel de párrafos y oraciones, para los lenguajes inglés – español. Se describe el método a seguir y algunos de los resultados obtenidos hasta el momento. Al final, se presentan los trabajos futuros relacionados a este proyecto.

**Abstract.** An implementation of the methodology for alignment of parallel texts at the paragraph level for English and Spanish languages is presented. The method and some preliminary results are described. Future directions of development are presented.

**Palabras Clave:** Alineación de textos, corpus paralelo alineado, detección de párrafos.

## 1 Introducción

Dentro de la lingüística computacional, existen campos de investigación que se apoyan, en gran medida, en recursos lingüísticos electrónicos tales como periódicos o revistas, siendo la traducción automática un ejemplo de ellos.

Mucho se ha dicho sobre la función de Internet como un gran corpus electrónico [5]; los investigadores cuentan con una gran cantidad de material de estudio en esta vasta cantidad de textos electrónicos, en sus muy diversos lenguajes, que se generan hoy en día y que se distribuyen en Internet.

Un campo de trabajo de interés actual es el que está conformado por los corpus paralelos. Dicho campo se ve beneficiado por la disponibilidad, en Internet, de los textos multilingües que conforman este tipo de corpus. Un corpus paralelo es un caso especial de un corpus multilingüe, el cual agrupa el mismo texto en diversas lenguas. En nuestro caso, manejamos textos en dos lenguas diferentes, a saber, el inglés y el español.

Para efectos de investigación, no basta con recopilar textos para formar un corpus paralelo; un material de este tipo no es de mucha utilidad pues los textos no son fácilmente procesados por los programas actuales usados en este campo. Para aprovechar al máximo el material, los textos deben estar estructurados de manera tal que se facilite su análisis.

De tal suerte, se pone en evidencia la necesidad de procesar los corpus paralelos para que queden en un formato útil para los investigadores y manejables para el software. Uno de los tratamientos más importantes que se aplica al corpus paralelo es lo que se conoce como la alineación. La alineación de los textos es un primer paso, muy necesario, que generará lo que se conoce como un corpus paralelo alineado. Alinear significa establecer la correspondencia entre los elementos estructurales de textos –párrafos, sentencias o palabras.

Un tipo de software que trabaja con textos paralelos y que requiere que los textos estén alineados es el que contiene los programas de concordancias [1].

Para entender la gran importancia de los corpus paralelos alineados, y por ende, del proceso de alineación, basta con ver la gran influencia que estos recursos han tenido en los avances alcanzados en las investigaciones sobre traducción automática, enseñanza de lenguas extranjeras [3] y lingüística contrastiva [6].

## 2 Textos Paralelos y su Alineación

Los textos paralelos son textos que tienen el mismo contenido semántico, pero expresado en lenguajes diferentes [8]. Reciben también el nombre de corpus paralelos, bilingües o bitextos. Son ejemplos de textos paralelos los manuales de productos electrónicos que presentan la misma información al usuario, pero en diferentes lenguajes.

El término “paralelo” no implica que los textos tengan una correspondencia exacta entre palabras, oraciones y/o párrafos; es decir, dos textos pueden estar completamente desalineados sin dejar de ser textos paralelos.

La alineación de un corpus paralelo consiste en la reestructuración de los textos de forma tal que se establezca una correspondencia entre los párrafos, las sentencias y/o las palabras de los textos involucrados. Considerando los lenguajes inglés y español, el alineado se puede llevar a cabo dividiendo en secciones el texto en inglés y definiendo, para cada una de estas secciones, una sección del texto en español que empate lo mejor posible con la sección en inglés [7], o al revés.

Algunos autores, entre ellos [7], distinguen cuatro diferentes niveles de alineación, a saber:

- **Alineación de 1er orden:** Alineación utilizada cuando los textos que conforman el corpus son muy cortos. Se denomina alineación a nivel de texto (completo).
- **Alineación de 2do orden:** La alineación de 2do grado se refiere a la alineación donde los segmentos por alinear son los párrafos.
- **Alineación de 3er orden:** La alineación de 3er grado se refiere a la alineación donde los segmentos por alinear son las oraciones.
- **Alineación de 4to orden:** La alineación de 4to grado se refiere a la alineación donde los segmentos por alinear son las palabras.

La tabla 1 proporciona un ejemplo, muy simplificado, de un texto alineado a nivel de oración.

**Tabla 1.** Ejemplo de texto alineado a nivel de oración

<b>Texto en inglés.</b>	<b>Texto en español.</b>
<s>BudaPesth seems a wonderful place, from the glimpse which I got of it from the train and the little I could walk through the streets.</s>	<s>Budapest parece un lugar maravilloso, a juzgar por lo poco que pude ver de ella desde el tren y por la pequeña caminata que di por sus calles.</s>
<s>feared to go very far from the station, as we had arrived late and would start as near the correct time as possible.</s>	<s>Temí alejarme mucho de la estación, ya que, como habíamos llegado tarde, saldríamos lo más cerca posible de la hora fijada.</s>

### 3 Recopilación del Corpus

Para crear el corpus paralelo se recopilaron obras literarias clásicas (novelas) en formato electrónico. Los textos se obtuvieron de Internet de las siguientes direcciones:

- (Textos en español) <http://www.artnovela.com.ar/>
- (Textos en español) <http://www.librodot.com/>
- (Textos en inglés) <http://www.promo.net/pg/>

Las traducciones al español fueron llevadas a cabo por varios traductores, lo que implica que hay diferentes estilos de traducción. Lo anterior puede ser considerado un factor importante al momento de probar la aplicación de alineación pues el software debe resultar eficiente sin importar el estilo o tendencia de los textos en cuestión. Un factor que dificulta el proceso de alineación es la inclusión, en la traducción, de comentarios que no son parte del original.

La tabla 2 muestra algunas de las obras literarias que conforman el corpus:

**Tabla 2.** Obras literarias usadas en el corpus paralelo

<b>Autor</b>	<b>Título inglés</b>	<b>Título español</b>
James, Henry	The turn of the screw	Otra vuelta de tuerca
Shelley, Mary	Frankenstein	Frankenstein
Stoker, Bram	Drácula	Drácula
Verne, Julio	Five weeks in a balloon	Cinco semanas en globo
Verne, Julio	From the earth to the moon	De la tierra a la luna

### 4 Detección de Segmentos de Texto

El presente trabajo implementa las bases para un mecanismo de alineación de textos a nivel de oraciones y de párrafos. Se ha desarrollado un programa que realiza el alineado a nivel de párrafos y se sigue trabajando en las técnicas de alineado que deben ser incluidas en este programa.

#### 4.1 Detección de párrafos

El trabajo de detección de párrafos es mucho más simple que el trabajo de detección de oraciones, lo cual resulta fortuito ya que la alineación a nivel de oraciones mejora



mucho si se realiza primero la alineación a nivel de párrafos. El algoritmo implementado maneja algunos problemas tales como la detección de varios delimitadores de párrafo (retorno de carro) consecutivos, los cuales deben ser ignorados. Otro posible caso que debe ser atendido es que, si en el texto se detecta una línea completa de espacios sin texto real que deba ser considerado como un párrafo.

Para verificar si el programa desarrollado identifica correctamente los párrafos de algunos textos de prueba, se comparó el número de párrafos detectados por nuestra implementación con el número de párrafos detectados por Word. Cabe aclarar que solo se está comparando el número de párrafos detectados.

Algo que representa un problema es el hecho de que la implementación considera como párrafos los títulos, encabezados, y otros elementos. Sin información alguna acerca del texto (metadatos), una posible manera de decidir si un segmento de texto es un párrafo, es considerando el tamaño, en caracteres, del mismo; los títulos o encabezados tienden a contener pocos caracteres respecto a los párrafos normales.

## 4.2 Detección de oraciones

Se está trabajando en el algoritmo para la detección de oraciones. El delimitador más conocido para las oraciones es el punto “.”, sin embargo, debido a que este elemento es usado también para otras funciones, no podemos basar el algoritmo en este único factor.

El algoritmo debe reconocer elementos tales como *Sr., Ing., Dr., 3.1416, etc., etc.*. Cualquiera de estos elementos puede ser, o puede no ser, el final de la oración. Si después de la sigla la siguiente palabra inicia con mayúscula se considera que la sigla finaliza la sentencia. Este razonamiento presenta problemas para las siglas que denotan a una persona pues “*Dr. Dante*” identifica a “*Dr.*” como el final de la oración. En el caso de los números la identificación es más sencilla pues si después del punto no se sigue otro número (3.) o símbolo se entiende que este número termina la oración.

También se cuida el hecho de que alguien puede omitir el punto final de oración al terminar un párrafo, en este caso se detecta y se etiqueta el fin de oración.

En la tabla 3 se presenta el resultado obtenido al analizar uno de los párrafos escritos en esta sección. Se puede observar el error comentado sobre “*Dr. Dante*”.

Tabla 3. Detección de sentencias

---

<b>&lt;numSent=1&gt;</b> El algoritmo debe reconocer elementos tales como <i>Sr., Ing., Dr., 3.1416, etc., etc.</i> . <b>&lt;/numSent&gt;</b> <b>&lt;numSent=2&gt;</b> Cualquiera de estos elementos puede, o puede no, ser el final de la sentencia. <b>&lt;/numSent&gt;</b> <b>&lt;numSent=3&gt;</b> Si después de la sigla la siguiente palabra inicia con mayúscula se considera que la sigla finaliza la sentencia. <b>&lt;/numSent&gt;</b> <b>&lt;numSent=4&gt;</b> Este razonamiento presenta problemas para las siglas que denotan a una persona pues “ <i>Dr.</i> ” <b>&lt;/numSent&gt;</b> <b>&lt;numSent=5&gt;</b> <i>Dante</i> ” identifica a “ <i>Dr.</i> ” como el final de la sentencia. <b>&lt;/numSent&gt;</b> <b>&lt;numSent=6&gt;</b> En el caso de los número la identificación es más sencilla pues si después del punto no se sigue otro número (3.) o símbolo se entiende que este número termina la sentencia. <b>&lt;/numSent&gt;</b>
--

---

## 5 Metodología de Alineación Empleada

Actualmente hay bastante trabajo de investigación respecto a la alineación. En la tabla 4 se enumeran algunas investigaciones referentes a la alineación a nivel de oraciones:

**Tabla 4.** Artículos desarrollados sobre alineación de textos a nivel de oraciones

Artículo	Lenguajes	Método
Brown et al. (1991c)	Inglés, Francés	# de palabras
Gale & Church (1993)	Inglés, Francés, Alemán	# de caracteres
Church (1993)	Varios	Señales 4-gram
Fung & McKeown (1994)	Inglés, Cantonés	Señales léxicas
Kay & Röscheisen (1993)	Inglés, Francés, Alemán	Léxico
Chen (1993)	Inglés, Francés	Léxico

Obsérvese que, de estos, ninguno considera al español dentro de sus investigaciones (aunque no se afirma que no exista trabajo al respecto). Esto puede ser un inconveniente si se considera las técnicas lingüísticas, pues éstas son dependientes del lenguaje, no así las técnicas estadísticas.

Se puede observar que las investigaciones sobre alineación de textos han seguido dos corrientes, a saber, la corriente estadística y la corriente lingüística.

Las técnicas estadísticas ofrecen una gran velocidad en el proceso de alineación y cierta independencia entre la técnica y los lenguajes que se están tratando. Su punto débil es que este método puede fallar si la traducción no cumple con la filosofía sobre la cual se apoya dicha técnica. El cumplir con la filosofía implica que la traducción debe tener alta semejanza estructural con el texto original; oraciones grandes (en caracteres) en el original tienden a ser traducidas a oraciones grandes y oraciones cortas tienden a ser traducidas a oraciones cortas [4].

Las técnicas lingüísticas tienden a ser mucho más lentas en el proceso de alineado, además de que resultan dependientes de los lenguajes tratados. No obstante, se espera que éstas generen mejores resultados al considerar el sentido de los textos. Por ejemplo, se pueden considerar nombres propios, datos numéricos o cognados para que, a partir de estos elementos, se tenga un punto de referencia entre las oraciones de ambos textos.

De [4] tomamos la idea de alinear los párrafos considerando sus tamaños en caracteres. [4] señala que, en un texto y su traducción, se observa una correlación respecto a sus tamaños. La correlación es una especie de escala que nos dice, por ejemplo, que a cada 2 caracteres del texto en inglés le corresponde 1 carácter de la traducción. La idea que tomamos es que, si dos textos (los originales, inglés y su traducción) tienen un valor de correlación (definido en 1)  $n$ , durante el proceso de alineado se anexarán párrafos de manera tal que se obtenga una correlación resultante y ésta será más cercana a  $n$  en cuanto mejor alineados estén los textos resultantes.

La correlación entre los textos (**correlación texto, Ct**) se basa en el número de caracteres del texto fuente (**caracteres texto fuente, Ctf**) y del texto destino (**caracteres texto destino, Ctd**). Esta correlación se define como:

$$Ct = \frac{Ctf}{Ctd} \quad (1)$$

Después de calcular **Ct** se inicia un proceso en el cual se selecciona un párrafo por cada texto y se calcula la correlación específica para estos párrafos (**correlación párrafo, Cp**). Lo que se busca es que **Cp** se acerque lo más posible a **Ct**.

Al calcular **Cp**, tendremos una de dos situaciones y acciones a realizar:

- **Cp < Ct**: Debemos anexar otro párrafo en el texto original (inglés).
- **Cp > Ct**: Debemos anexar otro párrafo en el texto destino (español).

Siempre que anexamos un párrafo en cualquiera de los textos, **Cp** se recalcula, con lo que dicho parámetro se acerca cada vez más a **Ct**. En esta operación se observa que hay ocasiones en que el orden se invierte (si **Cp < Ct** y se anexa un párrafo en inglés se obtiene **Cp > Ct**). Los textos quedan alineados por párrafos una vez que se ha reducido la diferencia entre **Cp** y **Ct** hasta un cierto nivel y sin que ocurra la inversión.

El programa desarrollado aun depende en gran medida de la intervención del usuario, la herramienta solo hace sugerencias sobre qué párrafo anexar y estas sugerencias las determina de acuerdo al valor **Cp** calculado en cada momento. En una aplicación semiautomática se debe realizar el alineado por si mismo y sólo en el caso en que se entre a un ciclo de inversiones de correlaciones se debe activar un modo de uso manual para que sea el usuario quien decida la alineación deseada.

## 6 Resultados

Con el trabajo realizado hasta el momento, se han obtenido algunos resultados en cuanto a la alineación de párrafos. Se muestra a continuación la aplicación del método a partir de un fragmento de "*Drácula*". Debido a las limitaciones de espacio, no se presentan las pantallas del software; solamente el texto, los valores de **Ct** y **Cp**, y la diferencia (**Dif**). Se presenta en la tabla 5 el párrafo en inglés y en la tabla 6 el párrafo en español, seleccionados inicialmente. En las tablas 7 y 8 se presentan el primer y segundo párrafos en español anexados. Debajo de cada una de las tablas 6, 7 y 8 se dan los valores calculados. En la tabla 9 se observa la inversión que ocurriría al anexar otro párrafo más, lo cual ya no deja alineado el texto en español con el de inglés, **Cp** se empieza a alejar de **Ct**.

**Tabla 5.** Párrafo inicial en inglés

---

3 May. Bistritz. \_\_Left Munich at 8:35 P.M, on 1st May, arriving at Vienna early next morning; should have arrived at 6:46, but train was an hour late. BudaPesth seems a wonderful place, from the glimpse which I got of it from the train and the little I could walk through the streets. feared to go very far from the station, as we had arrived late and would start as near the correct time as possible. The impression I had was that we were leaving the West and entering the East; the most western of splendid bridges over the Danube, which is here of noble width and depth, took us among the traditions of Turkish rule.

---

**Tabla 6.** Párrafo inicial en español

---

Bistritz, 3 de mayo. Salí de Múnich a las 8:35 de la noche del primero de mayo, llegué a Viena a la mañana siguiente, temprano; debí haber llegado a las seis cuarenta y seis; el tren llevaba una hora de retraso.

---

**Ct = 0.875; Cp = 2.953; Dif = 2.079**

---

**Tabla 7.** Primer párrafo sugerido y anexado (español)

---

Budapest parece un lugar maravilloso, a juzgar por lo poco que pude ver de ella desde el tren y por la pequeña caminata que di por sus calles. Temí alejarme mucho de la estación, ya que, como habíamos llegado tarde, saldríamos lo más cerca posible de la hora fijada.

---

**Ct = 0.875; Cp = 1.298; Dif = 0.424**

---

**Tabla 8.** Segundo párrafo sugerido y anexado (español)

---

La impresión que tuve fue que estábamos saliendo del oeste y entrando al este. Por el más occidental de los espléndidos puentes sobre el Danubio, que aquí es de gran anchura y profundidad, llegamos a los lugares en otro tiempo sujetos al dominio de los turcos.

---

**Ct = 0.875; Cp = 0.836; Dif = 0.038**

---

**Tabla 9.** Tercer párrafo no sugerido (español)

---

Salimos con bastante buen tiempo, y era noche cerrada cuando llegamos a Klausenburg, donde pasé la noche en el hotel Royale. En la comida, o mejor dicho, en la cena, comí pollo preparado con pimentón rojo, que estaba muy sabroso, pero que me dio mucha sed.

---

**Ct = 0.875; Cp = 0.619; Dif = 0.255**

---

## 7 Conclusiones

Se ha presentado una aproximación de alineación de textos, inglés - español, a nivel de párrafos. Se presentaron también algunos de los resultados obtenidos.

Hasta ahora sólo se ha usado el valor de correlación entre los textos para saber que párrafo debe ser anexado (en que idioma). Se observó, durante las pruebas, que el método funciona de manera adecuada siempre y cuando la traducción respete la estructura del texto original. Si un párrafo del texto original se ha dividido en  $n$  párrafos en la traducción, durante la alineación se unirán los párrafos segmentados (de acuerdo a **Cp**) para formar un solo párrafo que se corresponde con el original (los párrafos de las tablas 6, 7 y 8 se unen y alinean con el párrafo de la tabla 5).

El algoritmo de [4] emplea programación dinámica y probabilidades de alineación, elementos que no consideramos en este trabajo. Aún así, se observa la misma dependencia entre las estructuras de los textos que se mencionan en [4]. Se presenta un mal funcionamiento cuando esta estructura no se respeta, por ejemplo, en una prueba se omitió un párrafo en la traducción, lo cual no fue detectado por el programa, y se empezó a alinear párrafos que no tenían correspondencia entre sí. Este es un problema que se presenta en muchos de los métodos estadísticos ya probados, y hay técnicas que han ayudado a resolverlo, aunque no se ha logrado obtener un método 100% eficaz con métodos puramente estadísticos.

El programa aun no se ha automatizado todavía. La implementación sólo se limita a hacer los cálculos y sugerencias al usuario; si el usuario comete un error y acepta una alineación que no es correcta (como en tabla 9) el sistema no lo detectará y seguirá con el proceso, teniéndose un caso similar al descrito en el párrafo anterior (omisión de párrafo).

## 8 Trabajos futuros

Se debe implementar por completo el método estadístico. Se piensa que los mejores resultados en tiempos cortos en la alineación de texto se lograrán cuando se combinen las técnicas estadísticas y léxicas, esto es algo que se está considerando implementar.

Aunque no se mencionó en este trabajo, la alineación de textos generalmente implica la generación de los textos paralelos alineados y etiquetados. Se está estudiando el esquema de marcación a utilizar, específicamente se piensa usar el esquema CES (Corpus Encoding Standard) el cual es una aplicación de SGML.

Finalmente, cuando se haya implementado un método fiable se buscará la automatización de la herramienta, la cual solo debe pasar a un funcionamiento manual cuando detecte que no puede obtener una alineación correcta en pocas iteraciones.

## Referencias

- [1] Barlow, M. (2002). "ParaConc: Concordance software for multilingual parallel corpora". *Language Resources for Translation Work and Research*. 20-24.
- [2] Brown, P. F., Lai, J. C. & Mercer, R. L. (1991). "Aligning Sentences in Parallel Corpora". In *Proceedings of the Annual Meeting of the Association for Computational Linguistics – 91*, Berkeley, California.
- [3] Botley, S. P., McEnery, A. M. & Wilson A. (2000). "Multilingual Corpora in Teaching and Research". From the series *Language and Computers: Studies in Practical Linguistics*, No 22. Amsterdam; Atlanta, GA: Rodopi.
- [4] Gale, W. A. & Church, K. W. (1991). "A program for Aligning Sentences in Bilingual Corpora". In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Berkeley, California.
- [5] Galicia-Haro, S., A. Gelbukh, I. A. Bolshakov (2004). "Web-based Sources for an Annotated Corpus Building and Composite Proper Name Identification". *Atlantic Web Intelligence Conference (AWIC-2004)*. Lecture Notes in Computer Science, Springer-Verlag.
- [6] Johansson, S. (1997). "Using the English-Norwegian Parallel Corpus - a corpus for contrastive analysis and translation studies". In B. Lewandowska-Tomaszczyk & P.J. Melia (eds.) *Practical applications in language corpora*. Łódź: Łódź University. 282-296.
- [7] Mikhailov, M. (2003). "Parallel Corpus Aligning: Illusions and Perspectives".
- [8] Nevill, C.G., & Bell, T.C., (1992). "Compression of Parallel Texts". *Information Processing and Management*, Great Britain: Pergamon Press. 28(6), 781-793.

# Uso de Lattices para la Recuperación de Términos

Liset Fraguela C., Héctor Jiménez S.

Facultad de Ciencias de la Computación

B. Universidad Autónoma de Puebla

Edif. 135. Ciudad Universitaria, Puebla, Pue. CP 72570. México

Tel. (01222) 229 55 00 ext. 7212 Fax (01222) 229 56 72

lisetfraguela@hotmail.com, hjimenez@fcfm.buap.mx

Gerardo Sierra M.

Grupo de Ingeniería Lingüística

Universidad Nacional Autónoma de México

Torre de Ingeniería, Ciudad Universitaria, México, DF

GSierraM@iingen.unam.mx

**Abstract:** According to the formal concept theory, this paper presents the use of superior and inferior approximations for terminological information retrieval. The experiments show the need to incorporate synonym dictionaries to information retrieval systems, in order to expand user's keywords.

**Resumen:** Este artículo presenta el uso de las aproximaciones superior e inferior, con base en la teoría de conceptos formales, para la búsqueda de información terminológica a partir de la descripción del concepto. Se observa la necesidad de incorporar diccionarios de sinónimos a los sistemas de recuperación de información, a fin de expandir las palabras clave usadas por el usuario.

**Palabras clave:** Recuperación de información terminológica, lattices, búsqueda onomasiológica.

## 1. Introducción

En el campo de la recuperación de información terminológica, se conoce como búsqueda onomasiológica a aquella en la que se parte de la descripción del concepto para llegar al término que lo designa[8]. El problema es: a partir de la descripción de algún concepto obtener las palabras clave cuya definición se aproxime a la descripción dada. La propuesta que hacemos en este trabajo consiste en utilizar la aproximación en lattices para aplicarlo a Sistemas de Recuperación de Información. Para la aplicación particular en un diccionario onomasiológico, se realizó una prueba con un glosario o listado de conceptos de

la Física, con base en los trabajos que desarrolla G.Sierra y J.McNaught[9].

La teoría de conceptos formales fue introducida por R. Wille [2]. A partir de este planteamiento se han realizado varias aplicaciones en diferentes ámbitos [1]. La motivación para utilizar esta teoría es el hecho de que un concepto se representa por una pareja cuyo primer componente es el conjunto de ejemplares que denota el concepto, y la segunda componente es el conjunto de características que tienen todos los ejemplares del concepto. Además, consideramos los trabajos realizados en cuanto a la aproximación de conceptos [4][3].

Con estas aproximaciones es posible encontrar en el lattice de conceptos elementos que aproximan a conjuntos de objetos o a conjuntos de atributos. La aplicación que aquí presentamos considera la consulta del usuario como un conjunto de atributos y espera obtener un conjunto de objetos que aproximen a la descripción dada.

## 2. Antecedentes

Para precisar las ideas de concepto formal deberán considerarse las cotas superiores e inferiores de un subconjunto  $S$  de un conjunto ordenado  $P$ , asimismo, si existen, la mínima cota superior y máxima cota inferior de  $S$ :  $supS$  e  $infS$ , también denotadas por  $\bigvee S$  y  $\bigwedge S$ , respectivamente.

Dado un conjunto ordenado  $P$ , si  $x \wedge y$  y  $x \vee y$  existen para toda pareja  $x, y \in P$ , llamamos a  $P$  *lattice* (o *retícula*). Un contexto  $(G, M, \varphi)$  está formado por un conjunto de objetos,  $G$ , un conjunto de rasgos,  $M$ , y una correspondencia de  $G$  en  $M$ ,  $\varphi$ . Se define así un concepto como la pareja  $(A, B)$ , donde  $A \subset G$  es llamado extento del concepto,  $B \subset M$  es llamado intento del concepto. Las parejas  $(A, B)$  deben cumplir:

$$A = B' = \{g \in G \mid \varphi(g) = m, \forall m \in B\} \quad (1)$$

$$B = A' = \{m \in M \mid \varphi(g) = m, \forall g \in A\} \quad (2)$$

Una relación de orden parcial puede ser construida sobre los elementos del *lattice* de conceptos. Dados  $(A_1, B_1)$  y  $(A_2, B_2)$  decimos que  $(A_1, B_1)$  es *más particular* que  $(A_2, B_2)$ ,  $(A_1, B_1) \leq (A_2, B_2)$ , si y sólo si  $A_1 \subseteq A_2$ , o equivalentemente  $B_2 \subseteq B_1$ . Con lo anterior, en un contexto puede demostrarse que para cada pareja  $(A_1, B_1)$  y  $(A_2, B_2)$  existe el *sup* e *inf*, es decir  $L$  es un *lattice*. En un *lattice*  $L$ ,  $x$  es *inf-irreducible* si para cada  $y, z \in L$ ,  $x = y \wedge z$  implica  $x = y$  o  $x = z$ ; dualmente, un elemento  $x$  es *sup-irreducible* si para cualesquier  $y, z \in L$ ,  $x = y \vee z$  implica  $x = y$  o  $x = z$ .

Por otro lado, el enfoque de conjuntos rugosos [6] considera el empleo de relaciones de equivalencia o tolerancia para hacer aproximaciones a conjuntos. En un sistema de información  $(G, M)$ , se considera  $G$ , un conjunto de objetos y  $M$  un conjunto de atributos. En  $(G, M)$  se definen las aproximaciones inferior y superior de  $X \subset G$  como:  $\underline{X}_B = \{x \in G \mid [x]_B \subset X\}$  y  $\overline{X}_B = \{x \in G \mid [x]_B \cap X \neq \emptyset\}$ , respectivamente, donde  $[x]_B$  es la clase de equivalencia que induce  $B \subset M$ ,

las cuales se determinan por  $[x]_B = \{y | x.b = y.b, \forall b \in B\}$ . En el caso de las aproximaciones que se han intentado manejar en la teoría de conceptos formales hay variantes sobre el tipo de relación que induce las clases de los elementos de  $G$ , permitiendo, por ejemplo, relaciones de tolerancia (reflexivas y simétricas). Para este trabajo se empleó la aproximación propuesta por Keyun Hu *et.al.* [4], como veremos enseguida.

Dado el contexto  $(G, M, \varphi)$  y correspondiente *lattice* de conceptos  $L$ , se define  $gR = \{m \in M | \varphi(g) = m\}$  para  $g \in G$ . Así, puede definirse una relación binaria  $J$  en  $G$  como:  $g_1 J g_2$  si y sólo si  $g_1 R \subseteq g_2 R$ , donde  $g_1, g_2 \in G$ . Claramente,  $J$  es una relación de orden parcial en  $G$ . Denotamos las cotas superiores de  $g$  como la clase de orden parcial  $[g]$ , específicamente  $[g] = \{g^* \in G : g J g^*\}$ , y con  $P$  el conjunto de parejas  $([g], [g]')$ . Si  $SI$  es el conjunto de todos los elementos *sup-irreducibles* de  $L$  puede probarse que  $P = SI$  [4]. Similarmente, dado  $Rm = \{g \in G | \varphi(g) = m\}$ , para un contexto  $(G, M, \varphi)$ , la relación binaria  $K$  sobre  $M$  definida como:  $m_1 K m_2$  si y solo si  $Rm_1 \subseteq Rm_2$ , donde  $m_1, m_2 \in M$ , permite referirnos a la clase parcial de  $m$  con  $[m]$ :  $[m] = \{m^* \in M : m K m^*\}$ . De manera semejante a los elementos *sup-irreducibles* tenemos que si  $II$  es el conjunto de todos los elementos *inf-irreducibles* de  $L$  y  $Q$  es el conjunto de todos los pares  $([m]', [m])$ , entonces  $Q = II$ .

Dados  $B \subseteq M$ , la aproximación inferior (A-Inf) y la aproximación superior (A-Sup) de  $B$  con respecto a  $L$  son:

$$B_\bullet = \cap \{B \subseteq B^* \mid (B^{*'}, B^*) \in II, \} \quad (3)$$

y

$$B^\bullet = \cup \{B^* \subseteq B \mid (B^{*'}, B^*) \in SI. \} \quad (4)$$

Con estas aproximaciones es posible encontrar en el *lattice* de conceptos elementos que aproximan a conjuntos de objetos o a conjuntos de atributos. La aplicación realizada considera a la consulta del usuario como un conjunto de atributos y espera obtener un conjunto de objetos que aproximen a la descripción dada.

### 3. Aplicación en un diccionario onomasiológico

Se consideraron los siguientes datos de entrada al sistema: dos glosarios del área de Física, uno de ellos tiene una definición para cada vocablo y el otro tres definiciones; una colección de consultas y, también, fue utilizado un diccionario de sinónimos.

Un glosario está formado por una lista de registros, a la vez, cada registro está compuesto por el nombre del concepto o vocablo ( $D$ ), que sería la entrada del diccionario, y a continuación el conjunto de palabras clave ( $T$ ), que describen dicho vocablo. El vocabulario utilizado estará formado por las palabras clave que describen a todos los vocablos del glosario, eliminando de éste las palabras cerradas, y truncando el resto con un truncador de Porter adaptado al español.



Las clases parciales pertenecientes a los elementos de  $II$  se obtienen de una tabla inversa donde se representan por cada palabra clave de la definición,  $T_i$ , todos los vocablos donde ésta aparece. Las clases que pertenecen a los  $SI$ , por otra parte, se obtienen de una tabla donde aparece cada vocablo  $D_j$  representado por las palabras clave que componen su definición.

Partiendo de estas dos tablas se realizan las aproximaciones superior e inferior dada una consulta. Para la aproximación superior se utiliza la tabla del  $SI$  y para la aproximación inferior la tabla del  $II$ . Estas aproximaciones se obtienen realizando las operaciones que a continuación se indican.

### 3.1. Aproximación Inferior por atributo ( $T_\bullet$ )

Suponemos formada, entonces, la tabla  $II$ . Dada una consulta, que a su vez debe tener las mismas características que el texto (es decir, truncado, sin palabras repetidas y sin palabras cerradas), para cada palabra clave que compone dicha consulta se determina en qué definiciones de vocablos aparece. El próximo paso será traslapar el resultado anterior y la Aproximación Inferior (ecuación 3) estará dada por aquellos vocablos que contengan en su descripción a las palabras clave de la consulta. Así la ecuación 3 se reexpresa como:

$$T_\bullet = \cap \{T \subseteq T^* \mid (T^{*'}, T^*) \in II\} \quad (5)$$

### 3.2. Aproximación Superior por atributo ( $T^\bullet$ )

Se parte de una tabla de vocablos ( $SI$ ) representados con sus respectivas definiciones, donde el texto se encuentra truncado, sin palabras repetidas y sin palabras cerradas. La aproximación superior (ecuación 4) es la unión del conjunto de vocablos en el que las palabras clave de su definición están contenidas en la consulta.

Teniendo en cuenta que en muy pocos casos se van cumplir las contenciones exactas en ambas aproximaciones, se introduce el criterio de *mayoría*. Para la Aproximación Superior, el criterio consiste en aceptar como válido un vocablo donde la mayoría (la mitad del total más uno) de las palabras clave que lo definen aparezcan en la consulta. Y, en el caso de la Aproximación Inferior se refiere a que la mayoría de las palabras clave de la consulta estén contenidas en la definición de un vocablo.

Puede suceder que no se utilicen en la consulta las palabras clave que aparecen en las definiciones de los vocablos sino algún sinónimo de éstas, y es por ello que en una de las pruebas se hizo la representación de las palabras clave, tanto de las que componen las definiciones como de la consulta, con un diccionario de sinónimos.

### 3.3. Método directo

Esta primera prueba fue realizada con el contexto de conceptos del área de la Física. Las consultas utilizadas versaron sobre nociones de Mecánica y

Cinemática dadas por profesores de Ingeniería, estudiantes de computación y otras extraídas del libro de L.Landau /E.Lifshitz[5], las cuales se presentan en la tabla 1.

	<b>Consulta (truncadas)</b>	<b>Resp. Correcta (RC)</b>
C1	tercer planet sistem sol	<i>Tierra</i>
C2	choqu conserv cantidad movimient lineal energi	<i>Choque elástico</i>
C3	aplic fuerz provoc desplaz mism direccion	<i>Energía mecánica</i>
C4	movimient sistem describ line rect	<i>Movimiento lineal Movimiento rectilíneo (Rectilíneo)</i>
C5	cambi veloc tiemp	<i>Aceleración (Aceleración angular y Aceleración centrípeta)</i>

Tabla 1. Consultas utilizadas y respuestas correctas.

Para las respuestas correctas de la Consulta C5 se encuentran entre paréntesis los hipónimos de *Aceleración*. Mientras que la consulta C4 tiene tres posibles respuestas correctas, dos de ellas son sinónimos y la que se encuentra entre paréntesis es un hiperónimo de las anteriores.

Describiremos ahora los resultados de la aplicación de las fórmulas de aproximación a los glosarios y consultas.

Para la consulta C2 no hubo resultados. Se observó que en ocasiones el usuario no utiliza una descripción como las que aparecen en el glosario y, por lo tanto, no se encuentra la respuesta deseada. Esta consulta en particular fue obtenida en el libro L.Landau /E.Lifshitz[5]. Con respecto a los resultados de A-Inf se puede decir que la evocación aumentó en algunos casos (particularmente en la A-Inf de la consulta C5) y que la precisión se mantuvo igual. Los resultados obtenidos en esta prueba con respecto a las Aproximación Superiores (A-Sup) fueron los esperados. Para las tres primeras consultas no hubo resultados puesto que es difícil que las descripciones de los vocablos estuvieran contenidas en la consulta.

### 3.4. Uso de un diccionario de sinónimos

En esta prueba se utilizó un diccionario de sinónimos (41.5 Kb) para representar las palabras clave de la consulta y de las definiciones mediante una palabra "equivalente". La consulta CN homogenizada con el diccionario la denotamos con CNeq. El diccionario de sinónimos es una lista de registros, cada uno está compuesto por el sinónimo representante (un total de 523) y a continuación el listado de sinónimos de dicho vocablo. La sustitución de las palabras clave, tanto de la consulta como de las definiciones, por su sinónimo representante (en caso que lo tenga) se lleva a cabo de la siguiente forma:

1. Buscar la palabra clave de la consulta o de la definición dentro de las palabras en el listado de sinónimos.
2. Si la palabra clave se encuentra en el listado de sinónimos, entonces la palabra clave se sustituye por el sinónimo representante de la lista.
3. Si la palabra clave no se encuentra en dicho listado de sinónimos, se mantiene ésta en la representación.

## 4. Análisis de resultados

Las tablas 2 y 3 resumen los resultados de la aplicación de los dos métodos aplicados a las consultas dadas por la tabla 1. En estas tablas se dispone el número de respuestas correctas separado por una diagonal del número de respuestas entregadas por el sistema.

Al comparar los resultados de los métodos Directo y Sinónimos se podría decir que el uso de un diccionario de sinónimos para la representación, tanto de las consultas como de los glosarios, aumenta levemente la precisión con respecto al método directo. Esta mejora se intuía antes de aplicar este segundo método, sencillamente porque una misma consulta puede ser expresada por varios usuarios de manera diferente, y esta diferencia es en parte por el uso de sinónimos. Por ejemplo, para la consulta C3, en nuestro caso, la expresamos de la siguiente manera: *aplicación de una fuerza que provoca desplazamiento en una misma dirección*, sin embargo otro usuario podría expresar lo mismo diciendo: *generación de una fuerza que produce un movimiento en el mismo sentido*. En ambos casos estaríamos buscando que nuestro sistema nos devolviera como respuesta correcta *Energía mecánica o Trabajo mecánico*.

Aunque se puede considerar que hubo una mejora de los resultados del método de representación de palabras claves usando un diccionario de sinónimos con respecto al método directo, los resultados de A-Sup y A-Inf son muy parecidos. Por ejemplo, A-Sup aumentó la precisión utilizando el glosario de tres definiciones para cada vocablo en la consulta C2eq y la cantidad de resultados aumentó en las A-Inf de la consulta C5eq, al usar los dos glosarios y en la A-Sup de la consulta C4eq. Aparte de estos cambios, los demás resultados se mantuvieron igual a los del método anterior. Nos preguntamos, sin embargo, por qué no mejoraron aún más estos resultados. Un elemento que influye en dichas pruebas está relacionado con las entradas del diccionario de sinónimos utilizado. Otra limitación que tuvimos es que el diccionario de sinónimos un diccionario general y, por otra parte, no se considera el contexto de la palabra para hacer la elección del sinónimo correcto. Hay que tener en cuenta que partimos de dos glosarios donde como máximo contábamos con tres definiciones por vocablo. Sin embargo, siempre habrá más de tres formas de describir dichos vocablos.

Consulta	Directo		Sinónimo	
	A-Inf	A-Sup	A-Inf	A-Sup
C1	1/2	0/0	1/2	0/0
C2	0/0	0/0	0/0	0/0
C3	1/1	0/0	1/1	0/0
C4	1/2	2/2	1/2	2/2
C5	3/15	2/2	3/21	2/2

Tabla 2. Comparación de resultados de las A-Inf y A-Sup de los dos métodos en cuanto a precisión (mediante la relación  $P = R/NRes$ ); usando glosario de una definición por vocablo.

Consulta	Directo		Sinónimo	
	A-Inf	A-Sup	A-Inf	A-Sup
C1	1/2	0/0	1/2	0/0
C2	0/0	0/0	0/0	1/1
C3	1/1	0/0	1/1	0/1
C4	1/2	2/2	1/2	2/2
C5	3/19	2/2	3/27	2/2

Tabla 3. Comparación de resultados de las A-Inf y A-Sup de los dos métodos en cuanto a precisión (mediante la relación  $P = R/NRes$ ); usando glosario de tres definiciones por vocablo.

## 5. Conclusiones

En este artículo se presentó el uso de las aproximaciones superior e inferior, con base en la teoría de conceptos formales, con el fin de obtener los vocablos cuya definición se aproxima a la descripción de un concepto dado por un usuario. Los resultados fueron satisfactorios con el método Directo para un glosario de una definición. De manera general, el método mejora para glosarios con tres definiciones y aún más con el método de representación por Sinónimos. Estos resultados hacen evidente la necesidad de expandir la consulta dada por un usuario en una búsqueda de información, ya que una misma consulta puede ser expresada por los usuarios de manera diferente.

En este sentido, el análisis de los resultados nos permite proponer dos puntos principales para mejorar, en trabajos futuros, el uso de las aproximaciones superior e inferior en la obtención de términos en sistemas de recuperación de información. El primero es el uso de diccionarios de sinónimos especializados en el área que se están realizando las búsquedas. El segundo es contar con glosarios más extensos en cuanto a formas de describir un vocablo, incorporando, de ser posible, descripciones dadas por diferentes tipos de usuarios.

## Referencias

- [1] Hans-Hermann Bock (Ed.): *Classification, Data Analysis, and Knowledge Organization*, North-Holland, Amsterdam, 1991.
- [2] Bernhard Ganter & Rudolph Wille: *Formal Concepts Analysis*, Springer Verlag, 1999.
- [3] R. Kent: "Rough Concepts Analysis: A sythesis of rough set and formal concept analysis", *Fundamenta Informaticae* 27(1996), 169-181.
- [4] Keyun Hu, Yuefei Sui, Yuchang Lu, Ju Wang & Chunyi Shi: "Concept Approximation in Concept Lattice". PAKDD 2001, *LNAI* 2035, Springer Verlag, pp. 167-173 2001.
- [5] L. Landau & E. Lifshitz: *Curso Abreviado de Física Teórica, Mecánica y Electrodinámica*. Editorial Mir Moscú, 1979.
- [6] Z. Pawlak: *Rough Sets -theoretical aspects of reasoning about data*, Kluwer, 1991.
- [7] C.J. van Rijsbergen: *Information Retrieval*. University of Glasgow, pp. 114 - 117. Second Edition, 1999.
- [8] Gerardo Sierra & John McNaught: "Design of an onomasiological search system: A concept-oriented tool for terminology". *Terminology*. Vol. 6 (1), 2000.
- [9] Gerardo Sierra & John McNaught: "Natural Language System for Terminological Information Retrieval". CICLing 2003, *LNCS* 2588, Springer Verlag, pp. 541-552, 2003.

## **A Proposed Architecture for a Spoken Information Retrieval with Multimedia Databases**

Luis R. Salgado-Garza, Juan A. Nolazco-Flores, Pablo D. Díaz-López

Computer Science Department, ITESM, Campus Monterrey,  
Av. Eugenio Garza Sada 2501 Sur, Col. Tecnológico,  
Monterrey, N.L., México, C.P. 64849.  
{lsalgado, jnolazco, pdiaz}@itesm.mx

**Abstract.** This document describes the realization of a spoken information retrieval system and its application to word search into indexed multimedia databases. The multimedia database is build from a multiformat set of text, audio and video documents. The whole archive collection is indexed using preprocessing techniques to produce transcripts and indexing software tools to catalog them. The system uses a Java-based distributed client-server architecture. A Java applet is used to capture the audio signal for a spoken query, then it is transmitted to a server where an automatic speech recognition (ASR) software is applied to convert the signal into a transcribed hypothesis. Later, a query tool process the transcript sentence along with the indexed multimedia database and a set of pointers to documents are generated. Finally, a web page with links to the resulting documents, where queried words appear, are presented to the user.

**Resumen.** Este documento describe una arquitectura para un sistema recuperación de información con interfaz de voz y su aplicación a bases de datos multimedia. La base de datos propuesta se construye de documentos en formatos diversos, incluyendo texto, audio y video. La colección de archivos es indexada utilizando una etapa de preprocesamiento para generar transcripciones de texto, según el formato del archivo, y herramientas de indexado para catalogar dichas transcripciones. Se utiliza una arquitectura cliente-servidor distribuida, basada en tecnología Java. En la interfaz de usuario, una página de hipertexto utiliza un applet de Java para grabar la señal de audio correspondiente a una petición oral de búsqueda, la señal acústica es transmitida al servidor en dónde un reconocedor automático de voz (ASR) la convierte a una hipótesis de texto. Posteriormente, una herramienta de búsqueda procesa conjuntamente las palabras de la transcripción de voz y la información de la base de datos multimedia, produciendo apuntadores hacia documentos relevantes a la búsqueda. Finalmente, una página de hipertexto es generada para presentar al usuario enlaces hacia documentos de la base de datos relacionados con las palabras de su búsqueda.

**Key words:** SIR, multimedia, Java, ASR.

## 1. Introduction

The collection of written information in some sort of repository is a practice as old as civilization itself [1]. From ancient Mesopotamia and Alexandria to modern libraries around the globe, the need to collect information seems natural to the human kind. However, as time moves forward and new information arrives, we need new techniques to store and organize this heritage. Main concern is how to organize this large archives in order to have fast access to desired information, for this Information Technology has developed techniques for automatic classification, compression, storage and retrieval.

The everyday more accessible use of communications, such as radio, television and Internet, enable us to reach information from anywhere at anytime. We are immersed in a continuously growing world of information and we need tools to take advantage from it, in a simple and natural way to any user. The most natural mean to transmit information among humans is by voice, unfortunately a permanent recording of such class of information has been less common than text archiving, mainly because two reasons: the large amount of storage required for acoustics and its difficulties for classification or indexing. Nowadays, massive storage technologies are more affordable than ever and new algorithms for data compression permit to record hours of video and audio in just tenths or thousands of megabytes. These advances outline a future where multimedia document storage will surpass text-based archiving [2, 3].

In recent years, new techniques for text indexing and automatic information retrieval has been developed and successfully used over the Internet, focusing on digital documents containing text information. However, the future of these search engines requires its use on multimedia databases, if not, its application in mobile devices and hand-busy environments will be very limited. At present, several research groups have conducted efforts to develop search engines for multimedia repositories of information [4] commonly known as multimedia digital libraries. However, there is still much work to be done, particularly regarding speech information and spoken queries [5, 6].

In this paper we propose an architecture for information retrieval from multimedia documents, using spoken queries as user interface. The organization of this document is the following. Section 2 presents the architecture overview, with specifics about our realization. Section 3 shows the techniques for multimedia database construction and the use of preprocessing and indexing tools. In section 4 the spoken query technology is detailed, with specific emphasis in the use of automatic speech recognition (ASR) technology. Section 5 discusses our system setup and incipient experimental results. Finally, section 6 presents conclusions and tracks for future work.

## 2. System Overview

Our proposed implementation for the spoken query information retrieval system is based in a client-server architecture (Fig. 1), working as follows. A Java-enabled web front-end capture the acoustic information of the query, then it is transmitted as a

binary stream to a server-side servlet that invokes the audio preprocessing module and then the speech decoder. The decoder uses the acoustic and language models obtained from the training step. Using the Viterbi algorithm the decoder generates a hypothesis. This outcome is taken as the input for the query tool, and then our system produces a set of pointers to the documents where the queried words were found in the multimedia database. Finally, this result set is used to generate a web page (java server page – JSP) that is returned to the user, including the corresponding links to the document segments or audio/video file frames for subsequent consults.

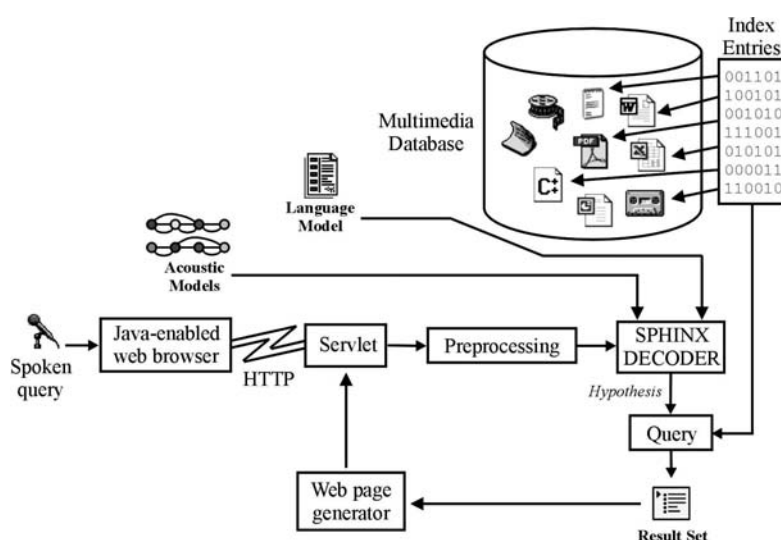


Fig. 1. General spoken query information retrieval system architecture.

In order to build a platform-independent user interface, the client-side module of our system will be implemented as a Java applet, which captured the acoustic stream uttered by the user to submit it later to the speech recognition and query system.

As already mentioned, on the server side, our system will use a servlet to receive the acoustic stream, as a wave file, and invoke the speech decoder to generate a transcription and then a query the multimedia database. The decoder is the CMU SPHINX-III system, running in the same Linux server as the main servlet, using acoustic and language models trained for Spanish language. For training and for recognition, the decoder preprocess the wave files to extract acoustic features as mel-frequency cepstral coefficients (MFCC) [7], then Hidden Markov Models (HMM) [8] are built for acoustic models of phones and triphones using the Baum-and-Welch algorithm. During recognition, the hypothesis generated by the decoder is supplied as input to the MG system query tool. Using an indexed multimedia database, a set of links to documents and specific pointers is generated, this information is used to build a dynamic web page, with references to the documents where queried words are found, and returned to the user.



Back on the client side, the returned web page is presented to the user with a choice of specific documents where the required information was located.

### 3. Multimedia Database Construction

The multimedia database used for information retrieval is a set of indexed documents, including plain-text, MS-Word, MS-Excel, MS-Powerpoint, PDF and audio/video files. For the non plain text documents, our system uses a preprocessing module, specific to each file format, to produce a raw transcription able to be processed and inserted into the indexed archive (Fig. 2).

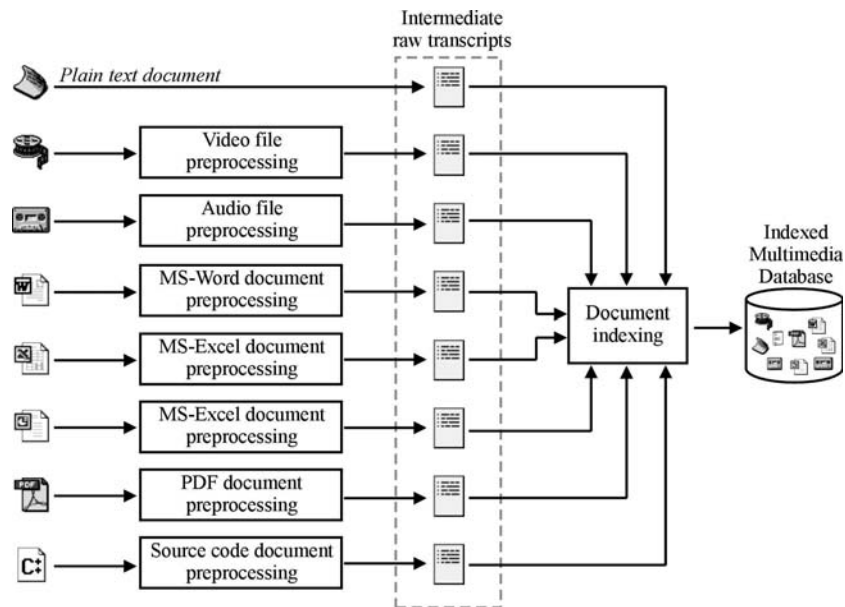


Fig. 2. Indexing procedure for video, audio, plain text and other formatted documents.

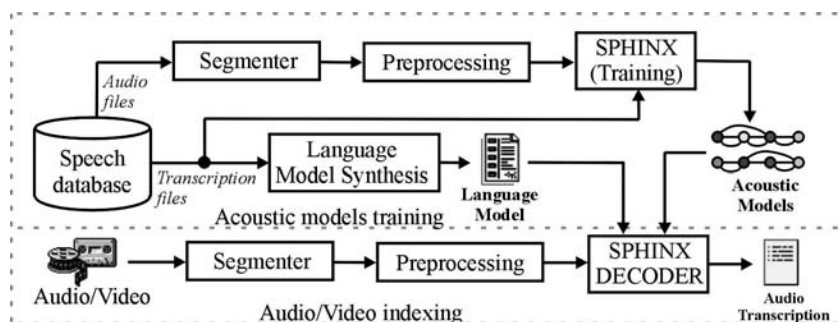
#### 3.1 Indexing procedure for files containing written text

For our information retrieval system, the queryable repository will be a database with several types of file formats. Particularly, from its generic source, we can identify two file categories, text-sourced and audio-sourced. The first class groups file format such as plain text, MS-Word, MS-Excel, MS-Powerpoint and PDF; the other relates to audio and video files (either in wav, mp3, mpeg or avi formats). For indexing, text-sourced files require a preprocessing to extract the plain text from their different

formats; for this a text extraction module is needed to analyze the raw file directly, identify its raw format and produce the corresponding plain text transcript, including neither font nor document style formats, also special characters are ignored as well as attached images.

### 3.2 Indexing procedure for audio and video files

Particularly, for audio and video files, the preprocessing stage is more complex and uses an automatic speech recognition (ASR) system in order to extract the audio transcriptions (Fig. 3).



**Fig. 3.** Audio/video file preprocessing using a speech decoder, and related acoustic and language models generation, previous document indexing.

The initial stage for audio-sourced files preprocessing deals with segment boundaries definition in the acoustic signals, for this we use the CMUseg [12] tool, to generate speech segments. Then, audio frames are preprocessed, the tool allows to generate static and dynamic information condensed as a feature vector. Each vector is used as the input for the SPHINX-III speech decoder, that is a HMM-based automatic speech recognition environment for continuous speech and large vocabulary tasks. Each hypothesis from the decoder represents its best transcriptions for the audio signal, the whole set is stored as the audio transcription archive.

In order to perform speech recognition, we train acoustic models using a speech database, containing several hours of audio, produced by different speakers for clean speech. As for decoding, during training analog signal, from each audio file in the database, is segmented, then sampled and converted to MFCCs. Finally, by the use of the Baum-and-Welch algorithm [9], the SPHINX-III software obtains the acoustic models. The Baum-and-Welch algorithm needs the name of the word units to train as well as the label and feature vectors, these files are generated in a preprocessing stage previous to acoustic models training. The SPHINX-III system setup allows us to select acoustic model either for a phone, a triphone or a word set, also we are able to model either discrete, semicontinuous or continuous acoustic models.

To improve recognition accuracy, the SPHINX-III system uses standard n-gram based language models (LM), for our implementation we obtained the LM using the

CMU-Cambridge statistical language model toolkit version 2.0 [10]. The LM aim is to reduce the perplexity of the task, by predicting the following word based in the words' history. Several techniques can be used to synthetize a LM [11], among them N-grams is the easiest technique with very good results. If all the n-grams are not contained in the language corpus, smoothing techniques need to be applied. In the CMU-Cambridge language model toolkit, unigram, bigrams or trigrams can be configured for this tool, as well as four types of discount model: Good Turing, Absolute, Linear and Witten-Bell.

Finally, the audio transcription files are indexed using the Managing Gigabytes (MG) [3] tools suite, that is a large scale inverted index based text retrieval system. The indexed database spans over the entire audio signal extracted from the audio or video file, including every word detected by the speech recognizer. This database indexes the video in a frame-time bases and will be used for the queries of the next stage.

#### 4. Spoken Queries

Our proposed design enable Internet access to our spoken query system. On the client side, we propose the use of a Java front-end because of its features of platform independence and minimal system's resource requirements. The spoken query applet uses the user's computer microphone to capture a query of isolated words. The audio signal is stored as a wave-data stream in memory and later transmitted to a servlet; on the server side, the servlet receives the acoustic stream via TCP/IP and stores it as a wave file. Using the received wave file as argument, the servlet invokes a shell script to start the decoding process for the query while in the client's screen a message for a wait period is displayed. The acoustic stream is preprocessed to calculate MFCCs, using a wave-to-features tool, and taken into the speech decoder. The best hypothesis is produced as the transcription for the query. The MG query tool uses the indexed multimedia database to locate the files, text or audio sourced, where each uttered word, as indicated by the decoder's outcome, appears. The result set is built as pairs of *document name* and *pointer*, for text-sourced documents the *pointer* is a reference to specific page and paragraph, while for audio and video archives the *pointer* is a time reference from the beginning of the file. Finally, a web page is dynamically produced by the servlet and returned to the user, with the links to the documents and pointers generated by the query; this information may be used to perform subsequent consults to the specific documents.

#### 5. Results

At present, we have a running prototype of the spoken information retrieval system using plain-text, audio and video database files with a text-based front-end. The speech recognizer (SPHINX-III) and the information retrieval (MG) modules are fully functional and integrated, while the Java interface and the servlets are still undergoing. Currently, a Java applet records acoustic information and store it in a

client-side file, this wave file is manually transmitted to the speech recognizer. In the server-side, our Java modules are still in a blueprint, but they are planned to execute in a Linux box running Java Virtual Machine and Tomcat, as a servlet container and JSP generator. In our proposal, the same machine will hold the speech decoder, but in general these modules could be deployed in different computers, using the network for communication.

The configuration of the SPHINX-III system for our experiments used 13 MFCCs and also their first and second derivatives, therefore the feature vector accounted for 39 elements. The speech lower and higher frequencies were set at 300 Hz and 7,000 Hz, respectively. The frame rate was set to 50 frames per second using a 30ms Hamming window. A 512 samples FFT length was used and the number of filterbanks was set to 40. Five states continuous HMM was used as acoustic modeling technique with mixtures of 16 gaussians per state. The training database for the speech recognition system had a vocabulary of 22,398 different words (409,927 words in 28,537 sentences). Triphones were used as the word unit and language modeling is based in word trigrams. The baseline configuration of our speech recognition system was as reported in [11], for our latest experiments we used a new version of the SPHINX-III system leading us to the scores reported in table 1.

**Table 1.** Word Error Rates for the speech recognition system.

<b>Experiment</b>	<b>Language Weight</b>	<b>WER</b>
Baseline	10.0	26.44%
Experiment 1	9.5	14.98%
Experiment 2	10.0	14.68%

With so emboldening results from the speech decoder, we were motivated to test its combined efficiency with the MG indexing and retrieval system in audio databases. For our experiments were performed spoken queries using single and multiple words, uttered by different speakers, independent from the ones in the training set for the ASR system. We found that the spoken query always performed correctly, reporting every file where the queried words were actually present or uttered (in audio-sourced files).

## 6. Conclusions

As seen from the results, the overall performance of the system showed that the integration of a speech recognizer, a text based document indexing and retrieval tools comprise an effective architecture for a spoken query system on multimedia databases. However, more work is needed in order test the system accordingly to relevance retrieval and to increase the robustness of the system to out-of-vocabulary (OOV) words and acoustic noise conditions. Worth to try is to use noise compensation methods, syntactically inspired language model techniques [11, 13] and n-Best list hypothesis from the decoder and evaluate the effect of these in the information retrieval cogency. We also propose the use of phonetic features into the document indexing algorithms (multifeatured information indexing) for audio and

video archives, because the use of this kind of acoustic information could lead to more comprehensive classification system for spoken information retrieval.

## References.

- [1] Krasner-Khait, B., "Survivor: The History of the Library", *History Magazine*, October 2001.
- [2] Chen, B., H.M. Wang, and L.S. Lee., "Retrieval of Broadcast News Speech in Mandarin Chinese Collected in Taiwan using Syllable-Level Statistical Characteristics", *Proceedings of ICASSP-2000*.
- [3] Witten, I.H., Moffat, A., and Bell, T.C., *Managing gigabytes: compressing and indexing documents and images*. Van Nostrand Reinhold, New York (1994).
- [4] Witten, I.H., Don, K.J., Dewsnip, M., and Tablan, V., "Text Mining in a digital library", *Journal of Digital Libraries*, 2003 (In Press).
- [5] Wolf, P.P.; Raj, B., "The MERL SpokenQuery Information Retrieval System: A System for Retrieving Pertinent Documents from a Spoken Query", *IEEE International Conference on Multimedia and Expo (ICME)*, Vol. 2, 317-320, August 2002.
- [6] K. Spärck Jones, G. J. F. Jones, J. T. Foote, and S. J. Young., "Experiments in spoken document retrieval", *Inf. Processing and Management*, 32(4):399-417, 1996.
- [7] Deller, J.R., Proakis, J.G., Hansen, J.H.L., *Discrete-Time Processing of Speech Signals*, Prentice Hall, Sec. 6.2, 1993.
- [8] Baker, J.K., "Stochastic modeling for automatic speech understanding", In D.R. Reddy, ed., *Speech Recognition*, New York: Academic Press, pp.521-542, 1975.
- [9] Dempster, A.P., Laird, N.M., Rubin, D.B., "Maximum likelihood for incomplete data via the EM algorithm", *J. Roy. Stat. Soc.*, Vol. 39, No. 1, 1-38, 1977.
- [10] Clarkson, P., Rosenfeld, R., "Statistical Language Modelling using the CMU-Cambridge Toolkit", *Proceedings of Eurospeech*, Rhodes, Greece, 1997, 2707-2710.
- [11] Luis R. Salgado-Garza, Richard M. Stern, Juan Arturo Nolasco-Flores, "N-Best List Rescoring Using Syntactic Trigrams", *Proceedings of MICAI 2004, Advances in Artificial Intelligence*, Springer-Verlag (LNAI 2972:79-88)
- [12] M. Seigler, U. Jain, B. Raj, R. Stern, "Automatic segmentation, classification, and clustering of Broadcast news audio", *Proc. Of the DARPA speech recognition workshop*, February 1997.
- [13] Hsin-Min, W., Berlin, C., "Content-based Language Models for Spoken Document Retrieval", *International Journal of Computer Processing of Oriental Languages (IJCPOL)*, Vol. 14, No.2, 2001.

# Aplicando la Clasificación de Texto en la Extracción de Información

Alberto Téllez-Valero<sup>1</sup>, Manuel Montes-y-Gómez<sup>1,2</sup>, Luis Villaseñor-Pineda<sup>1</sup>

<sup>1</sup>Coordinación de Ciencias Computacionales  
Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE)  
e-mail: {albertotellezv, mmontesg, villasen}@inaoep.mx  
72840 Puebla, México

<sup>2</sup>Departamento de Sistemas Informáticos y Computación  
Universidad Politécnica de Valencia  
e-mail: mmontes@dsic.upv.es  
46022 Valencia, España.

**Resumen:** La clasificación de texto es la actividad de etiquetar textos en lenguaje natural con categorías temáticas tomadas desde un conjunto previamente definido. En contraste, la creación de una representación estructurada de información tomada desde textos en lenguaje natural es conocida como extracción de información. En este artículo proponemos definir a la extracción de información como una tarea de clasificación de texto. La idea principal de esta aproximación es usar algoritmos de aprendizaje automático en lugar de un sofisticado análisis lingüístico. Con el propósito de mostrar los resultados de este método, presentamos un sistema llamado “Topo”. Este sistema permite extraer información de noticias que reportan desastres naturales.

**Palabras Clave:** Extracción de Información, Clasificación de Texto, Aprendizaje Automático.

## 1. Introducción

El crecimiento explosivo de documentos disponibles en formato digital complica su exploración y análisis. Por consiguiente, son necesarios nuevos métodos que ayuden a los usuarios a filtrar y estructurar la información relevante. Dos ejemplos de esta clase de métodos son la clasificación de texto y la extracción de información.

La clasificación de texto (TC, por sus siglas en inglés) es la asignación de documentos de texto en formato libre a una o más categorías predefinidas en base a su contenido [1]. Ella se usa en aplicaciones donde la organización de documentos es requerida. Por ejemplo, en el indexado y filtrado automático de textos.

Por otro lado, la extracción de información (IE, por sus siglas en inglés) es usada para estructurar la información considerada relevante que se encuentra en documentos de un dominio determinado. En otras palabras, el objetivo de un sistema de extracción de información es encontrar y enlazar la información relevante, ignorando la extraña e

irrelevante [2]. La principal aplicación de un sistema de IE es llenar una base de datos con información proveniente de textos sin un formato predefinido.

La clasificación de texto y la extracción de información son dos formas complementarias de procesamiento de texto. Sin embargo, su interacción y sinergia ha sido mínima. Muchos sistemas de IE usan métodos de TC para asegurar que el documento a ser procesado contenga la información esperada. En este artículo exploramos como TC e IE pueden integrarse más estrechamente. Básicamente, proponemos una aproximación de IE que usa técnicas de TC para guiar las decisiones de extracción. Con el propósito de mostrar los resultados de esta propuesta, presentamos el sistema "Topo" que permite llenar automáticamente una base de datos sobre desastres naturales a partir de noticias de periódicos electrónicos en español.

El resto del artículo es organizado de la siguiente manera. La sección 2 describe el trabajo relacionado. En la sección 3 presentamos nuestra aproximación para extraer información usando clasificación de texto. La sección 4 describe una aplicación real de nuestro método y muestra los resultados obtenidos. Finalmente, en la sección 5 concluimos la discusión.

## **2. Trabajo relacionado**

La disponibilidad de grandes colecciones de textos ha estimulado el uso de técnicas estadísticas en muchas tareas del procesamiento del lenguaje natural (ver [3] para un estudio). El uso de métodos estadísticos en aplicaciones de extracción de información está principalmente enfocado en la adquisición automática de los patrones de extracción, los cuales son utilizados para dirigir la tarea de extracción (ver [4] y [5] para un estudio). A continuación, se mencionan brevemente las aproximaciones para extracción de información que se consideran más importantes con respecto al presente trabajo. Además, se describe el enfoque de clasificación de texto basado en técnicas de aprendizaje automático, el cual es usado en esta investigación.

### **2.1 Extracción de información**

Existen pocos trabajos que utilizan métodos de TC para guiar la tarea de extracción de información. En general, estos desarrollos presentan arquitecturas que se basan en la hipótesis que analizando la combinación de palabras que rodean a la información de interés, es suficiente para encontrar los patrones de extracción requeridos. La ventaja de estas arquitecturas es que un profundo análisis lingüístico no es requerido, en su lugar, técnicas de TC son usadas para aprender los patrones buscados.

Este enfoque parece ser bastante útil, sin embargo sus aplicaciones actuales no reflejan del todo su potencial, así como sus deficiencias. Por ejemplo, en [6] esta aproximación fue usada para el problema de identificar afiliaciones de autores y analizar citas en publicaciones, y en [7] para extraer datos desde anuncios de empleo. En las tareas antes mencionadas se tiene cierta estructura en la información a ser extraída, por tal motivo no se puede concluir mucho acerca de su comportamiento en documentos con formato libre. Por otro lado, en [8] se presenta una aplicación que se encarga

de llenar una base de datos de contactos con direcciones de e-mail tomadas desde correos electrónicos, los cuales presentan diferentes formatos y estilos.

Esta última aplicación da una mayor perspectiva de los alcances de dicha aproximación, sin embargo su dominio de estudio se limita a extraer sólo un tipo de dato (direcciones de e-mail). Por tal motivo, es necesario extender este trabajo a un dominio más complejo, donde no sólo exista variedad de formatos y estilos, sino que además, la información a ser extraída sea de más de un tipo y de diferentes características, por ejemplo: fechas, cantidades, nombres de lugares y eventos, entre otras.

## 2.2 Clasificación de texto

La aproximación de TC basada en aprendizaje automático a ganado importancia y eventualmente ha llegado a ser la dominante. Las ventajas de esta aproximación son: una buena efectividad, un considerable ahorro en términos de mano de obra del experto, y cierta independencia del dominio [9]. La construcción del clasificador de texto inicia con una colección de documentos manualmente clasificados (el conjunto de entrenamiento), y entonces se aplican algunos métodos de recuperación de información y aprendizaje automático para adquirir una función de clasificación (el clasificador). Tal aproximación considera las siguientes tareas (consultar [1] y [9] para más detalle):

- **Indexado.** El objetivo es representar los documentos de texto en una forma adecuada de su contenido para su uso con el clasificador. La representación más comúnmente usada es el modelo de espacio vectorial, donde los documentos son representados por vectores de términos. Tales términos son usualmente determinados por medio de algún tipo de análisis morfológico. También, se utilizan varios esquemas de pesado para los términos.
- **Reducción de dimensionalidad.** Es común que el espacio de representación tenga una alta dimensionalidad. Así, con el propósito de evitar el sobre ajuste en el proceso de aprendizaje e incrementar su eficiencia y efectividad, es necesario considerar sólo un subconjunto de los términos originales. La selección de los términos más representativos es hecha por una función de características. Por ejemplo, la ganancia en la información y la estadística chi-cuadrada.
- **Aprendizaje.** Varias técnicas han sido aplicadas para construir el clasificador. Por ejemplo: clasificadores bayesianos como Simple de Bayes, árboles de decisión como C4.5, aprendizaje basado en memoria como K-Vecinos más cercanos, y métodos kernel como Máquinas de Vectores de Soporte. Sin embargo, ninguna domina sobre otra para todos los dominios.

## 3. Arquitectura para la extracción de información

La arquitectura propuesta considera tres tareas principales:

1. Filtrar los textos relevantes al dominio de estudio.
2. Detectar las entidades de información con posibilidad de ser extraídas.



3. Discriminar entre las entidades identificadas aquellas que proporcionan información útil para llenar la base de datos.

El diseño de la arquitectura consiste de un modelo estructural formado por tres componentes conectados en cascada (ver figura 1). A continuación se describe cada uno de los componentes.

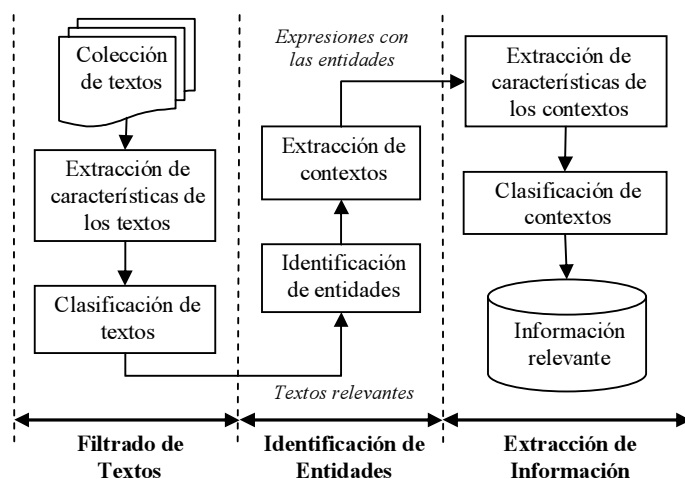


Fig.1 Arquitectura de Extracción de Información basada en TC.

### 3.1 Filtrado de textos

Debido a que es más probable que los textos pertenecientes al dominio de estudio contengan la información requerida, el objetivo de este componente es filtrar los documentos que se consideran relevantes. Una manera efectiva de llevar a cabo esta tarea es usando algún método de TC (ver sección 2.2).

Básicamente, el componente de filtrado de textos considera las siguientes tareas:

1. Transformar el texto de su formato inicial a una representación adecuada para la tarea de clasificación.
2. Aplicar el clasificador previamente construido para el dominio de estudio, con el propósito de separar los textos relevantes de los irrelevantes.

### 3.2 Identificación de entidades

La tarea de este componente es detectar, en los documentos considerados relevantes, aquellos segmentos de texto con alguna posibilidad de ser incluidos en la base de datos. A tales segmentos los vamos a llamar entidades. Para detectar las entidades decidimos hacer un análisis con expresiones regulares, esto con el propósito de hacer un mínimo uso de recursos lingüísticos.

Finalmente, una vez identificadas las entidades, la siguiente tarea es extraer la información que se encuentra próxima, a la cual llamaremos contexto. En la figura 2 se presenta un ejemplo.

### 3.3 Extracción de información

Una vez identificadas las entidades con posibilidad de ser extraídas, este componente se encarga de filtrar las que son consideradas útiles para llenar la base de datos. Para realizar tal selección se utiliza el contexto de la entidad junto con un TC previamente construido y entrenado para discriminar fragmentos del documento en las categorías de interés, las cuales generalmente se corresponden con los registros de la base de datos que deseamos llenar. En otras palabras, definimos a la tarea de extracción de información como una tarea de clasificación de textos, donde el propósito es clasificar entidades por su contexto en lugar de documentos por su contenido.

Es importante mencionar que en este componente se puede tener más de un clasificador de texto. Esto es debido a que en ocasiones es preferible especializar diferentes clasificadores en diferentes tipos de entidades. Por ejemplo, un clasificador para discriminar contextos de nombres y otro para discriminar contextos de cantidades. También, el mejor tamaño de los contextos para su correcta clasificación, así como el método de aprendizaje empleado, es algo que debe ser determinado mediante experimentación, y no precisamente debe ser el mismo para todos los tipos de entidades. En la figura 2 se ejemplifica el proceso de identificar entidades y de extracción de información.

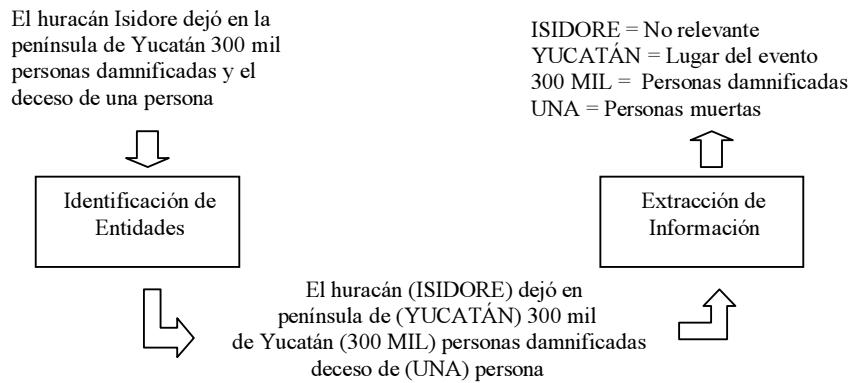


Fig.2 Proceso de identificar entidades y extraer información.

## 4. Caso de estudio: Noticias de desastres naturales

En esta sección introducimos un sistema de extracción de información basado en la arquitectura propuesta en la sección 3. El sistema es llamado “Topo”, y su propósito es extraer información relacionada con desastres naturales desde noticias de periódicos.

cos en español. Este caso de estudio está inspirado por el trabajo realizado por la Red de Estudios Sociales en Prevención de Desastres en América Latina<sup>1</sup>.

La base de datos a ser llenada por el sistema consiste en información relacionada con el evento (fecha, lugar y magnitud), en información relacionada con las personas (número de muertos, heridos o enfermos, desaparecidos, damnificados y afectados), información relacionada con las viviendas (número de viviendas destruidas y afectadas), e información relacionada con la infraestructura (número de hectáreas afectadas, pérdida económica, entre otras). Actualmente trabajamos con noticias que tratan los eventos de: huracán, incendio forestal, inundación, sequía y sismo. A continuación, presentamos las características técnicas de la implementación, y mostramos la forma de evaluar el sistema junto con los resultados obtenidos.

#### 4.1 Características técnicas

El conjunto de entrenamiento utilizado fue obtenido desde varios periódicos Mexicanos que están disponibles desde Internet. Esta colección consistió de 365 noticias que tratan sobre desastres naturales.

El componente de filtrado de textos incluye un clasificador entrenado con el conjunto previamente mencionado más una colección de 245 noticias consideradas irrelevantes<sup>2</sup>. Este clasificador implementa un indexado con una representación vectorial con entradas booleanas, es decir, cada texto es representado como un vector de ceros y unos que indican la presencia o ausencia de ciertos términos en el documento. El vector contiene 646 términos seleccionados mediante la función de características ganancia en la información. El esquema de aprendizaje utilizado es el clasificador Simple de Bayes.

Con respecto al componente de identificación de entidades, éste puede detectar nombres, fechas y cantidades utilizando un análisis con expresiones regulares. Además, se agregaron dos diccionarios para tratar las excepciones que ocurren en las expresiones con respecto a nombres y cantidades. Tales excepciones son las siguientes:

- Palabras reconocidas como una entidad de nombre únicamente porque inician con letra mayúscula (por ejemplo: artículos y preposiciones al inicio de una oración)
- Números reportados con letras en lugar de dígitos, o una combinación de ambos (por ejemplo: mil 500 personas)

En el componente de extracción de información se utilizan tres diferentes clasificadores, cada uno de ellos especializado en nombres, fechas y cantidades respectivamente. Para su entrenamiento se utilizaron un total de 1353 segmentos de textos –que representan contextos de las entidades de interés- tomados aleatoriamente desde el conjunto de entrenamiento. Los tres clasificadores implementan un indexado con una

---

<sup>1</sup> <http://www.desinventar.org/>

<sup>2</sup> Por textos relevantes entendemos todos aquellos que contienen información a ser extraída, mientras que los irrelevantes son los que contienen palabras o frases usadas comúnmente en la descripción de un fenómeno natural, pero que en estos casos se usan en contextos muy diferentes. Por ejemplo, la frase “ojo del huracán” en el contexto de “el presidente está en el ojo del huracán”.

representación vectorial con entradas nominales, es decir, cada contexto es representado por un vector de palabras. Debido a que el tamaño de los contextos que se tomaron para experimentar no es mayor de 14 palabras, no fue necesario hacer una reducción de dimensionalidad. Actualmente, los mejores resultados se han obtenido con contextos de 4 palabras (2 a la izquierda y 2 a la derecha) para clasificar nombres, y contextos de 6 palabras (3 a la izquierda y 3 a la derecha) para los otros dos casos. Los esquemas de aprendizaje con los que se han obtenido los mejores resultados son: k-Vecinos más cercanos para los nombres y Máquinas de Vectores de Soporte para los casos restantes.

Finalmente, cabe mencionar que el sistema “Topo” es una aplicación real implementada en Java utilizando Weka [10].

#### 4.2 Evaluación

Para evaluar el desempeño del sistema se utilizaron tres métricas: Precisión, cobertura y medida-F [1]. Las cuales son comúnmente usadas para evaluar sistemas de EI y TC.

Además, utilizamos la técnica de ejecutar 10 veces una validación cruzada con 10 pliegues (10-Fold Cross-Validation, en inglés) para la evaluación global. Esta técnica es la manera estándar de medir la tasa de error de un esquema de aprendizaje sobre un conjunto de entrenamiento [10]. La tabla 1 muestra el promedio de los resultados.

**Tabla.1** Resultados de evaluación.

<b>Componente</b>	<b>Precisión</b>	<b>Cobertura</b>	<b>Medida-F</b>
<b>FILTRADO DE TEXTOS</b>	<b>96%</b>	<b>95%</b>	<b>95%</b>
Huracán	100%	80%	89%
Incendio forestal	100%	100%	100%
Inundación	95%	100%	97%
Sequía	86%	97%	91%
Sismo	94%	98%	96%
Irrelevante	99%	96%	97%
<b>IDENTIFICACION DE ENTIDADES</b>	<b>88%</b>	<b>99%</b>	<b>93%</b>
Nombres	70%	100%	82%
Fechas	97%	100%	98%
Cantidades	98%	98%	98%
<b>EXTRACCION DE INFORMACION</b>	<b>80%</b>	<b>75%</b>	<b>77%</b>
Fecha del evento	63.3%	76.0%	69.1%
Lugar del evento	61.9%	47.0%	53.4%
Magnitud del evento	88.0%	88.0%	88.0%
Número de personas muertas	79.6%	74.7%	77.1%
Número de personas heridas o enfermas	94.7%	89.0%	91.8%
Número de personas desaparecidas	96.8%	75.0%	84.5%
Número de personas damnificadas	60.0%	53.8%	56.8%
Número de personas afectadas	71.4%	62.5%	66.7%
Número de viviendas afectadas	93.5%	72.5%	81.7%
Número de hectáreas afectadas	95.8%	92.9%	94.4%
Nombres irrelevantes	76.3%	85.6%	80.7%
Fechas irrelevantes	86.5%	77.8%	81.9%
Cantidades irrelevantes	69.8%	84.5%	76.5%

## 5. Conclusiones

Este artículo presenta una aproximación para construir sistemas de extracción de información. En general, la propuesta está soportada principalmente por un análisis de expresiones regulares y técnicas de clasificación de texto. Con esto se evita emplear cualquier tipo de análisis lingüístico sofisticado.

Nuestros experimentos demuestran el potencial de esta aproximación. Usando un conjunto de entrenamiento relativamente pequeño, alcanzamos una medida-F promedio de 77% para la tarea de extracción de información.

También, se puede concluir que la principal desventaja de esta propuesta es que no es posible extraer información expresada de forma implícita en los textos. Además, es complicado extraer y enlazar la información desde documentos que reportan más de un evento de interés. Creemos que tales problemas pueden ser parcialmente resueltos usando algún nivel de análisis lingüístico como etapa de pre-procesamiento, justo antes de aplicar la identificación de entidades.

## Agradecimientos

El presente trabajo fue parcialmente financiado por el CONACYT (Proyectos U39957-Y y 43990). Asimismo, el primer autor agradece al CONACYT por el apoyo otorgado a través de la Beca para Estudios de Maestría # 171610. Por su parte, el segundo autor agradece a la Secretaría de Estado de Educación y Universidades, España.

## Referencias

- [1] Aas K., Eikvil L., "Text Categorization: a Survey", Technical Report, Norwegian Computing Center (1999)
- [2] Cowie, J., Lehnert, W., "Information Extraction". Communications of the ACM, Vol. 39, No. 1 (1996) 80-91
- [3] Manning, C., Schütze, H., "Foundations of Statistical Natural Language Processing". MIT Press (1999)
- [4] Muslea, I., "Extraction Patterns for Information Extractions Tasks: A Survey". In Proceedings of the AAAI Workshop on ML for Information Extraction (1999)
- [5] Turno, J., "Information Extraction, Multilinguality and Portability". Revista Iberoamericana de Inteligencia Artificial. No. 22 (2003) 57-78
- [6] Bouckaert, R., "Low level information extraction". In Proceedings of the workshop on Text Learning (TextML-2002), Sydney, Australia (2002)
- [7] Zavrel, J., Berck, P., Lavrijssen, W., "Information Extraction by Text Classification: Corpus Mining for Features". In Proceedings of the workshop Information Extraction meets Corpus Linguistics, Athens, Greece (2000)
- [8] Kushmerick, N., Johnston, E., McGuinness, S., "Information Extraction by Text Classification". 17th International Joint Conference on AI (IJCAI-2001), N. Kushmerick Ed. Adaptive Text Extraction and Mining (Working Notes), Seattle, Washington (2001) 44-50
- [9] Sebastiani, F., "Machine Learning in Automated Text Categorization: a Survey". Technical Report IEI-B4-31-1999, Istituto di Elaborazione dell'Informazione (1999)
- [10] Witten, I. H., Frank, E., "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations". Morgan Kaufmann (2000)

# Comparación de dos métodos que determinan automáticamente el extracto de un texto

Hilario Salazar-Martínez, David Pinto, Héctor Jiménez-Salazar

Facultad de Ciencias de la Computación  
Benemérita Universidad Autónoma de Puebla  
C.U. 72570, Puebla, México

hilario\_cam@yahoo.com.mx, dpinto@cs.buap.mx, hjimenez@fcfm.buap.mx

**Abstract.** In this paper a comparison between two automatic text extraction corpus-based methods is presented. Each method determines the most representative sentences by assigning a score to each sentence on a plain text. The second method uses the text body to calculate the score, and the first method uses only the text head. Despite the first method uses less information, in our test, it got similar results than the second one. Therefore, in such cases that the text head is available, it is preferable to use the first method, due to, its complexity time. In this case, the first method is in  $O(n)$  and the second method is in  $O(n^2)$ ;  $n$  is the number of sentences in a plain text.

**Key Words:** Automatic Text Extraction, Mutual Information, Sense Relationship

**Resumen.** En este trabajo se comparan dos métodos para la generación automática de extractos de textos; ambos métodos utilizan pocos recursos lingüísticos y se apoyan en un corpus para obtener dicho extracto. Se usa una función de similitud para asignar una puntuación a cada oración y de esta manera determinar las oraciones más representativas. En el primer método, la función de similitud se aplica al encabezado del texto con respecto a cada oración del cuerpo del mismo. En el segundo, la función de similitud es aplicada a cada oración y su complemento. Aunque el segundo método utiliza más información, en nuestras pruebas, obtuvo un desempeño semejante al primero. Así, resulta conveniente utilizar el primero, ya que su comportamiento, con un texto de  $n$  oraciones, está en la clase  $O(n)$ , mientras que el segundo en  $O(n^2)$ .

**Palabras Clave:** Generación Automática de Extractos, Relaciones de sentido.

## 1. Introducción

Los volúmenes de información cada vez son mayores y con el surgimiento de Internet el manejo de éstos cobró mayor importancia, por lo que se hace necesaria la búsqueda de nuevas formas que ayuden a comprender eficazmente el contenido de un documento, sin tener que leer completamente el mismo. Por generación automática de resumen de un texto se entiende el proceso por el cual se identifica la información sustancial, proveniente de una fuente (o varias) para conocer de qué trata un documento sin necesidad de leerlo completamente, y producir una versión abreviada destinada a un usuario particular (o grupo de usuarios) y a una tarea (o tareas) específica.

Existen muchos enfoques y desarrollos para resolver el problema de la generación del resumen automático. En algunos enfoques se han generado diversas herramientas, algunas de las cuales hacen uso de recursos lingüísticos grandes. Otros más emplean recursos semánticos y otros, métodos estadísticos [4,12]. En general, la mayoría de los métodos se basan en recursos copiosos.

La generación automática del extracto de un texto, es una tarea particular en la generación automática de resúmenes, y de ahí la importancia de analizar las propuestas realizadas en este sentido. Daniel Marcu [3], por ejemplo, propuso un algoritmo que construye automáticamente un corpus para entrenar sistemas que determinan el extracto de un texto. Este algoritmo toma un conjunto de tuplas como entrada (Resumen, Texto) y genera el extracto correspondiente, es decir, el conjunto de cláusulas (oraciones) en el texto utilizado para escribir el correspondiente resumen. Para el desarrollo del algoritmo se realizó un experimento que fue evaluado por jueces. El experimento también sugiere estrategias de extracción para mejorar el desarrollo del sistema de resumen automático.

Debe reconocerse que la generación automática del resumen de un texto requiere mayor cantidad de recursos. Por ejemplo, el trabajo de Gustavo Crispino et al [7] propone una plataforma para resolver problemas de extracción de información y resumen automático. Este trabajo propone la definición de un modelo conceptual general de representación para conocimientos lingüísticos y el desarrollo de tareas especializadas que cooperen entre sí. En el modelo se identifican los conocimientos lingüísticos ubicándolos en sus contextos y organizándolos en tareas especializadas. Presenta, por un lado, la ventaja de permitir que el trabajo del lingüista se realice de manera independiente de su implementación informática, y, por otro, la de articular efectivamente en un mismo sistema informático los dos tipos de tareas.

Existen muchos trabajos que se apoyan en modelos de representación bien conocidos como el vectorial [5]. Así, pueden generarse resúmenes personalizados [8], cuyo fin es proveer al usuario de información contenida en textos de acuerdo con sus intereses.

Hemos abordado el problema de la generación automática de extractos mediante un algoritmo que ordena las oraciones contenidas en un texto según su similitud con el texto mismo [9]. El algoritmo está en  $O(n^2)$  donde  $n$  es el número de oraciones del texto. Los extractos obtenidos fueron evaluados por cuatro jueces humanos y arrojaron resultados prometedores. En el presente trabajo exponemos una variante inspirada en el trabajo de Daniel Marcu [3], que permite reducir el tiempo de cálculo de  $O(n^2)$  a  $O(n)$  sin detrimento de su eficacia.

La hipótesis de que una oración  $O_i$  de un texto  $T$  es más representativa que otra  $O_j$ , si la similitud entre  $O_i$  y  $T$  es mayor que la similitud entre  $O_j$  y  $T$ , conlleva a la experimentación con diversas técnicas a fin de obtener un conjunto de oraciones que puedan considerarse como el extracto de un texto. Cabe mencionar que el algoritmo determina el extracto de textos planos (sin estructura, etiquetas, u otra información).

En la sección 2 de este trabajo se introducen dos conceptos importantes para el entendimiento del trabajo desarrollado: el uso de la relación de sentido que se obtiene a través del corpus y el uso de la medida de información mutua como una forma de controlar la expansión de palabras usando contextos de un corpus. El método mejorado se expone en la sección 3. En la sección 4 se presenta una comparación entre los dos métodos mencionados usando 40 textos del dominio “política”. En la parte final de este artículo se presentan las conclusiones del trabajo.

## 2. Conceptos Básicos

El recurso en que se apoya el método que se presenta es el conjunto de relaciones de sentido de un término. Las relaciones del sentido de una palabra  $x$ , es el conjunto de todas aquellas palabras relacionadas con  $x$  [2]. En este caso, aproximamos las relaciones de sentido por los términos que ocurran en los contextos de  $x$ . Así, se están usando los términos de primer orden para representar  $x$  [1,6]. Se han realizado otras aplicaciones que utilizan esta representación [13].

Para obtener el extracto de un texto, identificamos las oraciones más representativas de él. Para este fin, usamos las relaciones de sentido y “expandimos” los términos de cada oración del texto; esto último permite aplicar una función de similitud que califique cada oración con un nivel de representatividad.

Por otro lado, consideramos importante refinar la aproximación a las relaciones de sentido que provee el corpus, con la finalidad de incrementar la precisión de los resultados. Así, hemos hecho uso del concepto de información mutua. La información mutua (IM) [11] proviene de conceptos teóricos sobre la información de sistemas y finalmente se traduce a una medida. La IM es entonces un cociente de asociación, para medir la norma de asociación de palabras.

Si dos palabras,  $x$  e  $y$ , tienen probabilidades  $\Pr(x)$  y  $\Pr(y)$ , entonces su información mutua,  $I(x, y)$ , se define como:

$$I(x, y) \equiv \log_2 \frac{\Pr(x, y)}{\Pr(x) \Pr(y)} \quad (1)$$

En este caso, las probabilidades  $\Pr(x)$  y  $\Pr(y)$  son estimadas calculando la frecuencia de ocurrencia relativa de  $x$  e  $y$  en un texto. La probabilidad conjunta,  $\Pr(x, y)$ , es estimada calculando el número de veces que ocurren  $x$  e  $y$  en la misma oración y dividiendo por el número oraciones del corpus ( $N$ ). Considerando que la frecuencia conjunta de términos pueda ser cero, la fórmula 1 queda reexpresada como:

$$M(x, y) = \log_2 \left( \frac{N \cdot fr(x, y)}{rf(x) \cdot rf(y)} + 1 \right), \quad (2)$$

donde  $rf(x)$  y  $rf(y)$  son la frecuencia de ocurrencia del término  $x$  e  $y$ , respectivamente, y  $fr(x, y)$  es la frecuencia de ocurrencia en una misma oración de los términos  $x$  e  $y$ .



### 3. Métodos de Extracción

Nuestro enfoque parte del supuesto de que el encabezado de un texto ofrece una buena aproximación al lector acerca del contenido del mismo y, por lo tanto, éste podría representar al texto [10]. Valiéndose de una función de similitud, se buscan entonces las oraciones  $O_i$  que tengan mayor similitud con el encabezado. Como puede verse, para determinar la puntuación de todas las oraciones, se realizan  $n$  cálculos de similitud para un texto de  $n$  oraciones. En tanto, si para calcular la puntuación de una oración  $O_i$  se realizan  $n$  cálculos de similitud, usando el cuerpo del texto, el algoritmo tendría un comportamiento cuadrático para asignar la puntuación a las  $n$  oraciones del texto.

Con la idea anterior se propuso realizar dos experimentos, en los que se hace uso de un corpus  $C$  (en este caso del dominio “política”) que permite representar los términos. Tanto  $C$  como el texto  $T$  deben ser preprocesados, obteniendo  $C_1$  y  $T_1$  respectivamente; en este caso, se eliminan las palabras cerradas y se trunca cada término. Además,  $T_1$  se descompone en su encabezado  $T_e$  y su cuerpo  $T_c$ . El corpus  $C_1$  se emplea para representar cada una de las oraciones con los términos de asociación de primer orden. Así, se obtiene lo que llamamos “oración representada”. La similitud en cada una de las pruebas se calcula con las oraciones representadas. A continuación se precisan las ideas antes expuestas.

#### 3.1. Representación de Texto

Las relaciones de sentido con respecto al corpus  $C_1$  son finalmente un conjunto de pares [9]:

$$V = \{(x, y) | x, y \in O, \text{ para una oración } O \in C_1\}. \quad (3)$$

La representación de una palabra  $x$  con base en  $V$  es:

$$\text{Expand}(x, V) = \{y | (x, y) \in V \wedge M(x, y) > 5\}. \quad (4)$$

La representación de una palabra es vista como un conjunto de palabras (sin repeticiones). A partir de la representación de una palabra, es posible representar una oración formando una tupla. Corresponde a  $O = (x_1, \dots, x_k)$  la representación:

$$\text{ExpandO}(O, V) = (\text{Expand}(x_1, V), \dots, \text{Expand}(x_k, V)). \quad (5)$$

La representación,  $T_2$ , del texto  $T_1$  es, entonces:

$$T_2 = (\text{ExpandO}(O_1, V), \dots, \text{ExpandO}(O_n, V)). \quad (6)$$

donde  $T_1$  está formado por  $(O_1, O_2, \dots, O_n)$ .

### 3.2. Extracto del Texto

Veamos ahora los métodos de extracto automático de un texto. El primero emplea el encabezado para determinar la puntuación de cada oración del cuerpo y, el segundo usa el mismo cuerpo para calcular esta puntuación.

**A: Encabezado.** Para calcular la puntuación de cada oración del texto, calculamos la similitud entre el encabezado del texto  $T$ ,  $T_e$ , y su cuerpo  $T_c$ . En esta tarea utilizamos el coeficiente de Jaccard. Lo anterior considerando la representación de cada una de estas oraciones:

$$sim(T_{e2}, O_{ic2}) = \frac{\#(T_{e2} \cap O_{ic2})}{\#(T_{e2} \cup O_{ic2})}, \quad (7)$$

donde  $O_{ic2}$  es la  $i$ -ésima oración representada del cuerpo de  $T_2$  y  $T_{e2}$  es el encabezado expandido de  $T_2$ .

**B: Complemento.** Este método se basa en el complemento de una oración [9], con respecto al texto, para elegir las oraciones más representativas. La puntuación de una oración se calcula con la siguiente simplificación del coeficiente de Jaccard:

$$sim'(O_{ic2}, \bar{O}_{ic2}) = \#(O_{ic2} \cap \bar{O}_{ic2}). \quad (8)$$

Aquí,  $sim'$  cuenta el número de elementos comunes de la oración  $i$  y su complemento en  $T_2$ ,  $\bar{O}_{ic2}$ ; es decir  $\bar{O}_{ic2}$  es  $T_2$  sin  $O_{ic2}$ .

## 4. Comparación de los métodos

El corpus utilizado en las pruebas realizadas consta de 375 textos, todos del dominio "Política". Dicho corpus tiene un tamaño de 2.7 kb, con 21559 pseudolexemas (considerando nombres propios) y 29741 oraciones. Cabe mencionar que los 40 textos usados en la prueba, no son parte del corpus.

La comparación de los métodos antes descritos se llevó a cabo tomando como punto de referencia el texto original. En esencia, se comparan las posiciones de las oraciones de un extracto ( $Ext$ ) con las posiciones de esas oraciones en el texto original. Así, cuando las posiciones de las oraciones extraídas por cada método son cercanas con respecto al texto original, entonces su representatividad es similar, y podemos concluir que los métodos obtienen resultados semejantes. Con esta idea se aplicaron funciones de cercanía ( $\mathcal{E}^*_{ora}$ ), a cada oración de  $Ext$  con cada  $O_i \in Ext$ . Esta medida se basa en la diferencia de la posición  $i$  de  $O_i \in Ext$  con la posición  $j$  de  $O_i$  en  $T_c$ . En suma, la valoración de una oración del extracto con respecto del texto original es:

$$\mathcal{E}^*_{ora}(T_c, O_i) = \frac{1}{|j - i| + 1}. \quad (9)$$

La evaluación del extracto del documento se obtiene promediando:

$$\mathcal{E}^*_{doc}(T_c) = \frac{\sum_{i=1}^5 \mathcal{E}^*_{ora}}{5}, \quad (10)$$

donde 5 fue un número arbitrario de oraciones que se tomaron para definir *Ext*.

En la siguiente tabla se muestran los resultados obtenidos después de promediar la evaluación de cada extracto de las 40 oraciones aplicando los métodos *A* y *B*.

**Tabla 1.** Semejanza relativa de los métodos

Método	<i>Cer<sub>ora</sub></i> de Ext					<i>Cer<sub>doc</sub></i>
	<i>O</i> <sub>1</sub>	<i>O</i> <sub>2</sub>	<i>O</i> <sub>3</sub>	<i>O</i> <sub>4</sub>	<i>O</i> <sub>5</sub>	
<b>A</b>	<b>0.190</b>	<b>0.126</b>	<b>0.172</b>	<b>0.211</b>	<b>0.198</b>	<b>0.179</b>
<b>B</b>	<b>0.235</b>	<b>0.184</b>	<b>0.151</b>	<b>0.133</b>	<b>0.173</b>	<b>0.175</b>

En la tabla 1 observamos un valor promedio de cercanía semejante para *A* y *B*. De la misma forma se calculó la semejanza entre los extractos (ec. 9). El resultado de este cálculo fue 0.66, lo cual nos permite afirmar que la eficacia de ambos métodos es similar y, por tanto, puede usarse con mayor ventaja el método *A*.

## 5. Conclusiones

Se compararon dos métodos para obtener el extracto de un texto. Los extractos arrojados por el sistema fueron evaluados con respecto a aquellos obtenidos por jueces humanos, sobre los mismos textos. Ambos métodos usan un corpus con el fin de representar las oraciones y, así, asignarles una puntuación sobre su representatividad acerca del texto. El método *A* requiere un tiempo en  $O(n)$  y el método *B*,  $O(n^2)$ , dado un texto de  $n$  oraciones. Que el encabezado sea revelador del contenido del texto es una hipótesis del método *A*. Mientras que *B* no depende del encabezado. Al aplicar una medida de cercanía con respecto al texto original fue posible constatar, en una prueba que usó cuarenta textos, que ambos métodos obtienen extractos similares. Se concluye, entonces, que el método basado en el encabezado del texto es ventajoso con respecto al tiempo de respuesta, ya que tiene una complejidad en tiempo lineal.

## Agradecimientos

Agradecemos los comentarios de los árbitros de este trabajo, así también el apoyo parcial recibido por parte del proyecto VIEP III 9-04/ING/G.

## Referencias

- [1] Ruge, Gerda: Combining corpus linguistics and human memory models for automatic term association, *Text information retrieval*, T. Strzalkowski (Ed.), Kluwer, 1999.
- [2] Lyons, J.: *Semantics*, Cambridge University Press, 1977.
- [3] Daniel Marcu: "The Automatic construction of large-scale corpora for summarization research", ACM-SINGIR '99, pp. 137-144, 1999.
- [4] F.C. Johnson, C.D. W.J. Paice, Black & A.P. Neal "The application of linguistic processing to automatic abstract generation", *SIGIR '94*
- [5] Gerard Salton, James Allan & Amit Singhal: "Automatic text decomposition and structuring", *Information Processing and Management*, V. 32, pp. 127-138, Elsevier, 1996.
- [6] Grefentette, Gregory: "Automatic thesaurus generation from raw text using knowledge-poor techniques", "Xerox", Grenoble Lab., 1995.
- [7] Gustavo Crispino, Jean-Luc Minel & Javier Couto "Contexto: Una plataforma para la extracción de información y el resumen automático de textos", *Proceedings of the 2nd. Workshop on Spanish Processing and Language Technologies*, pp. 153-157, Universidad de Jaén, España, 2001.
- [8] I. Acero, M. Alcojor, A. Díaz, J.M. Gómez & M. Maña "Generación Automática de Resúmenes personalizados", *Procesamiento de Lenguaje Natural 27* ", pp. 281-298, SEPLN 2001.
- [9] Salazar, H., Pinto, David & Jiménez, H. "Text extraction: a corpus-based approach", XXX Aniversario FCC-BUAP, noviembre 2003. ISBN: 968 863 711 4, pag. 92-94, 2003.
- [10] Salazar, H. *Obtención del extracto de un texto*, Tesis de Licenciatura en Ciencias de la Computación, FCC-UAP, 2003.
- [11] Kenneth Ward Church & Patrick Hanks "Word Association Norms, Mutual Information", *Computational Linguistics* Volume 16, Number 1, March 1990.
- [12] Kevin Knight & Daniel Marcu "Summarizing beyond sentence extraction: A probabilistic approach to sentence comprensión", Elsevier 2002, pp. 91-107, 2002.
- [13] Varaschin Gasperin, C. & Strube de Lima, V.L.: "Experiment on extracting semantic relations from syntactic relation", *Lecture Notes in Computer Science* ", Vol. 2588, Springer, 2003.

# Enhancing Flexibility of Named Entity Classifiers

Thamar Solorio and A. López López

Coordinación de Ciencias Computacionales  
Instituto Nacional de Astrofísica, Óptica y Electrónica  
Luis Enrique Erro #1,  
72840 Puebla, México

**Abstract.** Traditional methods for named entity classification are based on hand-coded grammars, lists of trigger words and gazetteers. While these methods have acceptable accuracies they present a serious drawback: if we need a wider coverage of named entities, or a more domain specific coverage we will probably need a lot of human effort to redesign our grammars and revise the lists of trigger words or gazetteers. We present here a method that uses Support Vector Machines for improving flexibility and accuracy of a traditionally-built named entity extractor. Experimental results show that this approach can be a very practical solution, increasing precision by up to 13.56% and recall by up to 38.41% without considerable human effort.

**Resumen.** Los métodos tradicionales para clasificación de entidades nombradas se basan en gramáticas codificadas manualmente, listas de palabras disparadoras y diccionarios. Aunque la exactitud de estos métodos es aceptable presentan una seria desventaja: si se requiere una cobertura más amplia o adaptar la cobertura de entidades a un dominio más específico probablemente se necesitaría de un gran esfuerzo humano para rediseñar las gramáticas y revisar las listas de palabras disparadoras y diccionarios. En este trabajo se presenta un método basado en Máquinas de Vectores de Soporte que facilita la adaptación de un extractor tradicional de entidades nombradas. Los resultados experimentales muestran que el método es una solución muy práctica, que no requiere de un trabajo manual pesado, obteniendo mejoras en precisión de hasta un 13.56% y en recuerdo hasta un 38.41%.

## 1 Introduction

While hand-coded grammars, gazetteers, chunkers, contextual rules and lists of trigger words provide a valuable source of information useful for building NE extractors, (e.g. [1–4]) they can become obsolete if no updating is performed. Another disadvantage of relying on this information is that the coverage of these tools might be too general or overly specific, thus achieving poor precision and recall when applied to more specific or different domains. However, they can

present a useful starting point in building accurate Named Entity (NE) classifiers. We believe that machine learning techniques can be used to build automated classifiers trained on traditional hand-built NE extractors. Then, instead of manually redesigning the NE extractors we can allow the classifiers to learn from tags assigned by the NE extractor. Hopefully the NE not covered by the extractor will be successfully classified by the learner, providing a faster/easier method for adapting named entity extractors.

We present here a new methodology for NE classification that uses Support Vector Machines (SVM) in order to enhance the flexibility and accuracy of a NE Extractor System (NEES). The NEES is considered as a black box, we are only interested in its output, which is used as one of the attributes in our learning scenario. Our proposed solution can be considered as a stack of classifiers where in the first stage a traditional hand-built NEES is used to assign possible tags to the corpus, then these tags are used by a SVM classifier to obtain the final NE tags. In addition, other attribute information used are the class values assigned to the 2 words to the left and right of the instance. Experimental results show that Support Vector Machines are successfully applied to this learning task increasing F-measure and accuracy.

## 2 Learning NE Classifiers

In this section we describe the setting of our learning scenario and present a brief introduction to Support Vector Machines.

As mentioned previously, we build our NE classifiers using the output of a NEES. Our assumption is that by using machine learning algorithms we can adapt NE extractors to different domains without a considerable effort, as opposed to that involved in extending or redesigning grammars and lists of trigger words and gazetteers.

In order to build a training set for the learner, each instance  $n$  is described by a vector of six attributes,  $\langle a_1, a_2, \dots, a_6 \rangle$ , where  $a_1$  and  $a_2$  are the classes assigned by the NEES to the first two words to the left of  $n$ ,  $a_3$  and  $a_4$  are the classes for the first two words to the right of  $n$ ,  $a_5$  is the class value of  $n$  predicted by the NEES and  $a_6$  is the true class value of  $n$ . Note that  $a_5$  and  $a_6$  will differ only when the base NEES misclassifies a named entity. There are two types of possible errors from the NEES that the learner will try to overcome. The first type occurs when the NEES assigns the wrong NE class to the NE; the other type occurs when the NEES fails to recognize the NE. When this is the case the NEES assigns a POS tag to the NE, although the same thing happens when the NEES needs to classify a non-NE word. It follows that the possible values for attributes  $a_1$  to  $a_5$  can take any value from the set of possible NEs, this is  $NE = \{Per, Loc, Org, Misc\}$ , plus the set of possible POS tags, which has over one hundred different tags; so the size of the feature space of our NE classification task is  $5 \times 10^{10}$ .

Having the high-dimensionality feature space described above has some serious drawbacks: the number of training examples needed in order to achieve

good accuracy is proportional to the size of the feature space; but increasing the amount of training examples to meet this requirement might be unfeasible. Another disadvantage is the high computational resources needed to use SVM with a feature space such as this.

In order to overcome these difficulties we decided to reduce the size of the feature space. We achieve this by generalizing the POS tags, i.e. instead of having tags for all the possible kinds of pronouns, we kept one tag  $P$  that encapsulates the set of pronouns. We did the same for each POS. Table 1 shows the resulting reduced set of feature values which has a size of  $16^5 * 4 = 4,194,304$ .

Feature Value	Description
Per	Person
Org	Organization
Loc	Location
Misc	Miscellaneous
N	Noun
A	Adjective
P	Pronoun
F	Punctuation mark
D	Determiner
V	Verb
S	Preposition
R	Adverb
T	Article
C	Conjunction
M	Numeral
I	Interjection

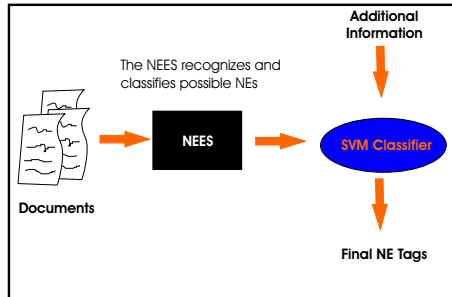
**Table 1.** Reduced Feature Values Set

A graphical representation of our proposed NE classifier is given in Figure 1. We can see that a NEES is used as a black box, the output of which is fed to the SVM classifier. The SVM classifier uses this information (in some cases also additional information such as POS tags, capitalization information and lemmas) and assigns the final NE classification. For information about SVM we refer the reader to [5, 6].

### 3 Experimental Results

Having introduced our proposed solution we continue describing in this section the experimental setting used to evaluate it. We begin by describing the data sets used, and continue presenting the results achieved. In our experiments we used the WEKA implementation of SVM [7].

Two data sets were used in the experiments, one was gathered by people in the NLP lab at our Institution. It consists of news in Spanish acquired from



**Fig. 1.** A graphical representation of our NE classifier.

different newspapers from Mexico that cover disaster-related events. This collection contains a total of 285 NEs. Although it is a small corpus (which is not so bad as it gives us the opportunity of manual revision) it is large enough for our experimental evaluation. The other corpus is that used in the CoNLL 2002 competitions for the Spanish NE extraction task. This corpus is divided in three sets: a training set consisting of 20,308 NEs and two different sets for testing, *testa* which has 4,634 NEs and *testb* with 3,948 NEs, the former was designated to tune the parameters of the classifiers (development set), while *testb* was designated as the one used to compare the results of the competitors. As in our setting there is no parameter tuning we performed experiments with the two sets.

In order to evaluate our solution we used the Named Entity extractor developed by Carreras and Padró [8]. They have developed a set of NLP analyzers for Spanish, English and Catalan that include practical tools such as POS taggers, semantic analyzers and NE extractors. This NEES is based on hand-coded grammars and lists of trigger words and gazetteer information.

Data Sets	Precision	Recall	$F_1$	Accuracy
Baseline	<b>90.80%</b>	66.2%	76.62%	62.10%
SVM Features	78.60%	89.95%	83.91%	73.38%
SVM Features+NEE	87.02%	90.83%	88.88%	80.00%
SVM NEE (*)	88.30%	<b>91.73%</b>	<b>89.96%</b>	<b>81.75%</b>
% of improvement (best vs baseline)	-2.95%	38.41%	17.41%	31.64%

**Table 2.** Results of NE categorization with our data set. The last row shows the percentage of highest improvement comparing the best result with SVM, marked with "\*\*", against the baseline.

We begin describing the results using our disaster corpus. We perform four different experiments and report the results in Table 2. We experimented here using 10-fold cross-validation. The first experiment, labeled Baseline is the result



of the NEES. As can be seen, it has very high precision, while recall and thus F-measure, are not so good. The Baseline system achieved an accuracy of 62.10%. A different experiment was performed using SVM trained on features like POS tags, lemma and capitalization information. These features are acquired automatically from the text using the Spanish analyzers mentioned above. Results from this experiment are also in Table 2 under label SVM Features. We can see that even though we are not using the NEES tags we achieve higher recall, F-measure and accuracy than the NEES. The results named SVM Features+NEE were obtained using the same features described in the previous experiment plus the NE tags assigned by the extractor system. This combination of features outperformed the previous results in all but one figure, achieving an accuracy of 80%. The last experiment performed with this corpus uses as features the output of the NEES, it does not use POS tags or additional information. Results are labeled SVM NEE, also shown in Table 2. As we can see, the methods using SVM outperformed the NEES in recall, F-measure and accuracy, the best results being those from training SVM on the extractor system tags. Precision of the NEES was the only figure that remained higher. However the improvements achieved by using SVM are as high as 38.41% in recall, 17.41% in F-measure and 31.64% in accuracy.

In Tables 3 and 4 we present results using the corpora from the CoNLL 2002 competition. In both tables we used the designated training set to build the classifiers. The first table shows the results of using the development set for testing. It can be seen that we achieve higher precision, F-measure and accuracy by using the NEES tags and additional features (POS tags, capitalization and lemmas), while the NEES by itself has the highest recall for this set. Table 4 shows the results of testing with the test set of CoNLL 2002, the SVM Feature+NEES classifier outperformed the Baseline method in three figures: precision, F-measure and accuracy.

<b>Data Sets</b>	<b>Precision</b>	<b>Recall</b>	$F_1$	<b>Accuracy</b>
Baseline	77.4%	<b>95.49%</b>	85.48%	74.64%
SVM Features+NEE (*)	<b>86.00%</b>	92.27%	<b>88.97%</b>	<b>80.14%</b>
SVM Features	67.00%	83.02%	74.12%	58.89%
SVM NEE	85.60%	91.31%	88.36%	79.15%
% of improvement (best vs baseline)	11.11%	-3.37%	4.08%	7.36%

**Table 3.** Results of NE categorization for CoNLL 2002 *development* set. The last row shows the percentage of highest improvement comparing the best result with SVM, marked with ”\*”, against the baseline.

Data Sets	Precision	Recall	$F_1$	Accuracy
Baseline	70.50%	<b>88.08%</b>	78.33%	64.38%
SVM Features+NEE (*)	80.06%	87.40%	<b>83.86%</b>	<b>72.21%</b>
SVM Features	71.60%	81.55%	76.23%	61.60%
SVM NEE	<b>80.20%</b>	86.57%	83.28%	71.35%
% of improvement (best vs baseline)	13.56%	-0.77%	7.05%	12.16%

**Table 4.** Results of NE categorization for CoNLL 2002 *test* set. The last row shows the percentage of highest improvement comparing the best result with SVM, marked with "\*", against the baseline.

## 4 Discussion

Most approaches to NE classification involve the use of hand-coded grammars, lists of trigger words and gazetteers. While the accuracies of such systems may be acceptable in some domains, it is very likely that their accuracy will suffer when used to classify NEs in documents from a more specific subject. Consider then the effort of extending the coverage of these approaches: we need to revise the grammar, or regular expressions, to adapt them in order to cover new NE instances. In addition, it is very likely that the lists of trigger words will need some changes, together with gazetteers. We have presented here an alternate solution where instead of redesigning internally the NEEs for new instances we need only to use a machine learning classifier, a SVM in this case, using the outputs of the current NEEs correcting them if pertinent and retrain. Our experimental results show that we can increase the overall performance of the NEEs with this approach as assessed by F-measure and accuracy. We observed from the experiments that there is a minor penalty either in precision or recall.

## 5 Acknowledgements

We would like to thank CONACYT for partially supporting this work under grants 166934 and U39957-Y.

## References

1. M. Arévalo, L. Márquez, M.A. Martí, L. Padró, and M. J. Simón. A proposal for wide-coverage spanish named entity recognition. *Sociedad Española para el Procesoamiento del Lenguaje Natural*, (28):63–80, May 2002.
2. P. Velardi, P. Fabriani, and M. Missikoff. Using text processing techniques to automatically enrich a domain ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems*, pages 270–284. ACM Press, 2001.
3. G. Zhou and J. Su. Named entity recognition using an hmm-based chunk tagger. In *Proceedings of ACL'02*, pages 473–480, 2002.
4. Radu Florian. Named entity recognition as a house of cards: Classifier stacking. In *Proceedings of CoNLL-2002*, pages 175–178. Taipei, Taiwan, 2002.

5. V. Vapnik. *The Nature of Statistical Learning Theory*. Number ISBN 0-387-94559-8. Springer, N.Y., 1995.
6. M. O. Stitson, J. A. E. Wetson, A. Gammerman, V. Vovk, and V. Vapnik. Theory of support vector machines. Technical Report CSD-TR-96-17, Royal Holloway University of London, England, December 1996.
7. I. H. Witten and E. Frank. *Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 1999.
8. X. Carreras and L. Padró. A flexible distributed architecture for natural language analyzers. In *Proceedings of LREC'02*, Las Palmas de Gran Canaria, Spain, 2002.

# Estudio Comparativo de Traductores Automáticos para QA multilingüe

Rita M. Aceves<sup>1</sup>, Luis Villaseñor<sup>1</sup> & Manuel Montes<sup>2</sup>

<sup>1</sup> Laboratorio de Tecnologías del Lenguaje  
Instituto Nacional de Astrofísica, Óptica y Electrónica  
Sta. María Tonantzintla, Puebla, México  
{rmaceves,villasen}@inaoep.mx

<sup>2</sup> Departamento de Sistemas Informáticos y Computación  
Universidad Politécnica de Valencia, España  
mmontes@dsic.upv.es

**Resumen.** El presente trabajo muestra los resultados hasta ahora alcanzados al usar cuatro traductores automáticos disponibles en la Web dentro de la problemática de la búsqueda de respuestas multilingüe. Los traductores fueron usados en la traducción de la pregunta, en nuestro caso del inglés al español, para posteriormente entregar dicha traducción a un sistema de búsqueda de respuestas en español para dominios abiertos en la Web. Estos primeros resultados, a pesar de realizarse sobre un corpus pequeño, son prometedores y comparables con el estado del arte en este campo.

## 1 Introducción

Un sistema de búsqueda de respuestas (BR) puede ser descrito como un tipo particular de motor de búsqueda que permite al usuario hacer las preguntas usando lenguaje natural en vez de un lenguaje de consultas artificial. Un sistema de BR encuentra la respuesta exacta en una colección de documentos, mientras que los buscadores clásicos regresan un conjunto de documentos enteros que tienen que ser analizados por el usuario.

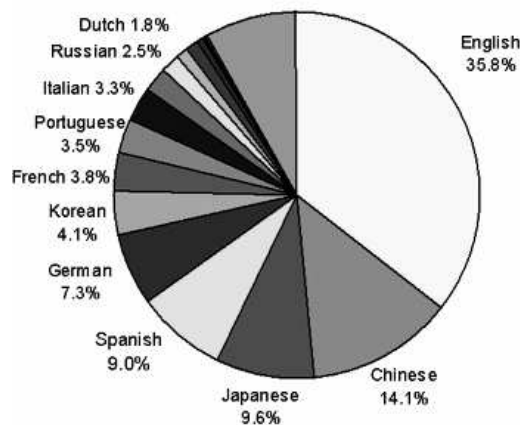
El interés de un sistema de búsqueda en la Web es encontrar las respuestas en una gran colección de documentos. Debido a que no existe sólo un lenguaje en la Web (véase la figura 1), los sistemas de BR multilingüe, son de gran relevancia. Un sistema de BR multilingüe, es aquel sistema donde el lenguaje de la pregunta es diferente al lenguaje del documento en el que se encuentra la respuesta. Este artículo presenta los resultados preliminares de un sistema de BR multilingüe para dominios abiertos sobre la Web, el cual responde preguntas factuales, formuladas en inglés, cuya respuesta se encuentra en documentos escritos en español.

En la actualidad los trabajos desarrollados en este contexto, usan diferentes recursos lingüísticos, como etiquetadores POS, extracción de entidades, relaciones semánticas, analizadores sintácticos, diccionarios, etc.; para, de este modo, entender la pregunta y las secciones concordantes en los documentos. Nuestra propuesta se

basa en un sistema prototipo de búsqueda de respuestas monolingüe para la Web, desarrollado en el laboratorio de Tecnologías del Lenguaje de la Coordinación de Ciencias Computacionales del INAOE [Castillo et al. 04]. Este sistema únicamente procesa preguntas formuladas en español, y la búsqueda se realiza en documentos en español. El sistema sólo se fundamenta en la redundancia de información en la Web [Brill et al. 01].

La idea básica de este trabajo es adecuar este prototipo, para poder realizar búsquedas multilingües. Para ello hemos realizado un estudio del comportamiento de algunos traductores automáticos disponibles en la Web. El objetivo de este estudio fue conocer hasta que punto influye la calidad de la traducción de las preguntas en un sistema de búsqueda de respuestas de este tipo, así como las debilidades de los traductores para esta tarea.

En la siguiente sección de este artículo explicaremos en detalle la arquitectura de nuestro sistema de BR multilingüe. En la sección 3 mostraremos los resultados obtenidos con el sistema; y por último, discutiremos acerca de las perspectivas de este trabajo.



**Fig. 1.** Lenguajes existentes en la Web (Marzo 2004), en un total de 729 millones de páginas Web. Fuente: <http://www.glgreach.com/globstats>

## 2 Arquitectura

La figura 2 muestra la arquitectura general de nuestro sistema de BR. Ésta consiste de cuatro módulos principales:

1. Traducción de las preguntas
2. Búsqueda en la Web
3. Cálculo de respuestas candidatas
4. Selección de la respuesta correcta.

Cabe mencionar que hasta el momento aún no se tiene completamente implementado el cuarto módulo; los resultados reportados en este trabajo fueron calculados al inspeccionar las respuestas candidatas manualmente. A continuación describiremos en detalle los tres primeros módulos.

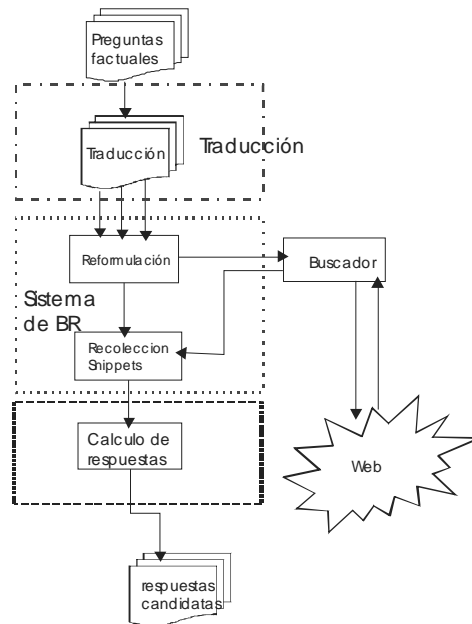


Fig. 2. Diagrama de la Arquitectura de nuestro sistema.

## 2.1 Traducción de las preguntas

Nuestro experimento intenta responder preguntas en inglés, realizando la búsqueda en documentos escritos en español disponibles en la Web. Así, el primer paso consiste en la traducción de las preguntas del inglés al español.

Actualmente se pueden encontrar varios traductores disponibles en la Web. Como se explicó en párrafos anteriores, uno de los objetivos de este trabajo es analizar la aplicación de estos traductores en la tarea de búsquedas de respuestas multilingües. De esta manera, se eligieron cuatro diferentes traductores: Google, Lycos, Webtranslation, Freetranslation. La selección de estos cuatro traductores se basó únicamente en su disponibilidad y facilidad de acceso en la red.

Como resultados de la traducción se obtuvieron cuatro conjuntos de preguntas todos ellos en español. Cada uno de estos conjuntos es resultado de aplicar uno de los traductores seleccionados sobre el conjunto de preguntas en inglés. La calidad de las traducciones obtenidas cubren todas las posibilidades desde las traducciones completamente correctas hasta las muy deficientes (véase la Tabla 1).

Cabe recordar que el tipo de preguntas hasta ahora tratadas en los sistemas de BR son básicamente preguntas cuya respuesta es un hecho: el nombre de una persona, una fecha, un lugar, etc. Sin embargo, la complejidad sintáctica de la pregunta no está limitada de ninguna forma.

Pregunta original	Traducciones
How much does a pill of extasy cost in Holland?.	¿Cuánto cuesta una píldora de “extasy” en Holanda?
	¿Cuánto hace una píldora del coste extasy en Holanda?
	¿Cuánto hace una píldora del coste extasy en Holanda?
	¿Cuánto hace una píldora del costo de extasy en Holanda?

**Tabla 1.** Ejemplos de la traducción de una pregunta

## 2.2 Búsqueda en la Web

Cada conjunto de preguntas fue utilizado para iniciar la búsqueda de respuestas en la Web. A continuación se explica brevemente el funcionamiento de los dos módulos principales de este sistema de búsqueda:

- Reformulación de la pregunta
- Recolección de extractos

### 2.2.1 Reformulación de la pregunta

Dada una pregunta, este módulo genera un conjunto de posibles reformulaciones de la pregunta. Las reformulaciones tienen por objetivo ampliar las posibilidades de recuperación de extractos pertinentes. Estas reformulaciones son palabras o secuencias de palabras que podrían ser usadas para enunciar la respuesta esperada.

El sistema puede realizar diferentes reformulaciones, “bolsa de palabras”, “corrimiento del verbo”, “componentes” y “componentes sin la primera y segunda palabra” [Castillo et al. 04], pero por ahora, sólo usamos la llamada “bolsa de palabras”. Por ejemplo, para la pregunta *¿Quién obtuvo el premio Nóbel de la paz en 1992?*, se obtiene la siguiente bolsa de palabras al eliminar las palabras vacías: *Obtuvo premio Nóbel paz 1992.*

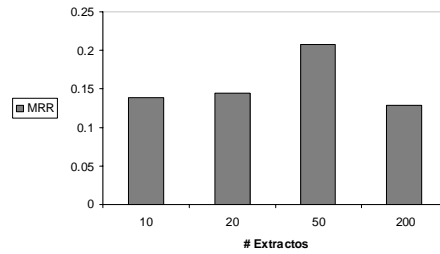
### 2.2.2 Recolección de extractos

Una vez que tenemos el conjunto de reformulaciones, se deben recolectar un número específico de extractos o “*snippets*”. Para ello usamos un motor de búsqueda en la Web, en nuestro caso Google. Como es de esperarse los primeros extractos regresados por el motor son más cercanos a la petición, y los últimos son más lejanos. Esto nos llevó a realizar varios experimentos con diferente número de extractos para tratar de establecer la cantidad más adecuada. En conclusión se optó por 50 extractos como cantidad suficiente para un buen rendimiento del prototipo. Para la evaluación de estos experimentos se usó el MRR (Mean Reciprocal Rank). El MRR es comúnmente

usado en este tipo de experimentos y trata de establecer que tan cerca está la respuesta correcta de ser la primera opción de entre el conjunto de respuestas candidatas. El MRR se calcula de la siguiente forma:

$$MRR = \frac{\sum_{i=1}^n r_i}{n},$$

donde  $n$  es el número total de preguntas y  $r_i$  es el recíproco del rango de la primera respuesta correcta en el conjunto ordenado de repuestas candidatas para la pregunta  $i$ . En la gráfica 1 se muestra el comportamiento general del MRR al incrementar el número de extractos. El comportamiento fue similar para todos los traductores y para todas nuestras preguntas.



**Gráfica 1.** Cantidad de extractos suficientes para un buen rendimiento del prototipo.

### 2.3 Cálculo de las respuestas candidatas

Las respuestas candidatas son aquellas expresiones que se presentan con mayor frecuencia dentro de la colección de “*snippets*”. El primer paso es calcular todos los  $n$ -gramas (desde uni-gramas hasta pentagramas) sobre el conjunto de extractos. Posteriormente, los  $n$ -gramas son ordenados en función de su frecuencia. Mientras más frecuentes más probabilidad de ser la respuesta correcta. El método usado para clasificar los  $n$ -gramas aplica un filtro basado en expresiones regulares y compensa la frecuencia en función de su repetición en cada clase de  $n$ -gramas. A continuación se describen los pasos del método:

1. Extraer los 50 uni-gramas más frecuentes que satisfacen un criterio tipográfico predefinido, esto es, las palabras con mayúsculas, los números y los nombres de los meses.
2. Determinar de todos los  $n$ -gramas, desde bi-gramas hasta pentagramas, únicamente aquellos que contienen los uni-gramas frecuentes.
3. Medir los  $n$ -gramas basados en su *frecuencia compensada*.

La frecuencia compensada de un  $n$ -grama  $g(n) = (w_1 \dots w_n)$  se calcula como sigue:

$$P_{g(n)} = \sum_{i=1}^n \sum_{j=1}^{\frac{n}{i}} \frac{f_{j(i)}}{\sum_{x=1}^{|G_i|} f_{x(i)}},$$



donde  $G_n$  es el conjunto de todos los n-gramas de tamaño  $n$ ,  $|G_n|$  indica la cardinalidad de este conjunto,  $i(j)$  es un n-grama  $i$  de tamaño  $j$  contenido en  $g(n)$ , y  $f_{i(j)}$  es la frecuencia de ocurrencia del n-grama  $i(j)$ . Un n-grama  $i(j)$  corresponde a la palabra simple del n-grama  $g(n)$ .

4. Mostrar los mejores n-gramas como respuestas candidatas.

### 3 Resultados

Para evaluar nuestros experimentos usamos un conjunto de 30 preguntas factuales, extraídas del corpus de preguntas en inglés del CLEF\*. Las respuestas esperadas para estas preguntas son de cuatro tipos: fecha, lugar, organización y persona. Las respuestas correspondientes también están definidas en este corpus.

La Tabla 2 compara los resultados obtenidos con cada una de los traducciones generadas por los traductores automáticos usados.

Tipo Pregunta	Traductor			
	Webtranslation	Freetranslation	Google	Lycos
Fecha	0.019	0.0209	0.042	0.099
Lugar	0.263	0.263	0.375	0.250
Organización	0.402	0.400	0.225	0.225
Persona	0.324	0.221	0.208	0.208
Todas	0.208	0.201	0.203	0.170

**Tabla 2.** MRR de cada traductor por tipo de pregunta

La tabla 3 muestra un segundo experimento donde se unieron los 4 conjuntos de extractos correspondientes a cada una de las traducciones. Para hacer nuestras evaluaciones además de medir el MRR calculamos la precisión como el número de respuestas correctas encontradas entre el número total de preguntas.

	Todas las traducciones	Webtranslation
MRR	0.161	0.208
Precisión	0.697	0.545

**Tabla 3.** Resultados al usar todos los traductores

Como puede observarse aunque la precisión aumenta, el MRR disminuye. Este comportamiento lo atribuimos al hecho de que al unir todos los datos, también se agregan datos irrelevantes alterando el ordenamiento de las respuestas candidatas.

\* The Cross-Language Evaluation Forum (CLEF), <http://clef-qa.itc.it/>

Es importante recalcar que los resultados hasta ahora alcanzados son comparables con los obtenidos por sistemas actuales de búsqueda de respuesta multilingüe. Por ejemplo, podemos mencionar [Neumann et al. 03] quienes desarrollaron una estrategia para la búsqueda de respuestas multilingüe basados en la Web. Este sistema está orientado a consultas en inglés ó alemán y búsqueda en alemán. El sistema difiere del nuestro pues usa varios recursos lingüísticos para la búsqueda de la respuesta (gramáticas LTAG, identificadores de Nombres Propios, así como WordNet). La tabla 4 muestra los resultados reportados para este sistema con los tipos de pregunta lugar y persona, para un conjunto de prueba de 39 preguntas.

Tipo de Pregunta	Neumann & Feiyu	Nuestra Propuesta
Lugar	0.216	0.375 (Google)
Persona	0.135	0.324 (Webtranslation)

**Table 4.** Tabla comparativa con los resultados de [Neumann et al. 03]

#### 4 Análisis de resultados y trabajo futuro

En general los resultados de este experimento resultaron ser alentadores, sobretodo si se toma en cuenta que nuestro método no se apoya en recursos lingüísticos adicionales a la traducción. Por otro lado, fue sorprendente comprobar que, a pesar de la mala calidad de las traducciones automáticas, es posible usar esta información para la búsqueda de respuestas. Por supuesto, aun falta realizar pruebas más completas, ampliando el conjunto de preguntas de prueba con otras reformulaciones, así como ampliar la cobertura a otro tipo de preguntas.

Entre los problemas que observamos de nuestro método podemos mencionar el tratamiento de las entidades nombradas. En estos casos, si la traducción de la entidad no es correcta la búsqueda se desvía completamente. Por ejemplo, en la pregunta: *Which is the name of Kim Sung II?* (¿Cuál es el nombre del hijo de Kim Sung II?) la entidad “*Kim Sung II*” fue incorrectamente transformada a “*cantado de Kim II*”. La misma situación se presenta con el caso de los acrónimos. Por ejemplo, la pregunta: *Who is the president of the FIAT?* (¿Quién es el presidente de la FIAT?), la traducción fue incorrecta en todos los traductores usados consiguiendo transformaciones como: ¿*Quién es el presidente de MANDATO?* Por supuesto, con este tipo de traducciones es imposible encontrar la respuesta correcta.

Además, como era de esperarse la calidad de la traducción impacta la búsqueda. En nuestro segundo experimento al reunir todos los extractos provenientes de las cuatro traducciones el MRR disminuyó. Básicamente la calidad de la traducción influye en los extractos recuperados. En este punto dos ideas se desean explorar. Por un lado, intentar determinar cual es la mejor traducción (o las dos mejores traducciones) utilizando un modelo de lenguaje. Así se tomarían sólo los extractos de aquellas traducciones con menor perplejidad. Por otro lado, utilizar un sistema de agrupamiento sobre los extractos, de manera a aplicar el cálculo de las respuestas candidatas únicamente sobre un subconjunto escogido de tales extractos.

Por último, como se mencionó aún falta realizar pruebas con un corpus mayor y considerando otros tipos de preguntas. Con los resultados hasta ahora alcanzados suponemos que serán necesarios filtros para el cálculo de las respuestas candidatas dependientes del tipo de respuesta esperada. Por ejemplo, se usará un filtro específico para el cálculo de fechas el cual explotará los elementos tipográficos propios de este tipo de información.

## 5 Agradecimientos

El presente trabajo fue parcialmente financiado por el CONACYT (Proyecto 43990 y beca No. 184663). Por su parte, el tercer autor agradece a la Secretaría de Estado de Educación y Universidades, España.

## 6 Referencias

- [Brill et al. 01] Brill E., Lin J., Banko M., Dumais S. and Ng A.(2001). Data-intensive question answering. In Proceedings of the Tenth Text REtrieval Conference (TREC 2001).
- [Castillo et al. 04] Alejandro del Castillo, Manuel Montes, Luis Villaseñor (2004), Encuentro Internacional de Ciencias de la Computación 2004, por publicar.
- [Neumann et al. 03] Neumann Günter, Feiyu Xu, Sacaleanu Bogdan (2003), Strategies for Web-based Cross-Language Question Answering. In proceedings of 2nd CoLogNET-ElsNET Symposium on *Questions and Answers: Theoretical and Applied Perspectives*, Amsterdam, 18 Diciembre, 2003, pp. 84-95.

# Formalización del Sistema de Nombres Hispanos

Noé Alejandro Castro Sánchez, José Ángel Vera Félix,  
Igor A. Bolshakov, Grigori Sidorov

Centro de Investigación en Computación (CIC),  
Instituto Politécnico Nacional (IPN)  
e-mail: {ncastro, jvera}@sagitario.cic.ipn.mx, {igor, sidorov}@cic.ipn.mx  
Av. Juan de Dios Batíz s/n, esq. Miguel Othón de Mendizabal,  
Unidad Profesional "Adolfo López Mateos", Edificio CIC.  
Col Nueva Industrial Vallejo, CP 07738, México, D. F.

**Resumen.** Determinamos la estructura de los nombres hispanos y analizamos la interacción de sus elementos en la formación de secuencias nominales correctas. Proponemos una gramática generativa para su descripción y un algoritmo para su identificación en textos.

**Abstract.** We define the structure of Hispanic names and analyze the interaction of their elements in the formation of the correct nominal sequences. We propose a generative grammar for their description and an algorithm for their identification in texts.

**Palabras Clave:** Secuencias Nominales Hispanas, Sistema de Nombres Hispanos, Gramática Generativa.

## 1 Introducción

Es de gran interés conocer la forma en que se combinan nombres y apellidos hispanos para constituir lo que nosotros denominamos **Secuencias Nominales (SN) Hispanas**, dado que, de esta manera, es posible reconocer nombres humanos en textos electrónicos, logrando mejorar la calidad de los procesos como la identificación de las palabras clave en los documentos [1], seguimiento de los entidades y actores en las historias [2] o noticias [9, 10] o etiquetado de las categorías gramaticales [11], así como automatizar procesos como la recuperación de trabajos científicos de bases de datos o Internet a través del análisis del título de los trabajos identificando los apellidos de los autores; o bien en el control automático de nombres de ciudadanos realizados por organismos oficiales, o bien la identificación automática de los nombres de las personas en las consultas a las bases de datos [12].

El objetivo de este artículo reside en proponer una gramática generativa que indique las reglas en las que se basa la formación de SN, y la realización de un algoritmo capaz de aplicar dichas reglas en la identificación de SN en documentos de texto electrónicos.

## 2 Estructura de las Secuencias Nominales

La SN que presenta una Secuencia de Nombres y una Secuencia de Apellidos, la denominamos SN Completa. Precisemos más a detalle cada una de éstas:

### 2.1 Secuencia de Nombres

- **Nombres Solos:** Son los elementos más simples (*José, Jazmín, Ramón*, etc.). Se dividen en Nombres Solos Masculinos (*Humberto, Luis*, etc.) y Nombres Solos Femeninos (*Jimena, Eva*, etc.).
- **Nombres Compuestos:** Surgen de la concatenación de un Nombre Solo y un Complemento Deado, el cual se forma por un Nombre Solo antecedido por la preposición *de* articulada (*del Carmen, de Jesús, del Refugio*, etc.). Son Nombres Compuestos *María de la Luz, José del Refugio, Andrea de los Ángeles*, etc.
- **Nombres Combinados:** Resultan de la concatenación de dos Nombres Solos de sexo opuesto. Ejemplo de tales estructuras son *José María, María José, José Trinidad*, etc., pero no se encuentran *Alina Raúl, Francisco Mariana*, etc.

### 2.2 Secuencia de Apellidos

En el sistema de nombres hispanos, los apellidos se conforman por dos grupos de elementos, denominados Apellido Paterno y Apellido Materno.

- **Apellidos Solos:** Representan los apellidos más simples y comunes (*Romero, Ávila, Castrejón, Gutiérrez* etc.).
- **Apellidos Deado:** Son los Apellidos Solos antecidos por la preposición *De* articulada (*Del Valle, De León, De los Cobos*, etc.).
- **Apellidos Deado Inseparables:** Surgen de la concatenación de dos palabras través de la preposición *de*, pudiendo ser una de éstas un Apellido Solo (*Montes de Oca*, donde *Montes* existe como apellido Solo, pero no *de Oca*).
- **Apellidos Articulados:** Constituyen palabras (no existentes como Apellidos Solos) precedidas por algún determinante (*La Rubia, La Moneda, La Chica*, etc.).
- **Apellidos existentes como Nombres:** Agrupa a todos aquellos elementos que existen como apellidos y nombres a la vez (*Santos, Félix, Santiago*, etc), o como apellidos y Complementos Deado (*De los Santos, De Román*, etc.).
- **Apellidos Compuestos:** Son estructuras creadas arbitrariamente, como *Robles y López, Ortíz-Landa, Tejada-y-Olivas, Romero Ovando*, formando cada agrupación un único apellido.

### 3 Gramática Generativa

Para dar inicio a nuestra gramática, en la siguiente tabla mostramos expresiones a los conceptos antes mencionados, y que serán utilizados en los siguientes capítulos:

**Tabla 1.** Expresiones a conceptos ya definidos

SecNml	Secuencia Nominal	ApPat	Apellido Paterno
		ApMat	Apellido Materno
SecNom	Secuencia de Nombres	ApSolo	Apellido Solo
SecAp	Secuencia de Apellidos	ApDeado	Apellido Deado
		ApDeadoInsprble	Apellido Deado Inseparable
NomSolo	Nombre Solo	ApArtdo	Apellido Articulado
Ncmsto	Nombre Compuesto	ApExsteNom	Apellido existente como Nombre
NCmb	Nombre Combinado	ApCmsto	Apellido Compuesto
CplmtoDeado	Complemento Deado	Sgmto( <i>j</i> )	Segmento <i>j</i> -ésimo

#### 3.1 Reglas estructurales

Nuestra gramática queda definida de la siguiente manera:

1. SecNml → SecNom(Sex) | SecAp | SecNom(Sex) SecAp
2. SecNom(Sex) → NomSolo(1, Sex) | NomSolo(1, Sex) NomSolo(2, Sex) | NomSolo(1, Sex) NomSolo(2, Sex) NomSolo(3, Sex) | NCmsto(1, Sex) | NomSolo(1, Sex) NCmsto(2, Sex) | NomSolo(1, Sex) NCmsto(2, Sex<sup>-1</sup>) | NCmb(Sex) | NomSolo(1, Sex) NCmb(Sex)
3. NCmsto(*j*,Sex) → NomSolo(*j*, Sex) CplmtoDeado
4. SecAp → Ap(pat) | Ap(pat) Ap(mat)
5. Ap(Prgtor) → ApSolo | ApDeado | ApDeadoInsprble | ApArtdo | ApExsteNom | ApCmsto
6. ApCmsto → Sgmto(1) Sgmto(2) | Sgmto(1) y Sgmto(2) | Sgmto(1)–Sgmto(2) | Sgmto(1)–y–Sgmto(2)
7. Sgmto(*j*) → ApSolo | ApDeado | ApExsteNom | *De* ApExsteNom | *De la* ExsteNom | *Del* ExsteNom | *De los* ExsteNom

*Nota 1.* *j* es variable con valores 1 o 2.

*Nota2.* En una misma regla, se tiene que NomSolo(1, Sex) ≠ NomSolo(2, Sex), NomSolo(2, Sex) ≠ NomSolo(3, Sex) y NomSolo(1, Sex) ≠ NomSolo(3, Sex)

*Nota 3.* Sex es variable de Sexo, con valores *mas* (masculino) o *fem* (femenino). Sex, presenta el mismo valor en una regla. Sex<sup>-1</sup> es complemento (valor opuesto) de Sex.

*Nota 4.* Prgtor es variable de Progenitor, con valores *pat* (paterno) o *mat* (materno).

*Nota 5* En la regla 2, en  $\text{NomSolo}(1, \text{Sex})$   $\text{NCmsto}(2, \text{Sex}^{-1})$ , el primer elemento del Nombre Compuesto (vea regla 3) y  $\text{NomSolo}(1, \text{Sex})$  deben formar un Nombre Combinado válido, por ser de sexo opuesto.

### 3.2 Símbolos terminales

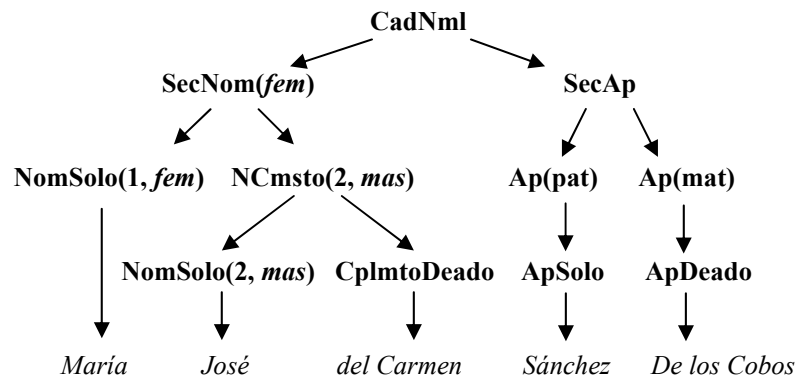
En la siguiente tabla mostramos los símbolos terminales de la gramática.

$\text{NomSolo}(i, \text{mas})$	→ <i>Noé</i>   <i>Javier</i>   <i>Omar</i>   <i>Adalberto</i>   ...
$\text{NomSolo}(i, \text{fem})$	→ <i>Rosario</i>   <i>Elva</i>   <i>Margarita</i>   <i>Jazmín</i>   ...
$\text{CplmtoDeado}$	→ <i>de la Luz</i>   <i>del Carmen</i>   <i>del Refugio</i>   <i>de Jesús</i>   ...
$\text{NCmb}(\text{mas})$	→ <i>José María</i>   <i>José Trinidad</i>   ...
$\text{NCmb}(\text{fem})$	→ <i>María José</i>   <i>María Jesús</i>   ...
$\text{ApSolo}$	→ <i>Romero</i>   <i>Paredes</i>   <i>Molina</i>   <i>Torres</i>   ...
$\text{ApDeado}$	→ <i>Del Valle</i>   <i>De León</i>   <i>De la O</i>   <i>De los Cobos</i>   ...
$\text{ApDeadoInsprble}$	→ <i>Montes de Oca</i>   <i>Cabeza de Vaca</i>   ...
$\text{ApArtdo}$	→ <i>La Rubia</i>   <i>La Chica</i>   <i>La Moneda</i>   ...
$\text{ApExsteNom}$	→ <i>Jorge</i>   <i>Santiago</i>   <i>Félix</i>   <i>Camilo</i>   <i>Alonso</i>   ...

La variable  $i$  de la fórmula  $\text{NomSolo}(i, \text{Sex})$  tiene valores 1, 2 ó 3.

### 3.3 Árboles de análisis sintáctico

Haciendo uso de árboles de análisis sintáctico, podemos mostrar gráficamente la estructura de SN. Sea la secuencia. *María José del Carmen Sánchez De los Cobos*:



## 4 Analizador Gramatical

El diseño de un analizador gramatical nos permite constatar si se cumplen o no las reglas que hemos definido en nuestra gramática. Éste analizador, debido a que es a

partir de los elementos más simples que espera conformar el nivel más alto en la gramática (el axioma), se denomina Ascendente.

#### 4.1 Denominaciones utilizadas en el programa

Las siguiente tabla muestra las denominaciones utilizadas por el programa y las definidas en nuestra gramática.

<b>Gramática</b>	<b>Programa</b>
SecNom(mas)	(SecNom,M)
SecNom(fem)	(SecNom,F)
NomSolo( <i>i</i> , mas)	(nomi,M)
NomSolo( <i>i</i> , fem)	(nomi,F)
Ap(pat)	(ap,pat)
Ap(mat)	(ap,mat)

Otras denominaciones independientes de la gramática proporcionan información más específica de los elementos analizados:

<b>Etiqueta</b>	<b>Significado</b>
( <i>SecTot, sexo, varFin</i> )	Etiqueta colocada al final de una SN.
(SecNom, <i>sexo</i> )	Variante 1. Etiqueta colocada al final de la Secuencia de Nombres
(SecNom, <i>sexo, varNom</i> )	Variante 2. Etiqueta colocada al final de la Secuencia de Nombres

Los valores y significado correspondientes a las variables señaladas en cursiva se muestran a continuación:

<b>Variable</b>	<b>Significado</b>
SecTot	Secuencia Total.
sexo	Sexo de la SecNom. Toma el valor M o F.

La variable *varFin* se coloca al final de una SN. Toma los siguientes valores:

<b>Valor</b>	<b>Significado</b>
Cmpl	[Secuencia] Completa. Indica que en la SN se aprecian Secuencias de Nombres y Secuencias de Apellidos
InCmpl	[Secuencia] Incompleta. En la SN, hace falta ya sea la Secuencia de Nombres o la Secuencia de Apellidos.
inndo	Indeterminado. En la secuencia analizada, los elementos presentes sí se reconocen como nombres pero no forman una estructura coherente.

En *varNom* encontramos los valores siguientes:



Valor	Significado
inndo	Indeterminado. Los elementos presentes sí se reconocen como nombres pero no forman una estructura coherente.
sexoCntrio	Sexo Contrario. Nombres de sexo opuesto no encontrados dentro del grupo de Nombres Combinados.
rep	Repetidos. En la Secuencia de Nombres existen nombres repetidos.
Larg	Larga. Secuencia de Nombres muy extensa (más de tres elementos).

## 4.2 Descripción del programa

Para la identificación de las SN, diseñamos diccionarios (bases de datos) de nombres y de apellidos. El sistema es totalmente dependiente de estas bases, aún cuando se utiliza una pequeña heurística para determinar Nombres Combinados, Apellidos Compuestos, etc. A grandes rasgos, el proceso seguido en el análisis es el siguiente:

En primera instancia, el sistema realiza un parseo por el texto electrónico, determinando si cada palabra encontrada pertenece o no a la base de datos; cuando se ha terminado de extraer una determinada secuencia, se aplica a ésta un segundo proceso de análisis más específico (proceso de etiquetado), que consiste en asignar a cada elemento su denominación correspondiente a la gramática elaborada, según el lugar que ocupe dentro de la SN extraída.

Algunos resultados arrojados por el sistema son los siguientes:

**Secuencia texto:** *José de Jesús Vélez De los Cobos*

**Resultado del programa:**

[ [[*José de Jesús*(nom1,M)]](SecNom,M)]  
 [*Vélez*(ap,pat)][*De los Cobos*(ap,mat)] (SecTot,M,Cmpl)

**Secuencia en texto:** *Brenda Rocío*

**Resultado del programa:**

[ [[*Brenda*(nom1,F)][*Rocío*(nom2,F)]](SecNom,F)] (SecTot,F,Incpl)

## 5 Conclusiones

Son muchos los factores que dificultan el desarrollo de un modelo que sea capaz de identificar y reconocer Secuencias Nominales Hispánicas dentro de textos electrónicos.

La ambigüedad inherente a los sistemas de nombres dificulta el proceso de modelación. No todo lo presentado en este artículo es concluyente, nuestro ámbito de análisis ha tenido que verse restringido por las variaciones tan extensas que existen en la formación de SN Hispánicas. Un ejemplo claro de esto se constata en Secuencias de Apellidos. Una práctica habitual hace tiempo era la concatenación de más de dos apellidos para constituir uno solo, por ejemplo, *Sánchez de Anda y Martínez Salgado*, *Martí y Zayas-Bazán*, *Bancés y Fernández-Criado*, etc. Obviamente, el tratar de abarcar todos estos casos (y más) resultaría en un entre marañado de reglas, en una complejidad realmente enorme.

Estos son sólo algunas características que pueden dificultar alcanzar el objetivo que se planteó en este trabajo. La más clara conclusión de esto es que un sistema de clasificación de nombres no puede ser automatizado en su totalidad, siempre será necesaria la intervención del usuario.

**Agradecimientos.** El trabajo fue realizado con apoyo parcial de CONACyT y PIFI.

## Referencias

1. Alexandrov, M., A. Gelbukh, and P. Makagonov. *A language-independent approach to evaluation of the document thematic structure*. Acta Academia 2000, Evrica, Chisinau, Moldova, 2000, pp. 177-189.
2. Calvo, H., A. Gelbukh. *Maintaining inter-sentential continuity of semantic indices with a knowledge base*. IEEE Conf. on Natural Language Processing and Knowledge Engineering, NLPKE-2003, Peking, China, IEEE Computer Society Press, p. 634–637.
3. Chen, H. H., S. J. Huang, Y. W. Ding, S. C. Tsai. (1998). *Proper Name Translation in Cross-Language Information Retrieval*. Proceedings of the 17<sup>th</sup> International Conference on Computational Linguistics. pp 232 – 236. <http://acl.ldc.upenn.edu/P/P98/P98-1036.pdf>
4. Chinchor, N. (1997). *MUC-7 Named Entity Task Definition*. Proceedings of the 7<sup>th</sup> Message Understanding Conference. [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/proceedings/ne\\_task.html](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ne_task.html)
5. Galicia-Haro, S. N., A. Gelbukh, I. A. Bolshakov. *Recognition of Named Entities in Spanish Texts*. Proceedings of International Conference MICAI'2004, Mexico City, Mexico. Lecture Notes in Artificial Intelligence, ISSN 0302-9843, Springer, 2004 (accepted).
6. Galicia-Haro, S. N., A. Gelbukh, I. A. Bolshakov. *Web-based Sources for an Annotated Corpus Building and Composite Proper Name Identification*. Proceedings of International Conference AWIC04, Cancun, Mexico. Lecture Notes in Computer Science, ISSN 0302-9843, Springer, 2004 (accepted).
7. Huang, F., A. Waibel. *An Adaptive Approach to Named Entity Extraction for Meeting Applications*. [http://www.is.cs.cmu.edu/papers/speech/HLT2002/hlt2002\\_fei.pdf](http://www.is.cs.cmu.edu/papers/speech/HLT2002/hlt2002_fei.pdf)
8. Mani, I., R. McMillian, S. Luperfoy, E. Lusher, S. Laskowski. (1993). *Identifying Unknown Proper Names in Newswire Text*. Proceedings of the Workshop on Acquisition of Lexical Knowledge from Text. pp 44 – 54. <http://acl.ldc.upenn.edu/W/W93/W93-0105.pdf>
9. Montes y Gómez, M., A. Gelbukh, A. López López. *A statistical approach to the discovery of ephemeral associations among news topics*. Proc. Database and Expert Systems Applications (DEXA-2001). Lecture Notes in Computer Science N 2113, Springer-Verlag, 2001, pp. 491-500.
10. Montes-y-Gómez, M., A. Gelbukh, A. López-López. *Detecting deviations in text collections: An approach using conceptual graphs*. Proc. Mexican International Conference on Artificial Intelligence (MICAI-2002). Lecture Notes in Artificial Intelligence N 2313, Springer-Verlag, 2002, pp. 176–184.
11. Morales Carrasco, R., A. Gelbukh. *Evaluation of TnT Tagger for Spanish*. ENC-2003, 4th Mexican International Conference on Computer Science. IEEE Computer Society Press, 2003, pp. 18–25.
12. Pazos R., R. A., A. Gelbukh, J. J. González B., E. Alarcón R., A. Mendoza M., A. P. Domínguez S. *Spanish natural language interface for a relational database querying system*. Text, Speech and Dialogue (TSD-2002). Lecture Notes in Artificial Intelligence N 2448, Springer-Verlag, 2002, pp. 123–130.

# Un Modelo de Representación basado en Sintagmas para Recuperación de Información

Miguel Rodríguez H., Héctor Jiménez S. & David Pinto

Fac. de Ciencias de la Computación

B. Universidad Autónoma de Puebla

C.U. 72570, Puebla, México

Tel. (+52222) 229 55 00 ext. 7212 Fax (+52222) 229 56 72

mrodriguez11@hotmail.com, hjimenez@fcfm.buap.mx, dpinto@cs.buap.mx

**Abstract.** Two experiments of document representation are presented in this paper. The first experiment used the term-phrase sense concept and the second one used EuroWordNet sense relationships. Additionally, the first experiment was carried out with three sets of 255 terms definitions of physics subject. In this case, each set was used to compose a corpus, a set of documents for representation and the evaluated queries. A set of 16 preprocessed queries, about 'earth' term, were executed in the second experiment, using both, the same corpus, and synonymy and synonymy-hyponymy sense relationships obtained from EuroWordNet. The results obtained show that the method that uses synonymy relationships is better than the others, despite the low value of precision.  
**Keywords.** Automatic Indexing, Term-phrase, Sense Relationship.

**Resumen.** En este trabajo se presentan dos experimentos de representación de documentos, usando por un lado el concepto del "sentido de un sintagma" y por otro las relaciones de sentido basadas en EuroWordNet. En el primer experimento se usaron tres colecciones con 255 definiciones de términos del área de física. La primera correspondió a la colección de documentos a representar, la segunda al corpus que proporcionó las relaciones de sentido y la tercera a la colección de consultas supervisadas. En el segundo experimento se efectuaron 16 consultas truncadas para el concepto tierra obtenidas de estudiantes de una Facultad de Ingeniería; en este caso, el sistema se apoyó en el mismo corpus del primer experimento y relaciones de sinonimia y sinonimia-hiponimia extraídas de EuroWordNet. Los resultados obtenidos muestran un desempeño bajo, debido principalmente al material empleado en el experimento, sin embargo se puede observar que el método que usa relaciones de sinonimia mejora los resultados sobre los otros dos métodos (corpus y sinonimia-hiponimia).

**Palabras Clave:** Indexado Automático, Sentido de un Sintagma.

## 1 Introducción

Hoy en día la demanda por información sigue creciendo al grado que la dependencia de las organizaciones e individuos sobre este recurso es crucial. Internet se

ha convertido en el principal proveedor de información y por lo tanto, juega un papel estratégico en la toma de decisiones. Sin embargo, es deseable por todos los usuarios tener acceso a la información que satisface mejor sus necesidades. Así, la construcción de buenos buscadores en esta abrumadora cantidad de textos, imágenes y otro tipo de medios representa un reto ineludible. Estamos frente al problema de hacer búsquedas de información con base en los conceptos que contiene un texto y no las palabras.

En este trabajo se experimentó con una representación de textos alternativa a la tradicional, la cual tomó en cuenta el contenido semántico de los textos con el propósito de ser utilizada en la recuperación de información. La idea principal en este trabajo es emplear el "sentido de un sintagma". Se implantó un sistema que representa un documento considerando los sintagmas que contienen los términos índices del documento. Las consultas se representan de manera semejante a la representación de un documento y la búsqueda se realiza en forma semejante a como lo hace el modelo booleano de recuperación de información.

A continuación se describe de manera sintética el contenido de las secciones que integran este trabajo. En la sección 2 se describen los trabajos relativos a relaciones de sentido, lo cual conforma la base sobre la que el presente fue realizado. En la sección 3 se plantea el método de representación de documentos con relaciones de sentido, extraídas de un *corpus* y de EuroWordNet (EWN), además, se plantea la forma en que se realizan consultas al sistema. En la sección 4 se describen las pruebas realizadas con diferentes conjuntos de consultas y diccionarios que proveen relaciones de sentido. Al final se presentan las conclusiones de este trabajo seguidas del apéndice con parte de datos y resultados tomados de las pruebas.

## 2 Antecedentes

Se han revisado varios trabajos que inciden en la noción de "asociación de términos". Esta idea se presenta en [11] vinculada con las áreas de Psicología y Lingüística. Además, a través del concepto de asociación de términos es posible hacer un planteamiento para explotar la semántica de los términos que aparecen en los textos. Por ejemplo, se ha visto que este concepto es útil en la construcción de tesoros [8], idea que efectivamente se ha comprobado en nuestro ámbito dentro del Laboratorio de Recuperación de Información de la FCC [3]. Los términos de asociación de primer orden, que son términos que rodean a una palabra en un texto, constituyen el punto de partida para representar la palabra en cuestión. Con este enfoque el consumo de recursos no es dispendioso, ya que únicamente es necesario un *corpus*; cuyo tamaño y balance influirá ciertamente en los resultados. Así, resulta atractivo investigar esta representación que es "más semántica" que la tradicional. Para ello han de realizarse experimentos diversos que respalden el valor de este enfoque.

Entre los experimentos realizados con esta representación podemos citar a aquél que clasificó frases nominales que ocurren en las respuestas a preguntas abiertas [9]. En este caso, se representaron las respuestas usando como *corpus* un

diccionario del español. Tal trabajo se sustenta en la propuesta que hace J. García [7] sobre el sentido de un sintagma. La propuesta se basa en la combinación que los seres humanos hacemos de las propiedades asociadas a los términos, de ahí la fórmula que se emplea para representar un sintagma (ver ec. 3). Otro trabajo en el que subyace la representación basada en términos de primer orden es el efectuado para obtener el extracto de un texto [12]. Cabe mencionar que hay otras propuestas realizadas por el grupo de RI de la FCC. Por ejemplo, la descrita en [5], la cual propone una forma alternativa de elección de índices. Los trabajos antedichos nos orientan a realizar una prueba de la representación con términos de primer orden y el sentido de un sintagma para conocer su efectividad en la recuperación de información.

### 3 Representación de Textos

La propuesta que se hace para representar un documento es tomar cada oración que contenga índices y utilizar la idea expresada por “el sentido de un sintagma” [7]. Por tanto, requeriremos de un medio que proporcione una aproximación a las relaciones de sentido de cada término. Hemos considerado un *corpus* para efectuar dicha aproximación. Veamos a detalle este planteamiento.

Partimos de aproximar las relaciones de sentido de una palabra  $x$  con las palabras que ocurran en los contextos de  $x$ . Tomamos como contexto de una palabra las oraciones donde ocurra la palabra. Para lograr lo anterior emplearemos un *corpus*  $C$ , y denotemos con  $C'$  el *corpus*  $C$  preprocesado; es decir una vez eliminadas las palabras cerradas y truncando sus términos. Definimos entonces las palabras  $y$  del contexto de  $x$  según la relación  $xVy$ , donde  $V$  es:

$$V = \{(x, y) | x, y \text{ ocurren en } S, \text{ con } S \in C'\}. \quad (1)$$

Debido a que los contextos son oraciones, definimos una oración como una tupla:

$$S = (t_1, t_2, \dots, t_k),$$

donde  $t_1, t_2, \dots, t_k$  son términos índices de un documento  $D$ . Debemos representar primeramente los términos índice de las oraciones con base en sus contextos obtenidos mediante la ecuación 1. Por tanto, nuestra representación  $\bar{t}$  para un término  $t$  sería:

$$\bar{t} = \{y | (t, y) \in V\}. \quad (2)$$

Un documento  $D_i$  está compuesto de oraciones  $S_{ij}$ , donde  $i$  representa al documento al cual pertenece la oración y  $j$  el número de oración. Así, podemos ver al documento  $D_1$  como:

$$D_1 = (S_{11}, S_{12}, \dots, S_{1n}).$$

De esta forma la  $j$ -ésima oración del  $i$ -ésimo documento se denotará como:

$$S_{ij} = (t_{ij1}, \dots, t_{ijn}).$$

Una vez obtenida la representación de los términos índice que integran las oraciones de un documento, mediante la ecuación 2, podemos obtener la representación  $\overline{S_{ij}}$  de una oración  $S_{ij}$ , realizando la intersección de todas las combinaciones de los términos representados y uniendo todas estas intersecciones. Esto se logra mediante la siguiente expresión[7][9]:

$$\overline{S_{ij}} = \bigcup \overline{t_{ijk}} \cap \overline{t_{ijl}}, \quad k \neq l \quad (3)$$

Así, el documento se representará por la unión de todas las oraciones representadas:

$$\overline{D_i} = \bigcup \overline{S_{ij}}. \quad (4)$$

Una consulta  $q$  al sistema deberá hacer uso de una representación que considere el significado de las consultas, al igual que se hizo con los documentos. Para realizar una consulta  $q = (q_1, q_2, \dots, q_k)$ , representamos los términos  $q_1, q_2, \dots, q_k$  de  $q$ , usando  $\overline{q_1}, \overline{q_2}, \dots, \overline{q_k}$ . Esto es equivalente a representar una oración usando la ecuación:

$$\overline{q} = \bigcup (\overline{q_i} \cap \overline{q_j}), \quad i \neq j \quad (5)$$

Finalmente, los documentos más relevante se obtienen utilizando un criterio derivado del modelo booleano: un documento representado que comparta mayor número de términos con la consulta representada tendrá mayor relevancia.

## 4 Pruebas Realizadas

Los datos con los que se realizaron las pruebas consistieron de un glosario de física de 255 vocablos<sup>1</sup> y tres definiciones por vocablo. Para cada vocablo del glosario, la primera definición se almacenó en un archivo que se usó como *corpus*. La segunda definición se almacenó en otro archivo usado como colección de documentos; se tomó como documento cada definición para formar el archivo de la colección. La tercera definición del vocablo fue almacenada en un archivo que constituyó una colección de consultas al sistema para una de las pruebas efectuadas. Todos estos archivos fueron preprocesados en la forma que se ha mencionado.

Se realizaron dos experimentos. El primero empleó la colección de 255 preguntas descrita en el párrafo anterior. El segundo, una colección de 16 preguntas dadas por estudiantes de Ingeniería Mecánica de la BUAP<sup>2</sup>. Ambos utilizaron como fuente para proveer relaciones de sentido el *corpus* (255 definiciones de física) y relaciones provistas por EuroWordNet (EWN). En el primer caso se utilizaron relaciones alternativas extraídas de EWN: sinonimia, hiponimia, hiperonimia, y la combinación de todas ellas.

<sup>1</sup> Recurso amablemente proporcionado por el GIL-II-UNAM, a través de G. Sierra M.

<sup>2</sup> Recurso amablemente proporcionado por L. Fragueta C. [6].

#### 4.1 Consultas del Glosario

Para las 255 consultas se realizaron las pruebas con las relaciones de sinonimia, hiponimia, hiperonimia y los tres tipos de relaciones contenidas en un sólo archivo para la representación que exige el método. Los resultados promediados se muestran en la tabla 2.

Como puede observarse hay una mejora sustancial usando como soporte para la representación de sintagmas solamente empleando las relaciones de sinonimia. Aunque se nota, también mejor desempeño al usar relaciones de hiponimia con respecto a todas y las de hiperonimia. Es importante notar que el empleo de un *corpus* para proveer relaciones de sentido tiene el desempeño más bajo. La alta precisión en el nivel de evocación uno, al usar sinónimos, puede observarse en la columna “Posición de la Respuesta Correcta”.

Todas estas observaciones deben aún comprobarse con mayor exhaustividad. Está por comprobarse, por ejemplo, si eligiendo términos de los contextos del *corpus* y filtrándolos con la medida información mutua hay una mejora. De igual forma, es necesario corroborar los resultados obtenidos con otras colecciones de textos.

#### 4.2 Consultas dadas por Estudiantes

Se trata de 16 consultas que son descripciones del concepto “Tierra”. La tabla 1 muestra estas consultas truncadas. Cada una de estas consultas fue hecha al sistema, utilizando el *corpus* y relaciones de sinonimia provenientes de EWN como apoyo a la representación por sintagmas. La tabla 3 muestra en resumen el promedio de los valores indicados para las dos formas de representación (*corpus* y relaciones de sinonimia), usando la fórmula 3.

Para estas consultas se realizó una prueba adicional, confrontando el modelo basado en sintagmas con el modelo vectorial. La tabla 4 presenta resultados promedios de las consultas realizadas con los dos modelos de recuperación de información. En ambos modelos se consideró el uso de sinónimos, y como se aprecia en el modelo vectorial el uso de sinónimos no provee mejora.

### 5 Discusión

En el presente trabajo se experimentó con una propuesta de representación semántica de textos con el fin de ser utilizada en la recuperación de información. Se partió de la idea del “sentido de un sintagma” y se implantó un sistema que realiza la representación considerando los sintagmas que contienen los términos índice de los textos (aquellos términos que resultan al final del preprocesamiento del texto). Además, las consultas se representan de manera semejante y la búsqueda se efectúa como lo hace el modelo booleano de recuperación de información.

Las pruebas se apoyaron con tres colecciones cada una de 255 definiciones del área de física. La primera correspondió a la colección de documentos, la segunda

al *corpus* y la tercera a la colección de consultas supervisadas. También se efectuaron 16 consultas truncadas para el concepto Tierra, obtenidas de estudiantes de la Facultad de Ingeniería Mecánica de la BUAP, apoyándose el sistema en un *corpus* y relaciones de sinonimia (extraídas de EuroWordNet).

Además de usar las relaciones de sinonimia, se utilizaron otras relaciones provenientes de EuroWordNet: hiperonimia, hiponimia y una colección que contenía a las tres anteriores.

Los resultados obtenidos muestran un desempeño absoluto bajo, debido principalmente al material empleado en el experimento. El número de elementos de la colección es de 255 y la mayoría de estos está compuesto de solamente una oración. Con respecto al número de documentos se utilizó un conjunto grande de consultas, cuyas características hacen difícil la recuperación exacta: por el origen coloquial de estas expresiones, el vocabulario de éstas es variado y, aunque se refieren al dominio, el tipo de lenguaje es informal y diferente al de los documentos. Aún así, consideramos importante realizar el experimento con este material. Todas estas observaciones afectan a ambos modelos probados.

En el caso del modelo basado en sintagamas, cuando se realiza la representación de los términos usando un *corpus* vemos que se introduce mucho "ruido" y por eso su desempeño es el más bajo. Como se espera, el empleo de sinónimos conlleva a una mejora, y al realizar la equivalencia entre términos el desempeño es casi el mismo al del modelo vectorial.

Pueden realizarse otras pruebas como la combinación de relaciones de sinonimia con hiperonimia, sinonimia con hiponimia e hiperonimia con hiponimia, aplicando la representación en diversos puntos del procesamiento dentro del Sistema de Recuperación de Información. Asimismo, experimentar con un *corpus* de mayor tamaño.

## Agradecimientos

Agradecemos los comentarios de los árbitros de este trabajo, así también el apoyo parcial recibido por parte del proyecto VIEP III 9-04/ING/G.

## Referencias

1. C.J. van Rijsbergen: *Information Retrieval*. University of Glasgow, pp. 114 - 117. Second Edition, 1999.
2. Gerardo Sierra & John McNaught: "Natural Language System for Terminological Information Retrieval". CICALing 2003, LNCS 2588, A. Gelbukh (Ed.) Springer Verlag, pp. 541-552, 2003.
3. Aceves, R., Castilla Y.: *Resultados del Laboratorio de Recuperación de Información de la FCC*. Facultad de Ciencias de la Computación. BUAP
4. Baeza-Yates Ricardo: *Modern Information Retrieval*, Addison Wesley. 1999
5. Reyes A. Berenice, Moyotl H. Edgar: *Reducción de términos índice usando el punto de transición*. Facultad de Ciencias de la Computación. BUAP
6. Fraguera Cuesta L.: *Análisis de una Representación de Textos usando Lattices*. Tesis de Maestría en Ciencias de la Computación. BUAP. 2004



7. García Fajardo Josefina: *Estructura conceptual y comunicación* , El Colegio de México 1995.
8. Grefenstette Gregory: *Automatic Thesaurus Generation from Raw Text using Knowledge-Poor Techniques*, Rank Xerox Research Centre.
9. Jiménez Salazr H., Morales Luna G.: *Domain membership and clasificación methods* FCC. BUAP, CINVESTAV-IPN
10. Rodríguez Hernández M.: *Representación de Documentos utilizando Relaciones de Sentido*. Tesis de Licenciatura. Facultad de Ciencias de la Computación. BUAP. Junio 2004.
11. Ruge Gerda: *Combining corpus linguistics and human memory models for automatic tem association*, AI Group, Institut fur Informatik, Munchen. 1999
12. Salazar Martínez H.; Jimenez Salazar H. & Pinto A. D.: *Extraction: a corpus-based Approach*. Facultad de Ciencias de la de la Computación. BUAP

## Datos y Resultados

**Tabla 1.** Consultas preprocesadas para el concepto Tierra.

Identificador	Consulta
$q_1$	tercer planet sistem sol vid diferent demas
$q_2$	mas efecto fuerz atraccion cuerpos
$q_3$	terc planet cercan Sol tard 24 hor gir propi eje rotacion cuenta divers cap recubr encuentr divers ecosistem encuentr divers tip biodivers
$q_4$	planet lug habit viv
$q_5$	planet habit Hombre unic vida
$q_6$	porcion material sol abund planet
$q_7$	convertir dia muramos
$q_8$	punt esencial lodo
$q_9$	planet sistem sol vivimos
$q_{10}$	primer cap cortez terrestr conoc tiempo
$q_{11}$	planet ocup terc lug cuatn posicion sistem sol Cuenta condicion propiedad favorec evol reproduccion vid organism
$q_{12}$	element natural capaz vida
$q_{13}$	nuev planet integr sistem solar
$q_{14}$	terc planet sistem sol satelit natural llam Luna
$q_{15}$	ecosistem conform plant animal element agua
$q_{16}$	terc enlac circuit

**Tabla 2.** Desempeño promedio del sistema usando diversas relaciones (255 consultas).

Relaciones	Promedio de Posición de la		Precisión	Evocación	$F_1$
	Documentos Recuperados	Respuesta Correcta			
sinonimia	48.07843	7.050980	0.050846	0.560784	0.037055
hiperonimia	132.6352	26.77254	0.01768	0.73725	0.015286
hiponimia	113.7803	23.60392	0.023224	0.694117	0.018432
todas	180.1921	45.2	0.014089	0.835294	0.012144
<i>corpus</i>	219.6117	78.32156	0.0065	0.898039	0.011090

**Tabla 3.** Representación por corpus y sinónimos (16 consultas de estudiantes).

Representación	Documentos Recuperados	Posición de la Respuesta Correcta	Precisión	Evocación	$F_1$
Corpus	151.125	41.875	0.00285	0.625	0.00568
Sinonimia	22.687	2	0.02444	0.5	0.04569
Sinonimia e hiponimia	23.25	2.312	0.02457	0.5	0.04589

**Tabla 4.** Modelo vectorial *vs* sintagmas (16 consultas de estudiantes).

Modelo	Precisión	Evocación	$F_1$
Vectorial	0.021	0.5	0.041
Vectorial sinónimos	0.021	0.5	0.041
Sintagmas <i>corpus</i>	0.002	0.6	0.005
Sintagmas sinónimos	0.024	0.5	0.045

# Pruebas con Algoritmos de Agrupamiento para generar una Base de Datos Léxica

Sofía Paniagua Rivera, Héctor Jiménez-Salazar & David Pinto

Facultad de Ciencias de la Computación  
B. Universidad Autónoma de Puebla,  
14 sur y Av. San Claudio. Edif. 135. Ciudad Universitaria,  
Puebla, Pue. 72570. México,  
Tel. (+52-222) 229 55 00 ext. 7212 Fax (+52-222) 229 56 72,  
sofiadp@terra.com.ve, hjimenez@fcfm.buap.mx, dpinto@cs.buap.mx

**Abstract.** Advances on constructing a Lexical Database (LDB) are presented in this paper. The goal is to use that LDB for Word Sense Disambiguation (WSD), thus, we present an evaluation of eight ambiguous sentences over a set of clusters (generated by the MOD-SLC algorithm). Each cluster is composed by features obtained by two methods: 1) using a efficacy formula and 2) using intersection by context-pairs. Although the corpus is small, the results encourage to proceed in this direction. The first method obtained a 75-percent accuracy and the second method obtained 62.5-percent accuracy.

**Keywords:** Clustering, Lexical Databases, Disambiguation

**Resumen.** En este trabajo se presentan avances sobre la construcción de una base de datos léxica (BDL) que estará conformada por grupos de sustantivos con su respectivos rasgos. Se espera que la BDL apoye en el proceso de desambiguación del sentido de una palabra (WSD). Por tanto, en este artículo se reporta la evaluación de ocho oraciones (con un término a desambiguar) sobre un conjunto de grupos generados a través del algoritmo MOD-SLC, con solución de rasgos por medio de dos métodos: uso de fórmula de eficacia e intersección a pares entre contextos. A pesar de usar un corpus sumamente reducido, los resultados son interesantes, ya que para el primer método se obtiene un 75% de precisión mientras que para el segundo método se obtiene un 62.5% de precisión.

**Keywords:** Agrupamiento, Bases de Datos Léxicas, Desambiguación.

## 1 Introducción

Una base de datos léxica es una base de conocimientos a gran escala, normalmente hecha a mano [7], y la construcción automática de ella es todavía una línea de estudio. Una base de datos léxica destinada, por lo general, a desambiguar el sentido de las palabras, también se considera como un diccionario que posee términos relacionados, cada uno con sus respectivas características (rasgos). La desambiguación del sentido de las palabras (WSD) ha sido un tema

de investigación en los últimos años y una tarea más en el procesamiento del lenguaje natural (PLN). El trabajo de desambiguación involucra, esencialmente, empatación del contexto de la instancia de una palabra a ser desambiguada con información de una fuente de conocimiento externa, o información acerca de contextos de instancias de la palabra anteriormente desambiguada. Se plantea la necesidad de crear un recurso que permita desambiguar palabras de textos en español y en dominios específicos. En este trabajo se presentan avances sobre la construcción de una base de datos léxica que estará conformada por grupos de sustantivos con sus respectivos rasgos. Las bases de datos léxicas tienen fundamentalmente información de términos y relaciones de significado.

El desarrollo de este recurso ha pasado por diversas fases. En la primera fase hubo necesidad de su uso en la traducción automática (MT), en donde se enfocó en la traducción de textos técnicos y de un mismo dominio. Después la MT fue orientada al desarrollo de diccionarios especializados. El sistema de Masterman(1962) utilizó una red semántica para derivar la representación de oraciones de conceptos de un lenguaje fuente [1]. Wilks(1990) usó las primitivas de Masterman para el lenguaje natural y fue uno de los primeros diseños para el tratamiento de la desambiguación de sentidos [2-4]. Otras aplicaciones se centran en la creación de diccionarios, por ejemplo, Lesk en 1986 [5] creó una base de datos donde asoció una firma a cada sentido en un diccionario. En algunos casos, se tiene un conjunto de palabras y su definición (diccionario de términos), en otros casos se tiene, además, la relación entre cada uno de los términos (por ejemplo Wordnet). El concepto de bases de datos léxicas no es reciente, sin embargo sigue en estudio; por ejemplo la base de datos léxica de EuroWordNet se está ampliando, utilizando nuevos conceptos basados en los algoritmos de Lesk, Wilks y otros precursores [6].

Los seres humanos tenemos un recurso similar a las bases de datos léxicas inmerso en nuestro conocimiento, y a partir de éste, es fácil identificar a que se refiere una palabra ambigua, siempre tomando como referencia el contexto en el cual se emite dicha palabra. Esto ha sido la base de muchos trabajos para desambiguar el sentido de las palabras y es el que seguimos en nuestro trabajo. Nuestras pruebas harán uso de corpora (homogéneo y heterogéneo) para proporcionar los contextos de los sustantivos a desambiguar. El paso más importante es agrupar automáticamente estos contextos con el fin de extraer características de cada uno de los sentidos de una palabra; pretendemos tomar cada grupo formado como un sentido. Así varios procedimientos ocupan un papel importante, como la función de similitud y el criterio de selección de características.

## 2 Planteamiento y solución

Tres técnicas están siendo implantadas con la finalidad de reportar su rendimiento en el ámbito de agrupamiento de contextos para sustantivos ambiguos. La primera técnica está fundamentada en el uso de un algoritmo voraz que utiliza una función de similitud basada en el valor del coseno del ángulo entre vectores representativos de cada contexto. La segunda técnica se le denomina MOD-SLC

(en la que se concentra el presente trabajo) y toma sus bases en la propuesta de Hassan et al [10] para fabricación de partes. La última técnica es una propuesta original de este trabajo cuyo objetivo es refinar la técnica MOD-SLC mediante iteraciones continuas, usando la técnica de temple simulado. Cada técnica hace uso de un mecanismo de representación distinto. Para representar los contextos de cada sustantivo, primero se extraen de cada contexto el mínimo número de palabras que describan al mismo (palabras representativas o rasgos), y se realiza de dos maneras. Una, calculando la eficacia y asumiendo que cada grupo está compuesto de solamente una palabra (la palabra a elegir). Se revisa para cada grupo cuál obtiene la mayor eficacia y a ese grupo se asigna la palabra como representativa de él. La otra forma de elegir las palabras representativas del grupo es obteniendo la unión de todas las intersecciones por pares de contextos (solución SR). A continuación se describe en detalle cada una de las técnicas implantadas en este trabajo.

## 2.1 Técnica MOD-SLC

El método SLC (Single Linkage Cluster; usado en biología) propuesto por Sneath [11], utiliza el coeficiente de similitud de Jaccard para encontrar la similitud entre bacterias. Por su parte la versión modificada del método SLC (MOD-SLC), utiliza una variante en el coeficiente de similitud de Jaccard conocido como coeficiente de similitud de Jaccardian o non-Jaccardian [10], el cual es una medida del nivel de empatamiento en donde el número de empatamientos es dividido por una cantidad normalizada. Este coeficiente tiene un término adicional en el numerador y es, básicamente, el número de parejas perdidas, el cual es dividido por la normalización de términos. En nuestro caso, hemos hecho uso de la medida de Jaccardian con la finalidad de medir el grado de similitud entre contextos [10]. El coeficiente de similitud queda definido como sigue:

$$SB_{ij} = (X_{ij} + \sqrt{X_{ij} * Y_{ij}}) / (X_i + X_j + X_{ij} + \sqrt{X_{ij} * Y_{ij}}),$$

donde  $X_{ij}$  es el número de unos en común entre los contextos  $i$  y  $j$  en la matriz de incidencia (palabras en común).  $Y_{ij}$  es el número de ceros en común (palabras que ambos contextos no tienen),  $X_i$  es el número de unos que están en el contexto  $i$  y no están en el contexto  $j$  (palabras diferentes).  $X_j$  se define de manera similar a  $X_i$  pero para el contexto  $j$ .

El procedimiento para la generación de los grupos se describe en la figura 1.

Posteriormente se procede a asignar palabras a cada grupo usando una función de eficacia. Para cada palabra, se verifican los contextos en los que ocurre y se calcula la eficacia entre estos contextos y para cada uno de los grupos generados. La medida de la eficacia global para los grupos generados, y que se usa también para la asignación de palabras a contextos se obtiene mediante la siguiente fórmula:

$$Eficacia = (e - e_0) / (e + e_1),$$

donde,  $e$  es la cantidad de unos en la matriz,  $e_0$  es el número de elementos excepcionales (fuera de los grupos) y  $e_1$  es el número de ceros dentro de los grupos.

- Paso 0** Generar la matriz de incidencia contexto-palabras.  
**Paso 1** Usando la matriz de incidencia crear la matriz de similitud de contextos.  
**Paso 2** Localizar el máximo valor de similitud ( $SB_{ij}$ ) en la matriz de similitud.  
**Paso 3** Asignar el contexto  $i$  y el contexto  $j$  al mismo grupo.  
**Paso 4** Eliminar la similitud ( $SB_{ij}$ ) de la matriz de similitud.  
**Paso 5** Si no están todos los contextos asignados regresar al paso 2.

**Figura 1.** Algoritmo MOD-SLC para agrupamiento de contextos.

La eficacia puede ser un buen mecanismo para medir la calidad de las soluciones obtenidas, ya que en el mejor de los casos, debería haber cero elementos excepcionales y ningún valor cero dentro de los grupos; lo que arrojaría un valor 1 y por tanto serían grupos bien definidos.

El experimento realizado en este trabajo se basó en el algoritmo MOD-SLC. Este algoritmo, como se mencionó anteriormente, permite agrupar los contextos obtenidos, para cada sustantivo del corpus, con base en una función de similitud. Para obtener estos contextos se utilizó un corpus heterogéneo. Se programaron las etapas de preprocesamiento, extracción de sustantivos, extracción de contextos, agrupamiento y desambiguación. En particular, se eligió el sustantivo “banco” como prueba y el proceso de prueba llevado a cabo con este sustantivo se describe a continuación.

Una vez generado el archivo de los contextos de la palabra ambigua “banco”, se procedió a obtener el vocabulario de sus contextos; esto debido a que el vocabulario es necesario para generar la matriz de incidencia contextos-palabras. Se procedió a aplicar el algoritmo MOD-SLC, obteniendo los grupos, y para cada grupo se extrajeron sus rasgos. Se utilizaron 8 oraciones de prueba en las cuales aparece la palabra “banco”. Las oraciones de prueba fueron extraídas del corpus del CIC<sup>1</sup>. Las palabras contenidas en las oraciones de prueba se presentan en la tabla 1. El proceso de desambiguación consistió en aplicar el cociente de Jaccard entre cada uno de los rasgos de los contextos obtenidos y la oración de consulta. Si el valor de similitud es igual a cero significa que la oración de consulta no pertenece a ningún grupo, en otro caso, simplemente se verifica el valor máximo obtenido (por grupo) y se concluye que el sentido de la oración de consulta corresponde a dicho grupo. Los resultados obtenidos se muestran en la tabla 2 y 3. Dichas tablas están estructuradas de la siguiente manera: en la primera columna se encuentra el número de oración a ser evaluada, en la segunda columna aparece el grupo al que debería pertenecer la oración (evaluación manual), en la tercera columna el grupo al que fue asignada (a través del cociente de Jaccard), y en la última columna se encuentra el valor de similitud entre la oración y el grupo a la que fue asignada.

---

<sup>1</sup> Recurso amablemente proporcionado por el Laboratorio de Procesamiento de Lenguaje Natural del CIC-IPN.

### 3 Resultados

En la tabla 2 se presenta el resultado de evaluar 8 oraciones sobre un conjunto de grupos generados a través del algoritmo MOD-SLC, con selección de rasgos por la fórmula de eficacia. Puede observarse que el número de oraciones asignadas correctamente a su grupo es 5, eso deja 3 oraciones que aparentemente no están siendo bien asignadas. Sin embargo, cabe aclarar que la oración 6 no debía ser asignada a grupo alguno, pues no existe un grupo que la represente (de acuerdo con los contextos que se obtuvieron del corpus). Por lo anterior puede decirse que, en general, el algoritmo asignó 6 respuestas correctas de 8 que se utilizaron para la desambiguación. En la tabla 3 se presenta el resultado de evaluar las mismas oraciones sobre un conjunto de grupos generados a través del algoritmo MOD-SLC, con selección de rasgos por intersección a pares entre contextos. Se puede observar que de las 8 pruebas realizadas, 5 de ellas son correctas y las 3 restantes están asignadas de manera incorrecta. Aunque nuevamente debemos tomar en cuenta que la oración 6 no debía ser asignada a grupo alguno, y sin embargo el sistema la coloca en el grupo que se espera más adecuado para dicha oración.

**Tabla 1.** Oraciones para realizar la prueba de desambiguación

No.	Consulta
1	El dolar puede cambiarse en varias instituciones financieras , tales como las casas de cambio o el banco
2	demanda a banco mundial ayuda monetaria para abatir pobreza de mujeres
3	Se acentúa la baja de las tasa financieras para depósitos en los bancos
4	Esta chica siempre toma asiento en el mismo banco
5	El banco no pudo estar cerrado sino el acreedor habría ido a otro lugar a cambiar el cheque
6	Esta evaluación cuenta con un banco de docencia curricular donde se brinda un seguimiento académico al profesor
7	Puesto que el banco rocoso se encuentra con pronunciada inclinación desde el río Troja
8	El banco condone parte de la deuda agrícola

### 4 Discusión

El trabajo presentado es el inicio de una herramienta que apoye el proceso de WSD. Para obtener el objetivo esperado y probar su efectividad se contempla realizar los siguientes pasos. En primer lugar, hacer pruebas con dos corpora, uno homogéneo y otro heterogéneo. Se usará enseguida un etiquetador de partes de discurso, para proveer mayor información a los algoritmos de agrupamiento. Asimismo, se hará uso de un lematizador para normalizar las palabras

**Tabla 2.** Resultados con la selección de rasgos utilizando eficacia de las palabras

Número de oración	Grupo correcto	Grupo asignado	Valor de la similitud
1	5	5	0.04
2	6	6	0.04
3	3	3	0.01
4	2	-	0.00
5	4	4	0.01
6	-	-	0.00
7	1	5	0.02
8	4	4	0.01

**Tabla 3.** Resultados con la selección de rasgos utilizando la solución SR

Número de oración	Grupo correcto	Grupo asignado	Valor de la similitud
1	3	3	0.43
2	3	3	0.25
3	3	3	0.5
4	2	3	0.25
5	3	3	0.17
6	-	3	0.33
7	1	3	0.2
8	3	3	0.2

que pertenecen a una misma familia morfológica, con ello se espera conseguir que el algoritmo de agrupamiento obtenga mayor eficacia al formar los grupos. Por ahora, están siendo probados los tres algoritmos de agrupamiento mencionados. Es importante destacar que el proceso de agrupamiento es complicado, debido a que existen realmente pocas palabras en común para aquellos contextos que se refieren a un mismo sentido, sin embargo, el algoritmo MOD-SLC ha mostrado resultados alentadores en el proceso de agrupamiento de contextos y se pretende mejorar esta técnica mediante la adición de heurísticas de refinamiento, como es el caso del temple simulado. Se tiene planeado usar la base de datos léxica en un sistema de recuperación de información para conocer la efectividad del recurso generado y probar con algunos otros métodos de agrupamiento como es el caso de CLC (Complete Link Clustering) y ALC (Average Link Clustering). Finalmente, este método podría vincularse con el proyecto de construcción de corpus a partir de la Web que sostienen actualmente los grupos de PLN del INAOE (Puebla, México) y de la Universidad Politécnica de Valencia (España).

## Agradecimientos

Agradecemos los comentarios de los árbitros de este trabajo, así también el apoyo parcial recibido por parte del proyecto VIEP III 9-04/ING/G.



## Referencias

1. Masterman, Margaret., "Semantic message detection for machine translation using interlingua", Stationery Office, London, pp. 437-475, 1962.
2. Wilks, Y. and Fass D., "Preference semantics: A family history", *Report M CCS*, pp. 90-194, 1990.
3. Wilks, Y. and Stevenson M., "The grammar of sense: Is word sense tagging much more than part-of-speech tagging?", *Technical report CS-96-05*, University of Sheffield, Sheffield, UK, 1996.
4. Wilks, Y. and Stevenson M., "Sense Tagging: Semantic tagging with a lexicon", 1996.
5. Lesk, Michael., "Automated sense disambiguation using machine-readable dictionaries: How to tell a pine cone from an ice cream cone", *SIGDOC Conference*, pp. 24-26, 1986.
6. Pedersen, T. and Banerjee, S., "An adapted Lesk Algorithm for Word Sense Disambiguation using WordNet", *CICLing 2002*, pp. 136-145, 2002.
7. Ide, N., and Veroni J., "Introduction to the Special Issue on Word Sense Disambiguation", *Computacional Linguistics*, Vol. 24, Number 1, pp. 1-29, 1998.
8. Yorick Wilks, Roberta Catizone: "Lexical Tuning". *CICLing 2002*, pp. 106-125, 2002.
9. Paolo Rosso, Francesco Masulli, Davide Buscaldi, Ferran Pla, Antonio Molina: "Automatic Noun Sense Disambiguation", *CICLing 2003*: pp. 273-276, 2003.
10. Hassan M. S., Reda M. S. A. A., Araby I. M., "Formation of Machine Groups and Part Families: A modified SLC Method and Comparative Study", *Integrated Manufacturing Systems*, pp. 123-137, 2003.
11. Sneath, P.H., "The application of Computers to Taxonomy", *Journal of General Microbiology*, Vol. 17, pp. 201-226, 1957.
12. Sneath, P.H., "Some Thoughts of Bacterial Classification", *Journal of General Microbiology*, Vol. 17, pp. 184-200, 1957.

# Unsupervised Learning of Word Combinations for Syntactic Disambiguation

Tania Lugo-Garcia, Alexander Gelbukh, Grigori Sidorov

Centro de Investigación en Computación, Instituto Politécnico Nacional,  
Av. Juan de Dios Bátiz, A.P. 75-476, C.P. 07738, México, D.F.,  
+52 (55) 57296000, ext. 56544, 56602, fax 5586-2936,  
tanialugo@correo.cic.ipn.mx,  
{gelbukh, sidorov}@cic.ipn.mx; www.gelbukh.com

**Abstract.** We propose a statistical method for syntactic disambiguation based on learning word combinations from unprepared unmarked text. It uses the statistics of sub-trees found in the dependency syntactic trees of the sentences in a large corpus. Namely, as features, we consider sub-trees consisting of two non-auxiliary words, one being governor and one dependent, possibly connected through a chain of auxiliary words, e.g., prepositions. The algorithm iteratively re-estimates the probabilities of the features and the probabilities (weights) of the variants of syntactic structures of the sentences of the corpus. The probabilities of the features learnt in this way are useful for disambiguating new (unseen) sentences. In addition, the features with high learnt probabilities constitute a dictionary of collocations of a given language, useful not only for syntactic disambiguation but also for part of speech tagging and similar tasks.

## 1 Introduction

Syntactic analysis, i.e., the construction of the structure of each sentence in a given text, is a difficult task when high quality is necessary. Its main problem is ambiguity: the computer usually builds more than one structure of each sentence and then it has to choose the correct one. Syntactic disambiguation is one of the most important tasks in natural language processing. We present a method based on statistics encountered in word combinations of a dependency tree.

Our method consists in building a database of sub-trees of a special form, with the statistical information gathered from raw unmarked texts.

We take as input a raw corpus, which we parse syntactically, obtaining all possible variants of parsing each sentence. We iteratively re-estimate the probabilities of the variants of parsing for each phrase and the probabilities (statistics) of the syntactic sub-trees found in the corpus, using one of them to re-estimate the other, and vice versa.

## 2 General Idea of the Method

Our method is based mainly on [1]. To resolve syntactic ambiguity, we first conduct multi-variant syntactic analysis, then estimate the probabilities of collocations found in the trees, then use these probabilities to re-estimate the probabilities of the variants, and repeat the process iteratively.

Our method is based in the linguistic knowledge acquired by the native speakers during the learning of their language. We believe that this method can be applied to many ambiguity problems, with an appropriate selection of features.

In our study we determine lexical subcategorization information (collocations) of verbs, adjectives and some nouns in Spanish. It is well-known that such knowledge is to be learnt from the data and cannot be specified a priori by rules or algorithms. Our model of syntactic ambiguity resolution is simple and easy to use and to compile the resources needed for it. We use the generative grammar formalism. The grammatical knowledge is coded directly a context-free grammar.

### 1. Linguistic Bases of the Method

Our method is based on the ideas of Galicia-Haro [1] but adapted to learning word combinations instead of prepositions. Our program uses various pre-processing modules:

- Morphological analysis and disambiguation. The POS tagging is useful for syntactic disambiguation. The POS tagging is the sub area of linguistic processing of texts which consider the study of the methods and algorithms to reduce the categorical ambiguity. Ambiguity of grammatical category refers to a word with several syntactic categories, i.e. *lie* can be a verb and also a noun. Knowing the POS of specific words helps to reject wrong variants of analysis. The disambiguation of this kind can reduce the amount of ambiguity that the syntactic analyzer has to deal with.

The LEXESP corpus we use has the categories according to the PAROLE standard [2].

- Syntactic analysis. Traditional methods employed to describe the syntactic level of natural languages are those based on generative grammars. In this approach, the description of the features of valences for the lexemes are made, mainly for the verbs. The actants (or complements) are described from a purely syntactic point of view. The information about the structure of the complements of a given verb are known as subcategorization. Each verb has its own set of complements that often have a given lineal order.

The follow examples illustrate subcategorization in Spanish:

- *To see* has a sub category *nominal group* (NG), i.e.: *Maria saw a spider.*
- *To give* has the sub category NG followed by PG, i.e.: *Maria gave a letter to his boyfriend.*
- *To put* also has the sub category NG followed by PG, the prepositional group is made by different prepositions: *on, under, etc.*

Information on subcategorization frames are useful to restrict the number of analyses generated by the syntactic analyzer.

The proposed method to obtain information on the collocations of the verbs, nouns and adjectives in Spanish is based in obtaining statistics of variants of analysis, similarly to [1], but in our case, the features to learn are combinations of verbs and individual words, with possible prepositions in between (in [1], the features are combinations of a word with a preposition). The selection of this kind of combinations or subcategorization frames is not arbitrary. These combinations are known to be mostly fixed for each specific word. The dictionary which we need to compile is thus a list of possible combinations of individual words, with possible prepositions in between. In its simplest form the list contains the entries like the follow ones for *to move*:

1. *move + table + to + corner*
2. *move + table + from + center + to + corner*

To solve the ambiguity of the links in sentences containing such combinations, we use the frequencies of the combinations in phrases of particular texts, that is, in the trees generated by the parser. In this method, for each phrase a weight (a probability) is determined for each variant of the syntactic structure. This weight is based on statistics of combinations used in the given language.

As a simplified example of such reasoning let us consider the case of disambiguation of POS. Suppose that the frequencies of different POS in the texts are:

$$P^+_{\text{noun}} = 0.4, P^+_{\text{adj}} = 0.4, P^+_{\text{verb}} = 0.2$$

Each variant consists of only one feature:  $v_1 = \{\text{adjective}\}$ ,  $v_2 = \{\text{verb}\}$ ,  $v_3 = \{\text{noun}\}$ . If this is all the information we have, then whit the result of the analysis  $V = \{\{\text{adjective}\}, \{\text{verb}\}\}$  of a word, we could conclude that, knowing

$$P^+_{\text{adj}} > P^+_{\text{verb}}$$

Then the correct variant should be adjective, because its weight is  $P(\{\text{adjective}\}) = 0.4 / (0.4 + 0.2) \approx 0.66$  while the weight for the  $P(\{\text{adjective}\}) = 0.2 / (0.4 + 0.2) \approx 0.33$ . Consider another example:  $V = \{\{\text{noun}\}, \{\text{adjective}\}\}$  and there any decision can not be made because weight are equal:  $P(\{\text{noun}\}) = P(\{\text{adjective}\}) = 0.5$ .

Suppose that the POS tagger used often reports some wrong variants for the words, and that it is done with the frequency of 0.9 for a noun, whit the frequency of 0.1 for an adjective, and never had reported a verb erroneously. Then for the result  $V = \{\text{adjective}\}, \{\text{noun}\}$  we can say that the right answer is *adjective* because both of them have the same probability and the parser make a minor error choosing an adjective.

With that reasoning, in our method we introduce two kinds of statistic weight:  $p^+$  and  $p^-$ . The weight  $p^+$  means the probability, is it, the frequency of occurrence of a particular combination with the specific governor word in the text, in a correct syntactic structure.

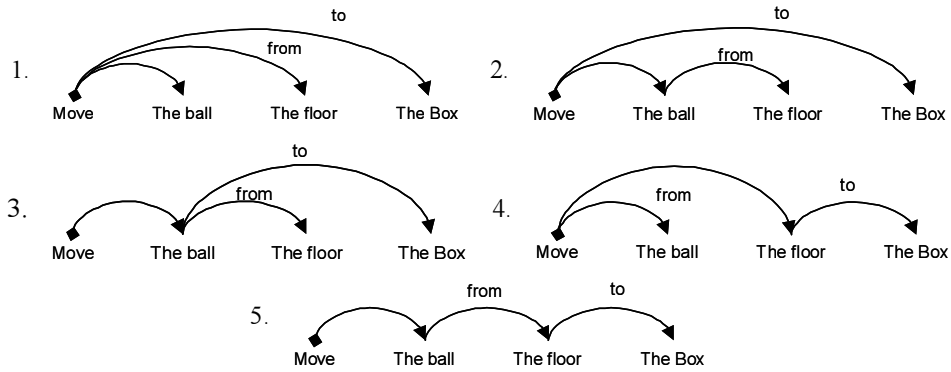


Figure 1. Variants of syntactic structure.

The  $p^-$  weight is more interesting than the previous one: it is the probability for the combination to occur in a structure that was constructed by the parser but is not correct. Consider an example:

*The girl moves the ball from the floor to the box.*

In Figure 1, the combination *the floor to the box* occurs twice in the incorrect variants 4 and 5; the combination *to the box* occurs just once in the correct variant, and 4 times in the wrong variants, so for this last combination  $p^+ = 1/5$  and  $p^- = 4/5$ .

In the model, the interdependences are not considered because including them creates troubles in the implementation (of space and time). Their inclusion should be deeply analyzed to finally decide if worth the effort. Another choice is instead taking probabilities, assign some initial weights to the variants and use them for computation.

The main feature of the model is that the event space is defined in two levels of granularity: lexical and syntactic. The lexical level is related to each word and at the syntactic level to the links that compound the combinations.

## 2. The Model

To obtain the formulas for the statistic weights of the variants of the syntactic trees of a phrase based on the combination that appears in each tree, a model of phrase generation is considered.

Suppose that all those combinations that we wish to obtain appear in the trees as abstract features of the tree. We enumerate these features i.e. the combinations “*to move*” +  $n_1$  + “*from*” +  $n_2$  + “*to*” +  $n_3$  is the feature number one, “*to accuse*” +  $n_1$  + “*of*” +  $n_2$  is the feature number two, etc. where  $n_1, n_2, \dots$ , etc. are specific nouns possibly linked to the head word by some prepositions. We denote those features as  $f_1, f_2$ , etc. Then the whole set of features or dictionary of these features is  $F$ .

In [1], the nouns linked to the verbs and the prepositions were not considered. We are considering them, in order to obtain more specific information for disambiguation. This is our new contribution as compared with [1].

Our interests are the statistics of  $f_i$  (the combinations); a simplification of the model is to omit the relationships between them. Then we consider a phrase  $\mathbf{P}$  as a set of such features,  $\mathbf{P} = \{f_{n1}, \dots, f_{nk}\}$ . I.e., for the phrase *The girl move the ball from the floor to the box*, we obtain the set  $\mathbf{P} = \{\text{move} + \text{the ball} + \text{from} + \text{the floor} + \text{to} + \text{the box}\}$ .

The generation model works this way: to generate a phrase  $\mathbf{P}$ , a source  $S$  which contains the feature  $f_i \in \mathbf{F}$  decides if this feature will be included or not in the phrase  $\mathbf{P}$  to be generated. The decision is made randomly, based on its probability  $p_i$ , which is associated in the dictionary  $\mathbf{F}$  to each of the features  $f_i$ . I.e. the generator can include the feature “*move + n<sub>1</sub> + from + n<sub>2</sub> + to + n<sub>3</sub>*” in one of each 1000 phrases  $\mathbf{P}$ , with the corresponding probability  $p_i = 0.0001$ .

Suppose then that the features are included or are not included in  $\mathbf{P}$  in an independent way (naïve independence assumption). We accept this assumption because for this method we do not have lexical knowledge of tags [3] or of lexical attraction characteristics [4].

Based on this method and in the independency hypothesis, we can compute the frequency of the appearance of a specific phrase  $\mathbf{P} = \{f_{n1}, \dots, f_{nk}\}$  in the output  $S$ , with the probabilities of the combinations. These are calculated in this way:

$$\alpha_n^{k,r} = \begin{cases} p_n^r, & f_n \in P \\ q_n^r, & \text{otherwise} \end{cases} \quad (1)$$

where:

- $p$  is the probability to choose the combination
- $q$  is the probability that the combination is not chosen; its value is  $q = 1 - p$
- $k$  is the number of the variant
- $r$  is 1 if it corresponds to the right variant (is represented by “+”) and is 0 if it corresponds to the wrong variant (is represented by “-“)
- $n$  is the number of combinations

So we have the following probabilities:

- $p_n^+$ , if  $f_n \in \mathbf{P}$  and  $k$  is the right variant
- $q_n^+$ , if  $f_n \notin \mathbf{P}$  and  $k$  is the right variant
- $p_n^-$ , if  $f_n \in \mathbf{P}$  and  $k$  is the wrong variant
- $q_n^-$ , if  $f_n \notin \mathbf{P}$  and  $k$  is the wrong variant

Then the probability of  $\mathbf{P}$  is:

$$P(\mathbf{P}) = \prod \alpha_n^{k,r} \quad (2)$$

since each of the features is included independently in the phrase  $\mathbf{P}$  with the probabilities  $\alpha$ . Where we denote  $r$  as follows:

$$\mathbf{r} = \delta_i^k \begin{cases} 1 & \text{if } k \neq i \text{ (right variant)} \\ 0 & \text{if } k = i \text{ (wrong variant)} \end{cases}$$

These probabilities can be seen as a matrix  $V$  with  $k$  rows, one per variant and  $n$  columns, one per combination. Then the values in the matrix are:

$$\alpha_n^{k,r} = \begin{cases} p_n^{\delta_i^k} & V_k[n] > 0 \\ q_n^{\delta_i^k} & V_k[n] = 0 \end{cases} \quad (3)$$

where  $V_k[n]$  represent the values of the probabilities of occurrence of the present combination,  $n \in V_k$ . If the combination  $n$  is present, then  $V_k[n] > 0$ , otherwise  $V_k[n] = 0$ .

The dictionary  $F$  of the features is so big that each of these specific features  $f_n$  is much more expected to be absent than to be present. This fact is inheriting to the texts, each word, except some conjunctions and prepositions very common, appears in a minority of the phrases of the text. In (2) the product is taken for each of the variants and for all combinations in  $F$ .

Consider the set of variants of the syntactic structure  $V = \{V_1, \dots, V_n\}$  built by the syntactic analyzer for the phrase  $P$ , it is possible to use the formula (2) for the disambiguation. Suppose that exactly one of them is the right variant. Let's be  $H_j$  the hypothesis that the variant  $V_j$  is the correct one. Let's be  $\xi$  the event of get exactly the set  $V$ , it is, the matrix  $V$ , as a result of the syntactic analysis. Then, by the Bayes formula, we have:

$$P(H_j | \xi) = P(\xi | H_j) \frac{P(H_j)}{P(\xi)} \quad (4)$$

For abbreviation, we denote  $P(H_j | \xi) = P_j$ , the probability of the variant  $V_j$  to be true. Since just one variant is true we can see that:

$$\sum_{V_j \in V} P_j = 1 \quad (5)$$

To obtain the value  $P(H_j | \xi)$  we consider that all the variants are noise except the correct one. Suppose that the hypothesis  $H_j$  is true, it is, that  $V_j = P$ , and the rest of the variants  $V_k$ , where  $k \neq j$ , are spurious variants, it is, noise. Suppose that the noise occurs in the set  $V$  independently from the true structure of  $P$ . Again, it is not correct at all, but even when it has not any useful information about the nature of its dependency, is a common simplification of the statistic methods.

Then

$$P(\xi | H_j) \sim \prod_{k=1}^K \prod_{n=1}^N \alpha_n^{\kappa, \delta_i^k} \quad (7)$$

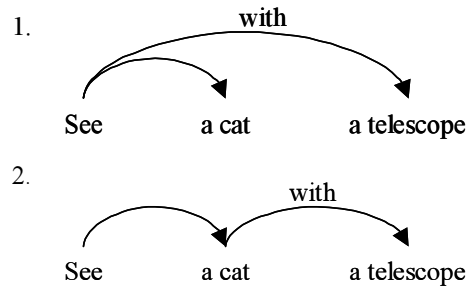
where  $N$  is the number of combinations and  $k$  is the number of variants, and  $\sim$  means *proportional*.

### 3. Example

The following simplified example is a bit artificial. Let us take a corpus consisting of the following phrase:

*I see a cat with a telescope.*

In the former model, we get the following results:



As we can see, they are very close but we know that the right analysis is 1. We expected to solve this kind of ambiguity including in the database the nouns related to the prepositions.

In the former model, two variants of analysis are built for the phrase, for the first one:  $V = \{V_1, V_2\}$ ,  $V_1 = \{see + with, cat, telescope\}$  and  $V_2 = \{see + with + telescope, cat\}$ . In fact  $V_1$  is the correct one.

By this corpus we can build the following dictionary  $F$  of combinations (for convenience, in the columns  $p^+$  and  $p^-$  we show the number of occurrences instead of frequencies):

<i>Feature f</i>	$p^+$	$p^-$
<i>see + with</i>	1	0
<i>see</i>	1	0
<i>see + with + telescope</i>	1	0
<i>cat</i>	1	1
<i>telescope</i>	1	0

Because the syntactic information does not include the nouns related to the verb and the prepositions, both analysis can be correct.

In our method the variants of the analysis will be:  $V = \{V_1, V_2, V_3\}$ ,  $V_1 = \{see + with + telescope, cat\}$ ,  $V_2 = \{see, cat + with + telescope\}$ ,  $V_3 = \{see + with, cat, telescope\}$  the dictionary  $F$  of combinations, will be the follow one:

<i>Feature f</i>	$p^+$	$p^-$
<i>See + with + telescope</i>	1	0
<i>See</i>	0	1
<i>see + with</i>	1	0
<i>Cat</i>	2	0
<i>Cat + with + telescope</i>	0	1
<i>Telescope</i>	1	0



Thus, we will not have the problem of syntactic ambiguity in this kind of phrases. Note that the method [1] cannot distinguish between the phrases *I see a cat with a telescope*, which has the correct interpretation [*I see [a cat] [with a telescope]*], and *I see a cat with a tail*, which has the correct interpretation [*I see [a cat with a tail]*]. Indeed, both phrases have as features the substructures *see + with*, *cat + with*, etc. Our method can easily distinguish between these phrases, since we consider the features *see + with + telescope*, *see + with + tail*, *cat + with + telescope*, *cat + with + tail*. The statistics of such elements helps to clearly distinguish between the two phrases, choosing the correct variant for each phrase (so that the selected variants of analysis of the two phrases prove to be different).

#### 4. Conclusions

We have suggested a generalization of the disambiguation method described in [1], to extend it to the combinations of specific words (collocations). What we expected to obtain, is a larger feature database which help us to disambiguate phrases like *I see a cat with a telescope* vs. *I see a cat with a tail*.

In [1], the nouns linked to the verbs were not considered. We are considering them, in order to obtain more detailed information for disambiguation of a phrase. The method addresses the problem of lack of information found in the former method. However, this is at the cost of greater space and time requirements and the need in larger corpora.

Possible future applications of our method are correction of semantic errors, part of speech tagging, and disambiguation in statistical grammars. In the future we will also consider other generalizations of our method, for example, to learning the selectional preferences of words as combinations of a specific word with words of specific grammatical categories (e.g., “*move* NOUN *from* NOUN *to* NOUN”) or semantic categories (“*move* OBJECT *from* PLACE *to* PLACE”) rather than specific words (“*move table from center to corner*”) as in the present work.

#### References

- [1] Sofia Galicia-Haro, Análisis sintáctico conducido por un diccionario de patrones de manejo sintáctico para lenguaje español. Ph.D. thesis, CIC, IPN, Mexico, 2000.
- [2] Civit, M. and I. Castellón. "Gramesp: Una gramática de corpus para el español". Magazine from AESLA, La Rioja, España, 1998.
- [3] Smadja, F. A. "Retrieving Collocations from Text: Xtract". *Computational Linguistics*, 19.1: 143--176, 1993.
- [4] Yuret, D. *Discovery of Linguistic Relations Using Lexical Attraction*. Ph.D. thesis. Massachusetts Institute of Technology, 1998.

# **Tecnología de objetos de aprendizaje**

Ma. Elena Chan Núñez  
Rafael Morales Gamboa  
Víctor G. Sánchez Arias  
(Editores)



## **Introducción**

La propuesta más importante hoy en día en el ámbito internacional del aprendizaje basado en tecnología es la organización de contenido educativo en la forma de objetos de aprendizaje. Existen un gran número de propuestas y proyectos en desarrollo, además de un creciente número de productos en desarrollo y operación.

La promesa de los objetos de aprendizaje es facilitar la reutilización, distribución y personalización de contenido educativo en Internet. Sin embargo, mucho trabajo hace falta todavía para acumular la masa crítica que haga de las promesas de la tecnología de objetos de aprendizaje una realidad. De particular importancia en este sentido es la elaboración, validación y adopción de un conjunto básico de estándares, modelos de referencia, recomendaciones y tecnologías que permitan alinear esfuerzos dispersos hacia un objetivo común: educación de calidad a bajo costo, accesible en todo lugar y momento.

El objetivo de este segundo taller sobre tecnologías de objetos de aprendizaje es continuar el esfuerzo del taller anterior (véase [www.iiie.org.mx/TOA](http://www.iiie.org.mx/TOA)) en formar una comunidad colaborativa, interesada en desarrollar y validar la tecnología de objetos de aprendizaje. De especial interés en esta ocasión es el intercambio de conocimientos y experiencias en relación a las diferentes propuestas internacionales para el desarrollo de contenidos y sistemas interoperables basados en objetos de aprendizaje.

## **Comité Científico**

Copresidentes:

Ma. Elena Chan Núñez, Univ. de Guadalajara (UdeG), [machancita25@hotmail.com](mailto:machancita25@hotmail.com)

Rafael Morales Gamboa, Inst. de Investig. Eléctricas (IIE), [rmorales@inf.ed.ac.uk](mailto:rmorales@inf.ed.ac.uk)

Víctor G. Sánchez Arias, Lab. Nac. Informática Avanz. (LANIA), [victor@lania.mx](mailto:victor@lania.mx)

Jorge Martínez Peniche, DGSCA-UNAM, [mpeniche@servidor.unam.mx](mailto:mpeniche@servidor.unam.mx)

Larisa Enríquez Vázquez, DGSCA-UNAM, [larisa@piaget.dgsca.unam.mx](mailto:larisa@piaget.dgsca.unam.mx)

Luz Maria Castañeda de León, DGSCA-UNAM, [luzcast@servidor.unam.mx](mailto:luzcast@servidor.unam.mx)

Leonel Iriarte Navarro, Universidad Agraria de la Habana, [lin@infomed.sld.cu](mailto:lin@infomed.sld.cu)

Ivonne Robledo Portillo, Univ. Autónoma de Ciudad Juárez, [irobledo@uacj.mx](mailto:irobledo@uacj.mx)

Lucio Teles, TELEStraining Inc., CANADA, [teles@telestraining.com](mailto:teles@telestraining.com)

Daniel Morán, Universidad Alicante, ESPAÑA, [dani.moron@ua.es](mailto:dani.moron@ua.es)

# Mundo Óseo para Niños

Misael Medina Hernández<sup>1</sup>, Edgar Ortíz Sánchez<sup>2</sup>, Marva Angélica Mora Lumbreras<sup>3</sup>

Universidad Autónoma de Tlaxcala  
Departamento de Ciencias Básicas, Ingeniería y Tecnología  
Laboratorio de Sistemas Distribuidos  
Calzada Apizaquito S/N C. P. 90300  
e-mail: mhmisa@hotmail.com<sup>1</sup>, edorsapa@hotmail.com<sup>2</sup>, marva@ingenieria.uatx.mx<sup>3</sup>

**Resumen:** En el presente artículo, se presenta un panorama general de la investigación y el desarrollo que se lleva a cabo en la elaboración del ambiente virtual de aprendizaje aplicado al sistema Óseo, con apoyo de contenido multimedia. Resaltando por una parte el valor educativo del ambiente virtual, que es el resultado de análisis previos sobre los destinatarios, los contenidos y el modelo didáctico en el que se enmarca el uso de la computadora. Por otra parte, se describe la importancia del uso de la tecnología multimedia, para el logro de los objetivos didácticos y pedagógicos. De esta manera, bajo las premisas mencionadas, se incrementa la interacción usuario-ambiente propiciando un entorno lúdico que favorece el estado de aprendizaje en el usuario. Este Software es de apoyo específico para 5to. Grado de primaria, y se basa en los temas relacionados con el mundo Óseo del libro de texto de la S.E.P, el propósito de este software es donarlo a las Escuelas primarias de Tlaxcala con el fin de que sea utilizado como una herramienta de apoyo.

**Palabras clave:** Constructivismo, Objeto de Aprendizaje, Software Educativo, Sistema Óseo

## 1. Introducción

Actualmente Universidades muestran un mayor interés en cuanto a la modernización de la educación.

Un error que ocurre a menudo en el deseo de modernizar la educación, es que se confunda una educación moderna, con el uso de la tecnología, con el simple hecho de digitalizar la información, por ejemplo, un libro en formato PDF para que los estudiantes tengan acceso a esa información, la impriman y la estudien. Otro error frecuente, es dar mayor importancia al aspecto gráfico, que al contenido de las aplicaciones, ya que es frecuente ver portales de Internet, visualmente impactantes, pero pobres en cuanto al contenido.

Insertos en la importancia que tiene el uso de las nuevas tecnologías en la educación en nuestros días, tratamos de centrar nuestros esfuerzos en desarrollar una aplicación

multimedia e interactiva con enfoque constructivista, apoyada en la necesidad de cautivar la atención de los usuarios, sin convertir la formación en algo aburrido y monótono.

## **2. Problemática**

Basados en el hecho de que los enfoques educativos tradicionales no cubren todas las necesidades educativas, proponemos que en lugar de memorizar hechos y deducir procesos podríamos dar énfasis a otras formas de enseñanza permitiendo la visualización de conceptos, eventos o interactuar con ambientes que muchas veces son completamente inalcanzables en condiciones reales.

## **3. Justificación**

La idea surge al darnos cuenta que los enfoques educativos tradicionales de la enseñanza, basados en el uso de libros de texto parece no ser suficiente para satisfacer y cubrir las necesidades de aprendizaje de los estudiantes, ya que no crea la suficiente motivación en los estudiantes para aprender.

Basados en los hechos anteriores y con la creencia de que el sistema óseo es una parte de la enseñanza que es un poco complicada de enseñarse y comprenderse, creemos que podemos aplicar a este un ambiente virtual de aprendizaje que facilite su percepción y acceso al mismo así como a su vez incluir una estrategia pedagógica que soporte el aprendizaje.

## **4. Estrategia Pedagógica**

El software educativo debe estar caracterizado con una determinada estrategia de enseñanza de una manera Implícita o Explícita: ejercitación y práctica, simulación, tutorías, uso individual, competición.

Cuando nos referimos al diseño y elaboración de un ambiente virtual con un determinado objetivo educativo, debemos tener presente la concepción acerca de cómo se producen los procesos de enseñanza-aprendizaje, que fundamenten el desarrollo del ambiente virtual y el cómo hacer viable su integración.

### **5.1 Constructivismo**

Los procesos de aprendizaje dirigido por el aprendiz son fundamentalmente una construcción en lugar de un proceso de asimilación; los significados y los contenidos son procesos de construcción y no de absorción [4].

Para la aplicación del constructivismo empleando la tecnología informática, se toma como referencia la representación del conocimiento. En el momento en que los usuarios tienen contacto con medios como videos, hipertextos, animaciones, simulación, descubren una manera diferente de representar la realidad del mundo que los rodea. De tal forma que se habitúan a representar o simbolizar todo cuanto esté a su alcance. Es decir, manipulando estos elementos de representación se le permite al estudiante crear su propia visión de la realidad [4].

## **5.2 Multimedia**

El uso de los multimedia nos presenta características y particularidades, con su capacidad de integrar paquetes de información en muy diversos formatos (texto, imágenes, sonido, vídeo, animaciones), lo que permite contextualizar el aprendizaje, y ubicar al usuario en entornos más asequibles [1].

A través de estas aplicaciones multimedia, a los usuarios se les permite obtener su propio ritmo de aprendizaje a la hora de ejercitarse en prácticas de repetición, identificación y afianzamiento de conocimientos previos, pero sobre todo presenta un nuevo aprendizaje, el de la adquisición de nuevas destrezas sobre el uso de medios tecnológicos como la computadora [2].

## **5.3 Interactividad**

Una de las claves del éxito de los ambientes virtuales de aprendizaje, se encuentra en su capacidad de generar situaciones interactivas, es decir, de provocar un diálogo entre el ambiente y el usuario que le permita seguir su propio aprendizaje. De este modo, los entornos multimedia han logrado romper con la transmisión unidireccional y lineal de la información, transformándose en vehículos de discusión a partir de los cuales se generan nuevas estrategias de búsqueda y acceso al conocimiento [2].

## **6. Mundo Óseo**

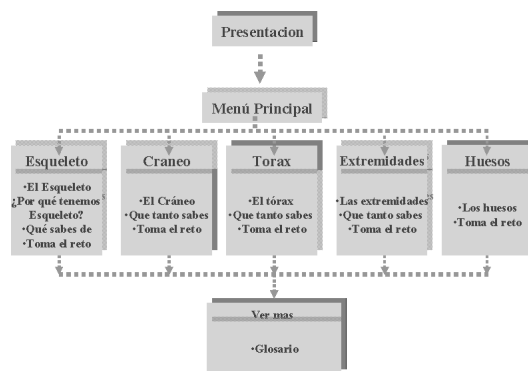
El objetivo de este proyecto es el de desarrollar un ambiente virtual de aprendizaje que funja como una herramienta didáctica, ágil y divertida, diseñado para los niños de 5to. Grado de primaria. Mundo Óseo se apoya de elementos multimedia, que transmiten la información de una manera atractiva para despertar el interés por parte del estudiante, con el fin de entender mas a fondo lo que es y representa el sistema Óseo.

El propósito es que el estudiante llegue a ser participante activo en el material que esta estudiando.



A continuación se hace la descripción del ambiente virtual así como de los diferentes módulos que lo componen.

En al Fig. 1, se observa el diagrama general del sistema, en él, se presentan de manera gráfica todos los componentes que conforman a Mundo Óseo así como las relaciones que se tienen.

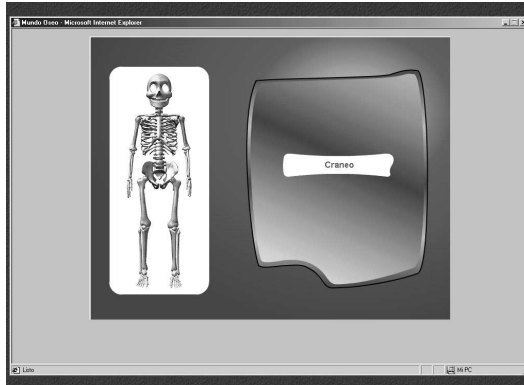


**Fig. 1** Diagrama General

El contenido temático se fundamenta en varios Aspectos:

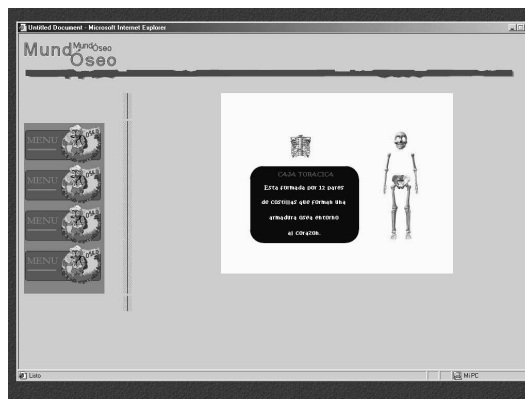
- Se basa en el programa de estudios de la educación primaria en la enseñanza del sistema óseo.
- Se basa en el empirismo de los profesores, al momento que se encuentran al enseñar este tema.
- Se hace uso de recursos didácticos como enciclopedias y laminas en el que nos hemos apoyado para documentar el ambiente.

En la Fig. 2, podemos observar el menú principal del sistema.



**Fig. 2** Menú Principal

En Mundo Óseo, se presentan tutorías que contemplan y ayudan a adquirir el conocimiento sobre el contenido temático a través de animaciones que le hacen entretenido su aprendizaje, un ejemplo se muestra en la Fig. 3.



**Fig.3** Contenido Temático

Además, Mundo Óseo cuenta con actividades que permiten al usuario reforzar y evaluar su conocimiento, logrando involucrar al usuario al tema que se está tratando. Un ejemplo de esto se muestra en la Fig. 4.

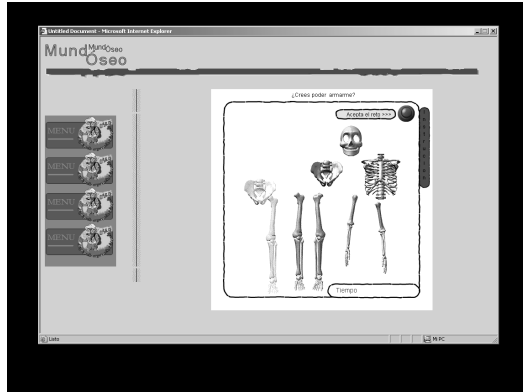


Fig. 4 Actividad de Refuerzo

La plataforma en que se basa el ambiente es una combinación de herramientas que permiten el manejo y la creación de contenidos multimedia. En la elaboración de tutorías se hace uso de SMIL para integrar estos contenidos. Las animaciones son hechas en 3D Studio Max, Flash MX JavaScript y finalmente páginas web, donde esta integrado todo el ambiente.

## 7. Aportaciones

Este trabajo nos presenta diversas estrategias en el manejo de contenidos conjuntamente con actividades e interacciones simultáneas apoyadas en un software educativo, donde el alumno aprende, manipula y aplica sus conocimientos previos construyendo sus propios conceptos.

Y nos reafirma el pensamiento de que un proceso de aprendizaje encuentra mejores condiciones cuando se tiene lugar en un medio activo en el que los alumnos participan en el propio proceso a través de la construcción de objetos y la noción de aprendizaje autónomo con el uso de entidades computacionales.

## 8. Conclusiones

Actualmente el software esta en proceso de entrega a tres escuelas primarias de Tlaxcala para su uso y evaluación del mismo, con el objeto de que exista una retroalimentación del mismo.

En ningún momento se ha pensado en sustituir los métodos y condiciones existentes con anterioridad sino en utilizar el producto como un complemento adicional del

proceso de enseñanza-aprendizaje, aprovechando las ventajas que este ambiente puede ofrecer.

Desde el punto de vista didáctico sobre el tema que aborda, consideramos que es posible consolidar los conocimientos sobre la estructura del Sistema Óseo y de la importancia que este tiene en nuestra vida.

Finalmente podemos decir que debemos utilizar y promover herramientas que propicien el aprendizaje, para interactuar con el conocimiento y hacerlo nuestro para que a su vez tengamos la capacidad de generar nuevo conocimiento. Solamente así, podremos decir que ya hemos modernizado la educación, reflejando una verdadera innovación en cuanto al método de aprendizaje.

## **9. Trabajos Futuros**

Actualmente la Universidad Autónoma de Tlaxcala esta desarrollando un software para el 5º. Grado de primaria, para que sea utilizado por las Primarias de Tlaxcala.

## **Referencias**

- [1] M. E. Del Moral y J. I. García, "DIMAS: diseño de módulos interactivos hipermediales para la explotación didáctica de los MASS MEDIA", Departamento de CC. De la Educación, Universidad de Oviedo, 1997.
- [2] M. E. Del Moral, "Aplicaciones multimedia para la formación del profesorado: diseño y evaluación". I Simposio Iberoamericano sobre Redes de Comunicación para la Educación, Mar del Plata, Argentina, 6-10 de Mayo de 1996.
- [4] J. Parra, et. Al., "Seviba: realidad virtual para el aprender a aprender. Una reflexión alrededor del desarrollo del software constructivita basado en pensamiento sistémico"
- [5] A. López, et. al. "Ambientes Virtuales de Aprendizaje. Sociedad Mexicana de Computación.
- [6] P. Ávila y M. D. Bosco, "Ambientes virtuales de aprendizaje. Unanuevaexperiencia", 20th. International Council for Open and Distance Education. Düsseldorf, Alemania, 1-5 Abril 2001.
- [7] A. M. Tejada, "Software Escolar Diseñado desde el Constructivismo: Estrategia Didáctica para Favorecer punto y modo de articulación del fonema /rr/ en niños preescolares con dislalias funcionales". Universidad Pedagógica Nacional Unidad Upn 099, D.F. Poniente, 2003.

# Objetos de Aprendizaje Adaptativos

José Mario García-Valdez<sup>1</sup>, Alfredo Cristóbal-Salas<sup>2</sup>,  
Antonio Rodríguez-Díaz<sup>2</sup>, Misael Isaac Ramírez-Noriega<sup>1</sup>

<sup>1</sup>Instituto Tecnológico de Tijuana, Departamento de Sistemas y Computación  
Calzada. Tecnológico s/n Fracc. Tomás Aquino,  
AP 1166 Tijuana, B.C. México 22000  
email: {mario, misael}@tectijuana.mx

<sup>2</sup>Universidad Autónoma de Baja California, Facultad de Ciencias Químicas e  
Ingeniería. Calzada Universitaria s/n, Delegación Otay, Tijuana, B.C. México 22390  
email: {cristobal, ardiaz}@uabc.mx

**Abstract:** Educational systems based on learning objects are currently available on the Internet, however the information presented to the user is not necessarily adequate because the context in which the user is immersed is not always considered. In this paper an extension to learning objects is proposed so they can automatically adapt to the context of use. Design details of the adaptation mechanism are presented and also a use case example. In preliminary results of our work the information presented is relevant to the users. Benefits to the learning process are expected and also an enhancement to the available learning objects.

**Resumen:** Existen varios sistemas en Internet que implementan objetos de aprendizaje. Pero vemos que la información que se muestra al usuario no necesariamente es la más adecuada y esto se debe a que el contexto en el que el usuario se desenvuelve no es tomado en cuenta. Es por esto que en este artículo se presenta una propuesta de extensión de los objetos de aprendizaje para hacer que se adapten al contexto del usuario de manera automática. Se presentan detalles de diseño como los módulos que integran el mecanismo de adaptación así como un ejemplo de uso de nuestra propuesta. Los resultados preliminares de nuestro trabajo muestran que el mecanismo de adaptación de la información presenta información relevante para el usuario y en el formato de su preferencia. Esto creemos que facilitará el proceso de enseñanza/aprendizaje y se aprovecharán los objetos de aprendizaje ya existentes.

**Palabras Clave:** Objetos de aprendizaje, Sistemas hipermedia adaptativos, Sistemas adaptativos.

## 1. Introducción

Tradicionalmente un instructor prepara su material didáctico utilizando contenido de distintas fuentes, selecciona distintos pedazos de información que posteriormente ensambla para formar el curso o clase a impartir. Los objetos de aprendizaje están basados en esta metodología viendo al contenido didáctico como un componente el

cual es diseñado para combinarse con otros para ser usados en diferentes contextos [3,2]. En general los objetos de aprendizaje [18] tienen las siguientes características:

- Cada objeto puede utilizarse de manera independiente.
- Un objeto de aprendizaje puede ser utilizado en diferentes contextos para múltiples propósitos.
- Los objetos de aprendizaje pueden ser agrupados en colecciones para después ser presentados con una estructura de curso tradicional.
- Cada objeto de aprendizaje tiene asociada información en forma de metadatos, que lo describe. Esto facilita su reutilización por medios automáticos.

De acuerdo al Modelo de Referencia de Objetos de Contenido Compartido (Sharable Content Object Reference Model - SCORM) [6] los objetos de aprendizaje llamados en este modelo Objetos de Contenido Compartible (Sharable Content Objects - SCOs) son una colección de uno o más datos. Estos datos son definidos como una representación electrónica de texto, imágenes, objetos de evaluación, o cualquier pieza de datos que puede ser representada por un cliente Web.

Otros modelos [7,16] proponen la agregación de elementos básicos para crear colecciones de objetos que se pueden relacionar con una estructura de un curso en particular. Nosotros definimos a un OA como un elemento dentro de una red de recursos utilizados para el aprendizaje. Los recursos son accedidos mediante metadatos estándar y al considerar a los objetos de aprendizaje dentro de una red es importante contar con la posibilidad de relacionarlos. La iniciativa de metadatos de Dublín (Dublin Core Metadata Initiative - DCMI) [14] define varios tipos de relaciones como: son *IsPartOf*, *References*, *Requires*, *IsBasedOn*, *IsA*, sin embargo para los objetos de aprendizaje se han propuesto otras relaciones como son: *IsPrerequisiteOf*, *HasApproved*, *Knows* [1]. Para nuestro modelo de objetos de aprendizaje se propone además ponderar ciertas relaciones.

Los objetos de aprendizaje normalmente son administrados por un *Sistema de Administración de Aprendizaje* (LMS, por sus siglas en inglés Learning Management System). Para poder intercambiar estos objetos de aprendizaje entre distintos LMSs es necesario contar con estándares abiertos los cuales permitan una plataforma común para su reutilización [6]. Ésta es la motivación de varias iniciativas de estandarización las cuales han propuesto diferentes modelos para lograr una mejor utilización de los objetos de aprendizaje [6,7,8,9]. Sin embargo, los elementos que tienen que ver con la personalización del contenido educativo son considerados de forma limitada. Es por esto que este artículo presenta una propuesta de extensión al concepto de objetos de aprendizaje para lograr adaptarlos al contexto del estudiante, considerando elementos adicionales encontrados normalmente en sistemas hipermedia adaptativos (SHA). En la siguiente sección se describen las principales ideas detrás de los SHA, en la sección 3 se describen las principales especificaciones abiertas para objetos de aprendizaje y en la sección 4 la propuesta de objetos de aprendizaje adaptativos, en la sección 5 presentamos el ejemplo de un objeto de aprendizaje adaptativo. Por último, presentamos las conclusiones y el trabajo futuro.

## 2. Sistemas Hipermedia Adaptativos (SHA)

Un sistema adaptativo es aquel que automáticamente modifica aspectos de su funcionalidad o interfaz para favorecer las necesidades de un usuario o grupo de usuarios [19]. Un SHA adapta el contenido y las posibilidades de navegación dentro de un sistema hipermedia considerando un modelo de usuario [20]. Técnicas de los SHA son utilizadas normalmente en sistemas de aprendizaje basado en Web para personalizar la presentación del material didáctico a las necesidades particulares de los usuarios. Benyon y Murray [19] identifican tres componentes básicos de un sistema adaptativo:

- El Modelo del Usuario. El cual es una representación del conocimiento y preferencias que el sistema “cree” que posee el usuario. El usuario puede ser un individuo, grupo de usuarios u otro sistema.
- El Modelo del Dominio. En este modelo se representa el conocimiento con que se cuenta para poder adaptar el sistema, es por ejemplo el conocimiento que se debe transmitir a un usuario.
- El Modelo de Interacción. Es una representación de la interacción deseada entre el sistema y los usuarios, captura la información generada por la interacción y define las inferencias, adaptaciones y evaluaciones que pueden ocurrir.

Brusilovsky [20] considera dos métodos de adaptación:

- Navegación Adaptativa. El contenido puede ser personalizado considerando el Modelo del Usuario, ajustando el nivel de detalle, el estilo o cambiando el medio de presentación. Las palabras, párrafos o incluso metáforas pueden adaptarse al usuario. En sistemas educativos el contenido es adaptado considerando el conocimiento y las metas del alumno.
- Contenido Adaptativo. De acuerdo al Modelo del Usuario, las opciones de navegación son modificadas, por ejemplo la liga a cierto contenido puede no estar disponible hasta que cierta meta sea alcanzada, o el color y otros aspectos de presentación cambian cuando se infiere que el usuario ya los conoce o ha visitado anteriormente. La ruta de navegación también es modificada considerando el estilo de aprendizaje o las preferencias del usuario. Los patrones de navegación de otros usuarios pueden considerarse para definir la navegación de usuarios similares.

## 3. Estándares para los Objetos de Aprendizaje

Las más importantes iniciativas de estandarización son la Advanced Distributed Learning Initiative (ADL-SCORM) [6], el Instructional Management System Project (IMS) [7], la Alliance of Remote Instructional Authoring Distribution Networks of Europe (ARIADNE) [8], y el IEEE Learning Technology Standards Committee [9]. Se han desarrollado especificaciones abiertas con distintos propósitos para los objetos de aprendizaje, a continuación mencionamos algunos:

*Learning Object Metadata (LOM)*. Este esquema de metadatos está basado en la

iniciativa Dublin Core [14] la cual describe de manera general a los diferentes recursos disponibles en el Web. El LOM define ciertos elementos adicionales para describir el contenido educativo. Este estándar puede ser referenciado por otros estándares los cuales pueden agregar otros medios que permitan explotar a los objetos de aprendizaje. [6].

*Modelo de agregación del contenido (Content Aggregation Model - CAM).* Define un modelo para el empaquetado y agregación de objetos de aprendizaje. Permite organizar a los objetos de aprendizaje en una estructura jerárquica y pasiva permitiendo su organización en forma de un curso; aunque actualmente no se tienen otro tipo de organizaciones [7].

*Información del estudiante (Learner Information - LI).* Nos permite describir a distintos tipos usuarios de los objetos de aprendizaje; tanto a los estudiantes como a los productores del contenido educativo. Los elementos están basados en accesibilidad, actividad, afiliación, competencias, metas, identificaciones, interés, cualidades, certificaciones y licencias, relaciones, llaves de seguridad así como transcripciones [7].

*Secuencia y navegación (Sequence and Navigation - SN).* Define un método para representar el comportamiento intencional de una experiencia de aprendizaje de tal manera que cualquier sistema de tecnología de aprendizaje (Learning Technology system – LTS) puede organizar actividades de aprendizaje discretas de manera consistente. También provee de un ordenamiento de comportamientos basado en reglas [7].

#### **4. Objetos de Aprendizaje Adaptativos (ALOs)**

Los objetos de aprendizaje (LO por sus siglas en inglés Learning Objects) son normalmente tratados como contenidos estáticos es por esto que los ALOs que proponemos en este artículo son una extensión de los LOs que además de tener las características de los LOs contienen los siguientes módulos que permiten su adaptabilidad al contexto del usuario.

*Modelo del dominio del conocimiento (Knowledge Domain Model - KDM).* El KDM representa el conocimiento aprendido por los usuarios. Para una universidad este modelo podría ser la información curricular del estudiante. Este modelo esta compuesto por una red de elementos del conocimiento (Knowledge Elements - KE) los cuales son los objetivos de aprendizaje más básicos.

*Modelo del usuario (User Model - UM).* El UM es un componente básico del sistema adaptativo, en este caso el perfil del usuario captura las características generales y las preferencias del estudiante, también el modelo del conocimiento del estudiante (Learner Knowledge Model - LKM), implementado como red semántica, representa el conocimiento actual o capacidades actuales del estudiante. Esta red es manipulada por reglas adaptativas encontradas en los ALOs.

*Sesión (S).* El componente sesión contiene información adicional que es generada en tiempo real cuando ocurre la interacción entre los ALOs y un estudiante. Por ejemplo, este componente es responsable de contabilizar el tiempo que les toma a los estudiantes responder una pregunta. Este componente asocia el modelo del estudiante, el modelo del contexto y el ALO.



Modelo del contexto (CM). Contiene el conocimiento sobre el contexto en el cual ocurre la interacción. En algunos casos utiliza un sistema de conciencia del contexto Context Awareness System - CAS) para capturar el contexto a través de algún hardware o aplicación. La información sobre el contexto puede ir desde el tipo de navegador que se está utilizando hasta la información biométrica del usuario.

Las Reglas de Adaptación (ARs) son utilizadas para personalizar la información que los ALOs proporcionan al usuario. Las ARs son reglas de producción que se activan y pueden manipular redes semánticas del LKM. Existen tres tipos de ARs:

*Reglas de navegación (NR)*. Son usadas para decidir que enlaces son habilitados o marcados por el usuario. Esto no es una decisión simple sino que hay varias propuestas diferentes a las reglas de producción como lo son las redes bayesianas [1, 17] así como los cálculos del tipo Gozinto. Estas propuestas toman en cuenta pre-requisitos ponderados entre los LOs. Existen otro tipo de relaciones que pudieran tomar en cuenta la relación *IsRelated* para producir un enlace si el estudiante está interesado en otro tópico. También si un ALO ha sido frecuentemente visitado entonces se pudiera presentar como primera opción o se pudiera presentar en otro formato.

*Reglas de presentación del contenido (CPR)*. Estas reglas escogen el ALO más apropiado dentro de varios ALOs para presentarlo al usuario considerando los componentes de adaptación.

*Reglas de Adaptación (AR)*. Estas reglas modifican el LKM añadiendo, borrando o modificando los pesos de los enlaces. También pueden ser activadas por la interacción del estudiante. También existen ALOs cuyo propósito es evaluar a los estudiantes. Por ejemplo, una AR añade un elemento de conocimiento al LKM del estudiante si la evaluación de la sesión con el objeto es mayor que algún cierto valor.

Estos modelos pueden estar empacados en un solo ALO o éste puede tener solo referencias a ellos, ya que todos los elementos son tratados como recursos. Esta propuesta difiere de otros sistemas adaptativos basados en OA [1,10,12,15] en los cuales el mecanismo adaptativo no se encuentra en los OA mismos.

## 5. Ejemplo de un ALO.



Fig 1. Recursos relacionados con el himno nacional mexicano.

Para clarificar el concepto de adaptabilidad de los LOs presentamos en esta sección un ejemplo de cómo se define un ALO. En la figura 1 se puede observar recursos de

información potenciales para definir un curso sobre el himno nacional mexicano. Para este ejemplo, contamos con tres imágenes: la bandera de México y los autores del himno nacional mexicano. También se incluye texto que contiene la historia del himno así como el texto del himno en dos versiones resumida y completa. La versión instrumental del himno está también disponible. Estos recursos pueden ser considerados como OA, etiquetados además con sus correspondientes metadatos, por ejemplo el LOM.

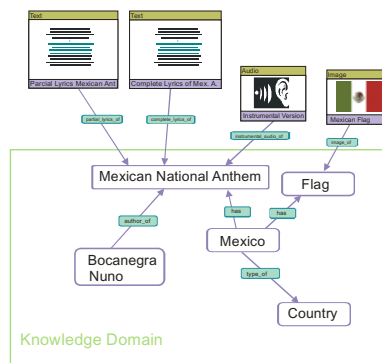


Fig 2. Relaciones entre los recursos y el dominio del conocimiento.

La figura 2 presenta las relaciones entre los componentes relacionados con el himno nacional mexicano contenidas en el dominio de conocimiento. Las relaciones usadas son: “letra parcial de”, “letra completa de”, “instrumentación de”, “imagen de”, “autor de”, “tiene”, “tipo de”. Para nuestra propuesta es importante que exista una conexión con ciertos dominios, ya que esto nos permite la selección de OAs a partir de elementos del dominio del conocimiento.

This figure shows an example of an Adaptive Learning Object (ALO) for the Mexican National Anthem. The interface includes several sections:
 

- Rules:** A section with a rule: "F Learner.Education = Basic THEN PartialLyrics.Enabled = TRUE".
- Learner Model:** A diagram showing the relationship between 'Flag', 'Mexico', and 'Country'.
- Knowledge Domain:** A diagram showing the relationship between 'Mexican National Anthem', 'Flag', 'Mexico', and 'Country'.
- Text (Aggregate):** A section containing the history of the anthem, starting with "Fue en el año de 1821 cuando se estrenó la primera composición de Himno Nacional suscrita por José Torrescano..." and mentioning figures like Andrew Davies Bradburn and Henry Hertz.
- Media Resources:** Images of Jaime Nuno and G. Bocanegra, the Mexican flag, and audio files for 'Partial Lyrics of Mex. A.', 'Complete Lyrics of Mex. A.', and 'Instrumental Version'.
- URI:** <http://www.hn.mx/himno/6/4/2001>

Fig 3. Ejemplo de un ALO

Al contar con un ALO podemos adaptar el contenido que presenta, en la figura 3 se puede observar que se cuenta con un modelo del usuario y además reglas de adaptación con estos recursos podemos especificar condiciones como la siguiente: suponemos que dos estudiantes de dos países distintos uno de ellos es México y ambos desean aprender el concepto de himno nacional entonces la información puede ser presentada al estudiante dependiendo de su nacionalidad ya sea en su lenguaje, con información pertinente a su país o con información que le sea mas común dentro de su ambiente social. En el caso de un estudiante mexicano, si éste es de nivel básico al mostrar la letra (en base a las reglas de adaptación) se mostraría el recurso de la letra parcial. Es importante mencionar que los recursos con que cuenta a su vez este ALO pueden ser otros ALOs como es el caso del texto de historia del himno nacional, este se adapta con sus propias reglas de adaptación. En este ejemplo básico no se muestran los modelos de sesión ni de contexto.

## 6. Conclusiones y trabajo futuro

En este artículo hemos presentado una extensión de a los objetos de aprendizaje para hacerlos adaptarse al contexto del usuario. Aunque la plataforma que implementa los objetos de aprendizaje adaptativos esta aún en desarrollo los resultados preliminares muestran que el contenido mostrado al usuario es mas adecuado. Presentando la información relevante para el usuario en el formato multimedia que más facilite su aprendizaje.

**Reconocimientos.** Este trabajo es parcialmente soportado por UC-Mexus bajo el programa de posdoctorado 2003-2004, SEP-PROMEP bajo el número de folio PTC-UABC-16, UABC con el número de proyecto 371. Cualquier opinión, descubrimiento, conclusión o recomendación expresada en este material solo manifiestan las ideas de los autores y no necesariamente reflejan el punto de vista de UC-Mexus, SEP o UABC.

## 7. Referencias

- [1] G.F. Knolmayer, "Decision Support Models for Composing and Navigating through e-Learning Objects", *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, 2003.
- [2] H.J. Redeker, "An Educational Taxonomy for Learning Objects", *Proceedings of the The 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03)*, 2003.
- [3] D.A. Wiley, *Connecting Learning Objects to Instructional Design Theory: A Definition, a Metaphor, and a Taxonomy In: The Instructional Use of Learning Objects*, Bloomington: Association for Educational Communications and Technology, 2000, pp. 3-23.
- [4] D.J. Mullier and M.B. Dixon, "Authoring Educational Hypermedia Using a Semantic Network", *Proceedings of ED-MEDIA 2000*, Montreal, Canada, 2000.

- [5] N. Henze and W. Nejdl, "Knowledge Modeling for Open Adaptive Hypermedia", *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems*, May 29-31, 2002, Malaga, Spain.
- [6] SCORM: Sharable Content Object Reference Model <http://www.adlnet.org>
- [7] IMS: Instructional Management System Project <http://www.imsproject.org>.
- [8] ARIADNE: Alliance of Remote Instructional Authoring and Distribution Networks of Europe <http://www.ariadne.unil.ch> .
- [9] LOM: Draft Standard for Learning Object Metadata, IEEE 1484.12.1-2002, Jul-15-2002.
- [10] P. Dolog, N. Henze and W. Nejdl, "Logic-Based Open Hypermedia for the Semantic Web", *International Hypertext Conference*, August 26-31 2003, Nottingham, UK.
- [11] L. Anido, M. Llamas, M.J. Fernández, J. Rodríguez, M. Caeiro and J. Santos, "A Standards-driven Open Architecture for Learning Systems", *Proceedings of the IEEE International Conference on Advanced Learning Techniques (ICALT'01)* 2001.
- [12] P. Mohan and C. Brooks, "Learning Objects on the Semantic Web", *Proceedings of the The 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03)*, 2003.
- [13] F. Concannon, M. Byrne and J. Shields, "Instructional Design and its implications for eLearning technologies", *Proceedings of EdTech 2003*, May 22-23 2003, Waterford, Ireland.
- [14] DCMI: Dublin Core Metadata Initiative <http://www.dublincore.org>
- [15] O. Conland, D. Lewis, S. Higel, D. O'Sullivan, V. Wade, "Applying Adaptive Hypermedia Techniques to Semantic Web Service Composition", *Proceedings of the Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2003)*, May 20, 2003.
- [16] C. Barritt, "*Reusable Learning Object Strategy Ver. 4.0*", Cisco Systems Inc. November 2001. [http://newsroom.cisco.com/dlls/innovators/e\\_learning/tech.html](http://newsroom.cisco.com/dlls/innovators/e_learning/tech.html).
- [17] N. Henze: "*Adaptive Hyperbooks: Adaptation for Project-Based Learning Resources*", PhD Thesis, University of Hannover, April 2000.
- [18] *About Learning Objects*, en la Universidad de Wesleyan  
(<http://learningobjects.wesleyan.edu/about/index.html>)
- [19] D.R. Benyon y D.M. Murray, "Adaptive Systems; from intelligent tutoring to autonomous agents". *Knowledge-based Systems*, 6 (3) 1993
- [20] P. Brusilovsky, Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6 (2-3), pp. 87-129. 1996.

# Propuesta de Metodología para elaboración de cursos en línea utilizando Objetos de Aprendizaje y Software Libre como entorno de aprendizaje.

Juan P Cardona S. y Francisco Álvarez R.

Universidad Autónoma de Aguascalientes  
Av. Universidad No. 940, c.p. 20100/Norte, México  
[fjalvar@correo.uaa.mx](mailto:fjalvar@correo.uaa.mx) [jpcardon@correo.uaa.mx](mailto:jpcardon@correo.uaa.mx)

**Palabras clave:** Aprendizaje asistido por Tecnología, Diseño de cursos en línea, Objetos de Aprendizaje, GNU GPL (General Public Licence.)

## Abstract

El presente trabajo se refiere a la propuesta de un curso en línea diseñado con Objetos de Aprendizaje como diseño instruccional; utilizando HTML como herramienta de diseño y como plataforma de herramientas de aprendizaje utiliza un software de licencia pública general (GNU GPL, 2004) basado en PHP<sup>1</sup>/MySQL<sup>2</sup> llamado Claroline (Claroline, 2001) el cual provee un entorno de aprendizaje colaborativo donde maestros e instituciones pueden crear y administrar cursos a través de la red Internet. La propuesta es una metodología para la integración de estos dos elementos con la libertad de definir el diseño instruccional del curso que para este caso se eligió el diseño de Objetos de Aprendizaje.

The present work focuses to the application of an on-line course designed with Learning Objects as educative design; using HTML as our design tool. As a platform of learning tools it uses a general public license (GNU GPL, 2004) software based on called PHP/MySQL: Claroline (Claroline, 2001) which provides an common learning environment where teachers and institutions can create and manage their courses through Internet. The proposal is a methodology for the integration of these two elements with the freedom of

---

<sup>1</sup> Lenguaje de instrucciones script diseñado especialmente para desarrollos en internet y puede ser mezclado con otros lenguajes para internet como HTML (HiperText Markup Language)

<sup>2</sup> Base de Datos con licencia pública general (GNU GPL) de mayor uso en el mundo por su rendimiento y seguridad.

defining the educational design of the course that in this case, the Learning Objects take place.

## **I. Introducción**

En el ámbito internacional de aprendizaje basado en tecnología, en la área de Educación Abierta y a Distancia, se están diseñando los cursos con los últimos avances de la Psicología Educativa como el diseño instruccional, en el caso de cursos en entornos virtuales el diseño instruccional más idóneo y mas utilizado es Objetos de Aprendizaje.

El mismo nombre de Objetos de Aprendizaje (OA) se utiliza para denominar a la descripción de un tipo genérico de dato, es decir datos acerca de los datos o lo que es lo mismo: metadata, este esquema -metadata se utiliza como unidad de almacenamiento y los datos almacenados bajo este esquema se conocen como Objetos, si la información es un apoyo al proceso de enseñanza-aprendizaje entonces será un Objeto de Aprendizaje.

Esta organización de la información hace posible compartir y explotar la información entre usuarios y entre sistemas y así cumplir uno de los principios del OA: que pueda ser “reusable, interoperable y con posibilidades de ensamble“(Wiley, 2001).

La información puede ser compartida automáticamente siempre y cuando se ajuste a algún estándar metadata, entre los estándares actuales está el LOM Learning Object Metadata creado por el Comité de Estándares de Tecnologías de Aprendizaje del IEEE, Institute of Electrical and Electronics Engineer, Inc., el DCIM Dublín Core Metadata Initiative, el SCORM (Sharable Courseware Object Referent Model, el IMS Information Management System y en México actualmente se esta trabajando en un Repositorio Distribuido de Objetos de Aprendizaje (REDOBA) que utilizará un esquema-metadata propio.

Para poner en operación un curso en línea es necesario contar con una plataforma o aplicación informática que contenga herramientas de apoyo al proceso de enseñanza-aprendizaje, esta plataforma puede ser comprada, rentada, hecha a la medida o hacer uso de plataformas de licencia pública.

El paso siguiente es propiamente poner en “línea” el curso presente, para esto es necesario una computadora propia que tenga acceso a Internet ó rentar el servicio de hospedaje del curso con alguna compañía que ofrezca este servicio.

La plataforma a utilizar deberá ser la más adecuada a las características del curso, para esta propuesta se eligió la plataforma Claroline por la sencillez de su manejo y la integridad de funcionamiento.

El proyecto Claroline fue inicialmente desarrollado por la Universidad Católica de Louvain, Bélgica por Thomas De Praetere, Hugues Peeters y Christophe Gesché, (UCL, 2001) y desde entonces lo desarrolla un red internacional de maestros y programadores que recicla programas completos o piezas de código encontrados en las vastas librerías que desarrolla la comunidad mundial de código libre, (Claroline, 2004).

## II. Justificación del uso de Claroline

En la elección de la plataforma Claroline se consideraron algunos indicadores sugeridos por el Gabinete de Tele Educación (GATE) de la Universidad Politécnica de Madrid (UPM, 2004), estos indicadores se agrupan en 5 categorías:

### 1. Información Técnica

#### 1.1 Características generales

Evalúa fundamentalmente si se adapta en lo básico a las necesidades y posibilidades de la institución

Aspecto a Evaluar	Evaluación de Claroline
a) Idiomas	Es multilinguaje
b) Numero de usuarios permitidos	Ilimitado, depende hardware
c) Numero de cursos	Ilimitado, depende hardware
d) Adaptable	Permite integrar material didáctico
e) Precio	Gratis
d) Servicio Técnico	Muy pobre

#### 1.2 Compatibilidad

a) Requisitos Hardware y Software	Ver Anexo A
d) Estándares IMS, AICC, ADL-SCORM, LOM	ADL-SCORM

#### 1.2 Robustez

Evalúa la fiabilidad del uso de la plataforma tanto desde el punto de vista de su consistencia como de su protección.

a) Integridad de funcionamiento	No fallas reportadas
---------------------------------	----------------------

### 2. Edición de Materiales

Es la gama de propuestas formativas por Internet..

#### 2.1 Edición de contenidos

Evalúa las herramientas y sus capacidades para la edición de contenidos pedagógicos

a) Capacidad de conocimientos técnicos requeridos y facilidad de aprendizaje	Básico
d) Manejo de archivos para trabajar en remoto	Si, mediante enlace

## 2.2 Edición de elementos de evaluación

Evalúa la riqueza de plantillas disponibles para generar elementos de evaluación y actividades de refuerzo

- |   |  |
|---|--|
| a) tipo de preguntas                    | La plataforma provee exámenes de opción múltiple, las demás actividades de refuerzo vendrían contenidos en el OA |
| b) Riqueza de recursos en cada pregunta | Pobre, depende del OA  |

## 2.3 Creación de otros elementos complementarios

- |                         |  |
|-------------------------|--|
| a) Tipos de actividades | Foros, chat, repositorio de documentos |
|-------------------------|--|

## 3. Proceso de Enseñanza -Aprendizaje

Prestaciones de la plataforma en lo referente al desarrollo de los cursos

### 3.1 Planificación de un curso

- |  |           |
|--|-----------|
| a) Guía del curso                                    | Si aplica |
| b) Integración de contenidos, actividades y recursos | Si aplica |

### 3.2 Supervisión y control del desarrollo del curso

- |   |                  |
|---|------------------|
| b) Capacidad de automatizar tareas de control | Si aplica        |
| d) Generación de un expediente del alumno     | En construcción. |

### 3.3 Herramientas específicas del alumno

- |                           |               |
|---------------------------|---------------|
| a) Acceso al material     | Aplica en 90% |
| b) Control de su progreso | Regular       |

### 3.4 Servicios de Comunicación

- |                       |           |
|-----------------------|-----------|
| a) Calendario         | Si aplica |
| b) Correo electrónico | Si aplica |
| c) Foros de Discusión | Si aplica |
| d) Trabajo en grupo   | Si aplica |
| e) Chat               | Si aplica |
| f) Pizarra compartida |           |
| g) Tablón de anuncios | Si aplica |
| h) Audioconferencia   | No aplica |
| g) Videoconferencia   | No aplica |

## 4. Administración y Gestión Académica

Todos los aspectos relacionados con la admisión y matriculación de los estudiantes en los diferentes cursos, gestión administrativa y elaboración de informes.



#### 4.1 Gestión de usuarios

- d) Gestión de la matriculación de los alumnos a las asignaturas Si aplica

#### 4.2 Ordenación Académica

- a) Capacidad para dar de alta cursos y convocatorias Si aplica  
b) Capacidad para obtener listados de alumnos por curso y convocatoria En construcción

#### 4.3 Informes

- a) Elaboración de estadísticas e informes sobre el trabajo y aprendizaje de alumnos En construcción

#### 5. Otras características

Son aspectos que en alguna medida son subjetivos y que están relacionados con la confianza que *a priori* nos ofrece la plataforma.

##### 5.1 Madurez y Difusión

- a) Historial de la Plataforma Consistente  
d) Evaluaciones realizadas por organismos e instituciones independientes en estudios comparativos. Evaluado por Cátedra UNESCO, Educación a Distancia, UNED

##### 5.2 Interface de usuario

Evalúa tanto la sencillez de manejo como la comodidad y amabilidad de uso

- a) Diseño y navegación Amigable  
d) Trabajo off-line, permite descargar el curso o parte de el. No implementado

##### 5.3 Orientaciones para su correcto manejo

Evalúa la cantidad y calidad del material disponible para aprender a utilizar correctamente la plataforma.

- a) Tutoriales y documentación No aplica  
b) Ayudas on-line No aplica

La plataforma tiene la fortaleza de contener la mayoría de herramientas de aprendizaje integradas en un solo sistema y proveer el código para adaptarlo a las necesidades regionales, las desventajas es el trabajo en investigar y aprender los aspectos técnicos para ajustarlo a las necesidades específicas ya que hay poco material documental de apoyo y muy poca asistencia técnica en estas plataformas; en términos generales son las mismas ventajas y desventajas de la mayoría del software de licencia pública.

La propuesta de esta metodología es resumir las actividades necesarias para lograr una plataforma ajustada a las necesidades locales o regionales y que incluya el material del curso en un formato de repositorio de Objetos de Aprendizaje.

### III. Arquitectura de la propuesta de aplicación

El sistema Claroline en la pantalla principal despliega categorías de área conocimientos donde se puede dar de alta cursos afines a cada categoría, por cada curso provee las herramientas de aprendizaje colaborativo como foros, repositorio de documentos, agenda, chat, asignación de áreas, ligas, administración de perfiles de usuarios conformando un entorno de aprendizaje amigable.

La propuesta de la aplicación consiste en utilizar una de las opciones de Claroline para enlazarla a un repositorio de objetos de aprendizaje que se utilizarán para los diversos cursos, además de temas relacionadas que se consideren que puedan apoyar el proceso de aprendizaje.

A continuación se muestra el esquema del sistema:

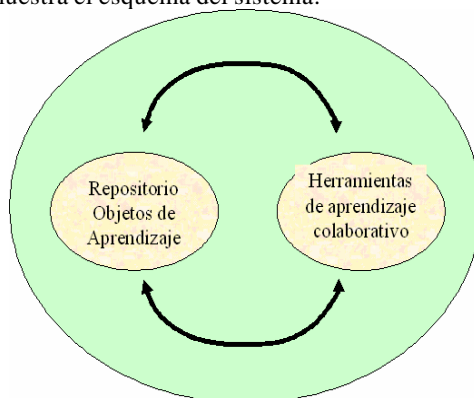
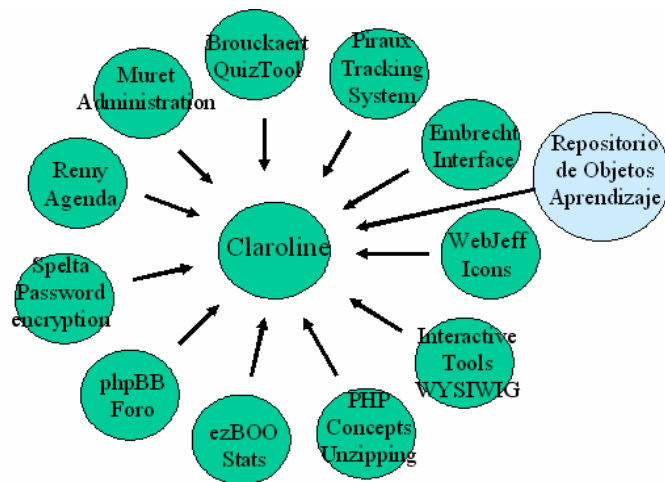


Fig. 1. Esquema de Interacción del entorno de aprendizaje

### IV. Arquitectura de Claroline e Integración del Repositorio de Objetos de Aprendizaje

Claroline es un sistema integrado con los principales programas y segmentos de programas que ha desarrollado la comunidad de código libre, por su misma arquitectura permite la interacción con otros componentes, como es el caso de un repositorio de Objetos de Aprendizaje, la siguiente gráfica muestra los programas y sus autores, además de la integración del repositorio de OA.



**Fig. No. 2** Integración de las distintas herramientas de aprendizaje en Claroline incluyendo el Repositorio de Objetos de Aprendizaje.

## V. Diseño y construcción de contenidos del curso basado en Objetos de Aprendizaje

El contenido del curso se guardará en un documento con un formato que sea rico en expresión gráfica, con hipertexto, que pueda enlazarse a otros recursos y compatible con el entorno donde se trabaja que es el Internet, estamos hablando de HTML que propiamente es la lengua nativa de Internet.

La Internet ó World Wide Web fue diseñada como un universo interactivo de información compartida a través de personas que puedan comunicarse entre si por medio de maquinas (Berners-Lee, 1996), para este fin se pensó hacerlo de la manera más simple utilizando un protocolo de comunicación HTTP y un lenguaje de formato de documentos HTML.

El protocolo con la velocidad suficiente para transferir hipertexto es el HTTP (HyperText Transfer Protocol), hipertexto se refiere a la presentación de información como una liga a una red de nodos donde los lectores son libres de navegar en un modo no lineal, lo que permite múltiples autores y que sea difícil de definir los límites de esta información (Nelson, 1965).

Las siglas de HTML son HyperText Mark-up Language; Mark-up es un término de imprenta que significa el conjunto de instrucciones estilísticas detalladas escritas en un manuscrito que debe ser tipografiado. Así, HTML podría ser traducido como "Lenguaje de Formato de Documentos para Hipertexto".

El material del curso específico deberá ser escrito en HTML y en unidades de aprendizaje u Objetos de Aprendizaje, con la intención de mejorar el proceso de enseñanza-aprendizaje; cada Objeto de Aprendizaje deberá contener: el insumo informativo, representación por diferentes canales perceptuales, contexto de uso, proceso o problema a resolver, las estrategias de aprendizaje, generación de producto de aprendizaje y cualquier otro apoyo al proceso de enseñanza-aprendizaje (Chan,2002).

Se pueden crear documentos HTML utilizando un editor de texto, pero se recomienda utilizar herramientas especializadas para este fin como Macromedia, FrontPage, WYSEWIG, etc.

## **VI. Evaluación de la propuesta**

El primer criterio son los indicadores de la UPM (UPM,2004) respecto a la plataforma, la evaluación del diseño instruccional del curso será mediante la aplicación del mismo a nivel medio superior, con una proporción 80% en línea y 20% presencial, esta proporción es por ser alumnos de primeros semestres y todavía no tienen la clara la mecánica del curso.

## **Conclusiones**

La revolución del mundo de la información con el desarrollo de tecnologías de computo y telecomunicaciones, han generado opciones como las licencias públicas de software, que son ideales para probar prototipos de cursos, de modelos pedagógicos, migración de materiales didácticos y creación de materiales didácticos adecuados a estos entornos virtuales antes de decidir si se utilizan plataformas comerciales o libres.

## **Referencias Bibliográficas**

- ADL SCORM <http://www.adlnet.org>, última visita Abril 2004
- Berners-Lee, Tim. 1996. Publicación especial de Computación IEEE Octubre 1996, el autor es creador del Internet y Director del Consorcio World Wide Web y científico M.I.T.
- Chan, Maria Elena 2002. Revista Apertura, Marzo 2002 Conferencias Magistrales del X Encuentro Internacionales de Educación a Distancia "Hacia la construcción de la sociedad del aprendizaje"
- Claroline, 2001. Software desarrollado principalmente por Thomas De Praetere, Hugues Peeters (peeters@ipm.ucl.ac.be), Christophe Gesché - aka Moosh (gesche@ipm.ucl.ac.be) en la Universidad Católica de Louvain, Bélgica.
- Claroline, 2004. <http://www.claroline.net/credits.htm>
- Dublín Core Metadata Initiative, <http://www.dublincore.org> última visita Abril 2004
- GNU GPL, 2004. GNU General Public License, proyecto que define la libertad de usar, modificar y distribuir software gratuitamente o con contraprestación. <http://www.gnu.org/gnu/thegnuproject.html>

IEEE 1484.12.1, Learning Technology Standard Committee, Learning Object Metadata standard (LOM), Draft Standard for Learning Object Metadata, Diciembre 2002.  
IMS Global Learning Consortium, Inc. <http://www.imsproject.org> última visita Abril 2004  
Nelson, Theodor Holm 1965, "A File Structure for the Complex, the Changing, and the Indeterminate", 20va conferencia anual de la Association of Computer Machinery (ACM), UCL, 2004. Universidad Católica de Louvain, Bélgica, universidad creada el 9 de Diciembre de 1425 a instancia del Duke Juan IV de Bravant. <http://www.ucl.ac.be/>  
UPM, 2004, Universidad Politécnica de Madrid, Gabinete de Tele Educación (GATE), <http://www.upm.es/>  
[http://www.gate.upm.es/plataformas/plataformas\\_vistas/documentos/indicadores.pdf](http://www.gate.upm.es/plataformas/plataformas_vistas/documentos/indicadores.pdf)  
Wiley, David A. (2001), Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy. Utah State University.

## **Anexo A**

Visitar el sitio <http://www.claroline.net/download.htm>, ahí se señala la versión mas estable, se recomienda la versión 1.4.2., los requerimientos para instalar Claroline en un servidor local o remoto son los siguientes:

### **Software**

- Servidor web (se recomienda Apache, aunque otros pueden funcionar)
- El software PHP 4.\* o recientes
- La base de datos *MySQL* versión 3.23 \* o recientes

### **Hardware**

- Computadora procesador desde Pentium 3
- 256 Mb Ram,
- Disco Duro desde 10 GB
- Conexión a Internet

# Guías de diseño para el desarrollo de objetos de aprendizaje

Jorge Aguilar Cisneros<sup>1</sup>, Jaime Muñoz Arteaga<sup>2</sup>, Saúl Pomares Hernández<sup>3</sup>

<sup>1</sup>Universidad Tecnológica de Puebla, Antiguo Camino a la Resurrección # 1002-A  
Zona Industrial, Puebla, Pue., México, C.P. 72300,  
jaguilar@utpuebla.edu.mx

<sup>2</sup>Departamento de Sistemas de Información de la Universidad Autónoma de  
Aguascalientes Av. Universidad # 940, Fraccionamiento Bosques, C.P. 20100,  
Aguascalientes, México.  
jmunozar@correo.uaa.mx

<sup>3</sup>Instituto Nacional de Astrofísica Óptica y Electrónica, Calle Luis Enrique Erro #1,  
Tonatzintla, Pue., México, Apartado Postal 51 y 216, 72000,  
spomares@inaoep.mx

**Resumen:** En este artículo se presenta una guía sobre los elementos que deben ser considerados cuando se diseñan y crean objetos de aprendizaje. La guía propuesta toma como referencia las características básicas de los objetos de aprendizaje: accesibilidad, reusabilidad e interoperabilidad. Se consideran además, aspectos pedagógicos para el diseño y creación de éstos. Los aspectos pedagógicos contemplados en este artículo se refieren a los diferentes estilos de aprendizaje presentes en los usuarios de los objetos que se desarrollen. Al desarrollar objetos de aprendizaje con cualidades pedagógicas se podrá garantizar de cierta manera el aumento en el porcentaje de retención de información de los usuarios y aplicado en un contexto educativo, disminuir el índice de reprobación. La creación de objetos de aprendizaje desde la perspectiva de este artículo, se evaluará en el dominio de las matemáticas universitarias.

**Palabras Clave:** Objetos de aprendizaje, Guías de diseño, Estilos de aprendizaje.

## 1. Introducción

Este artículo hace una contribución en el dominio de contenidos educativos digitales. Para ello presenta una propuesta para la construcción de objetos de

aprendizaje los cuales puedan ser insertados en distintos ambientes de aprendizaje.

Tres son las aportaciones de esta propuesta: i) la determinación de los principales elementos que deben constituir un objeto de aprendizaje, ii) el desarrollo de guías para el diseño y construcción de objetos de aprendizaje usando los elementos que se han determinado, y iii) la inclusión de atributos pedagógicos en los objetos de aprendizaje creados, los cuales se refieren a la consideración de diferentes estilos de aprendizaje.

En nuestra propuesta, considerar diferentes estilos de aprendizaje implica crear un objeto de aprendizaje que presente la información del mismo tópico, de diferentes formas: textual, gráfica, simulaciones, animaciones, etc.

Una vez que se han construido objetos de aprendizaje basados en las guías de diseño, estos deben conservar las características de: accesibilidad, reusabilidad e interoperabilidad.

## **2. Problemática**

Actualmente existen esfuerzos nacionales e internacionales para la descripción de objetos de aprendizaje a través de metadatos [1,2,3,4]. Sin embargo, un área que debe ser considerada para su investigación es la referente al desarrollo de guías de diseño de objetos de aprendizaje que permitan la construcción de éstos, de manera homogénea [5,6,7,8].

Por otro lado, investigaciones en el área educativa han permitido determinar que existe un grave problema de aprovechamiento en el dominio de las matemáticas en los diferentes niveles de enseñanza del sistema educativo nacional [17].

En este artículo se abordan los problemas descritos al proponer guías de construcción de objetos de aprendizaje. Los objetos de aprendizaje creados serán evaluados en cursos de matemáticas de nivel superior, se espera aumentar el porcentaje de aprovechamiento de esta materia y disminuir el índice de reprobación y deserción de este tipo de cursos.

## **3. Definición, características y clasificación de Objetos de Aprendizaje**

Actualmente no existe una definición única de objeto de aprendizaje(OA). En este artículo la definición que se adopta es: “un OA es cualquier contenido digital utilizado en un contexto de aprendizaje y accesible a través del Internet”[9]. Al igual que en la definición, en la clasificación no existe un estándar, la adoptada en este artículo es la que propone David Wiley[10].

Existen tres características básicas de un objeto de aprendizaje[11]: i)

*Accesibilidad.* El OA debe ser etiquetado, ésto permite su almacenamiento y referencia. *ii) Reusabilidad/Adaptabilidad.* Debe ser funcional para varios contextos de aprendizaje. *iii) Interoperabilidad.* Independiente del medio de entrega y del sistema de administración de aprendizaje.

Los tipos de OAs son [10]:

**1. Fundamental.-** Recurso digital individual, que no está combinado con ningún otro [12], ejemplo, un documento de texto sobre un sólo tópico.

**2. Combinación cerrada.-** Un conjunto de recursos digitales combinados por el autor en el momento del diseño del OA, estos recursos digitales no pueden ser accedidos individualmente, sino que se accesan en conjunto. Por ejemplo un video clip, que combina audio y video, no se puede acceder de forma independiente el audio o el video.

**3. Combinación abierta.-** Un conjunto de recursos digitales combinados por una computadora en respuesta a una solicitud. Estos también pueden ser reusados de forma individual. Por ejemplo, una página web generada dinámicamente que combina audio, texto, video.

**4. Presentación generativa.-** Un OA de presentación generativa tiene una alta reusabilidad intra-contextual (Pueden ser usados una y otra vez en contextos similares), pero tienen una baja reusabilidad extra-contextual (no pueden ser usados en dominios para los cuales no fueron generados).

**6. Generativa instruccional.-** Este tipo de OAs es concebido para evaluar la capacidad de un aprendiz de recordar una serie de pasos. En otras palabras, soportan estrategias instruccionales abstractas como por ejemplo: recordar y ejecutar una serie de pasos. Por ejemplo la identificación de la molécula de agua. Se presenta la teoría de los átomos que forman la molécula. Se muestra la teoría de cómo combinar estos átomos para obtener dicha molécula. Se pide al aprendiz que identifique esta molécula a partir de un dibujo.

Tomando en cuenta estos seis tipos de OAs, después de que se determinaron los elementos que deben de incluir los OAs, el tipo de OA que mejor se adapta al OA creado a partir de nuestra guía es el denominado combinación cerrada.

#### **4. Estilos de aprendizaje**

Existen diferentes teorías de estilos de aprendizaje, este artículo se basa en la teoría propuesta por David Kolb[14]. En ésta se asume que para adquirir un conocimiento se debe trabajar o procesar la información que se recibe. La información se puede recibir de dos formas: i) de una experiencia directa y concreta, ésta se tiene cuando hacemos, visualizamos o experimentamos algo; ii) de una experiencia abstracta, ésta se tiene cuando leemos acerca de algo o cuando alguien nos lo cuenta.

La experiencia que se tenga, concreta o abstracta se transforma en conocimiento cuando la procesamos de alguna de estas dos formas : i) reflexionando y



pensado sobre ellas, ii) experimentando de forma activa con la información recibida. En la figura 1 vemos gráficamente lo antes expuesto.

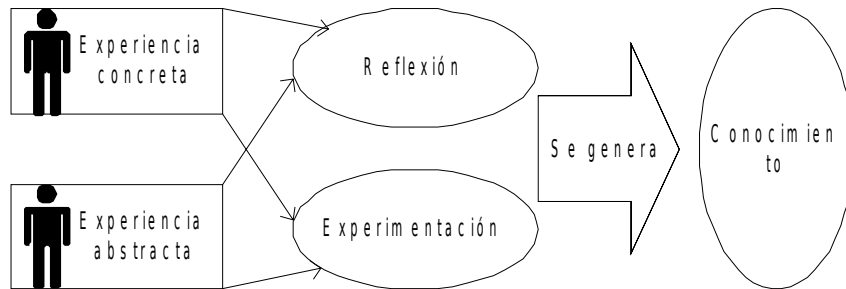


Fig. 1. Diagrama de transformación de experiencias en conocimiento.

## 5. Elementos de un objeto de aprendizaje

En la propuesta de este artículo, se sugiere que al desarrollar un OA, éste contenga 5 elementos: Teoría, experimentación, evaluación, colaboración y relación, como se muestra en la figura 2. De acuerdo a la clasificación de OAs, la propuesta que estamos presentando corresponde a la categoría combinación cerrada. Otros autores clasificarían a este OA, como compuesto[5]. Un objeto compuesto consiste de dos o más objetos fundamentales (Existen propuestas de desarrollo de OAs basadas en objetos fundamentales[6]).

Las siguientes subsecciones detallan los elementos de la propuesta.

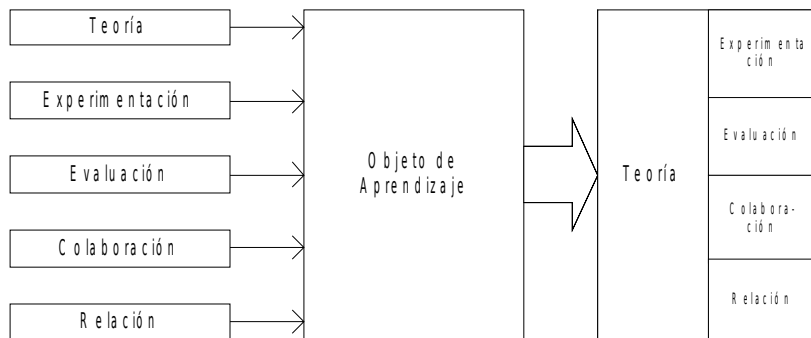


Fig. 2.- Esquema general de un objeto de aprendizaje

### **5.1 Teoría**

Esta área contiene información sobre el tópico del AO. Esta área favorece a los usuarios que tenga un estilo de aprendizaje abstracto debido a que encontrarán conceptos teóricos sobre el tema abordado.

### **5.2 Experimentación**

Esta área contiene animaciones, simulaciones etc. Esta área favorece a los usuarios que tengan un estilo de aprendizaje concreto. En esta área el usuario podrá experimentar y reflexionar los conceptos mostrados en el área de teoría. En esta área los usuarios obtienen una experiencia directa y concreta.

### **5.3 Evaluación**

En esta área se evaluará el conocimiento adquirido en el área de teoría y experimentación.

### **5.4 Colaboración**

En esta área se pueden hacer comentarios sobre el OA. Además, se pueden visualizar comentarios de otros usuarios. Mediante estos comentarios se hace una socialización del conocimiento. La colaboración se entiende como cualquier actividad que un par de individuos, o un grupo de gente hacen juntos [15].

### **5.5 Relación**

Esta área contiene ligas a otros OAs relacionados con el tópico. A través de ellas se pueden acceder otros OAs.

Estos 5 elementos permiten la construcción de OAs, en los cuales se considera el área pedagógica como parte fundamental para el proceso de aprendizaje. Para acceder estos objetos de aprendizaje, se ha desarrollado un sistema llamado ArROBA [16] que se describe en otros artículos.

## 6. Guías de diseño de objetos de aprendizaje

A continuación se presentan las guías de diseño para la construcción de un OA, en base a los elementos descritos en la sección anterior.

1. El OA debe especificar un objetivo general mediante el cual se ubique el contexto de aprendizaje. Este objetivo se localizará en el área de texto.
2. El OA desarrollado debe ser visualizado preferentemente en un espacio que alcance totalmente en la pantalla, esto quiere decir que se debe evitar el uso de barras de desplazamiento.
3. La teoría debe ser escrita siguiendo algún método instruccional, por ejemplo: i) *pistas tipográficas y subrayados*.- Señalamientos que se hacen en un texto para enfatizar y organizar ciertos elementos de contenido, ii) *ilustraciones*.- Representación visual de los conceptos expuestos en la teoría, iii) *preguntas intercaladas*.- Preguntas insertadas en un texto, a través de estas se mantiene la atención y retención del tópico estudiado. Se recomienda usar guías de colores para las fuentes y los fondos.
4. En el área de experimentación se deben colocar animaciones que pueden ser creadas a través de Flash o algún otro software que permita la animación, se pueden colocar applets que nos permitan la simulación o se pueden incluir videos.
5. El área de evaluación puede contener preguntas utilizando los siguientes tipos: i) *relación*.- Se presentan preguntas del lado izquierdo y sus respuestas del lado derecho (dentro del área de evaluación), el usuario hace una relación entre ellas, ii) *opción múltiple una sola opción válida*.- El usuario visualiza varias opciones para la respuesta de la pregunta y sólo una es verdadera, iii) *opción múltiple, múltiples opciones válidas*.- El usuario visualiza varias opciones para la respuesta de la pregunta y más de una son verdaderas, iv) *rellenar espacios*.- El usuario visualiza una pregunta y un espacio vacío en el cual debe escribir al complemento de la pregunta. Estas preguntas serán procesadas y retroalimentarán al usuario inmediatamente.
6. En el área de colaboración se visualizará un foro en el cual los usuarios puedan interactuar, también se puede agregar una herramienta que permita al usuario otorgar una calificación al OA.
7. En el área de relación el usuario encontrará ligas hacia otros OAs relacionados con el que se está estudiando, se recomienda no presentar más de 3 ligas con el objetivo de no ocasionar confusión y carga cognitiva.

Al seguir las guías antes referidas podemos obtener OAs como el que se muestra en la figura 3.

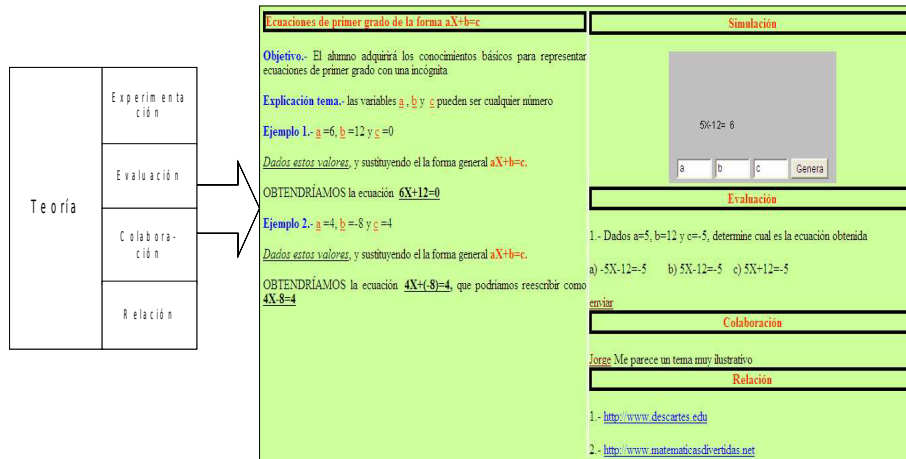


Fig. 3.- Creación de un OA con los elementos propuestos.

## 7. Evaluación

El marco de evaluación de los objetos de aprendizaje creados mediante las guías propuesta en este artículo, será el curso de Matemáticas I, impartida dentro del plan curricular de la carrera de Telemática de la Universidad Tecnológica de Puebla.

La etapa de evaluación se iniciará en el periodo septiembre-diciembre del 2004 y los resultados de esta evaluación se presentarán en artículos posteriores.

En esta evaluación se preguntará a los estudiantes sobre la experiencia que han tenido al interactuar con material presentado de esta forma, también se analizarán sus preferencias sobre la presentación en formato texto, sobre las animaciones, simulaciones y las evaluaciones. Se analizará el horario en que los objetos de aprendizaje fueron accedidos para determinar el interés por el material. Se evaluarán las propiedades pedagógicas presentes en los objetos de aprendizaje desarrollados para implementar, en caso que sea necesario, las adecuaciones pertinentes. Estos estudios nos permitirán determinar la efectividad de las guías de diseño propuesta.

## 8. Conclusiones

Contar con guías de diseño de objetos de aprendizaje facilitará el trabajo de los desarrolladores de contenido educativo en formato de objetos de aprendizaje.

El desarrollo de OAs compuestos o de combinación cerrada presentan ventajas sobre los objetos fundamentales o básicos, por que, además de poder ser reusables, permiten incluir características pedagógicas que no se pueden contemplar en

los objetos de aprendizaje fundamentales.

Existe mucho trabajo de investigación por desarrollar en el área del diseño de objetos de aprendizaje, en este artículo se presenta un avance para facilitar la producción de contenido educativo en el formato de objetos de aprendizaje. La evaluación de los objetos de aprendizaje desarrollado bajo el formato propuesto, nos permitirá marcar líneas de acción para mejorar la producción de los mismos en beneficio de los usuarios de ambientes de aprendizaje basados en Web.

## Referencias

- [1] IEEE 1484.12.1, Learning Technology Standards Committee, Learning Object Metadata standard , Draft Standard for Learning Object Metadata, Diciembre 2002.
- [2] Dublin Core Metadata Initiative, <http://dublincore.org/>, última visita, mayo 2003..
- [3] IMS Global Learning Consortium, Inc., <http://www.imsproject.org> , última visita julio 2004.
- [4] ADL SCORM, Advanced Distributed Learning, <http://www.adlnet.org> ,última visita julio 2004.
- [5] C. Claire, T. Boyle, *The development and deployment of multimedia learning objects*, Learning Technology Research Institute, London Metropolitan University, Learning Objects Symposium 2003, Educational Multimedia and Hypermedia, Honolulu, USA, june 2003.
- [6] T. Boyle, *Desing principles for authoring dynamic, reusable learning objects*, Learning Technology Research Institute, London Metropolitan University, ASCILITE 2002, Auckand New Zeland, december 2002.
- [7] A. Damle, *Learning Styles and Learning Objects*, The Ohio State University, LO: context and connections. 2003.
- [8] R. Roberts, *Creating Learning Objects*. SDSU Educational Technology, <http://coe.sdsu.edu/eet/Aricles/learnobject>, última visita julio 2004.
- [9] J Aguilar Cisneros, *Adminsitración de contenidos digitales para ambientes de aprendizaje basados en Web*, Departamento de Sistemas Computacionales. Universidad de las Amèricas Puebla, diciembre 2003.
- [10] D. A. Wiley, *Connectig Learning Objects to instructional design Theory: A definition* , Sistemas Computacionales. Universidad de las Amèricas Puebla, diciembre 2003.
- [11] P. R. Polsani, *Use and Abuse of Reusable Learning Objects*, University of Arizona, USA, Learning Technology Center, 2000.
- [12] D. A. Wiley, L. M. Nelson, *The fundamental object*, <http://wiley.ed.usu.edu/docs/fundamental.html>, 1998.
- [13] M. D. Merrill, *Instructional transaction theory (ITT): Instructional design based on knowledge objects*, . In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 397-424, 1999
- [14] H. Henke, *Learning Theory:Applying Kolb's Learning Style Inventory with computer based Training* , June 2001
- [15] L. Lipponen, *Exploring Foundations for Computer-Supported Collaborative Learning*, Department of Psychology University of Helsinki, Proceeding of: CSCL January 7-11,2002.

- [16] J. Aguilar Cisneros, J.L. Zechinelli Martini, J. Muñoz Arteaga. *Hacia la creación y Administración de repositorios de objetos de aprendizaje*. 4º Congreso Internacional de Ciencias de la Computación, Apizaco Tlaxcala, pp 245-250, septiembre 2003.
- [17] J. Fallad Chávez, *Software educativo para la asignatura de matemáticas 1 de nivel licenciatura*, XV Simposio intenacional de computación en la educación, SOMECE 1999.

# PROPUESTA DE IMPLEMENTACIÓN PARA GENERACIÓN DE MATERIALES MULTIMEDIA EN LÍNEA UTILIZANDO UN META ETIQUETADOR Y COMPONENTES OO CSPR(I) VERSIÓN 3.0 PARA RESOLVER LOS PROBLEMAS DE CONTENIDO, CONTROL, REUTILIZACIÓN Y SECUENCIACIÓN DINÁMICA

Rubén Peredo Valderrama, Leandro Balladares Ocaña, Cutberto García Tamayo

Instituto Politécnico Nacional  
Centro de Investigación en Computación  
Ave. Juan de Dios Bátiz S/N Esq. Miguel Othón de Mendizábal  
07738 México, D.F.  
Fax: 5586-2936; Tel.: 57296000 ext. 56611  
[peredo\[ballad\]@cic.ipn.mx](mailto:peredo[ballad]@cic.ipn.mx)

**ABSTRAC:** The development of materials for the Web has different stages and challenges, one of the fundamental challenges is the technological barrier, it simply is not a conversion of classic materials to HTML, it implies much more. From creating an interface and consistent functionality for all the materials, taking care of the bandwidth. With the purpose that the learning experience becomes a true experience multimedia and interactive attractive for the student. For all previously mentioned you development a denominated software Meta-Etiquetador (ME) written in Visual Basic .NET, being the fundamental piece of the same one the Components of Software Intelligent Reutilizables (CSPRI) v3 Object Oriented developed in ActionScript version 2.0. ME is the first phase of a tool CASE, that will allow to simplify the generation of rich educational materials with multimedia and interactividad with an interface and consistent functionality for education distance on-line.

**RESUMEN:** El desarrollo de materiales para la *Web* tienen diferentes etapas y retos, uno de los retos fundamentales es la barrera tecnológica, no se trata simplemente de una conversión de materiales clásicos a HTML, implica mucho más. Desde crear una interface y funcionalidad consistente para todos los materiales, cuidando el ancho de banda. Con la finalidad de que la experiencia de aprendizaje se convierta en una verdadera experiencia multimedia e interactiva atractiva para el alumno. Por todo lo anteriormente mencionado se desarrollo un software denominado Meta-Etiquetador (ME) escrito en *Visual Basic .NET*, siendo la pieza fundamental del mismo los Componentes de Software Reutilizables Inteligentes (CSPRI) v3 Orientados a Objetos desarrollados en *ActionScript* versión 2.0. El ME conforman la primera fase de una herramienta CASE, que permitirá simplificar la generación de materiales educativos ricos en multimedia e interactividad con una interface y funcionalidad consistente para educación a distancia en línea.

**Palabras Claves:** Meta-Etiquetador, CSPR(I) v3, educación a distancia en línea, OOP.

## **I. Introducción y estado de arte**

El requerimiento global urgente de recursos de aprendizaje, prácticos, reusables y de alta calidad, que involucra innovación y entrega de tecnología de aprendizaje, marca hoy en día la panorámica del desarrollo de sistemas WBT. Los principales aspectos en los que se centra la investigación son reusabilidad, accesibilidad, durabilidad e interoperabilidad de materiales didácticos y ambientes de educación virtual. Diferentes consorcios de universidades, instituciones gubernamentales y empresas están involucrados en el proceso de la búsqueda de soluciones del problema de integración adecuada de la tecnología con avances en la pedagogía. Las principales iniciativas son las siguientes: la Iniciativa de Conocimiento Abierta de MIT (*Open Knowledge Initiative* - OKI), el Aprendizaje Distribuido Avanzado Co -Lab (*Advanced Distributed Learning* - ADL) del departamento de defensa y la oficina de ciencia de la Casa Blanca y política tecnológica y el IMS Consorcio de Aprendizaje Global (*IMS Global Learning Consortium*). El IMS desarrolla especificaciones técnicas abiertas para apoyar aprendizaje distribuido. Estas especificaciones forman parte de los estándares y las herramientas que se están desarrollando en el marco del modelo de referencia SCORM (*Sharable Content Object Reference Model*), cuyo propósito es asegurar el acceso a educación de alta calidad y materiales de entrenamiento que pueden ser personalizados a las necesidades de los alumnos individuales y hechos disponibles cuando quiera y donde quiera que estos sean requeridos [1-5].

## **II. Metodología usada**

En el presente artículo, se expone el ME en conjunción con sus constituyentes los CSPR(I) v3, con el fin de simplificar el desarrollo de materiales didácticos multimedia y producir un material rico en multimedios, interactividad, retroalimentación y meta-etiquetado, para WBT. El modelo propuesto es compatible con las especificaciones de SCORM [4] y busca resolver los problemas de contenido, control, reutilización y secuenciación dinámica.

El resto del artículo está organizado de la siguiente manera. En la siguiente sección se presentan las bases del modelo CSPR(I) v3 *Object Oriented Programming* (OOP). La sección IV presenta los resultados experimentales de CSPRI v3 con el meta-etiquetador v1 con ejemplos de generación de módulos para los materiales. Finalmente, se discuten las conclusiones y trabajo futuro de la investigación del presente trabajo.

## **III. Programación Orientada a Objetos de los componentes CSPRI v3.0.**

La Programación Orientada a Objetos (en inglés, OOP) es el nombre dado a programas que hacen uso de objetos. Los objetos y la OOP son intrínsecos a la programación en *ActionScript* v 2.0 (AS 2.0), de *Flash MX* 2004. Para que un lenguaje sea OOP debe



permitir implementar tres características fundamentales que son: encapsulación, polimorfismo y herencia. Un objeto es una estructura que agrupa juntos propiedades y métodos relacionados. Un objeto típicamente encapsula sus funcionamientos, entendiéndose por esto que los detalles internos de cómo ejecuta sus funciones no son necesariamente visibles fuera del objeto. Un programa interactúa con un objeto vía sus interfaces. La herencia hace posible otro concepto clave de la OOP. El polimorfismo es una palabra elegante que significa: “muchas formas”. El polimorfismo por su parte, nos permite tratar objetos diferentes para que tengan funcionamientos diferentes de una manera uniforme, como la implementación de métodos con el mismo nombre. La herencia nos permite reutilizar código, con la finalidad de compartir características comunes entre clases. Una de las características cruciales de las clases dentro de la OOP avanzada es su habilidad para heredar propiedades y métodos. Una clase entera puede adoptar las propiedades y métodos de otra clase y adicionar sus características particulares. Las clases que sirven de molde para otras reciben el nombre de superclase, mientras que las que piden prestados propiedades y métodos de otra clase son conocidas como subclases. Se dice que subclase deriva de la superclase, porque hereda métodos y atributos de la superclase además de los suyos propios [6].

### **III.I. Encapsulación de componentes CSPRI v3.0.**

En AS 1.0 no hay una palabra reservada para la declaración de una clase como en C++ y Java. Se define un tipo especial de función, llamada función constructora, esta genera una nueva instancia de nuestra clase. Al definir la función constructora, se crea el molde de la clase o definición de la clase. La declaración de atributos y métodos de una clase también tenía algunas diferencias importantes con respecto a C++ y Java, primeramente se debe de utilizar la palabra reservada *this* en el caso de los atributos. Esta declaración de atributos se debe de hacer dentro de la función constructora, *this* hace referencia a la clase actualmente generada. En el caso de los métodos se debe de utilizar igualmente *this* seguida del nombre del método y sus parámetros dentro de la función constructora, y a continuación lo que debe de hacer el método.

En AS 2.0 ya hay una palabra reservada para la declaración de una clase como en C++ y Java denominada *class*.

Los métodos son funciones asociadas con los objetos y son típicamente usadas para ejecutar tareas relacionadas al objeto o acceder a propiedades del objeto, la sintaxis general para llamar a un método es la siguiente:

```
NombreObjeto.método(arg1,...,argn);
```

La manera de referenciar miembros de un objeto es con el operador punto (.) para acceder a los atributos y métodos del objeto, la sintaxis general es la siguiente:

```
NombreObjeto.miembro;
```

La Tabla 1 muestra un ejemplo de la declaración de una clase de un CSPRI v3 con su constructor.

<b>Tabla 1 Clase de unCSPRI v3 con su constructor.</b>
<pre> class Pregunta{ function Pregunta(preguntaXML:Number,Respuestacorrecta:Number, Textopregunta:String, respuestas:Array) { // Atributos Textopregunta = ""; respuestas = new Array(); Respuestacorrecta = 0; Respuestausuario = null; // Inicialización. private var Listarespuesta:Array = new Array(); if (arguments.length == 1 &amp;&amp; preguntaXML instanceof XMLnode)      { //Construye la pregunta usando XML. setRespuestaCorrecta(preguntaXML.attributes.RESPUESTA); setTextoPregunta(preguntaXML.attributes.TEXTO); for (var i=0; i &lt; preguntaXML.childNodes.length; i++) { Listarespuesta[i]= preguntaXML.childNodes[i].firstChild.nodeValue; } setRespuestas(Listarespuestas); } else { setRespuestaCorrecta(Respuestacorrecta); setTextoPregunta(Textopregunta); setRespuestas(respuestas); } } </pre>

### III.II. Polimorfismo en componentes CSPRI v3.0.

El polimorfismo permite a objetos distintos ser tratados de manera uniforme. Por ejemplo si utilizamos el método dibujar() para que cada objeto se dibuje, aun pensando que la función dibujar() para cada clase es única. La Tabla 2 muestra la sobrecarga del método evaluar de CSPRI v3 utilizando AS v2.0, en la versión 1.0 se tenía que utilizar prototype para llevar acabo esto.

<b>Tabla 2 Polimorfismo sobrecargando la función evaluar() de un CSPRI v3.</b>
<pre> CSPRIV3_FALSO_VERDADERO.evaluar = function(){ } CSPRIV3_RELACION_COLUM..evaluar = function () { } CSPRIV3_LLENADO_BLANCO.evaluar = function (){ } CSPRIV3_SELECCION_AREAS.evaluar = function () { } </pre>

### III.III. Herencia de componentes CSPRI v3.0

En la versión 2 de AS era necesario especificar una superclase para una clase dada, se debía de remplazar la propiedad prototype completamente de la clase con una nueva instancia de la deseada superclase, la herencia difiere de la forma de manejar herencia de C++ y Java. La sintaxis general usada para especificar una superclase es la siguiente:

```
SubclaseConstructor.prototype = new SuperClaseConstructor();
```

Donde *SubclaseConstructor* es la función constructora de la subclase y *SuperClaseConstructor* es la función constructora de la superclase. La Tabla 3 muestra un ejemplo de esto aplicado a un CSPRI v3.

**Tabla 3 Estableciendo *MovieClip()* como superclase de *Clip1* de un CSPRI v3.**

```
Clip1.prototype = new MovieClip();
```

Mientras que en la versión 2.0 de AS queda de la siguiente manera:

```
class Subclase extends SuperClase{ }
```

#### III.IV. Instanciación de objetos de componentes CSPRI v3.0

En la versión 1.0 de AS para crear una instancia de un objeto, se utiliza de manera regular el operador *new* con la función constructora, la cual inicializa el objeto, pero como veremos más adelante del trabajo, los componentes *Movie Clips* (MC) utilizan *duplicateMovieClip* y *attachMovie* para crear instancias de este tipo de objeto, la sintaxis general es la siguiente:

```
var manejador= new FunciónConstructora(arg1,...,argn);
```

Esta manera de instanciar objetos era muy similar a la manera de C++ y Java. Pero la instanciación de MC puede ser de dos maneras, la primera mediante la función:

```
duplicateMovieClip(target, Nombre, Profundidad);
```

En este caso era necesario tener una instancia previamente en el área de trabajo. Mientras que la segunda manera de instanciar un MC, es mediante la función:

```
Clip.attachMovie(simboloID, Nombre, Profundidad, [ObjetoInicializador]);
```

La diferencia fundamental entre estas dos funciones es que *attachMovie* no necesita de ninguna instancia previa en el área de trabajo. Mientras que en la versión 2.0 de AS se utiliza simplemente el operador *new*. La Tabla 4 muestra el código para instanciar una evaluación de relación de columnas tipo CSPRI v3.0.

**Tabla 4 Instanciación de una evaluación de relación de columnas tipo CSPRI v3.**

```
var ObjetoInicializador:RelacionColumnas = new RelacionColumnas("relacion_columnas", "rc1", 2, ObjetoInicializador);
```

### III.V. Utilización del objeto XML de *Flash* para la carga desde los archivos XML asociados al componente CSPRI v3.0

XML (Extensible Markup Language) es una manera de definir datos usando etiquetas que son similares a las etiquetas HTML, pero estas etiquetas no están predefinidas como en el caso de HTML. Se pueden crear etiquetas propias (o utilizar algún estándar de aplicación XML como en caso de SCORM) las cuales definen mucho mejor la estructura de datos propios como: *menu\_xml.xml*, *conceptos.xml* y *navegacion.xml*. El *menu\_xml* define la estructura del menú de opciones, *conceptos.xml* estructura el contenido y *navegacion.xml* estructura la secuencia de navegación del componente CSPRI v3 como se ve en la Fig.1, con esto se logra separar la parte de navegación del contenido, además de posibilitar la secuenciación dinámica de los materiales con vistas a su reutilización por medio del Sistema Manejador de Aprendizaje (por su siglas en ingles, LMS). Permittiéndonos seguir la estructura correcta XML (etiquetas anidadas una dentro de otras, todas las entradas teniendo etiquetas de apertura y cerrado), denominándose a este tipo de documentos XML “bien conformados”. Macromedia *Flash* puede leer filas que son escritas usando estas etiquetas. *Flash* no usa DTDs (*Document Type Definition*), los cuales son usados por otras aplicaciones de lectura XML para los llamados documentos XML “válidos” [7].

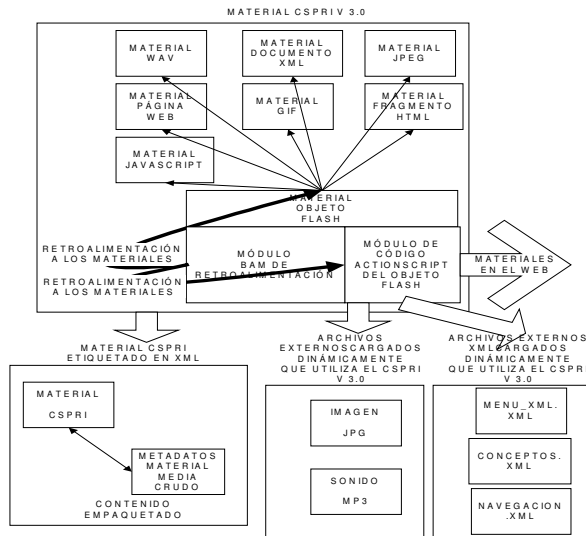


Fig. 1 Esquema general de los CSPRI(I) versión 3.

Los DTDs están siendo sustituidos por XML *Schema*, el mayor inconveniente de los DTD sobre los XML *Schema* es el no poder controlar el tipo de información que contiene un elemento o atributo en particular.

### III.VI. Pasos para leer contenido de un archivo XML dentro de un objeto XML de *Flash*

La lectura dentro de datos XML requiere de varios pasos, como en nuestro caso se requiere la lectura de tres archivos: `menu_xml.xml`, `conceptos.xml` y `navegacion.xml`. Se ejemplificara esto solo con el archivo `menu_xml.xml`, los pasos son:

1. Definir la función que ejecutara cuando los datos hayan sido cargados (`startMenu`).
2. Crear un nuevo objeto XML (`menu_xml`)
3. Establecer `ignoreWhite` a verdadero, así *Flash* no procesara retornos de carro y otros espacios en blanco dentro de la fila XML como nodos separados (`menu_xml.ignoreWhite = true`).
4. Mostrar a *Flash* cual función será usada cuando los datos han sido cargados (`menu_xml.onLoad = startMenu`).
5. Cargar los datos (`menu_xml.load("menu.xml")`).

**Tabla 5 Código para leer el archivo XML `menu_xml.xml` de un CSPRI v3.0**

```
function startMenu() {
    rootNode = menu_xml.firstChild;
    firstOption = rootNode.firstChild;
    currentOption = firstOption;
    showMenu();
}
menu_xml = new XML();
menu_xml.ignoreWhite = true;
menu_xml.onLoad = startMenu;
menu_xml.load("menu.xml");
```

### III.VII. Empaquetado del código de la clase y la librería de símbolos

Cuando se tiene una subclase MC en AS v 1.0 se pueden agregar características particulares por medio de la función constructora. Sin embargo, si buscamos compartir nuestra subclase con algún otro desarrollador. Tendríamos que explicarle varias cosas para que pudiera utilizarla. Sin embargo para simplificar su compartición se utilizan `#initclip` y `#endinitclip` válido también en AS ve 2.0, que deben de aparecer únicamente en el *frame* 1 de la línea de tiempo del MC. El código dentro del bloque es ejecutado una vez, y únicamente una vez por símbolo de MC, justo antes de que la primera instancia del símbolo sea creada. Al definir nuestra clase dentro del bloque, se garantiza que el constructor de la clase este disponible antes de que la primera instancia del símbolo aparezca. Por lo cual nuestra subclase debería de tener la forma general que se muestra en la Tabla 6.

<b>Tabla 6 Empaquetado de nuestra subclase tipo CSPRI v3.0</b>
<pre> #initclip // = = = = = // Función Constructora // = = = = = // Declaración de la superclase // = = = = = // Implementación de los métodos // = = = = = #endinitclip </pre>

### III.VIII. Programa principal de un CSPRI v3.

La llamada principal para un CSPRI v3. se ha tratado de asemejar al *main()* de un programa de C / C++, para entender de manera más clara su funcionamiento, como se muestra en la Tabla 7.

<b>Tabla 7 Código del programa principal main() de un CSPRI v3.0</b>
<pre> //===== // CARGA DE CLASES // ===== #include "cspriv3_examen.as" #include "cspriv3_examen_GULas" #include "cspriv3_examen_pregunta.as" //===== // main() //===== main();  function main () { ....;} </pre>

## IV. Presentan los resultados experimentales de CSPRI v3 con el Meta-Etiquetador v1

La Fig. 2 muestra el meta-etiquetador que es una aplicación hecha en *Visual Basic .Net*. Básicamente el ME v1 cuenta con funciones básicas tipo explorador de archivos de *Windows*. Además de permitir modificar los archivos de menú, contenido y navegación XML del componente CSPRI v3, permitiendo configurarlo de acuerdo a las necesidades del usuarios. La parte de prueba de este componente es básicamente ActiveX denominado Shockwave Flash, que permite manipular animaciones hechas en Flash, con su consecuente personalización. La Fig. 3 muestra el componente personalizado mediante los archivos XML de menú, contenido y navegación. Sobre todo se ha tratado de busca flexibilizar al máximo la construcción de los materiales.

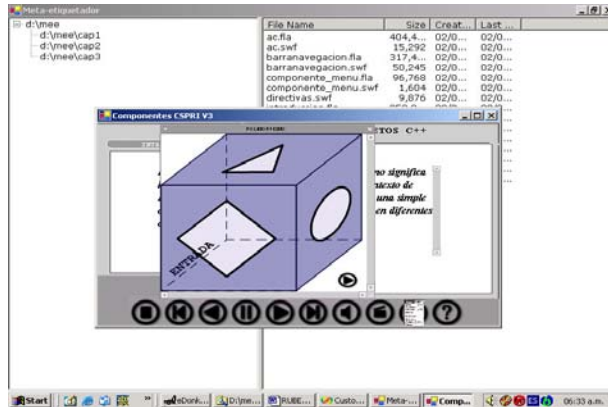


Fig. 2. Meta-etiquetador con CSPR(I)' v3.



Fig. 3. CSPR(I) v3 personalizado.

## V. Conclusiones y trabajo futuro

La utilización de componentes de software permiten simplificar de manera importante el desarrollo de materiales multimedia educativos para educación a distancia en línea. Además de permitir uniformizar los materiales en cuanto a presentación e interfaz hacia el usuario. Por el momento se tiene un módulo de navegación fijo, pero parte del trabajo futuro se basa en varios puntos fundamentales. Primero convertir todos nuestros componentes al paradigma OOP con Flash MX 2004 en conjunción con AS v 2.0 para poder reutilizarlos mejor, segundo insertar un API que ofrezca servicios desde los materiales hacia el servidor, además de permita tomar métricas de los alumnos y se vean reflejados en una secuenciación dinámica de los materiales, tercero agregarle

funcionalidades adicionales al ME para que pueda estructurar los materiales de manera jerárquica y construir las páginas *Web* de cada uno de los componentes para integrarlos a los materiales del desarrollador de material didáctico, por último construir más CSPR(I) v3 para poder estructurar mejores cursos, además de utilizar otras técnicas de enseñanza.

**Reconocimientos:** Agradecemos al IPN y al CIC por el apoyo económico prestado para el desarrollo de esta investigación.

## Referencias

- [1] Sheremetov, L. & Peredo, R., "Arquitectura de componentes de software programables reutilizables inteligentes para materiales multimedia en EVA", In Proc. of the Int. Conference CIC'01, D.F, México, 12-16 de Noviembre, 2001.
- [2] Rubén Peredo Valderrama, Oscar Fabela Cano, Edgar Jonathan Larios Tapia, Cutberto García Tamayo, Enrique Ramos Ordúñez,, "CSPR(I) versión 2.0 para materiales multimedia en educación a distancia", In Proc. of the Int. Conference ANIEI'02, Guadalajara, Jalisco, México, 23 - 25 de Octubre, 2002.
- [3] <http://web.mit.edu/oki/>
- [4] <http://www.adlnet.org>
- [5] <http://www.imsproject.org/>
- [6] Colin Moock. "ActionScript for Flash MX: The Definitive Guide". 2nd ed, USA. O'Reilly. 2002.
- [7] Dov Jacobson & Jesse Jacobson. "Flash and XML: A Developer's Guide". USA. Addison Wesley. 2001.



# Diseño e implantación de una interfaz interoperable para un patrimonio de recursos educativos basado en una red de acervos abiertos y distribuidos de objetos de aprendizaje

Víctor Germán Sánchez Arias <sup>1</sup>, Salvador López <sup>2</sup>, Rafael Morales <sup>3</sup>, Fredy Castañeda <sup>4</sup>

<sup>1</sup> Laboratorio Nacional de Informática Avanzada A.C. (LANIA)  
victor@lania.mx

<sup>2</sup> Fac. de Ciencias de la UNAM  
slm@ciencias.unam.mx

<sup>3</sup> Instituto de Investigaciones Eléctricas (IIE)  
rmorales@inf.ed.ac.uk

<sup>4</sup> Maestría en Ciencias de la Computación - LANIA  
fcs001@xico.lania.edu.mx

**Resumen.** En su sentido más amplio la educación es una actividad social en la que participan diferentes instituciones educativas y diversos tipos de participantes (estudiantes, profesores, tutores, administradores, etc.), y desde esta perspectiva consideramos que todo el conocimiento generado por esta comunidad debe ser compartido y estar al alcance de todos. Todo este conocimiento lo denominamos patrimonio educativo de una comunidad. Este trabajo se fundamenta en un modelo de e-educación, e-campus [2], que se basa en la construcción de un patrimonio de recursos educativos de uso común, compartidos y organizados en repositorios abiertos y distribuidos de objetos de aprendizaje a través de una red de instituciones y comunidades educativas que generan dichos objetos de uso común. Este trabajo es una continuación de un primer diseño general del patrimonio [2] y en este artículo se presenta una primera implantación de la interfaz interoperable para ese patrimonio de objetos.

**Palabras clave:** educación basada en tecnologías de la información, estándares para enseñanza, e-educación, objetos de aprendizaje.

## 1. Introducción.

En este trabajo presentamos los elementos de diseño de un patrimonio de recursos educativos basado en una red de repositorios distribuidos de objetos de aprendizaje. En el inciso 2, presentamos nuestro trabajo previo sobre el modelo de e-educación basado en diversos niveles de redes [2]. En el inciso 3, se presenta un diseño más detallado de la arquitectura del patrimonio basado en una red distribuida de repositorios [2]. En el 4 se presenta la implantación de un primer prototipo. En el 5 se presenta las principales conclusiones.

El modelo de e-educación basado en redes fue un proyecto realizado por LANIA y financiado por la UNESCO [1,2]. El primer prototipo del patrimonio, proyecto REDOBA [3] fue desarrollado conjuntamente con la Dirección General de Servicios de Cómputo Académico (DGSCA) de la UNAM, Universidad de Guadalajara (UdeG), el Instituto de Investigaciones Eléctricas (IIE) y el Laboratorio Nacional de Informática Avanzada (LANIA), el diseño e implantación de la red de repositorios fue realizada por las dos últimas instituciones y con la participación de la Facultad de Ciencias de la UNAM. El proyecto REDOBA fue financiado por la Corporación de Universidades para el Desarrollo de Internet 2 en México (CUDI).

## 2. Modelo de e-educación basada en redes.

El modelo de e-educación sobre el cual se basó el modelo de e-campus [1,2], se fundamenta en aplicaciones basadas en redes y la tecnología Web a través de cuatro niveles de red.

**1. Red de Servicios básicos de Internet-WEB.** Es el nivel más básico de la red y se conforma por todos los servicios de interconexión y de acceso a recursos de cómputo a través de una interfaz web (navegador) estándar.

**2. Red de conocimiento.** Esta red está conformada por el conjunto de repositorios distribuidos de recursos educativos organizados en unidades básicas de conocimiento denominadas objetos de aprendizaje.

**3. Red de aprendizaje.** Esta red conforma las comunidades de aprendizaje conformadas a su vez por tres tipos de usuarios: **P:** generador de conocimiento, **E:** estudiante que adquiere conocimiento y **A:** el que administra conocimiento, estudiantes y generadores de conocimiento (estas definiciones de funciones son sólo conceptuales pues una entidad podría tanto generar, adquirir, como administrar conocimiento).

**4. Red de apoyo basado en agentes.** Esta red está conformada por un conjunto de gentes (componentes computacionales basados en conocimiento y con capacidad de cooperar entre ellos) para apoyar a los diferentes usuarios de la red en los procesos de generación, adquisición, búsqueda y administración de conocimiento.

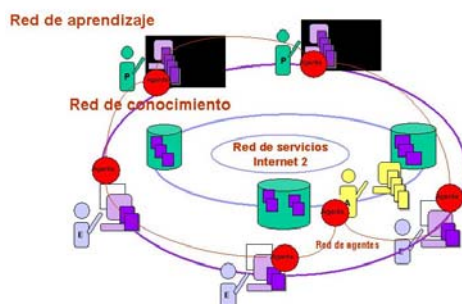


Fig. 1 Modelo de e-educación basada en redes

## 3. Diseño de la interfaz de interoperación del patrimonio.

El patrimonio como fue definido en [2], éste estaría conformado por un conjunto de repositorios de materiales educativos digitalizados distribuidos geográficamente y almacenados en los sitios locales de una comunidad específica de instituciones. El patrimonio sería el acervo común de la comunidad y cada centro produciría sus propios materiales y estos serían compartidos por la comunidad.

### 3.1 Premisas básicas consideradas para la implantación del patrimonio.

A partir de las consideraciones anteriores se definieron las siguientes premisas básicas de diseño para la implantación del patrimonio:

- Con capacidad de integrar material educativo digitalizado ya existente
- Con capacidad de integrar sistemas de *e-learning* ya existentes en particular del tipo abierto ("open source")
- Uso y operación independiente de las tecnologías de los soportes que lo implementen
- Acceso simplificado y estándar para implementar cualquier tipo de funcionalidad.

### 3.2 Red de acervos abiertos y distribuidos y su interfaz de interoperación

La interfaz interoperable es el elemento fundamental que nos permitirá conformar la red de repositorios heterogéneos distribuidos y abiertos. De esta manera cada acervo se estructurará a través de dos componentes:

- La interfaz interoperable de objetos
- Manejador local de objetos

La **interfaz interoperable** deberá estar orientada al manejo de objetos y ser utilizada por cualquier aplicación (por ejemplo para la generación, seguimiento y administración de cursos) independientemente de la implantación de las bases de objetos y permitir la interconexión con los otros acervos de la red. Las funciones básicas que debería ofrecer esta interfaz son:

- Agregar objetos a un repositorio local
- Obtener objetos (objetos específicos o búsquedas bajo cualquier criterio) de un repositorio o de todo el patrimonio.

El **manejador de objetos** será accesible a través de una interfaz estándar y la implantación se realizará de acuerdo a los manejadores de bases de datos específicos que se utilicen en la implantación. De esta manera se podrán integrar diversos tipos de manejadores de bases de datos a través de interfaces de acceso a objetos estándares. En la figura siguiente se muestra una red de acervos con sus dos componentes.

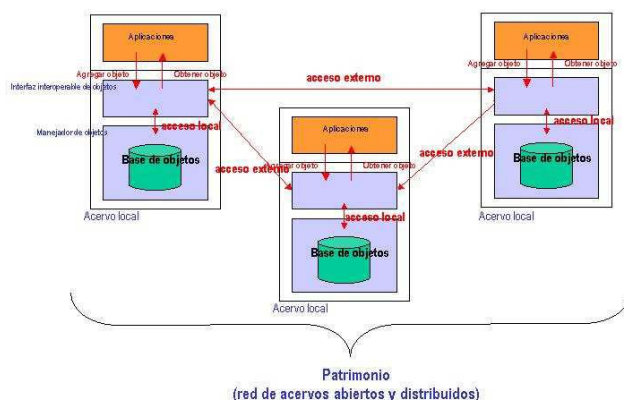
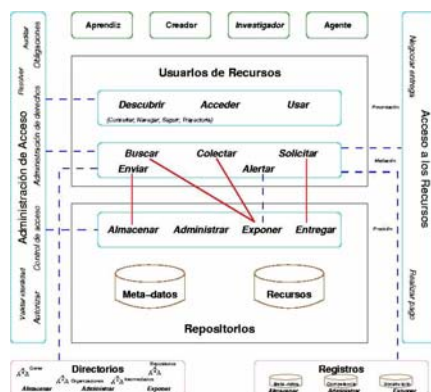


Fig. 2 arquitectura de la red interoperable de repositorios

### 3.3 La interfaz de interoperabilidad para la interconexión de los acervos distribuidos DRI-IMS

La pieza clave de nuestra arquitectura del patrimonio es la interoperabilidad entre sistemas manejadores de bases de datos heterogéneos. Para que esto sea posible fue necesario la utilización de estándares orientados a objetos de aprendizaje independientes de la implantación. Con este propósito nosotros empleamos la especificación IMS de Interoperación de Repositorios Digitales que proporciona recomendaciones para la interoperación de las funciones más comunes con el repositorio [9] que a continuación se presenta.



**Fig. 3 La interfaz de interoperabilidad DRI-IMS**

El núcleo de interacciones definidas en este modelo es el siguiente:

- Buscar/Exponer. El usuario especifica los atributos que servirán para discriminar sobre los meta-datos de los objetos de aprendizaje. La respuesta es un conjunto de meta-datos.
- Colectar/Exponer. Activación periódica del mecanismo de búsqueda.
- Enviar/Almacenar. Mecanismo para colocar objetos en el repositorio
- Solicitar/Entregar. Permite que un usuario de los recursos, que ya ha localizado algún meta-dato de su interés, solicite acceso al objeto de aprendizaje. La entrega es la respuesta del repositorio que proporciona acceso al recurso.
- (Alertar/Exponer). Interacción que aún no está completamente definida.

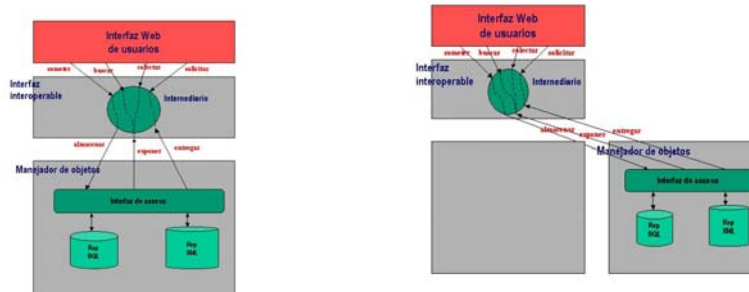
### 7.2.2 Diseño de la arquitectura para el patrimonio basado en el estándar DRI IMS

Las funciones de acceso a la interfaz y al manejador de la base de objetos son controladas por un componente denominado el intermediario.

El intermediario resolverá las peticiones de sus usuarios (cualquier tipo de aplicación basada en objetos de aprendizaje) a través de las siguientes funciones básicas de acceso: someter, buscar, colectar y solicitar.

Estas operaciones podrán ser realizadas al manejador local de objetos u otros distantes a través de las siguientes funciones de acceso a los manejadores de objetos: almacenar, exponer y entregar.

Todas estas funciones de interoperación usan el estándar por SOAP para una interconexión independiente de la implantación de los manejadores. En la siguiente figura se muestra el intermediario y las interacciones entre una aplicación (para el acceso al patrimonio basada en Web), la interfaz de interoperación y el manejador de objetos de aprendizaje.



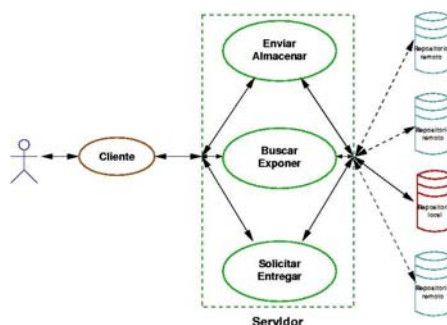
**Fig. 4 Arquitectura para el patrimonio basado en el estándar DRI IMS**

En esta versión de nuestra arquitectura solo estamos contemplando dos tipos de repositorios de objetos de aprendizaje los basados en archivos XML y en SQL.

## 4. Implantación de la interfaz de interoperabilidad y del manejador de objetos

### 4.1 Versión actual del sistema

En términos de arquitectura el bloque principal del repositorio se puede esquematizar como lo ilustra la siguiente figura:



**Fig. Implantación de la interfaz interoperable**

El cliente accede a los servicios del repositorio, que a su vez se encarga de realizar las operaciones con los objetos de aprendizaje. La comunicación entre el cliente y el servidor se implementa mediante un intercambio de mensajes SOAP, que en realidad ofrecen un servicio de comunicación tipo RPC, de tal suerte que el cliente realiza una petición y espera a que el servidor realice la tarea y entregue la respuesta.

El acceso a los objetos de aprendizaje actualmente se realiza en la misma computadora en la que se encuentran los servicios. En el caso más general este servicio se ubicará en otra computadora, ofreciendo un mejor nivel de distribución de las tareas.

Para el usuario es transparente la ubicación real de los objetos de aprendizaje, pueden estar en el repositorio local o pueden ser parte de otros repositorios. En términos gráficos del sistema se vería de la siguiente forma:

Los servicios que se ofrecen actualmente son:

#### **1. Agregar objetos al repositorio.**

Originalmente se había considerado que los objetos de aprendizaje sólo se podrían agregar de una forma: como paquetes que incluyen la descripción del objeto y el contenido. En base a las experiencias con los materiales de que se disponen, se incluyó la posibilidad de agregar solamente los metadatos o solamente el contenido.

#### **2. Buscar objetos con alguna característica especial.**

La búsqueda se realiza en base a una sintaxis especial, la del lenguaje de consulta XQuery. Esta sintaxis no es fácil de manejar para la mayoría de la gente, por lo que la interfaz de acceso al sistema ofrece la posibilidad de buscar sobre los principales campos de los metadatos.

#### **3. Extraer objetos del repositorio.**

Actualmente, la función de extraer objetos entrega el paquete completo que incluye tanto a los metadatos como al contenido mismo.

### **4.2 Tareas futuras**

Existe mucho trabajo que realizar para que el repositorio funcione mejor. Algunas cosas son nuevas funciones que se deben agregar al repositorio y otras son mejoras a las ya existentes. Por ejemplo:

1. Agregar un mecanismo de identificación de usuarios. Se necesita saber para cada usuario cuál es el conjunto de operaciones que le es permitido realizar con los objetos de aprendizaje que se encuentran en el repositorio.

2. Mejorar el mecanismo de comunicación entre repositorios. Si bien existe un intercambio básico de objetos, se ha visto la necesidad de definir un protocolo más robusto que garantice mejores niveles de funcionamiento; actualmente se está utilizando un mecanismo de comunicación tipo RPC, lo que obliga a que los repositorios trabajen en sincronía, cuando es muy posible que lo mejor sea que cada uno trabaje a su propio ritmo, lo que obligaría a usar un mecanismo de comunicación asíncrona.

### **4.3 Plataforma de desarrollo**

El prototipo de repositorio distribuido de objetos de aprendizaje, fue desarrollando con base en estándares, especificaciones, modelos y recomendaciones internacionales, como es el caso del estándar LOM del IEEE [5] sobre el contenido y estructura de los metadatos de los objetos de aprendizaje, la especificación de IMS [4], SCORM [6], de una representación de los metadatos en el lenguaje XML [9] las recomendaciones del W3C [10] para la implantación de servicios web con base en tecnologías abiertas tales como SOAP, WSDL y XQuery. El prototipo fue desarrollado sobre una plataforma de software libre, que incluye el sistema operativo Linux, el manejador de base de datos MySQL, Apache Tomcat como el manejador de componentes de Java, la biblioteca Axis, también de Apache, para la implantación del protocolo SOAP y la biblioteca XQuark Bridge para el acceso a base de datos utilizando XML y XQuery. Además, para el desarrollo del proyecto se utilizaron una gran cantidad de herramientas de software libre, además de software propietario.

## **5. Conclusiones.**

El uso software libre y estándares tuvo un costo: implantaciones más largas y menos estables, debido al uso de tecnologías en desarrollo (como SOAP, WSDL y XQuery) y de herramientas y plataformas que, aunque en muchos aspectos son excelentes, generalmente

incluyen menos funcionalidades que algunas de sus contrapartes comerciales, más costosas y propietarias. Los beneficios, sin embargo, más que justifican este costo.

Cabe enfatizar, sin embargo que el sistema desarrollado en este proyecto es un primer *prototipo*, que demuestra las posibilidades de un repositorio distribuido de objetos de aprendizaje, pero que tiene limitaciones importantes en cuanto a robustez, eficiencia, flexibilidad, capacidad y facilidad de instalación. Lo que se está entregando es una base excelente para trabajo futuro, con algunas decisiones importantes ya tomadas y evaluadas, para trabajar juntos en el desarrollo de un patrimonio de objetos de aprendizaje distribuido en la red CUDI que sea confiable, veloz, funcional y de fácil acceso.

### Otros repositorios

En la actualidad existen diversas iniciativas de repositorios como el del "OpenCourseWare2" del MIT (<http://ocw.mit.edu/>) que almacena cursos completos que no están definidos como objetos, pero existen otros que si se basan en la tecnología de objetos tales como **MERLOT** (<http://www.merlot.org/>), **CAREO** - Campus Alberta Repository of Educational Objects (<http://www.careo.org/>) o el **LOR-VCILT** University of Mauritius (<http://vcampus.uom.ac.mu/lor/>), sin embargo éstos son repositorios centralizados.

En cuanto a iniciativas de repositorios distribuidos, hay muy pocos y en general están en fase de desarrollo como por ejemplo, **ELENA/Edutella** Proyecto Europeo (<http://edutella.jxta.org/>) o el **eduSource** Canada Université Athabasca, Netera Alliance, New Media Innovation Centre, TéléEducation NB , Télé-Université , Université de Waterloo (<http://edusource.ca>). El primero utiliza el estándar RDF para la definición de objetos y el segundo utiliza los estándares IMS para la definición de sus objetos y IRD-IMS para la interoperabilidad entre repositorios. En este sentido **REDOBA**, propuesta de cuatro instituciones (UNAM, UdeG, IIE y LANIA), se apoya en los mismos estándares que el de eduSource pero sin contar con los mismos recursos financieros, este último proyecto cuenta para la fase de construcción de la red de repositorios con un financiamiento de \$ 8M de dólares durante 18 meses con la participación de seis universidades.

### Referencias.

- [1] Víctor G. Sánchez. Proyecto e-campus. Presentado a la UNESCO reunión del Colegio Internacional de la AFNIC, París 3 de Junio del 2002  
[http://www.lania.mx/~victor/presentaciones/unesco\\_2002\\_e-campus.ppt](http://www.lania.mx/~victor/presentaciones/unesco_2002_e-campus.ppt)
- [2] Víctor G. Sánchez. " Diseño de un patrimonio de recursos educativos basado en una red de acervos abiertos y distribuidos de objetos de aprendizaje" Memorias Taller Objetos de Aprendizaje, Fourth Mexican International Conference on Computer Science ENC03. Tlaxcala Mex. 8-12 Sep, 2003
- [3] Víctor G. Sánchez. Presentación del Proyecto e-campus al CUDI, Ciudad Juárez Chihuahua México 18 octubre 2002 [http://www.lania.mx/~victor/presentaciones/cudi\\_2002\\_campu.ppt](http://www.lania.mx/~victor/presentaciones/cudi_2002_campu.ppt)  
[http://www.lania.mx/~victor/pub\\_rec.html](http://www.lania.mx/~victor/pub_rec.html)
- [4] IMS Instructional Management System Global Learning Consortium Inc. (<http://www.imsproject.org/>)
- [5] IEEE LTSC Learning Technology Standards Committee (<http://ltsc.ieee.org>)
- [6] ISO/IEC JTC1/SC36 – Information Technology for Learning, Education, and Training (<http://jtc1sc36.org>)
- [7] ADL-SCORM Advanced Distributed Learning – The Sharable Content Object Reference Model (<http://www.adlnet.org>)
- [8] IMS: "IMS Digital Repositories Interoperability - Core Functions Information Model Version 1.0 Final Specification" [http://www.imsglobal.org/digitalrepositories/driv1p0/imsdri\\_infov1p0.html](http://www.imsglobal.org/digitalrepositories/driv1p0/imsdri_infov1p0.html), 13 January 2003.
- [9] Sitio Web XML: <http://www.w3.xml.org>
- [10] Sitio Web Servicios Web: <http://www.webservices.org>

# De LOM a XML y SQL, o cómo almacenar metadatos en un repositorio de objetos de aprendizaje

Rafael Morales<sup>1</sup> y Salvador López Mendoza<sup>2</sup>

<sup>1</sup> School of Informatics, Northumbria University,  
Pandon Building, Camden St, Newcastle upon Tyne, NE2 1XE, G.B.  
rmorales@unn.ac.uk

<sup>2</sup> Facultad de Ciencias, Universidad Nacional Autónoma de México  
Ciudad Universitaria, D.F., México.  
slm@fciencias.unam.mx

**Resumen.** En este artículo se describe el diseño e implementación del almacenamiento de metadatos conformes con el estándar IEEE LOM en REDOBA, un prototipo de repositorio distribuido de objetos de aprendizaje. Se analizan diversas opciones de diseño disponibles, indicando las ventajas y desventajas que cada una de ellas ofrece.

**Palabras clave:** Objeto de aprendizaje, metadatos, XML, modelo relacional.

## 1. Introducción

Uno de los aspectos más importantes a considerar en el diseño de un repositorio de objetos de aprendizaje, como es el caso de REDOBA [4], es el almacenamiento y acceso a los metadatos de los objetos, dado que el acceso a los metadatos es indispensable para buscar, localizar y seleccionar objetos de aprendizaje en el repositorio. En algunos repositorios, como MERLOT [3], únicamente se almacenan los metadatos y las referencias a los objetos correspondientes.

Para que un objeto (mediático, informativo o de aprendizaje, unidad de aprendizaje o red de conocimiento) pueda ser contenido en REDOBA es necesario que el contenido y formato de sus metadatos estén desarrollados conforme con las especificaciones de IEEE [2] e IMS [1] – conocidas de manera general como el modelo LOM para metadatos (por *Learning Object Metadata*). Esto significa, en particular, que los metadatos deben estar codificados en XML [7] de acuerdo con el esquema definido por IMS.

Existen varias maneras de almacenar información descrita originalmente en XML, entre las que cabe resaltar tres:

1. Almacenarla en archivos XML, utilizando el sistema de archivos del sistema operativo.
2. Convertir la estructura y el contenido del archivo XML en tablas, relaciones y registros de una base de datos relacional.
3. Almacenar la información en una base de datos especialmente diseñada para contener información originalmente en XML.



Cada una de estas opciones tiene ventajas y desventajas. La primera opción es fácil de implementar y no requiere de software adicional para el almacenamiento de la información, pero tiene las desventajas de hacer más difícil la búsqueda de información específica y ser lenta en su ejecución. La segunda opción, en cambio, permite el almacenamiento y acceso rápido a la información, por estar basada en la tecnología madura de las bases de datos relacionales; sin embargo, para su implementación es necesario definir mecanismos de conversión entre los archivos XML y su representación en la base de datos, mismos que deben ser ejecutados cada vez que se desea almacenar o extraer información del repositorio. La tercera opción combina las ventajas de las dos opciones anteriores, facilidad de implementación y velocidad de ejecución, pero los sistemas manejadores de bases de datos orientadas a XML son aún escasos y de creación reciente.

Para la implementación de REDOBA se han tomado las dos primeras opciones de almacenamiento de los metadatos; esto es, en su formato original, como archivos XML, y en una base de datos relacional. En las secciones siguientes se describen el diseño e implementación de ambas opciones.

Por convención, de aquí en adelante utilizaremos el término general *metadato(s)* para hacer referencia al caso particular de metadatos tipo LOM codificados en XML de acuerdo al esquema de IMS, excepto cuando explícitamente se exprese lo contrario.

## 2. Almacenamiento de metadatos como archivos XML

El modelo para repositorios distribuidos de objetos de aprendizaje [6] establece que los metadatos correspondientes a los objetos de aprendizaje deben de almacenarse dentro de una base de datos, de tal forma que se faciliten los procesos de búsqueda, recuperación y almacenamiento. Este modelo no especifica una base de datos en particular, en lo que sí es específico es en que se debe utilizar el lenguaje de consulta XQUERY para manipular los metadatos [5].

Durante el desarrollo de REDOBA se consideró la utilización de una base de datos relacional, a la que se accedería a través del lenguaje de consulta XQUERY; mediante la utilización de XQuark. Sin embargo, dados los problemas que se presentaron con la utilización de XQuark, se decidió implementar una versión del repositorio mediante el almacenamiento de la información en archivos XML con la finalidad de tener una versión funcional del prototipo del repositorio distribuido de objetos de aprendizaje. Esta decisión se tomó considerando que actualmente no se cuenta con una gran cantidad de objetos de aprendizaje, por lo que las operaciones con archivos no representan un problema en la eficiencia.

En esta versión del repositorio se decidió organizar la información de la siguiente manera:

- Un directorio de acceso general y cuya ubicación fuera accesible por todos.  
Dado que el prototipo ha sido desarrollado en el sistema operativo Linux, se decidió colocar el repositorio en el directorio `/var/redoba`. Al utilizarlo en un sistema operativo Windows este directorio ha cambiado a `\var\redoba`.
- Un directorio (MetaData) para almacenar los metadatos.

A cada metadato le corresponde un archivo en XML que contiene toda la información de dicho metadato.

- Un directorio (`Content`) para almacenar los objetos de aprendizaje. Cada uno de los objetos se almacena mediante un paquete ZIP, que contiene tanto el archivo con los metadatos como los elementos de que consta; estos pueden ser objetos mediáticos, informativos o de aprendizaje, así como redes de conocimiento o unidades de aprendizaje.

Esta versión del repositorio no utiliza el lenguaje de consulta XQUERY para especificar las operaciones sobre el repositorio, sólo se realizan búsquedas sobre los principales atributos de los metadatos; en particular se realizan búsquedas sobre título, descripción y palabras reservadas. Como ya se ha mencionado, la descripción de los metadatos se encuentra en archivos escritos en XML, por lo que se cuenta con una descripción de los metadatos en texto, esto facilita la labor de búsqueda, pues las operaciones se reducen a localizar patrones específicos dentro de un texto.

### 3. Almacenamiento de metadatos en una base de datos relacional

En principio, el proceso de diseño e implementación del almacenamiento de metadatos en una base de datos relacional se divide en dos etapas:

1. Diseño del modelo relacional de la base de datos, con base en el esquema de los metadatos y en requisitos y recomendaciones adicionales contenidos en LOM.
2. Especificación operacional de la correspondencia, o *mapeo*, entre los diferentes elementos y atributos contenidos en los metadatos y las correspondientes tablas y atributos en la base de datos relacional.

Un aspecto del problema que es importante mencionar aquí es que el esquema para la codificación de los metadatos en XML, tal como ha sido definido por IMS, es relativamente laxo. Esto es, muchas de las restricciones sobre los metadatos especificadas en LOM, tales como las longitudes de títulos y definiciones, la notación para representar diferentes lenguajes (naturales o artificiales) y el formato de fechas y tiempos, no existen en el esquema. Esto afecta, como se verá más adelante, tanto el diseño de detalle del modelo relacional como su mapeo con el esquema de los metadatos.

#### 3.1. Diseño del modelo relacional

La representación de los metadatos en XML está definida como parte de la especificación de IMS e implantada como un esquema XML, por lo que el diseño del modelo relacional para la base de datos debe permitir la representación completa y fidedigna de los metadatos así representados. Una manera fácil y rápida, incluso automática, de diseñar el modelo relacional consiste en definir una tabla para cada uno de los tipos complejos (*complexType*) definidos en el esquema XML, incluyendo llaves foráneas para representar la composición de tipos y, por consiguiente, la relación de inclusión entre los elementos de cada tipo en cada instancia del esquema. Sin embargo, esta manera de diseñar el modelo relacional tiene el inconveniente de no considerar ni información

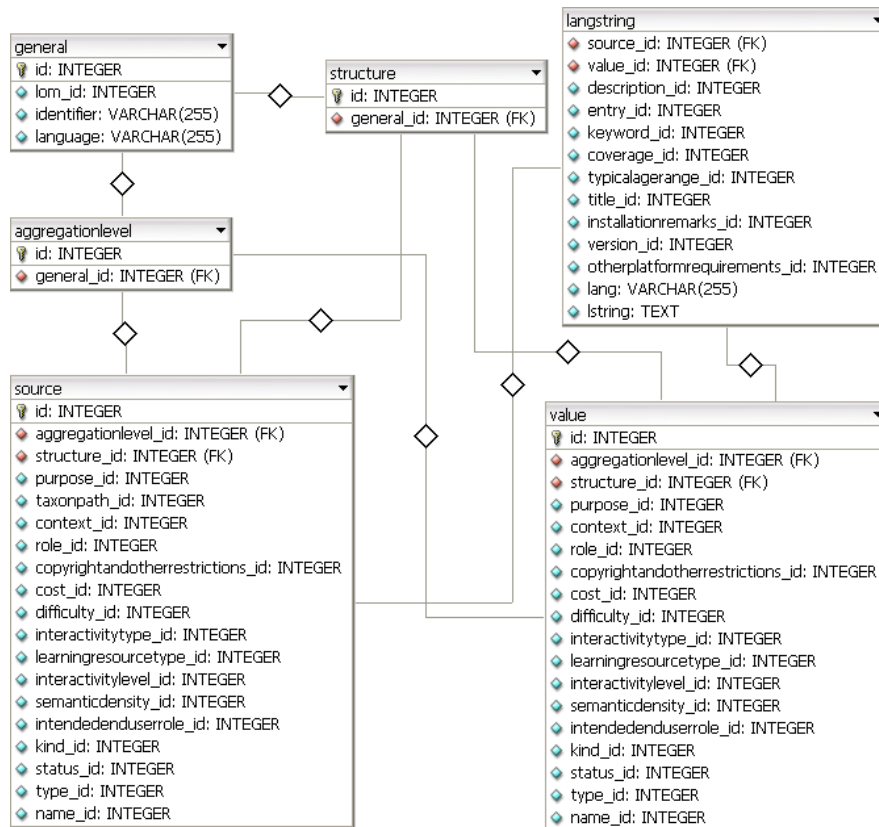
adicional, como la relativa a la semántica de los metadatos e incluso restricciones sintácticas expresadas en LOM y no en el esquema XML, ni aspectos de eficiencia de almacenamiento y acceso en la base de datos. Por ejemplo, la gran mayoría de los metadatos están etiquetados como `langstring` – un tipo complejo para representar texto en un lenguaje particular, natural o artificial – lo que significa que la mayoría de los datos y los accesos a la base de datos estarían concentrados en la tabla asociada al tipo `langstringType`, con un efecto negativo en el rendimiento de la base de datos. Otro problema de esta manera de diseñar el modelo relacional tiene que ver también con el tipo `langstringType`, que es aplicado para representar cadenas con longitudes que varían desde uno hasta dos mil caracteres (mínimo máximo recomendado), dependiendo de su significado y, por consiguiente, de su ubicación en la estructura de los metadatos. A manera de ejemplo, los subelementos *título*, *descripción* y *nivel de agregación* del elemento *general* en LOM se representan en XML usando `langstring`, como en

```
<general>
  <title>
    <langstring xml:lang="en">
      Draft Standard for Learning Object Metadata
    </langstring> ...
  </title> ...
  <description>
    <langstring xml:lang="en">
      Metadata is information about an object...
    </langstring>
  </description>
  <aggregationlevel>
    <source>
      <langstring xml:lang="x-none">LOMv1.0</langstring>
    </source>
    <value>
      <langstring xml:lang="x-none">2</langstring>
    </value>
  </aggregationlevel>
</general>
```

Sin embargo, mientras que la longitud en caracteres del valor del nivel de agregación no es más que uno (en el vocabulario definido en LOM), el título tiene generalmente una longitud de varias decenas de caracteres, en tanto que la descripción puede tener varios cientos o incluso miles de caracteres. Ubicar todos estos elementos en el mismo atributo de la misma tabla choca con su significado y puede ser ineficiente, o incluso inaceptable, en algunos manejadores de bases de datos relacionales.

Otra manera de diseñar el modelo relacional consiste en poner más atención a los elementos que conforman los metadatos que a los tipos de datos en el esquema. Esto es, realizar el diseño con base en la semántica y la organización de los metadatos, así como en las necesidades prácticas de almacenamiento y acceso a la base de datos. La diferencia entre los resultados obtenidos aplicando ambos métodos de diseño se puede apreciar comparando las figuras 1 y 2. Las tres tablas grandes – tanto en atributos como en número de registros – `source`, `value` y `langstring` del diseño por tipos de datos (figura 1) son combinadas y repartidas en tablas más pequeñas, como `aggregationlevel` y `struc-`

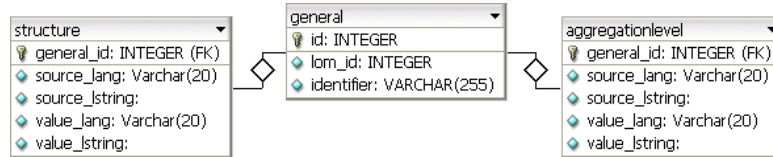
ture, en el nuevo diseño. Este nuevo diseño refleja mejor la semántica y organización de los metadatos, al mismo tiempo que hace un mejor balance de los datos en las tablas y del acceso a éstas en la base de datos. A cambio de estas ventajas, este diseño tiene que ser realizado manualmente y requiere más tiempo que el diseño por tipos de datos.



**Fig. 1.** Parte del modelo relacional diseñado con base en los tipos de datos contenidos en el esquema de los metadatos. Todos los atributos con terminación *\_id* son llaves foráneas, aunque solamente los utilizados por las relaciones visibles son marcados como tales.

### 3.2. Diseño del mapeo

Para implementar el almacenamiento de metadatos en una base de datos relacional en REDOBA se utilizan dos herramientas: la biblioteca BRIDGE de XQuark para la conversión entre archivos XML y bases de datos relacionales y el manejador de base de datos



**Fig. 2.** Parte del modelo relacional diseñado con base en la semántica y organización de los metadatos.

relacionales MySQL. La correspondencia o mapeo entre los metadatos y los atributos en las tablas de la base de datos de REDOBA se declara en BRIDGE utilizando un lenguaje XML, con `mapping` como raíz y los siguientes subelementos:

**element** Representa un elemento de los metadatos que se desea asociar con una tabla o atributo de una tabla en la base de datos.

**map** Introduce una tabla de la base de datos en correspondencia con un tipo de datos en el esquema XML (si aparece directamente bajo `mapping`) o con un elemento o atributo de los metadatos.

**attribute** Representa un atributo de un elemento de los metadatos que se desea asociar con una tabla o atributo de una tabla en la base de datos.

**generator** Es un elemento versátil, necesario para incluir información en la base de datos adicional a la contenida explícitamente en los elementos y atributos de los metadatos, tales como información que necesita ser calculada (mediante la ejecución de clases de Java cargadas bajo demanda) y relaciones entre tablas (mediante llaves primarias y foráneas).

Estos elementos pueden ser combinados de diversas maneras, según las necesidades impuestas por el esquema de los metadatos y por el modelo de la base de datos. Por ejemplo, un fragmento del mapeo del elemento `general` es

```
<?xml version="1.0" encoding="iso-8859-1"?>
<mapping
  xmlns="http://www.xquark.org/Bridge/1.0/Mapping" ...
  xmlns:ims="http://www.msglobal.org/xsd/imsmd_v1p2" ...>
  <map table="keyword" type="ims:keywordType" name="keywordMapping">
    <generator column="id" class="MyKeyGen"/>
    <element name="ims:langstring">
      <map table="key_plang">
        <generator column="keyword_id" ref="keyword.id"/>
        <generator column="lstring" variable="$NodeValue"/>
        <attribute name="xml:lang" column="lang"/>
      </map>
    </element>
  </map>
  <element name="ims:lom">
    <map table="lom"> ...
      <element name="ims:general">
        <map table="general">
```

```

<generator column="id" class="MyKeyGen"/>
<generator column="lom_id" ref="lom.id"/>
<generator column="identifier" class="ObjectIdentifier"/> ...
<element name="ims:keyword" map="keywordMapping">
  <map table="keyword_X_general">
    <generator column="keyword_id" ref="keyword.id"/>
    <generator column="general_id" ref="general.id"/>
  </map>
</element>
<element name="ims:structure">
  <map table="structure">
    <generator column="general_id" ref="general.id"/>
    <element name="ims:source">
      <element name="ims:langstring" column="source_lstring">
        <attribute name="xml:lang" column="source_lang"/>
      </element>
    </element>
    <element name="ims:value">
      <element name="ims:langstring" column="value_lstring">
        <attribute name="xml:lang" column="value_lang"/>
      </element>
    </element>
  </map>
</element> ...
</mapping>

```

Como se puede observar en este fragmento, en vez de mapear cada aparición del elemento `keyword` a la tabla correspondiente, se mapea el tipo `keywordType`. Este mapeo es invocado dentro del elemento `general` para mapear su subelemento `keyword`. En cambio, cada elemento `langstring` es mapeado explícitamente y de manera independiente, aunque todos estos mapeos siguen un mismo patrón, siendo la diferencia entre ellos los nombres de los atributos (y las tablas) a los que son mapeados. La diferencia entre los mapeos de `keyword` y `langstring` indica que el diseño final del model relacional se realizó tomando en cuenta *las dos* opciones de diseño descritas en la sección 3.1. Nótese también que se utiliza un objeto de la clase `MyKeyGen` para generar las llaves primarias de todas las tablas, mediante un `generator`, y otro `generator` se utiliza para crear llaves foráneas y así establecer las relaciones entre las tablas.

#### 4. Discusión general

El almacenamiento de los metadatos en archivos XML tiene la ventaja de que no requiere de un manejador de bases de datos, lo que permite su utilización en equipos con pocos recursos, tanto de procesamiento como de almacenamiento. Sin embargo para su utilización con grandes cantidades de objetos de aprendizaje (y sus metadatos) esta forma de funcionamiento no es la adecuada, pues fácilmente se pueden presentar problemas de inconsistencia en la información, o problemas para identificar de manera única los datos. Por otra parte, en esta versión existen limitaciones en cuanto a las ope-

raciones que se pueden realizar sobre los metadatos, sobre todo en cuanto a las búsquedas que se pueden realizar, están limitadas al título, la descripción o las palabras clave.

En cuanto al almacenamiento de los metadatos en una base de datos relacional, si bien el primer paso consistió en definir el modelo relacional, éste tuvo que ser revisado continuamente para acomodar a las restricciones impuestas explícitamente por las construcciones disponibles en BRIDGE para declarar el mapeo entre el esquema de los metadatos y el modelo relacional<sup>3</sup>. De manera general, se puede ver el diseño del almacenamiento en la base de datos como una instancia de tres esquemas – de los metadatos, de la base de datos y del mapeo – y el proceso de diseño como la búsqueda de un punto de equilibrio entre estos esquemas que produjera el resultado deseado<sup>4</sup>.

En la práctica, se ha utilizado una combinación de software comercial, gratuito y libre para implementar el almacenamiento de metadatos en REDOBA. Del lado del software libre cabe mencionar que el diseño de la base de datos fue realizado con DB DESIGNER 4 de fabFORCE, en tanto que el mapeo entre el esquema de los metadatos y la base de datos se ha llevado a cabo con BRIDGE de XQuark ejecutándose sobre la máquina virtual de JAVA distribuida gratuitamente por Sun Microsystems.

## 5. Conclusiones

En este artículo se han presentado aspectos del diseño del almacenamiento de metadatos tipo LOM en REDOBA, un prototipo de repositorio distribuido de objetos de aprendizaje. En particular, se discutieron aspectos relacionados con el almacenamiento como archivos XML, prácticamente tal como son recibidos por REDOBA, y con el almacenamiento en una base de datos relacional mediante un mapeo del esquema de los metadatos al esquema de la base de datos. Aunque este mapeo puede ser realizado automáticamente, se recomienda hacerlo manualmente, tomando en cuenta la organización de los metadatos, su semántica y las restricciones y recomendaciones descritas de manera textual en las especificaciones de LOM y su representación en XML, así como aspectos de eficiencia de almacenamiento y acceso en la base de datos.

## 6. Agradecimientos

Agradecemos a los desarrolladores y patrocinadores del gran número de herramientas de software libre y gratuito que se han utilizado para el desarrollo del trabajo que aquí se describe. En particular, nuestro agradecimiento va para el grupo XQuark por hacer disponible de manera libre BRIDGE.

El primer autor realizó parte de este trabajo como investigador en el Instituto de Investigaciones Eléctricas y con apoyo parcial de CONACyT y la Corporación Universitaria para el Desarrollo de Internet (CUDI), en el marco del proyecto CONACyT-CUDI correspondiente.

<sup>3</sup> También fue necesario acomodar a las restricciones impuestas por *bugs* descubiertos en la implementación actual de BRIDGE.

<sup>4</sup> La fase de implementación de este proceso se encuentra en un estado avanzado de desarrollo y se espera esté terminada para la fecha de realización del taller.

## Referencias

1. IMS Global Learning Consortium: IMS Learning Resource Meta-Data Information Model, Version 1.2.1 Final Specification (2001)
2. Institute of Electrical and Electronics Engineers: IEEE 1484.12.1 Draft Standard for Learning Object Metadata (2002)
3. Multimedia educational resource for learning and online teaching (2002). URL <http://www.merlot.org>
4. Morales R., López Mendoza S.: REDOBA: Un repositorio distribuido de objetos de aprendizaje (2004). Segundo Taller sobre Tecnología de Objetos de Aprendizaje
5. Riley K., McKell M.: IMS Digital Repositories Interoperability - Core Functions Best Practice Guide. Final specification, IMS Global Learning Consortium (2003). URL [http://www.imsglobal.org/digitalrepositories/driv1p0/imsdri\\_bestv1p0.html](http://www.imsglobal.org/digitalrepositories/driv1p0/imsdri_bestv1p0.html)
6. Riley K., McKell M.: IMS Digital Repositories Interoperability - Core Functions Information Model. Final Specification Version 1.0, IMS Global Learning Consortium, Inc. (2003). URL [http://www.imsglobal.org/digitalrepositories/driv1p0/imsdri\\_infov1p0.html](http://www.imsglobal.org/digitalrepositories/driv1p0/imsdri_infov1p0.html)
7. World Wide Web Consortium: Extensible Markup Language (XML) 1.0, 2nd edition (2000). URL <http://www.w3c.org/TR/2000/REC-xml-20001006>. W3C Recommendation





# **Ingeniería de software**

María de los Ángeles Arenas Valdés  
María Karen Cortés Verdín  
Juan Manuel Fernández Peña  
Jaime Muñoz Arteaga  
María de los Ángeles Sumano López  
(Editores)



## **Introducción**

El Taller de Ingeniería de Software tiene ya varias ediciones, siendo la primera de ellas en 1997 en la ciudad de Querétaro y la última el año pasado en la ciudad de Tlaxcala. Todas han tenido el denominador común de buscar el acercamiento entre la comunidad que se dedica a la Ingeniería de Software, ya sea en enseñanza, investigación o aplicación.

En esta ocasión se trató de averiguar cuáles son las prácticas de la Ingeniería de Software aplicadas en las empresas. Al parecer nuestras empresas están muy ocupadas y no hubo la respuesta esperada. Sin embargo, sí se tuvo una muy buena respuesta de la academia que se preocupa por infundir mejores prácticas en el desarrollo del software.

Un grupo de trabajos corresponde a experiencias en empresas, ya sea directamente o a través de proyectos que relacionan docencia con práctica profesional. Otros se orientan a propuestas de mejora en la propia docencia, para adecuarse a las necesidades de una industria de software en desarrollo.

En el transcurso del taller se presentaron las conclusiones de la edición pasada, se invitó a la Dra. H. Oktaba para la prenegación de Moprosoft, hubo la presentación de las ponencias y mesas redondas para el análisis de los resultados obtenidos y de nuevas estrategias a seguir.

## **Comité Científico**

Copresidentes:

María de los Ángeles Arenas Valdés, Universidad Veracruzana

María Karen Cortés Verdín, Universidad Veracruzana

Juan Manuel Fernández Peña, Universidad Veracruzana

Jaime Muñoz Arteaga, Universidad Autónoma de Aguascalientes

María de los Ángeles Sumano López, Universidad Veracruzana

Juan Manuel Fernández Peña

Guadalupe Ibarguengoitia González, Facultad de Ciencias, UNAM

Hanna Oktaba, Facultad de Ciencias, UNAM

Héctor Pérez González, Universidad Autónoma de san Luis Potosí

# La Ingeniería de Software como Proceso Permanente en la Enseñanza de la Computación

Ricardo Ruiz-Rodríguez<sup>1</sup>, Mario Alberto Moreno-Rocha<sup>2</sup>

Instituto de Electrónica y Computación  
Universidad Tecnológica de la Mixteca  
e-mail: {rruiz<sup>1</sup>, mmoreno<sup>2</sup>}@mixteco.utm.mx  
Km. 2.5 de la Carretera a Acatlima, C.P. 69000, Huajuapán de León, Oaxaca, México.

**Abstract:** Software engineering discipline has been taught traditionally like a single and essentially theoretical subject in the bachelor degree programs which cause a variety of problems in the students' professional education. This paper describes some problems that have been identified as well as the efforts the software engineering group at the Universidad Tecnológica de la Mixteca have been provided in order to solve them. Our group would like to share their experiences and get feedback from other related groups.

**Resumen:** Tradicionalmente, la ingeniería de software se ha enseñado en las universidades como una sola materia y con un carácter esencialmente teórico, lo cual ha sido origen de diversos problemas en la formación profesional de los estudiantes. Este artículo presenta algunas de las problemáticas identificadas en la práctica de la enseñanza de esta disciplina, así como los esfuerzos que el cuerpo académico de ingeniería de software de la Universidad Tecnológica de la Mixteca ha realizado para tratar de subsanarlos, al mismo tiempo, deseamos compartir experiencias y obtener retroalimentación de otros grupos de trabajo relacionados con la ingeniería de software.

**Palabras Clave:** Cuerpo Académico, Ingeniería de Software, Interacción Humano-Computadora, Usabilidad.

## 1. Introducción.

Sin lugar a dudas, una de las labores más complicadas en la academia es la de poder establecer un vínculo conveniente entre los intereses de los profesores, sus líneas de investigación o aplicación del conocimiento y la docencia. Afortunadamente la ingeniería de software tiene una maleabilidad tal que, si se utiliza de manera apropiada, puede ajustarse fácilmente a estas necesidades

En lo particular, el Cuerpo Académico de Ingeniería de Software (CASI) de la Universidad Tecnológica de la Mixteca (UTM) tiene tres líneas de investigación y/o aplicación del conocimiento las cuales son:

1. Metodologías Orientadas a Objetos
2. Interacción Humano Computadora

### 3. Desarrollo de Software Educativo

Desde sus inicios, la intención del CASI fue la de agrupar académicos con intereses afines y relacionados con la ingeniería de software para poder realizar investigación, proyectos, y poder participar conjuntamente en la formación de recursos humanos en el área, así como también el de jugar un papel importante en la promoción al desarrollo de la región de Huajuapán de León en Oaxaca.

Este artículo presenta un panorama general del papel de la ingeniería de software en la mayoría de las instituciones de educación superior, así como la propuesta del CASI y el papel que ha tenido dentro del contexto de la formación de recursos humanos a nivel licenciatura y maestría en la UTM.

## 2. Problemática.

El proceso usual de enseñanza para la generación de software en las universidades, se basa en un esquema tradicional y hasta cierto punto ortodoxo, en el que el proceso de desarrollo de software se presenta reducido a una sola etapa: la de programación. Si bien es cierto que los procesos de desarrollo formal en cuanto al ciclo de vida completo implican una etapa de programación, no debe perderse de vista que es solamente eso, una etapa y no un fin en sí mismo.

En uno de los libros clásicos de ingeniería de software [1] se comentan algunos de los aspectos involucrados en la generación de diferentes tipos y dimensiones de éste. Esto es importante mencionarlo ya que los desarrollos de software que se realizan en algunas de las materias básicas de la disciplina de la computación como lo son: introducción a la programación, estructuras de datos, paradigmas de programación, programación de sistemas, etc. carecen en su mayoría de los fundamentos importantes de la ingeniería de software respecto al análisis, diseño, pruebas y documentación por mencionar sólo algunas. En general, la didáctica de estas materias tiende a ser más bien enfocada hacia la labor de la programación en sí, desviando a los estudiantes y hacerlos pensar con ello, que los procesos de desarrollo de software funcionan de esa manera, y que a su vez, los patrones de solución aplicados a sus pequeños proyectos, son los mismos que funcionarán para todos los sistemas independientemente de su tamaño y tipo, asumiendo por consiguiente, que la ingeniería de software es en sí misma innecesaria o una materia con un interés puramente teórico.

Nuestra experiencia en este sentido muestra, que los estudiantes presentan una marcada renuencia hacia los conceptos y las metodologías propuestas de la ingeniería de software, debido a la inercia que tienen respecto a la forma en que han desarrollado software y que, en su experiencia, les ha dado resultado. Si a este hecho le agregamos el aspecto de que la ingeniería de software se imparte en general como una materia en semestres avanzados o en semestres intermedios, el problema se agrava.

La Fig. 1. ilustra el proceso de desarrollo de software presentado en [1]. La intención de presentarlo aquí es que, en nuestra experiencia, es el mismo patrón o al menos parecido al que se sigue en la enseñanza del proceso de desarrollo de software. En el esquema más simple de la amplia gama de posibilidades de la generación de software se tiene un programa; evoluciones de éste hacia productos más elaborados, robustos y sofisticados se tiene en productos de la programación y en la programación de sistemas, en donde se involucran ya otros aspectos más allá de la simple etapa de la programación, tales como la integración de interfaces, pruebas, documentación, etc. Sin embargo, un producto de la programación de sistemas entendido en su esquema más amplio, involucra aspectos y procesos propios de la ingeniería de software, por lo que es importante hacer conciencia de que desde los cursos básicos, los profesores tenemos la responsabilidad de integrar lo más formalmente posible, técnicas y metodologías de desarrollo de software, y no inducir a los estudiantes desde las primeras etapas de su formación, a actuar y pensar simplemente como programadores.



Fig. 1 Evolución del esquema de programación (adaptado de [1]).

### 3. Propuesta y plan de trabajo del CASI.

En la UTM en particular, el CASI ha pugnado y llevado a cabo la integración de técnicas y metodologías de desarrollo formales desde los primeros cursos de la ingeniería en computación, así como en todos aquellos cursos que impliquen de una u otra manera el desarrollo de software. Consideramos que no es conveniente esperar a que los alumnos lleguen a semestres intermedios en los que, de manera tardía, la disciplina de la ingeniería de software se imparte como una simple materia, ya que basados en la experiencia, podemos decir que es en esta etapa, en la que es más común encontrar a alumnos con vicios y malas prácticas en sus procesos de desarrollo, por lo



que tratar de modificar su modelo mental respecto a lo que ellos consideran que es el desarrollo de software, se presenta como una tarea muy difícil de enmendar.

Aunado a lo anterior, el CASI se vale de procesos y metodologías no muy comunes en los modelos de enseñanza tradicional pero ampliamente utilizadas en el campo de la ingeniería de software, como lo son las proporcionadas por la ingeniería de usabilidad y las heurísticas de Interacción Humano-Computadora (IHC). Estos procesos y metodologías han sido integrados satisfactoriamente como parte del proceso de enseñanza-aprendizaje en los últimos semestres de la carrera de ingeniería en computación, y se ha propuesto que se integren también como parte de los programas educativos en la formación profesional a nivel licenciatura en el área de la computación [2]. Así mismo, los procesos y técnicas de usabilidad se han incorporado también de manera satisfactoria en tres de las materias de la Maestría de Ingeniería de Software (MIS) que se imparte en la UTM: Desarrollo de Interfaces de Usuario, Desarrollo de Software Orientado a Web y Comercio Electrónico (véase la Fig. 2).

Semestre	Materia	Semestre	Materia
1	Complejidad de Algoritmos	4	Verificación y Calidad
	Metodologías de Desarrollo		Seminario de Tesis
2	Ingeniería de requerimientos	Optativas	Interacción Humano-Computadora
	Gestión de Proyectos de Software		Desarrollo de Software Orientado a Web
3	Desarrollo de Interfaces de Usuario	Comercio Electrónico	Software Empotrado
	Metodologías Orientadas a Objetos		Sistemas de Tiempo Real
3	Análisis y Diseño de Software I	Sistemas Distribuidos	Software para Sistemas de Seguridad
	Bases de Datos		
3	Métricas de Calidad para Software		
	Análisis y Diseño de Software II		
	Optativa I		
	Optativa II		

Fig. 2 Plan de estudios de la Maestría en Ingeniería de Software.

Es importante mencionar también que la experiencia demostrada en los cursos de desarrollo de interfaces, en los cuales se integran conocimientos de la IHC, han mostrado que, con un esfuerzo adicional de los alumnos en implementar nociones y guías de desarrollo ajenas a los conocimientos básicos de programación, el producto resultante es significativamente mejor. En nuestro Laboratorio de Usabilidad (UsaLab) los estudiantes aplican sus conocimientos adquiridos en clase, y obtienen una mejor retroalimentación de los usuarios a través de la realización de pruebas de usabilidad. Los resultados siempre han sido satisfactorios.

Por otro lado, la participación en proyectos reales y de mediana complejidad es un factor que consideramos importante en la aplicación de los conceptos de la ingeniería de software, por lo que el CASI se ha dado a la tarea en ello y participa activamente en proyectos internos de la UTM relacionados con la Universidad Virtual y la Biblioteca,

los cuales están intrínsecamente relacionados a las líneas de investigación del cuerpo académico o a las materias que los miembros del CASI imparten en la MIS [3, 4, 5]. Además, dentro de las materias impartidas a nivel licenciatura y maestría, se busca también motivar a los alumnos para que los trabajos desarrollados en sus materias deriven en artículos [6] y en algunos otros casos en tesis [7].

Es importante mencionar también que el CASI, asumiendo una de las misiones fundamentales de la UTM, participa activamente en actividades de promoción al desarrollo y ha comenzado a generar proyectos en los que participan alumnos y profesores, y que derivan en beneficios substanciales a la comunidad huajuapeña, tal es el caso del sistema de administración de cuentas del Sistema de Agua Potable y Alcantarillado de Huajuapán de León en Oaxaca (SAPAHUA) [8].

Este esfuerzo del CASI no es nuevo, ya que en otras ocasiones se han hecho propuestas y trabajos [9] que pretenden hacer alguna aportación en este sentido al compartir la experiencia que como cuerpo académico vivimos, y que como grupo de interés en ingeniería de software deseamos compartir; al mismo tiempo que se busca una retroalimentación útil de otros cuerpos académicos o grupos de interés afines con una mayor o mejor experiencia que la nuestra.

#### **4. Conclusiones.**

La ingeniería de software se vislumbra como una de las disciplinas más importantes en los próximos años en nuestro país, tanto en la industria como en la formación de recursos humanos capaces de aplicar esta importante área del conocimiento. En este sentido, todos los esfuerzos que estén siendo dirigidos hacia el fortalecimiento de esta disciplina, resultan indispensables para el intercambio de experiencias entre grupos interesados en el área de la ingeniería de software, por lo que resulta imprescindible el poder participar activamente en talleres y congresos en donde se discutan los elementos relacionados con diferentes tipos de experiencias en el desarrollo de software.

En este sentido, el CASI hace su mejor esfuerzo para introducir en las diferentes disciplinas de la formación profesional en computación aquellos aspectos de suma importancia en la ingeniería de software, y no esperar a que los alumnos tomen dicho curso para descubrir que detrás del proceso de desarrollo de software, existen metodologías y procesos que conllevan a un modelo de desarrollo más formal, estable y de mejor calidad.

Finalmente, se considera importante insistir en trabajar los conceptos y modelos de la ingeniería de software y, en la medida de lo posible, complementarlos con otros elementos como los de la ingeniería de usabilidad y la Interacción Humano-Computadora por ejemplo, y aplicarlos a su vez en la elaboración de proyectos derivados de problemáticas reales existentes en las propias instituciones, así como en proyectos externos que generen, eventualmente, recursos que financien las diferentes necesidades de los grupos de investigación.

## Referencias.

- [1] Brooks, Frederick P., Jr., "The Mythical Man-Month: Essays on Software Engineering", Addison Wesley 1995.
- [2] Moreno Rocha M A. "Adding Human Computer Interaction Studies into Informatics and Computing Engineering Bachelor Degrees in Latin America". Proceedings of the Conference on Human Factors in Computing Systems, Seattle, USA, 2001
- [3] Fernández y Fernández, C A; Moreno Rocha, M A. "Evolución del Entorno Tecnológico para la Enseñanza a Distancia en la Universidad Virtual de la UTM". IV Jornadas Sobre Informática y Sociedad, Barcelona, España.
- [4] Moreno Rocha M A y Hernández Rueda E. "Incorporando Dispositivos PDA en Educación a Distancia". Proceedings de Virtual Educa 2002, Valencia, España.
- [5] Fernández y Fernández, Carlos Alberto. "Aplicación del Modelo de Comercio Electrónico para la sistematización de la Biblioteca de la UTM". ANIEI. XV Congreso Nacional y I Congreso Internacional de Informática y Computación, Guadalajara, México.
- [6] Fernández y Fernández, Carlos Alberto; Mendoza Vásquez, Alejandra; Martínez Guzmán, David; Mendoza Ortiz, Eloy; Sumano Ortega, Pavel. "Ingeniería de Requerimientos aplicada a la Universidad Virtual de la UTM". ANIEI. XVI Congreso Nacional y II Congreso Internacional de Informática y Computación, Zacatecas, México.
- [7] Camarillo Barragán, J, Moreno Rocha, MA. "Aplicación de Estudios de Usabilidad a los Cajeros CFEmáticos", Proceedings of the Jornadas Informática y Sociedad 2002, Barcelona, España, 2002.
- [8] Solís Genesta, I. G. y Moreno Rocha, M. A. "Desarrollos de Software dentro del CASI: Proyecto SAPAHUA". Reporte presentado dentro de los Seminarios de Investigación de la Universidad Tecnológica de la Mixteca, Huajuapán de León Oaxaca, México, 2004.
- [9] Moreno Rocha, M. A. Ruiz Rodríguez Ricardo. Muñoz Arteaga, J. "Towards a Better Practice in Software Engineering Education and Academic Research: the CASI Considerations Proposal". IV Congreso Internacional de Ciencias de la Computación: Avances en Ciencias de la Computación, ENC 03, Tlaxcala, México, 2003.

# **Propuesta de una línea curricular de Ingeniería de Software en la Licenciatura en Sistemas Computacionales.**

Francisco Alvarez R., Juan Cardona S. y Alfredo Weitzenfeld R.

Av. Universidad No. 940, Universidad Autónoma de Aguascalientes,  
fjalvar@correo.uaa.mx, pedrok26@hotmail.com  
Rio Hondo No. , Instituto Tecnológico Autónomo de México,  
Alfredo@itam.mx

## **Resumen.**

El presente trabajo muestra una propuesta de prácticas de software insertadas en las materias del currículo de la Ingeniería en Sistemas Computacionales, para ayudar en la conformación de una línea curricular en ingeniería de software con el objetivo de ayudar al desarrollo de una industria nacional.

## **Palabras clave.**

Ingeniería de software, Industria de software, Línea curricular, Estrategias didácticas.

## **Introducción.**

Uno de los principales problemas a los que nos enfrentamos al momento de diseñar o revisar modelos curriculares de licenciaturas es la pertinencia con el mercado laboral, las tendencias en las áreas de conocimiento concernientes, y los modelos curriculares que emanan de las propias instituciones educativas.

El modelo curricular de la ANIEI (Asociación Nacional de Instituciones y Escuelas de Informática) define un profesional de la computación en cuatro perfiles, que van desde el capaz de planear, organizar, integrar y controlar servicios y sistemas de información de diversas organizaciones, hasta el dedicado al estudio y desarrollo de las ciencias de la computación, con capacidad de generar elementos para la concepción y creación de nuevos sistemas en entornos computacionales [1].

Dentro de estos perfiles profesionales se incluye una línea curricular enfocada a la “Programación y la Ingeniería de Software” que pretende formar al profesional de la computación en los diferentes paradigmas de programación y lenguajes, algorítmica e industria de software [2], sin embargo su alcance, instrumentación y operación en un currículum real no resulta ser tan claro, además de que no se apega completamente al Plan Nacional de la Industria de Software.

El “programa para el desarrollo de la industria de software” indica como visión lograr una industria competitiva internacionalmente, y asegurar su crecimiento en el largo plazo. Se espera situar a México como líder de esta industria en Latinoamérica para el año 2010 y convertirlo en líder desarrollador de soluciones de TI (Tecnologías de Información) de alta calidad y uso de software en el ámbito mencionado [8]. Una de las estrategias planteadas en este plan es la formación de recursos humanos competitivos, que marca líneas de acción concretas, tales como: a. la adecuación y mejoramiento dinámico y pertinente de los planes y programas de estudio, b. el impulso a sistemas, métodos, procesos e instrumentos que mejoren la enseñanza del desarrollo de software, enfatizando en la innovación, c. el fortalecimiento de mecanismos, instancias e instrumentos que eleven la efectividad y la eficiencia de la vinculación entre aprendizaje y desarrollo de software, etc. sin embargo se presenta nuevamente el problema de instrumentación de estas acciones en los programas educativos que están operando actualmente en todo el país.

Este trabajo pretende mostrar una instrumentación de las prácticas actuales recomendadas para desarrollar una industria de software en un programa educativo: Ingeniería en Sistemas Computacionales (I.S.C.) de la Universidad Autónoma de Aguascalientes, el cuál tiene una duración de 10 semestres y está clasificado como perfil “C” (profesional dedicado al estudio y desarrollo de las ciencias de la computación, que genera elementos para la concepción y creación de ambientes, facilidades y aplicaciones innovadoras de la computación dentro de entornos diversos de demandas por satisfacer) según los lineamientos de la ANIEI [1].

### **Propuesta de línea curricular para el desarrollo de la industria de software.**

Se propone una serie de prácticas para la industria de software que el alumno deberá cubrir a través de las distintas materias de programación y áreas afines, con el objetivo de formar las competencias necesarias en la profesionalización en ingeniería de software.

Algunas de las prácticas recomendadas por académicos de las áreas citadas [6][3] mencionan prácticas iniciales (Elementos de análisis de problemas, Diseño de solución, Revisión de diseño y código, Pruebas unitarias y de integración, Medición de tiempos, tamaños y defectos, Estándares de codificación, etc.), intermedias (Análisis y diseño orientado a objetos, Patrones de diseño de software, Diseño de interfaces de usuarios, PSP, Pruebas avanzadas, Administración de configuración de software) y avanzadas (Desarrollo en equipo con procesos definidos, Modelos de procesos de software y empresas de software), para el nivel de licenciatura.

De lo anterior las prácticas propuestas se clasifican en tres niveles los cuáles agrupan desde los aspectos básicos para la Ingeniería de Software hasta modelos de madurez de procesos de conocimiento difundido, cómo es CMM [7].

#### Prácticas Básicas.

- B1. *Análisis de problemas.* La práctica es abordada a través de la obtención de requisitos generales de proyecto: delimitación de objetivos y del problema, determinación de causas del problema, etc.
- B2. *Diseño de solución a problemas y de Interfaces de usuario.* Esta práctica es cubierta por técnicas para el diseño de interfaces de usuario, algoritmos y arquitectura básica del sistema.
- B3. *Revisión de diseño y código (personal y entre colegas).* Se aplican técnicas y formatos que permitan al alumno verificar su diseño y código, así como el de sus compañeros.
- B4. *Técnicas de pruebas unitarias y de integración de sistemas.*
- B5. *Métricas básicas (incluye Function Points).* Se comienzan a aplicar métricas para la estimación de tamaño y esfuerzo, a través de predicción sobre código desarrollado, productividad, errores y puntos de función.
- B6. *Estándares de codificación.* Conforme el estándar utilizado en el área de desarrollo de sistemas de la UAA, con el fin de que los alumnos aprendan a codificar y documentar de manera uniforme.

#### Prácticas Intermedias.

- M1. *Personal Software Process (PSP) [4].* Se enseña las técnica completa del proceso de software personal y su aplicación en un proyecto durante el curso.
- M2. *Team Software Process (TSP) [5].* Se enseña las técnica completa del proceso de software en equipo y su aplicación en un proyecto durante el curso.
- M3. *Métodos de análisis y diseño orientado a objetos basado en UML (Unified Modeling Language).*
- M4. *Patrones de diseño de software.*
- M5. *Tecnología JAVA avanzada.*
- M6. *Técnicas de pruebas avanzadas.*
- M7. *Administración de Configuración de Software.* Se pide a alumno dar el seguimiento de la administración de la configuración del software a través de formatos estandarizados durante el desarrollo del proyecto final del curso.

#### Prácticas Integradoras.

- I1. Desarrollo en equipo siguiendo un proceso de desarrollo definido para un cliente real.
- I2. Modelos de procesos de software.
- I3. Calidad y madurez de procesos de software.

La forma de instrumentación en el programa es mediante los cursos definidos en el programa educativo (ISC), es decir las prácticas y técnicas indicadas se integran cómo actividades regulares y obligatorias para los cursos: Lógica de programación, Programación I, II y III, Estructura de Datos, Lenguaje de Bases de Datos, Análisis y Diseño de Sistemas, Metodologías de Desarrollo e Ingeniería de Software.

Práctica	Materias por semestre.								
	1	2	3	4	5	6	7	8	9
B1.	Lógica de programación.								
B2.	Lógica de programación.								
B3.		Programación I.							
B4.			Programación II.						
B5.			Programación II.						
B6.		Programación I.							
M1. Inicial				Programación III.					
M1. Completo								Metodologías de desarrollo.	
M2. Inicial.							Lenguaje de B.D.		
M2. Completo								Metodologías de desarrollo.	
M3.					Análisis y Diseño	Estructura de Datos.			
M4.				Programación III.					
M5.				Programación III.					
M6.							Lenguaje de BD.		
M7.							Lenguaje de BD.		
I1.								Metodologías de Desarrollo	
I2.									Ing. de Software
I3.									Ing. de Software.
B1, B2, B3, B4, B5, B6, M4 y M5.						Estructura de Datos.			

**Tabla 1.** Instrumentación de prácticas de software en la línea curricular.

Lo cursos mencionados forman parte de una línea curricular en programación y desarrollo de software, con objetivos establecidos en el plan de estudios de ISC que indican las directrices sobre los contenidos, por ejemplo, el curso de “Análisis y Diseño de Sistemas”, indica que contiene todos los elementos teórico-práctico básicos para que puedan analizarse y solucionarse problemas de carácter organizacional utilizando metodologías actuales [9]. Se pretende entonces de forma adicional a los contenidos formales descritos incorporar las prácticas listadas anteriormente.

En el mismo ejemplo mencionado, se incorporarían métodos de análisis y diseño orientado a objetos basado en UML (Unified Modeling Language), además de las prácticas aprendidas anteriores (Análisis de problemas, Diseño de solución de problemas y de Interfaces de usuario, Revisión de diseño y código, Técnicas de pruebas unitarios y de integración de sistemas, Métricas básicas y Estándares de codificación), es decir las prácticas definidas para un primer curso deberán seguir constantes junto con las prácticas del segundo curso y así sucesivamente, como se muestra en la *Tabla 1*.

Adicionalmente se plantearon en esta Ingeniería prácticas profesionales que se desarrollan a partir del noveno semestre, se pretende integrar todas las prácticas indicadas en la *Tabla 1*, en un proyecto real con empresas desarrolladoras de software.

## **Resultados y Discusión.**

Hasta el momento se han desarrollado varias acciones con el fin de instrumentar la línea curricular propuesta, las cuáles van desde la capacitación para los profesores que dan cursos de las diferentes programaciones mencionadas con las prácticas básicas, hasta las modificaciones de programas de cursos avanzados en temáticas pertinentes en la industria de software (PSP, TSP y CMM).

Por otro lado se han comenzado a desarrollar proyectos de software con casos reales (todos dentro del Estado de Aguascalientes) que incluso son utilizados en las organizaciones y con los profesionistas para los que fueron creados; algunos de estos sistemas son: control médico, control escolar para primarias y secundarias, administración de proyectos de investigación, administración de ventas multinivel, etc.

La experiencia en el desarrollo de estos proyectos y sus correspondientes evaluaciones con los usuarios y clientes, ha permitido que el alumno experimente un compromiso formal, así como una planeación y seguimiento del proyecto con un nivel de rigor al menos básico.

Un aspecto que se ha observado es el entendimiento de los alumnos hacia la integración de un proceso de desarrollo de software y la importancia de éste en el logro de los objetivos de los proyectos planteados en los cursos.

## **Conclusiones.**

Uno de los principales problemas que se han comenzado a resolver es la integración de las prácticas de ingeniería de software desde el inicio de la carrera, a través de los cursos de la línea curricular de programación e ingeniería de software: Lógica de programación, Programación I, II y III, Estructura de Datos, Lenguaje de Bases de



Datos, Análisis y Diseño de Sistemas, Metodologías de Desarrollo e Ingeniería de Software, sin embargo es necesario reafirmar la actualización de las prácticas con el fin de adoptar nuevas propuestas y desarrollo tecnológico que beneficien a la industria de software.

## Referencias.

- [1] ANIEI, Perfiles Profesionales: Lic. en Informática, Lic. en Sistemas Computacionales, Lic. en Ciencias de la Computación y Ing. en Computación, Asociación Nacional de Institucionales de Educación en Informática, 2002.
- [2] CENEVAL, Guía de Examen para el egreso de la Licenciatura en Informática-Computación (EGEL Info-Comp), Centro Nacional de Evaluación para la Educación Superior A.C., Tercera Edición, 2003.
- [3] Fernández, L. y Montes de Oca, C. “Marco conceptual para la formación de recursos humanos en Ingeniería de Software”, Avances en Ciencias de la Computación, IV Congreso Internacional de Ciencias de la Computación, ENC’03. México, 2003.
- [4] Humphrey, W. *Introduction to the Personal Software Process*. Pearson Education. USA. 1997.
- [5] Humphrey, W. *The Team Software Process (TSP)*, Software Engineering Institute, 2000.
- [6] Oktaba, H. “Ingeniería de Software: trayecto personal desde la programación a la producción de software”, Avances en Ciencias de la Computación, IV Congreso Internacional de Ciencias de la Computación, ENC’03. México, 2003.
- [7] Paulk, M. et al., *The Capability Maturity Model: Guidelines for Improving the Software Process*, Software Engineering Institute, Carnegie Mellon University, Addison-Wesley. 1994.
- [8] Secretaría de Economía, *Programa para el Desarrollo de la Industria de Software*, 2002.
- [9] Universidad Autónoma de Aguascalientes, *Revisión Curricular de la carrera de Ingeniería en Sistemas Computacionales Plan 2001*, Dirección General de Docencia de Pregrado, 2001.

# Hacia una Especificación del Marco Conceptual para la Formación de Recursos Humanos en Ingeniería de Software

Gerardo Padilla, Carlos Montes de Oca  
{*gpadilla, moca*}@*cimat.mx*

Luis Felipe Fernández M  
*lfernand@uacj.mx*

Centro de Investigación en Matemáticas A.C.  
CIMAT  
Calle Jalisco S/N. Col. Mineral de Valenciana,  
Guanajuato, Gto., 36240

Universidad Autónoma de Ciudad Juárez,  
UACJ  
Instituto de Ingeniería y Tecnología  
Av. del Charro 610 nte. Cd. Juárez, Chih.

**Resumen:** El Marco Conceptual para la Formación de Recursos Humanos en Ingeniería de Software (MCFIS) es una herramienta usada como modelo de referencia y guía en la creación y evaluación de programas de formación en ingeniería de software. Uno de las metas del MCFIS es la definición un modelo de referencia estándar. Sin embargo, la descripción actual del MCFIS no permite realizar un análisis sistemático de la misma, no incluye criterios de interpretación del modelo ni permite una difusión y evolución similar a la que tienen los estándares de la industria. Este trabajo presenta una estrategia para llevar la descripción del MCFIS al nivel de una especificación del tipo industrial y resolver los problemas antes mencionados. Dicha especificación se basa en la estructura de un perfil del Lenguaje Unificado de Modelado (UML) que se extiende para manejar múltiples vistas del MCFIS. Denominamos a esta descripción como la Especificación MCFIS Versión 1.0 (E-MCFIS 1.0).

**Abstract:** The Conceptual Framework for Education of Human Resources in Software Engineering (MCFIS) is used as a reference model to create and evaluate software engineering academic programs. One of the MCFIS goals is the definition of a standard reference model. Nevertheless, the current MCFIS representation does not allow the analysis of the model itself, it does not include interpretation criteria for the model and it does not provide a way to share and evolve the model, as the industrial standards do. This paper describes an strategy to evolve the MCFIS representation to the level of an industrial specification. The proposed strategy uses a structure based on the Unified Modeling Language profiles (UML). This structure is extended to handle multiple views of MCFIS. We call this specification *Specification MCFIS Version 1.0* (E-MCFIS 1.0).

**Palabras Clave:** Educación, Ingeniería de Software, MCFIS, Especificación

## 1. Introducción.

El *Marco Conceptual para la Formación de Recursos Humanos en Ingeniería de Software* (MCFIS) [1] forma parte de los esfuerzos de instituciones académicas y de investigación para apoyar las estrategias descritas en el Programa para el Desarrollo de la Industria de Software [2]. Este programa tiene una serie de estrategias; una de ellas,

está íntimamente ligada a las instituciones de educación e investigación: “Educación y formación de personal competente en el desarrollo de software, en cantidad y calidad convenientes”. Partiendo de lo anterior, se generó el MCFIS [1].

El MCFIS es un modelo de referencia para la creación y análisis de programas para la formación de recursos humanos en el área de Ingeniería de Software. El MCFIS se representó mediante diagramas de clases usando el Lenguaje Unificado de Modelado (UML por sus siglas en inglés) [3]. La descripción del MCFIS se estructura con conceptos que tienen un papel preponderante en la formación de recursos humanos. El marco conceptual tiene la virtud de no imponer ideas preconcebidas de cada concepto y deja abierta la posibilidad de que a cada uno de ellos se le puedan asociar características propias. Esto significa que cada institución puede matizar el tipo de formación que ofrece de acuerdo a sus intereses y pertinencia. Por ejemplo, el MCFIS ha sido exitosamente usado para la planeación y análisis de un programa de postgrado en Ingeniería de Software en el Centro de Investigaciones de Matemáticas.

La descripción del MCFIS propuesta en [1] establece cierto tipo de estructura para el marco conceptual que describe. Sin embargo, el MCFIS carece de una estructura formal que permita: i) realizar un análisis sistemático del modelo mismo, ii) incluir criterios de interpretación del modelo y iii) permitir una difusión y evolución similar a la que tienen los estándares de la industria. Este artículo presenta una estrategia para resolver dichos problemas. La estrategia se basa en la creación de una especificación para el MCFIS, denominada E-MCFIS 1.0. La estructura, contenido e información de dicha especificación permitirá analizar el modelo de referencia, describir criterios de interpretación, establecer mecanismos para la evolución y mantenimiento del modelo, y facilitar la difusión y el proceso de su estandarización. La organización de esta especificación se basa en la estructura de los perfiles de UML e incorpora el uso de vistas de interés. Esta vistas usan un enfoque similar al utilizado en los métodos de evaluación de arquitecturas de software [4].

Este artículo se estructura de la siguiente manera: en la sección 2 se describen brevemente los perfiles de UML. En la sección 3 se presenta la estructura y características más importantes de la especificación E-MCFIS 1.0. Finalmente, se presentan las conclusiones y trabajo futuro relacionado con la versiones futuras del E-MCFIS.

## **2. Perfiles de UML.**

UML permite la construcción de modelos usando distintos tipos de diagramas que representan distintos aspectos del sistema (e.g., estructura, comportamiento, despliegue, ejecución). Una de las características más importantes de UML es su capacidad de extenderse para dominios particulares, esto es, UML permite definir un subconjunto personalizado de elementos de modelado con un significado particular para dominios específicos. A dichas extensiones se les denomina perfiles de UML. Se pueden encontrar ejemplos de perfiles en dominios tales como la prueba de software, sistemas en tiempo real, y CORBA entre otros. Un perfil define básicamente lo siguiente:

- Un *núcleo básico de conceptos* usado para el dominio en particular. Estos conceptos definen el lenguaje para ese dominio.
- Una *jerarquía de conceptos* basada en herencias, agrupaciones y niveles de abstracción. Dicha jerarquía organiza los conceptos de manera que permitan la reutilización, la comprensión y la creación de modelos.
- Un *conjunto de reglas de construcción y restricciones* relacionadas con el dominio específico. Dichas reglas permiten la creación de modelos bien formados para el dominio. Las *restricciones de los elementos de modelado* se especifican generalmente usando el lenguaje de restricciones de objetos (OCL por sus siglas en inglés) [3].
- En algunos casos, se incluye una *semántica para el análisis de los modelos*, que permite el análisis y manipulación de los modelos bajo un esquema formal.

### 3. E-MCFIS 1.0.

El E-MCFIS 1.0 representa la descripción del MCFIS con un formato de especificación. La organización de la especificación se basa en la estructura modificada de un perfil de UML. Dicha modificación agrupa ciertos elementos e incorpora el uso de *vistas* de interés, que han sido utilizadas en los métodos de evaluación de arquitecturas de software [4]. Las vistas permiten aislar los elementos relacionados con un interés en particular. Por ejemplo, la vista de Planta Académica, tiene como centro del análisis los aspectos de interés relacionados con la planta académica de la institución.

#### 3.1 Estructura básica del MCFIS

El MCFIS es un marco conceptual conteniendo 21 conceptos y 40 relaciones entre dichos conceptos. El MCFIS está representado por 5 diagramas de clases de UML. Cada diagrama se orienta bajo un concepto central (i.e., Ingeniería de Software, Planta Académica, Cuerpo de Conocimientos, Certificaciones y Egresados). Dicha organización de diagramas puede ser interpretada como un conjunto de vistas del modelo total.

El tipo de relaciones que define el MCFIS corresponde a cláusulas que definen interacciones entre los conceptos definidos. Por ejemplo, la Figura 1 presenta un fragmento de un diagrama del MCFIS donde una de las cláusulas presentadas en este diagrama establece que la *Ing. de Software es soportada en Modelos de Enseñanza*. Dicha cláusula asocia dos entidades conceptuales, la de Ing. de Software y la de Modelos de Enseñanza.

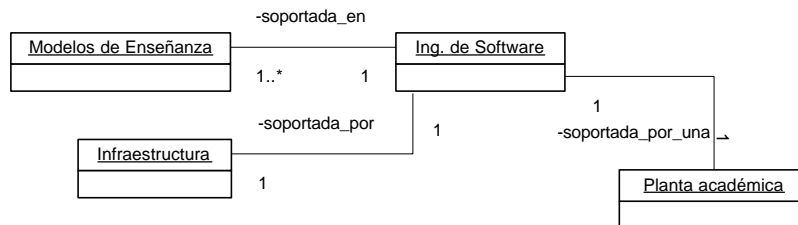


Figura 1. Modelo parcial del MCFIS.

### 3.2 Estructura conceptual del E-MCFIS

Desde el punto de vista conceptual, el MCFIS define un modelo compuesto de conceptos y relaciones. Las vistas vienen a representar secciones parciales de dicho modelo. Sin embargo, el MCFIS no prescribe ninguna estructura para organizar y describir estos elementos.

Esta situación limita y complica actividades de mantenimiento y evolución del modelo mismo. Es por esto, que se propone una estructura conceptual basada en el uso de un metamodelo del MCFIS (como se maneja en la definición del lenguaje UML). Partiendo de esto, se propone la estructura conceptual presentada en la Figura 2. Un núcleo que contiene conceptos (metaclases) que se usan para definir el MCFIS. Dichos conceptos se agrupan en los paquetes *Entidades*, *Vistas* y *Relaciones* (representados por el símbolo de una carpeta). Las flechas entre las carpetas denotan las relaciones de dependencia entre los paquetes; por ejemplo, MCFIS depende de los elementos de modelado contenidos en los paquetes Conceptos, Vistas y Relaciones. La Figura 2 muestra ejemplos de las clases conceptuales (o metaclases) de cada uno de los paquetes.

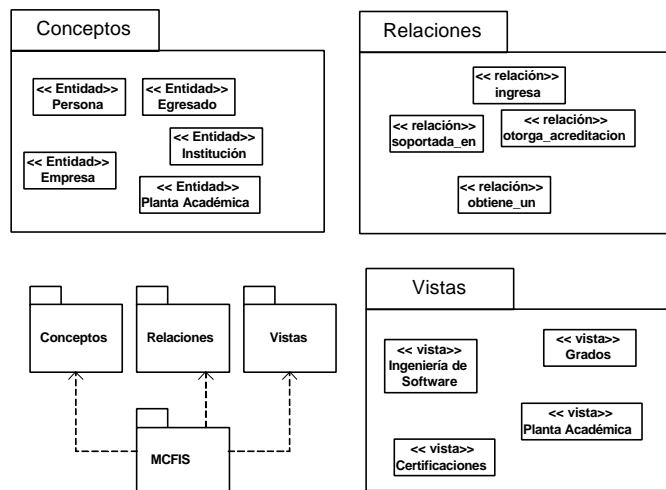


Figura 2. Estructura conceptual parcial del MCFIS.

### 3.3 Estructura de la especificación E-MCFIS 1.0

Partiendo de la estructura conceptual presentada en la sección anterior, se propone que la estructura de la especificación E-MCFIS 1.0 contenga los siguientes elementos.

- Estrategia
- Estructura de los paquetes
- Núcleo de conceptos básicos y relaciones
- Reglas de construcción
- Criterios de interpretación
- Vistas de los modelos
- Extensibilidad y evolución del modelo

En la sección *Estrategia* se presenta los motivos y enfoque seguido para la estructuración de la especificación, incluyendo la estructura conceptual ya mencionada. La sección *Estructura de paquetes* presenta una descripción de los paquetes que contienen los elementos de modelado de E-MCFIS. El *Núcleo de conceptos básicos y relaciones* presenta una descripción detallada de los elementos de modelado que incluyen conceptos y relaciones entre los mismos. Estas dos últimas secciones corresponden a los elementos “núcleo básico de conceptos” y “jerarquía de conceptos” de un perfil de UML. En la sección *Reglas de construcción* se presentan aquellas restricciones relacionadas con la construcción de modelos usando los conceptos y relaciones. Esta sección corresponde en gran manera al elemento “conjunto de reglas de construcción y restricciones” de un perfil de UML. La sección *Criterios de interpretación* contiene descripciones para precisar las interpretaciones de los modelos mostrados. Esto equivale en cierto modo al elemento “semántica” del perfil de UML. En la sección *Vistas de los modelos* se presentan las vistas del modelo MCFIS, como se describe en la subsección 3.4. Finalmente, se menciona en la sección *Extensibilidad* la estrategia y reglas a seguir si se desea extender o modificar el modelo MCFIS.

### 3.4 Vistas de interés

Las vistas de interés permiten ver bajo un enfoque particular los conceptos del modelo y las relaciones que existen. La descripción inicial del MCFIS[1] contiene cinco vistas definidas hasta el momento: i) Vista centrada en Ingeniería de Software, ii) vista centrada en la Planta Académica, iii) vista centrada en los Grados, iv) vista centrada en las Certificaciones y v) vista centrada en la Institución.

La información de cada vista dentro del E-MCFIS incluye:

- Propósito de la vista
- Conceptos que aparecen en la vista
- Relaciones que aparecen en la vista
- Escenarios genéricos de la vista
- Reglas semánticas de la vista

Los últimos dos puntos de la información contenida en la vista especifican mayor información del modelo en particular. Por ejemplo, se pueden definir escenarios concretos para situaciones particulares de la vista. Las reglas semánticas de la vista se refieren a restricciones que deben de cumplirse en la vista. Por ejemplo, si el modelo

general especifica cierta cardinalidad entre dos conceptos entonces la vista puede refinar una cardinalidad más específica para dichos conceptos.

### **3.5 Extensibilidad y mejora del modelo**

La estrategia para permitir la extensibilidad del modelo se fundamenta en la estructura conceptual presentada en la subsección 3.2. La capacidad de extensión del modelo se logra mediante la creación de un núcleo y extender el modelo a partir de éste. Así mismo, se deben de incluir algunas reglas adicionales para garantizar la consistencia de los modelos creados como instancia de E-MCFIS 1.0.

## **4. Conclusiones y Trabajo Futuro.**

Este trabajo ha presentado una estrategia para proveer al MFCIS de una estructura y organización que permita realizar un análisis sistemático del modelo mismo, incluir criterios de interpretación del modelo y permitir una difusión y evolución similar a la que tienen los estándares de la industria. Dicha estructura sigue el esquema básico usado en los perfiles de UML. Así mismo, se introduce el uso de vistas de los modelos que mejora la interpretación y análisis del modelo mismo.

Existen dos líneas de trabajo futuro. La primera, es generar la especificación E-MCFIS 1.0 en la que se vislumbra una serie de refinamiento en los conceptos, vistas y relaciones establecidas anteriormente. Dichos cambios provienen del análisis del modelo basándose en el uso de las vistas propuestas. La segunda, es aplicar el E-MCFIS a esfuerzos que se han estado haciendo por parte de la comunidad de investigación. Por ejemplo, E-MCFIS 1.0 incluirá los resultados del grupo de trabajo GT/MCFIS en la reunión celebrada el 31 de octubre de 2003, en el CIMAT, en la ciudad de Guanajuato, Gto.

## **Referencias**

- [1] L. F. Fernández, C. Montes de Oca, "Marco Conceptual para la Formación de Recursos Humanos en Ingeniería de Software", *4º Congreso Internacional de Ciencias de la Computación (ENC 2003): Avances en Ciencias de la Computación*, CIC-IPN y SMCC, Sep. 8-12, Tlaxcala, México, 2003, pp. 351-356.
- [2] Secretaría de Economía, Programa para el Desarrollo de la Industria del Software, versión Junio 2002. Disponible en <http://www.economia.gob.mx/?P=1128>.
- [3] OMG Unified Modeling Language Specification Version 1.5, Object Management Group, 2003, Disponible en <http://www.omg.org>.
- [4] P. Clements, R. Kazman, and M. Klein, *Evaluating Software Architectures: Methods and Case Studies*, Addison-Wesley, SEI Series in Software Engineering, 2002, pp. 368.

# Evaluación de la utilización de Rational Rose con el Proceso Unificado de Desarrollo de Software

Ángeles Arenas Valdés, Juan Manuel Fernández Peña, Ángeles Sumano López

Especialización en Ingeniería de Software, Universidad Veracruzana, Av. Xalapa S/N,  
Col. Obrero Campesino, C.P. 91020, Xalapa, Veracruz, México.  
{aarenas,jfernandez,asumano@uv.mx}

## Abstract

The Unified Software Development Process involves multiple artifacts' definition and handle, that must be managed because they suffer many transformations along the fundamental flows. It is convenient the use of a CASE tool, like Rational Rose. The present work shows how Rational Rose and Unified Software Development Process work together in the development of software systems in Xalapa and Veracruz harbor zone, inside of Veracruzana University's Software Engineering Specialty, also shows some compatibility problems between them and their solutions, and a Rational Rose user's evaluation.

## Resumen

El Proceso Unificado de Desarrollo de Software involucra la definición y manejo de múltiples artefactos, mismos que deben administrarse debido a la transformación que sufren desde un flujo fundamental a otro. Para ello, conviene utilizar una herramienta CASE, tal como Rational Rose. El presente trabajo muestra la forma como se están utilizando juntos, Rational Rose y el Proceso Unificado de Desarrollo de Software, en el desarrollo de sistemas de software de la zona Xalapa y puerto de Veracruz, dentro de la Especialización en Ingeniería de Software de la Universidad Veracruzana, los problemas de compatibilidad que presenta esta utilización, como se han superado algunos de estos problemas y la evaluación que se hizo a los usuarios de Rational Rose.

**Palabras clave:** CASE, PUDS, Rational Rose, modelos

## Introducción

En la Especialización de Ingeniería de Software de la Universidad Veracruzana, desde 2001, se utiliza el Proceso Unificado de Desarrollo de Software (PUDS), para el desarrollo del proyecto que cada alumno debe elaborar durante sus estudios y, como herramienta para automatizar las actividades del proceso, se hace uso de Rational Rose. La combinación de estos dos elementos ha permitido generar proyectos de una manera ordenada. Sin embargo, esta conjunción ha presentado algunos problemas de comprensión y adaptación. Algunos de ellos han logrado superarse y otros no.

PUDS involucra una serie de fases, dentro de cada una de ellas se trabaja en todos los flujos de trabajo fundamentales, los cuales utilizan diversos artefactos y diagramas que permiten ir modelando un producto de software hasta llegar a su implementación.



Para facilitar este proceso existen diversas herramientas CASE, como Rational Rose<sup>1</sup> que apoya a diversas metodologías basadas en UML (Lenguaje Unificado de Modelado), entre ellas al PUDS.

El presente trabajo está orientado a compartir las experiencias logradas en el uso de Rose como herramienta de modelado, bajo la metodología PUDS. Se divide como sigue: la herramienta Rational Rose, una breve explicación de los flujos fundamentales en PUDS, la forma de trabajo conjunto, y la evaluación que se hizo con usuarios de Rose.

## **1. Rational Rose.**

Rational Rose es un software poderoso que provee un conjunto de herramientas visuales para un desarrollo robusto de soluciones eficientes para necesidades reales de negocios cliente/servidor, empresas distribuidas y sistemas de tiempo real [3]. Esta herramienta da soporte a las metodologías que utilizan UML para crear sus modelos. Rose organiza sus modelos en cuatro vistas y se explican a continuación.

1. Vista de Casos de Uso, en la cual se incluyen los actores, Casos de Uso y diagrama de Caso de Uso del sistema.
2. Vista de componentes, que contiene información de las librerías, archivos ejecutables y otros componentes.
3. Vista de Despliegue, que muestra la arquitectura física del sistema, en ésta se incluyen: Procesadores, dispositivos (terminales, impresoras), dentro del diagrama de despliegue donde se especifican las configuraciones físicas de red
4. Vista lógica, en la que se implementa el comportamiento de los Casos de Uso, incluye diagramas de clases, diagramas de estados, diagramas de interacción (secuencia y colaboración) y paquetes.

## **2. Proceso Unificado de Desarrollo de Software**

Este proceso utiliza el Lenguaje Unificado de Modelado (UML); está dirigido por Casos de Uso, centrado en la arquitectura, iterativo e incremental [2].

PUDS presenta un marco de trabajo genérico para diversos tipos de sistemas de software. El desarrollo del producto, consta de cuatro fases: inicio, elaboración, construcción y transición. A su vez, cada una de las fases está formada por iteraciones y en cada iteración se realizan flujos de trabajo fundamentales; cada uno de ellos va transformando los Casos de Uso, es decir los requerimientos del usuario, a grandes rasgos de la siguiente manera:

1. Requerimientos. Se obtiene el modelo de Casos de Uso compuesto por actores y Casos de Uso. A partir de ellos se va ligando todo el sistema.
2. Análisis. Se realizan los Casos de Uso de análisis agrupándolos en diferentes Paquetes de Análisis (PA).

---

<sup>1</sup> Al referirnos a Rational Rose o simplemente Rose no se incluye la suite completa, únicamente la herramienta central.

3. Diseño. Compuesto por dos modelos: Diseño y Despliegue. Empieza estableciendo la arquitectura mediante el trazado de los PA a Subsistemas de Diseño y conforme al modelo de despliegue. Se hace la realización de Casos de Uso de diseño y se sitúan los diversos artefactos en los Subsistemas de Diseño correspondientes.
4. Implementación. Se mapean los Subsistemas de Diseño a Subsistemas de Implementación de manera biunívoca. Se trazan las clases de diseño a componentes. Se realiza el plan de construcciones.
5. Prueba. Se refinan los casos de prueba por cada Caso de Uso, se realizan las pruebas de integración de los Caso de Uso.

<b>ROSE PUDS</b>	<b>Vista de Casos de Uso</b>	<b>Vista Lógica</b>	<b>Vista de Despliegue</b>	<b>Vista de Componentes</b>
Requerimientos	- Modelo de Casos de Uso			
Análisis	- Modelo de Análisis (Paquetes de Análisis, Realización de Casos de Uso: diagramas de clase y de colaboración y flujo de sucesos y restricciones)			
Diseño		- Modelo de Diseño (Subsistemas de Diseño, Realización de Casos de Uso: diagramas de clases y de secuencia, flujo de sucesos y restricciones)	- Modelo de Despliegue	
Implementación				- Modelo de Implementación (Arquitectura de implementación, diagrama de cada Subsistema de Implementación)
Prueba	X	X	X	X

**Tabla 1. Relación de flujos de PUDS con Vistas de Rational Rose**

### **3. Utilización conjunta de PUDS y Rational Rose**

En la Especialización en Ingeniería de Software, los alumnos realizan, con el CASE Rational Rose, la mayor parte de los modelos que establece el PUDS en sus diferentes flujos fundamentales, a saber: requerimientos, análisis, diseño, implantación y pruebas. La relación de los flujos fundamentales y las vistas de Rose se hace como lo muestra la Tabla 1, debe notarse que el flujo de pruebas no tiene ningún artefacto incluido, pues Rose no brinda ningún apoyo.

Además, los modelos de análisis y Casos de Uso se ingresan en un archivo mdl diferente al de diseño e implementación, tal decisión se debe a dos razones:

- 1) Desde el punto de vista estético y de claridad. Dado que Rose va identificando cada artefacto con la vista donde fue creada, la lectura de los documentos se oscurece (por ejemplo, cambia los estereotipos de las clases).
- 2) La relación con PUDS. Los Paquetes de Análisis (PA) no serán, obligatoriamente, los Subsistemas de Diseño y, entre otros problemas, los nombres naturales puestos en los PA limitarían la nomenclatura para los Subsistemas de Diseño. Además, los Subsistemas de Diseño se corresponden uno a uno con los Subsistemas de Implementación.

### **3.1 Modelando con Rational Rose los productos de PUDS**

En general se trabaja de manera Top-down, pero al revisar los productos o hitos de las distintas iteraciones se recurre al modelado Bottom-up. Una vez ingresando a Rose, la realización de cada modelo es como sigue:

- 1) Para los modelos de Casos de Uso y de Análisis, en Rose se debe proceder de manera algo distinta, como sigue:
  - a) En el diagrama principal (main) de la Vista de Casos de Uso se incluyen los paquetes de análisis.
  - b) A cada paquete de análisis se le da clic derecho y se crea un diagrama de Casos de Uso, dentro del cual se hace el modelo de Casos de Uso de cada paquete.
  - c) A cada Caso de Uso se le asocian los diagramas de clases de análisis (estereotipos) y los de colaboración (pudiendo ser con clic derecho). Cada diagrama lleva asociados sus restricciones y flujos de sucesos respectivamente en el recuadro inferior izquierdo (de comentarios).
- 2) Para los modelos de diseño se realizan, en orden, las siguientes actividades:
  - a) Dentro de la Vista de Despliegue, se coloca el modelo de despliegue donde hay que colocar los nodos que conformarán la red, en que correrá el sistema a desarrollar, indicando el recurso de cómputo que representa: procesador (cubos sombreados) o dispositivo (cubo sin sombreado), se establecen las relaciones entre ellos, es decir el medio de comunicación, como Internet o Intranet entre procesadores y, para la conexión procesador y dispositivo, por ejemplo: lpt1: o bus. A cada procesador se le asignan los procesos que en él correrán. Este último paso permite que el desarrollador se percate de la necesidad de dividir el sistema en Subsistemas de Diseño y ya no, necesariamente, en PA, pero sirven de punto de partida.
  - b) Una vez identificados los diferentes procesos, en la Vista Lógica, se identifican los distintos Subsistemas de Diseño acomodándolos en la capa que les corresponde (específica o general de la aplicación, middleware o del sistema).
  - c) A cada Caso de Uso de diseño se le asignan varios diagramas y construyéndose en la raíz de la Vista Lógica:
    - i) El diagrama de clases y/o Subsistemas de Diseño que intervienen en el Caso de Uso, cuidando que los estereotipos coincidan con alguno del

- lenguaje a utilizar y que provengan o equivalgan a una clase del modelo de análisis.
- ii) Los diagramas de secuencia asociados a cada Caso de Uso de diseño se construyen al mismo nivel del diagrama de Subsistemas de Diseño y de clases.
  - d) Por último, se asocia a cada Subsistemas de Diseño un diagrama de clases arrastrando todos los artefactos que estarán en él, que no precisamente forman una unidad lógica, aunque sí física.
- 3) Utilizando la Vista de Componentes, para el modelo de implementación, se crean:
- a) Arquitectura del modelo de implementación (equivalente a la de diseño), que incluye los componentes que conforman cada Subsistema de Implementación indicando su estereotipo: exe, dll o bd (exclusivamente) y se establecen las relaciones de dependencia entre ellos
  - b) Para cada Subsistema de Implementación se hace su diagrama de componentes del lenguaje escogido y que contienen una o más clases de diseño.

## **4. Evaluación**

Tanto PUDS como Rational Rose han unificado diferentes notaciones y técnicas y a pesar de que pertenecen a la misma empresa, Rational Software Corporation, existen algunas complicaciones para desarrollar un proyecto combinándolas. Una de las razones más importantes es que el CASE de Rational Rose pretende dar soporte a todas las metodologías basadas en UML y no en especial a PUDS. A continuación, se muestran dos partes de la evaluación, una desde el punto de vista de usabilidad y la de problemas de modelado más evidentes.

### **4.1 Encuesta a usuarios**

Por el lado de usabilidad, una encuesta exploratoria basada en la que propone [4] y aplicada a doce alumnos de la Especialización en Ingeniería de Software, mostró que Rose es bastante aceptable en cuanto a facilidad de aprendizaje, capacidad de obtener los productos deseados de PUDS y mejorar la productividad. Sin embargo, apenas logró calificaciones entre 2 y 2.4, cuando la escala iba de 1 a 5, donde 1 era la mejor calificación. Aunque con pocos datos, la evaluación de quienes usan otras herramientas, como MS-Visio, dieron resultados semejantes.

### **4.2 Resumen de problemas**

En resumen, los problemas más molestos que se han tenido con Rose son los siguientes:

- a) Falta total de soporte al flujo de Pruebas, que obliga a usar otras herramientas más generales.

- b) Dentro del flujo de implementación no incluye soporte para un artefacto muy importante: el plan de integraciones; además su soporte para la generación de código no es muy cómodo y se limita a la parte estática de los modelos; en nuestro caso, además, no sirve para la mayoría de los proyectos que emplean Delphi y ha tenido poco uso cuando se desarrolla en Visual Basic.
- c) En diagramas de secuencia aún falla el manejo gráfico del alcance de la vida de los métodos (rectángulos verticales sobre la línea de vida), lo cual genera diagramas confusos, fragmentados.
- d) Si se requiere hacer un cambio en los diagramas de colaboración, es necesario borrar todos los mensajes anteriores y posteriores.
- e) Se requiere emplear dos archivos de proyecto: el de análisis y el de diseño, ya que si se trabajan juntos, todos los elementos serían confusos.

## Conclusiones

Durante tres años (generaciones) se ha utilizado el Proceso Unificado de Desarrollo de Software para el desarrollo de aplicaciones diversas. Para elaborar la documentación correspondiente se ha empleado Rational Rose. Sin embargo, se ha observado que muchas veces los alumnos se confunden, ya que no existe una equivalencia biunívoca entre PUDS y Rational Rose, esta última trata de dar soporte a diversas metodologías que tienen como base UML, por lo tanto no presenta los flujos de trabajo fundamentales que se siguen en PUDS, sino que presenta diferentes vistas dentro de las cuales debe colocar los artefactos de los diferentes flujos de trabajo.

No obstante, a pesar de todas las desventajas que surgen en la utilización de Rational Rose con PUDS, también existen algunas ventajas como es la consistencia entre los modelos, ya que si se realiza un cambio en algún elemento, este cambio se ve reflejado en todos los modelos donde aparece. Otro aspecto favorable es el amplio espectro de modelos que permite realizar, superando a otros modeladores como como MS-Visio y Model Maker.

Aunque lo más conveniente sería utilizar un CASE que permita seguir de manera equivalente cada uno de los flujos de trabajo fundamentales que propone PUDS, los autores estamos concientes de que el cambio y evolución de las metodologías harían que una herramienta de esta naturaleza se volviera obsoleta en corto tiempo.

## Referencias

- [1] Boggs Wendy, Boggs Michael, "UML with Rational Rose 2002", Sybex, 2003.
- [2]. Jacobson Ivar, Booch Grady, Rumbaugh, "El Proceso Unificado de Desarrollo de Software", Addison Wesley, 2000.
- [3] Quatrani Terry, "Visual Modeling whth Rational Rose 2002 and UML", Addison Wesley, 2003.
- [4] Ben Shneiderman, "Designing the user interface", Tercera edición, Addison Wesley Longman, 1998.

# El Desarrollo de la Industria de Software en Aguascalientes: Diagnostico y Oportunidades

Tirso Velázquez Navarro, Héctor Sánchez Herrera y Jaime Muñoz Arteaga

Depto. de Sistemas de Información de la Universidad Autónoma de Aguascalientes  
Av. Universidad # 940, fraccionamiento bosques, Aguascalientes, México  
[Tirso.Velazquez@inegi.gob.mx](mailto:Tirso.Velazquez@inegi.gob.mx), [hsanchez@aguascalientes.gob.mx](mailto:hsanchez@aguascalientes.gob.mx) y  
[jmunozar@correo.uaa.mx](mailto:jmunozar@correo.uaa.mx)

**Abstract:** The use of technology is becoming indispensable in our daily activities. For example the government, the enterprises and the schools could satisfy the client requirements if their services are offered throughout a web based information system. Then it is important to analyze what factors of the technology could facilitate the daily activities of clients. This paper analyzes the conditions of Aguascalientes's software industry in order to propose a set of recommendations for increase the production of software at local level.

**Resumen:** Actualmente es imposible imaginarse la vida cotidiana de la sociedad sin el involucramiento de las tecnologías. La economía de los países, los gobiernos, los individuos. etc. dependen en gran medida de la tecnología, por esto es, importante analizar los diferentes aspectos que influyen en el desarrollo tecnológico de los países. El presente documento presenta un análisis de las características particulares del estado de Aguascalientes y ofrece un conjunto de recomendaciones para incrementar su desarrollo de la industria de software.

## 1. Introducción.

El rol de la Información y las Tecnologías de Comunicación (TIC) en el crecimiento de la economía y los cambios sociales han recibido una considerable atención en los años recientes, particularmente en el debate de la “nueva economía”. La producción, difusión y uso de las TIC varían considerablemente entre y dentro de los países, aunque estas continúan con su propagación y su importancia económica ha crecido desde los años 90's. Indicadores confiables y comprensibles son necesarios para conducir los desarrollos en las nuevas tecnologías de información y entender sus impactos en nuestras economías y sociedades [1].

La contribución de la economía de la información sobre todo al crecimiento económico esta relacionado a la cantidad de los recursos destinados a las nuevas tecnologías de información, o sea en términos de consumo, infraestructura ó esfuerzos de innovación [1]. La diferencia entre los que tienen acceso a la información y los que no la tienen ha sido bautizada con el nombre de “Brecha Digital: la brecha entre individuos, familias, negocios y áreas geográficas de diferentes niveles socioeconómicos en relación a la oportunidades de acceso a las tecnologías de la información y la comunicación” [2].

Milton Mueller [3] afirma que la relación positiva que existe entre crecimiento económico e infraestructura en telecomunicaciones no representa una justificación sólida para implementar una política que promueva el acceso a los servicios. Los países más ricos tienen los mayores niveles de penetración en telecomunicaciones porque la riqueza genera mayor inversión en infraestructura, y no al revés.

En este sentido, en el presente artículo se hace un análisis de las características particulares del estado de Aguascalientes con el propósito de poder definir cual es el potencial en el desarrollo de la industria de software como parte de los desarrollos tecnológicos.

## **2. Aguascalientes y el entorno nacional.**

En lo que respecta a la “Economía Digital en México” se han realizado diferentes investigaciones de las que se pueden resaltar que el sector informático participa con 3.5% del total de la economía nacional. A su interior, el sector de mayor crecimiento fue el de las telecomunicaciones, con un incremento del 28.4% durante el año pasado, seguido del equipo y periféricos para procesamiento informático que lo hizo en 22.9%. El sector Servicios Profesionales en Informática creció 4.9% durante el año pasado a tasa anual y en términos reales. El desarrollo de la industria manufacturera informática se encuentra ampliamente concentrada regionalmente en los estados de la frontera norte y el occidente del país (Baja California, Chihuahua, Nuevo León, Sonora, Tamaulipas y Jalisco), de tal forma que estas 6 entidades en conjunto absorben el 69% de los establecimientos y el 94% del personal ocupado [4]. El parque instalado de computadoras personales (PC's) en México muestra un crecimiento constante. Se estima de manera preliminar que existen 82 equipos por cada mil habitantes, notablemente inferiores a la que observan nuestros principales socios comerciales como son Estados Unidos y Canadá con 658 y 487 computadoras por cada mil habitantes, en cada uno de ellos [4].

Con relación al aspecto telemático en México, a nivel nacional la densidad telefónica es de 11.2%. Existiendo cinco grandes grupos: el primero conformado por 5 entidades cuya densidad telefónica es superior a 15% (Baja California, Baja California Sur, Distrito Federal, Jalisco y Nuevo León); un segundo grupo de 11 estados con índices entre 10 y 15% (Aguascalientes, Colima, Coahuila, Chihuahua, Estado de México, Morelos, Querétaro, Quintana Roo, Sinaloa, Sonora y Tamaulipas); el tercer grupo de 6 entidades con índices entre el 8 y 9% (Durango, Guanajuato, Michoacán, Nayarit, Puebla y Yucatán); el cuarto grupo integrado por 7 estados, con índices entre 6 y 7% (Campeche, Guerrero, Hidalgo, San Luis Potosí, Tlaxcala, Veracruz, Zacatecas); y el último grupo conformado por 3 entidades con una densidad telefónica menor al 6% (Chiapas, Oaxaca y Tabasco). De acuerdo con los últimos datos reportados por el INEGI, Aguascalientes en el 2003 ocupó el lugar 14 entre los 32 estados del país con una densidad telefónica del 15.8% [6].

En las viviendas mexicanas solamente el 9.3% cuenta al menos con una computadora. Este último dato es altamente indicativo de la existencia de la denominada “Brecha

Digital”, acceso desigual que tienen las personas a las Tecnologías de la Información, situación que está causando importantes disparidades entre las entidades más desarrolladas y las que presentan mayores rezagos. Mientras en el Distrito Federal 21.6% de las viviendas poseen computadora, en el extremo opuesto siete entidades presentan porcentajes inferiores a 5 por ciento [4].

Entidad	Índice de Capital Humano	Índice de Aprendizaje e Innovación	Índice de Empresarialidad	Índice de Entorno Favorable	Índice de Inf. para la Economía Digital	Valor del Mercado	Índice de Capacidades Locales
Nacional	5.00	4.20	4.45	5.02	4.69	5.0	4.73
Aguascalientes	7.16	3.44	2.56	7.41	3.07	2.90	4.42

**Fig.1** Índices del desarrollo de la Industria de Software Nacionales vs. Estado de Aguascalientes

De acuerdo a la tabla anterior y al “Estudio de las Entidades con Programa de Desarrollo de la Industria de Software” [5], el estado de Aguascalientes cuenta con un índice de capital humano por arriba del nacional gracias al grado de escolaridad promedio de 7.9 años, superior al promedio nacional de 7.6. Otra de las fortalezas del estado radica en que cuenta con un índice de entorno favorable superior al nacional. Las fortalezas del estado no logran superar a sus debilidades ello resulta en un índice de capacidades locales (4.42) menor al nacional (4.73). Con el fin de mitigar esta problemática se fundó el Cluster de Tecnologías de Información en Aguascalientes el cual tiene como principal misión fomentar el desarrollo, transferencia y protección de las tecnologías de la información, buscando la apertura de mercados generando recursos propios. [9].

Hace 2 décadas se empezó a desarrollar la industria del software, la cual inició como parte interna de los procesos de producción (“embedded software”) en la industria automotriz o en la electrónica (Xerox y Texas Instruments), lo que originó el establecimiento de pequeños negocios locales de software. En esta perspectiva, los negocios de software empezaron a multiplicarse y ello condujo a que el gobierno de Aguascalientes en el año 2000 decidiera desarrollar un cluster de alta tecnología creándose una Asociación Civil que agrupa a las empresas del cluster del Software el cual se ha denominado “Innovatia”, constituido hoy día por 34 empresas, con ventas alrededor de 4.3 millones de dólares, a lo anterior se suma el hecho de que el gobierno decidiera crear una institución de apoyo tecnológico para apoyar el desarrollo de las tecnologías de la información en coordinación con el Consejo de Ciencia y Tecnología del Estado de Aguascalientes (CONCYTEA), contraparte local del Consejo Nacional de Ciencia y Tecnología (CONACYT), al cual se le denominó CEDITI (Centro de Desarrollo de Tecnologías de la Información).



### 3. Diagnostico de los núcleos de Aguascalientes

El resultado del estudio del cluster de tecnologías de información en Aguascalientes [5] se muestra en la figura 2. En esta figura se hace énfasis en la las fortalezas y debilidades que se tienen tanto en el ámbito nacional como el estatal, estas se definen por medio de las seis variables siguientes: núcleos de capital humano, aprendizaje e innovación, empresarialidad, entorno favorable, infraestructura para la economía digital y valor de mercado. Las variables son representadas por medio de una grafica de tres ejes en la que claramente se puede observar el comportamiento de México en su conjunto (color rojo) y el correspondiente al estado de Aguascalientes (color azul).



**Fig. 2** Fortalezas y debilidades del estado de Aguascalientes en el desarrollo de software.

La figura 2 muestra que en el estado de Aguascalientes el 8.17% de la población tiene instrucción superior y solo el 0.53% tiene maestría o doctorado (índice de *capital humano*). La normalización de las variables educativas da como resultado un índice de capital humano de 7.16, mientras que el índice a nivel nacional es de 5, es decir, los esfuerzos en capital humano en Aguascalientes han sido mayores que la media nacional. De las 88 instituciones de educación superior de todo el país que en 2002 contaban con postgrados en Informática, sólo 2 se ubican en Aguascalientes, lo cual impacta directamente en la formación de capital humano [7].

En cuanto al índice de *Aprendizaje e Innovación*, se puede comentar que los 7,668 investigadores que hay en el país, 36 de ellos se encuentran en Aguascalientes; también 2 de los 161 centros de investigación que existen en el país se encuentran en este estado; además, de los 5,771 registros de derechos de autor en programas de computo, 108 fueron hechos en Aguascalientes, y de las 13,791 becas otorgadas por el CONACYT en el 2000, 63 se otorgaron en esta entidad. La normalización de estos indicadores permite obtener un índice de aprendizaje e innovación de 3.44, siendo que el índice a nivel nacional es de 4.20, es decir, los esfuerzos en aprendizaje e Innovación no han sido suficientes.

En lo que concierne al índice de *empresarialidad*, los 2'726,366 establecimientos no agrícolas que existen en el país solo el 1% de ellos se ubican en esta entidad; y de las 539 empresas que lograron la certificación ISO-9000 en el año 2000, solo 9 de ellas se ubican en Aguascalientes. En este sentido, al normalizar estas variables, se obtiene un índice de empresarialidad de 2.56 el cual esta por debajo del índice de empresarialidad a nivel nacional el cual es de 4.45.

La variable de *entorno favorable* para el desarrollo de una economía digital, en el estado de Aguascalientes existen 123.8 líneas telefónicas por cada mil habitantes, además, el 6.39% de los hogares en Aguascalientes cuentan con PC e Internet siendo

que a nivel nacional el promedio es de 6.16%. En cuanto a la gestión de trámites empresariales, en Aguascalientes se necesitan solo 30 días para abrir un negocio solo superado por el estado de Sinaloa en donde solo se requieren de 29 días para abrir un negocio. En este sentido, el índice de entorno favorable es de 8.27, mientras que el índice a nivel nacional es de 4.98.

Finalmente en cuanto a la *Infraestructura para la Economía Digital*, solo el 1.12% del total de alumnos matriculados en las carreras de computación y sistemas se encuentran en Aguascalientes, solo 7 de las 531 escuelas de nivel superior se encuentran en esta entidad; por otro lado, del total de activos fijos en la industria de las tecnologías de la información solo el 0.31% se encuentra en este estado; y de las 2,095 empresas en la Industria de las Tecnologías de la Información solo 29 de ellas se encuentran en Aguascalientes. En este sentido, al normalizar estas variables se obtiene el Índice de Infraestructura para la economía digital con valor de 3.07, mientras que el Índice a nivel nacional es de 4.69. El mercado de TI en el estado de Aguascalientes es uno de los mercados más pequeños del país su valor fue en el 2001 de 11.6 millones de dólares que representa el .48% del valor del mercado nacional de TI. A partir de esto, se tiene que el índice de valor de mercado es de 2.90, muy por debajo del promedio nacional que es de 5

#### **4. Oportunidades para el Estado de Aguascalientes**

Con el fin de hacer frente a las debilidades (mostrada en la figura 2) en cuanto al desarrollo de software que presenta el estado de Aguascalientes propone aquí ciertas recomendaciones para mejorar a un mediano plazo los índices de “Aprendizaje e Innovación”, “Empresarialidad” y a un largo plazo la mejora de los índices de la “Infraestructura para la Economía Digital” y “Valor del mercado. Las recomendaciones que se proponen son las siguientes:

- Fortalecer y promover la capacitación especializada de los Profesionales Informáticos ya en ejercicio y de los que están en formación en herramientas informáticas de vanguardia sobre todo de aquellos de gran demanda en el mercado mundial, en el uso y aprovechamiento de las herramientas Case, y muy especialmente en el Capability -Maturity Model (CMM) con objeto de ir generando una cultura de calidad en el trabajo y de mejora continua del proceso de desarrollo del software.
- Proponer a las instancias de promoción tecnológica (CONACYT, CONCYTEA, IEA, SEP entre otros), el otorgamiento de becas y premios, que incentiven a las instituciones educativas a la implementación de estas materias como parte fundamental de sus programas de estudios profesionales informáticos.
- Por su parte las principales instituciones de educación superior de la región (v. gr. Universidad Autónoma de Aguascalientes (UAA), Universidad Bonaterra, Universidad Cuauhtémoc, Universidad Tecnológica (UNITEC), Instituto Tecnológico de Estudios Superiores de Monterrey Campus Aguascalientes, Instituto Tecnológico de Aguascalientes) pueden consolidar su vínculo con el medio industrial local a fin de poder responder a las necesidades de la demanda de

la sociedad y de esta manera generar una cadena de valores con la que se tenga una presencia competitiva en los mercados local, nacional e internacional.

Las medidas anteriores a mediano y largo plazo (de 2 a 6 años) proporcionarían la base necesaria y la posibilidad de poder promover el desarrollo de una “Industria de Software” exitosa para el estado ya que en este tipo de actividad, el Capital Humano y las capacidades de éste es lo más importante, además de que el entorno local y la cercanía con el mercado mundial más grande son factores favorables adicionales para su establecimiento. Nuestra propuesta apoyaría el desarrollo de la industria de software en Aguascalientes para competir por la maquila de proyectos de empresas internacionales muy particularmente de los Estados Unidos de Norteamérica, ofreciendo servicios de calidad comprobable a través de la certificación de capacidades y utilización de procedimientos de control de calidad del producto.

## 5. Conclusiones

El presente trabajo hacen ver que el estado de Aguascalientes cuenta con un fuerte potencial para el desarrollo de software en particular cuenta con un entorno favorable (con un índice de 8.27 sobre un 4.98 en el país) y con un importante capital humano (con un índice de 7.16 sobre un 5 en el país). Sin duda, es necesario resaltar también la necesidad de mejorar en los aspectos de empresarialidad, infraestructura para la economía digital y en el aprendizaje e innovación en los que el estado de Aguascalientes tiene índices por debajo de los presentados en el país. Una de las posibilidades para solventar las anteriores debilidades es el apoyo del estado para el desarrollo de las tecnologías de la información mediante el Centro de Desarrollo de Tecnologías de la Información (CEDITI) [5]. Este aspecto de gran relevancia deberá ser vinculado con al aspecto empresarialidad y los ofrecimientos de los tipos de conocimientos por parte de la academia, a fin de que pueda ser impulsado el desarrollo de la industria del software con características de gran calidad

## Referencias

- [1] OECD: Organization for Economic Co-Operation and Development, Measuring the Information Economy 2002-2003 pp. 3-94, <http://www.oecd.org>
- [2] Revista “Política Digital” Innovación Gubernamental, ISSN 1665-1669 Número 16 Abril / Mayo 2004; pp. IX – XI del Suplemento Especial Rayón. Editor responsable y Director General Andrés Hofmann ([andres@politicadigital.info](mailto:andres@politicadigital.info))
- [3] Muller, Milton L., Universal Service: Competition, Interconnection and Monopoly in the Making of the American Telephone System 1997.
- [4] Boletín Informativo Aguascalientes, Ags., 6 de Marzo de 2001, número 028/2001, pp. 2-5.
- [5] AMITI, Entidades con Programa de Desarrollo de la Industria de Software, pp. 8-9. 2003 <http://www.amiti.org.mx>
- [6] INEGI, “Densidad de líneas telefónicas fijas en servicio por entidad federativa, 1998-2003”,

- [7] INEGI, "Escuelas con Postgrado en Tecnología de Información y Comunicaciones por Entidad Federativa 1997/1998-2001/2002", [www.inegi.gob.mx](http://www.inegi.gob.mx).
- [8] Naciones Unidas-Consejo Económico y Social, Informe de síntesis sobre los grupos de la Comisión de Ciencia y Tecnología para el Desarrollo, Ginebra Suiza Mayo del 2003.
- [9] Gobierno del Estado de Aguascalientes, Secretaría de Desarrollo Económico, "Agrupamientos Industriales", <http://www.aguascalientes.gob.mx/sedec>

# Mejora de Calidad del Software Desarrollado por Publisoft de Aguascalientes en base al CMM

José Guillermo Torres Tovar, Maribel Romo Contreras y Jaime Muñoz Arteaga

Dpto. de Sistemas de Información de la Universidad Autónoma de Aguascalientes  
Av. Universidad # 940, Fraccionamiento Bosques  
C.P. 20100, Aguascalientes, México  
joseguillermotorres@hotmail.com, romo\_maribel@hotmail.com,  
jmunozar@correo.uaa.mx

**Resumen.** El presente artículo pretende analizar el nivel de madurez que se tiene en el desarrollo de software en la empresa Publisoft de Aguascalientes, esta empresa presenta un problema en la administración de recursos materiales y humanos durante el ciclo de desarrollo de sistemas. En el estudio que se presenta se propone el uso de CMM como herramienta para analizar el nivel de madurez, considerando particularmente a la empresa Publisoft para determinar en que nivel de madurez se sitúa y así poder proponer una serie de actividades que lleve a esta organización a una mejora continua para que desempeñe sus actividades de una manera más correcta y ordenada. Es de interés de la empresa identificarse y compararse con un modelo internacional sobre el desarrollo de software, lo cual permitirá identificar las principales actividades funcionales para obtener la mejora continua y la calidad en el desarrollo de software.

**Abstract.** The present paper tries to analyze the maturity level that the development of software has in the company Publisoft of Aguascalientes. This company has a problem in management of material and human resources during the system development cycle. In the study that appears sets out the use of CMM like a tool to analyze the maturity level, considering particularly to the Publisoft company to determine what the maturity level is and propose a series of activities that takes this organization to a continuous improvement to do the activities in the correct way and form. The company tries to identify itself and to compare itself with an international model on the development of software, which will allow to identify the main functional activities to obtain the continuous improvement and quality in the development of software.

## 1 Introducción

En los últimos años la industria mundial de desarrollo de software se ha preocupado por mejorar sus capacidades en el desarrollo de software de calidad. El Modelo de Capacidad de Madurez (CMM) [1] es un modelo reconocido para la calidad del desarrollo de Software. Las empresas hacen esfuerzos para implantar estos modelos y lograr la certificación o evaluación en alguno de ellos con el objetivo de obtener ventajas competitivas.

Dentro de los últimos cinco años las grandes corporaciones internacionales como Motorola, IBM y Volkswagen han introducido estas ideas en México y las han implementado en sus filiales en nuestro país. Esta situación ha despertado la inquietud y el interés de otras empresas mexicanas por mejorar sus propios procesos de desarrollo y no quedarse fuera del mercado.

En la presente investigación se analiza la empresa PubliSoft de Aguascalientes, esta empresa se dedica al desarrollo de software de escritorio, así como al diseño de páginas web, portales, etc. Actualmente cuenta con una metodología de administración de recursos humanos y materiales, sin embargo, esta empresa muestra deficiencias en la coordinación de desarrollo de proyectos. Se determinó que la metodología a seguir sea CMM, ya que ésta le permitirá conocer su nivel de calidad de software.

Dentro de la metodología de CMM podemos encontrar 5 niveles, de los cuales la tarea de esta investigación es identificar a través de una serie de puntos que se deben cumplir en cada uno de ellos, si alcanza a cubrir el nivel 2, que es nivel que en este momento quiere alcanzar la empresa; lo cual servirá como base de comparación cuando se lleve a cabo el estudio, los niveles son: Caos(1er nivel), Repetible(2do nivel), Definido(3er nivel), Medible(4º nivel), Mejora continua(5º nivel). [2]

## **2 Problemática**

En México a pesar de que muy pocas empresas de software están certificadas con alguna norma reconocida de Ingeniería de Software, se está dando a conocer el interés que está surgiendo año con año por lograr tener un reconocimiento de la calidad con que llevan a cabo sus actividades, así como del uso de herramientas que apoyan a la administración de recursos. [5]

Además México es considerado a nivel mundial como un posible competidor fuerte dentro de la industria de desarrollo de software [4], por todas las características que rodean al país (ubicación privilegiada, costos competitivos, cultura empresarial similar a la de EUA) [6]. Es por este interés que se realizó un estudio dentro de esta empresa, para la búsqueda de posibles elementos que estén deteniendo el pleno desarrollo de productos con calidad y con plena adaptación a las necesidades del cliente.

Los problemas con los que actualmente se enfrenta la empresa son en primer lugar que al no llevar una metodología global para el desarrollo de sistemas, se pueden ir presentando algunos sucesos inesperados para lo cual no se está preparado, como por ejemplo, que durante el desarrollo del software el tiempo que se tenía contemplado para el término del mismo va siendo diferido, ya que es muy común que el cliente solicite cambios dentro de los requerimientos del sistema, lo cual desestabiliza al equipo encargado de ese proyecto.

En segundo lugar, es que al aumentar el número de recursos utilizados en el desarrollo del mismo, aumenta su costo, el cual no siempre puede ser cargado al cliente que está solicitando esos cambios, por lo que la empresa pierde parte de sus utilidades, repercutiendo completamente en sus actividades internas.

### 3 Caso de Estudio

En el presente caso de estudio se investigó de manera externa a la empresa Publisoft de Aguascalientes, ya que es de su interés conocer su nivel de calidad de software, utilizando la metodología de CMM se determinarán los puntos importantes que se deben de cumplir dentro de las áreas claves de procesos (KPA's) del Nivel 2 (aseguramientos de calidad, admón.. de subcontratistas, admón.. de la configuración, seguimiento al proyecto, planeación del proyecto, administración de requerimientos) para determinar si se encuentra en que nivel 2 de CMM y así identificar que acciones deben de corregirse y/o mejorarse para llevar a cabo una organización correcta y ordenada de sus actividades y procesos que se llevan hasta el momento.

Se hizo una revisión de todas las herramientas que se utilizan para llevar a cabo la administración del desarrollo de sistemas, dentro de los cuales la compañía cuenta con un sistema computacional en la que se almacena toda la información del software, como por ejemplo: datos del sistema (nombre, fecha de inicio, nombre del cliente, costo) y datos de recursos humanos (nombre, puesto, actividades, proyectos en los que participo, perfil y habilidades).

Para presentar los resultados de la experimentación que se llevó a cabo mediante la revisión de las herramientas antes mencionadas, se irán presentando las áreas claves de procesos (KPA's) del Nivel 2 que se estarán estudiando, así como la identificación del cubrimiento de las mismas.

Los resultados son los siguientes:

Con lo que respecta al nivel 1 (Inicial) del modelo de CMM, se sabe que no se requiere llevar un proceso definido, por lo que todas las empresas que inician con su evaluación toman esta posición, sin tener que cumplir ningún objetivo o métrica; por lo que la empresa Publisoft inicialmente toma este nivel.

Con lo que respecta al nivel 2 (Repetido) del modelo CMM, las áreas claves de proceso (KPA's) a cubrir son las siguientes:

En la primera área clave de proceso que es la Administración de Requerimientos (RM), se determinó que no se cumple completamente con el mismo, debido a que al establecer un entendimiento común con el cliente sólo queda estipulado los requerimientos iniciales del proyecto de software, sin controlar el aumento o disminución de los mismos durante el desarrollo del proyecto, lo que nos conlleva a tener dificultades de mantener los planes, productos y actividades integrados en dichos requerimientos.

Con lo que respecta a la segunda área clave de proceso denominada Planeación del Proyecto (SPP), el cual cubre el establecimiento de planes y la administración del proyecto, la empresa lo lleva a cabo a través de un software propio, el cual determina tiempos, costos, asignación de recursos humanos y materiales, así como el estatus y avance del proyecto, lo que nos permite identificar que la empresa cumple completamente con este punto.

De acuerdo a la tercera área clave de proceso conocida como Seguimiento al Proyecto (SPTO), la cual tiene como propósito el garantizar la visibilidad adecuada del progreso real del proyecto. Para efectuar el seguimiento de los resultados y la ejecución del proyecto la empresa ya cuenta con una metodología de trabajo propia

que consta de las siguientes fases: Fase 1.- Acopio de información, fase 2.- Propuesta de diseño, fase 3.- Desarrollo de prototipo, fase 4.- Pruebas de aprobación y lanzamiento.

Además de que el software que maneja la empresa abarca el control tanto de los costos, así como de los tiempos en los que se va desarrollando el proyecto de software (Gráfica de Gant). También nos permite visualizar que recursos se están utilizando actualmente y cuando se van a liberar, lo que nos ayuda a determinar si se podrá contar con los recursos necesarios para el desarrollo de un nuevo software.

Por todas estas herramientas que utiliza la empresa se determinó que cumple con esta área, ya que si lleva acabo un seguimiento adecuado de sus proyectos.

Con lo que respecta a la cuarta área clave de proceso llamada Administración de Subcontratistas (SSM), esta no se lleva acabo, ya que no se subcontrata a ningún agente externo, ya que cuando surge la necesidad de iniciar un nuevo proyecto se evalúa si la empresa cuenta con los conocimientos técnicos y capacidad de desarrollo.

La quinta área clave de proceso conocida como Aseguramiento de la Calidad de Software (SQA), la empresa para medir su calidad lleva acabo varios tipos de pruebas: alfa (son pruebas que se realizan al producto final antes de salir al mercado, es decir, dentro de la misma empresa y en un ambiente controlado) y beta (son pruebas que se realizan a un producto final ya dentro del mercado, pero dejando en claro que es una prueba y que puede contener algún error), las cuales se llevan acabo antes del lanzamiento final, lo que le permite disminuir el riesgo de una mala funcionalidad dentro del mercado, lo cual inyecta calidad a nuestros productos.

Sin embargo, no se lleva acabo una adecuada difusión a las personas y grupos involucrados sobre las actividades y resultados de aseguramiento de la calidad, por lo que esta área es cubierta parcialmente.

Por último la sexta área clave de proceso llamada Administración de la Configuración (SCM), su propósito es establecer y mantener la integridad de los productos de software generados a lo largo del ciclo de vida del proyecto. De este punto se detectó que no llego a cumplirse por completo, ya que actualmente no se cuenta con una metodología de trabajo bien definida para el software (web y aplicaciones de escritorio) que principalmente se desarrolla en la empresa. Es decir, por una parte si se conoce y se usa metodologías para desarrollar software, y se tiene un sistema que ayuda a la gestión de proyectos, pero no se tiene una metodología global que permita una completa integración entre ambas partes. Esto conlleva a que no hay un control absoluto sobre los requerimientos y la carga de trabajo sobre cada uno de los proyectos que la empresa va desarrollando, y desafortunadamente las necesidades del cliente aumentan a medida que el proyecto avanza, lo que representa el principal problema de la empresa.

#### **4 Estrategias Para Lograr el Nivel 2 del CMM**

La posible solución después del análisis del nivel 2 de la metodología de CMM, que es el nivel no alcanzado por la empresa, se determinó que es muy difícil medir el grado en que puede crecer un proyecto o el aumento de las necesidades del cliente



durante el desarrollo del software, por lo que creemos que la mejor solución es determinar límites en cuanto a los cambios que se le pueden hacer al producto. Para lo cual se seguirían los siguientes pasos:

Al hacer contrato de inicio de desarrollo de software, determinar todos los elementos que va y no va a cubrir, es decir, redactar claramente los elementos en los que va a constar el software.

El contrato también debe contener ya sea el número de especificaciones que se pueden cambiar o un cierto porcentaje de cambio, esto dependiendo de lo que mejor se adapte a la cantidad y tipo de especificaciones de cada software.

Se debe determinar el tiempo y el número de recursos tomando en cuenta no sólo las especificaciones que se hicieron en el contrato, sino también a lo que se determinó como monto de posible cambio. En caso de superar éste se seguirá el siguiente paso:

Se da por terminado el contrato con lo que ya estaba especificado y se abre otro contrato regresando al paso número uno.

Con lo que respecta a los demás niveles de esta metodología no se tomarán en cuenta para este caso, ya que la inquietud de la empresa, al menos en este momento, es alcanzar el nivel 2 de CMM.



Fig. 1. Modelo para la mejora de calidad de software.

Para la aplicación de las estrategias señaladas anteriormente se propone el siguiente diagrama conceptual (figura 1) que muestra el proceso administrativo que cumple la empresa Publisoft para llevar a cabo el desarrollo de sus productos de software en unión con los procesos que se le anexaron para el cumplimiento de los requisitos del nivel 2 de CMM; todas aquellas actividades que no están subrayadas son las que fueron tomadas de cada una de las fases de la metodología de trabajo de Publisoft (obsérvese en el caso de estudio), y las actividades que están subrayadas son las que esta investigación sugiere para alcanzar y completar el nivel 2 de CMM, las cuales se observan en la Fig.1 y se explican a continuación:

Dentro de la fase de Recepción de información, las actividades son las siguientes:

(1) Determinación del porcentaje de cambio que soportará el proyecto: antes de firmar el contrato se debe estipular el porcentaje de cambio que puede llegar a modificarse el sistema; esto ayuda a la empresa a resolver el problema respecto a que el cliente no sobrepase un número considerado de cambios sobre su respectivo software.

(2) Realización de contrato: quedando determinadas cada una de las estipulaciones, así como otros factores (tiempo, costo y porcentaje de cambio), se realiza el cierre de contrato; esto deja estipulado por escrito toda la información requerida por el sistema, además de un dato muy importante como lo es el grado de cambio, lo que nos evitará posibles problemas con el cliente.

Dentro de la fase de diseño:

(3) Reestructuración del diseño original: dentro de esta actividad se realizarán todos los posibles cambios dentro del diseño, es decir, cambios (tamaño, colores, formas, tipografías, etc.) sobre pantallas. Esta actividad se debe hacer dentro de esta fase, ya que esto nos resolverá que se haga un cambio hasta el momento de hacer las pruebas finales, lo que nos haría retroceder demasiado en el desarrollo del sistema hasta la fase de diseño, causando gran pérdida de tiempo y desgaste de recursos.

(4) Diseño terminado y aprobado: se procederá a la realización de un documento para la aprobación que será firmando por ambas partes.

Dentro de la Fase de Pruebas de aprobación y lanzamiento:

(5) Determinación de posibles cambios del sistema (sólo si se requiere): al momento de hacer las pruebas directamente con el cliente, se determinarán las posibles modificaciones al sistema, determinando si el grado de cambio lo alcanza a cubrir o no.

(6) Determinación de un nuevo contrato: en caso de que las modificaciones sobrepasen el porcentaje que se había determinado, se decidirá si se hace un nuevo contrato para cubrir la parte de las modificaciones que no fueron cubiertas por el anterior.

Este diagrama se propone para empresas que se dediquen al desarrollo de software ya que es muy posible que siguiendo las actividades propuestas en el diagrama consigan obtener el nivel 2 de CMM.

## 5 Conclusiones

En este trabajo se detecto que la empresa Publisoft de Aguascalientes puede alcanzar el nivel 2 (repetible) del esquema de CMM, debido a que la actividad no alcanzada de seguir los estándares del proyecto y controlar los requerimientos y productos de trabajo, sólo es cuestión de una mejor administración de los recursos humanos y de una mejor planeación de tiempo, tomando en cuenta posibles alteraciones que puede sufrir el sistema. Del problema en la administración de recursos materiales y humanos, se puede determinar con un alto nivel de confianza, que la empresa logrará una metodología global que integre tanto al área de desarrollo de software (software de escritorio y páginas web), así como los recursos necesarios y la forma de trabajar con los clientes. Lo que busca la organización es una perfecta armonía para poder solucionar los problemas de sus clientes a través de un software de calidad y una correcta administración de sus áreas funcionales, quedando esta solución pendiente de llevarse acabo dentro de la empresa.

Para poder observar los resultados de la obtención del nivel 2 de CMM, se realizará un futuro estudio, cubriendo todos los puntos positivos y negativos que obtuvo la empresa en el desarrollo de sus sistemas al implantar las estrategias propuestas.

## Referencias

1. Kim Caputo, CMM Implementation Guide, Addison Wesley (1997).
2. The Capability Maturity Model, Guidelines for Improving the Software Process, SEI Series in Software Engineering, Addison Wesley (1995).
3. Capability Maturity Model Integration CMMI, V1.1, Software Engineering Institute, Carnegie Mellon, (2002), [www.sei.cmu.edu](http://www.sei.cmu.edu), (2004).
4. H. Oktaba. y G. Ibargüengoitia González, Software processes Modeled with Objects: Static View, Computación y Sistemas, Iberoamerican Journal of Computer Science, CIC-IPN, México, 1, 4, p.228-238 ,(1998).
5. Michael Cimet, Administración y Desarrollo de Aplicaciones, Experiencia y Lecciones Aprendidas en EDS de México, (2001)
6. Desarrollo de software, una oportunidad para México, amiti.org.mx, (2004).

# Administración Estratégica del Personal en base al People-CMM para la Mejora de la Productividad del Software

Teresa Cervantes López y Jaime Muñoz Arteaga

Dpto. de Sistemas de Información de la Universidad Autónoma de Aguascalientes  
Av. Universidad # 940, Frac. Bosques del Prado, C.P. 20100, Aguascalientes, México  
[terecervantes@prodigy.net.mx](mailto:terecervantes@prodigy.net.mx) y [jmunozar@correo.uaa.mx](mailto:jmunozar@correo.uaa.mx)

**Abstract:** Organizations are now facing a strong competence by positioning their products and services in the market, this requires of a workforce with the necessary capability which permit to achieve the best quality in relation to their competitors [3]. In Software development area this isn't an exception, in this area it is necessary experienced people and committed with the organizations people whose should be able to generate quality software in the appropriate moment that covers the necessities of the users' information completely. The present work proposes the People CMM model use as a guide in the best implementation practices to achieve the workforce strategy administration in order to improve the software productivity.

**Resumen:** Las empresas enfrentan una fuerte competencia por ubicar sus productos y servicios en el mercado, esto requiere de una mano de obra con la suficiente capacidad que permita lograr la mejor calidad en relación a los competidores [3]. En el área de desarrollo de software esta necesidad no es una excepción ya que se requiere de gente con experiencia y comprometida con las empresas que sea capaz de generar Software de calidad en el momento adecuado que cubra por completo las necesidades de información de los usuarios [3]. El presente trabajo propone el uso del Modelo People CMM como guía en la implementación de las mejores prácticas para lograr la administración estratégica de personal con el fin mejorar la productividad en el desarrollo de Software.

**Palabras Clave:** Administración estratégica del personal, modelo People CMM y Productividad en el desarrollo de software.

## 1. Introducción.

En la actualidad las empresas compiten en dos mercados uno por sus productos y servicios y otro por el talento requerido para producir o desempeñar estos. El éxito de una empresa en el mercado de sus productos y servicios está determinado por el éxito en el mercado del talento [3]. Debido a que el conocimiento requerido para fabricar productos y ofrecer servicios incrementa, la retención de empleados experimentados ha llegado a ser crítica para mejorar la productividad y tiempo de respuesta hacia el mercado.

En el área de desarrollo de Software la deficiencia es tan grande que las compañías están comenzando a ofrecer incentivos que antes fueron ofrecidos solo a los altos ejecutivos. Los líderes de las organizaciones se dan cuenta que su habilidad para competir está directamente relacionada a su habilidad para atraer, desarrollar, motivar, organizar y retener gente talentosa [3]. Las empresas han intentado aplicar diferentes técnicas en sus esfuerzos para lograr la administración estratégica del capital humano, pero muchas de ellas carecen de una guía ó plan para implementar dichas técnicas.

El presente trabajo se enfoca a la situación que se presenta de forma general en la administración del personal de desarrollo de SW en las empresas donde la retención de empleados es un problema que ha tenido consecuencias negativas en la efectividad y eficiencia necesarias para el logro de los objetivos y metas establecidas lo que representa pérdidas monetarias para las empresas por lo que se hace necesario un estudio de la problemática y posibles soluciones para solventar dichos problemas.

## 2. Problemática.

Un gran problema que presentan las empresas respecto a la administración de personal es que no se considera a esta como elemento estratégico, solo es considerada como elemento importante en el desarrollo de productos y prestación de servicios.

El modelo en la administración de personal que la mayoría de las empresas ha practicado por muchos años es un modelo donde todas las actividades relativas a la administración de personal se enfocan al cumplimiento de sus obligaciones guiadas por las políticas de la empresa y los requisitos legales establecidos por los organismos gubernamentales este modelo se ilustra en la figura 1.



**Fig.1** Modelo de la administración de personal tradicional.

Este modelo no permite a las empresas lograr una administración adecuada de las capacidades del personal ya que carece de un plan para el registro adecuado o inventario de las mismas lo que se traduce en falta de promoción y asignación de actividades de acuerdo a las capacidades del personal lo que a su vez repercute en la

falta de motivación e identificación de los empleados con la empresa que limita la productividad en el desarrollo de Software, causando además la fuga del personal.

### 3. Administración Estratégica del Personal en base al People-CMM.

Para mejorar la productividad actual en el área de sistemas de las empresas en cuanto a las actividades de análisis, diseño, desarrollo mantenimiento e Implantación de sistemas, el presente trabajo propone hacer un cambio en el enfoque de la administración de personal de tal forma que se considere al personal como elemento clave para el éxito en el cumplimiento de metas y objetivos del departamento de sistemas y por consiguiente de toda la empresa.

El presente trabajo propone la implementación de People CMM® (People Capability Maturity Model). Este modelo es una herramienta que ayuda a dirigir exitosamente los aspectos críticos del personal dentro de la organización. People CMM guía a las organizaciones en la mejora evolutiva de sus procesos para administrar y desarrollar el personal [1]. Ayuda a las organizaciones a evaluar la madurez de sus prácticas respecto al personal, establece un programa de desarrollo continuo de personal, establece prioridades para acciones de mejora, integra el desarrollo del personal con el proceso de mejora y establece una cultura de excelencia [1]. Su filosofía resalta la capacidad del personal la cual se relaciona estrechamente con el desempeño y competitividad de las empresas y que debe estar definida en relación a los objetivos de la organización [1].

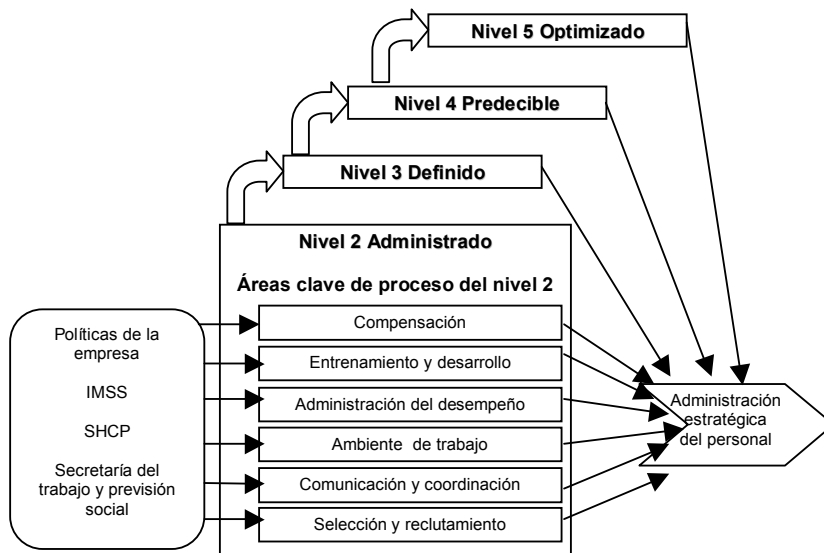


Fig. 2 Modelo de administración estratégica de personal utilizando People CMM.

El Modelo de People CMM está dividido en 5 estados o niveles que permiten remover por fases los impedimentos para la mejora continua. Cada nivel progresivo produce una transformación única en la cultura de la organización que le permite atraer, desarrollar, organizar, y retener la mano de obra [1].

El modelo propuesto para la administración estratégica de personal utilizando People CMM se ilustra en la figura 2.

A través de la implementación de People CMM se establecen procedimientos formales para la ejecución de las áreas claves de proceso en cada uno de los niveles del modelo que se describen en la figura 2. Esto permite lograr una mejora en la administración del personal de acuerdo al nivel como se describe a continuación:

*Nivel 2:* En este nivel, el hecho de establecer procedimientos formales permite a los administradores tomar la responsabilidad para administrar y desarrollar las capacidades del personal de la mejor manera. *Nivel 3:* Con este nivel se logra el desarrollo de las capacidades de los grupos de trabajo y alineamiento con los objetivos estratégicos del negocio. *Nivel 4:* Este nivel permite la integración y delegación de autoridad a los grupos de trabajo, y administración del desempeño cuantitativamente. *Nivel 5:* Con este nivel se llegará a la mejora continua y al alineamiento del personal, grupos de trabajo y capacidad organizacional [1].

En el modelo podemos observar que las actividades realizadas en la administración tradicional (figura 1) son las mismas que las áreas de proceso del nivel 2 (figura 2). La diferencia consiste en que las actividades en el modelo tradicional son guiadas como ya se comentó solo por las políticas de la empresa y los estatutos de gobierno, podría decirse que todas las empresas cumplen con estas medidas por lo que según el modelo People CMM todas ellas están en el nivel 1 que se denomina como inconsistente, por lo tanto para llegar al nivel 2 es necesario que todas esas actividades sean guiadas por el modelo People CMM donde estas actividades son consideradas áreas clave de proceso que se componen por las mejores prácticas lo que permitirá llegar a la administración estratégica del personal. Y una vez alcanzado el nivel 2 será más sencillo alcanzar niveles superiores de madurez en los cuales la administración estratégica de personal podrá mejorar constantemente.

## **4. Caso de estudio.**

### **4.1 Análisis y descripción del caso de estudio**

Como método para dar a conocer las ventajas que proporciona la implementación de People CMM se analizó la problemática presentada dentro del área de sistemas de un empresa local a manera de caso de estudio.

Analizaremos la situación de la empresa Big Pollo, S.A. de C.V. empresa del ramo alimenticio dedicada a la producción y venta de pollo. La situación a analizar comienza a partir de septiembre del año de 1998 cuando el área de sistemas toma un nuevo rumbo una vez asignado el nuevo gerente de sistemas. En este tiempo la empresa contaba con diferentes sistemas en cada una de las áreas desarrollados en diferentes plataformas y que funcionaban de manera aislada lo cual hacía difícil la comunicación e intercambio de información entre las diferentes áreas.

Para hacer frente a los nuevos retos de la empresa era necesaria una reestructuración o reingeniería de sistemas que permitiera la integración de todos los sistemas en un sistema de administración integral en el cual se incorporarían los sistemas de contabilidad, presupuestos, compras, ventas, facturación, almacén de materias primas, producción, almacén de ventas, nóminas y administración de personal; a este sistema integral se le dio el nombre de sistema 2000 ya que se pretendía terminar todos los sistemas antes de iniciar el año 2000 lo que permitiría también hacer frente al error del año 2000 con el que contaban varios de los sistemas que estaban operando.

Para soportar la integración de toda la información de la empresa de los diferentes módulos fue necesario un motor de base de datos robusto por lo que se compró INFORMIX, tanto el motor de base de datos como el lenguaje de desarrollo 4gl y por otra parte se compró Delphi para los módulos donde se requería la interfase gráfica. Todo esto requirió una gran inversión que los directivos de la empresa vieron con buenos ojos ya que esto les permitiría ser dueños completos de la información y tener el control del desarrollo que se adecuaría por completo a los procesos y necesidades de la empresa lo que ellos visualizaron como una estrategia competitiva a futuro, por lo que también se autorizó la contratación de personal para el desarrollo de este grande sistema ya que en ese tiempo el departamento de sistemas contaba solo con 6 personas quienes solo se dedicaban a dar mantenimiento a los sistemas existentes y servicio de soporte a usuarios.

Para el mes de febrero del año 1999 se contaba ya con un equipo de desarrollo de 20 personas quienes se organizaron de acuerdo al nivel de dificultad del módulo, cada modulo contaba con un líder de proyecto y una o dos personas de apoyo como analistas de sistemas. A todo el personal se le capacitó en todos los temas relacionados al desarrollo de sistemas: Cursos de Informix (Lenguaje 4gl, Conceptos generales y Administración de bases de datos y SQL para informix); Estándares de documentación y programación y capacitación en cuanto a procesos y estructura de la empresa. Todas estas actividades requirieron de un gran esfuerzo tanto en tiempo de los desarrolladores, usuarios y directivos involucrados como de recursos financieros por parte de la empresa, pero se perdieron de vista varios aspectos importantes en cuanto al personal como: La asignación de los empleados de sistemas a los módulos se realizó sin un análisis previo de sus capacidades, conocimientos y aptitudes; No se tomo en cuenta la importancia de retener a los empleados experimentados y capacitados; No se establecieron procedimientos para promover el desarrollo y crecimiento de los empleados en otras áreas de interés para ellos dentro de la empresa.



Los problemas anteriores parecen ser simples pero repercutieron en aspectos como: La falta de motivación de los empleados; La falta de identificación con la empresa; Se presentó un alto grado de renuncias; La reinversión en tiempo y recursos financieros para la capacitación; Retrasos en el tiempo de entrega de los sistemas, lo que repercutió en la productividad de la empresa.

Dada la problemática expuesta anteriormente y en base al análisis del caso se llegó a la conclusión de que el problema principal consiste en que no se visualizó al personal como un elemento clave para el éxito en el desarrollo del sistema 2000, se hace necesario un cambio de visión y una forma de hacerlo es tomando en cuenta las características del modelo People CMM. Consideramos que la implementación de los procedimientos formales para la administración de personal del nivel 1 (que se describió en la sección anterior) permitirá a esta organización mejorar en los siguientes aspectos: Las actividades asignadas a los empleados estarán más enfocadas a sus capacidades y aptitudes lo que permitirá una mayor productividad de los empleados; El logro de una mayor productividad se traducirá en adecuados tiempos de entrega y satisfacción a las necesidades de información de los usuarios de sistemas lo que traerá beneficios económicos a la empresa; Estos beneficios económicos de la empresa se verán reflejados en los trabajadores quienes recibirán también beneficios económicos y por lo tanto esto será una motivación más para ellos; Los trabajadores tendrán un mejor desempeño profesional lo que permitirá lograr una grande motivación para los empleados al ver que sus capacidades pueden ser promovidas y desarrolladas dentro de la empresa. La motivación económica y profesional alcanzada se traduce en una mayor productividad a nivel general en las empresas.

#### 4.2 Estimación de beneficios a alcanzar implementando People CMM

Existen ya muchas empresas que han implementado de manera exitosa el modelo People CMM, tal es el caso de empresas como Intel [7] y Cognizant [8] donde este modelo les ha permitido mejorar en gran medida la administración estratégica del personal.

En base en la experiencia de estas empresas se estima que la empresa Big Pollo puede alcanzar mejoras importantes en la eficiencia de la administración estratégica del personal con la implementación e Institucionalización de los procedimientos formales para la ejecución de las áreas claves de proceso del nivel 2 del Modelo People CMM . Cada una de estas áreas clave permitirá obtener una mejora en los porcentajes que se describen a continuación para cada una de las áreas clave de proceso.

**Tabla 1** Eficiencia a alcanzar por cada área clave de proceso

Áreas clave de proceso del Nivel 2	% de Eficiencia a alcanzar	Aspectos relevantes que permiten mejorar la eficiencia
Compensación	75%	Los procedimientos formales de esta área clave de proceso permiten dar a los individuos la remuneración y beneficios basados en su

		contribución y valor para la organización.
Entrenamiento y desarrollo	80%	Permite asegurar que todos los individuos tienen las habilidades para realizar sus tareas y proporcionar oportunidades de desarrollo.
Administración del desempeño	85%	Se establecen objetivos relacionados a la realización del trabajo contra las unidades individuales de desempeño que pueden ser medidas para comparar el desempeño contra los objetivos y mejorar continuamente el desempeño.
Ambiente de trabajo	80 %	Permite establecer y mantener condiciones adecuadas en el ambiente de trabajo y proporcionar los recursos que permitan a los individuos y grupos de trabajo realizar sus actividades eficientemente.
Comunicación y coordinación	80%	Se establece una comunicación oportuna a través de la organización y se asegura que los empleados tengan las habilidades para compartir información y coordinar sus actividades eficientemente.
Selección y reclutamiento	85%	Se cuenta con un proceso formal por el cual las actividades de un trabajo se relacionan a las unidades de recursos, lo que permite calificar, reclutar, seleccionar y asignar al personal una actividad específica de manera más eficiente.

Para alcanzar niveles óptimos en la eficiencia de la administración estratégica del personal es necesario implementar los niveles superiores del modelo People CMM.

## 5. Conclusiones y perspectivas.

El presente trabajo representa un análisis cualitativo de los beneficios que proporciona la implementación del modelo People CMM® (People Capability Maturity Model) para lograr la administración estratégica del personal y con esto una mayor productividad en el desarrollo de software. A pesar de que este estudio se enfocó a la productividad en el desarrollo de Software puede ser aplicado a cualquier área de la empresa donde sea necesaria la administración estratégica del personal. Es importante resaltar que es necesario un estudio cuantitativo de los beneficios que puede proporcionar la implementación de People CMM en la empresa en cuanto a costos y beneficios económicos. Así mismo es necesario evaluar la efectividad en la implementación de People CMM en un estudio cuantitativo exploratorio que permita a través de encuestas a empresas que han implementado People CMM evaluar el impacto y el nivel de satisfacción ofrecido para alcanzar la administración estratégica de personal.

## Referencias.

- [1] Bill Curtis, William E. Hefley, Sally A. Miller. "People Capability Maturity Model® (P-CMM®) Version 2.0.", July 2001.
- [2] Mark C. Paulk, Bill Curtis, Mary Beth Chrissis, Charles V. Weber. "Capability Maturity Model SM for Software Versión 1.1.", February 1993.
- [3] Mark C. Paulk, "Practices of High Maturity Organizations," The 11th Software Engineering Process Group (SEPG) Conference, Atlanta, Georgia, 8-11 March 1999.
- [5] Mark C. Paulk, "Using the Software CMM in Small Organizations," The Joint 1998 Proceedings of the Pacific Northwest Software Quality Conference and the Eighth International Conference on Software Quality, Portland, Oregon, 13-14 October 1998, pp. 350-361.
- [6] Mark C. Paulk, "Structured Approaches to Managing Change," Crosstalk: The Journal of Defense Software Engineering, Vol. 12, No. 11, November 1999, pp. 4-7.
- [7] Intel Information Technology. White Paper, "People Capability Maturity Model", February 2003. (<http://itpapers.techrepublic.com>)
- [8] Cognizant Technology Solution, [www.cognizant.com](http://www.cognizant.com), access date June 20, 2004

# Las Revisiones Técnicas Formales de los requerimientos y el análisis utilizando Áncora y el Proceso Unificado de Desarrollo de Software

Juan Manuel Fernández Peña<sup>1</sup>, María de Lourdes Hernández Rodríguez<sup>2</sup>, María de los Ángeles Sumano López<sup>1</sup>

<sup>1</sup>Especialización en Ingeniería de Software, Universidad Veracruzana, Av. Xalapa s/n,  
<sup>2</sup>LANIA, Av. Rebsamen # 80, Xalapa, Veracruz  
{jfernand@uv.mx, mhernand@lania.mx, asumano@uv.mx}

## Resumen

El Proceso Unificado de Desarrollo de Software involucra una serie de hitos en cada iteración y cada fase, mismos que deben ser valorados antes de continuar con las siguientes iteraciones y las diferentes fases. Por su parte Áncora tiene varias etapas en las que se crean diferentes hitos. La forma de evaluar el contenido de estos hitos es mediante el análisis de diferentes artefactos involucrados. El cómo, cuándo y qué revisar es el tema de este escrito. Las Revisiones de Avance y Revisiones Técnicas Formales se proponen como el punto de arranque para la evaluación; en ellas se han incluido los parámetros de calidad que deben alcanzarse, los artefactos involucrados y su forma de verificación. Todo este trabajo ha sido aplicado a varios sistemas de las empresas, públicas y privadas, a las que pertenecen los alumnos que llevan el posgrado de Especialización en Ingeniería de Software de la Universidad Veracruzana.

## Abstract

The Unified Software Development Process involves several milestones by iteration and by phase; they must be assessed before to proceed with the next. On the other hand, Áncora has several stages with several milestones. To assess the content of each milestone, all related artifacts must be analyzed. How, when and what analyze is the matter of this work. Progress Reviews and Formal Technical Reviews are proposed as starting point for assessment; in them are included quality parameters to reach, artifacts involved and verification procedure. The whole work has been applied to several systems from government and private enterprises, where Software Engineering Specialty's students work.

**Palabras clave:** Áncora, PUDS, RTF, calidad, artefacto,

## Introducción

Durante años se han sugerido una serie de pautas para la Revisión Técnica Formal (RTF) de avances dentro de los Sistemas de Software (SS), pero son pocas las veces que se especifica qué revisar, cómo y sobre todo cuáles parámetros de calidad se están asociando. Además, una RTF requiere un tiempo y un esfuerzo considerables, el llevarla a cabo da la impresión de que atrasa el trabajo de desarrollo, lo cual provoca que se vaya posponiendo indefinidamente. En este trabajo, se presentan dos tipos de revisiones: las de avance y las técnicas formales. Para mejor entendimiento de este trabajo, se divide de la siguiente manera: elementos metodológicos utilizados, propuesta y puesta en marcha de las revisiones y resultados obtenidos.

## 1. Elementos metodológicos empleados

Para llegar a establecer las propuestas aquí hechas se ha tenido que hacer una revisión de varios elementos metodológicos y de los artefactos involucrados en ellos. Aquí se presentan los más relevantes.

### 1.1 Revisión Técnica Formal

Una RTF es "una actividad de la garantía de calidad del software llevada a cabo por los ingenieros del software" [2]. Lo que se busca en una RTF varía según sus metas y los diferentes autores. En el caso de las revisiones de calidad "se refiere más a la validación de la interacción entre componentes y de conformidad con los requerimientos del usuario" [3].

### 1.2 Ciclo de vida de PUDS

PUDS [1] tiene un modelo de desarrollo especial de dos dimensiones. Una dimensión es la *del tiempo*, sobre el eje X, en donde se notan las fases y las iteraciones. En el eje Y se tienen *los flujos* de trabajo que se consideran en otros modelos de desarrollo y que en PUDS deben aplicarse para cada fase en grado mayor o menor como lo marcan los avances en cada fase (ver Tabla 1), lo cual permite obtener un producto en el que se revisan los avances por iteración y por fase, causando un mayor número de revisiones respecto a otros ciclos de vida.

### 1.3 Artefactos de Áncora y PUDS

Se han identificado 29 artefactos involucrados en los flujos de requerimientos y análisis de Áncora y PUDS. A continuación se explican los más relevantes, que son evaluados en las RTF, indicando su conexión con otros, ya sea del mismo flujo o de otros.

#### Artefactos de requerimientos, Explicación y Conexiones

1. Lista de los principales problemas. Resumen de los principales problemas separándolos en dos grupos: los que se pueden resolver utilizando computadora y los que no.
2. Guión de la Propuesta computacional. Breve introducción del contenido del documento y los guiones y diálogos propuestos. Cada pista será un paquete de análisis, cada escena un Caso de Uso y cada papel un Actor.
3. Diálogos de la propuesta computacional. Lista de acciones, expresadas en quintetas, aclarando alguna(s) quinteta(s) del guión de la propuesta computacional.
4. Prototipo Rápido. Manual Preliminar de operación del nuevo sistema.
5. Modelo E-R o Modelo Semántico. Modelo de Datos escogido, dependerá del SMBD que se piense usar. Identificadores de Puntos de Función
6. Bitácora de desarrollo. Plan de prueba y tiempo desarrollo de cada funcionalidad. Base para el Plan de Construcciones en implementación.
7. Tabla de Puntos de función sin ajustar. Resumen de los identificadores del método de Puntos de Función. Puntos para cada Caso de Uso

8. Tabla de modificadores. Con 14 modificadores del método de Puntos de Función y su grado de influencia en el sistema. Se transformarán en requerimientos no funcionales de cada Caso de Uso.
9. Modelo de casos de uso. Contiene actores, casos de uso y sus relaciones.
10. Caso de Uso. Una descripción de un conjunto de secuencias de acciones, incluyendo variaciones, que un sistema lleva a cabo y que conduce a un resultado observable de interés para un actor determinado.
11. Actor. Conjunto coherente de roles que desempeñan los usuarios de casos de uso. Iniciador de los diagramas de colaboración y de secuencia

#### **Artefactos de Análisis**

1. Modelo de análisis. Es una jerarquía de paquetes de análisis
2. Clase de análisis. Tres Estereotipos: boundary, control, entity
3. Realización de casos de uso del análisis. Colaboración dentro del modelo de análisis que describe cómo se lleva a cabo un caso de uso utilizando clases de análisis. Compuesto por: Diagramas de clases de análisis, Diagramas de interacción (colaboración), Flujos de sucesos - análisis, Requisitos especiales.
4. Descripción de la arquitectura. Contiene la Vista de la Arquitectura del Modelo de Análisis. Abarca: Clases, Paquetes, Realizaciones de casos de uso

## **2. Revisiones de Software en la EIS**

El uso del ciclo de vida marcado por PUDS marca, naturalmente, dos tipos de revisión la de la iteración o Revisión de Avances, marcada por Casos de Uso, y la de fase o Revisión Técnica Formal. Lo que PUDS deja abierto es ¿qué se debe revisar y en qué profundidad? Por lo anterior, en este trabajo se muestra la adaptación de estas revisiones en cursos de un programa de posgrado en el que a lo largo de un año y con sistemas de 100 a 300 puntos de función se ha logrado obtener.

### **2.1 Revisiones de Avances (RA)**

Éstas se llevan a cabo cada dos meses y se espera que lleven un cierto porcentaje de avance sobre los diferentes flujos fundamentales (ver Tabla 1). En cada RA se exploran diferentes artefactos según la fase en que se encuentre SS. En la Tablas 2 se muestran los artefactos revisados en dos RA, una transcurridos seis meses, en la que se debe llevar un 25% de avance y, la segunda donde se debe llevar un 50% de avance, después de ocho. La escala de calificación es de 1 = Bien, 0.5 = Regular y 0 = Mal.

<b>Fase</b>	<b>Requerimientos</b>	<b>Análisis</b>	<b>Diseño</b>	<b>Implementación</b>	<b>Prueba</b>
Inicio	40%	20%	10%	10%	10%
Elaboración	100%	80%	50%	50%	50%
Construcción		100%	80%	80%	80%
Terminación			100%	100%	100%

**Tabla 1. Porcentaje de avances en cada flujo fundamental .**

<b>Avances después de 6 meses</b>	<b>Avances después de 8 meses</b>
Distribución por Caso de Uso de Puntos de Función del Sistema	Tabla de Casos de Uso y Puntos de Función
Modelo de Casos de Uso	Modelo de capas del diseño <b>corregido</b> .
Diagramas de Colaboración y flujos de sucesos	Tabla de tiempos dedicados a cada flujo fundamental.
Restricciones para cada Caso de Uso	Diagrama de Secuencia exitoso de, al menos, un nuevo Caso de Uso
Modelo de capas del diseño	Diagramas de componentes de los subsistemas.
Diagramas de Secuencia exitosos	Casos de Prueba del Caso de Uso presentado
Ejecución del 25% Casos de Uso	Instalador automático
	Corrida del 50% de su sistema

**Tabla 2. Elementos para Revisión de Avances**

## 2.2 RTF en la EIS

Para realizar las RTF al concluirse cada flujo fundamental, se desarrollaron guías que permitan verificar los artefactos más relevantes de manera integrada. Esta revisión debe responder a una serie de atributos de calidad que deben observarse en los artefactos individuales (claridad, conformidad a reglas) y al encadenamiento de algunos de ellos (corrección, completez, trazabilidad). Del análisis de artefactos y parámetros de calidad, se realizaron los formatos para las RTF de requerimientos y la de análisis. Por falta de espacio, se muestran aquí sólo algunas de las Tablas (3 y 4) aplicadas tanto a la captura de requerimientos como al análisis en las que cada renglón se califica en la escala siguiente: 5: todas(os); 4: casi todos(as); 3: aproximadamente la mitad; 2: casi ninguno(a); 1: ninguno(a).

Además de los aspectos revisados con las tablas, se le pide al grupo de revisores que anote los tres aspectos más relevantes identificados, ordenándolos según la urgencia con que deben ser atendidos.

<b>Atributo observado</b>	<b>Concepto</b>
correcto	¿Cada caso de uso corresponde a una escena del guión de la propuesta computacional?
completo	¿Cada escena del guión de la propuesta computacional corresponde a un CU?
correcto	¿Cada actor corresponde a un papel?
completo	¿Cada papel corresponde a un actor?
completo	¿Cada requerimiento expresado en el guión y en tabla de modificadores de puntos de función, está asociado al menos con un caso de uso?
consistente	¿Los requerimientos no tienen conflictos entre sí?

**Tabla 3. RTF de requerimientos empleando: Modelo de casos de uso, guión y diálogos de propuesta computacional y modificadores de puntos de función**

<b>Atributo observado</b>	<b>Concepto</b>
Conforme	Cada caso de uso tiene un diagrama de clases de análisis
Conforme	Cada caso de uso tiene un diagrama de colaboración
Conforme	Cada diagrama de colaboración tiene un flujo de sucesos asociado
Completo	Cada entidad y relación m a n en el modelo E-R existe como clase en algún diagrama de clases
Conforme	Cada clase en los diagramas de clases tiene asignado un estereotipo (boundary, control o entidad)
Completo	Para cada clase en los diagramas de clase existe al menos un objeto en los diagramas de colaboración
Conforme	En los diagramas de colaboración existe al menos una colaboración etiquetada para cada objeto
Conforme	En los diagramas de colaboración todas las colaboraciones son iniciadas por actor, boundary o control
Conforme	En cada diagrama de colaboración el actor sólo se comunica con clases boundary, nunca con control o entidad
Conforme	La etiqueta de cada colaboración expresa un propósito preciso
Completo	Para cada diagrama de colaboración existe un flujo de sucesos
Conforme	Cada flujo de sucesos aclara el sentido de las colaboraciones asociadas
Trazable	Cada flujo de sucesos de análisis equivale al flujo de sucesos del caso de uso en el modelo de casos de uso

**Tabla 4. RTF en el Análisis empleando realización de casos de uso y Modelo Entidad Relación (o Modelo semántico)**

### 3. Resultados

Después de haber aplicado realizado las Revisiones de Avance (RA) y Revisiones Técnicas Formales (RTF), se tiene lo siguiente:

- a) Las RA permiten cuantificar de manera más objetiva el grado de avance de un proyecto. Sin embargo, se observa que muchas veces los desarrolladores se apresuran a cumplir avances de código sacrificando lo demás, en especial lo relativo a pruebas.
- b) Las guías de RTF ofrecen una herramienta que:
  - Identifica aspectos que deben cuidarse al desarrollar los modelos, asociados con atributos de calidad como corrección, conformidad y trazabilidad. Sin la guía, muchos aspectos tienden a olvidarse.
  - Permite una autoevaluación y control del proyecto.
  - Permite que otras personas (jefe, colegas) revisen proyectos ajenos de manera coherente, con criterios semejantes.
  - Saca a la luz aspectos que deben mejorarse en los modelos y en las prácticas de trabajo
- c) En la aplicación de las RTF se observó que:



- Los evaluadores requieren más sentido crítico: algunos suavizan calificación; pocos aportan sugerencias.
- Algunas preguntas resultaron oscuras, por su redacción o escala inversa; ya se modificaron.
- Algunos evaluadores agregaron aspectos no considerados, como la redacción.
- La escala empleada parece adecuada, se usaron todos los valores y no se reportaron confusiones.
- Los valores bajos de la escala, cuando se repiten en varios casos y persisten en distintos flujos, identifican puntos débiles genéricos de la metodología y de los cursos; un ejemplo es el de los requerimientos no funcionales, mal especificados y sin seguimiento.
- Entre desarrolladores con experiencia en dependencias de gobierno aparecen dos errores frecuentes: insistencia de incluir un papel o actor “jefe” puramente decorativo y el confundir modelo de casos de uso con jerarquía de menús.

## Conclusiones

Las RA y RTF presentadas han permitido avanzar hacia un proceso de revisión de hitos maduro y ligado a parámetros de calidad preestablecidos, siendo un apoyo valioso, tanto para desarrolladores como para responsables de proyecto.

Cuando un desarrollador juega también el papel de revisor, le es más fácil identificar sus deficiencias viéndolas en trabajo ajeno, lo que redundará en una mejora de su propio trabajo. También aprecia y aprende de las aportaciones de otros desarrolladores.

Aún falta refinar las herramientas, buscando un equilibrio entre detalle y simplicidad. Deben ampliarse para cubrir artefactos que aún permanecen desligados y cubrir los flujos de trabajo restantes. Más adelante hará falta comparar los resultados finales para analizar el impacto duradero de las RTF. Por el lado humano hace falta desarrollar más el sentido crítico de los desarrolladores.

## Referencias

- [1] I. Jacobson, G. Booch,, J. Rumbaugh, *El Proceso Unificado de Desarrollo de Software*, Addison Wesley, 2000.
- [2] R. Pressman, *Ingeniería de Software. Un enfoque Práctico*, Quinta edición, McGraw-Hill, 2002
- [3] I. Sommerville, *Software Engineering*, Quinta edición, Addison-Wesley, 1996
- [4] M. A. Sumano, *Método para el análisis de requerimientos de software con enfoque conjunto psicológico, social y lingüístico conducente al reuso*, Tesis doctoral. CIC, IPN. México, 2002

# Aplicando extensiones de UML para desarrollo de aplicaciones Web

Alejandro Botello Castillo,  
Sandra Dinora Orantes Jiménez

Laboratorio de Sistemas de Información  
Centro de Investigación en Computación  
e-mail: {botello, dinora}@cic.ipn.mx  
Av. Juan de Dios Bátiz s/n, esq. Miguel Othón de Mendizábal, Col. Nueva Industrial  
Vallejo, 07738, México, D.F.

**Abstract:** The Unified Modeling Language (UML) had provided a consistent and detailed foundation in the planning of phases involved to build software applications. Nevertheless, an actual trend is forward Web applications, and specifically, in Web Services, which take and apply component based environments concepts. In this work, we show experiences in the use of UML extensions applied to Web Services, taken as case of study the JDeveloper 10g software development tool, from Oracle Corporation.

**Resumen:** El Lenguaje Unificado de Modelado UML (Unified Modeling Language) ha proporcionado una base consistente y detallada en la planeación de las fases involucradas para la construcción de aplicaciones de software. No obstante, una tendencia actual es hacia las aplicaciones Web, y en específico, en los Servicios Web, los cuales toman y aplican los conceptos de ambientes basados en componentes. En este trabajo se muestran las experiencias en el uso de extensiones UML aplicado a los Servicios Web, tomando como caso de estudio la herramienta de desarrollo JDeveloper 10g, de Oracle Corporation[1]

**Palabras clave:** Servicios Web, UML, modelado.

## 1. Introducción

No es posible negar que Internet se ha convertido en la tecnología más prominente con lo que respecta al intercambio de información entre usuarios. Las empresas están apoyándose en el Web para desarrollar aplicaciones de negocios (e-business), de búsqueda de información, y de entretenimiento. Una de las principales razones de uso de Internet es que un usuario no necesita altos recursos de cómputo para ejecutar dichas aplicaciones; sólo basta que tenga un software navegador para Web (que inclusive se encuentra en versiones para dispositivos móviles) y la dirección URL (Unified Resource Locator, Localizador de Recursos Unificado) de interés para que pueda comenzar a interactuar de manera remota.

La rápida evolución de la Web impulsa al desarrollo de nuevas tecnologías, nuevos estándares, y por consiguiente, de nuevas herramientas de desarrollo, ya que las casas fabricantes de software no desean perder mercado de venta de sus productos[2].

Por otra parte, las metodologías para el desarrollo de software también han evolucionado, tratando de seguir a corto paso a las tendencias que emergen continuamente. Roger Pressman[3] discutió si valía la pena emplear un proceso de Ingeniería en las aplicaciones basadas en Internet, o bien, qué características tenían éstas que justificaran no utilizarlo. “Aplicar un proceso de ingeniería nunca es una mala idea, pero éste debería adaptarse a los requerimientos de cambio continuo y rapidez siempre presentes, en el proceso de desarrollo Web”.

Un modelo de desarrollo de aplicaciones que ha sido aceptado ampliamente es el Lenguaje Unificado de Modelado (Unified Modeling Language, UML), por su capacidad de expresar de manera concisa las etapas a seguir en la construcción de sistemas de software. No obstante la popularidad de UML, no se han contemplado la inclusión de características para el desarrollo en Web, tanto así que ha surgido una nueva rama de la ingeniería de software denominada 'Ingeniería Web'[4], en la cual se tratan de cubrir los aspectos importantes de las aplicaciones enfocadas a la Web.

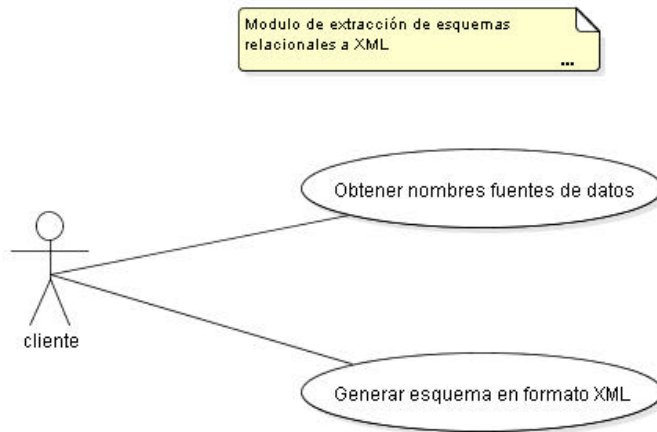
A continuación se presenta un ejemplo práctico en el que se aplicaron extensiones enfocadas a la creación de Servicios Web[5] usando UML.

## **2. Análisis**

En el marco de desarrollo de una aplicación integradora de esquemas de bases de datos relacionales, se ha desarrollado un servicio Web el cual establece una conexión a una fuente de datos, extrae sus metadatos (esto es, la definición y restricciones en la estructura de tablas y columnas de una base de datos) y los transforma a un documento siguiendo la especificación de XML Schema[6,7], con lo cual puede ser procesado por otro modulo integrador.

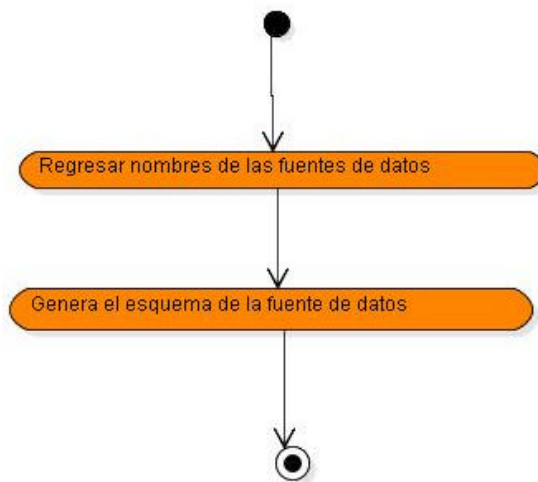
La herramienta de desarrollo JDeveloper, en su versión denominada 10g (la más reciente), ha incorporado muchas características para el desarrollo en Web, en especial enfocado a los servicios Web. En la parte de UML, se han incorporados los diagramas de casos de uso, de actividades, y de clases.

El caso de uso general del componente es simple, y se muestra en la Figura 1. Un modulo, proceso o usuario que requiere obtener el esquema de una fuente de datos, hace una solicitud al servicio Web para que bases de datos están definidas actualmente. Una vez que ha elegido alguna, solicita la generación del documento XML Schema correspondiente a esa fuente de datos.



**Fig.1** Diagrama de casos de uso para el módulo de extracción de esquemas.

El diagrama de actividades muestra la interacción entre el cliente y la invocación del servicio Web (Figura 2). Como se puede observar, la interacción es también simple, ya que lo único que necesita el cliente es llamar a los procesos que obtienen los nombres de las fuentes de datos y que generan el esquema correspondiente.



**Fig.2** Diagrama de actividades para el módulo de extracción de esquemas.

### 3. Implementación

La extracción de los metadatos de una fuente de datos se realiza mediante la invocación de la clase (escrita en el lenguaje de programación Java) MD2Schema, la cual se muestra en su diagrama de clases en la Figura 3.

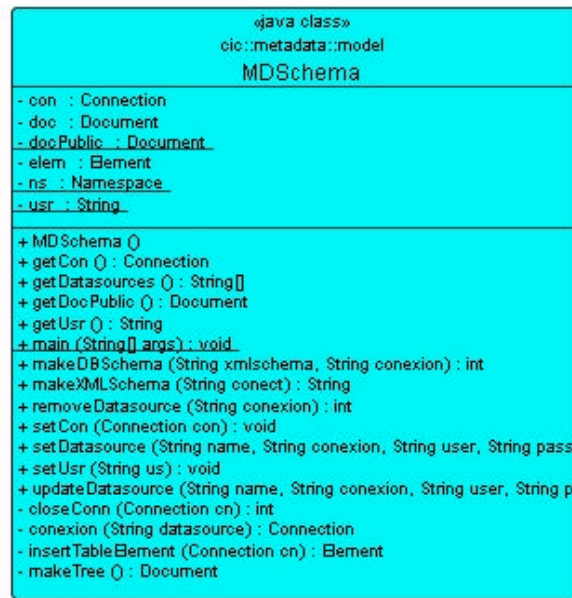


Fig. 3 Diagrama de clases para el módulo de extracción de esquemas.

Se observa de la figura anterior que la clase *MDSchema* consta de varios métodos, de los cuales *makeXMLSchema* y *makeDBSchema* son los que realizan la generación de esquemas, mientras que *getDatasources*, *setDatasource*, *updateDatasource* y *removeDatasource* se usan para la configuración de las fuentes de datos. Para generar el servicio Web es necesario indicarle a la herramienta que estos métodos conformarán sus operaciones; aquí es en donde se usan las extensiones UML de Servicios Web, añadiendo al diagrama una interfaz, que contiene sólo los métodos necesarios del servicio, y que automáticamente genera el código Java de dicha interfaz, así como el documento WSDL[8] (Web Services Description Language, Lenguaje de Descripción de Servicios Web), como se muestra en las Figuras 4 y 5.

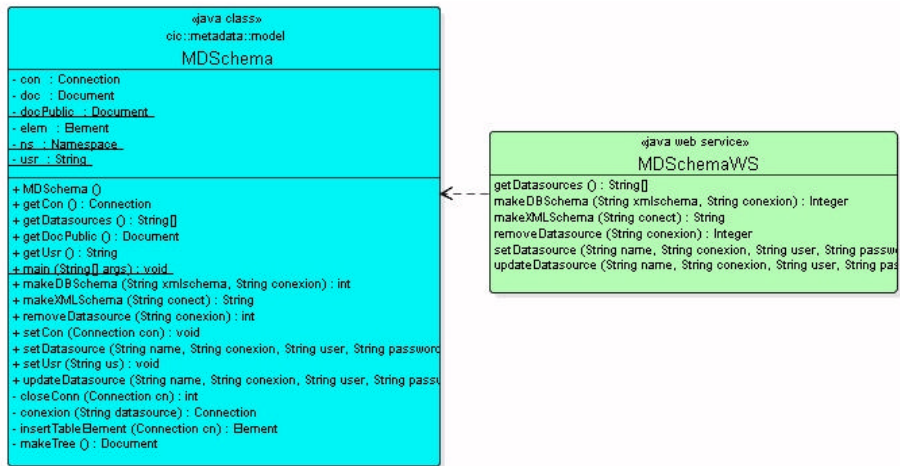


Fig. 4. Diagrama de clases para el módulo de extracción de esquemas, mostrando la extensión para Servicios web.

```

<?xml version="1.0" encoding="UTF-8" ?>
<!-- Generated by the Oracle Developer 10g Web Services WSDL Generator -->
<!-- Date Created: Thu Jul 13 21:24:44 CDT 2006 -->
<definitions name="MDSchemaWS" targetNamespace="http://cic.metadata.modelo/MDSchema.wsdl" xmlns="http://schemas.xmlsoap.org/wsdl/"
xmlns:cod="http://www.w3.org/2001/XMLSchema" xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
xmlns:tns="http://cic.metadata.modelo/MDSchema.wsdl" xmlns: xsi="http://www.w3.org/2001/XMLSchema" />
<types>
<message name="getDatabasesRequest" />
<message name="getDatabasesResponse" />
<part name="return" type="ns1:ArrayOfString" />
</message>
<message name="makeXMLSchemaRequest" />
<part name="connect" type="xsd:string" />
</message>
<message name="makeXMLSchemaResponse" />
<part name="return" type="xsd:string" />
</message>
<message name="setDataSourceRequest" />
<message name="setDataSourceResponse" />
<message name="makeDBSchemaRequest" />
<message name="makeDBSchemaResponse" />
<message name="removeDataSourceRequest" />
<message name="removeDataSourceResponse" />
<message name="updateDataSourceRequest" />
<message name="updateDataSourceResponse" />
<portType name="MDSchemaPortType" />
<binding name="MDSchemaBinding" type="tns:MDSchemaPortType"
soap:binding style="rpc" transport="http://schemas.xmlsoap.org/soap/http" />
<operation name="getDatabases" />
<soap:operation soapAction="" style="rpc" />
<input name="getDatabasesRequest" />
<output name="getDatabasesResponse" />
</operation>
<operation name="makeXMLSchema" />
<operation name="setDataSource" />
<operation name="makeDBSchema" />
<operation name="removeDataSource" />
<operation name="updateDataSource" />
</binding>
<service name="MDSchemaWS" />
<port name="MDSchemaPort" binding="tns:MDSchemaBinding"
soap:address location="http://localhost:8506/Metadata-Model-context-root/MDSchemaWS" />
</port>
</service>
</definitions>
  
```

Fig. 5. Documento WSDL generado de la definición del servicio web.

Para la invocación del servicio web generado se debe escribir la dirección donde se encuentre publicado en el navegador web. El resultado de la generación de un documento XML Schema para una base de datos de prueba se muestra en la Figura 6.

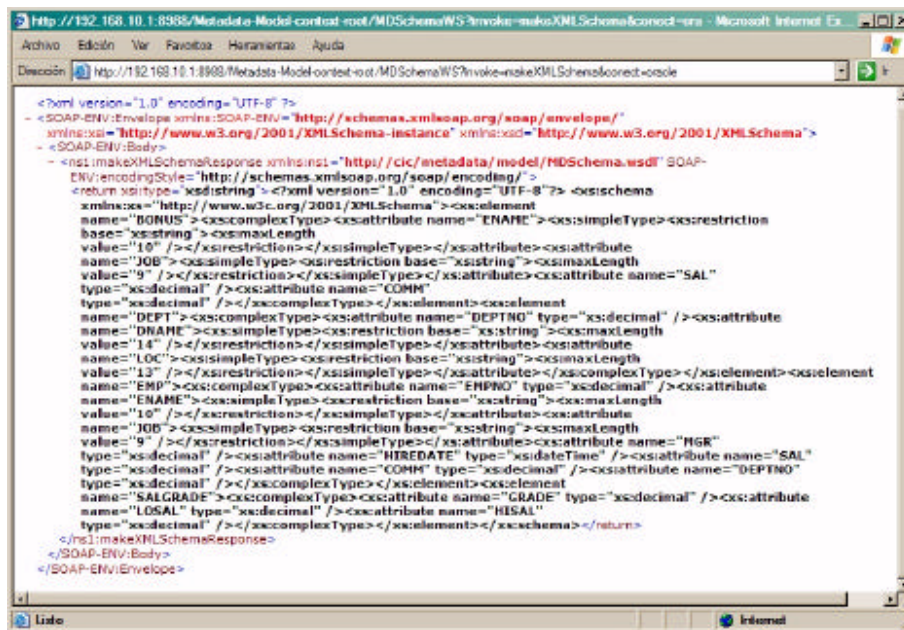


Fig. 6. Documento XML Schema obtenido de la solicitud al servicio Web.

Este documento es procesado posteriormente por un módulo que permite la integración de esquemas para base de datos, en una aplicación completa. La obtención de este documento ha sido de manera remota, a través de Internet, y sin requerir de instalaciones de software intermedio o adicional en la máquina cliente.

#### 4. Conclusiones.

Las herramientas de desarrollo de software van incorporando cada vez mayores características de nuevas tecnologías. Esta evolución ha dado pie a que los desarrolladores no tengan que depender de otras herramientas externas para realizar sus labores, además que las empresas estén incorporando estándares mundiales a sus productos, con lo que se obtiene reusabilidad de código y una mejor calidad en las etapas de desarrollo. Se espera que las nuevas tendencias, como los Servicios Web, se utilicen aún más en la construcción de software para la Web. Al momento de escribir este trabajo, se han evaluado tres herramientas para la generación de Servicios Web, siendo el JDeveloper 10g la única donde es posible extender las capacidades del UML para los Servicios Web.

## Referencias

- [1] Oracle Corporation, <http://www.oracle.com/>
- [2] Fournier, R. *Methodology for Client/Server and Web Application Development*. Yourdon Press, 1999.
- [3] R. Pressman, “*Can Internet-Based Applications Be Engineered?*” IEEE Software, September/October 1998, pag 104 – 110.
- [4] S. Murugesan, Y. Deshpande , S. Hansen, A. Ginige. “*Web Engineering : A New Discipline for Development of Web-Based Systems.*”, Lecture Notes in Computer Science 2016 Springer 2001, pag 3 – 13.
- [5] W3C Web Services, <http://www.w3.org/2002/ws/>
- [6] XML Schema Part 1: Structures, <http://www.w3.org/TR/xmlschema-1/>
- [7] XML Schema Part 2: Datatypes, <http://www.w3.org/TR/xmlschema-2/>
- [8] Web Services Description Language 1.1 Specification, <http://www.w3.org/TR/2001/NOTE-wsdl-20010315>
- [9] Pressman, R. “*Software Engineering: A Practitioner’s Approach. 5th ed.*” McGraw-Hill 2000.
- [10] Conallen, J, *Building Web Applications with UML*, Addison-Wesley, 1999.



# Elementos del Proceso Unificado de Desarrollo de Software aplicados al desarrollo del Sistema Integral de Gestión de Calidad y de Comercio Electrónico del Café-Veracruz (SIGC3).

María de Lourdes Hernández Rodríguez, Beatriz Gutiérrez Almaraz, Julio César Aguilar Cabrera

Calle Rebsámen Num.80 Col. Centro, CP 91000, Xalapa, Veracruz, México.  
{mhernand@lania.mx, bguiterr@lania.mx, jaguilar@lania.mx}

## Abstract

SIGC3 (Sistema Integral de Gestión de la Calidad y Comercio del Café-Veracruz) is an integrated system of applications: a manager for the processes of quality verification and certification, an e-business and commerce platform and applications for e-productivity and support and access control.

This article presents the application of some elements of Unified Software Development Process to the SIGC3 project, the results obtained and the perspectives.

## Resumen

El Sistema Integral de Gestión de Calidad y Comercio Electrónico del Café-Veracruz (SIGC3), consiste en un sistema integrador de aplicaciones: de administración de procesos en la verificación y certificación de calidad, de comercio y negocios electrónicos (e-bussiness), de aplicaciones de e-productividad (e-productivity) y de infraestructura y sistemas de acceso.

En este artículo se presenta la aplicación de algunos elementos del Proceso Unificado de Desarrollo de Software al SIGC3, los resultados obtenidos y las perspectivas.

Palabras clave: SIGC3, PUDS, UML, modelos.

## 1. Introducción

SIGC3 esta conformado por nueve módulos cada uno de los cuales tiene objetivos específicos que a su vez se suman a los objetivos generales y que en conjunto abarcan a todos los participantes de la cadena del valor del café (sector cafetalero). SIGC3 ganó en 2003 un concurso denominado ICT4BUS, del Banco Interamericano de Desarrollo (BID) y del fondo multilateral de inversiones (FOMIN), para la innovación en Tecnologías de la Información para pequeñas y medianas empresas. El proyecto fue

planteado para construirse en un periodo de 24 meses, financiado parcialmente por el BID.

Para su desarrollo se conjuntó un equipo altamente calificado en la programación de sistemas web, a quien se planteó el reto de generar un producto de calidad en un periodo de tiempo y presupuesto previamente establecido.

Una vez conformado el equipo de desarrollo se procedió a determinar la metodología aplicable al proceso de construcción del SIGC3. La metodología seleccionada fue el Proceso Unificado de Desarrollo de Software (PUDS)[1] que utiliza el Lenguaje Unificado de Modelado (Unified Modeling Language, UML[2]) para preparar todos los modelos de un sistema.

Los primeros aspectos a superar fueron:

- a) Convencer al equipo de las ventajas que representa la elaboración de modelos previos al enfoque de todos los esfuerzos sobre la fase de construcción.
- b) Elegir los elementos más relevantes de PUDS que de acuerdo a las características específicas de SIGC3 permitieran su aplicación.

Una vez superados estos aspectos se procedió a la aplicación de los elementos seleccionados descritos a continuación. Así como también se describirán los resultados obtenidos y finalmente las perspectivas.

## **2. Metodología**

SIGC3 es un sistema de integración de las funciones e interacciones de los diferentes grupos de usuarios o comunidades que participan en la cadena de valor del Café de calidad “Café-Veracruz”.

El proyecto SIGC3 se caracteriza por el hecho de que el grupo de usuarios a quien va dirigido no tiene necesidades concretas, a partir de las cuales se pueda realizar una recolección de requerimientos. Los diversos módulos del SIGC3 consisten en una propuesta de diversas herramientas tecnológicas para la apertura de canales y medios de comunicación de los diferentes grupos de usuarios. Por esta razón, algunos de los procesos sugeridos por PUDS para la fase de recopilación de requerimientos no son eficaces.

PUDS involucra muchos elementos que pueden ser implementados adaptándose a la naturaleza del proyecto, únicamente a nivel de artefactos pueden identificarse más de 100. Considerando las situaciones anteriores se inició una revisión de PUDS: su ciclo de vida, fases, versiones, iteraciones, flujos de trabajo, artefactos y roles. En nuestro

caso los aspectos que se consideraron para elegir los artefactos que se utilizarían fueron los siguientes:

### **2.1. Definición de una primera visión del sistema**

Este aspecto es el punto de partida del proceso de desarrollo, los objetivos previstos son: identificar y entender las principales necesidades del sector, representarlas y elaborar el modelo del dominio y del negocio que cubra dichas necesidades, además de contemplar elementos que pueden solventar un cambio a futuro en la funcionalidad del proyecto SIGC3.

El equipo de desarrollo se concentró en contestar a las preguntas [3]: ¿Cuáles son los términos clave? (glosario), ¿Cuál es el problema que se está tratando de resolver?, ¿Quiénes son los usuarios? ¿Cuáles son sus necesidades?, ¿Cuáles son las características del sistema?, ¿Cuáles son los requerimientos funcionales?, ¿Cuáles son los requerimientos no funcionales?, ¿Cuáles son las restricciones?

Los resultados de esta actividad son: el modelo de casos de uso, lista de requisitos genéricos no asociados a un CU específico, los actores, el prototipo de la interfaz de usuario y un glosario.

### **2.2. Definición de la arquitectura del sistema**

En este aspecto los objetivos que se pretenden alcanzar son: especificar detalladamente los requisitos (diagramas de clases de análisis y de colaboración), definir los subsistemas de análisis, identificar los elementos significativos de la arquitectura, detallar el funcionamiento de los Casos de Uso a través de clases de diseño y sus dependencias.

Las preguntas a resolver son: ¿Cuál es la mejor forma de implementar el proceso del negocio? ¿Cuáles son los elementos principales del proceso? ¿Cuáles pasos se pueden ahorrar para llegar a un resultado? ¿Cuál es el mecanismo establecido para que trabajen juntos? ¿Cuáles serán las entidades principales que debo contemplar y sus atributos? ¿Cómo construir la funcionalidad a implementar? ¿Qué tipo de elementos serían reutilizables? ¿En qué forma puedo agrupar las funcionalidades?

El proyecto SIGC3 contempla la funcionalidad de ser multiplataforma, por lo cual en la fase de diseño no es de vital importancia conocer el Sistema Operativo bajo el cual vivirá la aplicación ni la base de datos en donde se alojarán los datos de la misma.

Como resultado se obtienen el modelo de análisis, el modelo de diseño y el modelo de despliegue.

### **2.3. Generación y administración de un plan.**

A través de un plan general del sistema se esquematiza el proyecto, se establecen los recursos requeridos y se le da seguimiento a su progreso. Pueden establecerse diferentes tipos de planes en etapas distintas del proyecto como por ejemplo: plan de construcción, plan de pruebas, plan de aceptación del producto. En este caso lo importante no es elaborar el plan sino llevar a cabo su ejecución y control. Previendo posibles desviaciones y actuando oportunamente al respecto.

### **2.4. Elaboración y prueba del sistema**

Los objetivos en este aspecto son: implementar los elementos del modelo de diseño, generar un plan de integración y planificar, diseñar e implementar pruebas. Uno de los puntos principales es la generación incremental de código, conformando los componentes del sistema y verificando que se alcancen los requerimientos o funcionalidades previamente establecidas.

Los modelos obtenidos como resultado son: el modelo de implementación (componentes, subsistemas, plan de integración) y el modelo de pruebas (caso de prueba, componente de prueba, procedimiento de prueba, registro de defectos).

Las pruebas de rendimiento funcionan como uno de los elementos principales en esta categoría puesto que es de vital importancia conocer los límites de desempeño de los módulos del proyecto SIGC3. Los aspectos a considerar son: ¿Cuáles son los recursos necesarios? ¿Cuántas solicitudes puede procesar en cuánto tiempo?, si hay mucha demanda, ¿Cómo debe proceder? ¿Cuál es el mayor tiempo de respuesta?. Lo anteriormente descrito permite contemplar un panorama sobre el comportamiento del sistema y tomar acciones para que este sea óptimo.

Las pruebas de usabilidad de igual forma son una parte importante ya que el público objetivo a quien están dirigidos algunos de los módulos de SIGC3 cuenta con una formación informática mínima por lo que los aspectos: intuitivo en su uso, fácil de recordar y baja incidencia de errores cobran una especial relevancia.

### **2.5. Control y administración de cambios**

Una vez delimitados los alcances de la primera versión del sistema, se debe proceder a establecer claramente bajo que condiciones una solicitud de cambio o extensión puede ser aprobada. Puesto que se contempla el diseño del proyecto SIGC3 como un sistema flexible y multiplataforma, si un cambio no afecta significativamente el esquema bajo el cual se rige toda la arquitectura del sistema entonces tiene una alta probabilidad de ser aprobado.

## 2.6. Soporte a usuario final

El objetivo principal en este caso es generar la documentación de soporte necesaria para que el usuario pueda operar y mantener el sistema.

Adicionalmente debe generarse la documentación técnica y el repositorio de datos que permitan retomar el sistema y proponer la extensión de sus funcionalidades o la reutilización de los componentes de propósito general.

La elaboración de la documentación en las fases de análisis, diseño, implementación y pruebas le permitirán al usuario encargado de llevar el mantenimiento y administración de las diversas aplicaciones un mejor control de las mismas. Con la aplicación de la metodología PUDS el usuario que posea la documentación del proyecto será capaz de realizar los cambios que considere necesarios, así como corregir posibles defectos.

## 3. Resultados obtenidos

Los resultados obtenidos seis meses después de poner en operación los aspectos señalados en la sección anterior se presentan a continuación, en la Tabla 1 se puntualizan las actividades adicionales aplicadas a SIGC3 y en la Tabla 2 se presenta una lista de las ventajas o desventajas observadas.

**Tabla 1.** Actividades adicionales presentes en el desarrollo de SIGC3

<i>Artefacto al que se aplico</i>	<i>Actividades en SIGC3</i>
Modelo de negocio y de dominio Lista de características	<ul style="list-style-type: none"><li>Definición de un público objetivo y sus características</li><li>Análisis de opciones disponibles en el mercado con objetivos semejantes.</li></ul>
Modelo de casos de uso Modelo de análisis	<ul style="list-style-type: none"><li>Generación de un mapa de navegación</li><li>Identificación de dependencias entre diagramas (subdiagramas)</li></ul>
Modelo de diseño	<ul style="list-style-type: none"><li>Específicamente en el diagrama de capas de diseño, generalización de la capa de soporte</li></ul>
Modelo de implementación	<ul style="list-style-type: none"><li>Generación de diagrama de componentes a nivel general y específico</li><li>Evaluación, selección e integración de componentes middleware open-source.</li></ul>

**Tabla 2.** Ventajas / Desventajas observadas en la aplicación de PUDS

<b>Ventajas</b>
1. Identificación precisa de funcionalidades o alcances del sistema sobre el que se trabaja.
2. Lista “parcial” de componentes necesarios para cubrir las funcionalidades mencionadas.

3. Definición precisa del avance logrado a una fecha determinada
4. Registro detallado de costos y tiempos por actividad.
5. Identificación de elementos reutilizables
6. Robustez de la aplicación
7. Documentación del proceso de construcción de cada módulo
8. Implementación de una base sólida basada en estándares sobre la cual descansa el proyecto SIGC3.
9. El modelo de casos de uso resulto particularmente efectivo para establecer comunicación con el usuario final.
<b>Desventajas</b>
1. En una primera iteración, el modelo de implementación tiende a separarse del modelo de diseño, cuando el equipo de desarrollo utiliza herramientas nuevas.
2. Generalmente no se dispone de tiempo suficiente para realizar mas de dos iteraciones
3. Fue necesario invertir tiempo en capacitación sobre la metodología.

#### 4. Conclusiones.

Los elementos de PUDS llevados a la práctica han permitido obtener resultados tangibles y por tanto cuantificables en todas las fases del proceso de desarrollo, además de ayudar a medir la efectividad del producto final (código y/o componentes), contribuyendo a asegurar la calidad. Aunque no se ha aplicado en forma exhaustiva esperamos mejorar la efectividad en su uso y obtener de esta forma el máximo provecho posible.

Entre los aspectos identificados para mejorar sustancialmente el proceso de desarrollo están:

- La administración de procesos especificando
  - El flujo de control para cambios o extensiones en los productos
  - Estableciendo los canales de comunicación que permitan involucrar de manera adecuada a los integrantes del grupo de desarrollo
- La planificación y ejecución de pruebas, en especial de pruebas de regresión.

#### Referencias

- [1] I. Jacobson, G. Booch., J. Rumbaugh, *El Proceso Unificado de Desarrollo de Software*, Addison Wesley, 2000. ISBN 84-7829-036-2
- [2] Scout W.Ambler, *The Elements of UML Style*. Cambridge University Press 2003, ISBN 0 521 52547
- [3] Leslee Probasco, *The ten essentials of RUP*. The essence of an effective development Process. Rational Software White Paper 2000.  
<http://www-106.ibm.com/developerworks/rational/library/content/RationalEdge/dec00>



# **Problemas de la optimización en la computación paralela**

Andrei Tchernykh  
Klaus Ecker  
(Editores)





## **Introduction**

The papers in this chapter cover the areas of parallel and distributed computation, high-performance computing, parallel software methodologies, programming languages, resource management and scheduling, real-time systems, cluster and GRID computing.

The purpose of the workshop is to provide an open forum for the parallel computing and scheduling research communities to debate key issues of common interest, to give an overview of new developments, applications, and trends in parallel computing technologies, and to further the state of the art in parallel cluster and GRID computation.

Participants from France, Germany, Mexico, Poland, Russia, and USA presented unpublished results of their research.

We want to thank the persons whose active and steady work made the workshop possible and contributed to its success. In particular, we would like to thank to the members of the program committee for their useful suggestions regarding the workshop orientation and organization. Special thanks are given to the referees for their meticulous reading and evaluation of the submissions, and their valuable suggestions for the authors. We also gratefully emphasize the assistance of local organizers, secretaries, and students. Their valuable help provided the success of the workshop.

## Scientific Committee

### Co-Presidents:

Andrei Tchernykh, CICESE, Ensenada, Mexico, [chernykh@cicese.mx](mailto:chernykh@cicese.mx)

Klaus Ecker, TUC, Clausthal, Germany, [ecker@in.tu-clausthal.de](mailto:ecker@in.tu-clausthal.de)

Denis Trystram, Inst. of Inform. and Appl. Math., Grenoble, France, [trystram@imag.fr](mailto:trystram@imag.fr)

Dr. Andrei Tchernykh, CICESE, Ensenada, México, [chernykh@cicese.mx](mailto:chernykh@cicese.mx)

Dr. Klaus Ecker, TUC, Clausthal, Germany, [ecker@in.tu-clausthal.de](mailto:ecker@in.tu-clausthal.de)

Bruno Raffin, INRIA, Grenoble, France [raffin@imag.fr](mailto:raffin@imag.fr)

Victor Ivannikov, Inst. for System Programming, Moscow, Russia, [ivan@ispras.ru](mailto:ivan@ispras.ru)

Jacek Blazewicz, Inst. of Informatics, Politechnika Poznanska, Poznan, Poland,

[blazewicz@sol.put.poznan.pl](mailto:blazewicz@sol.put.poznan.pl)

### Special thanks to:

Juan Manuel Ramirez, Universidad de Colima, [jmramir@ucol.mx](mailto:jmramir@ucol.mx)

Sonia I. Serrano Barreda, Universidad de Colima, [sonser@ucol.mx](mailto:sonser@ucol.mx)

# A X Window Proxy Server for Display Wall

Jesus Alberto Verduzco and Bruno Raffin

Laboratoire Informatique et Distribution  
Project APACHE ID-IMAG CNRS - INPG - INRIA - UJF  
Grenoble, France

`Jesus.Verduzco@imag.fr`, `Bruno.Raffin@imag.fr`

**Abstract.** The necessity to have a surface of visualization of greater resolution than the one provided by a single monitor becomes evident when users interact with applications that generate significant amounts of graphics data, like scientific data visualization, virtual reality scenes, engineering design, financial graphics analysis, etc. This paper describes the first phase of the design and the implementation of a proxy server for the X Window server. The goal is to enable the execution of X applications, without code modification or recompilation on a large surface made of multiple screens or video-projectors. The proxy distributes the drawing requests of a client towards a set of servers that cooperate to construct a unique and high resolution image.

## 1 Introduction

Over the last 10 years the performance and capacity of most of the components of computers have dramatically improved. However the display resolution and size have slowly evolved. The interest to increase to the dimensions and resolution of the visualization surface has been present in these recent years. Motivated without a doubt by the popularization of applications that require to display large amounts of graphical data. Today, a common approach to increase the size and the resolution of the visualization surface consists in tiling several screens or projectors to build a single large and high resolution image. These systems are called display walls. Such display wall are usually driven by a PC cluster, each PC controlling one or few video-projectors. Figure 1 shows a display wall of four individual surfaces (called tiled). Each node of the cluster is responsible to generate the image corresponding to its tile and the communication network is used to transfer data between these nodes to keep the images synchronized. Interactions use a keyboard and mouse or other input devices like 3D wands, 3D tracker, etc. There exist several recent projects of display walls like the Scalable Display Wall [10] at the Princeton University [13], the Infinite Wall at the University of Illinois at Chicago [8], the Office of the future [15] at UNC, the Information Mural at Stanford [11], the Power Wall at the University of Minnesota [9]. Traditionally, visualization intensive applications use dedicated parallel computers like SGI-Onyx. The increase of performance and low cost of PCs turned the PC cluster into a viable alternative for such applications. Actually, most of display

walls are driven by PC clusters.

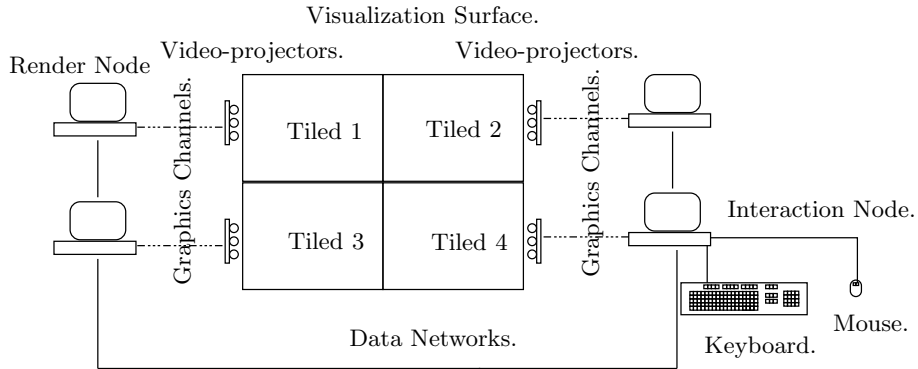
The research effort have been focused on developing software solutions to run 3D applications on PC cluster and display walls [7, 6, 12]. Fewer works were dedicated to run classical 2D applications on a PC cluster and display wall. Two projects, VNCwall [1] and Distributed multihead X [2], address this problem and propose each one its own approach. VNCwall is an extension of the software called VNC (Virtual Network Computing)[17]. VNC is aimed at displaying an application on a remote machine, taking advantage of a pixel level compatibility. The application first sends graphics orders to a VNC server that computes the image in a shadow frame buffer. Next, this image is sent for display to a VNC viewer running on a remote machine. VNCwall extends this principle by computing the full image on a modified VNC server that next splits this image in tiles and distribute them to the different nodes of the display wall. The rendering effort is then concentrated on the VNC server and the amount of data (the pixels) to send to the viewers is important. On the other hand, Xdmx is based on the X Window system and two extensions, Xnest [4] and Xinerama [3]. A front-end server appears to the applications like a classical X server, while this server appears to each node of the display wall (backend nodes) like a single client nesting a complete X server using Xnest. Although it is a functional system in constant improvement, his performance depends on Xinerama. In Xinerama most of the operations on the graphics applications are carried out with copy and paste operations. This produces performance troubles with applications that generate graphics intensively. Our goal is to provide a software solution that allows the applications based on the X Window system to be displayed on the display wall surface without any code modification or relinking. Our solution differs from previous works in that it uses a different approach to process and distributed the graphics primitives than Xdmx and VNCwall. Our software is light and under certain operating conditions we won in performance. This paper is organized as follow. In section 2, we briefly present the X Window system and Xinerama. In section 3 we analyze Xdmx and VNCwall. In the section 4 we described our solution. Finally in the sections 5 and 6 we discusse the preliminary results before to establish our conclusions.

## 2 Background

This section introduces the general concepts used throughout this document.

### 2.1 X Window

The X Window system [16] (commonly called X) is a server of windows that constructs the graphics of the X applications and manages the interaction of the user by means of a keyboard and mouse. X uses the client/server model. The application of the user represents the client (called X client) that sends messages (x11 messages) that contains *drawing request*. The server acts as an intermediary between client and the graphics resources of the local machine



**Fig. 1.** Architecture of a Display Wall.

where it is executed. In order to carry out the construction of the client image in the monitor, the X server processes the *drawing request* by means of a set of functions that writes in video memory (*framebuffer*) of the graphics card. The interactions of the user by means of the mouse and the keyboard are received in form of *events* by the server and sent towards the client, who processes them and returns the result to the server again. The communication between client and server is normalized between both parts by means of the X protocol. The X server support mechanisms to save offscreen images used when fractions of the windows are hidden, for example the regions that are not displayed in the monitor are eliminated by a *clipping* mechanism.

## 2.2 Xinerama

Xinerama is an extension of X. By means of Xinerama an instance of X in execution on one machine is able to manage a set of graphics cards and monitors like a single one and to make available for the X clients a large surface of visualization. To reach its objective, Xinerama intercepts and breaks up drawing requests and after repackage them for sending for each screen. An important problem in Xinerama is the limited number of graphics cards that can be installed in a PC (just one for AGP graphics cards). Another drawback is that the processing is supported by a single machine so that there is an important overload and a gradual decrease of performance if the size of the visualization surface is increased.

## 3 Related Work

This section analyzes the works related to our project.

### 3.1 VNCwall

VNCwall is a solution that allows applications based on X to use the visualization surface of display wall. VNCwall is based on VNC. There are three components in VNC:

- The Server is a version of the X Window server called Xvnc [5] that rendering graphics in main memory (called *shadow frame buffer*).
- The Viewer is the interaction point in front of the user. It displays the application graphics and allows the user interactions by means of the keyboard and mouse.
- The Protocol RFB is a mechanism for remote access to graphical user interfaces that allows to transmit groups of pixels between server and viewer.

As between server and viewer only pixels are transmitted, this approach has some advantages like portability (VNC supports X11, Window and Macintosh and run on the wide range of hardware). Having a viewer specific to a graphics-processing environment, an application can use the VNC server to render and display its graphics. Another important issue is that the *shadow frame buffer* grows in a proportional way to the size of the screen used by the viewer, so that this increases the memory consumption in the PC where the server is executed. A third problem is the network bandwidth consumed by the transfer of important amounts of pixels between server and viewer, although compression and caching techniques are used. This problem becomes serious when there is a client that modifies in an intensive way their graphics leading to an increase in the exchange of pixels between viewer and server, diminishing the efficiency of the protocol RFB.

### 3.2 Xdmx

Distributed multihead X is a solution that allows to distribute a desktop of the master X Window server (called front-end server) between a set of slave X Window servers (called backend servers). A X client using Xdmx is able to display their graphics on a display wall.

This project is based mainly on Xinerama and Xnest. Xnest is an extension of the X server that works like a proxy server. A proxy server is a process that takes advantage of the client/server X window architecture installed between the client and the server. It intercepts, modifies, discards or redirects the X11 messages.

In Xdmx, a single X server works like a proxy server connected to a set of back-end X servers. The request from clients are accepted by the front-server. Rendering is performed by the back-end servers.

Once a request is received from the front-end, it uses the Xinerama functionalities to process and distribute the client request to the backend servers. A problem in Xdmx is that the front-end is a complete server. Its installation, compilation and modification is a difficult process.

## 4 Proxy Server Overview

In this section we describe our approach. To distribute the display of a single X Window server, enable the request processing and display the X client graphics by a multiple servers. The X11 messages need to be transmitted towards each server. In the same way the messages generated from the user actions on the keyboard and mouse in the interaction servers must be recovered and directed to the client.

### 4.1 Software Architecture

In this section, we describe the component that we developed.

**Virtual Display** During the initial phase of execution, the proxy must consider the set of screens of the rendering servers like a global single graphics surface (we call it *virtual display*). To construct the *virtual display*, the proxy sends requests to the rendering servers to recover the graphics display characteristics. The display size is the main information used by the proxy to define the size of the *virtual display*. When a client sends an initial connection request to the proxy, this one returns the data that describes the graphics characteristics of the *virtual display*. Using this information the client can begin to send requests adapted to the resolution of the *virtual display*.

**X Request Processing** This functionality takes advantage of the proxy to intercept, modify and distribute the client requests.

- Request interception. It is carried out by the proxy that listen on a socket for the client requests. When the proxy receives a client initial request, it creates a socket to communicate with the client.
- Request modification. Normally the requests sent from the client to the server are adapted to the *graphics characteristics* of the display managed by this server. On the other hand, the server maintains structures that keep information about the identifiers of the resources created by the client. In our design, a single client maintains connections to a set of X servers using an intermediate process (the proxy). It is necessary that the proxy keeps the information from the resources created by the client in each of the rendering servers. Whenever the proxy sends a request it must adapt it to the corresponding server. Another central problem is the adjustment of requests to the dimensions of the *virtual display*. A client request contains the value of the coordinates of a point in the *virtual display*. We applied a transformation function to determine the corresponding coordinate into the coordinate system of each rendering server. The clipping operation performed by each X rendering server ensures that the windows image will be shown like a only piece on the *virtual display* dimensions.



- Request distribution. To send the client request from the proxy to the set of rendering servers, we consider two alternatives: We redirect the requests towards all rendering servers or only towards the rendering servers who will display the graphics. In this first phase, we applied the first alternative because we considered that the user can move the graphics between the screen of the rendering servers. This user operation it causes the interchange of fractions of the graphics between the rendering servers like intermediary the proxy. Then there is a communication cost implied by this operation. However with the applied procedure, each rendering server locally has a copy of the client graphics that can be displayed without perturbing the rest of the system. If a window image is displayed by two or more rendering servers, the modification of the coordinates carried out by the proxy and the operation of clipping applied locally by each rendering server allows to visualize the set of windows like a single image.

## 5 Experimental Results

To evaluate the performance of our implementation we used the *x11perf*[14] test suite. Our tests are performed on the GrImage platform located at the INRIA Rhône-Alpes. GrImage gathers 11 dual-Xeon 2.6 GHz PCs connected by a gigabit Ethernet network and a display wall of 8 projectors. Four PCs are dedicated to rendering, each one driving 2 projectors with a NVIDIA GeForce FX 5600 graphics board. The test application consists in applying two types of tests. It allows to evaluate the cost in request distribution, the request processing time by the proxy and the graphics rendering time.

- PutImage XY format 500x500 square. This test consists in clients sending a request to display an image of 500x500 pixels. This X11 request only transports pixels so that, its distribution towards the rendering servers implies an important cost of communication.
- Create child window and map using MapSubwindows. Using this test the client sends the request that transports the *graphics commands*. Then the cost of transport and process is less important so that the system performance is increased.

Tests were performed with a varying number of clients concurrently executing the benchmark and a varying number of rendering servers. All tests were performed in the same conditions with our approach and with Xdmx and the results are expressed in operations per second (ops).

The results with the test **Create child window and map using MapSubwindows** (table 1) show that the proxy server is at most 41% slower than Xdmx, because Xdmx optimization mechanism send the *graphics commands* only towards the rendering servers implied in its visualization, while the proxy server sends each one of them to the set of rendering servers. Nevertheless if we increased the number of active clients this percentage is reduced until 10%.

**Table 1.** Benchmark create and map window

	Xdmx			Proxy Server		
	1 Server	2 Servers	3 Servers	1 Server	2 Servers	3 Servers
1 Client	72,600	73,100	72,400	61,900	35,300	22,130
5 Clients	52,980	59,940	30,790	45,020	24,410	11,135
10 Clients	14,156	16,595	11,760	13,600	15,924	8,744

**Table 2.** Benchmark putimage500.

	Xdmx			Proxy Server		
	1 Server	2 Servers	3 Servers	1 Server	2 Servers	3 Servers
1 Client	30	30	31	43	31	23
5 Clients	13	7	11	14	12	9
10 Clients	5	6	6	12	10	10

The results when we applied the test **PutImage XY** (table 2) show that our system is 15% faster than the Xdmx.

These results show that the proxy server provides an efficient mechanism to process and to distribute the X11 request from the clients X. Our software is light, and processing request that transports pixels (*big request*) and with many active clients we reached equal performances to Xdmx

## 6 Conclusion and Future Works

In this paper we showed the first phase of the development of a proxy server that receives request from clients and distributes them towards a set of servers X Window. The goal is to enable the execution of unmodified X applications on a Display wall driven by a PC cluster.

Using this implementation the user have a tool that displays the application graphics on large surfaces, increasing the amount of information that can be perceived and improving the user interaction process. These functionalities are fundamental to facilitate the analysis of large data sets and sessions of cooperative work with multiple users.

This goal of this first prototype was to validate the functionality of the proposed design. In this first stage we implemented the treatment of a small representative subgroup of the X protocol requests. The results obtained are a strong motivation to finalize the implementation.

Future works will focus on improving the performance and to add mechanisms of interaction adapted to the dimensions of the visualization surface.

## References

1. Displaywall in a box software. <http://www.ncsa.uiuc.edu/TechFocus/Deployment/DBox/Doc/vnc.html>.
2. Distributed multihead x. <http://sourceforge.net/projects/dmx/>.
3. Xinerama extension to the x protocol. <http://sourceforge.net/projects/xinerama/>.
4. Xnest - a nested x server. <http://www.xfree86.org/current/Xnest.1.html>.
5. Xvnc - an x server providing vnc connectivity. <http://www.tightvnc.com/doc/man/Xvnc.1.html>.
6. J. Allard, V. Gouranton, L. Lecointre, E. Melin, and B. Raffin. Net juggler: Running vr juggler with multiple displays on a commodity component cluster. pages 275–276, March 2002.
7. P. Augerat, C. Goudeseune, H. Kaczmarek, B. Raffin, B. Schaeffer, L. Soares, and M. K. Zuffo. Commodity clusters for immersive projection environments. Siggraph 2002 Course, July 2002.
8. C. Cruz-Neira, D. J. Sandin, and T. A. Defanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. volume 27, pages 135–142, August 1993.
9. P. W. et al. The power wall. *Silicon Graphics Supercomputing '94*, pages 14–17, November 1994.
10. M. Hereld, I. Judson, and R. Stevens. Introduction to building projection-based tiled display systems. volume 20, pages 22–28, July/August 2000.
11. G. Humphrey and P. Hanrahan. A distributed graphics system for large tiled displays. *IEEE Visualization 99*, pages 24–29, October 1999.
12. G. Humphreys, M. Eldridge, I. Buck, G. Stoll, M. Everett, and P. Hanrahan. Wiregl: A scalable graphics system for clusters. In *Proceedings of SIGGRAPH 2001*, 2001.
13. K. Li, H. Chen, Y. Chen, D. W. Clark, P. Cook, S. Damianakis, G. ESSL, A. Finkelstein, T. Funkhouser, A. Klein, Z. Liu, E. Praun, R. Samanta, B. Shedd, J. P. Singh, G. Tzanetakis, and J. Zheng. Early experiences and challenges in building and using a scalable display wall system. *IEEE Computer Graphics and Application*, 20:671–680, 2000.
14. J. McCormack, P. Karlton, S. Angebrannt, C. Kent, K. Packard, and G. Gill. x11perf - x11 server performance test program. <http://www.xfree86.org/4.2.0/x11perf.1.html>.
15. R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image-based modeling. In *SIGGRAPH 98*, pages 179–188, July 1998.
16. R. W. Scheifler and J. Gettys. The x window system. *ACM Transactions on Graphics*, 5:79–109, 1986.
17. K. W. A. H. T. Richardson, Q. Stafford-Fraser. Virtual network computing. 2, January/February 1998.

# A polynomial approximation scheme for the $1 \mid r_j \mid L_{\max}$ scheduling problem with the guaranteed absolute error

Alexander A. Lazarev and Ruslan R. Sadykov

Kazan State University, Kremlevskaya Street 18, 420008 Kazan, Russia  
Alexandr.Lazarev@ksu.ru

**Abstract.** We consider NP-hard in the strong sense scheduling problem of minimizing maximum lateness on a single machine  $1 \mid r_j \mid L_{\max}$ . A polynomial approximation scheme has been developed. This scheme is based on the reduction of an instance to a polynomially solvable special case of the problem. We obtained the value of the maximum absolute error of such approximate solution for the certain variants of the scheme.

## 1 Introduction

First, we describe the problem. A set of jobs  $N = \{1, \dots, n\}$  has to be processed on a single machine. The machine can only process one job at a time and pre-emptions are not allowed. Each job  $j \in N$  has a release date  $r_j$ , a processing time  $p_j$  and a due date  $d_j$ .

Let  $\Pi(N)$  be the set of all possible schedules which include all the jobs from the set  $N$ .  $c_j(\pi)$  is the completion time of job  $j$  in schedule  $\pi$ . The lateness  $L_j(\pi)$  of job  $j$  in schedule  $\pi$  is the difference between the completion time of the job  $j$  and its due date:  $L_j(\pi) = c_j(\pi) - d_j$ . The objective is to minimize the maximum lateness over all schedules in  $\Pi(N)$ . Using standard scheduling notation [5], this problem is denoted as  $1 \mid r_j \mid L_{\max}$ .

The work on this problem began in the fifties in the last century. Lenstra et al. [13] have showed that the general case of the problem is  $\mathcal{NP}$ -hard in the strong sense. A number of polynomially solvable cases of the problem has been established beginning from the early Jackson result [8] for the special case  $r_j = 0, j \in N$ , when the EDD rule produces an optimal schedule. Special cases  $1 \mid prec; r_j \mid C_{\max}$  and  $1 \mid prec; p_j = p; r_j \mid L_{\max}$  and  $1 \mid prec; r_j; pmtn \mid L_{\max}$  with preceding constraints have been considered in the works by Lawler [10], Simons [18], Baker et al. [2], respectively. Hoogeveen [7] has suggested a polynomial algorithm for the special case of the problem, when the jobs satisfy to the condition  $d_j - p_j - \beta \leq r_j \leq d_j - \beta$  for some constant  $\beta$ . Another polynomially solvable case has been found by Lazarev [11]. A pseudopolynomial algorithm for a  $\mathcal{NP}$ -hard case of the problem ( $d_1 \leq \dots \leq d_n$  and  $r_1 \geq \dots \geq r_n$ ) has been developed by Lazarev and Shulgina [12].

Among the exact methods we can distinguish the Branch-and-Bound algorithm by Carlier [4]. This method is very good for solving the overwhelming majority of instances which one can encounter in practice. Earlier in the works [3, 15, 9] other exact approaches for the problem have been described.

In the literature one can find several approaches which can guarantee a certain relative error of the solution. Potts [17] has presented an iterative version of the extended Jackson rule (IJ) and has proved that  $L_{\max}(IJ)/L_{\max}^* \leq \frac{3}{2}$ . Hall and Shmoys [6] have modified the iterative version and have obtained the MIJ algorithm, which guarantees  $L_{\max}(MIJ)/L_{\max}^* \leq \frac{4}{3}$ . Also they have described two polynomial time approximation schemes which gives a solution with a value within a factor of  $(1 + \varepsilon)$  for any given  $\varepsilon$ . The first algorithm has the complexity  $O(n \log n + n(1/\varepsilon)^{O(1/\varepsilon^2)})$  and the running time of the second is  $O((n/\varepsilon)^{O(1/\varepsilon)})$ . Recently Mastrolilli [14] has presented improved  $\varepsilon$ -approximation scheme, whose running time is  $O(n + (1/\varepsilon)^{O(1/\varepsilon)})$ .

There are a few polynomial algorithms which guarantee a certain absolute error of the solution. For example, the Schrage algorithm produces a solution with an absolute error which is not more than  $\max_{j \in N} p_j$ . In this paper we present another scheme to find an approximate decision with a guaranteed absolute error. This scheme is based on the reduction of an instance to certain polynomially solvable case of the problem.

The paper is organized as follows. In Section 2 we obtain the maximum absolute error of the optimal solution of the reduced instance for the initial instance of the problem. Section 3 presents the scheme to find an approximate solution. Then the variants of the scheme based on two polynomially solvable instances are considered. For these variants we express the maximum absolute error of an approximate solution only in the initial parameters of jobs. Finally, in Section 4 some conclusions are drawn.

## 2 The absolute error of an approximate solution

Without lost of generality we assume that  $d_1 \leq d_2 \leq \dots \leq d_n$ .

Let  $L_{\max}^I(\pi)$  and  $c_j^I(\pi)$  be, respectively, the maximum lateness of a schedule  $\pi \in \Pi(N)$  and the completion time of a job  $j \in N$  in a schedule  $\pi$  for a instance  $I$  with the parameters of jobs  $\{r_j^I, p_j^I, d_j^I\}$ .

Consider two instances  $A$  and  $B$  of the problem. Let  $r_j^A = r_j^B$  and  $p_j^A = p_j^B$ ,  $j \in N$ . In other words, the instances only differ by due dates. Let  $\pi^A$  and  $\pi^B$  are optimal schedules for the instances  $A$  and  $B$ , respectively. The next lemma finds the maximal absolute error of the schedule  $\pi^B$  for the instance  $A$ .

**Lemma 1.** *For the optimal schedules  $\pi^A$  and  $\pi^B$  the following holds:*

$$\rho_d \geq L_{\max}^A(\pi^B) - L_{\max}^A(\pi^A) \geq 0, \text{ where}$$

$$\rho_d = \max_{j \in N} \{d_j^A - d_j^B\} - \min_{j \in N} \{d_j^A - d_j^B\}.$$

□

Consider two instances  $X$  and  $Y$  of the problem with the parameters of jobs  $\{r_j^X, p_j, d_j^X\}$  and  $\{r_j^Y = -d_j^X, p_j, d_j^Y = -r_j^X\}$ ,  $j \in N$ , respectively. The instance  $Y$  is said to be *inverse* to the instance  $X$ .

**Lemma 2.** *Let there are two schedules  $\pi^X \in \Pi(N)$  and  $\pi^Y \in \Pi(N)$ . In the schedule  $\pi^Y$  the jobs are sequenced in the reverse order in comparison with the schedule  $\pi^X$ . Then  $L_{\max}^X(\pi^X) = L_{\max}^Y(\pi^Y)$ .  $\square$*

*Consequence* If the schedule  $\pi^X$  is optimal for the instance  $X$ , then the schedule  $\pi^Y$  is optimal for the instance  $Y$ .

Consider two instances  $B$  and  $C$  of the problem. Let  $p_j^B = p_j^C$  and  $d_j^B = d_j^C$ ,  $j \in N$ . In other words the instances only differ by release dates. Let  $\pi^B$  and  $\pi^C$  are optimal schedules for the instances  $B$  and  $C$ , respectively. The next lemma finds the maximal absolute error of the schedule  $\pi^C$  for the instance  $B$ .

**Lemma 3.** *For the optimal schedules  $\pi^B$  and  $\pi^C$  the following holds:*

$$\begin{aligned} \rho_r &\geq L_{\max}^B(\pi^C) - L_{\max}^B(\pi^B) \geq 0, \text{ where} \\ \rho_r &= \max_{j \in N} \{r_j^B - r_j^C\} - \min_{j \in N} \{r_j^B - r_j^C\}. \end{aligned}$$

$\square$

Consider two instances  $A$  and  $C$  of the problem. Let  $p_j^A = p_j^C$ ,  $j \in N$ . In other words the instances may differ both by release and due dates. Let  $\pi^A$  and  $\pi^C$  are optimal schedules for the instances  $A$  and  $C$ , respectively. The next lemma finds the maximal absolute error of the schedule  $\pi^C$  for the instance  $A$ .

**Theorem 1.** *For the optimal schedules  $\pi^A$  and  $\pi^C$  the following holds:*

$$\begin{aligned} \rho &\geq L_{\max}^A(\pi^C) - L_{\max}^A(\pi^A) \geq 0, \text{ where} \\ \rho &= \max_{j \in N} \{r_j^A - r_j^C\} - \min_{j \in N} \{r_j^A - r_j^C\} + \\ &\quad \max_{j \in N} \{d_j^A - d_j^C\} - \min_{j \in N} \{d_j^A - d_j^C\}. \end{aligned} \tag{1}$$

$\square$

*Consequence*  $\rho \geq L_{\max}^C(\pi^A) - L_{\max}^C(\pi^C) \geq 0$ .

So  $0 \leq L_{\max}^A(\pi^C) - L_{\max}^A(\pi^A) \leq \rho = \max_{j \in N} \{d_j^A - d_j^C\} - \min_{j \in N} \{d_j^A - d_j^C\} + \max_{j \in N} \{r_j^A - r_j^C\} - \min_{j \in N} \{r_j^A - r_j^C\}$  and  $0 \leq L_{\max}^C(\pi^A) - L_{\max}^C(\pi^C) \leq \rho$ . Hence, if after an instance of the problem 1 |  $r_j$  |  $L_{\max}$  is solved, release or due dates are changed, then we can estimate the absolute error of the solution found for the new changed instance.

The obtained results can be also used for the approximate solution of an instance of the problem 1 |  $r_j$  |  $L_{\max}$ . The scheme for finding an approximate decision is considered in the next section.

### 3 The scheme for finding an approximate solution

Consider an arbitrary instance  $A$  of the problem  $1 | r_j | L_{\max}$  with the parameters of jobs  $\{r_j^A, p_j^A, d_j^A\}, j \in N$ . Consider also a polynomially solvable case of the problem when the parameters of jobs satisfy to the following linear constraints

$$\mathcal{A} * R + \mathcal{B} * P + \mathcal{C} * D \leq H, \quad (2)$$

where  $R^T = (r_1, \dots, r_n)$ ,  $P^T = (p_1, \dots, p_n)$ ,  $D^T = (d_1, \dots, d_n)$ ,  $\mathcal{A}, \mathcal{B}, \mathcal{C}$  - matrices of dimension  $n \times n$ , and  $H^T = (h_1, \dots, h_n)$  is some vector of dimension  $n$ . For example, for the special case when  $d_j = r_j + p_j + \beta$ ,  $j \in N$ , where  $\beta$  is a constant, the coefficients in the linear system (2) are the following:  $\mathcal{A} = \mathcal{B} = \mathcal{I}$ ,  $\mathcal{C} = -\mathcal{I}$ ,  $H^T = (\beta, \dots, \beta)$ . Here  $\mathcal{I}$  is the unitary matrix.

The idea of the approximation scheme consists in changing the release and/or due dates of the jobs of the instance  $A$  in such a way, that the reduced instance  $C$  with the parameters of jobs  $\{r_j^C, p_j^C = p_j^A, d_j^C\}, j \in N$ , satisfies the constraints (2). In this case it is possible to solve the instance  $C$  in polynomial time and the found optimal schedule  $\pi^C$  has the absolute error which is not more than the value of  $\rho$  for the instance  $A$ .  $\rho$  here can be found according to the formula (1).

Of course, when transforming the instance  $A$  into the instance  $C$  one should ensure that the value of  $\rho$  is minimized. The general way to minimize  $\rho$  is to solve the following linear programming problem:

$$\begin{aligned} \min & (x_d - y_d) + (x_r - y_r) \\ & y_d \leq d_j^A - d_j^C \leq x_d, \quad j \in N, \\ & y_r \leq r_j^A - r_j^C \leq x_r, \quad j \in N, \\ & \mathcal{A} * R^C + \mathcal{B} * P^C + \mathcal{C} * D^C \leq H. \end{aligned} \quad (3)$$

Whenever it is possible for the considered polynomially solvable case, a special algorithm for minimizing  $\rho$  should be developed.

Describe first a simple variant of the proposed approximation scheme. Consider the special case of the problem  $1 | r_j | L_{\max}$ , when  $d_j = \delta$ ,  $j \in N$ , where  $\delta$  is a constant. In other words all the due dates should be equal. For this special case, the schedule where the jobs are ordered by nondecreasing release dates is optimal. Denote such schedule as  $\pi_r$ . Obviously, one can find the schedule  $\pi_r$  in  $O(n \log n)$  time. To reduce an arbitrary instance of the problem to this special case, we should equate all the due dates with some constant. Set the linear programming problem of the type (3) for minimizing the maximum absolute error of the schedule  $\pi_r$  for the instance  $A$ :

$$\begin{aligned} \min & x_d - y_d \\ & y_d \leq d_j^A - \delta \leq x_d, \quad j \in N. \end{aligned} \quad (4)$$

The decision of the problem (4) is an arbitrary value of  $\delta$  and  $x_d, y_d$ , such that  $x_d - y_d = \max_{j \in N} d_j^A - \min_{j \in N} d_j^A = \rho$ . Thus in the considered case  $\rho$  does not depend on the value of  $\delta$ . So the maximum absolute error of the schedule  $\pi_r$  for the given instance  $A$  is the difference between the maximum and minimum due dates of  $A$ .

In two following subsections we consider other variants of the proposed approximation scheme. In these variants smaller values of the maximum absolute error of an approximate decision can be archived.

### 3.1 The variant based on the Lazarev's case

Consider the polynomially solvable case of the problem  $1 \mid r_j \mid L_{\max}$ , when the parameters of jobs satisfy the following conditions:

$$\begin{aligned} d_1 \leq \dots \leq d_n \quad \text{and} \\ d_1 - r_1 - p_1 \geq \dots \geq d_n - r_n - p_n. \end{aligned} \tag{5}$$

Lazarev [11] has proposed the algorithm of complexity  $O(n^3 \log n)$  to solve instances of this special case.

Let an instance  $A$  with the parameters of jobs  $\{r_j^A, p_j^A, d_j^A\}$ ,  $j \in N$ , is given, which does not satisfy the conditions (5). We apply the proposed approximation scheme for this instance. Change due and release dates of the instance  $A$  in such a way, that the reduced instance  $C$  with the parameters of jobs  $\{r_j^C, p_j^C = p_j^A, d_j^C\}$ ,  $j \in N$ , satisfies the conditions (5). Let  $\pi^C$  is an optimal schedule for the instance  $C$ .

The next theorem justifies the algorithm which reduces the instance  $A$  to the instance  $C$  and minimizes the maximum absolute error of the schedule  $\pi^C$  for the instance  $A$ . The theorem also expresses the value of this maximum absolute error  $\rho$  in terms of the parameters of jobs of the instance  $A$ .

**Theorem 2.** *Let a given instance  $A$  of the problem  $1 \mid r_j \mid L_{\max}$  does not satisfy the constraints (5). Then there exists an instance  $C$ , which satisfies constraints (5), such that*

$$\rho = \max_{i,j \in N} \min\{d_j^A - d_i^A, d_j^A - d_i^A + r_i^A + p_i^A - r_j^A - p_j^A\}, \tag{6}$$

where  $\rho$  is computed by the formula (1). And there is no other instance which satisfies constraints (5), such that the value of  $\rho$  is smaller than in (6).  $\square$

Based on Theorem 2 we present Algorithm 1, which reduces the instance  $A$  to the instance  $C$  and guarantees the minimum value of the maximum absolute error  $\rho$ .

The complexity of Algorithm 1 is  $O(n \log n)$ .

*Example.* Let we have an instance  $A$  of the problem  $1 \mid r_j \mid L_{\max}$ . The parameters of jobs of the instance  $A$  are shown in Table 1. The jobs are ordered by nondecreasing  $r_j^A + p_j^A$ .

Begin the execution of the algorithm.  $N = \{1, \dots, 8\}$ . After the loop on the lines 2-4 we have  $r_j^C := r_j^A$ ,  $p_j^C := p_j^A$ ,  $\forall j \in N$ .

$k = 1$ .  $j_1 = \arg \min_{j \in N} d_j^A = 3$ .  $N_1 = \{1, 2, 3\}$ . After the loop on the lines 2-4 we have  $d_j^C = d_3^A = 11$ ,  $\forall j \in N_1$ . For now  $N \setminus N_1 \neq \emptyset$ .



---

**Algorithm 1** The algorithm reduces an arbitrary instance to the instance, which satisfies the Lazarev's special case, and minimizes  $\rho$

---

```

1: Renumber in the order by nondecreasing  $r_j^A + p_j^A$ .
2: for  $j := 1$  to  $n$  do
3:    $r_j^C := r_j^A$ ;  $p_j^C := p_j^A$ 
4: end for
5:  $k := 1$ 
6:  $j_1 := \arg \min_{j \in N} d_j^A$ 
7:  $N_1 := \{j \in N \mid j \leq j_1\}$ 
8: for all  $j \in N_1$  do
9:    $d_j^C := d_{j_1}^A$ 
10: end for
11: while  $N \setminus N_1 \setminus \dots \setminus N_k \neq \emptyset$  do
12:    $k := k + 1$ 
13:    $j_k := \arg \min_{j \in N \setminus N_1 \setminus \dots \setminus N_{k-1}} d_j^A$ 
14:    $N_k := \{j \in N \setminus N_1 \setminus \dots \setminus N_{k-1} \mid j \leq j_k\}$ 
15:   for all  $j \in N_k$  do
16:      $d_j^C := \min\{d_{j_k}^A, d_{j_{k-1}}^C - r_{j_{k-1}}^A - p_{j_{k-1}}^A + r_j^A + p_j^A\}$ 
17:   end for
18: end while

```

---

**Table 1.** The parameters of jobs of the instance  $A$

$j =$	1	2	3	4	5	6	7	8
$r_j^A$	0	3	2	1	5	3	5	7
$p_j^A$	4	1	3	5	3	5	4	2
$d_j^A$	14	12	11	15	14	13	18	16

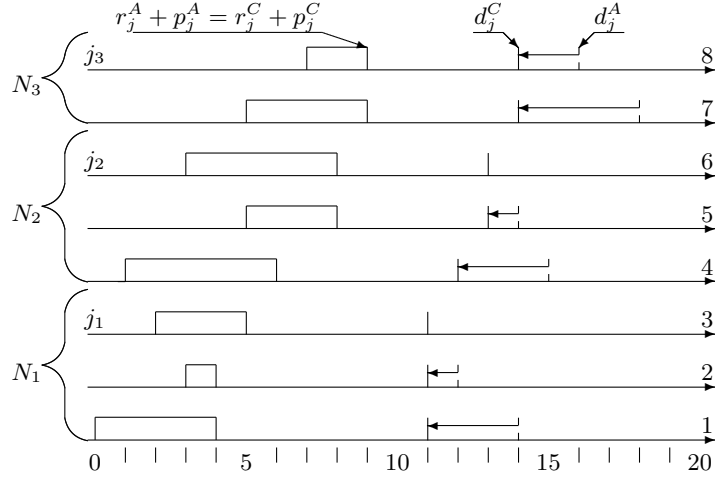
$k = 1$ .  $j_2 = \arg \min_{j \in N \setminus N_1} d_j^A = 6$ .  $N_2 = \{4, 5, 6\}$ .  
 $d_4^C = \min\{d_6^A, d_3^C - r_3^A - p_3^A + r_4^A + p_4^A\} = 12$ .  $d_5^C = \min\{d_6^A, d_3^C - r_3^A - p_3^A + r_5^A + p_5^A\} = 13$ .  
 $d_6^C = \min\{d_6^A, d_3^C - r_3^A - p_3^A + r_6^A + p_6^A\} = 13$ . For now  $N \setminus N_1 \setminus N_2 \neq \emptyset$ .  
 $k = 2$ .  $j_3 = \arg \min_{j \in N \setminus N_1 \setminus N_2} d_j^A = 8$ .  $N_3 = \{7, 8\}$ .  
 $d_7^C = \min\{d_8^A, d_6^C - r_6^A - p_6^A + r_7^A + p_7^A\} = 14$ .  $d_8^C = \min\{d_8^A, d_6^C - r_6^A - p_6^A + r_8^A + p_8^A\} = 14$ .

We have now  $N \setminus N_1 \setminus N_2 \setminus N_3 = \emptyset$  and the algorithm is completed.

So, we obtained the instance  $C$ , which satisfies the conditions (5). The parameters of jobs of the instance  $C$  are shown in Table 2. Let find the value of maximum absolute error in for the considered example:  $\rho = \max_{j \in N} \{d_j^A - d_j^C\} = d_7^A - d_7^C = 4$ . The employment of Algorithm 1 is also exhibited in Figure 1.

### 3.2 The variant based on the Hoogeveen's case

As another variant of the approximation scheme consider the Hoogeveen's case [7], when the parameters of jobs satisfy the following conditions for some constant



**Fig. 1.** Example of the work of Algorithm 1

**Table 2.** The parameters of jobs of the reduced instance  $C$

$j =$	1	2	3	4	5	6	7	8
$r_j^C$	0	3	2	1	5	3	5	7
$p_j^C$	4	1	3	5	3	5	4	2
$d_j^C$	11	11	11	12	13	13	14	14

$\beta$ :

$$r_j \in [d_j - p_j - \beta, d_j - \beta] \quad \forall j \in N. \quad (7)$$

There is the algorithm of complexity  $O(n^2 \log n)$  for solving the instances which belong to this special case. The constraints (7) can be rewritten in the following way. There exist a constant  $\beta$ , such that:

$$\beta \in [d_j - p_j - r_j, d_j - r_j] \quad \forall j \in N. \quad (8)$$

Let an instance  $A$  with the parameters of jobs  $\{r_j^A, p_j^A, d_j^A\}$ ,  $j \in N$ , is given, which does not satisfy the conditions (7). We apply the proposed approximation scheme for this instance. Change due and release dates of the instance  $A$  in such a way, that the reduced instance  $C$  with the parameters of jobs  $\{r_j^C, p_j^C = p_j^A, d_j^C\}$ ,  $j \in N$ , satisfies the conditions (7). Let  $\pi^C$  is an optimal schedule for the instance  $C$ .

Again, the next theorem justifies the algorithm which reduces the instance  $A$  to the instance  $C$  and minimizes the maximum absolute error of the schedule  $\pi^C$  for the instance  $A$ .

**Theorem 3.** *Let a given instance  $A$  of the problem  $1 | r_j | L_{\max}$  does not satisfy the constraints (7). Then there exists an instance  $C$ , which satisfies constraints (7), such that*

$$\rho = \max_{i,j \in N} \{d_j^A - r_j^A - p_j^A - d_i^A + r_i^A\}, \quad (9)$$

where  $\rho$  is computed by the formula (1). And there is no other instance which satisfies constraints (7), such that the value of  $\rho$  is smaller than in (9).  $\square$

Based on Theorem 3 we have the simple Algorithm 2, which reduces the instance  $A$  to the instance  $C$  and guarantees the minimum value of the maximum absolute error  $\rho$ . The complexity of Algorithm 2 is  $O(n)$ .

---

**Algorithm 2** The algorithm reduces an arbitrary instance to the instance, which satisfies the Hoogeveen's special case, and minimizes  $\rho$

---

```

1:  $j^* = \arg \max_{j \in N} \{d_j^A - r_j^A - p_j^A\}$ 
2: for  $j := 1$  to  $n$  do
3:    $r_j^C := r_j^A$ 
4:    $p_j^C := p_j^A$ 
5:    $d_j^C = \max\{d_{j^*}^A - r_{j^*}^A - p_{j^*}^A + r_j^A, d_j^A\}$ 
6: end for

```

---

## 4 Conclusion

In this paper we presented the polynomial approximation scheme which allows to find a solution with a guaranteed absolute error. The scheme is based on reducing a given instance to a polynomially solvable case of the problem. For two variants of the scheme the values of the maximum absolute error of the scheme solution have been found and expressed in terms of parameters of jobs of the initial instance.

It is interesting now to compare the absolute error, guaranteed by the proposed scheme and the absolute error, provided by other methods which find an approximate solution. Classes of instances could be obtained, where variants of the proposed scheme guarantee the best absolute error.

## References

1. J. Adams, E. Balas, D. Zawack. The shifting bottleneck procedure for job shop scheduling. *Manag. Sci.*, 34:3 (1988), pp. 391-401.
2. K.R. Baker, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan. Preemptive scheduling of a single machine to minimize maximum cost subject to release dates and precedence constraints. *Oper. Res.*, 31:2 (1983), pp. 381-386.
3. K.R. Baker, Z.S. Su. Sequencing with due dates and early start times to minimize tardiness. *Naval Res. Logist. Quart.*, 21 (1974), pp. 171-176.

4. J. Carlier. The one-machine sequencing problem. *Europ. Jour. of Oper. Res.*, 11:1 (1982), pp. 42-47.
5. R.L. Graham, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan. Optimization and approximation in deterministic sequencing and scheduling: a survey. *Ann. Discrete Math.*, 5 (1979), pp. 287-326.
6. L.A. Hall, D.B. Shmoys. Jackson's rule for one-machine schedulings: Making a good heuristic better. *Math. Oper. Res.*, 17 (1992), pp. 22-35.
7. J.A. Hoogeveen. Minimizing maximum promptness and maximum lateness on a single machine. *Math. Oper. Res.*, 21 (1996), pp. 100-114.
8. J.R. Jackson. Scheduling a production line to minimize maximum tardiness. *Manag. Sci. Res. Project*, UCLA, Research report, 43 (1955).
9. B.J. Lageweg, J.K. Lenstra, A.H.G. Rinnooy Kan. Minimizing maximum lateness on one machine: computational experience and applications. *Statist. Neerlandica*, 30 (1976), pp. 25-41.
10. E.L. Lawler. Optimal sequencing of a single machine subject to precedence constraints. *Manag. Sci.*, 19:5 (1973), pp. 544-546.
11. A.A. Lazarev. *Effective algorithms for certain scheduling problems with due dates on a single machine*. Ph.D. Thesis, Kazan State University, (1989), in Russian.
12. A.A. Lazarev, O.N. Shulgina. A pseudopolynomial algorithm for the NP-hard scheduling problem of minimizing maximum lateness. *Proceedings of the 11-th International Baykal workshop "Optimization methods and applications"*. Irkutsk, Russia, (1998), pp. 163-167, in Russian.
13. J.K. Lenstra, A.H.G. Rinnooy Kan, P. Brucker. Complexity of machine scheduling problems. *Annals of Oper. Res.*, 1 (1975), pp. 343-362.
14. M. Mastrolilli, Efficient Approximation Schemes for Scheduling Problems with Release Dates and Delivery Times. *J. of Scheduling*, 6:6 (2003), pp. 521-531.
15. G. McMahon, M. Florian. On scheduling with ready times and due dates to minimize maximum lateness. *Oper. Res.*, 28 (1980), pp. 1436-1441.
16. L. Péridy, E. Pinson, D. Rivraux. Using short-term memory to minimize the weighted number of late jobs on a single machine. *Europ. J. of Oper. Res.*, 148 (2003), pp. 591-603.
17. C.N. Potts. Analysis of a heuristic for one machine sequencing with release dates and delivery times. *Oper. Res.*, 28 (1980), pp. 1436-1441.
18. B.B. Simons. A fast algorithm for single processor scheduling. *19-th Annu. Symp. Found. Comput. Sci.*, New York: Ann. Arbor, Mich., (1978), pp. 246-252.

# Solution Algorithms for the Total Tardiness Scheduling Problem on a Single Machine

Alexander Lazarev<sup>1</sup>, Alexander Kvaratskhelia, Andrei Tchernykh<sup>2</sup>

<sup>1</sup> Kazan State University, Russia

Alexandr.Lazarev@mail.ru

<sup>2</sup> CICESE, Mexico

chernykh@cicese.mx

**Abstract.** We study the classical NP-hard in the ordinary sense single-machine total tardiness scheduling problem  $1 \parallel \sum T_j$ . New polynomial and pseudo-polynomial solvable cases of the problem are presented and respective algorithms are proposed with no more than  $O(n^2 \sum p_j)$  time.

## 1 Introduction

We consider the single-machine total tardiness problem. Given a jobset  $N = \{1, 2, \dots, n\}$  with  $n$  jobs have to be scheduled on a single machine without preemptions. The machine can handle only one job at a time and is available for processing at a start time of machine  $t_0$ , typically  $t_0 = 0$ . Processing time  $p_j > 0$  and due date  $d_j$  are given for a job  $j \in N$ , by default we assume  $p_j \in \mathbb{Z}^+$ ,  $d_j, t_0 \in \mathbb{R}$ . A schedule  $\pi$  is defined as a permutation of the jobset's  $N$  elements  $\pi = (j_1, j_2, \dots, j_n)$ .

The problem calls for construction an optimal schedule  $\pi^*$  which minimizes function  $F(\pi) = \sum_{j=1}^n \max\{0, c_j(\pi) - d_j\}$  where  $c_j(\pi)$  is the completion time of job  $j$  in a schedule  $\pi$  (let  $\pi = (j_1, j_2, \dots, j_n)$  then  $c_{j_1}(\pi) = t_0 + p_{j_1}$  and  $c_{j_k}(\pi) = c_{j_{k-1}}(\pi) + p_{j_k}$  for  $k = 2, 3, \dots, n$ ).

The problem is known to be NP-hard in the ordinary sense [1]. Lawler proposed pseudo-polynomial algorithm in  $O(n^4 \sum p_j)$  time [2] to solve arbitrary instance. Among solution approaches for the problem, the decomposition based algorithms are widely indicated and studied [2–7]. The paper [9] showed that all known constructive heuristics and a number of decomposition heuristics for the total tardiness problem present arbitrarily bad approximation ratios.

In the paper we present four algorithm for solution the problem in the case  $p_1 \geq p_2 \geq \dots \geq p_n$  and  $d_1 \leq d_2 \leq \dots \leq d_n$ .

We use the following notation:  $x = \langle \{p_j, d_j\}_{j \in N}, t_0 \rangle$  is an instance of the problem with parameters of jobs  $p_j, d_j$ , jobset  $N$  and start time  $t_0$ ;  $\{\pi\}$  is the set of sequenced in schedule  $\pi$  jobs;  $(i \rightarrow j)_{\pi^*}$  means that processing of job  $i$  precedes to processing of job  $j$  in schedule  $\pi$  (i.e.  $c_i(\pi) < c_j(\pi)$ );  $\Pi^*(N', t')$  is the set of optimal schedules for instance  $\langle \{p_j, d_j\}_{j \in N'}, t' \rangle$ ,  $N' \subseteq N$ ,  $t' \geq t_0$ .

## 2 Partitioning procedure

In the paper we proposed polynomial and pseudo-polynomial algorithms for all sub-cases of the case:

$$\begin{cases} p_1 \geq p_2 \geq \dots \geq p_n \\ d_1 \leq d_2 \leq \dots \leq d_n \end{cases} \quad (1)$$

We present the following solution approach for the case (1). The initial jobset  $N$  is partitioned into subsets and, depends on the number of subsets, solution algorithm are proposed.

---

### Partitioning procedure

---

0.  $k := 1, \alpha_k := 1;$
1. **for**  $j = 2, 3, \dots, n$  **do**:
  - if**  $d_j - d_{\alpha_k} > p_j$  **then**
    - $\beta_k := j - 1;$
    - $k := k + 1;$
    - $\alpha_k := j;$
  - end if**
- end for**
2.  $\beta_k := n.$

---

The procedure partitions the jobset  $N$  into  $k \leq n$  subsets  $M_1, M_2, \dots, M_k$  where  $M_i = \{\alpha_i, \alpha_i + 1, \dots, \beta_i\}$ ,  $\alpha_1 = 1, \beta_k = n$ ,  $M_1 \cup M_2 \cup \dots \cup M_k = N$  and  $M_i \cap M_j = \emptyset$  if  $i \neq j$ .

For example, if  $N = \{1, 2, 3\}$ ,  $p_1 = 10, p_2 = 10, p_3 = 2$ ,  $d_1 = 7, d_2 = 9, d_3 = 10$  then the procedure constructs subsets  $M_1 = \{1, 2\}$  and  $M_2 = \{3\}$ .

In the following sections we present solution algorithms for the four sub-cases depends on value  $k$ :

- pseudo-polynomial algorithm *B-1* for the case  $k = 1$ ;
- pseudo-polynomial algorithm *B-k* for the case  $1 < k < n$ ;
- polynomial algorithm *C-1* for the case  $k = 1$  and  $d_n - d_1 \leq 1$ ;
- polynomial algorithm *B-n* for the case  $k = n$ .

Algorithm *B-k* can be used to solve an arbitrary instance in the case (1) with  $O(kn \sum p_j)$  time. If  $k = 1$  (i.e.  $d_n - d_1 \leq p_n$ ) then we can use algorithm *B-1* which runs in  $O(n \sum p_j)$ . If  $d_n - d_1 \leq 1$  then algorithm *C-1* constructs an optimal schedule in  $O(n^2)$  time. And finally, if  $k = n$  then algorithm *B-n* constructs an optimal schedule in  $O(n^2)$  time.

### 3 Pseudo-polynomial algorithms

To describe pseudo-polynomial algorithms *B-1* and *B-k* we need to define values  $d_j(t)$ :

$$d_j(t) = d_j - d_n + t - t_0, \quad j \in N.$$

Due to (1), for each  $t \in \mathbb{R}$  we have  $d_1(t) \leq d_2(t) \leq \dots \leq d_n(t)$ . Let  $\pi_l(t)$  and  $F_l(t)$  be an optimal schedule and optimal total tardiness value for the parametric instance  $\langle \{p_j, d_j(t)\}_{j \in N_l}, 0 \rangle$  where  $N_l = \{l, l+1, \dots, n\}$ .

Well known property of the total tardiness problem states that two instances  $x_1 = \langle \{p_j, d_j\}_{j \in N}, t_0 \rangle$  and  $x_2 = \langle \{p_j, d_j + C\}_{j \in N}, t_0 + C \rangle$ , where  $C$  is arbitrary constant, have the same sets of optimal schedules, i.e. the instances  $x_1$  and  $x_2$  are equal.

Due to this property, initial instance  $x = \langle \{p_j, d_j\}_{j \in N}, t_0 \rangle$  and corresponding parametric instance  $\langle \{p_j, d_j(t)\}_{j \in N}, 0 \rangle$  are equal if  $t = d_n$ .

The main idea of the algorithms *B-1* and *B-k* is the following. Consecutively, for  $l = n, n-1, \dots, 1$ , schedules  $\pi_l(t)$  are constructed and values  $F_l(t)$  computed for each point  $t \in \mathbb{Z}^+$ . Optimal schedule for the initial instance  $x$  is  $\pi_1(d_n)$ .

Notice that if  $d_n \notin \mathbb{Z}^+$  we construct and solve instance  $\langle \{p_j, d'_j\}_{j \in N}, t'_0 \rangle$ ,  $d'_j = d_j - \Delta$ ,  $t'_0 = t_0 - \Delta$  where  $\Delta = d_n - \lfloor d_n \rfloor$ ,  $\lfloor d_n \rfloor$  is the integer part of value  $d_n$ . In this case  $d'_n$  is integer value and as we showed above schedule  $\pi_1(d'_n)$  is optimal schedule for instance  $\langle \{p_j, d_j\}_{j \in N}, t_0 \rangle$ .

#### 3.1 Case $k = 1$

In the case  $k = 1$ , parameters of jobs satisfy conditions:

$$\begin{cases} p_1 \geq p_2 \geq \dots \geq p_n \\ d_1 \leq d_2 \leq \dots \leq d_n \\ d_n - d_1 \leq p_n \end{cases} \quad (2)$$

We proved the following main property for the case (2).

**Property B-1.** *Every schedule  $\pi = (\pi_1, i, \pi_2, l, \pi_3, j, \pi_4)$ , where  $l < \min\{i, j\}$ , can be eliminated from consideration without lost of optimality.*

Property B-1 predicates that there exists an optimal schedule  $\pi^*$  where for all  $l = n-1, \dots, 1$

- either  $(l \rightarrow j)_{\pi^*}$  for all  $j \in \{l+1, \dots, n\}$ ,
- or  $(j \rightarrow l)_{\pi^*}$  for all  $j \in \{l+1, \dots, n\}$ .

Based on this property the Algorithm *B-1* of solution the problem in the case (2) are proposed.

---

**Algorithm B-1.** Solution the problem in the case (2).

---

0.  $\pi_n(t) := (n)$ ,  $F_n(t) := \max\{0, p_n + t_0 - t\}$ ;  
1. **for**  $l = n - 1, n - 2, \dots, 1$ :  
 $\pi^1 := (l, \pi_{l+1}(t - p_l))$ ,  $\pi^2 := (\pi_{l+1}(t), l)$ ;  
 $F(\pi^1) := \max\{0, p_l - d_l(t)\} + F_{l+1}(t - p_l)$ ;  
 $F(\pi^2) := F_{l+1}(t) + \max\{0, \sum_{j=l}^n p_j - d_l(t)\}$ ;  
 $F_l(t) := \min\{F(\pi^1), F(\pi^2)\}$ ;  
 $\pi_l(t) := \arg \min\{F(\pi^1), F(\pi^2)\}$ .  
**end for**  
2.  $\pi^* := \pi_1(d_n)$  and  $F(\pi^*) := F_1(d_n)$ .
- 

For each  $l < n$  and  $t$  optimal schedule  $\pi_l(t)$  is the better of schedules  $(l, \pi_{l+1}(t - p_l))$  and  $(\pi_{l+1}(t), l)$ . If  $t \leq t_0 + p_n$  then all jobs are late and  $\pi_l(t)$  is SPT sequence  $(n, n - 1, \dots, l)$ ; if  $t \geq t_0 + \sum_{j=1}^n p_j$  then no more than one job is late and  $\pi_l(t)$  is EDD sequence  $(l, l + 1, \dots, n)$ . Due to SPT and EDD sequences can be computed before algorithm runs, we need to process only integer points  $t : t_0 + p_n < t < t_0 + \sum_{j=1}^n p_j$ . Algorithm B-1 construct schedule  $\pi_1(d_n)$  in  $O(n \sum p_j)$  time.

### 3.2 Case $1 < k < n$

In this case, parameters of jobs satisfy conditions

$$\begin{cases} d_1 \leq d_2 \leq \dots \leq d_n \\ p_1 \geq p_2 \geq \dots \geq p_n \\ d_{\beta_1} - d_{\alpha_1} \leq p_{\beta_1} & \alpha_1 = 1 \\ d_{\beta_2} - d_{\alpha_2} \leq p_{\beta_2} & \alpha_2 = \beta_1 + 1 \\ \dots \\ d_{\beta_k} - d_{\alpha_k} \leq p_{\beta_k} & \beta_k = n \end{cases} \quad (3)$$

**Property B-k/1.** Every schedule  $\pi = (\pi_1, j, \pi_2, l, \pi_3, l+1, \pi_4)$ , where  $j > (l+1)$ , can be eliminated from consideration without lost of optimality.

**Property B-k/2.** Every schedule  $\pi = (\pi_1, i, \pi_2, l, \pi_3, j, \pi_4)$ , where  $l < \min\{i, j\}$  and  $i, j$  belong to the same subset  $M_\nu$ , can be eliminated from consideration without lost of optimality.

Properties B-k/1 and B-k/2 predicate that there exist an optimal schedule  $\pi^*$  where job  $l$  is processed either before all jobs of set  $\{l + 1, \dots, n\}$ , or after job  $(l + 1)$  and is not processed between jobs  $i, j : l < \min\{i, j\}$ , jobs  $i, j$  belong to the same subset  $M_\nu$ .

Based on these properties, Algorithm B-k of solution the problem in the case (3) are constructed. Due to high complication of formal description of the algorithm, we show the main idea. Algorithm B-k starts as algorithm B-1 by



constructing schedules  $\pi_n(t)$ . Then for every  $l = n - 1, n - 1, \dots, 1$  and  $t$ , the algorithm constructs schedules  $\pi^1, \pi^2, \dots, \pi^g$ ,  $g \leq k$ ,  $g$  is count of positions for job  $l$  predicated by the properties in schedule  $\pi_{l+1}(t)$ , and select the best one which is schedule  $\pi_l(t)$ . Algorithm *B-k* constructs schedule  $\pi_1(d_n)$  in  $O(kn \sum p_j)$  time.

## 4 Polynomial algorithms

### 4.1 Case $k = 1$ and $d_n - d_1 \leq 1$

Suppose that parameters of jobs satisfy conditions

$$\begin{cases} d_1 \leq d_2 \leq \dots \leq d_n, \\ d_n - d_1 \leq 1, \\ t_0 \in \mathbb{Z}^+ \end{cases} \quad (4)$$

without restrictions  $p_1 \geq p_2 \geq \dots \geq p_n$ .

In this case we proved the following properties which allow to construct polynomial solution algorithm. Consider sub-instance with jobset  $N' \subseteq N$  and start time  $t_0$ . Let  $S = t_0 + \sum_{j \in N'} p_j$  and  $z = \lfloor d_n \rfloor$  ( $z$  is the integer part of value  $d_n$ ).

**Properties C-1.** *There exists an optimal schedule  $\pi^* \in \Pi(N', t_0)$  such that  $\pi^* = (\pi, j^*)$  and  $\pi \in \Pi^*(N' \setminus \{j^*\}, t_0)$  where*

1. *if  $S - p_j > z$  for all  $j \in N'$  then  $j^* = \arg \max_{j \in N'} \{d_j : p_j = \max_{i \in N'} p_i\}$ ;*
2. *if  $S - p_j \leq z$  for some  $j \in N'$  then  $S - p_{j^*} \leq z + 1$ ;*
3. *if  $S \leq z$  then  $j^* = \arg \max_{j \in N'} \{d_j\}$ .*

---

**Algorithm C-1.** Solution the problem in the case (4).

---

0.  $S := t_0 + \sum_{j=1}^n p_j$ ,  $N' := N$ ,  $\pi^* := \emptyset$ ,  $\pi_D := (1, 2, \dots, n)$ ;
1. **while**  $\{j \in N' : S - p_j \leq z\} = \emptyset$  **and**  $N' \neq \emptyset$  **do**:  
 $\pi^* := (j^*, \pi^*)$ , where  $j^* := \arg \max_{j \in N'} \{d_j : p_j = \max_{i \in N'} p_i\}$ ;  
 $S := S - p_{j^*}$ ,  $N' := N' \setminus \{j^*\}$ ,  $\pi_D := \pi_D \setminus \{j^*\}$ ;  
**end while**
2. **if**  $N' = \{j\}$  **then**  $\pi^* := (j, \pi^*)$ ; **stop**.
3. **for all**  $j \in N' : S - p_j \leq z + 1$  **do**:  
**for all**  $i \in N' \setminus \{j\}$  **do**:  
 $\pi_{ij} := (\pi_D \setminus \{i, j\}, i, j)$ ;  
**end for**  
**end for**
4.  $\pi^* := (\pi, \pi^*)$ , where  $\pi := \arg \min_{i, j} F(\pi_{ij})$ .

---

Algorithm *C-1* starts with jobset  $N$  and due to Properties C-1 on each step determines the job which has to be processed after all unsequenced jobs  $j \in N'$  and before sequenced jobs  $j \in \{\pi^*\}$  in the optimal solution. Algorithm *C-1* constructs an optimal schedule  $\pi^*$  in  $O(n^2)$  time.

## 4.2 Case $k = n$

Consider the case of the problem  $1 \parallel \sum T_j$  where

$$d_j - d_{j-1} > p_j, \quad j = 2, 3, \dots, n, \quad (5)$$

without restrictions  $p_1 \geq p_2 \geq \dots \geq p_n$  and  $p_j \in \mathbb{Z}^+$ .

In this case we present polynomial algorithm based on the following property.

Consider sub-instance with jobset  $N' \subseteq N$  and start time  $t' \geq t_0$ ,  $N' = \{j_1, j_2, \dots, j_m\}$  and  $d_{j_1} \leq d_{j_2} \leq \dots \leq d_{j_m}$ . Let  $j^* = \arg \max_{j \in N'} \{d_j : p_j = \max_{i \in N'} p_i\}$  and  $S_\alpha := t' + p_{j_1} + p_{j_2} + \dots + p_{j_\alpha}$ ,  $\alpha = 1, 2, \dots, m$ .

**Property B-n.** *There exists an optimal schedule  $\pi^* \in \Pi^*(N', t')$  such that*

$$(j \rightarrow j^*)_{\pi^*} \text{ for all } j \in \{j_1, j_2, \dots, j_{\alpha^*}\} \setminus \{j^*\} \text{ and} \\ (j^* \rightarrow j)_{\pi^*} \text{ for all } j \in \{j_{\alpha^*+1}, \dots, j_m\}.$$

where  $\alpha^* = \min\{\alpha \in \{1, \dots, m\} : p_j + d_j \leq S_\alpha < d_{j_{\alpha+1}}, j \in \{j^* + 1, \dots, j_\alpha\}\}$ , in addition  $d_{j_{m+1}} = +\infty$ .

In the notation of decomposition approaches [2–7], Property B-n means that there exists an optimal schedule  $\pi^*$  where the largest processing time job  $j^*$  is processed on the first "suitable" position  $\alpha^*$  and the sub-instance  $N', t'$  is decomposed by the position  $\alpha^*$ .

---

### Procedure ProcB-n ( $N', t'$ )

---

0. Given set of jobs  $N' \subseteq N$ ,  $N' = \{j_1, j_2, \dots, j_m\}$  and start time  $t' \geq t_0$ ,  $d_{j_1} \leq d_{j_2} \leq \dots \leq d_{j_m}$ ;
  1.  $S_\alpha := t' + p_{j_1} + p_{j_2} + \dots + p_{j_\alpha}$ ,  $\alpha = 1, 2, \dots, m$ ;  
 $j^* = \arg \max_{j \in N'} \{d_j : p_j = \max_{i \in N'} p_i\}$ ;
  2.  $\alpha^* := \min_{\alpha=1, \dots, m} \{\alpha : d_j + p_j \leq S_\alpha < d_{j_{\alpha+1}}, j \in \{j^* + 1, \dots, j_\alpha\}\}$ ;
  3.  $N_1 := \{j_1, \dots, j_{\alpha^*}\} \setminus \{j^*\}$ ,  $t_1 := t'$ ;  
 $N_2 := \{j_{\alpha^*+1}, \dots, j_m\}$ ,  $t_2 := S_{\alpha^*}$ ;
  4. Return schedule  $\pi^* := (\pi_1, j^*, \pi_2)$  where  
 $\pi_1 := \mathbf{ProcB-n}(N_1, t_1)$ ,  
 $\pi_2 := \mathbf{ProcB-n}(N_2, t_2)$ .
- 

**Algorithm B-n.** Solution the problem in the case (5).

---

$$\pi^* := \mathbf{ProcB-n}(N, t_0)$$


---

Algorithm B-n constructs an optimal schedule  $\pi^*$  in  $O(n^2)$  time.

## 5 Conclusion

In the paper we consider NP-hard in the ordinary sense single-machine total tardiness scheduling problem. The complex research of the case (1) has been made. Four algorithms are proposed to solve all possible sub-cases of the case with no more than  $O(n^2 \sum p_j)$  time. In the future we will study the hypothesis that the general case of the problem (without restrictions (1)) can be solved with no more than  $O(n^2 \sum p_j)$  time (that is essentially less than in the Lawler algorithm).

## References

1. J. Du and J. Y.-T. Leung: Minimizing total tardiness on one processor is NP-hard. *Math. Oper. Res.*, **15** (1990), pp. 483-495.
2. E.L. Lawler: A pseudopolynomial algorithm for sequencing jobs to minimize total tardiness *Ann. Discrete Math.*, **1** (1977), pp. 331-342.
3. C.N. Potts and L.N. Van Wassenhove: A decomposition algorithm for the single machine total tardiness problem, *Oper., Res. Lett.*, **1** (1982), pp. 177-182.
4. S. Chang, Q. Lu, G. Tang, and W. Yu: On decomposition of the total tardiness problem, *Oper. Res. Lett.*, **17** (1995), pp. 221-229.
5. W. Szwarc, F. Della Croce and A. Grosso: Solution of the single machine total tardiness problem. *J. Sched.*, **2** (1999), pp.55-71.
6. W. Szwarc, A. Grosso and F. Della Croce: Algorithmic paradoxes of the single machine total tardiness problem. *J. Sched.*, **4** (2001), pp. 93-104.
7. A.A. Lazarev: Efficient algorithms of decisions some problems of theory of scheduling for single machine with deadline terms of service jobs. PhD Thesis, 1989, (in Russian).
8. H. Emmons: One machine sequencing to minimizing certain function of job tardiness. *Oper. Res.*, **17** (1969), pp. 701-715.
9. F. Della Croce, A. Grosso, V. Paschos: Lower bounds on the approximation ratios of leading heuristics for the single-machine total tardiness problem. *J. Sched.* **7** (2004), pp. 85-91.

# Dynamic Analysis and Trace Simulation for Data Parallel Programs in the ParJava Environment.\*

Victor Ivannikov, Sergey Gaissaryan, Arutyun Avetisyan, Vartan Padaryan, and Hennadiy Leontyev

Institute for System Programming Russian Academy of Sciences, 25 B.  
Kommunisticheskaya st., Moscow, 109004, Russia  
{ivan, ssg, arut, vartan, gena}@ispras.ru  
<http://www.ispras.ru>

**Abstract.** Program profiles allow to define sequence of models simulating the traces of a data parallel Java programs with various levels of precision. The simulation of a data parallel program's trace is called symbolic execution. It allows to analyze interposes communications of a data parallel program and evaluate its execution time, when given number of processor nodes is used. Evaluation of interposes communications helps to perform additional tuning of a program to detect and remove possible bottlenecks preventing to achieve needed level of its scalability. Another use of dynamic analysis and symbolic execution results is the rollback recovery tool based on coordinated check-pointing algorithms. The algorithms were improved in order to apply them to high performance computations.

## 1 Introduction

There is a great deal of interest in using Java for parallel computing. The most part of current research is devoted mainly to MPI interface implementations for Java [1, 2].

When a parallel program is developed it is necessary to assure not only its correctness, but also its efficiency and scalability. For this purpose it is useful to know some dynamic properties of the program (profiles, traces, slices, etc.). This information may help to modify the program in order to improve its parallel features. Profiles and traces of the program show the order and the interference of its interprocess communications.

Model of a data parallel program providing its symbolic execution may be used in checkpointing algorithms as well.

Java language and JVM specification do not provide any explicit recovery facilities. JVMTI interface [3] recently published by SUN allows to create some facilities analyzing and controlling the states of JVM. These facilities can be used for saving and recovering sequential Java-programs. There are some well-known techniques which allow to apply such facility to parallel programs as well [4].

---

\* This work is supported by RFBR, grants 02-01-00961, 02-07-90302

Occurrence of an advanced environment allows to increase the efficiency of a parallel program using other facilities integrated in the environment.

The paper has the following structure. In section 2 we give description of facilities supporting modeling execution of data-parallel program in ParJava Environment. In section 3 the aim of the recovery protocols is discussed. Section 4 presents description of checkpointing algorithm. In section 5 we compare the related systems with corresponding facilities. In section 6 some conclusions are made.

## 2 Symbolic execution of a data-parallel program

Investigation of dynamic properties of a parallel program (communication profile, traces, etc.) requires repeated execution of the program using various sets of initial data and various sets of processor nodes, in order to gather statistics. The advantage of symbolic execution against actual execution of a parallel program is that the first one can be performed much faster and using only one processor.

Symbolic execution allows to estimate limits of scalability and hence to calculate expected execution time for a program using various numbers of processor nodes during the stage of its development, using dynamic properties of the parallel program.

We define several models of symbolic execution, which have various precisions in ParJava. Changing execution model we can change the time necessary for symbolic execution. The most precise model uses frequency and time profiles obtained using typical input data (the input data of a parallel program is called typical if the trace determined by this input data contains all significant statements of the parallel program). We assume that user has idea of typical input data of his parallel program. The profiles should be calculated using actual execution of the program on a parallel computer system. However, less precise models using profiles obtained by the execution of the program using one processor and model input data are defined. Use of such "rough" models allows to reduce program development overheads.

As mentioned above symbolic execution of a parallel program needs its exact frequency and time profiles. To define frequency (or time) profile we use the notion of a *program step* (we call a *step of a program* calculation of one of its expressions or execution of one of its basic blocks). The following types of basic blocks are considered in ParJava: sequence of assignments, condition, function call, and communication function call. Frequency (or time) profile of a sequential program is defined as mapping which associates each step of the parallel program with the frequency (or time) of its execution. The profile of a sequential program may be represented as a vector with dimension equals to number of program steps. The profile of the SPMD-program is represented by a matrix whose lines are profiles of this program's parts executed in parallel.

Symbolic execution allows to estimate limits of scalability and hence to calculate the expected execution time for the program during the stage of its development. The advantage of symbolic execution against actual execution is that the symbolic execution can be performed much faster and using only one processor.

Changing the precision of execution model we can change the time necessary for symbolic execution. There are defined several execution models of various precisions

in ParJava. The most precise model uses frequency and time profiles obtained using typical input data. These profiles must be calculated using actual execution of the program on parallel computer system. However, less precise models using profiles obtained by the execution of the program using one processor and model input data are defined. Use of “rough” models allows to reduce program development overheads.

Symbolic execution is performed by the *symbolic interpreter*, which uses: the control flow graph of a parallel program being executed, its dependency graph, and its frequency and time profiles. During the symbolic execution of the program its basic blocks that lie on the path defined by the specified input data, are interpreted symbolically. The symbolic interpretation of each basic block is performed by interpreting the assignment statements included in this block, calculating the transfer condition, and selecting the basic block which should be executed next. The execution starts from the entry block and ends with the exit block of the program. Function calls as well as the calls to communication functions form separate basic blocks. When the basic blocks containing calls to communication functions are executed the a priori information about the communication subsystem is used. The information allows to determine the duration of transfer of  $n$  bytes from one processor node to another. Such information for communication subsystems on the basis of Ethernet and Myrinet is inputted into the system, and the rules of such information input for other communication subsystems are specified. Transfer time is estimated according to the model LogGP [5] of computer network. There is opportunity to use any other communication model. If user wants to use communication model other than LogGP he must add to symbolic executor the library according to the model he is going to use.

The interpreter allows the quick interpretation of the program executed on the cluster on one processor that significantly shortens the time necessary for its development and debug.

Symbolic execution allows to estimate program execution time for each number of cluster nodes as well as maximal number of nodes for which scalability is hold.

### **3 Recovery algorithms in ParJava**

Problem of reliability assurance of long-time-working applications is urgent for parallel programs, because node failure brings application to restarting from beginning. Wide spread cluster decisions often base on common PC, in the role of computing nodes, that decrease fault tolerance of whole system. Therefore that class of computing systems especially needs facilities supporting restoration system from intermediate point of execution, where crash took place. Applying rollback-recovery protocols can solve that problem.

Unlike sequential, parallel programs make substantial demand to checkpointing systems – saved data should present consistent state of parallel program. Also they should save state of communication environment – messages, which were sent, but weren't received before failure, should be saved and then restored. Notion of consistent state is considered in detail in works [6].

It is known three classes of checkpoint-based recovery algorithms [4], which allow save consistent states: *uncoordinated*, *coordinated*, *communication-induced*.

Some types of parallel programs don't accept some classes of algorithm. In this case, scientific calculating applications are considered. They occupy all accessible memory on computational node (approx. 1Gb at common cluster, at the present moment), works on large quantity of nodes and actively transmit messages each other. Uncoordinated algorithm stores multiple of  $N^2$  checkpoints, where  $N$  – the total number of processes in a system. It makes up huge amount of data for such type of application.

There are two (rather independent) criteria for the choice of the program area holding a checkpoint: (1) achievement of integrity of the program state being stored; (2) minimization of volume of data representing the state of a parallel program being stored.

The first criterion is satisfied if coordinated protocols are used. In that case we minimize the total number of checkpoints. To satisfy the second criterion it is necessary to choose such program area within which comparatively small amount of data is actual. The checkpointing method with coordinated protocol may choose a place for a checkpoint within a parallel loop, according to the first criterion – achievement of integrity of the program state, which is stored. In such a case a large amount of data being actual only inside the loop will be stored, so the second criterion is not satisfied. It is evident that in aspect of the second criterion areas before or after a loop are more appropriate for checkpoints, because in such cases it is not necessary to store variables used for temporary data within the loop.

Thus, it seems useful to choose a place for a checkpoint in such program areas where the amount of used storage (size of JavaVM heap) is minimal (or at least less than in other areas). Such areas may be detected analyzing the structure of the control-flow graph using symbolic execution. Detection of such areas is able to increase the efficiency of the coordinated protocol [7]. In existing classification [4] such modified protocol may be related to communication-induced model-based.

The proposed analysis of a parallel program is based on JavaVM heap size profiling. As parallel program profiling using target computer system is inefficient due resource demands, we perform profiling of a model of the program using symbolic interpreter.

The analysis of the dynamics of used memory may be divided into the following two problems: (1) keeping track of the scope of each variable, and (2) keeping track of the amount of storage attached to the variable.

The first problem is solved by extending the notion of Java-method. Evidently, the lifetime of the names of variables defined within a loop does not exceed the lifetime of the loop. We add to control-flow graph structure the blocks representing extension/reduction of the context (they are placed consequently before the entry and after the exit of a loop). It makes possible precise definition of the scope bounds of each variable. In most cases the storage allocated to such variable is released at the exit of a loop. The storage may stay allocated if the reference to it was passed out of the loop.

Such situations are the subject of the second problem. Variables of primitive Java-types are processed, as well as special reference type, containing reference to auxiliary object describing the amount of storage occupied. In that case the modeling

of reference assignment is come to the change of number of references to such object-descriptor. When the number of references becomes equal to zero, storage is freed. The similar technique is used in garbage collector of JavaVM.

## 4 Checkpointing algorithm

The proposed variant of the algorithm considers the separate process of the parallel program. Several restrictions are imposed on the program: the program should contain only one thread and should not use native methods (the user part of code is meant). Let us consider that parallel program is already analyzed and intervals where it is possible to save the state of the process are discovered. The first stage consists in automatic source to source transformation: the instrumental insertions – calls of `ChPt` method of save/restore subsystem are added and several calls of user's methods are transformed as well. If user method is called, which directly or indirectly contains call of method `ChPt` and that call of is part of composite expression, then that expression (main expression) decomposed into three statement expressions. The first one contains calculations that are located “before” call of user method in main expression. Result of sub expression stores in temporary variable. Second part is call of user method and its result is also stores into temporary variable. And the third part is the main expression, where sub expressions isolated into first and second part are substituted by temporary variable.

The result of such code transformation is then method `ChPt` loads into stack of JVM, stack doesn't contain any intermediate results of expressions being calculated. So, all valid data are stored into method local variables and JVM heap.

Then program is started, special flag `STATUS` located in the run-time subsystem takes on a value `RUN`. Then execution reaches method `ChPt` and value of flag `STATUS` equals `RUN`, checkpoint is created.

Process of creating of checkpoint consists of two stages: storing stack of method calls and storing subset of objects of JVM heap.

Each element of JVM stack contains index of current instruction of byte-code (value of PC register then next method was called) and values of local variables. Stack is stored separately from JVM heap into XML-based text file that render to user ability of analysis and editing of program state. While stack storing each variable, which has object reference, transmit value to service structure. Then XML-representation of stack is prepared, such variable will be contain a reference to service structure. Hereinafter we will refer to XML-representation as checkpoint stack.

After JVM stack, data from JVM heap are dumped. Store subsystem serializes service structure, mentioned earlier, by general algorithm provided JVM. This supports dumping only reachable from live local variables objects.

While program restoring next actions are executed. Byte-code of methods located in checkpoint stack is instrumented. Two types of insertions are placed into byte-code. In beginning of methods *trampolines* are placed, portion of code executed then `STATUS=RESTORE`. *Trampolines* jump control flow to specified point of byte-code. If program doesn't contain recursive methods, each method is included in stack only once and there is only one jump address in trampoline. In other case, some methods



are presented in stack several times and trampoline should have jump array. For each recursive method special static counter is created. Jump addresses are sorted in array by order of entries of stack. Then execution passes trampoline counter increases by one.

Second type of insertions is located directly before next method call. At that method call shifts in byte code. Insertion restores values of local variables of current method frame and also snap into action only if `STATUS=RESTORE`.

While restoring program starts with loading of JVM heap dump. Values of local variables of simple data types are restored by data directly placed in checkpoint stack. Variables refers to object obtain their values from restored by standard deserialization service data structure, mentioned earlier.

Then restore subsystem sets flag `STATUS` into value `RESTORE` and calls lowest method of checkpoint stack. Thus, execution reaches quick enough call of method `ChPt` – work of processed methods consist just of jumps and restoring of local variables. If execution reaches method `ChPt` and `STATUS=RESTORE` this method only changes value of flag into `RUN`. Then returns from `ChPt` application continue its normal work.

The information about instrumental code arranged in the program (position in bytecode and size) is recorded in the descriptor stored together with a checkpoint. When repeated recovery is performed these data are used on the stage of bytecode instrumentation. All insertions made before are removed, and the stack of a checkpoint, containing data considering the insertions is modified in corresponding way. Then the same actions as during the first recovery are performed.

Consider an example that demonstrates described above tools. Parallel program calculates heat transmission in homogeneous stick. It uses explicit scheme and works over matrix 3000x3000 elements at each process of parallel program. Data grows up with processors being used in such manner that program uses 80% memory of calculating node. Table 1 shows precision of symbolic execution and resources it consumes. The computations were performed using ISP RAS cluster with following parameters: 8 dual Athlon 1533 MHz having 1 Gb memory, communication network Myrinet, full duplex hypercube, bandwidth 2 Gb/sec, control network – Fast Ethernet. Instrumental machine is Pentium IV 2666 MHz with 512 Mb memory.

**Table 1.** Precision of symbolic execution.

Processors	2	4	6	8	10	12	14	16
Real exec. (sec.)	569	584	584	590	592	590	593	595
Symbolic exec. (sec.)	571	583	584	586	590	591	591	594
Consumed time (sec.)	1	3	3	5	7	7	9	11

Also, we measured overhead produced by checkpointing (Table 2). Program runs over one calculating node with variable dimension of matrix. Recovery subsystem produced checkpoint every 30 seconds. Despite of frequent dumping, checkpointing overhead restricts to 5%.

**Table 2.** Checkpointing overhead.

Data size (dimension of matrix)	1000	1500	2000	2500	3000	3500
With checkpointing (sec.)	63	143	252	403	586	810
Without checkpointing (sec.)	62	142	249	395	569	778

## 5 Related Works

At the present moment, most of checkpointing systems exist as subpart of implementation of some communication library, such as Lam/MPI [8].

Lam/MPI is widespread communication library, which implements MPI. It's included in Red Hat Linux distributive packages. Begins from version, Lam/MPI contains facilities to recovery parallel program after crash, based on Berkeley Lab Checkpoint/Restart (BLCR)[9].

Another work [7] investigates in general principles of placing checkpoints in program, but doesn't consider technology of state saving. Offered method works in programming languages, where arrays are first class citizens, for example Fortran 90. Special tool reveals program's interval, which are optimal for checkpointing. These intervals contain minimal quantity of arrays. Revealing uses results of live variables and control flow analysis. Used technique protects parallel loop nests from checkpointing, and saves up to 73% of time, spent to process checkpoint.

All well-known integrated environments for developing parallel programs, support programming for Fortran and C languages. We may emphasize such environments as AIMS (NASA Advanced Supercomputing Division) [10], TAU (University of Oregon) [11], and DiP (Universitat Politècnica de Catalunya) [12].

The above-mentioned systems employ post-mortem analysis of the program. These systems provide wide opportunities for trace viewing, gathering of parallel applications profiles and operating with statistics. The similar tools are realized in the ParJava environment.

AIMS is the environment most similar to ParJava. AIMS, the same as ParJava, provides the symbolic execution of the SPMD-program. But during the instrumenting of the program in AIMS, unlike in ParJava, the quantity and position optimization of calls to instrumental functions is not executed. The interpreting program is presented internally in AIMS in BDL language in which the presentation of the program is compiled from the blocks of three kinds: "cycle", "communication" and "sequence". In the ParJava environment the program is presented by the control flow graph that supplies a more adequate interpretation. Moreover, in the ParJava environment the mechanism that gives an opportunity to control the process of the symbolic execution is realized. It is also possible to gather the trace of the symbolically executed program, to monitor the states of the program parts executed in parallel and to determine the deadlock automatically.

## 6 Conclusion

The ParJava environment enables to develop portable scalable SPMD Java-programs, providing the application programmer with software tools that help him to investigate expected execution time, scalability limits and some other properties of the parallel Java-program in development time. The ParJava environment provides some software tools that help application programmer to improve his parallel program properties.

Now, checkpoint recovery protocol is released as prototype in ParJava environment. Coordinated usage of tools can decrease development and debugging time of parallel program on unreliable computing systems. Applying of program analysis as static, as dynamic allows to making checkpointing overhead small enough.

ParJava is installed on several clusters: ISP RAS, Joint Supercomputer Center (Moscow), and Moscow University. Presently some applications (flood modeling [13]) are being developed using ParJava.

## References

1. S. Mintchev. Writing Programs in JavaMPI. //TR MAN-CSPE-02, Univ. of Westminster, London, UK, 1997.
2. Tong WeiQin, Ye Hua, Yao WenSheng. PJMPI: pure Java implementation of MPI. //Proceedings of the 4th International Conference on High Performance Computing in the Asia-Pacific Region, 2000.
3. JavaTM Virtual Machine Tool Interface (JVMTI). <http://java.sun.com/j2se/1.5.0/docs/guide/jvmti/index.html>
4. M. Elnozahy and L. Alvisi and Y.-M. Wang and D.-B. Johnson. A survey of rollback-recovery protocols in message passing systems. / Technical report, Carnegie Mellon University, CMU-CS-96-181. October 1996.
5. R. Martin, A. Vahdat, D. Culler, T. Anderson. The Effects of Latency, Overhead and Bandwidth in a Cluster of Workstations. //In Proc. 24th Int. Symp. on Com. Arch. (ISCA'97), June 2 - 4, 1997, pp 85 - 97.
6. O.Babaoglu and K. Marzullo. Consistent global states of distributed systems: Fundamental Concepts and mechanisms. / Distributed Systems, Addison Wesley pp.55-96 1993.
7. Sung-Eun Choi and Steven J. Deitz. Compiler Support for Automatic Checkpointing. // In Proceedings of the 16th Annual International Symposium on High Performance Computing Systems and Applications, 2002.
8. LAM/MPI Parallel Computing. <http://www.lam-mpi.org/>.
9. Berkeley Lab Checkpoint/Restart (BLCR). <http://fig.lbl.gov/checkpoint/>.
10. J. C. Yan and S. R. Sarukkai. Analyzing Parallel Program Performance Using Normalized Performance Indices and Trace Transformation Techniques. //Parallel Computing. Vol. 22, No. 9, November 1996. pages 1215-1237
11. S. Shende, A. D. Malony, J. Cuny, K. Lindlan, P. Beckman and S. Karmesin. Portable Profiling and Tracing for Parallel Scientific Applications using C++. //Proc. SPDT'98: ACM SIGMETRICS Symposium on Parallel and Distributed Tools, pp. 134-145, Aug. 1998.
12. J. Labarta, S. Girona, V. Pillet, T. Cortes, and L. Gregoris. Dip: A parallel program development environment. // In Euro-Par, Vol. II, pages 665-674, 1996.
13. L. Hluchy, V.D. Tran, D. Froehlich, and W. Castaings, Methods and Experiences of Parallelizing Flood Models, The 10th EuroPVM/MPI conference. LNCS 2840. Sept. 2003, Venice. pp. 677-681.

# Clusterweb – a WEB-based cluster management interface

Arutyun Avetisyan, Oleg Samovarov, Dmitry Grushin, Andrey Ryzhov

Institute for System Programming Russian Academy of Sciences, 25 B.  
Kommunisticheskaya st., Moscow, 109004, Russia  
{arut, samov, grushin, ryzhov}@ispras.ru

**Abstract.** Clusterweb - a WEB-based interface to cluster management systems is discussed. Each cluster management system has many command line options, scriptable parameters, and tools, which are very difficult to learn for the typical user. WEB-based interface simplifies job submission and management as well as access to cluster from non-UNIX environments. Typical job management includes file uploads and downloads which are usually performed using FTP protocol being not secure, so for secure file transfers additional software must be installed. WEB-based interface does not require any additional software and can be used from any HTML compliant browser. All data transfers can be made secure if server and client browser support SSL. WEB-based interface to cluster can retain information concerning uploaded job files and used command scripts, which could be very convenient for all users accessing cluster. Moreover, WEB-based interface can support different user customizations and have nice graphical look. Clusterweb is implemented using Java and XSL stylesheets on the top of Tomcat servlet container. Tomcat is free software and can be easily integrated into the Apache Web server. Interface features include job management, command file management and user settings management. The interface can be used with different cluster management systems, which have command line interface or can be accessed through API services. Future versions may include integration with Grid environments.

## 1. Introduction

Clusterweb is a WEB based interface to cluster management systems. Many cluster management systems have a lot of command line options, scriptable parameters, and different tools, which are very difficult to learn for the typical user. WEB-based interface simplifies job submission and management. It also could be very useful for users that access cluster from non-UNIX environments, for example Windows and Mac machines.

Typical job management includes file uploads and downloads which are usually performed using FTP protocol, which is not secure. Simple FTP client exists in all major OS distributions, on the other hand for secure file transfers additional software

must be installed. From this point of view WEB interface does not require any additional software and can be used from any HTML compliant browser. All data transfers can be made secure if server and client browser support SSL.

Besides this, WEB-based interface to cluster can retain information concerning uploaded job files and used command scripts, which could be very convenient for all users accessing cluster. Moreover, WEB-based interface can support different user customizations and have nice graphical look.

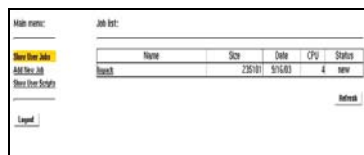
Clusterweb is implemented using Java and XSL stylesheets on the top of Tomcat servlet container. Tomcat is free software and can be easily integrated into the Apache Web server. Interface features include job management, command file management and user settings management.

The interface can be used with different cluster management systems, which have command line interface or can be accessed through API services. Future versions may include integration with external systems, i.e. Grids.

## 2. Clusterweb features

Clusterweb supports submission and management of parallel jobs to cluster by means of communicating with underlying management system. Clusterweb provides higher level of abstraction for users who submit and manage jobs, then using a command line based interface. Each job is represented as a number of files, compressed in a single ZIP archive<sup>1</sup>. Job has a name, size, date, number of CPU and status attributes. Job may have other specific attributes, which depend only on the target management system.

Every user has a virtual home directory, which holds all submitted jobs. User can upload new job, delete or restart already finished jobs. User's home page is shown on the figure below.



Name	Size	Date	CPU	Status
test	23510	01/16/03	4	new

**Fig. 1.** User's home page

Home page contains main menu and a list of jobs. Name is an ordinary label for a job; therefore, two jobs with the same name may exist in the list. Each job is represented by a single ZIP archive file, which contains all necessary data for job execution on a target parallel machine. The size attribute shows the size of the corresponding job archive file in bytes. Date attribute shows the date when the corresponding job has been added to the list. CPU is the number of processing

---

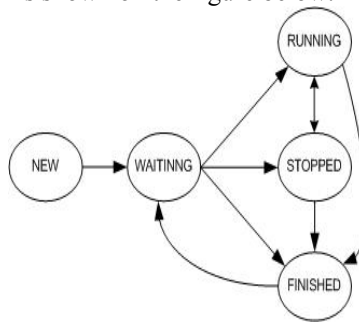
<sup>1</sup> ZIP compression is quite fast and many users already have ZIP compatible programs that can open .zip files.

elements required by the job. There are five possible values for the status code; all of them are shown in the table below.

**Table 1.** Job status codes

New	Job has been submitted to the list.
Waiting	Job has been started and is waiting for its execution time.
Running	Job is running.
Stopped	Job has been stopped.
Finished	Job finished.

Job state transition graph is shown on the figure below.



**Fig. 2.** Job state transition graph

Submission of new jobs is performed in two steps. First, user should create a single ZIP archive file, with all necessary data for job execution. This file will be further decompressed by the system into temporarily directory on a target machine. The temporarily directory will be set as a current directory for the job. That means users should avoid using absolute path names in submitted programs. The second step is to create a command file for the job.

Command file depends on the underlying management system; hence, creation of the command file may be troublesome for end users. Clusterweb can automatically create a command file for a job based on job's specific attributes. Command file may have several versions to support different cluster management systems. Clusterweb chooses an appropriate command file automatically, each time job is executed on a target cluster machine. This functionality will be implemented in upcoming version of Clusterweb. At present job archive file should contain a file named start, which is a command file for the job.

New job wizard is accessible through the "Add New Job" menu. An example page is shown on the figure below.

**Fig. 3.** New job submission

To submit a new job, user should enter name, description, required number of processing elements and job archive file. Job description is optional. Job archive file should contain all data necessary for job execution and a command file named start. If the file does not exist, the user should assign command file to the job manually. Job can be started in two ways: either automatically after upload, or manually later. By default, starting time for a job is manual.

Before starting the job on a target machine, the system will decompress job archive file into a temporarily directory. After decompression, the system will try to execute (in terms of the underlying management system) file with the name start. Job standard output and error streams should be redirected to files named stdout and stderr correspondingly. If the system fails to find these files, the contents of the corresponding output streams will not be available.

User can manage command files used for job execution. Command files are called scripts<sup>2</sup>. The scripts are organized in groups. User can add or delete script groups. User can delete only empty groups and scripts, which are not assigned to jobs. Each script has name, size and date attributes. Two scripts with the same name may exist in the list. Size attribute identifies the size of the script in bytes. Date attribute shows the date when the corresponding script has been added to the list. During job upload user can create a new script or select some script from the list and edit it if necessary.

Job archive file may contain executable programs or source codes, which need to be compiled before job execution. In this case, all necessary commands should be added to the start script.

---

<sup>2</sup> This is a script file in terms of the underlying management system, but it may be not shell-compatible.

### 3. Clusterweb architecture

Clusterweb system is built using Java 2 frameworks and services for business logic components and Servlet API for front-end on the top of the Apache Tomcat 4.1 servlet container.

Clusterweb architecture is a variation of the Model-View-Controller (MVC) design pattern [5]. MVC organizes an interactive application into three separate parts: one for the application model with its data representation and business logic, the second for views that provide data presentation and user input, and the third for a controller to dispatch requests and control flow.

The MVC design pattern provides a host of design benefits. MVC separates design concerns (data persistence and behavior, presentation, and control), decreasing code duplication, centralizing control, and making the application more easily modifiable. MVC also helps developers with different skill sets to focus on their core skills and collaborate through clearly defined interfaces [5].

The system includes three layers: Presentation Layer (View), Business Layer (Model) and Workflow Layer (Controller).

The purpose of the Presentation Layer is to provide a mechanism to create presentation documents (HTML) with the following goals kept in mind.

The system must be able to handle a number of locale-specific versions of each page. The system must choose the most appropriate page version transparently to the underlying processing layers

The stylesheets must be completely separate from the application logic, so that it were possible to modify the layout, perform localization and otherwise edit stylesheets without any impact on the code. Such separations of concerns implies that people without Java programming skills can customize layout as far as it doesn't affect screen navigation, essential data passed and other functional aspects.

Although the current version does not deal with presentation media other than HTML, it is possible that the forthcoming versions will. Hence, it should be possible to add another presentation media with no impact on core business logic and minimal impact on other layers.

In order to satisfy the goals above, XML/XSL-based approach was chosen as an underlying technology for the Presentation Layer. Presentation layer is responsible for rendering of the XML documents created by the underlying layer using the most appropriate chosen stylesheet. Both Presentation and Business layers coexist within Workflow layer. Workflow Layer is designed to keep Business Objects in the Business Layer free from state management and interaction with the Presentation Layer. Its responsibility is to convert user actions into calls to stateless transactional Business Objects, gather data and prepare XML documents to be rendered by the Presentation Layer, which is summarized on the picture below.



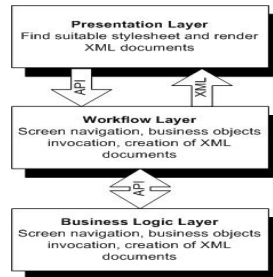


Fig 4. Clusterweb layers

Workflow Layer is responsible for screen navigation, i.e. it decides what screen to show next according to user action. Unlike the Presentation, Workflow Layer includes both framework parts and pluggable pieces of code, which implement functionality specific to particular pages and page groups. Workflow Layer also works as a “defense perimeter”: this is where the request parameters are parsed, validated, and permissions are checked. The calls from Workflow Layer to Business Objects must ensure that according to a particular Business Object contract, the call will not violate the security constraints.

Workflow layer corresponds to a common Three-Tier Architecture. The main servlet called WFServlet, analyzes incoming request and extracts the value of request parameter, which holds page identifier. Based on the extracted id, WFServlet invokes appropriate request and response handlers based on user-computable conditions. The picture below illustrates this request-response processing.

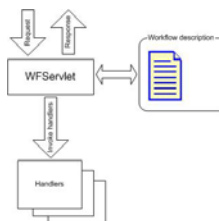


Figure 5. Workflow details

Handlers modify single XML document (DOM object), which is used for constructing response. Each handler is independent from others. All interaction between handlers is performed through conditions and HTTP session.

#### 4. Interaction with cluster management system

To start working with Clusterweb user should supply a login name and a password. Clusterweb has a stand-alone access list, which is checked for existence of the provided user name and password.

In the current version, registration of new user is performed by submission of request form, which is checked by the system and send to a system administrator.

After successful authentication user can start working with Clusterweb. When the user submits a new job, the system opens Secure Shell (SSH) connection to the desired cluster server and executes corresponding command line utilities. All necessary files are transferred with SCP utility. This approach requires that the user under which Tomcat server is run should have public key SSH access to the cluster machine with the desired user id. Hence, Tomcat SSH public identity key should be copied to the corresponding user's `authorized_keys` file on the cluster machine. When Clusterweb does a secure shell login into the user's system account it creates working directory named `clusterweb`. Each uploaded job is given a subdirectory inside this directory. When the job finishes, all related files are removed from the directory.

## 5. Related work

There are many WEB-based interfaces designed to help cluster management, but most part of them can be used only for cluster monitoring than for submitting user jobs [1, 2, 4]. Such systems are designed to provide flexible WEB-based interface and tools for secure cluster administration. The systems include automated alarms and remote notification through email or SMS alerts, up-to-the minute statistics of the cluster status, and historical resource utilization data [4].

PBSWeb is a simple WEB-based interface to the Portable Batch System (PBS), which supports both local and remote users, maintains a simple history database of past job parameters, and hides much of the complexities of the underlying scheduler [3]. PBSWeb is implemented using CGI Perl scripts on the top of the Apache server.

## References

1. Ganglia Toolkit, <http://ganglia.sourceforge.net>
2. ClusterWorX 2.0, Linux NetworX, <http://www.linuxnetworx.com>
3. George Ma and Paul Lu. "PBSWeb: A Web-based Interface to the Portable Batch System", 12th IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS), Las Vegas, Nevada, U.S.A., November 6-9, 2000, pp. 24-30
4. The Cluster Management System, Streamline Computing, <http://www.streamline-computing.com>
5. Sun BluePrints, "Designing Enterprise Applications with the J2EETM Platform, Second Edition", chapter 11



# **Robótica móvil**

Angélica Muñoz Meléndez  
Victor Ayala Ramírez  
(Editores)



## Introducción

La robótica como disciplina de estudio ha recibido en México un apoyo limitado. Son muy pocos los proyectos de investigación que reciben algún tipo de financiamiento de parte de algún organismo diferente a las universidades y a los centros de investigación<sup>1</sup>.

Esta falta de apoyo puede constatarse no sólo en la atribución de recursos, sino también en el número de grupos de investigación y de empresas trabajando en el área. El directorio de robótica y automatización de México<sup>2</sup> y las asociaciones mexicanas de robótica y de mecatrónica reúnen, respectivamente, a siete empresas mexicanas interesadas en el desarrollo y en la venta de sistemas robóticos industriales y educativos, y a profesionales de no más de una veintena de institutos y de universidades de todo el país.

Los grupos mexicanos que trabajan en robótica sufren también de la dispersión de esfuerzos. Prueba de ello es, por mencionar sólo un ejemplo, la multitud de concursos y torneos que, de manera independiente, se organizan en diversas ciudades.

De acuerdo a este panorama, no parece equivocado deducir que en México se invierte muy poco en la robótica y que, lo poco que se invierte, se fragmenta en esfuerzos individuales. Si ya es desafortunado que los centros de investigación, las empresas y las universidades no estemos capitalizando nuestras experiencias, lo es más que no conjuntemos nuestros esfuerzos por alcanzar objetivos comunes.

Con estas inquietudes en mente nos dimos a la tarea de organizar un evento para reunir a la comunidad de roboticistas mexicanos. Nuestra motivación fue promover la conformación de una comunidad de roboticistas sólida: que se conoce, se reúne periódicamente a discutir sus trabajos, y es capaz de organizar y de colaborar en proyectos conjuntos.

Los objetivos específicos que nos trazamos con la organización de este taller fueron los siguientes:

- Crear un foro para el intercambio de ideas entre investigadores, estudiantes y profesionales trabajando en el área de robótica móvil en México.
- Evaluar la situación de la investigación en el área.
- Proponer líneas de acción conjuntas, como pueden ser la organización de eventos nacionales, talleres y simposia, que ayuden a difundir el trabajo desarrollado por roboticistas mexicanos.
- Servir de catalizador de nuevas alianzas de cooperación académica-industrial en el área de robótica móvil.

Para conformar el taller, invitamos a la comunidad trabajando en el área de robótica móvil en México a enviar artículos técnicos describiendo resultados de su investigación, los cuales fueron sometidos a la evaluación de un comité de programa inte-

---

<sup>1</sup> De 2002 a la fecha, por ejemplo, sólo se ha reportado la aprobación de un proyecto directamente relacionado con la robótica, en el marco de los fondos mixtos y sectoriales de Conacyt (fuente: [www.conacyt.mx/fondos/resultados.html](http://www.conacyt.mx/fondos/resultados.html)).

<sup>2</sup> <http://directorio.com.mx/robotica-automatizacion/>

grado por reconocidos expertos. Invitamos también a la comunidad a enviar presentaciones describiendo sus grupos y sus líneas de trabajo.

El resultado de esta convocatoria fue modesto, considerando el número de artículos recibidos. Sin embargo, los artículos técnicos aceptados para su presentación son de gran calidad y nos dan una idea del amplio panorama de la investigación que se hace en el área de robótica móvil en nuestro país. Las presentaciones de dos grupos trabajando en el área de robótica no hacen sino confirmar la gran variedad de temas abordados por los roboticistas mexicanos.

Estamos seguros que este taller es apenas el inicio de una fructífera relación entre los roboticistas mexicanos. Nos nos queda más que agradecer a los organizadores de los talleres del Encuentro Internacional en Ciencias de la Computación 2004, así como a los miembros de nuestro comité de programa por su valiosa ayuda. Todos ellos nos apoyaron de manera entusiasta en la organización de este evento.

## **Comité Científico**

Copresidentes:

Víctor Ayala, Universidad de Guanajuato

Angélica Muñoz, Instituto Nacional de Astrofísica, Óptica y Electrónica

Michel Devy, Laboratoire d'Analyse et d'Architecture des Systèmes

Olac Fuentes, Instituto Nacional de Astrofísica, Óptica y Electrónica

José Joel González, Universidad de Guanajuato

José Luis Gordillo, Instituto Tecnológico de Estudios Superiores de Monterrey

José Negrete, Universidad Nacional Autónoma de México - U. Veracruzana

Antonio Marín, Universidad Veracruzana





# Object Identification and Location by Using Simple Sonar Sensors

Luis Enrique Arámbula Miranda<sup>1</sup>

<sup>1</sup> Universidad Autónoma de Aguascalientes, Ave Universidad 940,  
Aguascalientes, Ags., México  
[arambula@correo.uaa.mx](mailto:arambula@correo.uaa.mx)

**Abstract.** Three simple objects: walls, corners and edges can be used to model indoor environments. These objects are typically detected either by using a complete echo signal analysis or by means of image processing. However, many mobile robots are equipped with simple sonar systems. This paper analyses some issues about the object classification by using only the time-of-flight (TOF) parameter. In order to collect enough information, the robot made turns in 1.5 degrees steps. The optimal number of positions for reading was 27. Since our robot has 8 sonars, the input vector size is 216. Three specialized Multi-Layer Perceptron (MLP) networks for identification, distance and angle estimation were developed. The accuracy achieved was 99.7, 93 and 99 % respectively. Thus, we argue that simple sonar-based robots can identify and locate walls, corners and edges.

## 1 Introduction

Nowadays, roboticists are looking for the best approach to recognize the environment that surrounds a robot. Many efforts have been made in both vision and sound. Although vision is a promissory field, we are strong believers that sound has some advantages over vision. First, sound transducers are simpler and cheaper than vision transducers. Second, the image processing demands a lot of computational effort. Third, sound can be more useful in certain light conditions. Finally, there is a possibility to recognize 3-D objects without complex stereo vision systems [1]. The use of sound for image recognition has achieved many interesting results in several areas including medical ultrasound, undersea sonar, seismic imaging and nondestructive evaluation, among others. So, we argue that sonars for object identification in mobile robots can be another successful area for sound applications.

Although several researchers are using sonar sensors to identify objects, most of them use a complete echo-signal analysis [2] [3]. Others use complex sonar arrays [4] [5] [6] [7]. However, several mobile robots are equipped with simple sonars. They only report the time-of-flight (henceforth TOF). Our work addresses the object identification using only the TOF, another echo features such as amplitude, phase shifting and second echo pulse are ignored.

## 2 Previous Related Work

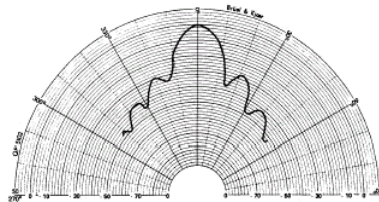
Barshan and Kuc exploited the differences in echo amplitude and ranges [8]. They used two transmitters and two receivers. Peremans *et al.* used one transmitter and three receivers to process several echoes [4]. They were able to distinguish and locate planes, corners and edges. Kleeman and Kuc discriminated planes, corners and edges with a two-transmitter two-receiver array by using templates [2]. They also identified objects in the background.

Gorman and Sejnowski reported the first use of sonars and neural networks for classification [9]. They classified undersea targets using a MLP. Watanabe and Yoneyama achieved an amazing 3-D object recognition system with an 8x8 array [1] . Dror *et al.* identified cubes and pyramids using one microphone and one speaker moving around an object [10]. The objects were illuminated with a multi-frequency bat-like sound. Barshan *et al.* classified and located planes, corners, edges and cylinders using a MLP with outputs digitally encoded [3].

## 3 Object Identification Using Sonars

### 3.1 Typical Problems with Sonars

When working with sonars, one must accept the characteristics involved on it. Firstly, the beamwidth is not sharp, the main lobe is about 30° wide, see Fig. 3.1. Secondly, the sound waves reflected on objects suffer attenuation, deviation and reflection.



**Fig. 3.1** Propagation pattern of the Polaroid 6500 series sonar. (Polaroid, 1990 [11]).

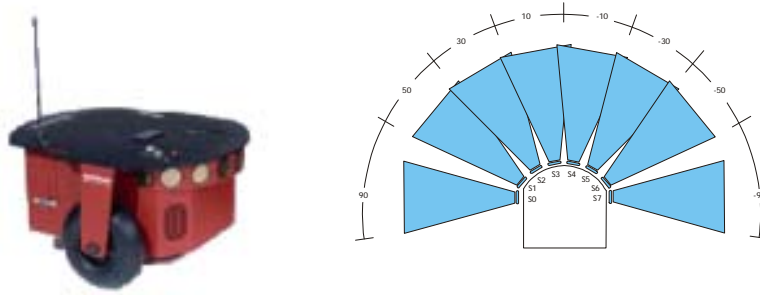
According to Nehmzow [12], the sonar problems can be summarized as follows:

- i) *Uncertainty of exact position.* Any object within the sensible region reflects an echo. The exact distance and angle are unknown.
- ii) *Specular reflections.* This error arises when a sonar ultrasound wave is reflected by several surfaces until it reaches the sensor.
- iii) *Crosstalking.* If several sonars are being used, the emissions can interfere among themselves. This problem also arises when several robots are working closely.

We can also say that successful object detection depends on the size, the shape and the texture of the object.

### 3.2 The Sonars in the Pioneer II Mobile Robot

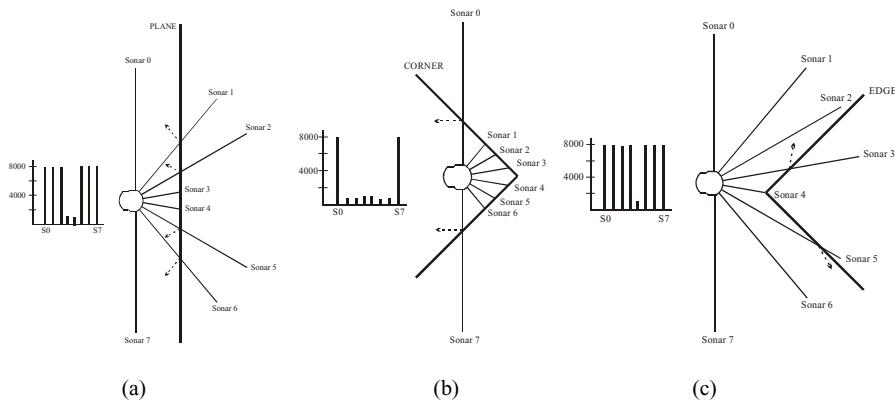
There are 8 sonars in the robot's front. They are distributed as depicted in Fig. 3.2. The only available data is the TOF from each sonar. It is not possible to fire one sonar and hear the echo in another, i.e. each sonar hears only its own pulse.



**Fig. 3.2** Pioneer II mobile robot and the sonar covered regions. (after Pioneer II Operations Manual, ActivMedia Robotics, 2000 [13])

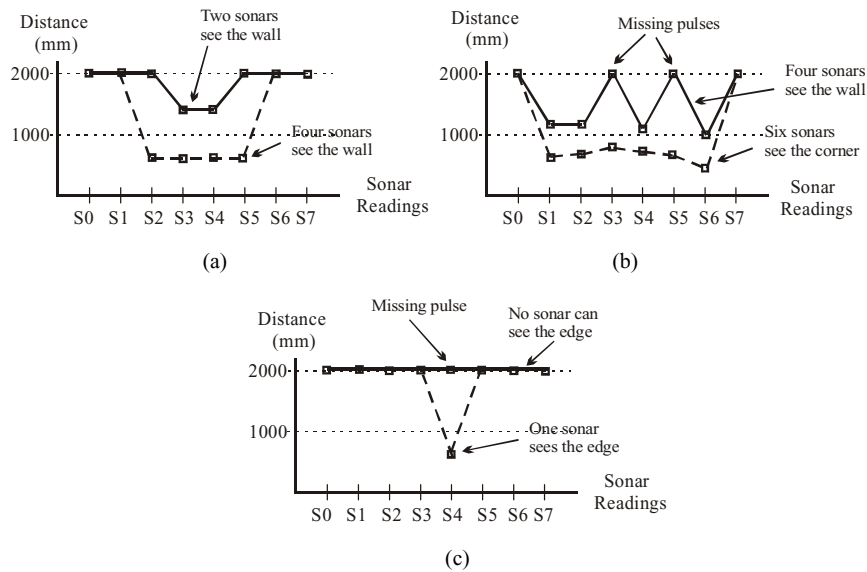
### 3.3 The Robot Facing The Three Objects

The model used is a straight line at the centre of the sonar axis. Only 2-dimensional data is available. The length of this line is the TOF reported. Therefore, we assumed that the object reflecting the echo is at the middle of the lobe. Fig. 3.3 illustrates our sonar model with the robot facing a wall, a corner and an edge. When facing a wall, two sonars at the centre read a small value (see Figure 3.3.a). When facing a corner, six sonars read small values (see Figure 3.3.b). Finally when facing an edge, only one sonar reads a small value (see Figure 3.3.c).



**Fig. 3.3** Sonar readings obtained by facing a wall (a), a corner (b) and an edge (c).

The actual readings are slightly different (see Fig. 3.4). Bear in mind that any reading bigger than 1200 mm is set to 2000 mm. It can be seen that the pattern depends on the distance between the object and the robot. In Fig. 3.4(a) the Sonar 2 (S2) and Sonar 5 (S5) see the wall only when it is close enough. If the wall is far, only Sonar 3 (S3) and Sonar 4 (S4) detect the wall. Fig. 3.4(b) shows that Sonar 3 and Sonar 5 cannot read their pulse. Hence, they are considered missing pulses. Fig. 3.4(c) shows that the edge is detected by the Sonar 4 only when it is close enough. When the edge is far away, no sonar can see the edge.



**Fig. 3.4** Actual sonar readings obtained when the robot is facing a wall (a), a corner (b) and an edge (c).

To solve the missing pulses problem, we proposed that several readings at different angles should be made. This problem was previously solved scanning each 0.522 degrees [14], or each 1.8 degrees [15] [3]. Another approach reads the echoes at different transducers [2]. Our proposal is to rotate the robot and acquire sonar readings from the 8 transducers each 1.5 degrees. In order to find the optimal vector size, we tested: 8, 80, 216 and 960 values.

### 3.4 Data Collection and Preparation

The data was collected at the Broker Lab in the University of Essex, UK. The walls corners and edges were assembled using cardboard of 30 cm of height. This cardboard has a very smooth texture, and for some angles it is invisible for the sonars. The robot was placed initially at 600mm from the object, then it made a turn while reading the sonars. After this, it moved away 10 cm from the object and made a new turn. This process continued until the robot was at a distance of 1200 mm. After data collection, the values were normalised to a maximum value of 1. Finally, the readings were inverted to enhance the importance of small values.

Example:	Acquired vector:	2968	8768	8768	623	584	8768	8768	8768
	Normalized vector:	1	1	1	0.28	0.25	1	1	1
	Inverted vector:	0	0	0	0.72	0.75	0	0	0

### 3.5 Network Architecture

In a similar work to ours, Barshan *et al.* [3] proposed a 232:100:21 network. The 21 output neurons had different purposes: the first seven are for type discrimination, the next seven encode the target range, and the last seven encode the angle. Our proposition has two main differences:

1. We used three networks, each one of them for a specific task. We argue that by separating the networks, the learning process is focused on mapping only a specific feature.
2. Instead of encoding digitally the angle and the distance, we used an analogue output neuron. Hence, we used only one neuron in the output layer.

For object type it was used a 216:20:3 network. The number of neurons in the input layer was determined by the input vector described before. The number of neurons in the hidden layer was determined using a process named as *enlarging*. It consists in starting with a relatively small number of units, then the size is incremented until the network learns. The number of output neurons is determined by the number objects that our system classifies. The valid outputs are: 001 for planes, 010 for corners, and 100 for edges. Any other values are invalid. Then, it means a ‘not recognized’ object.

The network for distance estimation is a 216:20:1 network. These values were found similarly to the previous network. The output neuron used the sigmoid function. Therefore, its output range swings from 0 to 1. This output indicates the normalized object distance, i.e. 0 for 0mm and 1 for 1200mm.

For angle estimation, we used a 216:12:1 network. It is similar to the distance estimation network except in the hidden layer. The output neuron used also the sigmoid function. It was also normalized to the 0 to 1 range. In this case, 0 for –40 degrees and 1 for 40 degrees.

We used *error back-propagation algorithm*, which is based on the *error-correction learning rule* (see Haykin [16]). A set of 3066 input vectors was generated and presented to the network. Each 10 epochs, a set of 1533 input vectors (half the training set size) not included in the training set was presented to the network to avoid over-training.

## 4 Experimental Results

The network were trained 30, 150 and 200 epochs for type, distance and angle respectively. The values found in the tables below reflect percentage of accuracy. Table I shows the accuracy for object classification. Notice that for 8 inputs the accuracy is low. Also notice that for 216 and 960 inputs the results are nearly the same. Table II shows a comparison among the four MLP distance estimators, the tolerance is 15 cm. Notice that the highest accuracy rate is for the 216-inputs MLP.

Table III shows the results for angle estimation. A 10° tolerance was used. Notice that for 8 and 80 inputs the average accuracy is nearly the same. For 216 input the accuracy is almost 100%. But for 960 inputs, the value is very low. A possible explanation for this low figure is that we used few training vectors.

**Table I** Accuracy rate for object classification.

Object	8 in	80 in	216 in	960 in
<b>Walls</b>	83.76	95.39	100	100
<b>Corners</b>	90.48	100	100	100
<b>Edges</b>	41.26	95.39	99.32	99.41
<b>Average</b>	71.83	96.93	99.77	99.8

**Table II** Accuracy for distance estimation with a tolerance of 15 cm.

Object	8 in	80 in	216 in	960 in
<b>Walls</b>	94	88	95	95
<b>Corners</b>	85	89	98	97
<b>Edges</b>	88	71	86	74
<b>Average</b>	89	82	93	89

**Table III** Accuracy rate for angle estimation using a tolerance of 10°

Object	8 in	80 in	216 in	960 in
<b>Walls</b>	76	93	100	51
<b>Corners</b>	76	88	100	37
<b>Edges</b>	98	70	99	35
<b>Average</b>	83	84	99	41

In Fig. 4.1 the first three plots illustrate the desired distance in red colour (stairs) and the MLP output in blue colour (peaks). Each step represents a distance, from 600 mm to 1200mm, in 100mm steps. The horizontal axis indicates the number of sample. The vertical axis represents the distance between the robot and the object. The fourth plot compares the accuracy for the 3 objects. In Figure 4, it can be seen that it was easier to estimate the distance to the corners than to the walls and corners. Also, the estimation of the edges distance has low accuracy. Further, notice that the bigger the tolerance, the greater the accuracy. Finally, observe that for tolerances beyond 20cm the accuracy remains nearly at 100%.

Fig. 4.2 depicts the plots produced by the 216-input angle estimation MLP. Seven positive peaks and seven negative peaks can be seen. Each one represents the distance at which the robot was trained. The rising slope indicates different angles for the same distance. Notice that the desired output and the actual output are almost the same, which reflects an excellent angle estimation accuracy. The fourth plot shows that the edges are the most difficult object to locate. Notice that after 10 degrees of tolerance the accuracy is 100%.

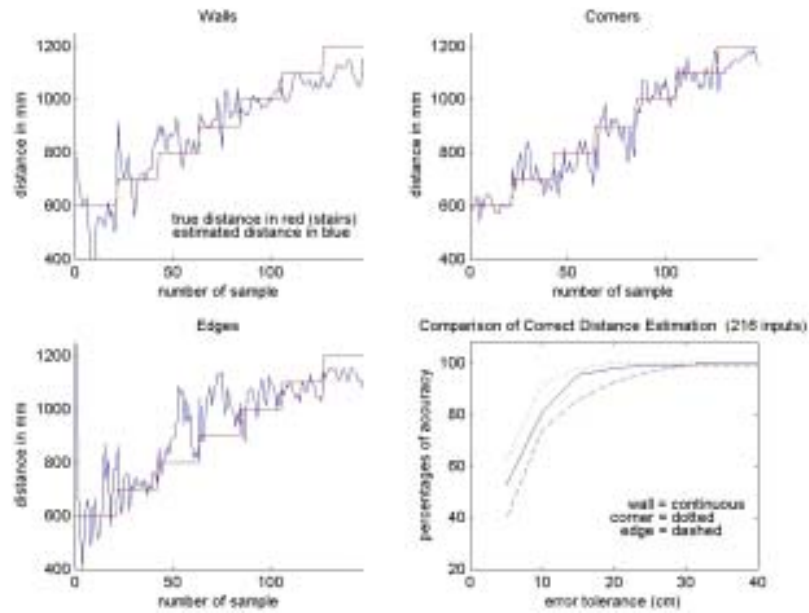


Fig. 4.2 Plots for distance estimation using a 216-input MLP

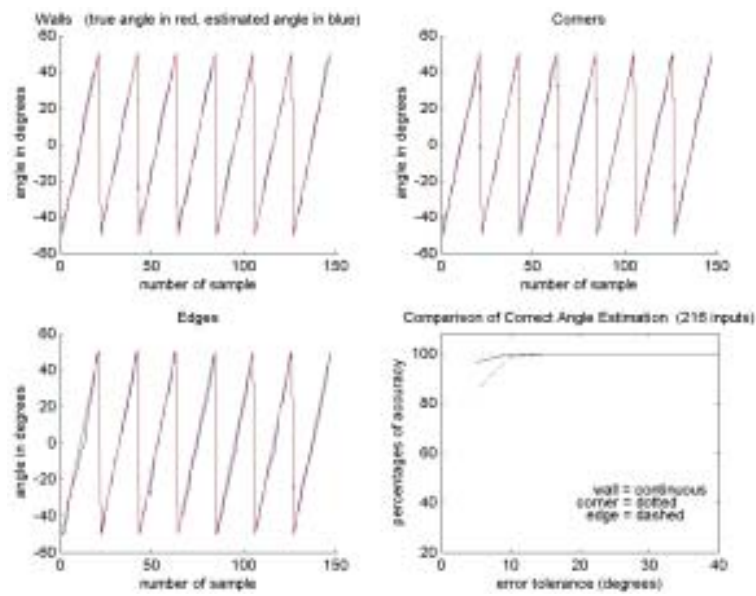


Fig. 4.2 Angle estimation plot for 216-input MLP



## 5 Conclusions

Traditional sonar-based classifiers rely on a complete echo signal analysis and on sonar arrays. By using a robot with simple sonar data we have found that it is possible to achieve good classification results. To compensate for the lack of available data, we proposed to turn the robot and acquire sonar data each 1.5 degrees. Different sizes of data vectors were compared to find the optimal solution. Specifically, we analysed 8, 80, 216 and 960 input vector size. The best results were obtained using 216 values; the scanning time was around 9 seconds. The MLP classifier could also estimate the distance and the angle. This network was divided into three sub-networks, one for each task. This method showed an average accuracy of 97 %.

The future work could be aimed to compare other approaches, for example supervised SOMs or Fuzzy C-means clustering. Additional future work is needed to compare those techniques under different circumstances, for example, employing a different robot. Finally, the inclusion of other objects (e.g. cylinders or acute corners) will be useful for the classifier.

Despite its bad reputation, there is no doubt that sonars will be used in today and tomorrow mobile robots, but not only for obstacle avoidance. It seems that they will be used as the main or one of the main elements for object recognition. Paraphrasing the John J. Leonard [17] challenge: *What is a robot with a sophisticated vision system in an environment full of glass walls, compared with a robot equipped with a sonar-based identification system?*

## References

- 1 Watanabe, S. and Yoneyama, M.: An ultrasonic visual sensor for three-dimensional object recognition using neural networks, IEEE Trans. on Robotics and Automation, Vol. 8, No. 2 (1992) 240-249.
- 2 Kleeman, L. and Kuc, R.: Mobile robot sonar for target localization and classification, Intl. Journal of Robotics Research, Vol.14 (1995) 295-318.
- 3 Barshan, B., Ayrulu B. and Utete S. W.: Neural network-based target differentiation using sonar for robotics applications, IEEE Trans. on Robotics and Automation, Vol. 16, No. 4 (2000) 435-442.
- 4 Peremans, H., Audenaert, K., and Van Campenhout, J.: A high-resolution sensor based on tri-aural perception, IEEE Trans. On Robotic and Automation, Vol. 9, No. 1 (1993) 36-48.
- 5 Peremans, H., Walker, A., and Hallam, J.: 3-D object localization with a binaural sonarhead, inspiration from biology, IEEE Int. Conf. On Robotics and Automation, (1998) 2795-2800.
- 6 Wilkes, D., Dudeck, G., Jenkin, M. and Milios, E.: Multi-transducer sonar interpretation, Proc. IEEE International Conference on Robotics and Automation (1993) 392-397.
- 7 Cao, A. and Borenstein J.: Experimental characterisation of Polaroid Ultrasonic sensors in single and phased array configuration, in The UGV Technology Conference at the 2002 AeroSense Symposium, Orlando Fl, April, 1-5 (2002).
- 8 Barshan, B. Kuc, R.: Differentiating sonar reflections from corners and planes by employing an intelligent sensor, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12, No 6, (1990) 560-569.
- 9 Gorman, R. P. and Sejnowski, T. J.: Learned classification of sonar targets using a massively parallel network, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 36, No. 7, (1988) 1135-1140.

- 10 Dror, I. E., Zagaeski, M. and Moss, C. F.: Three-dimensional target recognition via sonar: a neural network model, *Neural Networks*, Vol. 8, No. 1 (1995) 149-160.
- 11 Polaroid: Ultrasonic Ranging System – User Guide, Polaroid Corporation, Ultrasonic Components Group, 119 Winsor Street-2B, Cambridge, MA, 02139, USA (1990).
- 12 Nehmzow, U.: *Mobile Robotics: A Practical Introduction*, Springer-Verlag, London (2000)
- 13 ActivMedia Robotics: *Pioneer II Mobile Robots Operations Manual*, Peterborough, NH. (2000)
- 14 Leonard, J. L. and Durrant-Whyte, H. F.: Mobile robot localization by tracking geometric beacons, *IEEE, Trans. on Robotics and Automation*, Vol.7, No.3 (1991) 376-382.
- 15 Lee, D., C.: The map-building and exploration strategies of a simple, sonar-equipped, mobile robot; an experimental, quantitative evaluation, PhD Thesis, University College, London. (1995)
- 16 Haykin, S.: *Neural Networks, A comprehensive foundation*, Upper Sadle River, Prentice-Hall, NJ, (1999)
- 17 Leonard, J. L.: Directed sonar sensing for mobile robot navigation, PhD Thesis, University of Oxford. (1990)

## Una arquitectura de software para la integración de aplicaciones en robótica móvil

*José Antonio Pacheco-Sánchez<sup>1</sup>, Roberto Valdivia-Beutelspacher<sup>2</sup>, Oscar Mayora-Ibarra<sup>1</sup>*

*1 Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Cuernavaca  
Av. Paseo de la Reforma 182-A Lomas de Cuernavaca, C.P. 62589 Temixco Morelos  
México.*

*00379336@academ01.mor.itesm.mx*

*omayora@itesm.mx*

*2 Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Ciudad de  
México Calle del Puente 222 Ejidos de Huipulco, C.P. 14380 México D.F., México.  
valdivia@itesm.mx*

**Resumen.** Investigaciones recientes en el campo de la robótica, han tratado el problema de crear un modelo arquitectónico robusto para la integración de aplicaciones robóticas basadas en componentes. Para poder lograr lo anterior, se requiere identificar las necesidades que dicha arquitectura deberá satisfacer para implementar exitosamente la unificación de aplicaciones individuales dentro del contexto de la arquitectura; asimismo, es indispensable contar con una infraestructura y marco de trabajo que se tomará como base para los servicios o aplicaciones que se requerirán en futuros desarrollos.

**Abstract.** Recent research projects in the robotics field have treated the problem of proposing an architectural robust model for the integration of robotic applications based on components. To be able to achieve such integration, it is needed to identify the requirements that the above architecture will have to satisfy in order to implement successfully the unification of individual applications inside the context of the architecture, likewise, is indispensable to rely on an infrastructure and framework to take as basis for the services or applications that will be needed in future developments.

**Palabras clave:** Arquitectura de software, robótica móvil, sistemas distribuidos, modelo de capas.

## **1 Introducción**

En la actualidad se trabaja en desarrollar sistemas robóticos colaborativos cuya misión sea realizar tareas complejas con una gran seguridad y confiabilidad en medios ambientes reales, los cuales son por naturaleza muy dinámicos, desconocidos e incluso hostiles. Para lograr la seguridad y confiabilidad necesaria, se requieren incorporar formalmente diversos aspectos del campo de los sistemas computacionales tendientes a establecer un modelo arquitectónico más robusto del sistema multi-robótico. Entre las áreas consideradas están: los sistemas distribuidos colaborativos, las redes móviles, los sistemas inmersos (embedded) en tiempo real y los sistemas tolerantes a fallas.

Incorporar estos aspectos se debe hacer tanto internamente en los componentes de software de cada robot (intra-robot), como externamente en la interacción entre el grupo de robots entre sí (inter-robots) y la interacción con los seres humanos (humano-robot), así como con los otros elementos externos que influyen en el sistema robótico.

### **1.1 Cátedra/Proyecto de Robótica Móvil en el Tecnológico de Monterrey**

El uso de robots en la vida diaria está dejando de ser una ficción. Los desarrollos en robótica indican que en una o dos décadas existirá toda una nueva industria puesto que los robots serán artículos de uso común tal como ha sucedido con las computadoras.

El surgimiento de este tipo de robots implica que deben operar en ambientes reales, que normalmente son más complejos y dinámicos. Esto trae consigo importantes retos de razonamiento con incertidumbre, planeación, coordinación entre robots y comunicación humano-robot entre otras.

A partir de este entorno surge la cátedra/proyecto de robótica móvil, formada por un grupo interdisciplinario de investigadores en el Tecnológico de Monterrey en los Campus Cuernavaca y Ciudad de México, cuyo fin es lograr avances significativos en estas áreas y contribuir al desarrollo de los robots del futuro.

### **1.2 Desarrollos Actuales**

Dentro de la cátedra de robótica móvil se realizan proyectos de investigación en diversas áreas específicas, de acuerdo al primer taller organizado dentro de la cátedra [Sucar 2003] y que se describen a continuación:

- Aprendizaje por refuerzo.
- Construcción de mapas y localización de robots.
- Reconocimiento visual de ademanos.

- Sistema de navegación por medio de comandos vocales.
- Detección y evasión de obstáculos.
- Laboratorio virtual de robótica.
- Agente emotivo de interfaz para tutores.

La arquitectura que se pretende desarrollar permite incorporar formalmente -bajo una estructura modular y robusta- los diversos proyectos en las áreas listadas anteriormente.

## 2 Arquitecturas de software

Según la organización IEEE, dentro de su estándar 1471-2000, una arquitectura de software es considerada como la organización fundamental de un sistema, formada principalmente por componentes, las relaciones entre ellos y el ambiente, y los principios que gobiernan su diseño y evolución.

Se han propuesto diferentes paradigmas de arquitecturas para sistemas de razonamiento, mismas que son adaptables en entornos robóticos.

- **Arquitecturas basadas en reglas:** Este paradigma es una extensión de los sistemas expertos tradicionales que satisfacen los problemas para el control de aplicaciones en tiempo real. El conocimiento se encuentra en forma de reglas, mismas que pueden ser invocadas por eventos externos.
- **Arquitecturas de pizarrón:** El modelo de pizarrón es muy similar a la manipulación de datos en sistemas expertos, con la diferencia de que los procesos comparten una base de datos común llamada pizarrón.[Carver 1992]
- **Autómatas situados:** Este paradigma se asemeja al software de tiempo real convencional. Un autómata situado es una máquina de estados finitos cuyas entradas son obtenidas a través de sensores conectados con el ambiente y sus salidas están conectadas a actuadores. Los autómatas situados presentan dos tendencias:

*Automatón:* compilación fuera de línea, empezando por una representación explícita.

*Basado en la conducta:* procedimiento de diseño manual basado en una descomposición y acorde con las diversas conductas que el robot debe mostrar. [Rosenschein 1995]

- **Arquitecturas BDI (Belief-Desire-Intention):** Las arquitecturas BDI se inspiran en un modelo cognitivo del ser humano [Bratman 1987]. Los agentes (programas que se ejecutan sobre la arquitectura cuyo modelo toma una serie de precondiciones que alimentan una acción, misma que genera un resultado y una serie de postcondiciones) utilizan un modelo del mundo, una representación de

cómo se les muestra el entorno. El agente recibe estímulos a través de sensores ubicados en el mundo. Estos estímulos modifican el modelo del mundo que tiene el agente (representado por un conjunto de creencias). Para guiar sus acciones, el agente tiene *Deseos*. Un *deseo* es un estado que el agente quiere alcanzar a través de *intenciones*. Aunque la formulación inicial es de Bratman, fueron Georgeff, Rao y Kinny [Kinny 1997] quienes formalizaron este modelo y le dieron visos de metodología.

- **Arquitectura por Capas:** Rodney Brooks [Brooks 1986] propone una solución al problema de control en robots descomponiendo el mismo en capas de acuerdo a un cierto nivel de comportamiento. Brooks demostró las ventajas de descomponer el sistema de control en tareas de comportamiento ligadas a su respectiva capa. El resultado fue una arquitectura que puede soportar múltiples servicios de manera simultánea, garantizando con ello, una respuesta potencialmente satisfactoria las tareas o los servicios robóticos.

Cada uno de los paradigmas mencionados anteriormente posee ventajas y desventajas. La elección adecuada entre cada uno de ellos depende de ciertos factores [Ingrand 2003] entre los que podemos mencionar:

**Programabilidad:** Que la arquitectura sea fácilmente programable (tanto del punto de vista del programador como del usuario final)

**Integración:** Que se puedan establecer mecanismos transparentes de comunicación entre los componentes.

**Autonomía, adaptabilidad y coherencia:** Se refiere a que el robot ejecute las acciones, afine y adapte sus planes y sus comportamientos en función de los objetivos y del ambiente que lo rodea.

**Reactividad:** Que los componentes que integran la arquitectura sean capaces de reaccionar apropiadamente ante estímulos específicos y situaciones particulares.

**Robustez y tolerancia a fallas:** La capacidad de explotar la redundancia de recursos, fuentes de información y procesos, son tareas que la arquitectura deberá permitir, además de contar con un sistema de tolerancia a fallas, que nos permita recuperar la funcionalidad del robot en entornos semi controlados.

**Seguridad:** Se requieren mecanismos que garanticen la seguridad de las aplicaciones robóticas, puesto que algunas veces, el robot puede verse implicado en situaciones donde la seguridad es un factor muy importante

**Extensibilidad:** La arquitectura debe permitir agregar nuevas funcionalidades que puedan ser fácilmente adaptables a las existentes.

El modelo de arquitectura que más se aproxima a estas características, es la *arquitectura por capas* [Brooks 1986], que ha sido probado y es actualmente el modelo más utilizado para entornos robóticos.

## 2.1 Modelo de capas

La arquitectura de 3 capas, ha sido estudiada desde los años 80s, cuando Nilsson propuso descomponer en elementos funcionales el sistema de control de robots autónomos móviles: un sistema sensorial, un sistema de planeación y un sistema de ejecución; donde el trabajo del sistema sensorial era interactuar directamente con los sensores, recibiendo las entradas de éstos y generar un modelo de estado. El sistema de planeación, toma ese modelo de estado y la meta a cumplir para establecer una misión o plan; por último, el sistema de ejecución, toma el plan y realiza las acciones necesarias para llevar a cabo la misión [Nilsson 1980].

## 2.2 Arquitectura LAAS

La arquitectura propuesta por el Laboratory for Analysis and Architecture of Systems [Alami 1998] se encamina a producir un marco de trabajo abierto (framework open source), proveyendo una base funcional para el control general en robots. El software está diseñado para ser *independiente de la plataforma*, pero además es planteado como independiente de la aplicación. Uno de los puntos interesantes es que la arquitectura propuesta tiende hacia la integración de componentes de software.

### Estrategias utilizadas

- *Orientación a objetos*, como la base para formar librerías de clases ("módulos").
- *Patrones de software*, dado que esta propuesta hace uso y rediseña patrones existentes, se evita perder tiempo en la generación de nuevos diseños o reingeniería de malos diseños.
- Enfoque hacia una arquitectura de componentes, y no hacia una *arquitectura de sistema*.

### 3 Arquitectura Propuesta

Proponemos seguir el modelo basado en capas, con una capa funcional, o de más bajo nivel, una capa de ejecución y una capa de decisión. Aunado a esta configuración, cada capa se divide en subcapas, principalmente la capa de ejecución, las cuales se encuentran en la fase de definición de servicios. El manejo de errores y la tolerancia a fallas, es manejada por una capa adicional que se proyecta a las tres capas principales como una capa común, o una capa sobredimensionada sobre éstas.

Lo que se pretende es llegar a un modelo conceptual de control que cubra no solo el control de movimientos del robot, sino todos los tipos de control. En este *patrón de control*, la mayor parte se trabaja en *tiempo real*; el nivel más bajo controla cuidadosamente el control de movimientos instantáneos (“llamado control reactivo”); la capa media realiza la *planeación local* y la ejecución de tareas; y la capa mas alta realiza la *planeación global, recuperación de errores, y “razonamiento”*.

#### 3.1 Modelo basado en servicios (modelo extendido)

El modelo basado en servicios, nace principalmente por la necesidad de cómputo demandante que requieren ciertas aplicaciones (como el reconocimiento de rostros, reconocimiento de voz, etc.). Lo que se pretende lograr es mantener un robot “ligero”, donde las aplicaciones arriba mencionadas, puedan ser puestas a disposición en servidores remotos con gran capacidad de carga, tanto de información, como de comunicaciones.

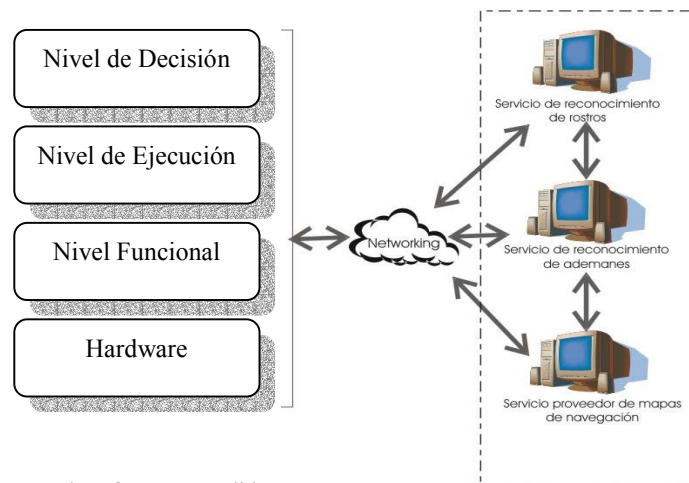
Los proyectos que se especifican en la sección 1.2, han sido analizados para su posible implementación bajo un esquema de servicios robóticos, esto quiere decir, que algunos proyectos serán implementados a través de servidores en una red, mismos que recibirán peticiones desde las entidades robóticas, las analizarán y ejecutarán de acuerdo a las restricciones del servicio y posibilidades en cuanto a recursos se refiere. Una vez ejecutada la petición, se generará una respuesta que es enviada al origen de la petición.

El modelo de arquitectura extendida, definirá:

- El tipo y formato de intercambio de datos entre los servicios y las entidades robóticas.
- Los modelos de enlace de comunicaciones alámbricas e inalámbricas que ofrezcan la robustez necesaria para soportar todo el sistema.
- Especificaciones de requerimientos en tiempo real que deberán cubrir las aplicaciones que se convertirán en servicios.



- Un modelo de arquitectura distribuida que soporte la interacción entre los servicios y las entidades robóticas, así como también la interacción entre los servicios entre sí.



**Figura 1.** Arquitectura de software extendida

Algunas de las estrategias a seguir para lograr una adecuada y correcta integración de componentes, son principalmente el detectar patrones, componentes, rutinas y métodos comunes de software para crear bibliotecas precompiladas de uso general.

#### 4 Capa de decisión

La capa de decisión esta ubicada en la parte superior del modelo de 3 capas, ésta incluye capacidades deliberativas tales como la planeación de tareas, reconocimiento de situaciones, detección de fallas de alto nivel, etc., contiene cuando menos un supervisor de tareas, el cual está comunicado hacia el nivel más bajo, a través de la capa de ejecución para poder enviar las peticiones. Es el responsable de supervisar los planes o la ejecución de procedimientos mientras al mismo tiempo inicia las reacciones a eventos del nivel más bajo.

- El nivel de la decisión está constituido por todos los procesos que requieren un cierto conocimiento global de la(s) tarea(s) y el contexto de su ejecución. Involucra las capacidades de planeación y toma de decisiones.
- Este nivel abarca dos entidades: *un planificador y un supervisor*. El *planificador* produce la secuencia de las acciones necesarias para alcanzar una

tarea o una meta dada. El *supervisor* interactúa con el nivel siguiente, controlando la ejecución del plan y reaccionando a los eventos entrantes. [Alami 1998]

## 5 Capa de ejecución

Localizado en medio de las capas del modelo. El nivel de ejecución, revisa las peticiones enviadas al nivel funcional, así como también los recursos utilizados. Este nivel está sincronizado con los módulos del nivel funcional, en el sentido de que puede ver todas las peticiones que llegan a ellos y los reportes que se generan y envían desde los módulos hacia el origen de las peticiones. Este nivel funciona como un filtro que establece qué peticiones serán aceptadas o rechazadas para pasar al nivel más bajo de acuerdo al estado actual del sistema y al modelo propuesto por el usuario. Los requerimientos temporales de este nivel están estrechamente relacionados con sistemas en tiempo real.

Este nivel es una interfaz entre el nivel de decisión y el nivel funcional. La capa de ejecución es puramente reactiva sin ninguna capacidad de planeación. Recibe desde la capa de decisión las secuencias de las acciones a ejecutar, posteriormente, selecciona, parametriza, y sincroniza dinámicamente la función o funciones adecuadas en la capa funcional. [Ingrand 2002]

## 6 Capa funcional

Se localiza en el nivel más bajo e incluye todas las capacidades integradas a las acciones y percepción del robot. Estas funciones de procesamiento y ciclos de control (procesamiento de imágenes de bajo nivel, control de movimiento, etc.) serán encapsuladas en componentes comunicados entre sí. Cada componente aporta un número de servicios y aplicaciones disponibles a través de peticiones recibidas por los sensores de acuerdo a su naturaleza. Una vez completada o abortada la tarea, se genera un reporte con el estatus, mismo que es enviado al origen de la petición. Los datos serán producidos por los módulos, como la posición actual y velocidad (provenientes del componente de navegación), o la trayectoria actual (proveniente del componente de planeación de trayectorias). Tales datos pueden ser accedidos desde otros componentes y niveles superiores. Los requerimientos temporales de los componentes dependen del tipo de tratamiento que realicen; esto quiere decir que algunos componentes deberán cumplir requerimientos temporales más estrictos que otros.

La capa de ejecución contiene un grupo de acciones básicas del robot, implementando funciones de procesamiento y tareas orientadas hacia planeación, visión, localización,

control de movimiento, etc.; así como también funciones de supervisión programables (condición-reacción) usadas para desencadenar acciones reactivas. [Alami 1998]

## **7 Tolerancia a fallas**

Cabe destacar, que un punto muy importante en el desarrollo de nuestra arquitectura, consiste precisamente en generar un modelo robusto de tolerancia a fallas; éste mismo es concebido cómo una capa que se sitúa en un contexto global dentro del sistema de capas de servicios, o bien, como una capa “envolvente” que interactúa directamente con el sistema, permitiendo una supervisión y control reactivo inmediato dentro de cada capa. Algunas características del modelo son:

- El modelo es concebido como una interfase con todos los componentes y los diferentes niveles, asegurando que todas las peticiones que pasan a través de la capa sean consistentes respecto a las necesidades del sistema. Por ejemplo, se debe asegurar que una petición remota de que el robot se empiece a mover, no se ejecute si el estado actual del robot está en modo de control directo.
- Manejar eficientemente los recursos del sistema y garantizar que cualquier petición que conduzca a un estado de sobrecarga o inconsistencia, sea controlada adecuadamente.
- Actuar en forma síncrona con la capa funcional, dado que la comunicación en este nivel es llevada a cabo en su totalidad en tiempo real, asegurando así una vista consistente del estado de los componentes en el nivel funcional.
- Analizar los requerimientos específicos de tiempo real y tolerancia a fallas para cada uno de las aplicaciones o servicios.

## **8 Conclusiones**

Este artículo presentó las características esenciales que debe considerar una arquitectura de software que soporte diversas aplicaciones de sistemas multirobóticos. Se analizaron diversas clases y se concluye que la mejor opción es una arquitectura de capas extendida, que considere tres niveles: funcional, de decisión y de ejecución, con una capa subyacente que realice las funciones de tolerancia a fallas. La incorporación de un modelo basado en servicios sobre una red de área local, se plantea como una respuesta hacia los altos costos computacionales que demandan algunos servicios, mismos que al ser distribuidos realizan una tarea colaborativa dentro de la arquitectura de software.

## 9 Referencias

- [Alami 1998] R. Alami, R. Chatila, S. Fleury, M. Ghallab, F. Ingrand, *An Architecture for Autonomy*, International Journal of Robotics Research, Special Issue on Integrated Architectures for Robot Control and Programming, Vol. 17, No. 4, pp. 315-337, April 1998.
- [Bratman 1987] Bratman, M. E.: *Intentions, Plans, and Practical Reason*. Harvard University Press. 1987.
- [Brooks 1986] Rodney A. Brooks. *A robust layered control system for a mobile robot*. IEEE J. Robotics and automation, RA-2(1):14-23, March 1986
- [Carver 1992] Carver, N. and Lesser, V. R. (1992) *The Evolution of Blackboard Control Architectures- Technical Report 92-7*. Department of Computer Science, University of Massachusetts. 1992
- [Ingrand 2002] Félix Ingrand, Frédéric Py, *An Execution Control System for Autonomous Robots*, IEEE International Conference on Robotics and Automation, May 11-15, 2002, Washington D.C., USA
- [Ingrand 2003] Félix Ingrand: *Architectures Logicielles pour la Robotique Autonome*, Journées Nationales de Recherche en Robotique, October 8-10, 2003
- [Kinny 1997] Kinny, D., Georgeff, M., and Rao, A, *A Methodology and Modelling Technique for Systems of BDI Agents*, In *Agents Breaking Away: Proceedings of the Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World, (LNAI Vol 1038): 56-71*. Springer-Verlag, Berlin 1997 .
- [Nilsson 1980] Nils J. Nilsson, *Principals of Artificial Intelligence*. Ed. Tioga Press, Palo Alto, California 1980.
- [Rosenschein 1995] Stanley J. Rosenschein, Leslie Pack Kaelbling: *A Situated View of Representation and Control. Computational Theories of Interaction and Agency, 1995*.
- [Sucar 2003] Enrique Sucar Succar., *Primer taller de Robótica móvil, Cátedra de robótica móvil ITESM Cuernavaca Morelos México 2003*.

# Deterministic sampling: A framework for Probabilistic Roadmap Planners

Abraham Sánchez<sup>1</sup>, René Zapata<sup>2</sup>, and David Pinto<sup>1</sup>

<sup>1</sup> Facultad de Ciencias de la Computación, BUAP  
14 Sur esq. San Claudio, CP 72550  
Puebla, Pue., México  
`asanchez, dpinto@cs.buap.mx`

<sup>2</sup> LIRMM, UMR5506 CNRS, 161 rue Ada 34392,  
Montpellier Cedex 5, France  
`zapata@lirmm.fr`

**Abstract.** We present our current progress on designing sampling strategies in sampling-based motion planning field. The basis for our methods is the replacement of random sampling by a deterministic one [12]. Sampling-based methods use only information from a collision detector as they search the configuration space. The development of efficient collision detection algorithms, has resulted in the introduction of a number of powerful motion planning algorithms, capable of solving challenging problems. Key open issues and areas of future research are also discussed.

## 1 Introduction

Motion planning is concerned with computing collision-free motion for objects in an environment populated with obstacles. It has its origin in robotics, where planning collision free motion to achieve a specified goal is a fundamental characteristic of autonomous robots. Motion planning has application in many other areas, such as design for manufacturing, computer animation, medical surgery simulation, and computational biology.

Randomization is the most effective technique for reducing the high cost associated with motion planning of moving objects with many degrees of freedom. Recently, a class of randomized algorithms called Probabilistic Roadmap Methods (PRMs) has successfully solved difficult motion planning problems. PRM approach is composed by a learning phase and a query phase. In learning phase, a graph (*the roadmap*) is generated in the free space. In query phase, a simple planning algorithm is used to link both, initial and goal configurations to the graph, and then a standard graph search algorithm is performed to find the shortest path. The goal of the roadmap is to represent as faithfully as possible the connectivity of the free space.

The most difficult problems for PRM planners are those for which the robot must pass through a narrow passage to reach the goal. In such cases a big quantity of samples must be generated to get a high enough density of samples in the narrow passages. There are several sophisticated sampling strategies that

can solve this difficulty [5], [1], [4], [3], [10], [15]. Many of these strategies require complex geometric operations that are difficult to implement in high-dimensional configuration spaces.

In this work, we propose deterministic sampling strategies in the PRM framework. This approach remarkably improves the PRM-based planners, by replacing random sampling. Indeed, we can obtain the best possible asymptotic convergence rate, which is shown to be superior to that obtained by randomization. Sampling can be seen as an optimization problem in which a set of points is chosen to optimize some criterion of uniformity. Also, deterministic sampling can be thought of as a sophisticated form of stratified sampling.

## 2 Deterministic sampling

The historical origin of discrepancy theory is the theory of uniform distribution developed by H. Weyl and other mathematicians in the early days of the 20th century. While the latter deals with the uniformity of infinite sequences of points, the former deals with the uniformity of finite sequences. Finite sequences always have some irregularity from the ideal uniformity due to their finiteness. *Discrepancy* is a mathematical notion for measuring such irregularity. Let  $X = [0, 1]^d \subset \mathbb{R}^d$  define a space over which generate samples. Consider designing a set,  $P$ , of  $n$   $d$ -dimensional sample points  $\{x_0, x_1, \dots, x_n\}$  in way that covers  $X$ . Let  $\mathcal{R}$  be a collection of subsets of  $X$ , called a *range space*. Let  $R \in \mathcal{R}$  denote one such subset. The formal definition of discrepancy is as follows:

$$D_n(P, \mathcal{R}) = \sup_{R \in \mathcal{R}} \left| \frac{P \cap R}{n} - \lambda(R) \right|, \quad (1)$$

where  $\lambda$  denotes the Lebesgue measure on  $X$  and the supremum is taken over all axis-parallel boxes  $R$ . A detailed analysis of the discrepancy can be found in [9] and in the references therein.

Similarly to the notion of discrepancy, it is possible to quantify the denseness of  $n$  points, the *dispersion*, which is defined by

$$d_n(P, \delta) = \max_{x \in X} \min_{1 \leq i \leq n} \delta(x, x_i), \quad (2)$$

It was introduced by Hlawka (1976) and later investigated in more general form in [9]. Above  $\delta$  denotes any metric.

Although dispersion has been given less consideration in the literature than discrepancy, it is more suitable for motion planning. Dispersion has been developed to bound optimization error; however, in PRM-based planners, it can be used to ensure that any corridor of a certain width will contain sufficient samples [8].

## 3 Probabilistic roadmap planner

The algorithm for constructing the graph is shown in pseudo code.

```

BUILD_PRM()
1  $n \leftarrow 0$ 
2 while  $n < N$ 
3    $q \leftarrow \text{SamplingStrategy.generateConfiguration}()$ 
4   if  $q \in \mathcal{CS}_{free}$ 
5      $U \leftarrow \text{NeighborStrategy.findNeighbors}(q)$ 
6      $V \leftarrow V \cup q$ 
7     foreach  $u \in U$ 
8       if  $\text{LocalPlanner.verifyPath}(q, u) = \text{true}$ 
9          $E \leftarrow E \cup (q, u)$ 
10     $n \leftarrow n + 1$ 

```

By applying deterministic sampling techniques, our approach replaces the random sample generator that appears in Line 3 in the roadmap construction algorithm. We can consider two variations: 1) a version for multiple-queries, where the roadmap is precomputed, and 2) a Lazy version that performs collision checking during the search, which is suitable for single-query [12].

## 4 Experimental comparison between random and deterministic sampling

Much research has been done on motion planning for non-holonomic car-like robots (see [6] for a review). Within PRM framework, Švetska and Overmars used RTR paths as local method [14]. An alternative to RTR local method is to use a local method that constructs the shortest car-like paths that connects its argument configurations [11]. Another randomized strategy that has been used for non-holonomic planning is the RRT approach [7].

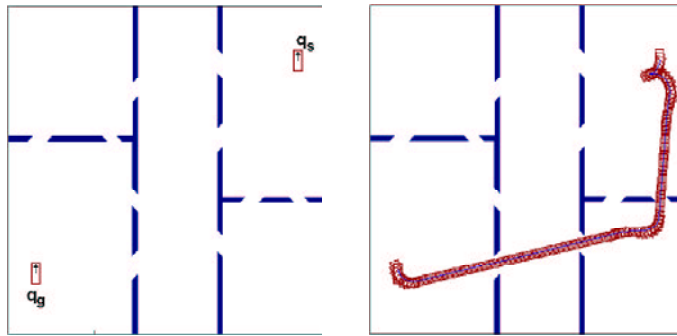
Since a randomized algorithm rarely behaves identically after repeated trials, it is hard to compare a randomized algorithm with a deterministic one. This difficulty is met by comparing the performance of the deterministic PRM version (DRM) with the “average” performance of  $k$  iterations of the classic PRM.

Table 1 shows the results of experiments performed on a challenging scene (Figure 1). The connection radius and the number of neighbors are given respectively in first and second columns. The number of nodes required to find a path is shown for the different versions of DRM and 100 averaged trials of the PRM. In all experiments, we have used Halton/Hammersley(H/H), Sobol (S) and Faure (F) points for DRM; pseudo-random numbers were generated using the linear congruential generator. The last column indicates the improvement factor of deterministic sampling over random sampling, in terms of the number of nodes. In Figure 1, we show a path computed in this scene. We observed during the experimental tests that the election of the connection radius and the number of neighbors have an important influence in the roadmap construction step.

Computing paths using the local method is the most time-consuming step in the roadmap construction algorithm. We would like to avoid such computations

**Table 1.** A comparison between PRM and DRM for a narrow corridor problem.

Rad	K	PRM	DRM-H/H	DRM-F	DRM-S	Factor
0.10	7	417	350	350	350	1.19
0.15	10	283	350	200	350	1.42
0.20	15	243	250	200	250	1.22
0.30	15	214	225	200	235	1.07
0.40	10	121	125	100	100	1.21



**Fig. 1.** This scene is difficult, the only way to get from the start configuration to the goal configuration, is by passing through some of the narrow passages. A path computed is shown.

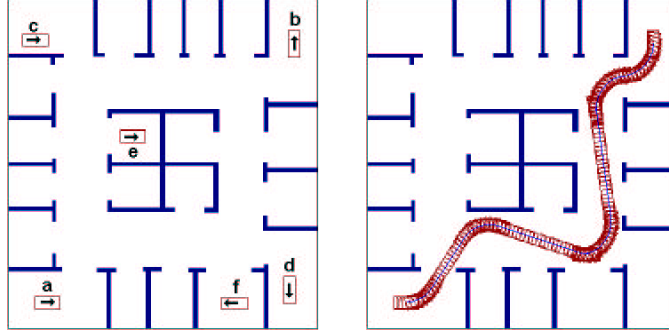
as much as possible. A recent variant called lazy-PRM has been proposed [2], the idea is not to test whether the paths are collision-free unless it is really needed. The goal of this variant is to minimize the number of collision checks.

The tests are performed in a “corridor-like” scene, consisting of 21 “rooms” and a “hallway” (see Figure 2). The main difficulty is the large number of narrow passages that connect the rooms to the hallway. The rectangular robot can only just pass through. Also, it is difficult for the robot to reverse its orientation when it is in the hallway. As a query test set, we take  $\{(a, b), (c, d), (e, f)\}$ . We have used Halton/Hammersley, Sobol and Faure points as inputs to both algorithms to solve this test set. Lazy-PRM uses the linear congruential method for generating pseudo-random points. For lazy-PRM, we performed 150 trials on each test set.

Table 2 shows the minimum, maximum and average number of nodes for lazy PRM in the first three columns. The last column shows the number of nodes used by lazy DRM. In this case, the three low-discrepancy sequences used the same number of nodes. Table 3 compares computation times.

A difficulty becomes when use lazy PRM approach, without prior knowledge about the particular problem, it is difficult to determine how many vertices should appear in the roadmap. Another difficulty is how many nodes should be added if query fails.





**Fig. 2.** A corridor-like scene, and its configuration test set. A path computed by the planner is shown.

**Table 2.** Comparisons of the number of nodes for lazy PRM vs. lazy DRM.

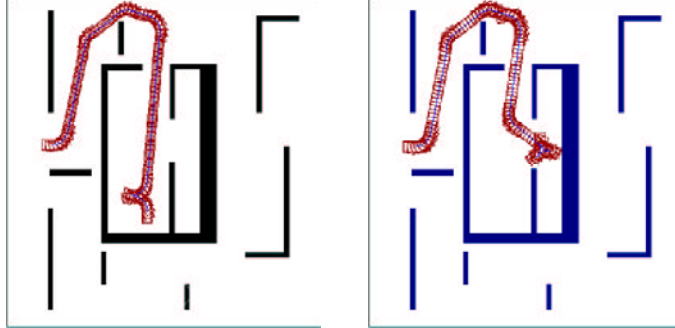
Test	Min	Max	Avg	Lazy DRM
(a,b)	80	170	128	100
(c,d)	100	150	125	100
(e,f)	100	200	147	100

We know that lattices have a regular, well-defined neighborhood structure. This allows the initial roadmap to be implicitly defined with little or no pre-computation because all vertices, neighboring vertices and edges are defined implicitly by lattice rules [8], [12]. Due to that lattices have discrepancy bounds that are as good as the best bounds for non-lattice sample sets, we can make a lazy LRM (lattice-based roadmap) that simultaneously obtains low-discrepancy benefits observed in the lazy DRM and the dramatic reduction in pre-computation time.

Figure 3 shows a very difficult scene used to compare lazy LRM with other planners (Lazy PRM and lazy DRM). It contains many narrow passages, it provides little freedom of movement for the robot, and any path solving the problem will be very long. Table 4 gives a comparison of running times for different versions.

**Table 3.** Comparison of the running times for lazy PRM vs. lazy DRM.

Test	Min	Max	Avg	PreCmp	PreCmp	Query	PreCmp	Query	PreCmp	Query
				H/H	H/H	S	S	F	F	
(a,b)	0.46	1.58	1.62	3.15	2.40	0.89	2.85	0.82	1.97	0.62
(c,d)	0.65	2.56	1.46	2.95	2.44	0.88	2.07	0.68	2.37	0.52
(e,f)	0.87	3.40	1.58	6.09	2.26	0.76	2.45	0.57	2.75	0.64



**Fig. 3.** Two feasible computed paths.

**Table 4.** Comparisons between lazy LRM, lazy DRM and 50 trials of lazy PRM.

	lazy PRM	H/H	lazy DRM F	S	lazy LRM
PreCmp	10.14	5.18	4.59	11.63	0.65
Query	0.96	0.66	0.68	0.57	0.57
Nodes	210	140	160	250	100

## 5 Discussion

After discussing some difficulties of uniform sampling, we have verified that deterministic techniques can be beneficial if we compare them to the random methods that have dominated sampling-based motion planning. However, it is more challenging to face these difficulties in a non-uniform sampling context. The main motivation for non-uniform sampling is based on the possibility of determining whether some areas of the  $\mathcal{C}$  are more important than others. If this is the case, then we should sample them with a higher density, as efficiently as possible. Several authors have discussed the importance of generating samples around narrow passages [5], [1], [3], [4], [10], [15].

An important difference between simple-query and multiple-query planners was considered in [5]. The first sampling-based methods used the main idea of searching solutions of single query. PRM was introduced as a pre-calculated data structure that could be used to quickly answer many queries in the same environment. If we analyze the time needed to solve a query after constructing a roadmap, we can then say that an advantage exists regarding dedicating time toward pre-calculation procedures. However, the time invested is important only if a particular application uses many planning queries for the same environment. An important issue that deserves more research is the idea of how to incrementally construct a data structure whose transformation velocity increases with the number of queries made.

It has been seen that there are benefits for examining deterministic alternatives of random approaches frequently used in sampling-based motion planning. We believe that sampling-based motion planning is better than combinatorial algorithms primarily because they are based more on sampling than on randomness (which we believe to be incidental). In fact, using randomness could lead to a lower understanding of key issues in searching and motion planning. Therefore, we believe that new advances in motion planning will occur as a result of detailed studies of these key issues, more than with creative applications of randomness.

There are, however, original advantages for some kinds of randomness which we should take advantage of. It is possible to design new methodologies for searching with a limited or appropriate randomness.

## 6 Conclusion

The key for the success of contemporary algorithms of motion planning is not the randomness or heuristics, but the fact that these algorithms are based on sampling and consequently avoid the complexity of building  $\mathcal{C}_{obs}$  representations, which several planners have had to implement.

We believe that randomization is useful in many contexts. Its value, nevertheless, depends greatly on the paradigm within it is used. Randomization does not appear to be advantageous in the PRM context, according to our experiments in non-holonomic motion planning and theoretical analysis presented by the authors in [8], [12]. Deterministic sampling enables the DRM, lazy DRM and lazy LRM to be *resolution complete* [8]: if it possible to solve the query, they eventually solve it.

The lattice-based PRM shows dramatic performance improvements, primarily because it exploits the neighborhood structure of the lattice to avoid the pre-computation required by the randomized lazy PRM.

PRM is a highly-flexible method. It is possible to apply it to some particular robot type, all that is needed is a local method that computes feasible paths for this robot type and some induced metric. Much research in PRM field has been done, not only in robotics, its application has extended to other fields as animation, computer games, virtual environments, and maintenance planning and training in industrial CAD systems.

In this work we have not tried to dismiss to all developed work in the PRM framework but rather we presented *some simple ideas* to obtain better results. The different improvements suggested are difficult to compare, since each author used his or her own implementation of PRM and used different test scenes, both in terms of environment and the robot type used.

## References

1. N. A. Amato, B. O. Burchan, L. K. Dale, C. Jones, and D. Vallejo. OBPRM: An obstacle-based PRM for 3D workspaces, *Proc. of the Workshop on Algorithmic Foundations of Robotics*, pp. 155-168, 1998.

2. R. Bohlin and L. Kavraki. Path planning using lazy PRM, *Proc. of the IEEE Robotics and Automation Conference*, pp. 521-528, 2000.
3. V. Boor, M. Overmars, and F. Van der Steppen. The gaussian sampling strategy for probabilistic roadmap planners, *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 1018-1023, 1999.
4. D. Hsu D, L. Kavraki, J. C. Latombe, R. Motwani, and S. Sorkin. On finding narrow passages with probabilistic roadmap planner, *Proc. of the Workshop on the Algorithmic Foundations on Robotics*, pp. 141-154, 1998.
5. L. Kavraki, P. Švestka, J. C. Latombe, and M. Overmars. Probabilistic roadmaps for fast path planning in high dimensional configuration spaces, *IEEE Transactions on Robotics and Automation*, 12(4), pp. 566-580, 1996.
6. J. P. Laumond (Editor). Robot motion planning and control, *Lecture Notes in Control and Information Sciences*, Springer Verlag, 1998.
7. S. M. LaValle and J. J. Kuffner. Randomized kinodynamic planning, *Proc. of the IEEE Robotics and Automation Conference*, pp. 473-479, 1999.
8. S. LaValle and M. Branicky. On the relationship between classical grid search and probabilistic roadmaps. *Proc. of the Workshop on the Algorithmic Foundations of Robotics*, 2002.
9. H. Niederreiter. *Random number generation and Quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, 1992.
10. C. Nissoux, T. Siméon, and J. P. Laumond. Visibility based probabilistic roadmaps, *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, pp. 1316-1321, 1999.
11. J. A. Reeds and R. A. Shepp. Optimal paths for a car that goes both forward and backwards, *Pacific Journal of Mathematics*, Vol. 145, No. 2, pp. 367-393, 1990.
12. A. Sánchez L. Contribution à la planification de mouvement en robotique: Approches probabilistes et approches déterministes (In French), *PhD thesis*, Université Montpellier II, 2003.
13. A. Sánchez L. A deterministic sampling approach to robot motion planning, *Fourth Mexican International Conference on Computer Science*, IEEE Computer Society, pp. 300-307, 2003.
14. P. Švestka, and M. H. Overmars. Motion planning for car-like robots using a probabilistic learning approach, *The International Journal of Robotics Research*, Vol. 16, No. 2, pp. 119-143, April 1997.
15. S. Wilmarth S, N. M. Amato and P. R. Stiller. MAPRM: A probabilistic roadmap planner with sampling on the medial axis of the free space, *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 1024-1031, 1999.

# Analyzing Multi-Robot Box-Pushing

Angélica Muñoz Meléndez<sup>1</sup> and Alexis Drogoul<sup>2</sup>

<sup>1</sup> INAOE, Luis Enrique Erro No. 1  
72840 Tonantzintla Puebla, México  
munoz@inaoep.mx

<sup>2</sup> LIP6 - UPMC, Case 169 - 4, place Jussieu  
75252 Paris Cedex 05  
alexis.drogoul@lip6.fr

**Abstract.** In this paper, various experiments on multiple robot coordination are described. These experiments were conducted to identify experimentally how the use of communication can improve the performance of a box-pushing task by avoiding robot interferences. In these experiments, an intuitive approach consisting in observing and analyzing the outcome of several hypotheses is adopted.

## 1 Introduction

What is it necessary to constitute a colony of robots? What kind of mechanisms should be designed in order to enable a group of individualistic robots to perform actions jointly?

A mobile robot have to cope generally with highly dynamic environments, in which it is very hard to foresee all the potential situations that the robot may encounter. Specify in advance all the patterns of interaction between a mobile robot and its environment, as well as those patterns among various mobile robots is then, impossible.

Explaining the collective behavior of a group of robots is also complicated. The behavior of a robot colony may have properties that only occur at global level. Then, to derive the global behavior of a group of mobile robots directly from their individual control programs does not seem to be feasible.

For these reasons, we adopt an intuitive approach to research experimentally the elements that are required when designing robot colonies. For that, a very specific problem is formulated and then elements are added gradually to enable the establishment, the demonstration and reformulation of hypothesis. The purpose of this research is to get insight into the design trade-offs of multi-robot coordination. For that, various experiments on box-pushing were conducted using two mobile robots.

The paper is organized as follows, section 2 addresses related work, section 3 and 4 describes and discusses experiments and finally, section 5 concludes with a summary.

## 2 Related work

The coordination of a group of robots has been analyzed from different perspectives. Todt et al. [13] analyze and compare various methods proposed for multi-robot coordination. These methods deal with the problem of coordination intended to avoid collisions among robots. Collisions are usually avoided by organizing the trajectories that were previously planned by the robots [11].

Jäger [4] reports work on colonies of cooperative cleaning robots. In his experiments, each robot of the colony is controlled by a multi-agent architecture, where agents are in charge of tasks such as tracking objects and navigation. The area to be cleaned is divided into sections and each section is dynamically allocated to one robot. His method relies on local communication and has been tested using simulated robots mainly.

Parker [9] proposes a distributed architecture for multi-robot systems. Her architecture combines action-selection mechanisms and adaptation to enable a group of physical heterogeneous robots to perform cooperative tasks.

There are few works concerning the identification of the aspects and requirements to consider when designing robot colonies. Goldberg and Mataric [3], for example, examine the incidence that interference between robots has in the performance of a control system. They also use this interference as a criterion for evaluation of control systems with interesting results. Hayes [5] for his part, studies performance metrics for robot colonies. He analyses particularly the trade-offs between group size and efficiency in collective search tasks.

Finally, box-pushing is one of the tasks that has been explored by roboticists. Several planning algorithms [1, 2] have been proposed for individual box-pushing in simulated environments. Experiments on box-pushing using two or more physical robots have been reported [6, 7, 12, 15] in cooperative contexts mainly, that is, in situations where several robots push the same object. In this paper, the coordination of two robots that execute individually box-pushing in the same environment is explored.

## 3 Experiments

This section describes various experiments that were conducted in order to identify experimentally the key elements that are required to achieve multi-robot coordination. Experiments were conducted in such a way that a task, initially in charge of one robot, was gradually complexified towards a multi-robot context.

The research described in this paper is part of the MICRobES<sup>1</sup> project [10]. The experiments were conducted using two Pioneer 2-DX mobile robots of ActivMedia©, provided with odometers, bumpers, sonars, radio modem, video-camera and onboard computer.

---

<sup>1</sup> MICRobES is an acronym in French for Implementation of Robot Collectivities in a Social Environment.

### 3.1 Problem and settings

Box-pushing is the application chosen to explore the constitution of a robot colony. In this, robots have to search and move objects from one place to another place of the environment.

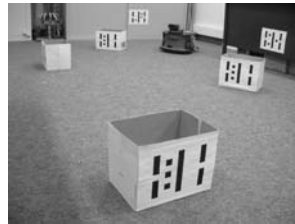
Robots use a cue-based recognition system developed by Viel [14] in order to differentiate visually significant elements of the environment. Both, the objects that robots have to transport and the places where the store and the supplier are located are indicated by cues consisting of bar codes (see figure 1).

The objects that robots have to transport in these experiments are boxes of  $20\text{cm} \times 30\text{cm} \times 23\text{cm}$ . A robot does not know where these objects are located, but it is able to recognize one of the box faces visually.

Robots have also a repertoire of actions and behaviors that are activated by external stimuli they perceive. Robots are able to execute simple actions such as `turn`, `go forward` and `go back`. More complex actions called behaviors such as `wander`, `line up`, `surround box`, `push box towards store`, `wait` and `send a message` have been implemented in and tested on our robots (for descriptions see the appendix).

Robots are programmed following a situated approach and therefore they do not have detailed information about their environment. They know, for instance, a general description about what an obstacle is from their sonar readings.

**Fig. 1.** A robot in an environment containing four objects and two bar codes over the wall. Bar codes consist of three black vertical bars on a white background. The patterns of a cue are composed of assembled black or white square-shaped unitary elements



### 3.2 Individual box-pushing

A first group of experiments designed to implement and test individual box-pushing is described below.

**Experiment 1. Foraging in a corridor.** The environment of this experiment is a corridor of  $1.70\text{m} \times 10\text{m}$ . At each extreme end of the corridor is placed a bar code, one indicates the supplier and the other indicates the store. Boxes are introduced into the corridor by a human operator in front of the supplier. These boxes are presented in such a way that the bar code side of boxes is oriented towards the robot, leaving enough space between the box and the supplier for the robot. In this environment there are mobile obstacles such as walkers, but there are not fixed obstacles that block the robot movements.

The robot is randomly located in the environment. Its goal is to transport boxes from the supplier to the store. For that, the robot should look for the supplier and stay close to it in order to detect and transport boxes with a minimum delay. Figure 2 shows the Finite State Machine used to control the robot and figure 3 shows a sequence of actions generated by this controller.

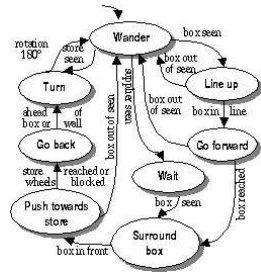


Fig. 2. Foraging in a corridor

Boxes		Time	
total	collected	waiting	total
1	1	0'00"	2'24"
2	1	0'00"	3'37"
5	3	1'26"	11'28"

Table 1. Experiment 1

Table 1 shows the results of three experiments of individual box-pushing. The total number of boxes that were presented to the robot, as well as the number of boxes that were accurately transported are indicated. Experiments lasted the timed required to transport all the boxes whose total number was known by the robot. A robot takes 2'24", 3'37" and 3'20" to push a box in environments where 1, 2 and 5 boxes were introduced respectively. That is the result of dividing *actual time* (total time – waiting time) into the number of boxes that were transported from the supplier to the store.



Fig. 3. Individual box-pushing in a corridor. From left to right, a robot perceives the box, reaches and surrounds it. Next the robot pushes the box towards the bar code located at the extreme end of the corridor

### 3.3 Coordinated box-pushing

A second group of experiments designed to extend individual box-pushing is described below. Through these experiments we inquire answers regarding possible



mechanisms of interaction between two robots that have to complete the same task within the same environment. We want to *discover* robot requirements in order to perform collectively a task that robots manage to do individually, rather than force robots to coordinate and avoid collisions.

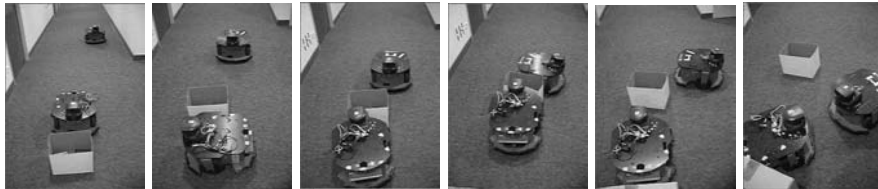
### Experiment 2. Individualistic foraging

**Hypothesis 1.** *A robot that is able to perform a task individually, should be able to perform the same task in a multi-robot context, because no collective effort is required to perform this task.*

The environment used in this experiment is exactly the same of experiment 1 but here, two robots are randomly located in the environment. Both robots, whose goals are to push boxes from the supplier to the store, are controlled by the Finite State Machine of figure 2.

As we could have foreseen, the presence of a second robot within the same environment affects the accomplishment of tasks. Because of robots do not have any way to recognize a partner, an important number of collisions happen (see figure 4).

Table 2 gives us an idea of the influence that collisions between robots have in their performance. The number of boxes that were accurately transported is indicated in bold characters. In the best case, a robot succeeded one of two attempts to transport a box. Both robots spent 6'47" and 28'43" to transport respectively 1 and 2 boxes, the sum of actual time for both robots divided into the number of boxes transported from the supplier to the store.



**Fig. 4.** Sequence of actions executed by individualistic foraging robots. From left to right, a robot tries to surround a box that is transported by another robot. Robots collide and damage the box, that is finally abandoned

**Experiment 3. Spatially coordinated foraging.** Most collisions in experiment 2 resulted of a spatial conflict produced by two individualistic robots unable to share their environment if designers do not provide them an explicit way to do that. Third experiment is designed to prove a new hypothesis.

**Hypothesis 2.** *An explicit division of the environment should be enough to avoid collisions between robots performing the same task within the same environment.*

For this experiment, the environment described in experiment 1 is divided in two sub-corridors and each one is allocated to one robot. In the extreme ends of the corridor, bar codes indicating the supplier and the store are placed. In the middle of the corridor, a double faced cue hangs from a string situated  $60\text{cm}$  from the floor. This cue indicates the store for the robot situated in the first sub-corridor and the supplier for the robot situated in the second one.

The robots' goal is to push boxes through the corridor, where each robot is in charge of one sub-corridor. Boxes are introduced into the environment by a human operator who is placed in the extreme that corresponds to the supplier.

Table 3 summarizes the results of three experiments where robots  $p03$  and  $p01$  were in charge of sub-corridors. In the best case, robots succeeded one of five attempts to transport the same box. A success was here a box accurately transported to the second store. Two robots in a divided corridor spent  $10'38''$  and  $14'41''$  to transport the same box, of a total of 2 and 3 boxes respectively. Robots got a slightly better score but they still had problems of coordination.

The line that divides the corridor is a critic zone which robots cannot deal with. It happens often that the second robot reaches and tries to push a box that is been pushed by the first robot. Indeed, the controller of the second robot triggers the action `go forward` when a box is perceived, producing a collision (see figure 5). This situation could be solved if `go forward` was triggered once the box is perceived under a fixed threshold. This solution is however partial, because the calculation of distances to mobile objects is always approximate.



**Fig. 5.** Sequence of actions executed by robots foraging in a divided corridor. From left to right, robot  $p03$  pushes a box while robot  $p01$  waits in front of its own supplier.  $p01$  tries to reach the box before than  $p03$  finished, producing then a collision

**Experiment 4. Spatio-temporally coordinated foraging.** Last experiments were based on the same control mechanism. We wanted to explore the possibilities of individual box-pushing and test the robustness and weakness of a general behavior in a multi-robot context.

We have seen that, even though an explicit division of the environment was introduced, robots were not always able to coordinate their actions. Previous results showed that robots acting within the same environment need an explicit mechanism for coordination.

**Hypothesis 3.** An explicit mechanism for synchronization is required for multi-robot coordination.

In this experiment, synchronization is achieved through a simple mechanism of direct communication between robots. Robots' controller is here modified by introducing a behavior that enables one robot -the *server*- to send a message to another robot -the *client*- (see figure 6).

The robot in charge of the first sub-corridor, the server, behaves more or less in the same way described previously; however this robot sends here a message to its partner once it has completed its task. The robot in charge of the second sub-corridor, the client, does not longer executes *wander* as its first action. It waits and only starts to move if it receives a message of its partner indicating that a box is waiting in the middle of the corridor.

Table 4 shows the results of our last experiments. From the number of boxes that were collected, in bold characters, we confirm that robots have improved their performance. They succeeded, in the best case, two of five attempts to transport the same box. This improvement is due to their competence to communicate directly with each other, that enables them to avoid the collisions produced in experiment 3. The sum of actual time for both robots divided into the number of boxes transported from the supplier to the store was 6'36" to transport 1 box; 7'45" and 9'30" to transport 2 boxes of a total of 2 and 3 boxes respectively.

Total Robot boxes	Robot	Boxes		Time	
		detected	collected	wait.	exp.
1	p01	1	<b>0</b>	0'00"	1'45"
	p03	2	<b>0</b>	0'00"	1'47"
2	p01	2	<b>0</b>	2'20"	4'56"
	p03	2	<b>1</b>	0'48"	4'59"
3	p01	5	<b>1</b>	0'14"	14'46"
	p03	5	<b>1</b>	0'36"	14'47"

Table 2. Experiment 2

Total Robot boxes	Robot	Boxes		Time	
		detected	collected	wait.	exp.
1	p03	1	<b>1</b>	0'24"	3'43"
	p01	1	<b>0</b>	2'14"	3'48"
2	p01	2	<b>1</b>	0'00"	5'10"
	p03	3	<b>1</b>	0'00"	5'28"
3	p01	5	<b>2</b>	0'42"	9'05"
	p03	4	<b>1</b>	2'47"	9'05"

Table 3. Experiment 3

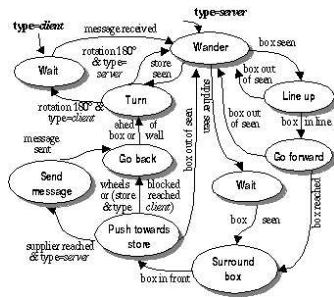


Fig. 6. Foraging with communication

Total Robot boxes	Robot	Boxes		Time	
		detected	collected	wait.	exp.
1	p01	2	<b>1</b>	0'00"	4'19"
	p03	1	<b>1</b>	1'58"	4'15"
2	p01	3	<b>2</b>	1'26"	6'06"
	p03	2	<b>2</b>	2'59"	6'04"
3	p01	3	<b>2</b>	0'00"	6'02"
	p03	2	<b>2</b>	2'31"	5'59"

Table 4. Experiment 4

## 4 Discussion

Even though the competences of robots involved in previous experiments are not exactly the same, we can compare and discuss their results. The purpose of this comparison is to figure out the role that communication has in multi-robot interaction.

Figures 7, 8 and 9 illustrate the behaviors that were executed by robots during the foraging of three boxes in experiments 2, 3 and 4 (rows 5 and 6 of tables 2, 3 and 4). These experiments lasted for 14'47", 9'05" and 6'02".

In experiment 2 according to figure 7, the number of fails -the triggers of behavior `line up` and `push box` that did not succeed- is important. There are less attempts that did not succeed in experiment 3 as indicated in figure 8. In addition to the problems of coordination that were discussed, a robot may lose boxes for other reasons such as an instability problem caused by box placements and boxes out of the sight of robots.

Collisions between robots disappeared in experiment 4 when we introduced communication, as we can see in figure 9. This figure shows also that synchronization avoids useless movements of *p03*.

Is communication essential to achieve multi-robot coordination? In experiments where robots did not communicate, when an individualistic foraging robot met a partner, it reacted as if it were an obstacle. However, when they met while pushing a box they ignored themselves and collide.

A thorough coordination of the actions executed by robots may avoid these situations, as we can see in experiment 3. We think, however, that parameters tuning should ease by communication. Thus, we conclude that communication is desirable for achieving collective actions, even if communication is not necessarily achieved in a direct form.

## 5 Conclusions

We described our research in collective robotics applying an intuitive approach. Four experiments of foraging robots in single and multi-robot contexts were presented. We conducted these experiments *naively*, a problem was defined and a number of hypothesis -apparently obvious- were formulated, explored and demonstrated experimentally, analyzing the outcomes of several tests.

We consider that this approach is useful when analyzing multi-robot coordination strategies. The main contribution of our experiments is that the requirements of a multi-robot system were detected experimentally, rather than imposed *a priori* by designers.

Future work will focus on extending our box-pushing mechanisms to deal with collective box-pushing, a task involving the joint efforts of a robot colony.

## Appendix

Description of behaviors used in the experiments (for more details see [8]):

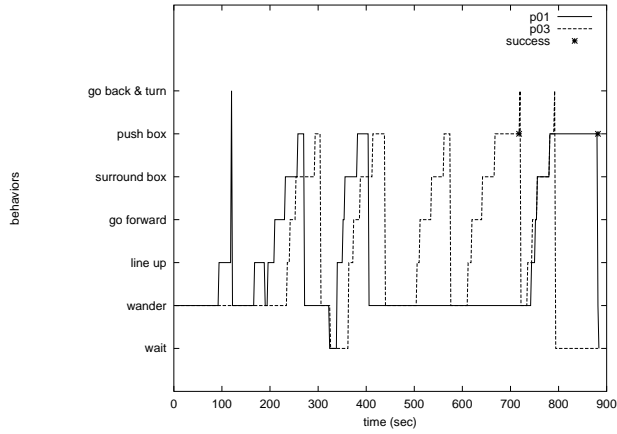


Fig. 7. Behaviors of two individualistic foraging robots

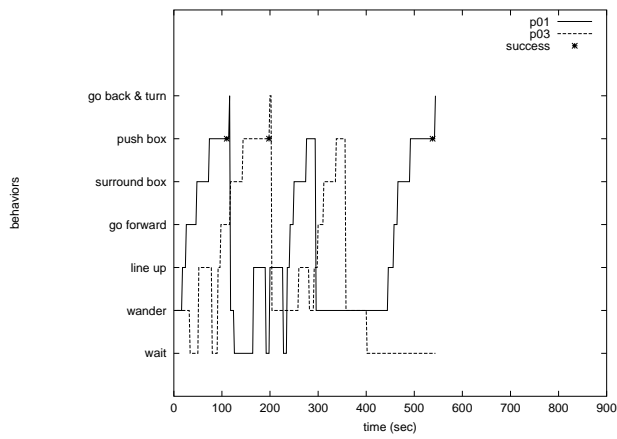


Fig. 8. Behaviors of two spatially coordinated foraging robots

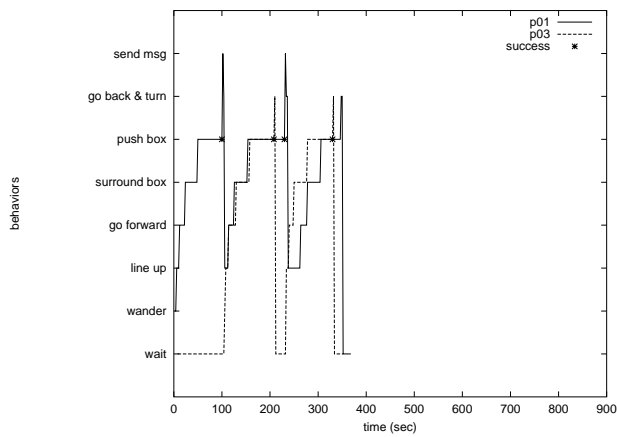


Fig. 9. Behavior of two spatio-temporally coordinated foraging robots

- **Wander.** The robot explores its environment randomly avoiding obstacles. Robot moves at a speed of  $250\text{mm}/\text{sec}$ .
- **Line up.** The robot aligns itself with a bar code and stands in line.
- **Push towards store.** The robot pushes a box detected in front by its sonars. Robot pushes in a straight line at a speed of  $100\text{mm}/\text{sec}$  until it is at a given distance from a given bar code.
- **Wait.** The robot remains at the same position.
- **Surround box.** The robot goes towards a box and moves to the opposite side of the bar code side. Robot moves at a speed of  $200\text{mm}/\text{sec}$ .
- **Send a message.** The robot sends a message to a given robot. Robots communicate via sockets, transmitting simple messages such as “box”.

## References

1. O. Ben-Shahar, E. Rivlin: Practical Planning for Rearrangement Tasks. *IEEE Transactions on Robotics and Automation*, 14(4), 549-559, 1998.
2. E.D. Demaine, R.A. Hearn, M. Hoffmann: Push-2-F is PSPACE-Complete. *Proc. of the 14th Canadian Conference on Computational Geometry*, 31-35, 2002.
3. D. Goldberg, M.J. Mataric: Interference as a Tool for Designing and Evaluating Multi-Robot Controllers. *Proc. of the National Conf. on Artificial Intelligence, AAAI*, 637-642, 1997.
4. M. Jäger: Cooperating Cleaning Robots. *Distributed Autonomous Robotics Systems 5*, H. Asama et al. (eds), 237-246. Springer Verlag, 2002.
5. A.T. Hayes: How Many Robots? Group Size and Efficiency in Collective Search Tasks. *Distributed Autonomous Robotics Systems 5*, H. Asama et al. (eds), 289-298. Springer Verlag, 2002.
6. C.R. Kube, E. Bonabeau: Cooperative Transport by Ants and Robots. *Robotics and Autonomous Systems*, 30, 85-101. Elsevier, 2000.
7. M.J. Mataric, M. Nilsson, K.T. Simsarian: Cooperative Multi-Robot Box Pushing. *Proc. of the 1999 IEEE Int. Conf. on Intelligent Robots and Systems*, 556-561, 1995.
8. A. Muñoz-Meléndez: Coopération Située: une Approche Constructiviste de la Conception de Colonies de Robots. PhD thesis, University Pierre et Marie Curie, 2003.
9. L.E. Parker: Heterogeneous multi-robot cooperation. PhD thesis, MIT, 1994.
10. S. Picault, A. Drogoul: The MICRobES Project, an Experimental Approach towards Open Collective Robotics. *Distributed Autonomous Robotic Systems 4*, L.E. Parker et al. (eds), 481-482. Springer-Verlag, 2000.
11. S. Qutub, R. Alami, F. Ingrand: How to Solve Deadlock Situations within the Planning Paradigm for Multi-robot Cooperation. *IEEE/RSJ Int. Conf. on Robotics and Automation*, 1610-1615, 1997.
12. A. Sudsang, F. Rothganger, J. Ponce: Motion Planning for Disc-shaped Robots Pushing a Polygonal Object in the Plane. *IEEE Transactions on Robotics and Automation*, 18(4), 550-562, 2002.
13. E. Todt, G. Raush., R. Suárez: Analysis and Classification of Multiple Robot Coordination Methods. *Proc. of the 2000 IEEE Int. Conf. on Robotics and Automation*, 3158-3163, 2000.
14. P.-E. Viel: Reconnaissance Individualisée de Robots et de Zones. MSc thesis (DEA de Robotique et Systèmes Intelligents), University Pierre et Marie Curie, 2001.
15. S. Yamada, J. Saito: Adaptive Action Selection without Explicit Communication for Multi-robot Box-pushing. *Proc. of the 1999 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 1444-1449, 1999.

# Current work at the Robotics group of the Universidad de Guanajuato

Víctor Ayala Ramírez and Raúl E. Sánchez Yañez

Universidad de Guanajuato FIMEE  
Tampico 912, Col. Bellavista  
Salamanca, Gto.  
{ayalav,sanchezy}@salamanca.ugto.mx

## 1 Introduction

At Universidad de Guanajuato at Salamanca, robotic techniques are developed in LaViRIA (Laboratory of Vision, Robotics and Artificial Intelligence). Our work is directed towards developing methods that integrate computer vision techniques and artificial intelligence techniques (mainly based in soft computing) in robotics applications and its implementations in the real robots. Main applications we are pursuing are related with agricultural robotics and mini-robotics.

Our group is composed of three researchers and two others in formation and about fifteen students developing thesis for the two academic programs to whom our laboratory contributes, to know: the B. Eng. program in Electronics and Communications, where it is possible to follow a Computer Engineering major; and the Master program in Electrical Engineering option Instrumentation and Digital Systems that includes a research theme on signal processing.

## 2 Main research themes

### 2.1 Computer vision

- Texture analysis (detection, classification and segmentation).

Texture analysis is essential for scene interpretation in outdoor images.

We have developed works [1] on texture detection, where the main concern is to detect if there is some texture on an image. Regions where texture is detected can then be segmented and processed in the scene interpretation stage. Our approach is based on the application of fuzzy inference rules to the homogeneity measures over some neighborhood window of the images. Our method has been validated on synthetic compositions of texture patches.

Concerning texture classification, we have developed two approaches: one that uses fuzzy inference rules [2] and another that uses GLCM but interpreting them using a Bayesian framework [3]. Both methods were tested on images from the Brodatz texture database.

- Object recognition and modeling.

Object recognition is very important in robotic vision applications as in visual servoing or visual tracking tasks. We have developed some methods using Hausdorff distance to compare object models [4]. Another method [5] [6] describes regions in an image as color patches in the CIE Lab color space. We have also worked in active object recognition using mutual information [7] where we take advantage of an active sensor to assign a class label for an object by choosing views that provide optimal information to reduce ambiguity. This procedure is carried out using a mutual information framework and a Markovian decision process.

- Motion analysis.

Motion analysis is the main component of dynamic tasks in robotic vision. We work on the motion detection problem [8] where we use a Gaussian mixture approach to detect motion zones that could be used as target initializations in a target tracking system.

Another work [9] deals with the problem of interaction of several modalities of tracking methods in order to optimize the visual tracking task of a mobile robot by following target characteristics and tracking performance over the temporal evolution of the scenes acquired by a mobile robot.

## 2.2 Soft computing

- Soft computing libraries.

We have developed object-oriented libraries that implement basic functionalities of soft computing techniques. Fuzzy logic and genetic algorithms [10] can then be implemented in our systems easily in order to develop intelligent systems for robotics applications.

- Model identification using fuzzy systems.

We have also worked on modeling systems from input/output numerical evidence [11]. This has led us to develop some original methods in fuzzy clustering and to implement some other classical methods as the Fuzzy-C means method.

- Genetic algorithm (GA) applications.

Some work is in progress to solve a packaging problem of arbitrary shapes by using genetic algorithms [12]. This approach is intended to help leather and textile industries, two very important manufacturing sectors in our region. We are developing also methods for path planning in robotics that optimize trajectories for mini-robots by considering uncertainty in the actual path execution in the fitness function.

## 2.3 Robotic applications

- Mini-robotics navigation and mapping

In this subject, we have worked on handling uncertainty in robot position [13] and how to implement in mini-robots from the Lego Mindstorms series [14]. We have also used in the past a robot simulator environment developed



at our laboratory that uses a reactive navigation strategy using fuzzy logic. In the near future, we will have an all-terrain robot that will serve us to validate our previous approaches to robot navigation and mapping.

- Visual servoing and tracking  
A fuzzy approach to visual servoing and its optimization by means of genetic algorithms has been dealt by Pérez [15]. Color target tracking has been developed by Montecillo [5]. In its work, target has been represented by fuzzy rules that use values in the CIE Lab color space.
- Parametric shape detection using GA  
Genetic algorithms were used in a parametric object recognition problem by Garcia [16]. In this work, circle and ellipses are detected in real-world images. Some extensions for this method will be applied in industrial metrology problems.
- Modality control of active sensors.  
We have worked in the recent years in the modality selection problem for active sensors (also known as active sensing). Our initial work was directed towards path servoing problems [17] [18]. We have recently addressed the active object recognition problem [19] [7].
- Robot architectures  
Some works have dealt the sensor/actuator interaction in a robot architecture [20] [21] [22]. Our approach is based on the concept of the intelligent sensor(actuator) that can be in charge of high level tasks. This enables the robot to use low bandwidth when communicating with these devices.

### 3 Main research projects

- Topological navigation in structured and semi-structured environments.  
We are participants of the French-Mexican Laboratory on Informatics (LAFMI). We are partners with the Laboratory of System Analysis and Architecture in Toulouse, France to develop techniques for the topological navigation of mobile robots [23]. A first application of these techniques will be demonstrated on a high-clearance agricultural vehicle designed at FIMEE.
- Fuzzy texture detection for topological navigation.  
In this project, we use fuzzy approaches for texture detection, segmentation and classification in natural environments [24]. Main tasks include road extraction and place recognition.
- Active perception in robotics.  
Active perception for outdoor robotics has been the object of a Ecos Nord project [25]. This project has contributed mainly to fund mobility programs for researchers in the partner robotics teams during four years.  
Some active perception functionalities for mobile robots have been developed in a CONACYT project [26]. This project has contributed with funding to buy some vision and computing processing equipment.

## 4 Conclusions and prospectives

Our group has established a steady flow of human resources trained in our laboratory. We are working to set up a doctoral program and, in parallel, to build some partnerships with other programs in Mexico and abroad to have doctoral students that could help us in the development of our laboratory. We are working towards our consolidation as a research group in robotics applications. We have already developed some academic research and we have established some partnerships with other laboratories around the world. We need however to develop knowledge transfert skills in order to impact Mexican industrial environment and to become more independent with relation to funding issues by obtaining some contracts with Mexican industries.

## References

1. Del Moral Perea, A.: Detección de textura en imgenes usando un modelo difuso de inferencia. M.E.E. thesis, Universidad de Guanajuato FIMEE (2004)
2. Ramírez Chiu, M.A.: Una aproximación difusa a la caracterización de propiedades de texturas visuales. M.E.E. thesis, Universidad de Guanajuato FIMEE (2004)
3. Ayala-Ramírez, V., Obara-Kepowicz, M., Sánchez-Yáñez, R.E., Jaime-Rivas, R.: Bayesian texture classification method using a random sampling scheme. In: Proc. of the 2003 IEEE Int. Conf. on Systems, Man and Cybernetics, Washington, DC, IEEE Press (2003) 2065–2069
4. Ayala-Ramírez, V., Parra, C., Devy, M.: Active tracking based on hausdorff matching. In: Proc. 15th Int. Conf. on Pattern Recognition (ICPR'2000). Volume 4., Barcelona, Spain, IEEE Computer Society Press (2000) 706–709
5. Montecillo Puente, F.J.: Sistema de seguimiento de objetos en tiempo real mediante caracterización difusa del color. B. Eng. thesis, Universidad de Guanajuato (2003)
6. Montecillo-Puente, F.J., Ayala-Ramírez, V., Pérez-García, A., Sánchez-Yáñez, R.E.: Fuzzy color tracking for robotic tasks. In: Proc. of the 2003 IEEE Int. Conf. on Systems, Man and Cybernetics, Washington, DC, IEEE Press (2003) 2769–2773
7. Trujillo-Romero, F., Ayala-Ramírez, V., Marín-Hernández, A., Devy, M.: Active object recognition using mutual information. In: Proc. MICAI 2004: Advances in Artificial Intelligence. Volume 2972 of LNAI., Mexico City, Mexico, Springer Verlag (2004) 672–678
8. Rincón Casián, J.A.: Detección de movimiento usando mezclas de gaussianas (in progress). B. Eng. thesis, Universidad de Guanajuato (2004)
9. Marín-Hernández, A., Ayala-Ramírez, V., Devy, M.: Collaborative tracking for robotic tasks. In: Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'2003). Volume 1., Las Vegas, NE, IEEE Press (2004) 272–277
10. Pérez-García, A., Ayala-Ramírez, V., Jaime-Rivas, R.: Una librería genérica de algoritmos genéticos para aplicación en problemas de optimización. In: Memoria 5to. Congreso Mexicano de Robótica, San Luis Potosí, Mexico (2003) 141–146
11. Sánchez-Yáñez, R.E., Ayala-Ramírez, V., Jaime-Rivas, R.: Fuzzy models for system identification. In: Proc. of the 2003 IEEE Int. Conf. on Systems, Man and Cybernetics, Washington, DC, IEEE Press (2003) 3330–3335
12. Ponce Pérez, A.: Empaquetamiento de formas arbitrarias usando algoritmos genéticos (in progress). B. Eng. thesis, Universidad de Guanajuato (2004)

13. López Padilla, R.: Tareas de navegación de un robot móvil bajo incertidumbre. B. Eng. thesis, Universidad de Guanajuato (2004)
14. Martínez Labrada, E.: Estrategias de navegación y construcción de mapas usando mini-robots (in progress). B. Eng. thesis, Universidad de Guanajuato (2004)
15. Pérez García, A.: Servocontrol de una cámara activa usando técnicas geno-difusas. B. Eng. thesis, Universidad de Guanajuato (2002)
16. García Capulín, C.H.: Detección de formas paramétricas en imágenes (in progress). M.e.e. thesis, Universidad de Guanajuato FIMEE (2004)
17. Razo-Razo, M.A., Ayala-Ramírez, V.: Exploración coordinada de un ambiente mediante métodos de potencial. In: Memoria CONIELECOMP 2003, Puebla, Mexico, IEEE Mexico (2003) 201–205
18. Trujillo Romero, F.: Control de modalidades de una cámara activa usando funciones de utilidad. M.E.E. thesis, Universidad de Guanajuato (2002)
19. Pérez-Mendoza, S., Ayala-Ramírez, V., Pérez-García, A., Sánchez-Yáñez, R.E.: Visual servoing for micro-robotic platforms. In: Proc. 14th Int. Conf. on Electronics, Communications and Computers (CONIELECOMP'2004), Veracruz, Mexico, IEEE Computer Society Press (2004) 18–22
20. Dorado Prieto, A.: Interfaz de comunicación del sistema de instrumentación de un robot móvil. M.E.E. thesis, Universidad de Guanajuato (2002)
21. Alvarado Martínez, T.: Especificación de un actuador inteligente para tareas robóticas. M.E.E. thesis, Universidad de Guanajuato (2003)
22. Pérez Mendoza, J.S.: Servocontrol visual para plataformas mini-robóticas. B. Eng. thesis, Universidad de Guanajuato (2003)
23. Ayala Ramírez, V., Devy, M.: Concepción de funciones de navegación topológica de robots móviles en ambientes estructurados y semi-estructurados. Project funded by LAFMI, Universidad de Guanajuato and Laboratoire d'Analyse et d'Architecture des Systèmes (2002-2004)
24. Sánchez Yáñez, R.E.: Navegación topológica de robots utilizando análisis difuso de textura. Project funded by SEP-PROMEP, Universidad de Guanajuato (2003-2005)
25. Jaime Rivas, R., Devy, M.: Percepción activa para robótica de exterior. Project funded by ANUIES (México) and Comité ECOS (France) No. M99M01, Universidad de Guanajuato and Laboratoire d'Analyse et d'Architecture des Systèmes (1999-2003)
26. Ayala Ramírez, V.: Concepción de funciones de percepción activa. Project funded by CONACYT I-39328A, Universidad de Guanajuato (2002)

# Grupo de Robótica del INAOE

Angélica Muñoz Meléndez

INAOE - Ciencias Computacionales  
Luis Enrique Erro No. 1  
72840 Tonantzintla Puebla, México  
munoz@inaoep.mx

**Resumen** En este artículo se presentan los intereses y las actividades del Grupo de Robótica del Instituto Nacional de Astrofísica, Óptica y Electrónica. Este laboratorio forma parte de la Coordinación de Ciencias Computacionales del instituto y en él participan tres investigadores principales y siete estudiantes de maestría trabajando principalmente en áreas relacionadas con la robótica móvil.

## 1. Introducción

El Grupo de Robótica del INAOE tiene como objetivo la búsqueda de soluciones novedosas a problemas de robótica en áreas como percepción y control de robots móviles, concepción de robots adaptativos, y diseño y control de sistemas multi-robot.

El grupo se compone de tres investigadores principales y siete estudiantes de maestría asociados, quienes desarrollan actualmente su trabajo de grado bajo la supervisión de los primeros. El grupo como tal se conformó recientemente con la incorporación de dos nuevos miembros a la planta de investigadores de la Coordinación de Ciencias Computacionales del instituto.

Las actividades de los miembros del grupo se articulan alrededor de dos ejes, la investigación y, la docencia y la difusión del conocimiento, las cuales se describen a continuación.

## 2. Investigación

Las actividades del grupo de roboticistas de la Coordinación de Ciencias Computacionales del INAOE se centran esencialmente en el desarrollo de investigación básica y aplicada en temas como percepción robótica, arquitecturas de control, técnicas de aprendizaje y coordinación multi-robot.

El grupo busca proponer proyectos que permitan la integración de estudiantes del INAOE y externos a sus actividades de investigación. Dos son las áreas en las que se desarrollan principalmente estos proyectos: técnicas de aprendizaje automático y robótica colectiva.

## 2.1. Técnicas de Aprendizaje Automático

En esta área se busca proponer métodos que permitan a un robot desarrollar o ajustar sus habilidades en interacción con su medio ambiente, con mínima o nula intervención del programador.

Los trabajos desarrollados por los investigadores que trabajan en esta área se centran en el desarrollo de técnicas que permitan al sistema de control de un robot aprender a mapear las entradas de sus sensores con las salidas de sus efectores para producir comportamientos complejos, como la manipulación [1] o la evasión de obstáculos [4].

Los investigadores de esta área han aplicado también diversas técnicas de aprendizaje automático en la navegación de robots móviles basada en visión [2] y basada en múltiples sensores [5].

## 2.2. Robótica Colectiva

La robótica colectiva busca diseñar sistemas compuestos de múltiples robots capaces de resolver conjuntamente un problema. En ella convergen disciplinas tales como la inteligencia artificial distribuida y la robótica móvil.

Un sistema multi-robot está compuesto regularmente por varios robots homogéneos y de bajo costo. Estos sistemas están orientados a la resolución de problemas para cuya solución un único robot no es suficiente o resulta muy costoso -en términos de diseño y de tiempo por ejemplo-, como pueden ser el transporte de objetos voluminosos, la manipulación de material peligroso, la exploración y vigilancia de terreno.

Los proyectos de robótica colectiva del laboratorio son dos:

- Generación y mantenimiento de formaciones en robótica colectiva
- Diseño de mecanismos auto-reconfigurables

El primer proyecto se interesa por las formaciones de robots simulados. Las formaciones son estructuras espaciales típicamente lineales o poligonales constituidas por un conjunto de robots en donde cada uno representa un punto o vértice de la estructura en cuestión, y las líneas entre los vértices representan los contornos de la estructura.

El segundo proyecto se interesa por el diseño y el control de prototipos físicos de sistemas con capacidades de auto-reconfiguración, sistemas multi-robot capaces de modificar y de reorganizar su morfología para adaptarse al ambiente en el que están inmersos, sin intervención humana.

La coordinación de movimientos es un aspecto clave de los sistemas multi-robot. Los investigadores de esta área se interesan en el desarrollo de métodos descentralizados de coordinación de sistemas multi-robot [3]. El interés por explorar estos métodos radica en su potencial para coordinar sistemas multi-robot en ambientes parcialmente desconocidos y en ambientes conteniendo obstáculos.

## 3. Docencia y Difusión

Entre las actividades de los investigadores que participan en el Grupo de Robótica, la docencia y la difusión ocupan también un lugar importante.

En lo que se refiere al apoyo a los posgrados en ciencias computacionales que ofrece al INAOE, los investigadores del grupo están a cargo de la impartición de materias como *Introducción a la Robótica*, *Aprendizaje Automático* y *Procesamiento Digital de Señales*. Otras materias que también se imparten en los posgrados relacionadas con la robótica son *Interacción Humano-Computadora*, *Introducción a las Tecnologías del Lenguaje* y *Análisis Digital de Imágenes*. Estos cursos son facultativos y se imparten durante el primer año de estudios de la maestría o del doctorado.

Los investigadores del grupo se interesan también por la organización de eventos que contribuyan a la difusión del conocimiento. Los concursos robóticos, por ejemplo, han probado ser una excelente vía para difundir el trabajo realizado en las aulas y en los laboratorios por investigadores interesados en áreas del conocimiento tan diversas como las matemáticas, la física, la electrónica y la computación. A través de estos eventos es posible acceder a un gran público integrado por niños y jóvenes que se sentirá quizás atraído por esta manera de “aprender jugando” y más tarde, por la ciencia.

En agosto de 2004, el Grupo de Robótica co-organiza con la Maestría en Inteligencia Artificial de la Universidad Veracruzana el *Primer Torneo Mexicano de Robots Limpiadores* [6]. Diecisiete equipos provenientes de doce instituciones universitarias del país confrontarán amistosamente las técnicas y los enfoques propuestos para programar robots que deben limpiar una habitación simulada.

## 4. Perspectivas

Entre las perspectivas del grupo destacan la incorporación de un mayor número de estudiantes a los proyectos a cargo de los investigadores principales.

La interacción con otros grupos de roboticistas trabajando en México figura también entre nuestras perspectivas. Nuestro grupo está dispuesto a participar y a colaborar en la constitución de redes temáticas y de excelencia en el área de robótica, y a sumarse en general a todo esfuerzo encaminado a crear foros de intercambio de ideas entre investigadores roboticistas mexicanos y de la región.

## Referencias

1. Fuentes O., Nelson R.C. (1988) Learning Dextrous Manipulation Strategies for Multifingered Robot Hands Using the Evolution Strategy. *Machine Learning* Vol. 31 (1-3), pp. 223-237. Kluwer Academic Pub.
2. Lara Ramírez J.E. (2004) Robot Navigation and Map Construction Using Machine Learning and Probabilistic Models. Tesis de Maestría en Ciencias Computacionales, INAOE.
3. Muñoz Meléndez A., Drogoul A., Viel P.-E. (2004) Collective Behavior as Assembling of Spatial Puzzles. *MICAI 2004: Advances in*

- Artificial Intelligence*, pp. 901-910, Monroy R. et al. (eds). Lecture Notes in Artificial Intelligence, Springer-Verlag.
4. Salim A., Fuentes O., Muñoz A. (2004) Development of Local Perception-Based Behaviors for a Robotic Soccer Player. Por publicarse en *Proceedings of the IX Ibero-American Conference on Artificial Intelligence (IBERAMIA)*. Lecture Notes in Artificial Intelligence, Springer-Verlag.
  5. Rao R.P.N., Fuentes O. (1998) Hierarchical Learning of Navigational Behaviors in an Autonomous Robot using a Predictive Sparse Distributed Memory. *Machine Learning* Vol. 31 (1-3), pp. 87-113. Kluwer Academic Pub.
  6. Primer Torneo Mexicano de Robots Limpiadores:  
<http://ccc.inaoep.mx/~torneo-de-robot/>

# **Cómputo móvil**

José Antonio García Macías  
Oscar Iván Lepe Aldama  
Rolando Menchaca Méndez  
(Editores)





## Introducción

Las computadoras fijas han dejado de ser el medio preponderante para procesar y comunicar información. En la actualidad son comunes los dispositivos móviles tales como organizadores personales, computadoras portátiles (laptops, tablets), teléfonos celulares inteligentes y otros nuevos dispositivos que aparecen día con día. Nos encontramos con dispositivos que se comunican a través de redes inalámbricas y que permiten la movilidad de sus usuarios. Esta movilidad tiene gran impacto no solamente en las redes y sus protocolos, sino también en las aplicaciones, en la formas de colaboración utilizando la tecnología, en la interacción con los dispositivos, etc. El cómputo móvil es la rama de las ciencias computacionales que investiga las arquitecturas, organizaciones, algoritmos, estructuras de datos, protocolos de comunicaciones y técnicas de construcción de sistemas (ingeniería de software) que eficazmente permitan sacar el mejor provecho de los dispositivos móviles y las redes inalámbricas para ofrecer donde y cuando sea servicios de información particularizados a las condiciones del entorno del usuario. Y en este sentido el cómputo móvil es hoy todavía un ámbito emergente, que presenta un número importante de retos científicos y tecnológicos.

Algunos de los tópicos relevantes que se tratan en el taller son: cómputo ubicuo, modelado de usuarios y personalización, cómputo consciente del contexto, funcionamiento desconectado y débilmente conectado, agentes, redes ad-hoc, microarquitecturas para ahorro de energía, por nombrar solo unos cuantos. El propósito del taller es reunir a investigadores, estudiantes y profesionales del cómputo móvil, y proveer un foro para que miembros de las comunidades académica e industrial discutan y presenten trabajos o propuestas de trabajo sobre tópicos referentes al cómputo móvil, redes inalámbricas y temas relacionados. En principio, el alcance del taller es de carácter nacional, pero son bienvenidas las contribuciones que se puedan recibir del extranjero.

Este taller es una continuación de los esfuerzos que iniciaron en el primer taller, el cual se llevó a cabo durante el ENC 2003 en Tlaxcala. En este segundo taller se recibieron 15 trabajos, de los cuales tras un proceso de revisión se han aceptado 10 para su presentación.

## **Comité Científico**

Copresidentes:

Oscar Iván Lepe Aldama, CICESE, Ensenada, México

J. Antonio García Macías, CICESE, Ensenada, México

Rolando Menchaca Méndez, CIC-IPN, D.F., México

Jesús Favela, CICESE, Ensenada, México

Gabriel González, CENIDET, Cuernavaca, México

Guillermo Licea, UABC, Tijuana, México

Bárbaro J. Ferro Castro, ITESM-CCM, D.F., México

Victor Sosa - CENIDET, Cuernavaca, México

Matías Alvarado Mentado, IMP, D.F., México

Felipe Rolando Menchaca García, CENAC-IPN, D.F., México

Luis Enrique Vizcarra Corral, UABC, Ensenada, México

Roberto Gómez Cárdenas, ITESM-CEM, Edo. de México, México

Luis Villaseñor González, CICESE, Ensenada, México

# Service Discovery in Mobile Ad-hoc Networks by Extending the AODV Protocol

Dante Arias-Torres, José Antonio García-Macias

Departamento de Ciencias de la Computación  
Centro de Investigación Científica y Educación Superior de Ensenada, CICESE.  
e-mail : {darias, jagm}@cicese.mx  
Km. 107 Carretera Tijuana – Ensenada, CP 22860, Ensenada, B.C. México.

**Abstract:** In recent years, mobile devices have become increasingly popular; also, wireless communication technologies such as WiFi and Bluetooth have achieved commercial success. These facts have made possible the creation of networks with mobility support, such as mobile ad-hoc networks (MANETs). In this field, there are still several challenges that need to be overcome; one of these is automatic service discovery. Existing solutions such as Jini, SLP, and others can not be applied to MANETs, mainly due to the frequent motion of nodes. In this paper we propose a service discovery architecture based on extension to the AODV routing protocol that takes advantage of the route discovery mechanisms to also perform service discovery.<sup>1</sup>

**Resumen:** En los últimos años, los dispositivos móviles se han popularizado, así mismo, las tecnologías de comunicación inalámbricas como WiFi y Bluetooth han logrado éxito comercial. Estos acontecimientos han hecho posible la creación de redes con soporte a la movilidad, como las redes ad hoc móviles (MANETs). En este campo, aun existen muchos retos que tienen que ser superados; uno de estos es el descubrimiento automático de servicios. Soluciones existentes tales como Jini, SLP, y otras no pueden ser aplicadas a las MANETs, principalmente debido al frecuente movimiento de los nodos. En este artículo se propone una arquitectura para el descubrimiento de servicios basada en la extensión del protocolo de enrutamiento AODV que toma ventaja de los mecanismos de descubrimiento de rutas de este protocolo para realizar el descubrimiento de servicios.

**Keywords:** Mobile Ad-hoc Networks, Service Discovery, AODV.

## 1. Introduction.

The technological revolution of recent years encompasses a radical change in society; we are in the transition from the personal computer era (one device per person) to the era of ubiquitous computing where people will use, at the same time and sometimes

---

<sup>1</sup> We wish to acknowledge the financial support provided by the Mexican Science and Technology Council (CONACyT).

without knowing it, many digital platforms, anywhere, anytime [16]. In such an environment, interactions between devices will be carried out spontaneously and transparently, being very common the formation of mobile ad-hoc networks (MANETs) [2] with these devices. In these scenarios, the resources (shared or not) can be represented as services, which can be offered by devices (such as printers, sensors, personal computers, etc.) or by specialized software [10]. The presence of ad-hoc networks between devices and the representation of resources as services does not suffice to have an ubiquitous computing environment; it is also necessary to have a mechanism that allows discovering available resources. The problem of service discovery has been studied extensively, and there are existing solutions such as Jini [12], SLP [8], UPnP [6], etc. These mechanisms have been developed for fixed networks with pre-established infrastructure, thus, they can not be applied to networks with limited resources, where nodes move frequently and where there is not a fixed computing infrastructure available (e.g., ad-hoc networks) [1]. For this reason, the interest of provide solutions to service discovery in this networks have increased, and there are existing solutions like NOM [4], Konark [10], etc. In this paper we present a solution that extends a routing protocol for mobile ad-hoc networks (i.e., the AODV protocol) in such way that it also performs service discovery. This document is organized as follows: section 2 describes the AODV protocol, focusing on unicast routing; section 3 describes service discovery by extending AODV, section 4 describes a service discovery architecture that uses the mechanism proposed in section 3, while section 5 presents future work and conclusions.

## 2. AODV.

Ad-hoc On Demand Vector (AODV) routing was designed specifically to be used in mobile ad-hoc networks. This protocol is categorized as an on-demand routing protocol, which means that routes between nodes are constructed only when they are required by the nodes. These routes are maintained as long as required. To ensure the validity of routes, AODV uses sequence numbers, much in the same way as DSDV [14]. Additionally, it is a cycle-free, autonomous protocol, which can scale to a high number of nodes [15].

Type	J	R	G	D	U	Reserved	Hop Count
RREQ ID							
Destination IP Address							
Destination Sequence Number							
Originator IP Address							
Originator Sequence Number							

**Fig.1** Format of an RREQ message.

As in all on-demand protocols, routes in AODV are discovered by a request/respond process. Each time a route is required between two nodes, the requesting node sends a route-request (RREQ) message through the network. Nodes receiving this RREQ update their information about the requesting node and set pointers in their routing

tables towards the node from which they received the first copy of RREQ; this reverse route is set to send back a route-reply (RREP) message, from the destination node (or from an intermediary node that knows a route to the destination). The format of RREQ messages is shown in Figure 1. For more detailed information refer to RFC 3561[13].

Type	R	A	Reserved	Prefix Sz	Hop Count
Destination IP address					
Destination Sequence Number					
Originator IP address					
LifeTime					

**Fig.2** Format of an RREP message.

A node receiving a RREQ message can transmit (by unicast) a route reply message (RREP) to the requesting node if: it is the destination node, or if it has a valid route to the destination node (the sequence number is greater or equal to the one contained in the RREQ message). Otherwise, the node retransmits the RREQ message. All nodes keep track of the RREQs they have processed. If the same RREQ is received again, it is then discarded. RREP messages have the format shown in Figure 2. For more detailed information refer to RFC 3561 [13]. As an RREP is propagated back to the requesting node, all nodes along the way set pointers to the destination node (a forward route is set to send packets from the requesting to the destination node). In the route discovery process, pointers that are set but are not in the forward route, are eliminated. Once the requesting node receives the RREP, it can send packets to the destination node using the forward route set by this message. Later, if the requesting node receives an RREP with a greater sequence number or an RREP with the same sequence number but with a lower hop count, it updates its routing information to the destination node and starts using that route.

### 3. Service Discovery with AODV.

To achieve service discovery, we take and extend the ideas proposed by Charles Perkins and Rajeev Koodli[11]. To take advantage of AODV's route discovery process, summarily described above, to also perform service discovery. It is necessary to extend the formats of the RREQ and RREP messages; new fields should be defined, but also actions that nodes along the network should perform when receiving these extended messages. With this proposal, resources are efficiently used, as an already existing process (route discovery) is used to perform an additional one (service discovery). The proposed solution is as follows: whenever a node requires a service, it should lookup its services table, which contains information about services it provides as well as those provided by other nodes. The information about others' services is acquired when the node participates in a service discovery process, while the information about the services it offers is set when the node is initialized (or when a new service is initialized within the node). Figure 3 shows the structure of the services table. Each row in this table contains the service identifier (a string that uniquely identifies the service), its IP address, a lifetime, and a list of attributes according to the service. Given the great

variety of existing schemas for naming services [5][7][9], we do not assume the use of any particular schema. A lifetime is used to keep information up to date, which is mandatory in ad-hoc networks where there are frequent changes in topology.

Service ID	IP Address	LifeTime	Attributes List
------------	------------	----------	-----------------

**Fig.3** Service table.

If a node meets the following conditions: it has in its services table the information about the service it requires, it has a valid route to the service provider, and the service lifetime is still valid, then it can contact the service provider to use the service (the way to invoke the service is out of the scope of this article); otherwise, the nodes initiates a service discovery process, sending an RREQ message with extensions, known as SREQ. Recall that these extensions are fields added to a regular RREQ, so it is possible to locate the service (IP address of the service provider), as well as a route to it, simultaneously.

Type	Length	Service ID Length
Service ID		
Attributes List		

**Fig.4** Format of extensions to RREQ messages for service discovery.

In Figure 4 we show the format of extensions to RREQ messages, which include a Type field to identify a SREQ message (it has a value of 20), the Length of the extension, the Service ID Length, a Service Identifier and lastly an Attributes List with optional value, used to include restrictions in the service. When a node receives an SREQ it executes the following actions. First, it should determine if it has in its services table a valid association (service name, IP address) for the required service; that is, if it provides the service or if it knows a valid route to a node that can provide it. If any of these criteria are met, the node should issue an RREP message with extensions for service discovery, also known as SREP, and send it to the requesting node. Of course, both SREQ and SREP messages are transmitted through the network according to the rules of the AODV protocol, so the route to the provider of the service will be found, as well as the reverse route. The format of the extension to RREP messages for service discovery is shown in Figure 5. The fields included in the extensions are: a Type to identify SREP messages (with a value of 21), the Length of the extensions, a service Lifetime, the Length of the service URL, and finally the URL itself. If a node contains an association (service name, IP address) to a service, but does not have a valid route to it, then it assigns this IP address as the address to resolve issuing an SREQ.

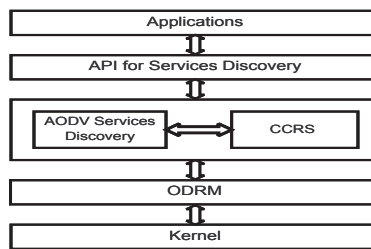
Type	Length	LifeTime
URL Length	URL	

**Fig.5** Format of extensions to RREP messages for service discovery.

Any node receiving an SREQ with a valid destination address should send an SREP if it has a route to the destination node, or if it knows an equivalent route to the service required. Otherwise, if it does not have information about the service required, or a route to the destination node, it will only forward the SREQ message. If a node receives an SREP and has information about the service required, it should compare its lifetime with the one contained in its services table; if the information in the table is more updated, then it should discard the message received and issue an SREP with that information. Otherwise, it will simply send the message to its destination.

#### 4. Service Discovery Architecture.

The on-demand service discovery mechanisms that we propose are available to application programmers through an API that can enable applications to request, register, and deregister services. The overall architecture is shown in Figure 6.



**Fig.6** Service Discovery Architecture.

At the top are applications that require the use of services, so they can call functions in the API for service discovery (API-SD); this will allow access to mechanisms and information found in AODV with extensions for service discovery (AODV-SD) and in the Common Cache Registry Server (CCRS). AODV-SD is responsible for locating services and routes to them, as described in the previous section, while the CCRS manages the services table. The On Demand Routing Module (ODRM) is located at the bottom, and handles the interaction between AODV-SD and the operating system's kernel. This architecture is based on MobileMAN [3], a framework for protocol development and test for mobile ad-hoc networks, proposed by Helsinki University of Technology (HUT). We are currently testing this architecture, which uses Linux as the development platform. The next step is to build applications that take advantage of this platform, in order to evaluate their benefits and performance.

#### 5. Conclusions and future work.

Given their wide range of application, mobile ad-hoc networks will be increasingly used in different environments. Service discovery is an important feature in this type of



networks. We have presented a proposal for service discovery in ad-hoc networks, which does not incur additional traffic, as it takes advantage of the already present process of route discovery to extend it and also discover services. The current implementation of our proposal still needs to be tested in order to evaluate its performance and allow us to experiment with variations of the mechanisms proposed.

## References

- [1] Liang Cheng. Service advertisement and discovery in mobile ad hoc networks. In Workshop on Ad hoc Communications and Collaboration in Ubiquitous Computing Environments, New Orleans, Louisiana, USA, 2002.
- [2] Scott Corson and Joseph Macker. Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations. RFC 2501 IETF 1999.
- [3] Jose Costa-Requena. Mobileman ad hoc framework. In Technical Report, Helsinki University of Technology, 2004.
- [4] Diego Doval and Donald O' Mahony. Nom: Resource location and discovery for ad hoc mobile networks. Proceedings of the first annual Mediterranean ad hoc networking workshop, Med-hoc-Net, Italy 2002.
- [5] UPnP Forum. UPnP ServiceType:V Service Template Version 1.01. 2004.
- [6] UPnP Forum. UPnP Device Architecture v1.0.1 Draft, 2004.
- [7] Arnt Gulbrandsen, Paul Vixie, and Levon Esibov. A DNS RR for specifying the location of services (DNS SRV). RFC 2782 IETF 2000.
- [8] Erik Guttman and Charles Perkins et al. Service Location Protocol, Version 2. RFC 2608 IETF 1999.
- [9] Erik Guttman, Charles Perkins, and James Kempf. Service Templates and Service Schemes. RFC 2609 IETF 1999.
- [10] Sumi Helal, Nitin Desai, Varun Verma, and Choonhwa Lee. Konark – a service discovery and delivery protocol for ad-hoc networks. In Proceedings of the Third IEEE Conference on Wireless Communication Networks (WCNC), New Orleans, Louisiana, USA, 2003.
- [11] Rajeev Koodli and Charles E. Perkins. Service Discovery in On-demand Ad Hoc Networks. Internet-draft IETF. 2002.
- [12] Sun Microsystems. Jini Specifications v2.0, 2004.
- [13] Charles Perkins, Elizabeth M. Royer, and Samir Das. Ad Hoc On Demand Distance Vector (AODV) Routing. RFC 3561 IETF 2003.
- [14] Charles E. Perkins and Pravin Bhagwat. Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In ACM SIGCOMM'94, Computer Communications Review 24, 1994.
- [15] Charles E. Perkins and Elizabeth M. Royer. Ad hoc on-demand distance vector routing. In Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications New Orleans, Louisiana, USA, 1999.
- [16] Mark Weiser. The computer for the twenty-first century. In Scientific American, 1991.

## Lucrn: A middleware for application development in spontaneous networks of Bluetooth devices

J. Alfonso Briones-García<sup>‡</sup>, Silvana Bravo-Hernández,  
J. Oscar Olmedo-Aguirre

Departamento de Aplicación de Microcomputadoras<sup>‡</sup>,  
Instituto de Ciencias, Universidad Autónoma de Puebla,  
e-mail: jabriones@mac.com, 49 Poniente 1102. Col Reforma Agua azul,  
Apdo. Postal 461, CP 72000, Puebla, Puebla, México.

Departamento de Ingeniería Eléctrica, Sección Computación,  
Centro de Investigación y de Estudios Avanzados del IPN, CINVESTAV.  
e-mail: sbravo@computacion.cs.cinvestav.mx, oolmedo@delta.cs.cinvestav.mx  
Av. IPN 2508, CP 07300, DF, México.

**Abstract:** A spontaneous network is a wireless network that operates without the support of any fixed infrastructure and where the mobile nodes --laptops, pda, cell phones-- configure themselves to form the network. The main problem in this kind of network is that applications do not have mechanisms to support: the casual creation of the network, the nodes mobility, and the exchange of information between them. In this paper we present the middleware Lucrn. The middleware comprises a suite of protocols that extends the Bluetooth stack, providing to applications the means to support the dynamic environment imposed by the spontaneous network. The middleware has been implemented in Mac OS X and Palm OS, and has been used to create several applications, one of them is a shared whiteboard that works between Bluetooth enabled laptops and personal digital assistants.

**Resumen:** Las redes espontáneas son redes inalámbricas que sin tener que depender de una infraestructura física o de servicios preestablecida pueden formarse entre dispositivos móviles como computadoras portátiles, teléfonos celulares y asistentes digitales. El principal problema que se presenta en estas redes es que las aplicaciones no cuentan con los mecanismos necesarios para tratar con: la formación casual de la red, la movilidad de los nodos, y el intercambio de información entre ellos. En este artículo se explica el *middleware* Lucrn el cual está compuesto por un conjunto de protocolos que extienden la pila de protocolos de Bluetooth y proporcionan a las aplicaciones los mecanismos necesarios para soportar el ambiente de red dinámico impuesto por las redes espontáneas. El *middleware* ha sido implementado en las plataformas Mac OS X y Palm OS, y ha sido usado para el desarrollo de diferentes aplicaciones, una de las cuales es un pizarrón electrónico compartido el cual funciona entre computadoras portátiles y asistentes digitales equipados con tecnología inalámbrica Bluetooth.

**Palabras Clave:** Middleware, Bluetooth, Spontaneous Networks, Ad-hoc networks.

## 1. Introduction

In recent years, due to technological advances in VLSI, microelectronics and telecommunications it has been possible to build mobile devices equipped with wireless technology. This convergence between wireless communications and mobile devices has created a new area in computer science known as *mobile computing*. This area try to solve the problems that rise when the communication devices are mobile, resource limited and the communication technology is wireless [1].

In this area we found a special kind of network called *ad-hoc network*. This network consists of a collection of wireless nodes that cooperatively form a network that operates with no support of any preplanned (wired) infrastructure or (DHCP) services. An important subclass of ad-hoc networks is *spontaneous networks* that has well defined properties [2]:

- These networks are formed in a casual way.
- These networks have a limited extent in both space and time.
- The nodes of the network are computer laptops, PDAs, and cell phones.
- The communication technology includes Bluetooth, IEEE 802.11 and infrared.

Spontaneous networks have been proposed for scenarios where a group of people come together and use wireless computing devices for some computer-based collaborative application. For example, applications like a shared whiteboard could be very useful for meetings and educational scenarios [2]. Nowadays a wireless communication standard that has the mechanisms needed to create spontaneous networks is Bluetooth [3]. However, despite that this wireless communication technology has proliferated, applications like shared whiteboards are difficult to find in spontaneous networks of mobile devices. This problem is due to the lack of mechanisms that applications need to interoperate in a mobile environment.

The article is organized as follows. In the next section we present the problems that application developers have to face in a spontaneous network environment. Then, we discuss the middleware and the suite of protocols that comprises the middleware. Next, we explain the middleware implemented in Mac OS X and Palm OS. Finally, we present our results and some concluding remarks.

## 2. Spontaneous networks challenges

Applications running on a spontaneous network are characterized by problems that are not present in, for example, local area networks based on IEEE 802.3. We have classified these problems in three groups:

- **Problems generated by the spontaneous network formation.** In a spontaneous network, we cannot assume that a computer provides a service like DHCP for automatic network interface configuration, because it cannot be

determined before hand neither the nodes belonging to the network nor the services they offer. To cope with this problem we use some of the mechanisms that Bluetooth technology offers supplemented with some protocols that announce the addition of a Bluetooth device into the spontaneous network.

- **Problems generated by the node mobility.** Node mobility conveys some problems to applications because the limits of the network are loosely defined and it is possible that the network could be fragmented. Therefore, application developers cannot assume that some services are always available and therefore they need mechanisms to know when a node has entered or left the spontaneous network.
- **Problems generated by the communication technology.** Communication technology very often has characteristics that can limit the performance of spontaneous network applications. For example, Bluetooth technology only offers the creation of both point-to-point links and point-to-multipoint links. These kinds of links imply that is not possible to exchange information among all the nodes of the spontaneous network. Consequently some mechanisms are needed for exchange of information among the nodes of the network.

In summary, the main problem is the lack of a middleware that could be implemented in mobile devices to support interoperation among applications running on spontaneous networks. In the next section, we present the design of *Lucrn*, a middleware that comprises a suite of protocols that extend the Bluetooth stack and provides to mobile device applications the means to interoperate on a spontaneous network.

### 3. The middleware Lucrn

In a high level context we identify Lucrn as a software layer that sits between the applications and the operating system. The middleware has been designed and built upon the mechanisms offered by the Bluetooth technology. Essentially, Bluetooth is a suite of protocols designed to enable wireless communication on devices like portable digital assistants, laptop computers and cell phones, among others. Actually, Bluetooth is a public specification [4] [5] that is the base of the standard IEEE 802.15.1[6].

Bluetooth introduces novel mechanisms like *Device discovery* [9] and *Service discovery* [10]. These two mechanisms are used by the middleware to establish in an ad-hoc manner the wireless links between devices.

The middleware comprises four protocols that we explain in the following sections.

### 3.1 Device Announcement Protocol

This protocol suppose that a L2CAP (logical link and adaptation protocol) channel [10] has been established and once the channel is created it sends special messages in order to:

- Announce the creation of a new L2CAP channel in the network.
- Synchronize the Bluetooth active devices with the new node.

To perform these actions it sends the messages: **LUCRN\_NEW\_DEVICE** and **LUCRN\_SYNC\_DEVICES**.

### 3.2 Routing Protocol

Bluetooth allows the creation of point-to-point and point-to-multipoint communication links that is not a broadcast medium like IEEE 802.11 or IEEE 802.3. In Bluetooth, the communication is bidirectional between the master and the slave that share the channel. The same happens in point to multipoint links the communication is between the slaves and the master but could not be between the slaves.

Taking this into account we create a routing protocol that allows the exchange of information between any nodes that belong to the network. The protocol works along with the *device announcement protocol* and the *mobility protocol* in order to update the routing tables that each device holds.

The valid message for this protocol is **LUCRN\_DATA\_MSG**.

### 3.3 Mobility Protocol

The goal of the mobility protocol is to announce that a node has left the network. Each time that a node receives this message it must update its routing table. This protocol only works for single nodes that only have one adjacent link. This is important because if the device has more than one adjacent link it means that is the master of a piconet.

The valid message for this protocol is **LUCRN\_DISCONNECTING\_DEVICE**.

### 3.4 Collaboration Protocol

This is an high level protocol designed to provide mechanisms that facilitate collaboration among applications running on spontaneous networks. The protocol was designed to allow the exchange of actions that control applications. For example, if a user has drawn or erased a figure this action will be forwarded to all the members of the spontaneous network.

This message is encapsulated in the payload area of **LUCRN\_DATA\_MSG**. The message valid for this protocol is **LUCRN\_COLL\_ACTION**.

Once the protocols that conform the middleware were designed, they were implemented on Mac OS X and Palm OS platforms. In the next section we explain the main components of these implementations.

## **4. Middleware implementations**

The middleware has been implemented in Mac OS X and Palm OS platforms.

### **4.1 Mac OS X implementation**

The Mac OS X implementation was developed using the object-oriented language Objective-C. The implementation comprises a set of classes that handles publishing of services, the communications channels and the exchange of information among the nodes. The main class is Lucrn that provides methods to discover and to connect to services, to publish services, to send unicast and broadcast messages and to query the devices currently available in the network. From a developer point of view, the implementation is an object-oriented framework that can be easily integrated in programming projects to develop applications for spontaneous network.

### **4.2 Palm OS implementation**

In Palm OS, the implementation was done using the C programming language, an important issue because of the minimal size required for the executable program to meet the severe memory constraints. The implementation consists of a library that allows to the programmer to discover and connect to a service, to send unicast and broadcast messages, and to query about the devices present in the network.

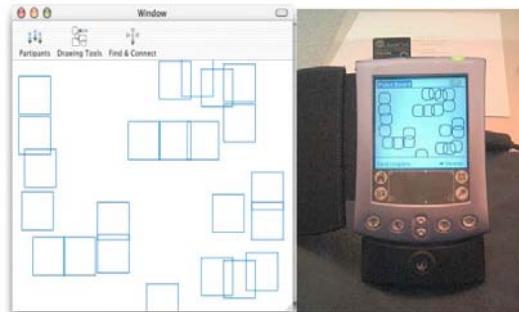
The exchange of information between different layers of the implementation is done through function pointers, which are assigned and implemented by the modules that form the application.

## **5. Results**

The results obtained so far are very encouraging. Lucrn is a middleware composed of a set of protocols used to develop applications to be deployed in spontaneous networks of heterogeneous devices. With this middleware we have developed several applications exploiting the wireless capabilities of each device.

By using the middleware we have developed the following applications:

- *A system to exchange messages between heterogeneous devices.*
- *A remote control for the Mac OS X iTunes and Keynote applications.*
- *A shared whiteboard.* (Figure 1)



**Fig. 1.** A shared whiteboard between laptops and PDAs

## 6. Conclusions

In this paper, we have described Lucrn, a middleware that extends the Bluetooth protocol stack to offer enhanced services to applications deployed on mobile spontaneous networks. The enhanced functionality provided by Lucrn greatly simplifies the formation of networks that do not require any pre-planned infrastructure nor permanent user intervention for device reconfiguration. The applications developed so far have shown that Lucrn can be used for a number of application domains like distributed information systems, computer supported collaborative work and pervasive computing. The results obtained are encouraging and a number of directions for further research are being discussed and explored.

## References

- [1] M. Satyanarayanan, "Fundamental challenges in mobile computing", Proc. 15th ACM Symp. Principles of Dist. Comp, May 1996
- [2] L.Feeney and B. Ahlgren and A. WesterLud, "Spontaneous networking: An Application Oriented Approach to ad-hoc networking", IEEE Comm. Magazine, June, 2001.
- [3] R.Bruno and M.Conti and E. Gregori, "WLAN technologies for Mobile Ad-Hoc Networks", Proceedings of the 34th Hawaii International Conference on Systems Sciences, 2001.
- [4] Bluetooth SIG, "Specification of the Bluetooth System. Core", Version 1.1, February 2001, [http://www.bluetooth.com/pdf/Bluetooth\\_11\\_Specifications\\_Book.pdf](http://www.bluetooth.com/pdf/Bluetooth_11_Specifications_Book.pdf)
- [5] Bluetooth SIG, "Specification of the Bluetooth System. Profiles", Version 1.1, February 2001, [http://www.bluetooth.com/pdf/Bluetooth\\_11\\_Profile\\_Book.pdf](http://www.bluetooth.com/pdf/Bluetooth_11_Profile_Book.pdf)
- [6] Chatschik Bisdikian, "An overview of the Bluetooth Wireless Technology", IEEE Comm. Magazine, December 2001
- [7] Rajeev Shorey and Brent A. Miller, "The Bluetooth technology: Merits and Limits", IEEE Proceedings of ICPWC'2000, 2000
- [8] BlueZ, "Official Linux Bluetooth protocol stack", <http://www.bluez.org/>
- [9] T. Salonidies et al, "Proximity awareness and fast connection establishment in Bluetooth", Mobile and Ad Hoc Networking and Computing, 2000. MobiHOC. 2000 First Annual Workshop on , 2000 , Page(s): 141 -14
- [10] Brent A. Miller and Chatschik Bisdikian, "Bluetooth Revealed. The insider's guide to an open specification for Global Wireless Communication", Prentice Hall PTR, Upper Saddle River, NJ 07458, 2001

# Arquitectura de Colaboración Mediante Dispositivos Móviles Aplicada a la Administración del Conocimiento

Chadwick Carreto Arellano<sup>1</sup>, Rolando Menchaca García<sup>2</sup>, Rolando Menchaca Mendez<sup>2</sup>, Agustín Gutiérrez Tornes<sup>2</sup>

<sup>1</sup>Escuela Superior de Computo, <sup>2</sup>Centro de Investigación en Computación  
Instituto Politécnico Nacional

e-mail : <sup>1</sup>[ccarretoa@ipn.mx](mailto:ccarretoa@ipn.mx), <sup>2</sup>[fmenchac@ipn.mx](mailto:fmenchac@ipn.mx), <sup>2</sup>[rmen@ipn.mx](mailto:rmen@ipn.mx)

Av. Juan de Dios Bátiz s/n esq. Miguel Othón de Mendizábal,  
Unidad Profesional Adolfo López Mateos, C.P. 07738, México D.F.

**Resumen:** En el presente trabajo se muestra el desarrollo de una arquitectura que permita trabajar en un ambiente colaborativo desde dispositivos móviles, apoyada en el uso de agentes de software, para brindar una funcionalidad orientada a la administración del conocimiento aplicable a cualquier proceso de enseñanza-aprendizaje. La arquitectura permitirá administrar el conocimiento, de tal forma que automatice el proceso de recopilación de información, realice búsquedas de la información de acuerdo a un perfil y necesidades específicas, y la presente, ya sea en un formato para dispositivo móvil o en un formato tradicional para PC. La arquitectura pretende poner a disposición de los usuarios, herramientas para el trabajo en grupo que pueden ser utilizadas en cualquier momento y lugar a través de un dispositivo de capacidad limitada con conexión a Internet.

**Palabras Clave:** Trabajo Colaborativo, Dispositivo Móvil, Agente de Software, Administración de Conocimiento, e- learning.

## 1. Introducción.

Actualmente el aprendizaje y el conocimiento son parte esencial en toda organización y comunidad; y al hablar de comunidad es necesario que el conocimiento se distribuya y se aplique, de otra forma es inútil. La clave de la evolución social radica en la generación, difusión y distribución del conocimiento.

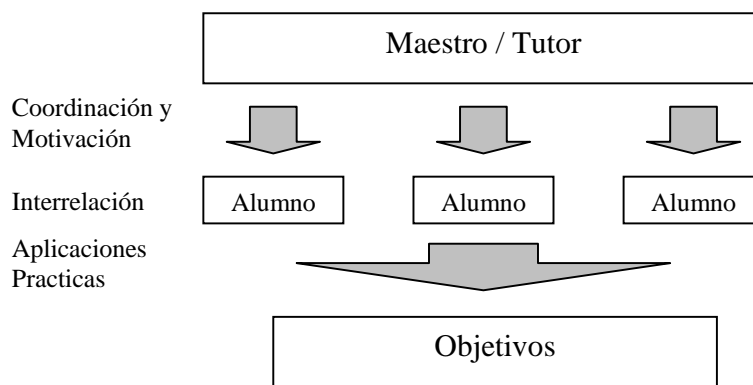
A través del tiempo esta distribución de información (comunicación) se ha dado por diferentes medios. En los últimos años, la comunicación inalámbrica (wireless) está experimentando un crecimiento acelerado. Las nuevas tecnologías permiten aumentar la calidad no sólo en el sector productivo, sino también en el sector educativo que ha sufrido un cambio drástico desde la aparición de la computación y últimamente la Internet; esto ha permitido la creación de innovadoras técnicas de enseñanza y la aparición de material de apoyo y referencias de fácil acceso.

Las tecnologías inalámbricas permiten a mayor número de personas comunicarse e intercambiar información sin la necesidad de un aula de cómputo o una computadora; esto permite introducir el concepto de colaboración digital, que complementa a la presencial.



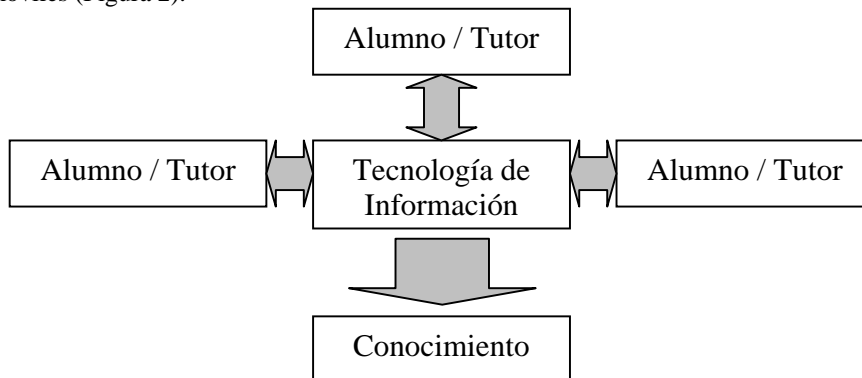
El aprendizaje es un campo en el que bien se podría aplicar la tecnología inalámbrica, ya que permite establecer un ambiente de colaboración que no se reduce a un aula o a estar sentado frente a una computadora; incluso el salón se podría convertir en laboratorio o la cafetería en biblioteca.

Esta tecnología proporciona cambios fundamentales en la forma en que se difunden las ideas, el conocimiento y la comprensión [1]. La colaboración entre alumnos puede generar grandes ventajas educativas si se manejan adecuadamente, a través de un esquema de aprendizaje colaborativo apropiado para la interacción con tecnología móvil. El esquema general de aprendizaje colaborativo es (Figura 1):



**Figura 1.** Esquema de aprendizaje tradicional

En este esquema claramente se observa la necesidad de que las entidades se interrelacionen de manera directa en un espacio cerrado. Se propone entonces una reestructuración del esquema anterior para omitir esta necesidad aplicando tecnologías móviles (Figura 2).



**Figura 2.** Nuevo esquema de aprendizaje

La comunicación inalámbrica (en especial en México) está en sus primeros pasos, mientras que la educación a través de estos dispositivos (móviles o inalámbricos) está en su etapa embrionaria a nivel mundial. Hasta ahora los conceptos *e-learning* y

*colaboración digital* no se han generalizado para dispositivos móviles inalámbricos, pero es deseable dado que la principal ventaja de su aplicación en la educación es la transparencia; es decir, el acceso *anytime, anywhere* (a cualquier hora y en cualquier lugar) a una fuente de conocimiento [2].

La razón de que se dificulta esta interacción a través de tecnologías inalámbricas es precisamente sus limitantes: ancho de banda, presentación de la información, capacidad de procesamiento y memoria, entre otras. Algunas empresas proponen soluciones como la creación de un protocolo (WAP) que permite incorporar estándares internacionales como HTML y JavaScript; las compañías de telefonía celular han aportado soluciones como la creación de redes con mayores servicios y dispositivos con mejores capacidades. Sin embargo, sigue latente la problemática de adaptar aplicaciones que sean útiles y aceptadas por los usuarios.

Existen algunos trabajos [3] [4] [5] que pretenden resolver, no propiamente la problemática que planteamos, pero sí alguno de sus aspectos, ya sea promover la administración de información o conocimiento, manejar grupos de trabajo colaborativo, o trabajar alguno de los aspectos con dispositivos móviles.

A continuación, en la sección 2 se describe la Arquitectura Propuesta, así como las características de los elementos que la componen. La sección 3 explica el diseño y la forma de implementación técnica del modelo. Finalmente, en la sección 4 da una conclusión de trabajo expuesto y establece el trabajo a futuro.

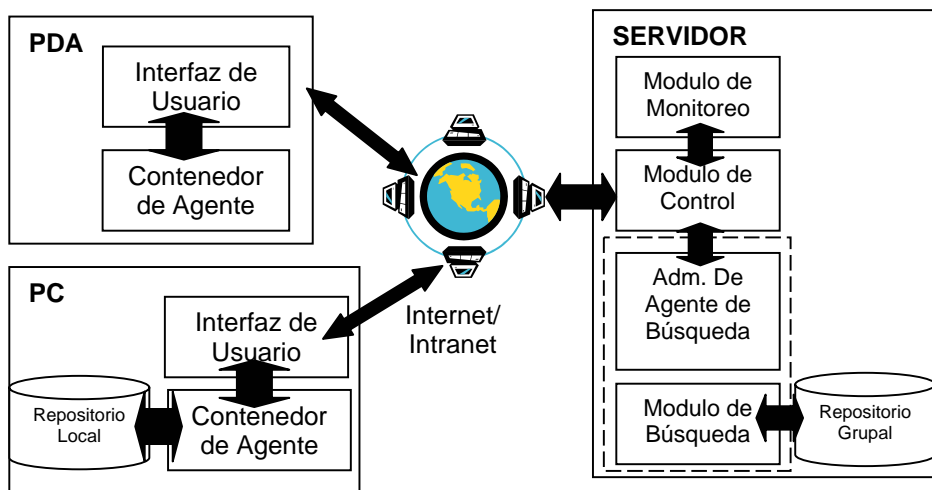
## **2. Arquitectura del Sistema de Colaboración.**

Se busca que la arquitectura proporcione una funcionalidad para los usuarios de búsqueda de información, recuperación de documentos, publicación de documentos y envío de mensajes dentro de un grupo de trabajo. Lo cual le permitirá establecer un esquema de cooperación que apoye el desarrollo de su aprendizaje a través de la adquisición de conocimiento de manera autodirigida, responsable y consciente. Este proyecto utiliza las tecnologías inalámbricas en una aplicación que permita la administración del conocimiento, entendiéndose por ello el proceso mediante el cual se alcanza el objetivo (planeación, organización, dirección, control) de manejar la información (documentos, publicaciones, comentarios, mensajes) construida a partir del conocimiento de los usuarios previendo sus necesidades futuras de aprendizaje. Como se observa en la figura 3, la arquitectura del sistema está dividida en cinco módulos, de los cuales dos forman el servidor del sistema, dos módulos en los clientes y el módulo correspondiente a la base de datos se subdivide para crear un repositorio en cliente PC y en el servidor, con algunas diferencias entre ambos. A continuación se describe el funcionamiento de los módulos que conforman la arquitectura:

### **2.1 Módulo de Administración**

En este módulo es donde llegan las peticiones de los miembros de un grupo de trabajo y se encarga de atenderlas, por esta razón tiene acceso a Internet. Aquí también se reconoce desde que tipo de dispositivo se conecta el miembro del grupo de trabajo ya

sea desde una PC o una PDA, además de esta comunicación, existe otra con el módulo de recolección y análisis de información, la cual se utiliza para realizar, de manera indirecta, consultas a la base de datos del sistema y búsquedas en Internet. La administración de perfiles y autenticación de miembros se realizará en este módulo haciendo peticiones de consulta o modificaciones a la Base de Datos del grupo donde están almacenados los perfiles de los miembros del grupo.



**Fig.3** Arquitectura del Sistema de Colaboración.

## 2.2 Módulo de Búsqueda

Encargado de hacer las consultas a la Base de Datos del grupo, ya que es el único módulo que tiene conexión directa a la Base de Datos, y regresar los datos al módulo de administración. También tiene la función de darle un formato estándar a la información (XML) para ser almacenada y que pueda ser regresada en diferentes formatos (HTML ó WML); el módulo no cuenta con acceso a Internet por lo que necesita establecer comunicación con el modulo de búsqueda en Internet para hacer búsquedas de este tipo, estos documentos provenientes de Internet son analizados por el módulo de administración basándose en el perfil del usuario. Este modulo también tiene la función de realizar búsquedas en Internet y regresarlas al modulo de administración.

## 2.3 Módulo del Agente de Usuario

Se encuentra en ambos tipos de clientes (PDA y PC). Debido a que tiene comunicación con el servidor, vía Internet, se encarga de hacerle llegar las peticiones del usuario, además es donde se almacena temporalmente (lo que dure la sesión) un identificador de usuario.

## **2.4 Módulo de Interfaz de Usuario**

Contiene las herramientas disponibles para el usuario: búsqueda de información, recuperación de documentos, publicación de documentos y envío de mensajes. Se comunica con el módulo del Agente del usuario para llevar a cabo estas funciones.

## **2.5 Modulo de Base de Datos y Conocimiento**

Esté modulo se divide en dos partes, una que trabaja en el servidor y la otra que opera sobre los clientes del sistema. La que está presente en el servidor contiene la información y perfiles de los miembros y los documentos del grupo y la que se encuentra en el cliente de tipo PC y contiene los documentos recientes de un miembro en particular para que puedan ser examinados sin estar en línea.

Una vez definida la Arquitectura en la siguiente sección se presentara su diseño y la implementación de la misma.

## **3. Diseño e Implantación de la Arquitectura**

El diseño de la Arquitectura se basa en la Arquitectura del sistema y la implementación sobre una plataforma de Agentes Móviles que trabajan para administrar la información y documentos que se guardan, consultan y generan en el sistema.

La implantación es un problema sustancial que hasta ahora se ha resuelto con la utilización de una infraestructura de agentes que permitan identificar potenciales colaboradores y realizar tareas de búsqueda de manera autónoma. Se define un ambiente colaborativo que maneje dos tipos de agentes inteligentes, los agentes de usuario y los agentes mediadores.

Estos agentes trabajan sobre una plataforma de movilidad permitiendo la comunicación con múltiples usuarios y el manejo de múltiples documentos e información. La habilidad de interactuar, esto es, de intercambiar información y conocimiento de los agentes es lo que nos lleva a su aplicación en entornos colaborativos.

La implementación muestra el desarrollo y codificación de la arquitectura planteada con base en su diseño y modelado, dándole una funcionalidad a través de herramientas para el trabajo en grupo. La arquitectura se desarrollo completamente en Java trabajando con la funcionalidad de Jade para la creación de lo agentes y con la API de Google para permitirnos realizar búsquedas mas especializadas sobre la información solicitada y requerida de acuerdo al perfil de nuestro usuario.

La Arquitectura se ha implementado para su prueba en un entorno real dentro de la Escuela Superior de Computación del Instituto Politécnico Nacional, donde ha permitido (en una fase experimental) la comunicación e intercambio de información entre una comunidad de alumnos y maestros lo cual permite asegurar la aplicabilidad de la arquitectura así como su flexibilidad para adaptarse a distintos tipos de intercambio de información sobre diferentes grupos de colaboración.

## 4. Conclusiones y Trabajo a Futuro

Esta arquitectura resulta de un proyecto de investigación que engloba dos aspectos importantes de las tecnologías de la información; por un lado los ambientes colaborativos de trabajo y por el otro, las tecnologías de cómputo Móvil. Los primeros son útiles debido a que permiten a los usuarios de un sistema o aplicación tener conciencia de que aún estando en su propia estación de trabajo forman parte de un grupo y las actividades se pueden realizar conjuntamente. En un ambiente colaborativo, los usuarios pueden estar en constante interacción lo que brinda retroalimentación, fomenta discusiones y agiliza el cumplimiento de un mismo fin. Las tecnologías de cómputo móvil, por su parte, nos brindan la capacidad de trasladar el ambiente a cualquier lugar y en cualquier momento lo que desaparece la limitante de encontrarse frente a una computadora para poder ingresar al sistema. El tipo de arquitectura desarrollada puede ser útil a varios propósitos. En esta ocasión se demostró su funcionalidad a través de una aplicación orientada al aprendizaje colaborativo, sin embargo cambiando los módulos de base de datos y de interfaces se puede cambiar el tipo de aplicación. Se pensó en una aplicación orientada al conocimiento debido a la convicción de que el crecimiento social, cultural, económico, está fuertemente asociado con el nivel de conocimiento adquirido y aplicado. Asimismo, el presente trabajo se enfoca a la adquisición de conocimiento o aprendizaje en un nivel de estudios superiores o aprendizaje en adultos, ya que el aprendizaje colaborativo se adapta a las necesidades de este sector. Así, una conclusión importante a la que se ha llegado es que la arquitectura es totalmente flexible y aplicable a múltiples áreas y diferentes necesidades de usuarios. La aplicabilidad, entonces, podrá cambiar de manera sencilla. Como trabajo a Futuro es importante resaltar que esta arquitectura por su flexibilidad modular puede implementarse sobre diversos campos de investigación y educación pero no se descarta que también pueda emplearse en áreas de distribución de información a grupos.

## Referencias

- [1] T Román Julián, Rebeca, Domínguez López Alejandro (2002). "Integración Efectiva de los Sistemas de Información, un modelo de apoyo". CNCIIC Octubre 2002. Guadalajara, México.
- [2] Ayala Gerardo, Intelligent Agents for a Lifelong Learning Environment in Information Technologies, International Journal of Continuing Engineering Education and Life-Long Learning, 2002.
- [3] Kutay Cat, Interact Software Architecture, University of New South Wales, Australia 2003.
- [4] Peredo Rubén, Et. al., Arquitectura Computacional para Sistemas de Aprendizaje Colaborativo, Instituto Politécnico Nacional-Centro de Investigación en Computación, México.
- [5] Rebollo Miguel, Et. al., CALIOPE: Una Arquitectura para Aprendizaje Autónomo Colaborativo en Entornos no Presenciales, Universidad Politécnica de Valencia, 2000.

# Mobile Access to Digital Libraries via Generic Interfaces

Nohema Castellanos, J. Alfredo Sánchez

Universidad de las Américas-Puebla,  
e-mail : {sp089077, alfredo}@mail.udlap.mx  
Cholula, Puebla, C.P. 72820, Mexico

**Abstract:** We have designed an architecture aimed at facilitating access to digital library (DL) resources from mobile devices. Our architecture includes the notion of a unique specification used to dynamically build interfaces for an appropriate deployment on each of various target mobile devices in an automated fashion. We describe our architecture and report on the results being obtained for its first instantiation.

**Resumen:** Se describe una arquitectura que facilita el acceso a recursos de bibliotecas digitales (BDs) desde dispositivos móviles. La arquitectura incluye la noción de una especificación única que se usa para construir interfaces dinámicamente y se despliegan adecuadamente en cada uno de los diversos dispositivos móviles objetivo de manera automática. Este documento describe la arquitectura propuesta y reporta los resultados que se han obtenido con una primera instancia de este concepto.

**Keywords:** Digital libraries, generic interfaces, XML, mobile devices, OAI.

## 1. Introduction

Digital libraries (DLs) are virtual spaces that comprise huge digital repositories and a wide range of services, user environments and interfaces, intended to support learning and collaborative research activities. In order to assist DL users in dealing with such a complex system, we posit that they should be able to create virtual spaces from which they can access DL resources, manage contents relevant to their interests and organize them according to their needs and preferences. We term these environments *personal spaces*. We have been exploring the notion of federated personal spaces in the context of federation of DLs that operate under OAI-PMH [5]. Three cooperating institutions participate currently in our project, [6] but the approach is extensible to all OAI-PMH collections.

Although we have instantiated successfully the concept of personal space for deployment with conventional workstations, we would like to make our DL resources accessible to users from a wide range of wireless and portable devices as well. However, various differences of mobile devices with respect to conventional computers make existing user interfaces difficult to use and would make it necessary to redesign user interfaces for each kind of device. Thus, we have

designed a system architecture that includes the notion of “generic” user interfaces and automates the generation of user interfaces for diverse portable devices, such as PDAs and cellular phones.

## 2. Related Work

The increasing importance of mobile devices has motivated research regarding their use for accessing digital libraries. Important work has focused on usability issues. For example, [11] and [12] present results of a study on the usefulness of handheld devices in reading. [9] outlines the need for user-centered design in mobile computing and presents the results from two preliminary studies aiming to solicit user feedback regarding mobile DL applications. [8] discusses some cultural and socio-technical concerns which arose with two small-screen demos implemented on the Greenstone Digital Library. The user interface design of a client system for Digital Video Libraries access and presentation in a PDA was discussed in [10].

In addition to usability issues, supporting mobile access to DLs involves the provision of dynamic interfaces that render appropriately on displays smaller than those of desktop. This makes it difficult to use existing user interfaces and would make it necessary to redesign user interfaces for each kind of device. To provide users with web pages that are well-suited to the limited screen and capabilities, [1] describes four major approaches: device-specific authoring; multiple-device authoring; client-side navigation; and automatic re-authoring. Device-specific authoring involves designing user interfaces for each kind of device; in multiple-device authoring, a range of target devices is identified, and mappings from a single source document to a set of rendered documents are defined to cover the devices within the range; in client-side navigation, the user is given the ability to interactively navigate a single web page by altering the portion of it that is displayed at any given time; finally, automatic re-authoring involves developing software which can take an arbitrary web document designed for the desktop, along with characteristics of the target display device, and re-author the document through a series of transformation. This transformation is referred to as *transcoding* [4].

A new approach has emerged recently which proposes XML-based markup languages as the basis for creating *generic interfaces*. A generic interface is specified in such a way that its appearance may vary from device to device while its functionality remains the same for each of the involved devices [3]. UIML (User Interface Markup Language) [2] is representative of this approach. UIML describes generic user interfaces in a device-independent manner and mapped to a specific target language. As many versions of the generic interface have to be constructed as target languages are desired. These multiple descriptions constrain the generality in the design of user interfaces [10]. We address some of these drawbacks in our approach.

### 3. Dynamic Generation of Interfaces for Digital Libraries

We have designed a conceptual architecture aimed at accessing DL resources from mobile devices. Our approach to create dynamic interfaces that deploy correctly on small screens combines the notion of generic interface and the automatic re-authoring approach, as a transcoding process is applied to a unique specification—which we term *generic interface*—to dynamically build interfaces for appropriate deployment on each of various target mobile devices. In contrast with the approach of UIML, these interface specifications should not need to be mapped to the specific target languages, as our interface specifications are device-independent. As noted in Fig. 1, the architecture consists of three main components, namely:

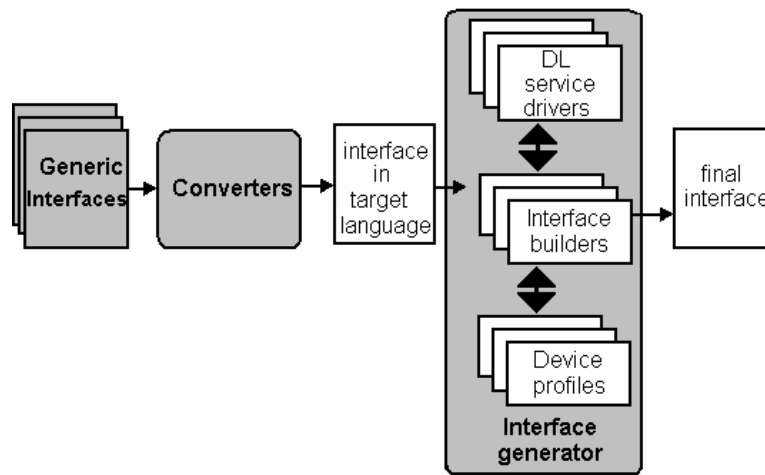


Fig. 1 Conceptual architecture for interface generation

*Generic interfaces* contain the descriptions of user interfaces. These descriptions must be written in a generic format and in a device-independent manner.

*Converters* transform the interface descriptions into the specific target languages. This component contains one converter for each target language.

The *interface generator* builds the interfaces that are ultimately presented to the user. This component generates interfaces by adapting the code produced by the converter to the corresponding DL service and the personal configuration of the users. This is what provides our generic interfaces with independence from a specific target language. We conceive the interface generator as consisting of three major sub-components: (1) *Device profiles*: Each device has its own profile, which describes its features: family (cellular phone, PDA, voice browser, etc), vendor, model, screen size, interaction modality (keyboard, touch screen, etc), languages and browser; (2) *DL service drivers*: Each module provides functionality for accessing a specific service provided of the DL; and (3) *Interface builders*: These modules adapt the interfaces produced by the converter to the device features, the corresponding DL service and the user configuration.



## 4. Prototypical Implementation

We have produced a prototypical implementation of the proposed architecture for providing mobile access to our personal space environment, hence we refer to this implementation as Portable Personal Spaces (PoPS). PoPS fosters mobile access to DLs using XML technologies and a client-server approach for accessing our DL's services. In correspondence to the proposed architecture, our prototype consists of the following major components:

*Generic personal space (GPS):* This is the set of generic interfaces for the personal space environment. Device independence for these interfaces is accomplished by using a generic XML format.

*Converter:* This component transforms the interface descriptions into code which is written in the language used by the target device. It comprises two sub-components: XSL stylesheets and an XSLT processor.

*Interface generator:* This component builds the interfaces that ultimately are presented to the user. It consists of three elements:

*Device profiles,* which are written in XML. Each profile is kept in one single XML file. At present, PoPS contemplates three devices (iPAQ Pocket PC, Palm Tungsten C and Siemens S45) and five features: family, vendor, model, language and browser.

*DL service drivers,* which allow for accessing the different services provided by the federated DLs. Access to federated collections is accomplished by relying on a distributed search platform we refer to as MAIDL. This platform implements distributed information retrieval services through mobile agents that travel around the nodes in the federation seeking information that may be relevant for user queries [7].

*Interfaces builders (IB)* adapt the interfaces produced by the converter to the devices features described in the profiles, the respective service of the federated DLs (for instance, search results) and the user configuration.

## 5. Generating interfaces for PDAs and mobile phones

We have successfully implemented the key architectural components of PoPS, which in turn have produced prototypical interfaces for personal spaces that can be displayed on PDAs and cellular phones. PoPS works as follows:

Each of the elements of the GPS interface is kept in a single XML file. The converter transforms GPS these elements into their XHTML (for PDAs) or WML (for WAP phones) equivalents. An XSLT processor transforms each XML file of the GPS into WML or XHTML according to the instructions provided by the respective XSL stylesheet. Each of the target languages requires its own XSL stylesheet. Thus, the converter has two XSL stylesheets: one for XHTML and another for WML. Therefore, in order to convert the GPS to a new language, only the appropriate XSL stylesheet of the new language needs to be designed. There is

no need to update any of the GPS specifications. Finally, the interface generator builds interfaces for personal spaces by adapting the code produced by the converter to the user preferences taking device limitations into account. A controller (implemented as a server) receives the requests from the client, detects what kind of device it is and instructs where that request should be forwarded.

## 6. Evaluation

We applied PoPS to the 23 generic interface elements that make up the personal space environments described previously. PoPS was able to generate the 46 specific interfaces (23 in WML and 23 in XHTML) using the generic specifications. In all cases the conversion was successful and the generated interfaces were rendered correctly in the corresponding mobile device. Thus, we have successfully produced interfaces for personal spaces that can be displayed on PDAs and cellular phones. This prototypical implementation makes it possible to search information, using PDAs or mobile phones, in the federation of digital collections discussed earlier. Users personalize the resulting interfaces by choosing from a list of icons (in a PDA) or labels (in a cellular phone) that represent the various services provided by the federated DLs. This is illustrated in Fig. 2.



**Fig. 2** Interfaces generated by PoPS

Also, our prototypical implementations have undergone usability tests focused on our primary target community of users. The testing environment for the PDA interfaces was an iPaq H3870 Pocket PC that accesses PoPS through our university's wireless network. For the cellular phone interfaces, the testing environment was the Siemens S45 5.0.2 browser provided by the Openwave SDK 5.1 emulator to access PoPS via the Internet. During the tests using a PDA, we observed that participants adapted quickly to the *mouseless* interaction and were able to perform the tasks they intended or were asked to do with only minimal guidance. Most of the users showed a great enthusiasm and curiosity about the use of mobile devices to access our DL.

## 7. Conclusion

At present, our research group is collaborating in a joint project to create an infrastructure for distributed digital libraries to allow mobile access to multimedia collections. Also, we are working in an online converter that extracts HTML code from arbitrary web pages and transforms them into their XHTML, WML and VoiceXML equivalents.

The contribution of this work is related not only to porting our personal space environments to wireless devices but also to defining an architecture for facilitating the generation of user interfaces that can be displayed on a variety of mobile devices. Our preliminary results show that PoPS is a promising approach to make DL resources accessible from mobile devices.

## Acknowledgments

This work has been supported by the University Corporation for Internet Development (CUDI) through its Program for Advanced Applications 2004 CONACyT (Project number 35804-A and G33009-A).

## References

- [1] Bickmore, T.W., Schilit, B.N., "Digstor: Device-independent Access to the World Wide Web", Proc. of the 6th International WWW Conference, 1997.
- [2] Harmonia, Introduction to UIML. <http://www.uiml.org>, 2000.
- [3] Mayora-Ibarra, O., "Generation of Device Independent User Interfaces", Proc. of International Workshop on Research & Development of Human Communication Technologies for Conversational Interaction and Learning, 2002, pp 1-3.
- [4] Mohan, R. J., Smith, C.S., "Adapting Multimedia Internet Content For Universal Access", IEEE Transactions on Multimedia, 1999, pp 104-114.
- [5] OAI, "Open Archives Initiative", Available at: <http://www.openarchives.org>, 2003.
- [6] Reyes-Farfán, N., Sánchez, J.A., "Personal Spaces in the Context of OAI", Proc. of the JCDL, Houston, Tex., May 2003, pp 182-183.
- [7] Sánchez, J. A., Nava, S., Fernández, L. Chevalier, G., "Distributed information retrieval from web-accessible digital libraries using mobile agents" Upgrade 3, 2, Special Issue on Information Retrieval and the Web, April 2002, pp 37-43.
- [8] Buchanan, G., Jones, M., Marsden, G., "Exploring Small Screen Digital Library Access with the Greenstone Digital Library", Proc. ECDL, Springer-Verlag, 2002, pp 583-596.
- [9] Jones, M. L. W., Rieger, Treadwell, P., Gay, G., "Live from the Stacks: User Feedback on Mobile Computers and Wireless Tools for Library Patrons". Proc. of the Fifth Conference on digital libraries, ACM, San Antonio, TX USA, 2000, pp 95-102.
- [10] Lyu, M., Yen, J., Yau, E., Sze, S., "A Wireless Handheld Multi-modal Digital Video Library Client System", MIR'03, November 7, 2003, Berkeley, USA, pp 231-238.
- [11] Marshall, C, Ruotolo, C. "Reading-in-the-Small: a study of reading on small form factor devices", Proc. of the JCDL, ACM, Portland, USA, July 2002, pp 56-64.
- [12] Marshall, C., Price, M., Golovchinsky, G., Schilit, B., "Introducing a Digital Library Reading Appliance into a Reading Group", ACM DL 1999, pp 77-84.

# **μCronos: Un micro Núcleo con Soporte para Adaptabilidad Segura de Bajo Nivel**

Víctor D. Castillo Díaz, Rolando Menchaca-Méndez

Centro de Investigación en Computación  
Instituto Politécnico Nacional

e-mail : [VictorCastilloEn@acm.org](mailto:VictorCastilloEn@acm.org), [rmen@cic.ipn.mx](mailto:rmen@cic.ipn.mx)

Unidad Profesional “Adolfo López Mateos”. D.F., México.

**Abstract:** A microkernel is composed of a minimum set of abstractions that are closely related to the underlying hardware. These abstractions can be used as an infrastructure to build over them more specialized services that meet particular necessities. This kind of architectures allows operating systems to dynamically change its structure because all these specialized services are independently implemented. This characteristic is particularly valuable when we have to meet the requirements of applications executed in mobile devices with limited resources that are deployed in dynamic environments such as those described in ubiquitous scenarios. In this paper we present some key issues in the design and implementation of operating systems for ubiquitous environments. We also purpose a microkernel designed to solve these problems. We also show the results of the tests applied to our system.

**Resumen:** Un micro núcleo está compuesto de un conjunto mínimo de abstracciones que están fuertemente ligadas al hardware subyacente. Estas abstracciones pueden ser usadas como una infraestructura sobre la cual se pueden construir servicios más especializados que cubran necesidades particulares. Este tipo de arquitecturas permiten a los sistemas operativos que las utilizan, cambiar dinámicamente su estructura. Esta característica es particularmente valiosa cuando se quiere cumplir con los requerimientos de aplicaciones que se ejecutan en dispositivos móviles con recursos limitados y que interactúan en ambientes altamente dinámicos como los descritos en los escenarios de cómputo ubicuo. En este artículo se presentan algunos de los problemas clave que se presentan en el diseño e implementación de sistemas operativos para ambientes ubicuos y se presenta un micro núcleo diseñado para solucionar estos problemas. Se presentan los resultados de pruebas comparativas de desempeño.

**Palabras Clave:** Micro Núcleo, Extensiones Confiables, Seguridad.

## **1. Introducción**

Los primeros micro núcleos (μNúcleo) ofrecían flexibilidad, seguridad y modularidad, sin embargo no tuvieron el éxito esperado debido principalmente al pobre desempeño que el paso de mensajes generaba durante la transición de un espacio de direcciones protegido a otro.

Una segunda generación, ha logrado minimizar el problema del desempeño lo suficiente como para darles viabilidad práctica [1] (se logró bajar el costo del paso de de 100 a 10 s). Sin embargo, no se ha logrado desarrollar un SO que cubra las necesidades de las nuevas aplicaciones que se ejecutan en ambientes sumamente cambiantes y con fuertes restricciones en cuanto a los recursos de cómputo. En este sentido, consideramos que otro de los servicios básicos de un  $\mu$ Núcleo, debe ser su capacidad para reconfigurarse dinámicamente y así ajustarse a los cambios en su entorno. Otra de las principales limitantes para el desempeño de los SO basados en  $\mu$ Núcleo es la sobrecarga causada por la protección de hardware requerida para implementar diferentes espacios de direcciones. Para solucionar estos problemas proponemos un esquema de extensiones confiables, así como esquema de administración de memoria basado en un espacio de direcciones simple.

El resto de este artículo está organizado de la siguiente forma. La sección 2, describe las principales propiedades de la arquitectura de  $\mu$ Núcleo. En la sección 3, se presenta la arquitectura del sistema. En la sección 4 presentamos los resultados preliminares de las pruebas realizadas a nuestro  $\mu$ Núcleo. Finalmente la sección 5 presenta nuestras conclusiones y trabajos futuros.

## **2. Arquitectura de $\mu$ Núcleo**

El pobre desempeño de los primeros micro-núcleos fue producido principalmente por la reutilización y adaptación del código de SO preexistentes. Con esto en mente, la segunda generación fue diseñada e implementada a partir de cero y esta vez el desarrollo se enfocó en aprovechar las características específicas de la arquitectura y obtener un conjunto mínimo de mecanismos altamente optimizados. Representantes de esta nueva generación son L4Ka [2] y QNX [3]. El conjunto mínimo de abstracciones que un  $\mu$ Núcleo normalmente implementa está formado por: Hilos, Espacios de Direcciones, y Mecanismos de Comunicación entre Procesos (IPC).

Mientras que los sistemas basados en la arquitectura  $\mu$ Núcleo ofrecen la mejor relación entre protección y flexibilidad, tienen serios problemas de eficiencia. Por su parte los sistemas operativos extensibles ofrecen flexibilidad y eficiencia pero requieren de herramientas adicionales que aseguren la seguridad de las extensiones [4]. En el diseño de la arquitectura de  $\mu$ Cronos intentamos conjugar las ventajas de ambas arquitecturas para lograr un sistema que sea a la vez confiable, flexible y eficiente.

## **3. Arquitectura de $\mu$ Cronos**

Para definir nuestra arquitectura partimos del análisis de las principales tecnologías de extensión [4] y arquitecturas de sistemas operativos [5-10]. Como resultado, se definieron los principales servicios de  $\mu$ Cronos:

- **Extensiones confiables:** Define el conjunto de servicios que hace posible extender la funcionalidad del sistema, es responsable de obtener módulos certificados de un repositorio, verificar si provienen de un proveedor certificado para proceder a asignarle recursos y registrarlo dentro de la infraestructura adaptable.
- **Administración de memoria simple:** Permite hacer uso del espacio de direcciones virtuales gestionado por el micro núcleo; incluye toda la funcionalidad relacionada con el multiplexaje y administración del espacio de direcciones.
- **Hilos libres:** Provee la funcionalidad necesaria para gestionar la creación, administración y destrucción de hilos.
- **Infraestructura adaptable:** Ofrece facilidades para registrar y compartir objetos.
- **Acceso a repositorios de módulos de extensión:** Por medio de este servicio, los proveedores certificados pueden publicar módulos certificados para que otros sistemas puedan hacer uso de ellos.
- **Almacén de llaves:** Por medio de este servicio se registran las llaves certificadas y sus propietarios, así como su revocación en un momento dado.

Los detalles sobre cada una de las abstracciones básicas serán descritos en las siguientes secciones.

### 3.1. Extensiones Confiables

En pruebas comparativas realizadas a diversas tecnologías de extensión Small y Seltzer [4] concluyeron que las tecnologías interpretadas son muy lentas para responder a eventos del núcleo, a su vez los sistemas basados en protección por hardware generan demasiada sobrecarga por causa del constante número de mensajes y cambios de contexto, siendo entonces las tecnologías compiladas y con protecciones por software las más adecuadas para extender el núcleo de forma eficiente.

Por lo anterior nuestra tecnología de extensión se basa en un esquema de confianza compuesto por un criptosistema de llave pública, un almacén de llaves y repositorios de módulos. Bajo este esquema, una extensión confiable es implementada como un módulo firmado por un proveedor que a su vez es respaldado por una autoridad certificadora que tiene la responsabilidad de autenticar la identidad de los proveedores certificados. A continuación describiremos con más detalle cada uno de los módulos que componen el servicio de extensiones confiables.

***Criptosistema de Llave Pública.*** Las propiedades clave que el criptosistema de llave pública [11] aporta a las extensiones seguras son:

- **Autenticidad:** es posible validar la fuente de un módulo; es decir, que fue generado por un desarrollador propiamente identificado.
- **Integridad:** se asegura que un módulo no fue modificado.
- **No repudiación:** El creador de un módulo no puede negar la autoría del mismo.

**Módulos.** Un módulo es una entidad ejecutable que se apega a normas de construcción predefinidas y opcionalmente exporta una o más interfaces para que éstas sean usadas por otros módulos cooperativos. Típicamente estos módulos contienen funcionalidad extra que puede ser adicionada dinámicamente al núcleo del sistema operativo. Dado que los módulos son verificados para garantizar que provienen de una fuente confiable que asegura el correcto funcionamiento de la misma, la posibilidad de que un error en un módulo pueda afectar otras partes del sistema es reducida.

**Extensiones Confiables.** Por medio de un almacén de llaves y la autoridad certificadora es posible controlar, agregar o revocar permisos a los proveedores de servicios que estén certificados. El modelo soporta 3 niveles de confianza: total, restringido o nulo. Es decir, podemos permitir que se descarguen módulos de cualquier proveedor certificado, definir un grupo restringido de ellos o no permitir las extensiones. Para realizar una extensión confiable se necesita:

- Localizar el módulo que implementa la extensión deseada.
- Verificar que el módulo contenga la firma digital de un proveedor certificado y que su llave pública se encuentre en el almacén de llaves autorizadas.
- Cargar el módulo certificado en el sistema.
- Registrar el módulo en la tabla de objetos de la infraestructura adaptable.

### 3.2. Espacio de Memoria Simple

Nuestra arquitectura propone el uso de un espacio de direcciones simple (*Single Address Space*) que complementado con las extensiones confiables (*Trusted Extensions*) permite un modelo de protección fuerte no intrusivo en el que las extensiones no tienen que enfrentarse con la sobrecarga causada por el esquema de protección por hardware. Esto hace factible mejorar el desempeño y flexibilidad mientras que por medio del esquema de confianza se disminuye la posibilidad de que las extensiones realicen operaciones invalidas.

En nuestro esquema, un fragmento del espacio de direcciones es considerado como el contenedor de un objeto [12]. El sistema mantiene información relacionada con los contenedores de objetos en dos partes: La primera en el mapa de espacios de direcciones gestionado por el administrador de memoria (donde se maneja información administrativa como posición, tamaño, etc.). La segunda en la tabla de objetos que es gestionada por el administrador de la configuración y que controla las referencias a objetos entre diversas entidades del sistema.

### 3.3. Hilos Libres

En nuestra arquitectura no existe el concepto de proceso y los hilos comparten un único espacio de direcciones. De esta manera los hilos pueden compartir código y datos (objetos) sin la sobrecarga impuesta por atravesar espacios de direcciones protegidos que caracteriza a los procesos. Este esquema favorece la adaptabilidad pues trata al hilo

como la unidad básica de extensión, haciendo posible las extensiones de grano fino [13] en el sistema. Por otro lado, si es requerido, un hilo puede ser extendido hasta convertirse en un proceso. El modelo de hilos libres, ofrece ventajas importantes sobre los hilos convencionales:

- Tienen pocos problemas de contención gracias a que comparten un espacio de direcciones único.
- Promueven un mayor promedio de aciertos en el TLB por medio de un conjunto único de referencias a páginas frecuentemente usadas por todo el sistema y no las de un conjunto por cada aplicación como sucede normalmente en los procesos.
- Se eliminan los costosos *TLBs flushing*.
- Estos hilos pueden enlazar y usar cualquier procedimiento o estructura de datos.

### 3.4. Infraestructura Adaptable

Esta abstracción define a una estructura de datos responsable de almacenar las interfaces de los módulos cargados y descargados sobre demanda. Actúa como un repositorio de interfaces por medio del cual se mantiene la visibilidad global de los módulos enlazados en el sistema. Las extensiones son implementadas como módulos bien definidos que sólo interactúan por medio de sus interfaces. Una vez que se permite instalar una extensión, ésta puede acceder sólo a los componentes publicados en la tabla de objetos que es controlada por el administrador de la configuración.

## 4. Pruebas

Actualmente se cuenta con una versión funcional para arquitectura IA-32 y se planea desarrollar 2 más para las arquitecturas ARM e IA-64. La versión actual logra un desempeño aceptable con respecto de sistemas similares. En nuestras pruebas seleccionamos 4 sistemas: Mungi, L4Ka, Aegis y Linux para mostrar el notable incremento en el desempeño obtenido gracias a los avances tanto del hardware como del software de sistema. Los resultados se muestran en la Tabla 1.

Servicio	Cronos P4m, 1.8GHz	Mungi R4600, 100MHz	L4Ka R4600, 100MHz	Aegis DEC3100 16.7 MHz	Linux R4600, 100MHz
threadCreate()	1.24	48	10	N/A	N/A
threadDelete()	0.42	48	N/A	N/A	N/A
objectCreate()	0.49	60	N/A	N/A	N/A
objectDelete()	0.15	150	N/A	N/A	N/A
bind()	0.62	N/A	N/A	N/A	N/A
call()	0.02	30	0.86	10.4	160
Procedure call	0.01	N/A	N/A	0.42	2.9
System call	0.71	4.6	N/A	2.9	6.3

**Tabla 1.** Desempeño de los servicios básicos.



## 5. Conclusiones y Trabajos Futuros

Este artículo describe el diseño de  $\mu$ Cronos, un  $\mu$ núcleo para SO dinámicamente extensibles. Este  $\mu$ núcleo incorpora varias características novedosas, como el uso de una infraestructura de llave pública para aumentar la confiabilidad de extensiones dinámicas, así como un espacio de direcciones simple que brinda flexibilidad y mejoras en el desempeño; además del administrador de la configuración que en colaboración con los hilos escalables, favorece la reconfiguración dinámica del sistema.

Finalmente, consideramos que los sistemas modernos y en particular aquellos que se ejecutan en dispositivos con recursos restringidos requieren reconfigurabilidad en el nivel del sistema para obtener un mejor desempeño y flexibilidad, y la arquitectura de Cronos es una alternativa para soportar estos requerimientos.

## Referencias

- [1] J. Liedtke, "Toward Real Microkernels", *Communications of the ACM*, Vol. 39, No. 9, Sept. 1996, pp. 70-77.
- [2] J. Liedtke, et al, "The L4Ka Vision", University of Karlsruhe, System Architecture Group, Karlsruhe, Germany, April 2001.
- [3] D. Hildebrand, "An architectural overview of QNX", 1st USENIX Workshop on Microkernels and Other Kernel Architectures, Seattle, WA, April 1992, pp. 113-126.
- [4] Christopher A. Small and Margo Seltzer, "A Comparison of OS Extension Technologies", *Proceedings of the 1996 USENIX Technical Conference*, San Diego, CA, January 1996.
- [5] Deepack Gupta, "Extensible Operating Systems", Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, April 2001.
- [6] Gernot Heiser, Kevin Elphinstone, Jerry Vochtelloo, Stephen Russell and Jochen Liedtke, "The Mungi Single-Address-Space Operating System", *Software: Practice & Experience*, July 1998.
- [7] Dawson R. Engler, "The Exokernel Operating System Architecture", Massachusetts Institute of Technology, Depto. of Electrical Engineering and Computer Science, Doctoral Thesis, October 1998.
- [8] Ivan Bowman, Saheem Siddiqi, Meyer C. Tanuan, "Concrete Architecture of the Linux Kernel", Department of computer Science, University of Waterloo, Feb 1998.
- [9] Jean-Philippe Fassino, Jean-Bernard Stefani, Julia Lawall and Gilles Muller, "THINK: A Software Framework for Component-based Operating System Kernels", *USENIX Annual Technical Conference*, Monterey, California, USA, June 2002.
- [10] Andrew S. Tanenbaum, "A Comparison of Three Microkernels", Dept. of Mathematics and Computer Sciences, Vrije Universiteit, Amsterdam, The Netherlands, 1995.
- [11] R. Rivest, A. Shamir, L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", *Communications of the ACM*, Vol.21, No.2, Feb. 1978, pp. 120-126.
- [12] Daniel Cutting, "Protection Domains and Threads in Mungi", School of Computer Science and Engineering, The University of New South Wales, Sidney2052, Australia 1998.
- [13] Trent Jaeger, Jochen Liedtke and Nayeem Islam, "Operating System Protection for Fine-Grained Programs", *Proceedings of the 7th USENIX Security Symposium*, San Antonio, Texas, January 1998.

# Context-Aware Middleware for Pervasive and Mobile Ad Hoc Environments

Hector A. Duran-Limon

Department of Computing Science, Tecnológico de Monterrey (ITESM), Campus Guadalajara, México.

Gordon S. Blair, Adrian Friday, Thirunavukkarasu Sivaharan, Maomao Wu, Paul Okanda, Carl-Fredrik Sørensen

Computing Department, Lancaster University, Bailrigg, Lancaster LA1 4YR, UK

**Abstract:** Recent advances in the area of mobile ad hoc computing and pervasive computing have driven the emergence of new challenges. For example, the “Intelligent Environment” or “Smart Environment” has become one of the key research areas in the pervasive computing arena. Mobile ad hoc scenarios also include time-critical applications such as an autonomous vehicle system in which vehicles are able to operate independently and cooperate with each other to avoid collisions. In this paper, we present a context-aware middleware architecture for the support of both pervasive and mobile ad hoc environments.

**Resumen:** Avances recientes en el área de computación móvil ad hoc y cómputo penetrante (pervasive computing) han dado cauce al surgimiento de nuevos retos. Por ejemplo, el “Ambiente Inteligente” se ha convertido en una de las áreas claves de investigación en cuanto a cómputo penetrante se refiere. Por otro lado, escenarios de computación móvil ad hoc incluyen aplicaciones de tiempo real tales como sistemas autónomos de vehículos, mismos en los que los vehículos son capaces de operar de manera independiente y cooperar entre ellos para evitar colisiones. En este artículo presentamos una arquitectura de middleware para el apoyo de cómputo penetrante y ambientes móviles ad hoc.

**Keywords:** mobile ad hoc networks, pervasive systems, real-time systems, middleware

## 1. Introduction

Recent years have witnessed advances in the enabling technologies for mobile computing, such as the increasingly mature end-systems, various kinds of wireless communication protocols, and mobile networking technologies. As a result new challenges have emerged in both the pervasive and mobile ad hoc computing arenas. Regarding the former, one of the key research areas in recent years is the “Intelligent Environment” or “Smart Environment”. A personalized intelligent room, for example, can set the room temperature according to the occupier’s preference, play his beloved music, and make his favourite coffee, etc.

In addition, examples of mobile ad hoc applications include air traffic control systems whereby thousands of aircraft are proactively coordinated to keep them at safe distance from each other, direct them during takeoff and landing from airports, and ensure that traffic congestions are avoided. Another example is an autonomous vehicle system in which vehicles are able to operate independently and cooperate with each other to avoid collisions. This kind of systems is time-critical and needs to be provided with both timeliness assurances and adaptation mechanisms that lead the system to a safe state in case of unexpected changes introduced in the environment.

This paper provides a solution to three important issues: a) support for sentience and intelligent behaviour, b) support for high mobility and pervasiveness, and c) real-time support for mobile ad hoc environments. These challenges are addressed in an integrated manner by our middleware architecture as shown below. We are not aware of any work addressing the three issues altogether.

The paper is structured as follows. Section 2 introduces the main challenges of the targeted applications. Relevant related work is mentioned in section 3. Section 4 then presents the overall middleware architecture. Section 5 shows how the challenges are addressed by our middleware architecture. The implementation of a prototype is presented in section 6. Finally, some concluding remarks are provided in section 7.

## **2. Challenges**

### **2.1 Challenge 1: Provide intelligent behaviour to pervasive systems**

In order to make pervasive computing systems intelligent, it is crucial for them to first be sentient – they should be able to sense the changes in the environment and make autonomous decisions on how to react. For example, a “smart room” should be able to obtain the room temperature, the noise level, the light intensity, how many persons are in the room, etc. Moreover, they should also have the capability to process and analyse the sensory data to deduce high-level context and infer long-term personal preferences of the users. By utilising the contextual information and personal preferences, the pervasive computing environments can make autonomous decision according to their own control logic. This decision making process might need certain degree of intelligence, and it might also need to resolve conflicts in the pervasive computing environment. For example, if there are more than two persons in a “smart room”, the system should be able to decide what temperature to set to minimize the disturbance, what kind of music to play to maximize satisfaction, etc.

### **2.2 Challenge 2: Provide support for high mobility and pervasiveness**

We envisage hundreds and even thousands of autonomous entities being involved in the targeted applications. Therefore, suitable communication models are required to handle a many-to-many type of communication. The communication service should be reliable and efficient, i.e. the amount of resources used by the communication protocol should be maintained at acceptable levels. These systems should also be scalable due to the fact that the number of participants may increase considerably over time. Furthermore, the level of mobility used by the system should be considered as this can have an impact on the communication system. For instance, research has shown that communication protocols used in low mobility conditions are not optimal for high mobility scenarios [1]. In addition, mobile ad hoc networks are characterised by being highly unpredictable. Communication delays between nodes may vary unexpectedly as the number of hops to reach the destination changes. Moreover, a geographical area may unexpectedly become congested, resulting in lack of communication resources. Periods of disconnection are also likely to happen at any time due to the conditions of the geographical area.

### **2.3 Challenge 3: Provide QoS guarantees in mobile ad hoc environments**

In the current state-of-the-art in mobile ad hoc network technologies, it is not feasible to offer hard real-time guarantees for communication resources except in special cases. Such guarantees can only be offered if certain specific conditions are met such as a scenario whereby a limited number of nodes are moving in an obstacle-free area at the same speed and direction. Although, hard real-time guarantees can still be provided for local resources such as CPU, only soft real-time guarantees can be offered for communication resources in most cases whereby these resources are dynamically allocated according to deadlines. Another tough challenge arises here: how can we deal with applications demanding hard real-time guarantees when the underlying infrastructure is only capable of offering soft real-time guarantees.

### 3. Related work

The Gaia project [2] developed at the University of Illinois is a distributed middleware infrastructure that provides support for ubiquitous computing. Although Gaia shares several common design goals with CORTEX, Gaia's main intended application domain is confined to fixed intelligent environments and lacks the support for real-time mobile ad hoc application scenarios. The EasyLiving project [3] from Microsoft focuses on development of architectures and technologies for intelligent environments. It identifies several research efforts required on a variety of fronts, including middleware, geometric world modelling, sensing capabilities, and service description. The middleware, called InConcert, identifies the importance of having an asynchronous communication model for the coordination of entities contained in the environment. However, this approach does not fully address the concerns of real-time mobile ad hoc environments and pervasive application scenarios.

Regarding multicast in mobile ad hoc networks, previous research [1] has recognised that most existing algorithms perform inadequately when high mobility is present in the environment. The main reason for these protocols to fail is that these protocols maintain shared state in the nodes in the form of routes and adjacent information, which are rapidly outdated due to high node mobility. In contrast, our multicast protocol is based on a probabilistic flooding algorithm with damping, which does not maintain shared state in nodes.

Finally, to the best of our knowledge, no previous work presents an integrated solution to the areas of autonomous intelligent behaviour, pervasive mobile systems and real-time mobile ad hoc computing.

### 4. Overall Middleware Architecture

Central to the CORTEX architecture is the notion of a *sentient object* [4-6], which is defined as an entity that is able to both consume and produce events. That is, sentient objects are entities that receive events, process them and generate further events. Input events are received from either sensors or other sentient objects (local or remote). Similarly, output events are sent either to actuators or other sentient objects (local or remote). Sentient objects are autonomous entities that are able to sense their environment. Interestingly, sentient objects have a proactive role in that they are capable of making decisions and performing some actions (i.e. generate further events) based on the information sensed. Hence, sentient objects include a control logic which realises the decision-making mechanism.

The environment support for the interaction of sentient objects is also conceptualised as a componentised middleware platform. In fact, the middleware is structured in terms of component frameworks (CFs) [7]. Essentially, component frameworks are "collections of rules and interfaces that govern the interaction of components 'plugged into' them" [7]. In other words, a component framework is a reusable architectural design targeting a specific domain whereby desired architectural properties and invariants are enforced. The publish/subscribe CF realises the CORTEX event model [8]. The functionality of the control engine of a sentient object is provided by the context CF. Facilities for multicast in ad hoc environments are then provided by the multicast CF. The QoS management CF arbitrates the allocation of resources and provides facilities for monitoring and adaptation of QoS. Lastly, the resource management CF controls the resources used by all the CFs [9]. The following section shows how the challenges outlined in section 2 are addressed by our middleware architecture.

## **5. Approach to Challenges**

### **5.1 Support for Autonomous Intelligent Behaviour**

The autonomous intelligent behaviour is achieved by the sentient objects. A sentient object has been defined as “an entity that both consumes and produces software events, and lies in some control path between at least one sensor and actuator” [9]. The internal architecture of a sentient object is realised by the context CF, which consists of four components: sensory capture, context management, inference engine and the learning engine. Getting contextual information from raw sensory data is the main task of context acquisition, and the major issues in the area of sensory capture are data filtering and sensor fusion. The context management component deals with the representation of context information. The inference engine component is actually the brain of a sentient object, and it has some form of a decision-making ability and intelligence. In order to make decisions, some rules need to be applied on the current context.

The learning engine component is complementary to the inference engine component, and it can dynamically plug-in different machine learning algorithms that make the system able to learn rules from previous context states and from interaction with the user. The learned rules can either be injected into the inference engine or used by the learning component itself to make decisions.

### **5.2 Support for High Mobility and Pervasiveness**

We believe the publish/subscribe or event based communication paradigm is well suited to address the issues of pervasive systems in a mobile ad hoc environment. Moreover, the publish/subscribe model is well recognised to support anonymous and asynchronous communication requirements [10-12]. Hence, many-to-many communication scenarios are well supported by the anonymous dissemination of information. In addition, asynchronous communication is ideal in systems where frequent disconnections are likely to happen thus to avoid blocking conditions. We have adopted the event model STEAM [8] as it addresses a number of core issues with regard to publish/subscribe systems in mobile ad hoc networks. Briefly, an implicit event model for ad hoc networks is defined in which there is not an event broker or mediator, instead brokering functions are implemented at both the consumer and the producer side.

The issues of high mobility in ad hoc environments are tackled by our probabilistic multicast protocol. This protocol specifically targets ad hoc environments where high node mobility and a frequently changing number of group members are present. Each node decides if it should forward a flooding message according to a probability  $p \in (0,1]$  which is updated according to the number of duplicates that a node has received from its neighbouring nodes. This effectively minimises the number of unnecessary duplicate messages without sacrificing reliability as we have found experimentally through simulations. The second mechanism, which is called damping, aims to eliminate the number of unneeded duplicates by allowing nodes to wait for a random, small time interval before they will actually forward a message.

### **5.3 Real-Time Support for Mobile Ad Hoc Environments**

In order to support QoS guarantees in mobile ad hoc environments, a high probability of meeting deadlines has to be offered. We also believe that such infrastructure has to be adaptable and flexible to deal with the highly dynamic and unpredictable nature of the environment. Resource management plays an important role in this adaptation process in terms of both resource awareness and dynamic reallocation of resources. Crucially, fail-safe mechanisms are needed to bring the application to a safe state when timing failures are detected. Therefore, QoS management support is required for monitoring QoS violations and triggering both adaptation and fail-safe

procedures when required. In addition, QoS management is required to arbitrate the allocation of network resources whereby admission control tests are performed and more resources are conceded to tasks with shorter deadlines and higher criticality..

Monitoring of QoS violations and adaptation procedures are carried out by the QoS management CF. This is achieved by the use of the *timely computing base (TCB)* [13] which is a framework that provides crucial time-related services. More specifically, TCB supports the detection of timing failures and enforces the timely execution of fail-safe procedures and adaptation strategies. Fail-safe procedures ensure that the system is taken to a safe state when a critical failure is detected. A coverage stability facility is provided where a high probability of meeting a deadline can be defined. In the case of going below a specified value, adaptation procedures are triggered (e.g. a redistribution of network and CPU resources).

Network QoS management is achieved as follows. Firstly, every node is able to listen to traffic as the dissemination of packets is carried out by using an application level multicast protocol. Secondly, available bandwidth is fairly distributed among the nodes within a transmission area. For this purpose, a fully distributed protocol is used. Thirdly, a weighted fair scheduling policy [14] is used to allocate bandwidth within a single node to multiple service classes, each one of them associated with a particular *task*. A task [15] is a logical unit of computation which has an amount of resources allocated. A service class (i.e., a task) can be partitioned into sub classes (i.e., subtasks). In addition, classes and subclasses have a priority value. More critical classes have a higher priority. Different policies can be defined in case of resource contention. A detailed description of the network QoS management protocol can be found in [15].

## 6. Implementation

As mentioned earlier, the implementation has been carried out using OpenCOM and running on Windows CE 3.0. A prototype of a STEAM-like publish/subscribe system has been built. A particular instantiation of the resource framework was also developed to extend the event system with resource management capabilities. Support for the management of CPU resources is provided at this stage. The resource management system offers dispatching predictability of real-time events. CPU reservations are organised in rounds. A round (or dispatch table) contains a number of time slots which are assigned to hard and soft real-time tasks. The framework has been realised by a two-level scheduling model implemented in Windows CE. A prototype of the Context CF provides the facility for the inference engine functionality implemented in CLIPS and sensor fusion functionality based on different fusion algorithms [16].

These CFs form parts of a cooperating sentient vehicle test bed [16] that has been built. The autonomous vehicles have the objective of travelling along any given path, defined by a set of GPS waypoints, i.e. a virtual circuit. Every vehicle that travels along the virtual circuit cooperates with other vehicles by inter-vehicle communication in order to avoid collisions.

The sentient vehicle is a modified Juggernaut 2 (Ford F-350) RC Car. The RC module has been removed and the car is self-controlled by an HP iPAQ Pocket PC via a RS232 connector. The car is augmented with a GPS module for location sensing, a electronic compass for sensing its orientation, and 8 units of ultrasonic sensors (with 3cm to 3m range) to sense the presence of physical objects. The Pocket PC has 2 WLAN cards in ad hoc mode. The more powerful WLAN card is for the exclusive use of TCB control channel and the other WLAN card is for the event channels (payload) used for inter-vehicle communications. The two WLAN operate in non-overlapping 802.11b channels. The on-board network is used on the sentient vehicle to broker data between the sensors, actuators and iPAQ. The on-board network is a bespoke ring

topology with single break failure resilience. Importantly, the design enables addition of further devices onto the ring in a plug-and-play fashion making it extensible. The test bed contains sentient vehicles and a laptop-based traffic light.

## 7. Concluding remarks

We have presented a middleware architecture for the support of both pervasive and mobile ad hoc computing environments. More specifically, the sentient object model has been introduced as a central concept in the CORTEX architecture. This model provides key features for allowing both context-awareness and intelligent autonomous behaviour. In addition, the middleware is constituted by a number of component frameworks. Autonomous intelligent behaviour is addressed by both the sentient CF and the context CF. In addition, mobility and pervasiveness issues are tackled by the publish/subscribe CF, the multicast CF and the service discovery CF. Lastly, the issues of time criticality in mobile ad hoc environments are addressed by both the QoS management CF and the resource management CF.

We are currently using the vehicle test bed to evaluate the middleware architecture. Ongoing work also includes the simulation of the network QoS management protocol and the experimental evaluation of the scheduling system.

## Reference

- [1] Lee, S.-J., et al. *A Performance Comparison Study of Ad Hoc Wireless Multicast Protocols*. in *Proceedings of IEEE INFOCOM 2000*. 2000. Tel Aviv, Israel.
- [2] Roman, M., et al., *Gaia: A Middleware Infrastructure to Enable Active Spaces*. IEEE Pervasive Computing, 2002. **Oct-Dec**: p. pp. 74-83.
- [3] Brumitt, B., et al., *EasyLiving: Technologies for Intelligent environment*. Handheld and Ubiquitous Computing, 2000. **September**.
- [4] CORTEX. *Preliminary Definition of the CORTEX Programming Model*. CORTEX Project. IST-2000-26031, Deliverable D2. March 2002.
- [5] Verissimo, P., et al. *CORTEX: Towards Supporting Autonomous and Cooperating Sentient Entities (2002)*. in *In Proceedings of European Wireless 2002*. 2002. Florence, Italy.
- [6] Verissimo, P. and A. Casimiro. *Event-Driven Support of Real-Time Sentient Objects*. in *Eighth IEEE International Workshop on Object-oriented Realtime Dependable Systems (WORDS 2003)*. 2003. Guadalajara, Mexico.
- [7] Szyperski, C., *Component Software: Beyond Object-Oriented Programming*. 1998, Harlow, England: Addison-Wesley.
- [8] Meier, R. and V. Cahill. *Steam: Event-based Middleware for Wireless Ad Hoc Networks*. in *In Proceeding of the International Workshop on Distributed Event-Based Systems (ICDSC/DEBS'02)*. 2002. Vienna, Austria.
- [9] CORTEX. *Preliminary Specification of Basic Services and Protocols*. CORTEX Project. IST-2000-26031. Deliverable D5. February 2003.
- [10] OMG, *CORBAServices: Common Object Services Specification*. 95-3-31. 1998, Object Management Group.
- [11] Sun, *Java Distributed Event Specification*. 1998, Sun Microsystems, Inc. <http://www.javasoft.com/products/javaspaces/specs>.
- [12] Carzaniga, A., Rosenblum, D. and Wolf, A., *Design and Evaluation of a Wide-Area Event Notification Service*. ACM Transactions on Computer Systems, 2001. **19**(3): p. pp 332-383.
- [13] Verissimo, P. and A. Casimiro, *The Timely Computing Base Model and Architecture*. Transaction on Computers - Special Issue on Asynchronous Real-Time Systems, 2002. **51**(8).
- [14] Demers, A., S. Keshav, and S. Schenker, *Analysis and simulation of a fair queueing algorithm*. Journal of Internetworking Research and Experience, 1990: p. pp. 3-26.
- [15] Duran-Limon, H.A. and G.S. Blair. *A Resource and QoS Management Framework for Real-Time Event Systems in Mobile Ad Hoc Environments*. in *In 9th IEEE International Workshop on Object-oriented Real-time Dependable Systems (WORDS 2003F)*. 2003. Capri Island, Italy.
- [16] Sivaharan, T., et al. *Cooperating Sentient Vehicles for Next Generation Automobiles*. in *MobiSys 2004 Workshop on Applications of Mobile Embedded Systems*. 2004. Boston, MA, USA.

# Seamless IP micro-mobility

Christian P. García-Martínez, J. Antonio García-Macías

Departamento de Ciencias de la Computación  
Centro de Investigación Científica y Educación Superior de Ensenada, CICESE.  
e-mail : {cpgarcia, jagm}@cicese.mx  
Km. 107 Carretera Tijuana – Ensenada, CP 22860 AP 2732 Ensenada, B.C. México

**Abstract:** Mobility protocols allow mobile devices to roam within networks, performing handovers between wireless cells. Context transfer mechanisms aim for performance improvements in the process of handover, so that the applications running on mobile nodes can operate with minimal disruption, thus allowing seamless mobility. In this paper we propose mechanisms that enable context transfer between access routers offering internet connectivity for mobile nodes; our proposal includes means for representation of contexts for services. We use two candidate services (MLD and header compression) as examples to evidence the benefits of context transfer.<sup>1</sup>

**Resumen:** Los protocolos de movilidad permiten a los nodos moverse entre redes realizando *handovers* entre células inalámbricas. El mecanismo de transferencia de contexto intenta realizar mejoras al proceso de *handover*, de tal forma que las aplicaciones que están corriendo en el nodo móvil sigan operando con la menor alteración posible, logrando así una movilidad transparente. En este artículo proponemos mecanismos que permiten llevar a cabo la transferencia de contexto entre enrutadores de acceso; nuestra propuesta incluye la representación del contexto de los servicios. Presentamos dos servicios candidatos como ejemplo para evidenciar los beneficios de la transferencia de contexto.

**Palabras Clave:** Context Transfer, Micro-mobility, Seamless mobility.

## 1. Introduction.

The Internet Protocol (IP) has become the de facto standard for connectivity of different types of nodes all over the world. Mobile devices that move within an IP-based network introduce new problems, such as, the problem of how to route the packets to Mobile Nodes (MNs) when these change their point of attachment to the network. Mobile IP [3] proposes a mechanism that solves this problem, however, it is not possible to achieve seamless mobility due to the high cost of signaling and registration procedures incurred. That is why, many people are working on how to improve the efficiency of mobile IP, designing extensions that aim at better performance in certain scenarios. Specific characteristics have been identified when

---

<sup>1</sup> Financial support for this project has been provided by the Mexican Council for Science and Technology (CONACyT).



the mobility is within an administrative domain; this has led the research toward the study of micro-mobility [1].

Context transfer has emerged as one of the problems that must be solved to achieve seamless mobility in IP networks. In general, context is the information on the current state of a service, required to re-establish the service on a new subnet without having to perform the entire protocol exchange between the MN and the Access Router (AR) from scratch. In this paper we propose mechanisms that enable context transfer between access routers offering Internet connectivity for mobile nodes; our proposal includes means for representation of contexts for services. RFC 3374 [8] describes the main reasons why Context Transfer procedures may be useful in IP networks.

The remainder of the paper is organized as follows: Section 2 overviews related work. Section 3 describes potential customers for Context Transfer Protocol (CTP). Section 4 describes the CTP; there, we also propose a format to represent the context of two candidate services (MLD and HC). Section 5 discusses our experience with its implementation; we provide conclusions and mention some future work.

## **2. Related Work.**

The work of Koodli and Perkins [12] was one of the first that mentioned context transfer in mobile networks. There are a couple of Internet drafts on generic context transfer protocols [9, 11] proposed by the IETF Seamless Mobility (Seamoby) working group [5]. These proposals have not been evaluated through analytical, simulation or test bed means; the exact deployment scenarios are not clear yet. It is also unclear how these protocols represent the different types of context meant to be transferred. The IRTF Micro-mobility research group was working on these topics, but they have been dormant for about a year. There is a new IRTF Mobility Optimizations (MOBOPTS) [6] research group that is working on this but to date, they haven't presented any related proposal.

## **3. Candidate services for context transfer.**

Mobile nodes improve the performance of their connections across wireless media by establishing various kinds of state (context), in order to use the available bandwidth securely and economically [8]. For example, a particular node establishes the Header Compression (HC) service by collecting enough information about its data streams to get synchronized with its AR and then be able to send and receive compressed packets. This state information is needed to keep the service available. Once the mobile move to another subnet, it is required to re-establish the service by the same process it used to initially establish it, and delay-sensitive traffic may be seriously impacted. An alternative is to transfer the context of the service to the new subnet so that the service can be re-established quickly, rather than establish it from scratch. The context transfer may be advantageous in minimizing the impact of host mobility on certain cases. In the

following paragraphs we describe two of the candidate services we consider good examples to evidence the benefits of context transfers.

**Multicast Listening Discovery (MLD).** We consider the transfer of context of MLD [13] as an example of how context transfer can improve the performance of an IP layer handover. The purpose of MLD is to enable each IPv6 router to discover the presence of multicast listeners on its directly attached links, and to discover specifically which multicast addresses are of interest to those neighboring nodes. This information is then provided to whichever multicast routing protocol is being used by the router, in order to ensure that multicast packets are delivered to all links where there are interested receivers. MLD is a good example to show the benefits of context transfer because every node must perform two MLD messaging sequences per subscription on the wireless link to establish itself as an MLD listener. If this state information is transferred to the new router it is possible to avoid these messaging sequences.

**Header Compression (HC).** The use of bandwidth constrained links in wireless networks suggests the use of bandwidth saving HC schemes. A number of HC schemes have been proposed for existing Internet protocols, one of them is Robust Header Compression (RoHC) [2]. In some cases, for example using IPv6/UDP/RTP, the headers can be reduced from 84 bytes to 1 byte, and therefore, can reduce the load on the wireless link by 50-70% [4]. The HC process is stateful; the compressor and the decompressor, both maintain records to represent the expected content of protocol headers, so that headers fields that have the expected values do not need to be retransmitted, since they can be inferred. Establishing HC generally requires from 1 to 5 packets exchanges between the last hop router and the mobile node with full or partially compressed headers before full compression is available [2]. During this period, the MN will experience an effective reduction in the available bandwidth equivalent to the uncompressed header information sent over the air. An approximate quantitative estimate for the amount of savings in handover time can be obtained as follows. A RTP/UDP/IPv6 message contains 84 bytes only in their headers; as we said before, from 1 to 5 messages with full headers are required. Let's say 3 messages are sent before the MN and the AR are synchronized, that means about 252 bytes in headers. Some representative numbers can be obtained by assuming bandwidths of 20 kbps or 100 kbps. With these two bit rates, the savings from not having to send these messages with full headers are 110 and 20 ms. If there are more data streams being compressed, the amount of time saved could be significantly more.

#### **4. Context Transfer Protocol**

We believe that it may be advantageous to bundle the context transfer signaling between the MN and network, or between access routers, together with other handover signaling. We use a micro-mobility protocol that had simplicity as one of its main design requirements [7]. When a handoff request is sent, specific information is piggybacked to the new access router indicating that context transfer will take place (if any). Previous work has explored the issue of finding out if it is possible to perform

context transfer [10]; it also describes mechanisms for Candidate Access Router Discovery (CARD), which allows knowing the characteristics of the new access router before the handover.

#### 4.1. CTP overview

Context transfer takes place in response to a context transfer trigger; which could be a handoff request or an explicit context transfer request sent to the previous AR (pAR). The MN initiates the process by sending a Context Transfer Request (CTR) message to the pAR. Once the pAR has received the message, it generates a Context Transfer (CT) message, which contains all the state information associated to the mobile node and capable of being interpreted by the nAR. The nAR receives the context and relocates the state information in order to offer the same treatment to the data flows as the pAR did. Performing context transfer in advance of the MN attaching to the new AR (nAR) can increase handover performance. In order to make this happen, certain conditions must be met. For example, pAR must have sufficient time and knowledge about the impending handover. This has proven to be feasible [10].

Every message has the format illustrated in figure 1, the header is the same for all messages and is comprised by a Type field, which indicates the type of the message (e. g. CTR); a Length field that contains the size of the complete message; a Flags field that contains indicators as Version, need for Acknowledgement, sequence number; and an authorization token required to ensure that the MN has the rights to exchange messages within the domain. The Data Block varies depending on the type of message.

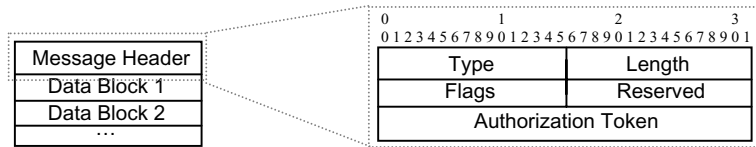


Fig. 1: Generic message format.

The context data structure needs appropriate encapsulation for communication and processing, figure 2 shows the format of a data block for a CT message. The field Context Type indicates the interpretation of the next context blocks. The context type defines how the state information is organized. In section 4.2 we present a specific representation for the state information of MLD and Robust Header Compression. The amount of information in a Context Data Block is not fixed, due to the fact that a node could have any number of data streams being treated by the same service. For instance: multiples audio/video sessions using RoHC.

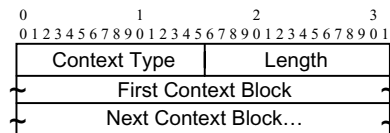


Fig. 2: Context Data Block format.

The messages will use UDP as the transport protocol for the Context Transfer Request and Context Transfer Data messages. Signaling messages will use ICMP.

#### 4.2. Feature Context Representation

Every context type requires a different way of representation due to the differences between the state information of the services. Despite these differences, we can organize the context in a generic way as follows: a MN has associated data streams, data streams could have associated different services, and each service has associated state information which is called service context. We have identified and represented the context of two candidate services (MLD and RoHC). Figure 3a shows the specific data that need to be transferred to enable the MLD service in a nAR; this is what is carried on the Context Block field showed in figure 2. The Subnet Prefix on the nAR Wireless Interface field contains a subnet prefix that identifies the interface on which multicast routing should be established, as sometimes a router has more than one interface to the same link and it is necessary to perform the router part of MLD only over one of those interfaces. The Subscribed IPv6 Multicast Address field contains the multicast address for which multicast routing should be established.

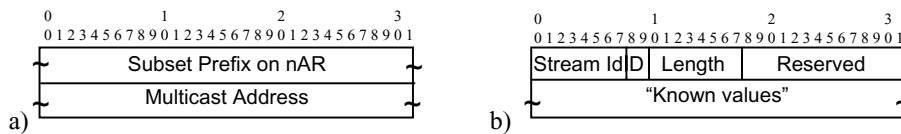


Fig. 3: a) Representation of MLD context. b) Representation of RoHC context.

In order to determine the information that needs to be transferred to enable the RoHC service in the nAR, it is necessary to know how the header fields can be classified based on their behavior during the session lifetime. RFC 3095 [2] allow us to distinguish those fields that can be inferred and those that are known. Once they are identified it is necessary to get the values for each instance of the fields, and then associate each instance with a data stream ID. This is how the context feature data for this service is built. This information is sent to the nAR so it can reconstruct the headers, gathering the data received in the context transfer, the data inferred by the service and the data contained in the compressed headers.

#### 5. Conclusions

To date, many interesting proposals have been made for using context transfer to speed up handovers. Some of them are complex enough to discard any short-time applicability. That is why our proposal is meant to be simple but functional for a controlled environment, enabling the identification of requirements that only experimental tests can provide. We also take advantage of the way the micro-mobility protocol works, reducing the traffic during the handover. The reason why we use RoHC instead of other proposed schemes for HC is because its design and performance

[4] helps to evidence the benefits of context transfer. On the other hand, due to the fact that, currently, there is not any freely available implementation of a RoHC scheme, determining the advantages of context transfer for header compression through test bed means, requires carefully designed experiments. We believe that more experimental results are needed to lead the research in this area. We are working on the deployment of this framework in order to evaluate our proposal through simulations and experimental means; this has led us to face new problems that must be solved in order to achieve seamless mobility. The current work can be extended by applying our approach to different candidate services, which also implies identifying the context of the candidate services, as well as defining structures to carry it through the network.

## References

- [1] A.T. Campbell, J. Gomez, S. Kim, C.Y. Wan, Z.R. Turanyi, and A.G. Valko, "Comparison of IP micromobility Protocols", *IEEE Wireless Communications*, 9(1):72–82, 2002.
- [2] C. Bormann, "Robust Header Compression (ROHC)", RFC 3095, IETF, July 2001.
- [3] C.E. Perkins, "Mobile IP Support", RFC 2002, IETF, October 1996.
- [4] C. Westphal and R. Koodli, "IP Header Compression: A Study of Context Establishment", *IEEE Wireless Communications and Networking Conference*, New Orleans, USA, 2003.
- [5] IETF. Seamless Mobility (SeaMoby) Working Group.  
Homepage:<http://www.ietf.org/html.charters/seamobycharter.html>.
- [6] IETF. IP Mobility Optimizations (Mobopts) Research Group.  
Homepage:<http://www.irtf.org/charters/mobopts.html>
- [7] J. Antonio García Macias, "Mobile Communication Architecture with Quality of Service", Ph. D. Thesis, Institute National Polytechnique de Grenoble, January 2002.
- [8] J. Kempf, "Problem Description: Reasons for Performing Context Transfers between Nodes in an IP Access Network", RFC 3374, IETF, September 2002.
- [9] J. Loughney, "Context Transfer Protocol", Internet Draft, IETF, January 2004.
- [10] Juan M. Oyoqui and J. Antonio Garcia Macias, "Context Transfer for Seamless Mobility", 4th Mexican International Conference on Computer Science, (ENC) Tlaxcala, Mexico, Sept. 2003.
- [11] R. Koodli R. and C. E. Perkins, "A Context Transfer Protocol for Seamless Mobility", Internet Draft, IETF, August 2002.
- [12] R. Koodli R. and C. E. Perkins, "Fast Handovers and Context Transfers in Mobile Networks", *ACM Computer Communication Review*, vol. 31, number 5, 2001.
- [13] S. Deering, "Multicast Listener Discovery for IPv6", RFC 2710, IETF, October 1999.

# Requirements analysis for ubiquitous computing system to be used during a congress<sup>\*</sup>

Oscar-Iván Lepe-Aldama<sup>‡</sup>, Luis-Enrique Vizcarra-Corral<sup>†</sup>, Edgar Arriola-Arreola<sup>‡§</sup>,  
Angel Galindo-Rodríguez<sup>‡§</sup>, Frankli-Augusto Pérez-Navidad<sup>‡§</sup>

<sup>‡</sup>Centro de Investigación Científica y de Educación Superior de Ensenada  
Dirección de Telemática

Km. 107 Carretera Tijuana-Ensenada, Ensenada 22860, Baja California, México  
[olepe,earriola,agalindo,anavidad]@cicese.mx

<sup>†</sup>Universidad Autónoma de Baja California, Facultad de Ciencias

Km. 104 Carretera Tijuana-Ensenada, Ensenada 22860, Baja California, México  
luisvi@uabc.mx

**Abstract.** *In this paper we present a scenario-based requirements analysis for a prototype of an ubiquitous computing system: a handheld-computing supported cooperative work system to be used during a congress or any other kind of technical or scientific people gathering. The prototype is working on several iPAQ h3955 running JVM for Windows CE and on some Zaurus SL-5600 running JVM for Linux Embedded; all connected by a WiFi WLAN.*

## 1. Introduction

In a provocative article [1], Arnold Kling questions: “Where is the next generation of ‘killer applications’ that will drive mainstream adoption of technologies that are tantalizingly close to realization, such as wireless Internet access, pervasive computing, and radio on a chip?... A reasonable guess is that the next wave of compelling applications will be in the realm of what [some] call social software;” where social software is, basically, software that supports group communications. Our research collaboration between CICESE and UABC exploits this idea and proposes the construction of a handheld computer-supported cooperative work system for driving our investigations on ubiquitous computing. Mark Wiser in its seminal article [2] calls ubiquitous computing “the next wave” in computing. Computers will pervade the user surroundings, becoming an integral part of the physical environment in a non intrusive way, and will use context information to dynamically build appropriate computing models for provide information and services at the right place in the right time. The main idea of our investigation in ubiquitous computing is to use applications as beacons for signaling technology improvement opportunities.

This paper presents our experiences building a research experimental environment—

---

<sup>\*</sup> Financial support for this work has been provided in part by a grant from Mexico’s University Consortium for the Development of Internet (CUDI), and the National Council for Science and Technology (CONACYT).

<sup>§</sup> Mr. Arriola-Arreola, Mr. Galindo-Rodríguez and Mr. Pérez-Navidad are Bachelor students from UABC’s Computer Science program carrying out a bachelor thesis fellowship at CICESE.

a social software to be use during a congress or any other kind of technical or scientific people gathering. Congress scenarios define no trivial collaboration and communication processes between people, which impose infrastructure requirements, both hardware and software, not well provisioned nowadays. In a related paper [3], one of our colleagues explores some of these infrastructure deficiencies and proposes a technology improvement. The application prototype, here presented, relies on handheld computing, wireless networking, and multiagent middleware, aimed at constructing a ubiquitous computing environment for a congress. The main contributions of this paper are to discuss the requirements analysis of this social software.

## **2. Social software for a congress environment**

Congresses provide to attendants opportunities to meet with colleagues, to interchange information with them, and to set work relationships, all this for improving professional performance and accelerate knowledge advancement. However, it is well known that it is difficult to exploit all these opportunities within a regular congress time frame, because it is not easy to find people with corresponding interests or personal profile; or to select the right paper presentation at multitrack congresses. Moreover, congress management (i.e., attendants sign-in; last-minute schedule changes; localization of people and sites; congress related leisure activities coordination) also requires collaboration and communication between selected people. All these problems get accentuated because people in a congress moves freely around the congress building. From this it should be evident that congress attendants would greatly improve congress results if they could count with a computing system that assists them with all this information technology requirements. The computing system we refer must support user mobility and provide the right information at the right time.

Fortunately, handheld computing is mainstream nowadays. Most if not all congress attendants count with handheld computing devices like personal digital assistants, advanced cellular phones, and pocket or tablet personal computers. These handheld devices are equipped with Personal Information Management software—agenda, calendar, and to-do list—which provides a ubiquitous software infrastructure upon which one may build social software. On the other hand, most if not all of these devices carry some type of wireless data communications hardware and software—Bluetooth, WLAN (802.11a/b/g) or WMAN (GPRS). At the same time, WMAN networks cover all of the busiest cities in the world and business buildings there are equipped with WLAN base stations. All this provides a ubiquitous data communications infrastructure upon which one may build social software.

## **3. Typical scenarios for a congress environment**

Here we present three of the scenarios we identified may take place in a not far future congress. We use these scenarios, and some others, for driving the design of the application prototype. Clearly, a ubiquitous computing environment could be made to pro-

vide the following functionality.

**Arriving to the congress.** Upon arriving to the congress location, a congress attendant must perform a series of time consuming activities in order to get registered and schedule its work and leisure activities. Registration involves filling out forms with personal information, even in the case attendants registered itself in advance. Surely, in this last case the registration procedure takes less time but the processes cannot be avoided—the congress' organization requires to have precise information of who has come a who has not.

**Scheduling activities.** Attendants, after registering and receiving the congress program, should start to schedule its work and leisure activities. Most of the times, the difference between obtaining excellent results and obtaining just good or bad ones is having a good congress schedule. Making a schedule for a small congress may not be a problem, if one only considers the size of the information at hand. But things may get complicated if we consider that almost always there is at least one paper presentation that gets cancelled after the congress program was printed. On the other hand, for large congresses, it should be evident that attendants would greatly appreciate if they could have an automatic tool for scheduling its time based on her profile and current interests.

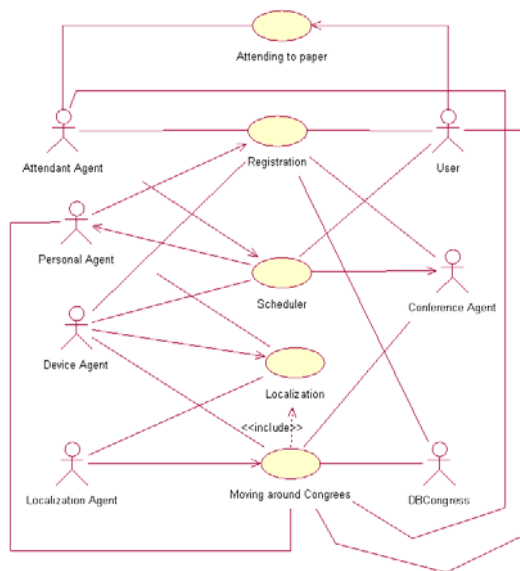
**Moving around congress location.** As stated before, meeting with people, known and unknown, is one of the things congress attendants look forward to do. But how could a congress attendant easily find his peers, specially the ones he doesn't know? One can argue that he could do it at paper presentation rooms. After all, all of the attendants to a particular paper presentation should share interests and profiles. However, during the time span of a paper presentation there may not be enough time to meet all the interesting people attending. Consequently, we think it is important to have a tool that would facilitate a person to locate peers based on her profile and current interest independently of people location. Besides, if the congress includes a commercial exhibition, congress attendants expect to locate just products they care about, as well as exhibitors expect to show their products only to potential clients. On the other hand, in general, congress attendants are not acquainted with the current congress location. Thus, the congress organization must provide proper environment signals and indications to easy the localization of places to congress attendants.

**Attending to paper presentations.** During a paper presentation, attendants take notes and acquire an opinion of the speaker. Wouldn't it be nice if attendants could instantly and quietly share, on their command, notes and opinions? Moreover, this notes and opinions could be shared with congress attendants with similar interests and profiles but not attending to the particular paper presentation.

#### 4. Use cases

Upon the scenarios we identified, we came up with the Use Cases Model shown in Figure 1. This Model is based on the assumption that an autonomous agent middleware will be used. Autonomous agents are technologies from the field of Artificial Intelligence that have been successfully applied in several domains, including ubiquitous computing. For instance, agent technologies have been considered as an alternative for





**Figure 1**—Use cases model

user's PIM; detecting the user state within the infrastructure; superseding the user when the user is off-line; and maintaining communication channels with device agents located at devices near the user's physical location. We envision these personal agents would be hosted by computer hardware of some Ubiquitous Computing Service Provider.

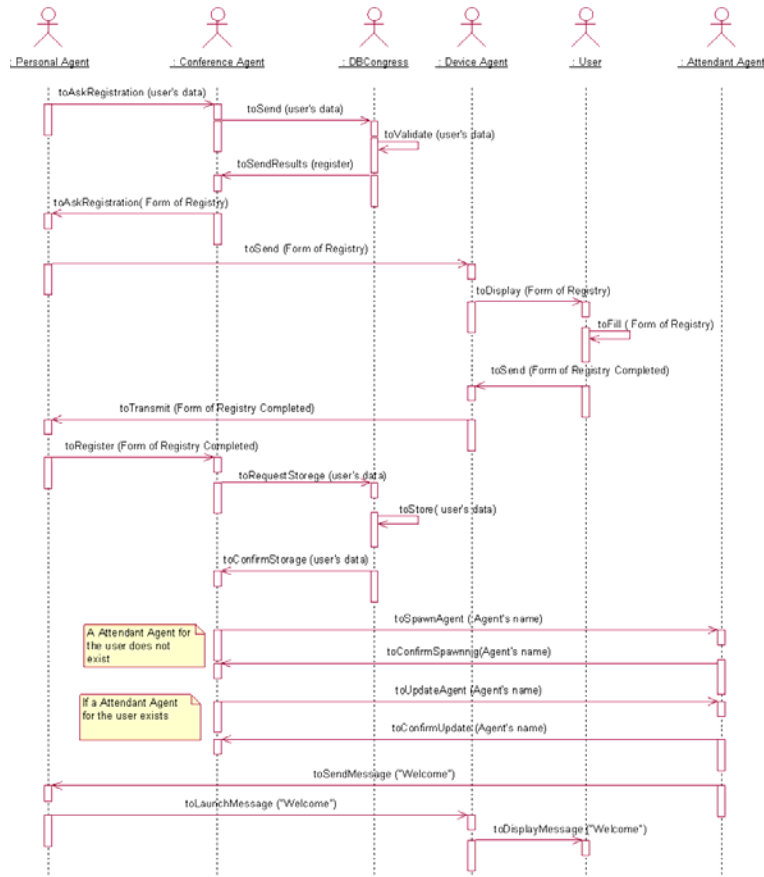
An attendant agent is a congress attendant's software proxy. It implements knowledge and algorithms specially design for a congress environment: congress scheduling, locating peers, and interfaces with the congress agent. It defines and maintains a tight bound with the corresponding personal agent for getting access to the congress attendant's PIM, state and location.

We envision each computing device hosting a device agent implementing knowledge and algorithms specially design for conveying information between any other agent and the input/output hardware. It must implement an ontology for publishing device capabilities so that the information it sends and receives has appropriate formats. This is important for not having to convert data formats locally at the expense of draining computer power and battery lifetime. It also provides secure means for identifying persons and bounding the device with the identified person's personal agent. This scheme supports the paradigm of application on demand, like .NET, where no application program is required to be permanently installed within the device. Applications are accessed through the infrastructure by means of the bound between the device agent and the user's personal agent.

The conference agent is the software component implementing knowledge and algorithms for managing the conference. It holds the knowledge of people registered in advanced, registered congress attendants, and congress services. It implements an interface with personal agents for signing on tasks. After signing on its user, personal agents

service discovery in pervasive computing environments [4]. Other research [5] has proposed the use of agent technology to wrap functionality to obtain context information (such as user's profile, preferences, activity, location and used devices), which influence the way a service is accessed or service results are presented to the user. Other authors [6] use agent technologies to manage and keep mobile user's data in a secure and safe space in a network's fixed node.

A personal agent is a person's software proxy. It must be a generic agent in the sense that we expect there good be millions of these agents within a global ubiquitous computing infrastructure. It is responsible of managing the



**Figure 2**—Sequence diagram for the Registration Use Case

interact with the congress agent by means of an attendant agent. It is responsible of spawning attendant agents.

We envision a network of interacting localization agents within the global ubiquitous computing infrastructure implementing knowledge and algorithms for providing location context to agents, and for localizing persons and devices. One can argue that GPS equipped handheld devices and proper software would be sufficient for implementing these functionalities. However, one thing is for a device to know its place on Earth and another is to provide the user (and the device agent) with timely and detailed information of people and services around him. GPS, or any other hardware mean for determine Earth coordinates, is a first step. Localization agents would translate longitude, latitude and altitude measurements to information like: “Welcome to Auroras Hotel, host of the annual XYZ congress. Here is your site map. You are here.”

## 5. Sequence diagrams

We have design sequence diagrams for all use cases. Due to space constraints, here we only show the one for the Registration Use case—see Figure 2. The user enters the congress building and his device's WLAN networking hardware detects a new signal and thus establishes a new network session, including sessions with local services. Then, the device agent queries the local localization agent and informs the user's personal agent of the new environment, possible including its current GPS readings. Upon knowing the new environment the user's personal agent communicates with the conference agent. The personal agent knows by context that it should communicate with a conference agent; for instance, the user had annotated in its agenda it is attending a conference at its present location, or the local localization agent informs the personal agent there is a conference going on and the latter determines the congress' topic matches its user profile. The user's personal agent negotiates with the conference agent the user registration, including any input required by the user. Upon successful completion, the conference agent spawns a new attendant agent and binds it to the user's personal agent. The first task for the attendant agent is to negotiate with the user's personal agent and the user device's device agent the presentation to the user of a personalized welcome message.

## 6. Concluding remarks

Today's successful histories in information technology research and innovation come from research projects aiming at providing people with solutions to better its life. Our research group is aiming at building social software applications and ubiquitous computing environments that will drive our investigations in software engineering and computer networking.

## References

- [1] Kling, A. "Social Software," Tech Central Station, [www.techcentralstation.com/04103C.html](http://www.techcentralstation.com/04103C.html), 2000.
- [2] Weiser, M. "Some computer science issues in ubiquitous computing," *Communications of the ACM*, Special issue on computer augmented environments, Vol.36, No.7, pp.75-84, July 1993.
- [3] Macías-Estrada, A.; et-al. "An analysis of the communications support of the JADE/LEAP agent platform," submitted 2004.
- [4] Campo C. "Service discovery in pervasive multi-agent systems", In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagents Systems, AAMAS 2002, Bologna Italy (2002)*.
- [5] Laukkanen M.; et al. "Tourists on the move", In *Proceedings of CIA 2002, LNAI 2246*, Springer-Verlag, Berlin Heidelberg New York, 2002, 36-50.
- [6] Villate Y.; et al. "Keep your data safe and available while roaming", *Mobile Networks and Applications, MONET'02, Special issue on pervasive computing, Vol. 7 No. 4, 315-328*, ACM Press, New York, 2002

# Administrador de Candados para Transacciones Anidadas sobre Dispositivos Móviles

Jorge Martínez-Muñoz<sup>1</sup>, Luis A. Gama-Moreno<sup>1,2</sup>,

<sup>1</sup> Laboratorio de Inteligencia Artificial

Centro de Investigación en Computación, IPN.

e-mail : {george, lgama}@correo.cic.ipn.mx

Juan de Dios Bátiz esq. Othón de Mendizábal S/N Uni. Prof. Adolfo López Mateos,  
CP 07738, México, D.F., México.

<sup>2</sup> Instituto Tecnológico de Zacatepec

Calzada Tecnológico S/N, Zacatepec, Mor. C.P. 62780, México

**Abstract:** Transaction processing is a Distributed Systems branch aimed to study and preserve data consistency. Beyond flat models –where objects are stored at the same host, there is the distributed approach enclosing Nested Transactions. Executing these latter over mobile devices instead of fixed hosts, implies the necessity for new models capable of dealing with mobile computing traits. In this work, we introduce a lock manager supporting concurrency control, the transaction processing component that manages multiple access requests over shared resources.

**Resumen:** El procesamiento de transacciones es una rama de Sistemas Distribuidos orientada al estudio y preservación de la consistencia de datos. Más allá de los modelos tradicionales –donde los objetos están en un mismo servidor, se encuentran los modelos distribuidos al que pertenecen las transacciones anidadas. La ejecución de estas últimas sobre dispositivos móviles en lugar de estaciones de trabajo fijas, conlleva la necesidad de nuevos modelos que se adapten a las características del cómputo móvil. Este trabajo presenta un administrador de candados como soporte el control de concurrencia, elemento que permite coordinar la ejecución simultánea de múltiples peticiones de acceso para un recurso compartido.

**Palabras Clave:** Transacción anidada, Administrador de Candados, Control de Concurrencia.

## 1. Introducción.

El concepto de transacción puede ser encontrado en casi cualquier tipo de procesos de negocios. En [2] se presenta una definición de transacción la cual es un mecanismo para preservar la consistencia de un grupo de objetos en proceso. En [1] se introduce una definición formal de transacción como un conjunto de operaciones de lectura/escritura definidas bajo un orden parcial. Los modelos de transacciones están enfocados en la preservación de las propiedades ACID: atomicidad, consistencia, aislamiento y durabilidad [5]. Sin embargo, en recientes trabajos [3] se ha propuesto la

relajación de los niveles de aislamiento, con el objetivo de incrementar el número transacciones concurrentes a la vez que se mantiene la consistencia de los datos.

En el presente trabajo, se describe un administrador de candados orientado a resolver conflictos por acceso concurrente en un grupo de dispositivo móviles que ejecuta transacciones anidadas. Este mecanismo introduce una noción de persistencia en el sentido que transacciones fallidas, realizarán nuevos intentos dentro de restricciones de tiempo real para completar sus operaciones, sin dejar de lado la condición de que existen transacciones adicionales ejecutándose de manera simultánea. El mecanismo de control forma parte de un monitor para transacciones anidadas actualmente en desarrollo.

## 2. Estado del arte.

### 2.1 Transacciones Anidadas.

En [9] se introduce el concepto de transacciones anidadas (TA). En dicho modelo, una transacción puede contener cierto número de sub-transacciones y cada sub-transacción, a su vez, puede contener sub-transacciones. La transacción completa conforma una estructura arbórea, donde el nodo en el tope del árbol es llamada raíz [4][10]. La principal ventaja en TA es que la transacción raíz es capaz de validar con el trabajo confirmado de algunas de sus sub-transacciones aunque otras presenten fallas. El abandono de una sub-transacción en cualquier nivel de la TA, causará que todas las sub-transacciones hijas abandonen, esta es la razón del por qué las sub-transacciones hijas sólo tienen las propiedades Atomicidad, Consistencia y Aislamiento pero no la de Durabilidad [4]. Únicamente la transacción raíz puede cumplir con las propiedades ACID. La Figura 1.a. muestra un ejemplo de una transacción anidada.

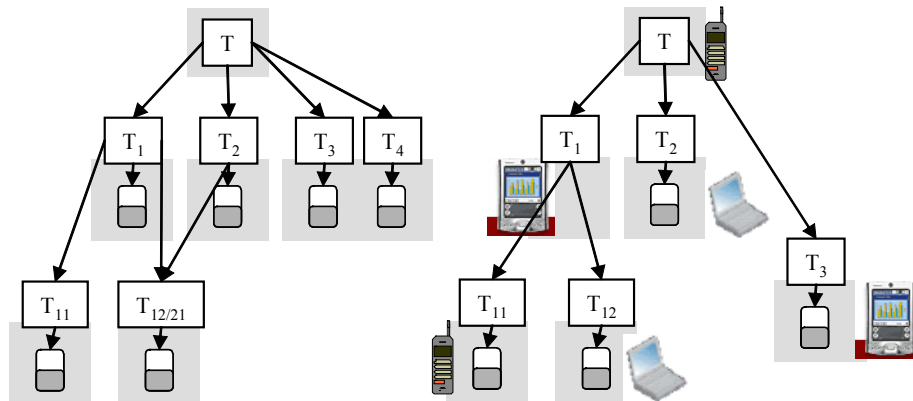


Fig. 1 (a) Transacción Anidada.

(b). Transacción Anidada Móvil.

El modelo TA es adecuado para aplicaciones que involucran aplicaciones móviles ya sea mediante conexiones celulares o inalámbricas [4]. Por ejemplo, una transacción formada por sub-transacciones, ofrece mayor rendimiento y tolerancia a fallas. Las sub-transacciones pueden ser ejecutadas en paralelo a la vez que mantienen control local de su confirmación o cancelación.

### **2.3 Transacciones Anidadas Móviles.**

En [3] se introduce el modelo de Transacciones Anidadas Móviles (TAM). Éste es una combinación de los conceptos de TA y computación móvil (CM), en particular enfocado a un grupo de dispositivos (*hosts*) móviles (HM). En [9], se menciona que únicamente los nodos en el nivel inferior (hojas), pueden acceder a objetos de la base de datos, mientras que la definición en [5], permite también a los nodos intermedios ejecutar operaciones de lectura/escritura sobre datos. Esta es la diferencia principal que puede ser apreciada en la Fig. 1.b.

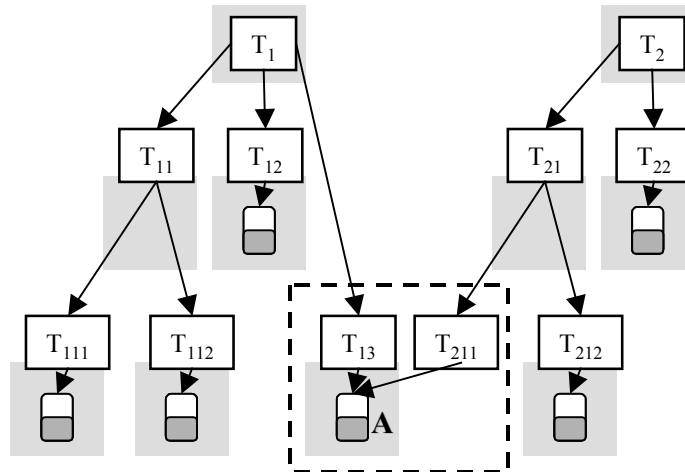
### **2.4 Trabajos relacionados.**

En [7] se presenta un modelo de simulación para medir el impacto de ciertas características en las transacciones tales como número de hojas y niveles de profundidad con respecto al desempeño del sistema. En [8] se introduce el concepto de lista de requerimientos, con el objetivo de proveer tiempos de ejecución cortos y un mejor rendimiento del sistema basándose en intra-transacciones. En [6] se proponen mecanismos de cooperación y comunicación para explotar el paralelismo en intra-transacciones, y se introduce el concepto “Herencia descendente de bloqueos”. En [11] se proponen soluciones a tópicos de control de concurrencia y recuperación basadas en agentes.

## **3. Control de concurrencia**

### **3.1 Control de concurrencia basado en candados.**

El mecanismo de control que presentamos, está orientado a resolver conflictos que se originan si dos o más nodos de alguna transacción intentan acceder al mismo objeto. Se tienen presentes las condiciones que ponen en riesgo la operación de un dispositivo móvil. Por ello, el control de concurrencia debe estar diseñado para administrar eficientemente recursos compartidos bajo restricciones de tiempo. El mecanismo, permite además tratar con posibles desconexiones y baterías agotadas. El desempeño global, resultado de aplicar el control de concurrencia basado en candados, resulta de la ejecución de operaciones simples implementadas localmente por cada nodo. Estas son activadas durante la interacción del sistema en el procesamiento de la transacción.



**Fig. 2.** Ejemplo de acceso concurrente.  $T_{13}$  se encuentra usando A cuando  $T_{211}$  intenta acceder el mismo objeto. El Administrador de Candados permite resolver el conflicto por medio de algunas sencillas reglas.

El mecanismo propuesto se ilustra en la Figura 2. Donde, la sub-transacción  $T_{13}$  de  $T_1$  ya ha bloqueado un objeto de la base de datos, posteriormente una segunda sub-transacción  $T_{211}$  de  $T_2$  intenta acceder al mismo objeto.

El Administrador de Candados cuenta con una tabla donde se almacena información sobre los objetos en uso y los candados correspondientes en el dispositivo anfitrión. Los candados se mantienen durante el tiempo que requiera la transacción global para concluir y son inmediatamente liberados cuando se recibe el *commit* o *abort* final desde el Nodo Raíz.

Supongamos que existe una petición de candado compartido para acceso de lectura por parte de  $T_{13}$ , el Administrador registrará qué transacción y objetos se encuentran ligados durante el tiempo necesario hasta que se complete la transacción. Si  $T_{211}$  quiere leer el mismo objeto, el Administrador le concederá acceso compartido dado que no hay conflictos entre ambas peticiones.

Los problemas surgen cuando al menos una de las transacciones requiere acceso exclusivo dado que necesita escribir sobre los datos. Bajo tales condiciones, se introducen las siguientes reglas para tratar con peticiones de acceso que entran en conflicto por su naturaleza.

Para ilustrar el ejemplo arriba mencionado, supongamos que en la Figura 4, el Administrador del nodo  $T_{13}$  ya ha registrado a la transacción para acceso exclusivo sobre A. Ahora, si una segunda transacción  $T_{211}$  llega con la intención de usar A, ésta recibirá una notificación de rechazo por lo que  $T_{211}$  entrará en un estado de espera.

A continuación, dos eventos pueden ocurrir:  $T_{13}$  libera el objeto A y  $T_{211}$  es notificado para que continúe con sus operaciones asignadas por su padre  $T_{21}$ ; o bien,  $T_{21}$  alcanza el fin de lapso después del cual verifica si  $T_{211}$  pudo completar sus operaciones o no, pues esta información es requerida para que  $T_{21}$  la envíe hacia su respectivo superior. Dado que  $T_{211}$  no tuvo éxito en la ejecución de sus operaciones, la sub-transacción será abortada y  $T_{21}$  tomará una decisión sobre si aún es posible ejecutar

*commit* a pesar de la falla de  $T_{211}$ . Esta decisión se realiza considerando la clasificación de acciones y operaciones que se introduce en la siguiente sub-sección.

La información de todas las sub-transacciones, eventualmente llega al Nodo Raíz  $T_2$ . Si el nodo fallido ( $T_{211}$ ) desencadenó un *abort* global debido a su prioridad, entonces el coordinador raíz debe programar un nuevo intento para completar las tareas inconclusas. Antes de ejecutar la siguiente transacción, debe revisarse la configuración definida por el usuario. De esta última depende si la ejecución se realiza de manera automática o habrá necesidad de autorización por parte del usuario.

### 3.2 Toma de decisiones en solución de conflictos por concurrencia.

Durante el procesamiento de la transacción, se pueden encontrar objetos de la base de datos que ya han sido bloqueados por otros clientes. Para resolver este tipo de conflictos, se propone el siguiente mecanismo. Se trata de una clasificación de acciones y operaciones en: obligatorias y opcionales, fuertes y débiles. La combinación de estas características y la decisión que aplica cuando alguna de las operaciones pertenecientes a una sub-transacción no puede ser ejecutada exitosamente se muestran en la Tabla 1.

Tipo	Subtipo	Descripción
Obligatoria	Fuerte	Se aborta la transacción.
Obligatoria	Débil	Se realiza un nuevo intento inmediatamente, a menos que el usuario cancele la operación.
Opcional	Fuerte	Se ignoran las operaciones fallidas para esta transacción, pero se programa una transacción posterior a menos que el usuario cancele la operación.
Opcional	Débil	No se realizan intentos adicionales.

**Tabla 1.** Clasificación de acciones para dar soporte al control de concurrencia en el procesamiento de transacciones anidadas. La columna de descripción muestra qué debe realizarse en caso de que las operaciones y sub-transacciones fallen.

La clasificación anterior auxilia en la toma de decisiones de los nodos de la transacción. Esto puede ocurrir cuando se intenta acceder a un objeto que ya ha sido asegurado por otro nodo, e incluso cuando se pierde la comunicación con alguno de los participantes. El nodo padre puede seleccionar una vía alterna para operar con un conjunto de descendientes distinto al original. Si la operación fallida es obligatoria-débil u opcional-fuerte, se puede realizar un nuevo intento, ya sea dentro de la misma transacción o programando una transacción futura. Aquí se introduce la noción de persistencia en el sentido de que se realizarán nuevos intentos para completar transacciones fallidas, siempre que esto sea posible y así lo desee el usuario. El mecanismo propuesto opera como auxiliar para el control de bitácoras de operación y recuperación después de una falla.



## 4. Conclusiones

En este artículo se describieron modelos para mejorar el rendimiento y tolerancia a fallas de aplicaciones que procesan transacciones en dispositivos móviles. El modelo de transacciones anidadas en particular, permite la división de una transacción en sub-transacciones incrementando de esta forma, la concurrencia en las aplicaciones. Se introdujo un administrador de candados para la coordinación de transacciones que intentan acceder a objetos previamente bloqueados con acceso exclusivo. Con ello se busca disminuir los efectos de “abrazos mortales”. Finalmente, se presentaron políticas para la solución de conflictos dentro del árbol de transacciones clasificando las operaciones en obligatorias u opcionales.

## Referencias

- [1] P. E. Bertino, B. Catania, A. Vinai. “Transaction Models And Architectures. Encyclopedia of Computer Science and Technology”, Vol.38, Marcel Dekker, 1998, pp. 316-400.
- [2] G. Coulouris, J. Dollimore, T. Kindberg T. “Distributed Systems”. Addison Wesley, 2002.
- [3] L. A. Gama, M. Alvarado. “Concurrency control for Read-Only in Mobile Nested Transactions”. Proc. of Intelligent Computing in the Petroleum Industry ICPI’03, associated with IJCAI’03, 2003.
- [4] L. A. Gama, M. Alvarado. “Mobile Nested Transactions for Nomadic Teams”. Special Issue on Intelligent Computing for Petroleum Industry. Expert Systems with Applications, Vol. 26-4. Elsevier, 2004.
- [5] J. Gray, A. Reuter. “Transaction Processing: Concepts and Techniques”. Morgan Kaufmann Publishers, Inc., 1993.
- [6] T. Härder, K. Rothermel. “Concurrency Control Issues in Nested Transactions”. The VLDB Journal - The International Journal on Very Large Data Bases, Vol. 2, Issue 1, Springer-Verlag New York, Inc., January 1993, pp. 39-74.
- [7] H. S. Hassanein, M. E. El-Sharkawi. “Performance Modeling of Nested Transactions in Database Systems. Proc. of the 2000 conference of the Centre for Advanced Studies on Collaborative research, IBM Press, 2000, p. 4.
- [8] M.L. Lo, C.V. Ravishankar. “A Concurrency Control Protocol for Nested Transactions”. IBM Centre for Advanced Studies Conference. Proc. of the 1992 conference of the Centre for Advanced Studies on Collaborative research, Vol.2, IBM Press, 1992, pp. 67-80.
- [9] J.E.B. Moss. “Nested Transactions: An Approach to Reliable Distributed Computing”. MIT Press, Cambridge, MA, 1985.
- [10] J. Martinez, M. Alvarado. “Mobile Nested Transaction Monitor based on MultiAgent Systems: Workflow Layer and JADE ontologies”. Proc. of Workshop on Intelligent Computing ’04, associated with MICA I ’04, 2004.
- [11] E. Pitoura, B. Bhargava. “A Framework for Providing Consistent and Recoverable Agent-Based Access to Heterogeneous Mobile Databases”. ACM SIGMOD Record, Vol. 24, Issue 3, ACM Press, September 1995, pp. 44-49.

# Autonomous Agents for Supporting Mobility in AmI Environments

Marcela Rodríguez Urrea<sup>1,2</sup>, Jesús Favela<sup>1</sup>, Edgar A. Martínez<sup>1</sup>

<sup>1</sup>Centro de Investigación Científica y de Educación Superior de Ensenada  
{marcerod,favela,edgar}@cicese.mx

<sup>2</sup>Universidad Autónoma de Baja California

**Abstract.** In this paper, we present how we use autonomous agents as a useful abstraction for the design and development of an Ambient Intelligence medical environment. Hospitals are characterized by the high mobility of their medical staff which need to opportunistically access information and services. Our system includes a handheld computer with an agent wrapping a trained backpropagation neural-network that estimates the user's location, and a client to access information from the Hospital Information System that is relevant to the user's current location. We illustrate how we implemented the location-estimation agent by using the SALSA middleware which enables developers to implement systems that support ambient intelligence.

**Keywords.** Ambient intelligence, autonomous agents, location-aware services

## 1 Introduction

Ambient Intelligence (AmI) is the vision of smart spaces in which computational artifacts become invisible, present whenever we need it, adaptive to mobile users and context, enabled by simple and effortless interactions, and autonomously acting to support users' activities and goals [3]. For this, AmI systems need intelligent capabilities to be reactive to context and learn from users' behavior in order to provide high quality services based on their preferences. For this, AI techniques such as learning algorithms, pattern matchers and agent technology, may be used in the implementation of these intelligent systems. We realized that AmI environments are characterized for the autonomy, reactivity, distribution, collaboration and adaptation of its artifacts, and in this sense, they share the same characteristics as agents. Agents have been considered as a technological alternative to cope with some of the challenges in implementing smart spaces, such as problems related with the discovery of services, security and adaptation of information. However, the use of autonomous agents has not been explored as an abstraction tool for the design and construction of ambient intelligence environments. To determine if autonomous agents are the appropriate metaphor for implementing intelligent environments, we designed and built a handheld-based system to support the location-aware access to medical information in hospitals. In hospital environments, medical workers are highly mobile. They need to move to locate other colleagues, patients and resources, such as medical equipment. Additionally, the information required by these specialists depends of their location. It has been considered that mobility is a key concept that deserves special attention when

designing AmI spaces [3]. In this paper, we describe how we use autonomous agents to address mobility issues, in particular, the estimation of the user's location. We illustrate how the location-aware hospital information system was built by using the SALSA agent middleware, which is flexible enough to enable the implementation and evolution of AmI applications [4][5].

In the next Section, we illustrate with a scenario our vision of AmI in hospitals, in which mobile users opportunistically access information and services. Section 3 describes recent efforts aimed at estimating users' location, and Section 4 describes our approach based on a wireless LAN and a back-propagation neural network which is wrapped in an autonomous agent. Section 5 illustrates the implementation of the location-estimation agent, and its integration of the location-aware hospital system. Finally, Section 6 presents our conclusions.

## 2 AmI in Healthcare Environments

The AmI vision was shaped through a set of scenarios in which people are at the forefront of the Information Society [3]. Following this approach, we have identified scenarios with the purpose of discovering how AmI systems can support the activities of medical staff in their working environment [4]. Next, we present one of the scenarios that guided our design of the location-aware hospital information system CHIS system.

*While Dr. Diaz is checking the status of a patient (bed 222), he realizes that he should request a laboratory study. Using his handheld, he makes this request through the patient's electronic clinical record. When the chemist responsible for taking samples for the analysis, visits the internal medicine area, the application in his handheld realizes he is in a different hospital floor and informs him that in bed 222 a patient requires laboratory analysis. When the chemist stands in front of the patient, his handheld lists the samples he must take and the type of analysis to be performed. Once he performs these analyses, he adds the results to the patient's clinical record. Afterwards, when Dr. Diaz visits the patient on his next round, the lab results will be displayed on his handheld. On the basis of these results, he re-evaluates the patient and decides to fill a medical note requesting the nurse in charge to increase the doses of the patient's medication.*

In this scenario, the system is continuously estimating the location of Dr. Diaz and communicating this information to the Hospital Information System, which updates the physician's location on all other users' handhelds. After a few seconds of Dr. Diaz being near bed 222 the medical record of this patient is displayed on his PDA. In the case of the chemist, when he moves to the floor where the patients are located, his handheld updates his location, and then, provides him with information of the patients pending laboratory analysis on that floor. Thus, the system adapts to the context, and in particular to the location of its user. Undoubtedly, location is important to understand the context of mobile users. Location becomes a useful indexing information from which to infer the overall context that a system will use to provide services and information to mobile users.

### 3 Current Approaches to Location-estimation

Estimating the location of a user has been a subject of considerable attention in recent years. Advances in Global Positioning Systems (GPS) allow mobile computers to determine its location with an approximate accuracy of 10 meters by using triangulation of the signals received from multiple satellites. Passive RFID tags have become increasingly popular to track and identify products and people. With a read distance generally limited to three feet, a high concentration of tags would be required to continuously track the location of people. There are also active RFID tags with transmission distances of over 100 feet that use a battery to power the chip's circuitry and broadcast a signal to a reader. This solution however, requires the installation of a dedicated infrastructure besides the computer network as is the case with most of the solutions discussed above.

Of particular interest are location estimation techniques that make use of an existing wireless LAN infrastructure, since they have better scalability and less installation and maintenance costs than ad-hoc solutions. These methods use the RF signal strength between a mobile device and several access points of the wireless LAN, to estimate location. The RADAR location system uses an IEEE 802.11 wireless LAN and an empirical method based on the nearest neighbor algorithm [1]. Similarly, the Nibble system uses the signal-to-noise ratio, which is more stable than signal strength to compute the distance to the access point. Nibble uses a Bayesian network to estimate the probability of the mobile being at one of a set of discrete locations [2].

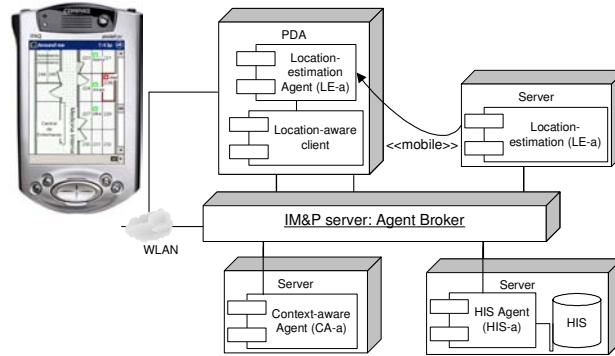
### 4 Neural Networks for Location Estimation

We are using backpropagation neural networks trained to map RF signals from a WLAN to 2D coordinates [5]. Once trained, the neural network can be used to classify incoming patterns into labeled classes. The best results of the location estimation were obtained with an output layer of 2 neurons corresponding to the X and Y coordinates we want to estimate, 5 neurons in the input layer and 16 neurons in the hidden layer. We use the signal to noise ratio as input, and a variation of the backpropagation algorithm (conjugate gradient with Polak-Ribière updates) which fulfills several conditions that guarantee that an optimum learning coefficient was found. With this configuration, we obtained an average error of 2.0947 meters and we observed that 80% of the output patterns are within 3 meters of their target. More details of the experimental setup and results are presented in [5].

### 5 Integrating the Location-estimation Agent to the Location-aware Hospital System

As illustrated in Figure 1, the main components of the location-aware system are autonomous agents implemented with the SALSA middleware [4]. These agents communicate through an Agent Broker, which instantiates an Instant Messaging &

Presence Server. In the handheld computer resides the Location-aware Client. It was conceived as an autonomous agent that notifies the user's location to other users and agents, provides mobile users with information relevant to their location, and allows them to fill requests and communicate with other members of the staff. Its interface is based on the Instant Messaging (IM) paradigm, through which users are notified of the availability of other users and their location. In the handheld also resides the Location-estimation Agent (LE-a) which obtains the user's position (X,Y coordinates), and informs it to the Location-aware Client. The Hospital Information System Agent (HIS-a) acts as proxy of the HIS, which manages and stores the patient's clinical records and other data relevant to the hospital. This agent provides access to information contained in the HIS, and monitors its changes. Finally, the Context-aware agent (Ca-a) delivers the messages that depend on contextual variables for their delivery.



**Fig. 1** Architecture of the Location-aware hospital system.

As mentioned above the location estimation component is a SALSA agent that wraps in its reasoning module a trained neural network to estimate the user's location. Figure 2a illustrates the LE-a class diagram by using the AUML (Agent Unified Modeling Language) notation. It specifies the fields that define the *state* of the agent, which are the information perceived (SNR- Signal to Noise Ratio) and the output information (coordinates X,Y). To generate this output, the agent has the *capability* of estimating the user's position, and then it has to *act* by notifying the user's position to the location-aware client. To do this, the LE-a uses the *communicative acts* of SALSA. The agent may also decide to update his reasoning component when it detects that the name of the Access Points has changed

Figure 2b shows how the components of the location-estimation agent (LE-a) interact to estimate the user's location. These components are instances of SALSA classes. The agent's perception module receives the SNR (Signal to Noise Ratio) through the *PassiveEntityToPerceive* object, which represents the memory of the wireless network card. The developer implemented this interface to read data from the WLAN card, wrap the data in an *Input* object, and then, notify it to the *PassivePerception* component of the LE-a. When the SNR value is changed, the *PassivePerception* object generates a *StateChangeEvent* which is captured by the reasoning component. Then, the *Reasoning* component makes a new estimation of the user's location based on the *Input* received, and invokes to the *Acting* component. It communicates this information to the location-aware application, which is also an

agent residing on the user's PDA. Since the signal strength decays considerably from floor to floor, one estimation agent is trained for each hospital floor. When the strength of the signal from the access points in one floor goes below certain threshold, the LE-a asks the AgentDirectory to request information of the server in which the LE-a that holds the reasoning component that estimate his location in that floor resides. Thus, by using the mobility attribute of agents, the LE-a on the PDA updates its reasoning component.

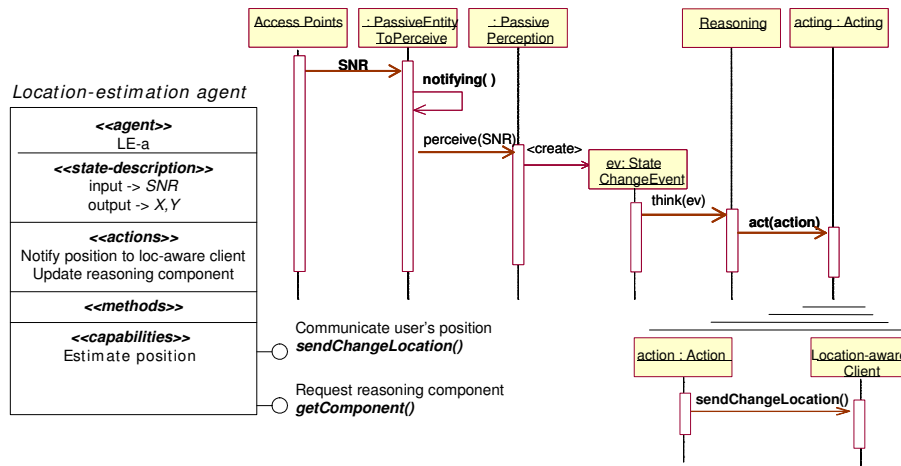


Fig. 2 AUML class diagram and sequence diagram of the location-estimation agent

The location-aware application receives the user's X,Y coordinates through its *PassivePerception* object. This information will be communicated through the *StateChangeEvent*. Then the reasoning component maps the X,Y coordinates to an area identifier (bed number, room, etc.) and the acting component communicates this new location to the rest of the system's agents and users through its instant messaging client.

In Figure 3(a) is depicted the implementation of the reasoning component of the LE-a by using the SALSA classes. For this, the developer created the class *ReasoningLE\_a* by extending the *Reasoning* class of SALSA. As it is an abstract class, its *think* method was overwritten to process the input perceived and then, to indicate the agent what action should execute. For instance, if the SALSA event created by the *PassivePerception* component was a *stateChangeEvent*, it indicates that a new *input* has to be processed by the *estimateLocation()* method which implements the trained Neural Network. Then the acting component is invoked to execute the *communicateNewLocation\_Action* object, which implements the SALSA Action class, to pass the new user's location to the *PassivePerception* of the Location-aware client. Once the developer implemented the Perception, Reasoning, and Action components of the agent, he has to create the LE-a as illustrated in Figure 3(b). The agent was created by extending the SALSA Agent class and by invoking its *create* method which will automatically produce the components of this agent.

## 6 Conclusions

We used autonomous agents as a technique to model and design the system. Thus, the main components of the location-aware hospital system were identified as agents that respond autonomously in accordance with the context surrounding the activities performed at the hospital. To implement the system, we used the agent SALSA middleware which enables developers to easily add new functionality to AmI systems, by allowing the integration of components represented by autonomous agents. We are currently evaluating the core features of SALSA by implementing several AmI applications. Among these applications, we are integrating to this system Wall-Sized Displays to support the access of semi-public medical information that may not be adequately displayed on the users' PDA. We will need to enhance the location-estimation mechanism in order to predict if the user wants to use de Public Display.

```
import SALSA.*;
public class ReasoningLE-a extends SALSA.Reasoning{
    public ReasoningLE-a(SALSA.Agent a){
        super(a);
    }

    public void think(EventObject ev){
        SALSA.Events.Event event = (SALSA.Events.Event) ev;
        //if the SNR changed
        if (event.getType() == event.stateChangeEvent) {
            coordinates = estimatesLocation(ev.input);
            //Invokes the action component
            agent.acting.act(new communicateNewLocation_Action());
        }
        //if the new neural network for that floor was obtained
        else if (event.getType() == event.ArriveComponentEvent){
            agent.acting.act(new integrateComponent_Action());
        }
        . . . .
    }
}
```

(a)

```
import SALSA.*;
public class IE_a extends Agent{
    ReasoningLE_a rsn;

    public IE_a(){
        //Specify the reasoning component
        rsr=new ReasoningLE_a(this);
        . . .
        //Create the components of the agent
        this.create(rsn);
    }

    public static void main(String[] args) {
        //Creates an instance of the IE-a
        IE-a agent=new IE-a();
    }
}
```

(b)

**Fig. 3** (a) Implementation of the LE-a reasoning component and (b) the creation of the LE-a

## References

- [1] P. Bahl and V.N. Padmanabhan. "RADAR: An in-building RF-based location and tracking system" in IEEE INFOCOM, 2000, Tel-Aviv, Israel.
- [2] P. Castro, P. Chiu, T. Kremenek, and R. Muntz. "A Probabilistic Room Location Service for Wireless Networked Environments", In Proc. of Ubicomp Conference, Springer-Verlag, 2001, pp.18-34.
- [3] M. Lindwer, D. Marculescu, T. Basten, R. Zimmermann, R. Marculescu, S. Jung, E. Cantatore. "Ambient Intelligence Visions and Achievements: Linking Abstract Ideas to Real-World Concepts" In Proc. of Design, Automation and Test in Europe Conf., 2003.
- [4] M. A. Muñoz, M. Rodriguez, J. Favela, V.M. Gonzalez, and A.I. Martinez-Garcia. "Context-aware mobile communication in hospitals," IEEE Computer, Vol. 36, 2003, pp. 38-46.
- [5] M. Rodriguez, Jesus Favela, Miguel A. Muñoz, and Edgar A. Martínez. "Location-aware Access to Hospital Information and Services," Submitted to IEEE Trans. on Information Technology in Biomedicine, 2004.

# **Bioinformática y biología computacional**

Jesús Antonio González Bernal  
Mauricio Osorio Galindo  
Javier Garcés Eisele  
(Editores)





## Introducción

Bioinformática y biología computacional son áreas de investigación interdisciplinarias encontrándose en la intersección entre ciencias naturales, ciencias computacionales y tecnologías de la información. La bioinformática aplica principios de ciencias y tecnologías de la información para que los datos abundantes, diversos y complejos del área de la biología se vuelvan más fáciles de comprender y más útiles. La biología computacional desde el punto de vista matemático y computacional analiza preguntas teóricas y experimentales de la biología. Si bien la bioinformática y la biología computacional son áreas separadas, existe también una intersección significativa entre ambas. De acuerdo al NIH *Biomedical Information Science and Technology Initiative Consortium* se pueden definir estas áreas como sigue:

- **Bioinformática:** Investigación, desarrollo o aplicación de herramientas y métodos computacionales para expandir el uso de datos biológicos, médicos y del área de la salud. Esto incluye aquellos para adquirir, almacenar, organizar, archivar, analizar o visualizar estos datos.
- **Biología Computacional:** El desarrollo y la aplicación de métodos teóricos y técnicas de análisis de datos, modelado matemático y simulación computacional para el estudio de sistemas biológicos y sociales, de su estructura y su comportamiento.

La investigación en estas áreas ha mostrado un gran crecimiento en los últimos años, este crecimiento se debe principalmente a los grandes volúmenes de datos biológicos que se han recopilado y a la necesidad de analizarlos. El interés hacia esta área ha motivado a investigadores de diferentes instituciones a trabajar en alguno de los diversos temas derivados de la bioinformática. El objetivo de este taller es dar oportunidad a investigadores mexicanos de reunirse a discutir temas de interés para formar lazos de colaboración cuando sea posible y al mismo tiempo fortalecer la investigación en el área de bioinformática en México.

En esta primera edición del taller de Bioinformática y Biología Computacional se recibió un total de 8 artículos, de los cuales fueron aceptados 5 después de un proceso de evaluación llevado a cabo por el comité revisor del taller. Los trabajos recibidos fueron elaborados por investigadores mexicanos, abarcando diversas temáticas que podrán ser expuestas durante las sesiones de trabajo del taller. Con esto se busca el intercambio de ideas para fortalecer la investigación en esta área.

## Comité Científico

Copresidentes:

Jesús Antonio González Bernal, I.N.A.O.E., Puebla, [jagonzalez@inaoep.mx](mailto:jagonzalez@inaoep.mx)

Mauricio Osorio Galindo, U.D.L.A.P., Puebla, [josorio@mail.udlap.mx](mailto:josorio@mail.udlap.mx)

Javier Garcés Eisele, U.D.L.A.P. / Lab. Clínicos de Puebla, [jgarces@mail.udlap.mx](mailto:jgarces@mail.udlap.mx)

Carolina Yolanda Castañeda Roldán, U.D.L.A.P., [ccastane@mail.udlap.mx](mailto:ccastane@mail.udlap.mx)

Nicandro Cruz Ramírez, Lab. Nac. de Informática Avanz., Xalapa, [ncruz@lania.mx](mailto:ncruz@lania.mx)

Olac Fuentes Chávez, I.N.A.O.E., [fuentes@inaoep.mx](mailto:fuentes@inaoep.mx)

Ma. del Pilar Gómez Gil, U.D.L.A.P., [pgomez@mail.udlap.mx](mailto:pgomez@mail.udlap.mx)

Iván Olmos Pineda, I.N.A.O.E., [iolmos@inaoep.mx](mailto:iolmos@inaoep.mx)

Geneveva Vargas Solar, CNRS, Francia, Laboratorio Logiciels Systèmes Réseaux /  
investigadora invitada, CENTIA-UDLAP,

[geneveva@mail.udlap.mx](mailto:geneveva@mail.udlap.mx), [Geneveva.Vargas@imag.fr](mailto:Geneveva.Vargas@imag.fr)

# Clasificación No Sesgada de Proteínas con Redes Neuronales y Computación Evolutiva

Angel Kuri<sup>1</sup>, Martha Ortiz<sup>2</sup>, Lucía Castellanos<sup>1</sup>, Carlos Ponce<sup>1</sup>, Christopher Gasca<sup>1</sup>, Oscar Herrera<sup>3</sup>

<sup>1</sup>Instituto Tecnológico Autónomo de México

<sup>2</sup>Universidad Autónoma Metropolitana

<sup>3</sup>Instituto Politécnico Nacional

**Abstract.** The automated classification of proteins has been subject of interest for the past years. In all the attempts reported in the literature we have found that the criteria which determine the position of a given protein in the purported classification is a function of subjective considerations. In this work we report a methodology which takes advantage of the state-of-the-art tools of soft computing to attempt non-biased classification of proteins in *E. coli* and *S. cerevisiae* but which may be applied to, virtually, any organism. The method is based on an adaptive metric following the so-called self organizing maps and the search for patterns in the sequences of aminoacids of the proteins applying genetic algorithms.

**Keywords.** Protein classification, aminoacids, clustering, pattern identification, soft computing.

**Resumen.** La clasificación automática de proteínas ha sido un tema de interés en los últimos años. En todos los enfoques reportados en la literatura hemos encontrado que los criterios que determinan la posición de una proteína dada en la clasificación propuesta es función de consideraciones subjetivas. En este trabajo reportamos una metodología que aprovecha las herramientas de punta de la computación suave para lograr la clasificación no sesgada de proteínas en *E. coli* y *S. cerevisiae* pero que puede ser aplicada, virtualmente, a cualquier organismo. El método se basa en una métrica adaptativa nacida de la aplicación de los llamados “mapas auto-organizados” y en la búsqueda de patrones en las secuencias de aminoácidos (AA) de las proteínas usando algoritmos genéticos.

**Palabras clave.** Clasificación de proteínas, aminoácidos, agrupamiento, identificación de patrones, computación suave.

## Introducción

En el ámbito de la Biología Molecular (BM), uno de los problemas relevantes es la adecuada clasificación de proteínas. El desarrollo reciente de la BM ha permitido generar enormes volúmenes de información cuyo tratamiento y explotación constituyen un problema de alta complejidad. Actualmente existen más de 150 genomas secuenciados y este número continúa creciendo. Aplicando diversos métodos de estadística clásica y aproximaciones apriorísticas, se han propuesto diversas

clasificaciones de familias de proteínas, identificación de regiones conservadas entre grupos (*signatures*) y modelos de inferencia de estructura tridimensional, entre otros. Se usan por ejemplo árboles probabilísticos de sufijos para predecir la probabilidad de que una proteína pertenezca a una familia, basada en una o más regiones significativas conservadas [1]. Se utilizan algoritmos tradicionales de clasificación, como en [2], donde se presenta un mapeo de proteínas a un nuevo espacio de representación que captura las relaciones secuenciales entre los diferentes AAs, consideradas como las características significativas y posteriormente se aplica el algoritmo *k-means* para hacer el agrupamiento. En [3] se construye una matriz de similitudes entre pares de proteínas usando el algoritmo heurístico BLAST [4] y posteriormente se aplica el algoritmo de *clustering* de Markov (MCL) para encontrar las familias de proteínas. Otra estrategia para hacer clasificación de proteínas ha sido el uso de redes neuronales. Como se sabe, existen diversos modelos que se aplican en la solución de problemas de clasificación supervisada, como las redes de perceptrones y las máquinas de soporte vectorial [5]; o bien, en clasificación no supervisada, como en el caso de los mapas autoorganizados [6, 7]. En este sentido, también se han desarrollado diversos trabajos que aplican redes neuronales bayesianas [8], redes neuronales en cascada [9], redes neuronales jerárquicas [10] y redes neuronales recurrentes [11], para clasificar proteínas. Sin embargo, es importante subrayar que todos los trabajos mencionados se encuentran limitados porque las características y propiedades de las proteínas que consideran, están determinadas a priori por los expertos en biología molecular. Es decir, el investigador propone una hipótesis que debe validar. Obviamente, la variedad y complejidad de las hipótesis están limitadas por las relaciones funcionales que el investigador es capaz de visualizar. Una alternativa a estas metodologías clásicas es aquella (derivada de la computación suave) en la que las hipótesis no dependan del proponente, sino que sean “generadas” de manera automática. Por tanto, el objetivo de este trabajo es plantear una metodología, utilizando las herramientas de punta de la computación suave, para lograr una clasificación no sesgada de proteínas en *E. coli* y *S. cerevisiae*.

## Metodología

El problema de clasificación no sesgada basada en la expresión de las proteínas en AAs puede reducirse, conceptualmente, a lo siguiente: a) La identificación de grupos de proteínas que compartan características comunes y b) La determinación de las razones estructurales por las cuales las proteínas en cuestión se agrupan de la manera indicada. La limitante fundamental en los enfoques ya mencionados parece poder salvarse si abandonamos la idea de establecer criterios de clasificación (que denominamos “sesgados”) fundamentados en ideas preconcebidas de cómo debemos de observar las proteínas para lograr su clasificación. En este sentido, hay dos asuntos que considerar: 1) Cómo lograr la caracterización de las proteínas de manera no sesgada y 2) Cómo lograr lo anterior sin apelar a medidas de agrupamiento que, a su vez, impliquen sesgos en dicho agrupamiento. La respuesta a la segunda pregunta ha sido intentada en el pasado aplicando técnicas de clasificación no supervisada basadas en redes neuronales (RNs). Particularmente interesante para nosotros ha sido la aplicación de los llamados

“mapas auto-organizados” (o SOMs, por su acrónimo inglés). En este tipo de RNs el conjunto de datos se presenta a la red como una tabla de características de interés y ésta, vía el algoritmo de Kohonen [12], ubica en un mapa n-dimensional (típicamente  $n=2$ ) los elementos que comparten características “similares”. Lo interesante de este método es que la medida de similitud no se explicita; el algoritmo establece una métrica dinámica que varía con los datos. La respuesta a la primera pregunta (la determinación de las características) ha sido respondida en nuestro grupo apelando a características estructurales de las proteínas expresadas como cadenas de AAs, como se discute más adelante. La explicación de los agrupamientos puede lograrse apelando a un argumento más sutil, basado en las técnicas de compresión de datos sin pérdida. Como se menciona en [13] existen dos motivaciones principales para buscar la compresión en un conjunto de datos: 1) El uso eficiente de recursos de almacenamiento y ancho de banda y 2) La posibilidad de explicitar la estructura en una secuencia. Se sabe [14] que las técnicas de compresión de datos tradicionales son inadecuadas para comprimir estructuras proteínicas y que la inclusión de criterios bioquímicos muestra que hay poca dependencia de Markov en la proteína. De esta manera, los esquemas de compresión de datos tradicionales se reducen a modelos de orden cero. Si queremos subsanar esta limitante de los métodos de compresión clásicos, debemos abandonar los modelos de bajo orden. Hemos propuesto, por ello, esquemas de compresión sin pérdida basados en la detección de patrones arbitrarios [15]. En [16] demostramos que el problema de encontrar dichos patrones es NP-completo y su solución requiere de la aplicación de meta-heurísticos [tales como los algoritmos genéticos (AGs)]. En resumen, pues, una exitosa metodología de clasificación de las proteínas y la explicación de dicha clasificación debe resolver no solamente el problema de la clasificación no sesgada, sino también el de la detección de patrones embebidos en las proteínas pertenecientes a cada uno de los grupos. En las siguientes dos secciones hacemos una breve exposición de los principios detrás de los SOMs y los AGs.

### **Mapas Auto-Organizados**

El principal objetivo de un mapa auto-organizado es transformar un patrón de datos de entrada con dimensiones arbitrarias en un mapa discreto multidimensional y ejecutar esta transformación de una manera ordenada topológicamente. Cada patrón de entrada presentado a la red típicamente consiste de una región localizada de actividad versus un fondo estable. La localización y naturaleza de cada región usualmente cambia de una realización a otra con respecto al patrón de entrada. Todas las neuronas en la red deben ser expuestas a un número suficiente de realizaciones del patrón de entrada para asegurar que el proceso de autoorganización tiene oportunidad de madurar apropiadamente. Con el objetivo de lograr esta organización de neuronas, éstas son colocadas de acuerdo con los patrones de entrada por medio de vecindades, por lo que las neuronas no son adaptadas individual sino conjuntamente con sus neuronas vecinas. De esta manera, las neuronas pueden, inicialmente, ser organizadas en grupos para representar aproximadamente patrones similares; en las siguientes fases una estructura más fina para los grupos de neuronas es desarrollada [19]. En nuestro caso las clases corresponden a grupos de proteínas que comparten rasgos de similitud.

## Algoritmos Genéticos

El problema de optimización que está tácito en los procesos antes mencionados es NP-completo y ello hace imposible encontrar, en general, la solución óptima. En años recientes se han desarrollado técnicas denominadas “evolutivas” que permiten abordar con éxito este tipo de problemas. Es imposible discutir las razones teóricas detrás de esta afirmación en un trabajo de esta índole. Simplemente dejamos constado aquí que puede demostrarse matemáticamente que una de dichas técnicas (los AGs) converge al óptimo global de cualquier problema de optimización estática (ver, por ejemplo, [20]).

## Prospectiva

La solución del problema que nos ocupa es de importancia en varios sentidos. Desde el punto de vista teórico su solución, por supuesto, implica el descubrimiento de relaciones estructurales de orden superior. Este descubrimiento, tácitamente, implica el hallazgo de relaciones de tipo filogenético, biológico y, en última instancia, a nivel de la química subyacente. Pero trabajos como [17] indican que, puesto que la replicación de nucleótidos es esencialmente un fenómeno de transferencia de información, las presiones evolutivas que dan forma a la composición del alfabeto (A, C, G, T) de los nucleótidos pueden no estar confinadas a cuestiones físico-químicas y que la codificación de los errores de transcripción puede haber jugado un papel restrictivo en los procesos evolutivos. En ese contexto, es claro que no podemos buscar los orígenes de estructuración proteínica (que nace de la codificación de los codones en AAs) apoyándonos en consideraciones meramente funcionales. Más aún, si, como se afirma en [13] la inclusión de criterios basados en principios bioquímicos no arroja luz acerca de la estructura de las proteínas, es lógico buscar otros principios.

## Clasificación

Nuestro enfoque al problema consiste en buscar características estructurales en las cadenas de AAs sin apelar a criterios explicativos apriorísticos. En ese sentido decidimos establecer rasgos estructurales y: a) Trabajar con proteínas expresadas como secuencias de AAs; b) Asignar a cada AA en una proteína un conjunto de coordenadas en espacios de 1, 2 y 3 dimensiones simultáneamente; c) Establecer un criterio de dispersión para considerar aquellos casos en los que los AAs en cuestión aparezcan más de una vez en la secuencia; d) Establecer una medida del contenido de información de cada AA y e) Establecer un descriptor único para la proteína, de manera tal que dicho descriptor considerara a todos los AAs caracterizados de acuerdo con (b-d). Así, cada proteína se caracteriza por un vector de 260 rasgos conformados como sigue: Consideramos cada proteína representada en 1, 2 y 3 dimensiones; obtenemos las coordenadas de los AAs en dichas representaciones y, para cada una de ellas, calculamos la desviación absoluta. Además, registramos la frecuencia del AA en la proteína. A cada AA corresponden, pues, 13 rasgos; de manera que para una proteína tenemos un total de 260 de ellos. Este conjunto de vectores se proporciona a un SOM

el cual, de acuerdo con su entrenamiento, aglutina conjuntos de proteínas de manera “natural”. Los resultados preliminares arrojan, a la fecha, comportamientos estadísticamente interesantes: para grupos de proteínas de *S. cerevisiae* elegidos aleatoriamente, los *clusters* determinados se repiten consistentemente. Es decir, en una colección arbitraria de proteínas del mismo organismo se repiten los mismos agrupamientos. Además, y de manera significativa, las proteínas encontradas en cada uno de los *clusters* comparten características biológicas que no podemos detallar en este trabajo por razones de espacio.

### **Patrones**

El segundo paso consiste en aplicar los algoritmos de compresión basados en patrones y no en colecciones de cadenas, tal como se describe en [16]. Este problema ha sido atacado en [18] con resultados alentadores sobre datos diseñados *ad hoc* para probar la eficiencia de los AGs ya discutidos. Después de aplicar el algoritmo sobre conjuntos determinados (como se describió en el apartado precedente) hemos encontrado patrones similares en grupos similares. La metodología propuesta, pues, ha quedado ampliamente justificada y nos permite pensar que una posible solución al problema de clasificación de proteínas es reducible a uno, fundamentalmente, de cómputo. Una vez logrado lo anterior, la participación de expertos en BM nos llevará a una mejor comprensión de los procesos bioquímicos y funcionales involucrados.

### **Conclusiones**

Indudablemente la agenda planteada depende del concurso multidisciplinario en las áreas de computación, teoría de la información, inteligencia artificial, redes neuronales, computación evolutiva y de la biología molecular en su primera etapa. Posteriormente requerirá de otras áreas para validar los resultados obtenidos y, en su caso, llevarlos a cristalizar aplicaciones en el área médica, biológica y química. Además de los problemas técnicos, otro más básico con el que hemos tenido que contender ha sido la heterogeneidad de los dialectos técnicos que los cultivadores de cada una de las áreas antes mencionadas y que, directa o indirectamente, han participado en el proyecto manejan. De la conclusión de los desarrollos planteados deberá desprenderse el diseño de sistemas de aplicación muy concretos y aplicados a problemas de interés vario.

### **Referencias**

- [1] Sun, Z., Deogun, J. S. (2004) “Local prediction approach for protein classification using probabilistic suffix trees”. ACM International Conference Proceeding Series. Proceedings of the second conference on Asia-Pacific bioinformatics - Volume 29. Dunedin, New Zealand. Pages: 357 - 362.



- [2] Guralnik, V., Karypis, G. (2001) "A scalable algorithm for clustering protein sequences BIOKDD01". Workshop on Data Mining in Bioinformatics (with SIGKDD01 Conference) Pages: 73-80.
- [3] Enright, A. J., Ghosh, J., Peng, K. (2003) "An efficient algorithm for large-scale detection of protein families". *Nucleic Acids Research*, 30(7):1575-1584.
- [4] <http://www.ncbi.nlm.nih.gov/>
- [5] Liao, L., Stafford N. W. (2003) "Combining Pairwise Sequence Similarity and Support Vector Machines for Detecting Remote Protein Evolutionary and Structural Relationships". *Journal of Computational Biology*. 10 (6): 857-868.
- [6] Andrade, M.A.; Casari, G., Asari, C., Sander and Valencia A. (1997) "Classification of protein families and detection of the determinant residues with an improved self-organising map". *Biol. Cybern.* 76, 441-450.
- [7] Pollock, R., Toby, L., Watts, M. (2002) "A Kohonen Self-Organizing Map for the functional classification of proteins based on one-dimensional sequence information". *Proceedings of IJCNN* p. 189-192
- [8] Wang, J.T.L.; Quicheng, M.A.; Shasha, D, Wu H., Cathy. Application of Neural Networks to Biological Data Mining: a case study in protein sequence classification. <http://www.cs.dal.ca/~cblouin/Blouin/lectures/p305-wang.pdf> .
- [9] Pasquier, C., Promponas, V. J., Hamodrakas, S. J. (2001) PRED-CLASS: Cascading Neural Networks for Generalized Protein Classification and Genome-Wide Applications. *PROTEINS: Structure, Function, and Genetics* 44: 361-369.
- [10] Pasquier, C., Hamodrakas, S. J. (1999) "An hierarchical artificial neural network system for the classification of transmembrane proteins". *Protein Engineering*. 12 (8): 631-634.
- [11] Chen, J., Chaudhari, N. S. (2003) "Protein family classification using second-order recurrent neural networks". *Genome Informatics* 14: 520-521.
- [12] Kohonen, T., (1988), "Self-Organizing and Associative Memory", 3rd. Ed. , Springer-Verlag.
- [13] Nevill-Manning, C.G. and Witten, I.H. (1999) "Protein is incompressible" *Proc. Data Compression Conference*, J.A. Storer and M. Cohn (Eds .) IEEE Press, Los Alamitos, CA, pp. 257-266.
- [14] Loewenstein, D., and Yianilos, P. (1997), "Significantly lower entropy estimates for natural DNA sequences", *Proc. Data Compression Conference*, IEEE Press, pp. 151-160.
- [15] Kuri, A., (2004), "Pattern Based Lossless Data Compression", *WSEAS Transactions on Communications*, Issue 1, Vol. 3, pp. 42-49.
- [16] Kuri, A. and Galaviz, J.,(2004), "Pattern-based Data Compression", *Lecture Notes on Artificial Intelligence*, LNAI 2972, Springer-Verlag, pp. 1-7.
- [17] Mac Dónaill, D.A., (2003), "The Role of Error-Coding in Shaping the Nucleotide Alphabet: Nature's Choice of A, U, C and G", *Proc. of the 25th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, IEEE Press, Cancún, México.
- [18] Kuri, A. and Herrera, O., (2003), "Transformación de un mensaje generado por una Fuente No Ergódica a otro generado por una Fuente Ergódica para Compresión de Datos", 14ª Reunión de Otoño de Comunic., Comp. y Electrónica, ROC&C 2003, IEEE, Acapulco, Nov. 2003.
- [19] Kohonen, T. *Self-Organizing Maps*, Second Edition, Springer, pp. 85 – 90, 1997.
- [20] Kuri, A., (2002), "Algoritmos Genéticos", Fondo de Cultura Económica.

# Experimental Analysis of a Genetic Algorithm for the Sequencing by Hybridization Problem

Luis C. González-Gurrola and Carlos A. Brizuela

Computer Science Department, CICESE Research Center  
Km 107 Carr. Tijuana-Ensenada, Ensenada, B.C., México  
{gurrola, cbrizuel } @ cicese.mx

**Abstract.** This paper deals with the experimental analysis of a genetic algorithm for a computational biology problem. The problem appears in the computational part of a new deoxyribonucleic acid (DNA) sequencing procedure denominated sequencing by hybridization (SBH). The method is widely used for Single Nucleotide Polymorphism analysis and its general usage for sequencing purpose depends mainly on the development of good algorithmic procedures solving the computational phase of this approach. One of the parameters of this method is the length of the oligonucleotides. We show by means of experiments, that modifying this parameter, a recently proposed genetic algorithm generates high quality solutions for large instances.

**Resumen.** Este artículo trata sobre el análisis experimental de un algoritmo genético para un problema de biología computacional. El problema aparece en la fase computacional de un nuevo procedimiento de secuenciación de ADN denominado secuenciación por hibridación (SBH). Este método es ampliamente utilizado para el análisis de los *polimorfismos de los nucleótidos únicos* (SNP's) y su utilización para propósitos de secuenciación de ADN depende principalmente del desarrollo de algoritmos que resuelvan la fase computacional de este método. Uno de los parámetros de este método es la longitud de los oligonucleótidos. En este trabajo se muestra por medio de experimentos que modificando este parámetro, un algoritmo genético propuesto recientemente genera soluciones de alta calidad para casos de gran tamaño.

**Keywords:** GA's, DNA Sequencing by Hybridization, DNA Chip.

## 1. Introduction

The DNA sequencing is a very important problem in molecular biology. It refers to the identification of an unknown short DNA sequence generally of  $100 \leq n \leq 500$  base pairs. Among the existing methods for sequencing we are going to concentrate in a relatively new one denominated sequencing by hybridization (SBH) [1]. This method has been already widely used for **Single Nucleotide Polymorphism** (SNP) analysis and its general usage for sequencing purpose depends mainly on the development of good algorithmic procedures solving the computational phase of this approach [2].

The SBH problem is composed of two stages: the biochemical experiment and the computational stage. For the biochemical experiment we are given an unknown DNA

sequence, composed by a sequence of nucleotides from a set of four of them: A (adenine), C (cytosine), G (guanine), and T (thymine). From this sequence we need to detect all oligonucleotides (short sequence of nucleotides) of length  $l$ , which properly ordered compose the unknown sequence. The current technology allows oligonucleotide length  $l$  up to 30 nucleotides [10]. The detection is performed using a DNA chip (or DNA array) which contains all fragments (oligonucleotides) of length  $l$ , i.e.  $4^l$  fragments. A single strand of the unknown DNA sequence treated with a fluorescent substance is deposited on the DNA chip. During the biochemical experiment, complementary fragments (A-T and C-G) of length  $l$  come together i.e. hybridize. After the hybridization experiment and by reading a fluorescent image of the chip, we can obtain the spectrum, which is the set of oligonucleotides composing the unknown DNA sequence. Here, the second stage of SBH starts, i.e. given the spectrum and the length of the unknown DNA sequence, find the order of the oligonucleotides such that consecutive elements always overlap on  $l-1$  nucleotides. When the hybridization is performed without errors the spectrum includes all length  $l$  oligonucleotides originally in the unknown DNA sequence. However, many factors do not allow the experiment to be run error free. If the spectrum does not include one or more oligonucleotides originally in the DNA sequence, then we have an experiment with negative errors. On the other hand, if the spectrum includes oligonucleotides that are not present in the original sequence then we have an experiment with positive errors.

The remainder of the paper is organized as follows. Section 2 briefly describes the state of the art. Section 3 states the problem we are dealing with. Section 4 indicates the methodology proposed to approximately solve the problem. Section 5 presents the experimental setups and results. Finally, section 6 states the conclusions and points out some ideas for future research.

## 2. State of the Art

There have been many attempts in trying to find an efficient method to approximately solve the computational part of the SBH problem. For the ideal experiment case it has been shown that the problem can be solved in polynomial time [8] by reducing it to finding an Eulerian path in a directed graph. However, when only negative or only positive or both errors are present, the problem becomes strongly NP-hard [5]. A Tabu Search based algorithm for the latter case is proposed in [4]. Then a sophisticated heuristics that improves the results in [4] is presented in [2]. At the same time a genetic algorithm [6] is proposed to effectively and efficiently deal with positive and negative errors. In a previous work, Blazewicz and colleagues [3] present a branch and bound method to deal with positive and negative errors. This algorithm obtains its best performance with only positive errors, however, it has problems handling negative errors produced by repetitions of oligonucleotides of length  $l$ . Zhang et al. [10] employ a different technique to effectively handle negative errors generated by repetitions of oligonucleotides of length  $l$ . In the authors recent work [7], a genetic algorithm denominated *Sequencing GA* which is an improvement over the genetic algorithm proposed by Blazewicz et al. [6], is presented. In the comparison of

both algorithms, these were applied to spectra of  $100 \leq |\mathbf{S}| \leq 500$ . In this work, the spectra size is increased up to 1000 oligonucleotides and by modifying one parameter of the SBH method, the solution quality remains high.

### 3. Problem Statement

The computational part of the SBH problem consists of a set  $\mathbf{S} = \{s_1, s_2, \dots, s_k\}$  of equal length ( $l$ ) strings  $s_i$ 's over an alphabet  $\Sigma = \{A, C, G, T\}$  and a number  $n$  representing the length of the unknown sequence. Each  $s_i$  is always a fragment of the original sequence  $N$ , whenever the experiment is error free ( $|\mathbf{S}| = n - l + 1$ ). However, in general,  $s_i$  may represent a fragment that is not in the original sequence (positive errors), furthermore, there may be fragments in the original sequence  $N$  that do not appear as a string  $s_i$  in  $\mathbf{S}$  (negative errors). The problem is to find a sequence  $L$  of length no greater than  $n$  such that the number of used strings  $s_i$ 's is maximized, and therefore the differences between  $N$  and  $L$  minimized.

By means of experiments the Sequencing GA has shown its superiority over the GA proposed by Blazewicz [7]. However, when the spectrum size increases to  $|\mathbf{S}|=600$  and above, the similarity score of the solutions generated by both algorithms, regarding the original DNA sequence, decreases drastically below an acceptable level. This phenomenon can be avoided if the length of the oligonucleotides used in the SBH biochemical stage is increased. Both algorithms used  $l=10$ , if we modify this length, we show by means of experiments that the similarity score of the solutions generated by the Sequencing GA are over 99% using spectra up to  $|\mathbf{S}|=1000$ .

### 4. Methodology

The algorithm is proposed in the authors recent work [7], it is based on the one proposed by Blazewicz and colleagues [6]. Details of this algorithm can be found in [7].

### 5. Experimental Setup and Results

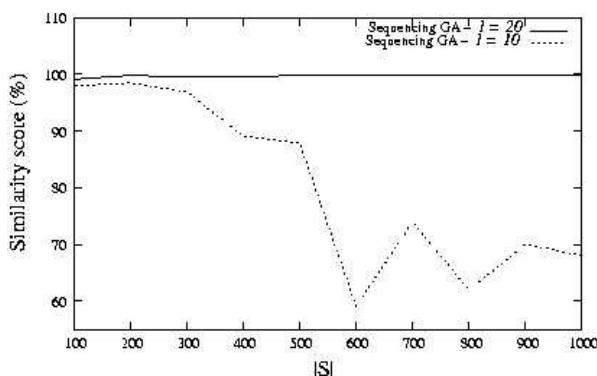
The Sequencing GA has been applied to real DNA sequences. All spectra used in the experiment have been derived from the DNA sequences coding human proteins taken from GenBank, National Institute of Health, USA. Their accession numbers are given by X05908, X02994, X51841, X06537, X07820, D11428, Y00093, X02160, X14894 and X06985, respectively. Each derived spectrum was modified with the introduction of 40% of errors (20% negative and 20% positive). The spectra have been sorted alphabetically, thus no information about the original order of oligonucleotides from its original sequence is known.

The lengths of the sequences have been set to  $109 \leq n \leq 1009$ . For each case, 10 different instances were generated. The algorithm was run 30 times, and it was

applied to each of the 10 instances. All the experiments have been performed on a PC station with Athlon XP 2.0 GHz processor, 512 MB RAM and Linux Mandrake 9.1 operating system.

In order to know the effect of the length of oligonucleotides in the sequence reconstruction, these are set to  $l=10$  and  $l=20$ , respectively.

The main criterion we want to compare is the similarity score; Figure 1 shows the results for this comparison. The continue line is for  $l=20$  and the dotted line is for  $l=10$ .



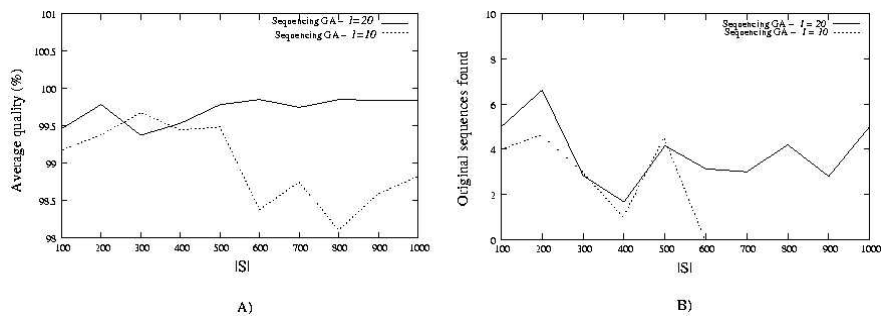
**Fig. 1.** Similarity score for the sequences generated by the Sequencing GA using  $l=10$  and  $l=20$

The similarity score indicates the percentage of similarity between the sequence generated by the Sequencing GA and the original sequence (if both sequences are identical, then the similarity score is 100%). In order to compute this criterion, we compared the solutions generated by the GA with the original sequences using a pairwise alignment algorithm [9]. We can see in this figure the differences among the sequences generated by the same algorithm, but using a different oligonucleotide length. Definitely, an increment in the length of oligonucleotides can lead to very good (more than 99%) similarity scores for all sequences. An interesting point is that no matter what the instances size are, the similarity score remains close to the optimal value.

For the cases of  $l=10$ , the similarity score decreases due to the existence of groups of oligonucleotides (adjacent in the original DNA sequence) missed in the spectrum. Therefore, false oligonucleotides (positive errors) or other oligonucleotides (which correspond to other sequence location) can be selected for reconstructing the candidate DNA sequence.

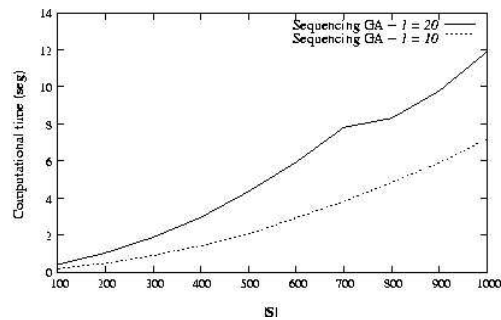
Figure 2(A) shows the average quality obtained by the Sequencing GA for  $l=10$  and  $l=20$ . Here, quality is the number of oligonucleotides in the spectrum used by the GA for generating a solution. Notice that, the optimal quality is the number of real (not corresponding to positive errors) oligonucleotides in the spectrum for a given case. In this figure, we can see that while the quality obtained by the Sequencing GA using  $l=10$  decreases as the instance size grows, using  $l=20$  maintains in average a 99.5% of the optimal quality.

Figure 2(B) shows the number of original sequences found per run at each instance size. Notice that at each run, the Sequencing GA is applied over 10 instances, therefore, the ideal case is to find at each run 10 original sequences. From  $|S|=600$  and above the Sequencing GA using  $l=10$  does not obtain any of the original sequences, whereas with  $l=20$  it can obtain two or more of them.



**Fig. 2.** A) Average similarity score over 30 runs at each instance size using  $l=10$  and  $l=20$ . B) Number of original sequences found per run using  $l=10$  and  $l=20$

Finally, a comparison of the computational time is presented. Figure 3 shows the differences in computational time for the cases of  $l=10$  and  $l=20$ . Obviously the larger the oligonucleotides are, the longer the computational time will become. But, at least for these instances size, we consider we can sacrifice the differences in time in order to obtain more accurate sequences.



**Fig. 3.** Computational time of the Sequencing GA using  $l=10$  and  $l=20$

## 6. Conclusions

A recently proposed genetic algorithm is analyzed when real size oligonucleotides are introduced to solve the sequencing by hybridization problem. In this previous work the candidate sequences similarity score decreases below an acceptable level for 600

$\leq |S| \leq 1000$  and  $l=10$ . We show by means of experiments that an increment in the length of  $l$  improves this score.

The feasibility of using this algorithm to complete both stages, biochemical and computational, in the SBH method is aimed as future research. It would also be interesting to extend this algorithm to deal with the shotgun sequencing method in order to reconstruct sequences with repeats.

### **Acknowledgements**

This research has been partially supported by LAFMI Project REF J000-0686/2003.

### **References**

- [1] W. Bains and G.C. Smith. A novel method for nucleic acid sequence determination. *Journal of Theoretical Biology*, 135:303-307, 1988.
- [2] J. Blazewicz, P. Formanowicz, F. Guinand and M. Kasprzak. A Heuristic Managing Errors for DNA Sequencing. *Journal of Bioinformatics*, 18(5):652-660, 2002.
- [3] J. Blazewicz, P. Formanowicz, M. Kasprzak, W. Markiewicz and J. Weglarz. DNA Sequencing with Positive and Negative Errors. *Journal of Computational Biology*, 6:113-123, 1999.
- [4] J. Blazewicz, P. Formanowicz, M. Kasprzak, W. Markiewicz and J. Weglarz. Tabu search for DNA Sequencing with false negatives and false positives. *European Journal of Operational Research*, 125:257-265, 2000.
- [5] J. Blazewicz and M. Kasprzak. Complexity of DNA Sequencing by Hybridization. *Journal of Theoretical Computer Science*, 290:1459-1473, 2003.
- [6] J. Blazewicz, M. Kasprzak and W. Kuroczycki. Hybrid Genetic Algorithm for DNA Sequencing with Errors. *Journal of Heuristics*, 8:495-502, 2002.
- [7] C.A. Brizuela, L.C. González and H.J. Romero. An Improved Genetic Algorithm for the Sequencing by Hybridization Problem. In *Applications of Evolutionary Computing*, LNCS, volume: 3005, pages:11-20, 2004.
- [8] P.A. Pevzner. 1-tuple DNA Sequencing: Computer Analysis. *Journal of Biomolecular Structure and Dynamics*, 7:63-73, 1989.
- [9] J. Stoye. Multiple Sequence Alignment with the Divide-and-Conquer Method. *Gene*, 211(2):GC45-GC56, 1998.
- [10] J. Zhang, L. Wu and X. Zhang. Reconstruction of DNA sequencing by hybridization. *Journal of Bioinformatics*, 19(1):14-21, 2003.

# Optimización de la Geno-tipificación de Papilomavirus Humano como un problema de Selección de Características utilizando PCA y RNAs

M. Rodríguez-Muro, P. Gómez-Gil, C. Castañeda-Roldán, J. Garcés-Eisele, J. Lozano-Yécora, M. Osorio-Galindo

Universidad de las Américas  
Sta. Catarina Mártir, Cholula, Puebla.  
72820 México

marianorodriguez@mexico.com, {pgomez, ccastane, jgarces, ci205690, josorio}@mail.udlap.mx

**Resumen.** Se presenta el problema de optimizar la geno-tipificación de Papilomavirus Humanos como un problema de selección de características. Se propone el uso del Análisis de Componentes Principales como técnica de selección de características sobre el problema de tipificación de ADN apoyado por el desempeño de redes neuronales artificiales para la evaluación de las selecciones.

## 1. Introducción

La geno-tipificación de Ácido Desoxirribonucleico (ADN) del Papilomavirus Humano (VPH) se puede realizar a través del método conocido como Polimorfismo en la Longitud de los Fragmentos de Restricción (RFLP). Dicha técnica se basa en la aplicación de enzimas sobre las cadenas de ADN de la muestra del virus [8].

Existen más de 200 diferentes enzimas disponibles para geno-tipificación y más de 60 variantes del VPH. Muchas de estas enzimas generan información redundante entre ellas por lo que, para identificar una muestra, el analista debe escoger un pequeño grupo de enzimas que sean suficientes para identificar determinado grupo de virus; si después de aplicar las enzimas sobre una muestra aún quedan dudas sobre la identidad del virus, el analista escoge otra(s) enzima(s) que le permita(n) concluir la identificación. Es este proceso el que deseamos optimizar. El procedimiento de selección es empírico y, debido a que el número de variantes del VPH y el número de enzimas para su identificación son elevados, escoger un conjunto de enzimas mínimo con el que sea posible identificar con certeza cada tipo viral resulta muy difícil y en algunos casos imposible. Además este problema no es exclusivo de la geno-tipificación del VPH, y en otros casos el número de reactivos y genotipos puede ser mayor aumentando la dificultad de la selección empírica.

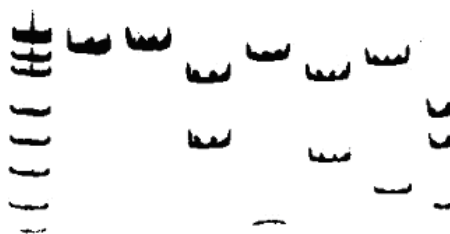


El problema se puede atacar como un problema de *selección de características*, es decir, la selección de un subconjunto de características tal que el desempeño del clasificador sea máximo [6]. Proponemos utilizar el Análisis de Componentes Principales (Principal Component Análisis, PCA) usado como técnica de selección de características [11],[3]. El PCA es una técnica de análisis multivariante que se utiliza generalmente para problemas de reducción de la dimensionalidad, es relativamente simple y fácil de revisar por lo que esperamos nos dé una pauta de medición para el futuro análisis de otras técnicas.

## 2. Tipificación de VPH a través de RFLP

El método de RFLP utiliza *enzimas de restricción* para identificar el ADN. Éstas reconocen una subcadena y cortan una cadena de ADN en donde ocurra dicha subcadena. Así se obtienen fragmentos de diferentes tamaños; son el número de fragmentos y el tamaño de éstos lo que nos permite identificar la muestra [8].

Para obtener el número y el tamaño de los fragmentos se utiliza una técnica llamada *electroforesis*. Dicha técnica genera patrones de restricción como el mostrado en la figura 1, donde cada columna representa un patrón de restricción. Éstos se generan del desplazamiento de los fragmentos de ADN en un gel después de aplicarles una diferencia de potencial. El desplazamiento está en función del tamaño de los fragmentos por lo que podemos basarnos en esta medición y en la cantidad de fragmentos para identificar parcialmente la muestra.



**Fig. 1.** Patrones de restricción para diferentes tipos de VPH, generados con la enzima de restricción DdeI [8]

Debido a que los patrones de restricción se forman con base solamente en el tamaño de las moléculas, es posible tener patrones de restricción iguales para virus diferentes utilizando la misma enzima. De esta manera tenemos que si dos enzimas A y B generan el mismo patrón para cierto virus, ubicándolo en un genotipo *G1* o *G2*, es necesario utilizar una enzima C que sabemos que genera un patrón diferente para los genotipos *G1* y *G2* para poder identificar correctamente la muestra.

### 3. Trabajo relacionado

Se ha demostrado que encontrar el número mínimo de enzimas necesarias para identificar un grupo de virus, como el del VPH, es un problema NP-completo [8]. Sin embargo, puede atacarse heurísticamente de diferentes maneras.

En [4] se utiliza la Teoría de la Información de Shannon para crear un algoritmo de selección basándose en la cantidad de información contenida por enzima; en este trabajo se concluye que es posible clasificar los 50 tipos virales con los que se trabajó con solamente 2 enzimas. Sin embargo, en la práctica, estas enzimas son poco útiles por generar frecuentemente digestiones parciales, lo que limita su uso.

En [9] se utiliza un algoritmo ávido y un algoritmo ramifica-limita para realizar la selección. En el mejor de los casos, utilizando 139 enzimas y 48 genotipos virales, el algoritmo ávido seleccionó 2 enzimas; el algoritmo ramifica-limita no fue probado con este caso debido a su extensión. En el peor de los casos, utilizando 17 enzimas y 19 genotipos virales, el algoritmo ávido seleccionó 11 enzimas; el algoritmo ramifica-limita seleccionó 6 enzimas partiendo de las 11 enzimas seleccionadas por el algoritmo ávido.

Entre los beneficios que se obtendrían de esta investigación, y de los demás trabajos que se están efectuando sobre el tema, está la reducción de costos de tiempo y recursos utilizados en la geno-tipificación [4].

### 4. Selección de Características en la Tipificación de VPH

El problema de selección de características ha sido atacado fuertemente en los últimos años y está íntimamente ligado con el reconocimiento de patrones. Tsamardinos [12] señala que no existe una definición totalmente aceptada del problema pero propone una que creemos es muy adecuada para los propósitos de esta investigación:

**Definición 1** [12] Un problema de selección de características o variables es una tupla  $\langle X, \Phi, T, A, M \rangle$ , donde  $X$  es una muestra de patrones de entrada definidos sobre un conjunto de características  $\Phi$ ,  $T \in \Phi$  es una variable objetivo,  $A$  es un algoritmo de clasificación que produce un modelo de predicción para  $T$  dado  $T$  y  $X$ ; y  $M$  es una métrica de desempeño del modelo del clasificador y de las características seleccionadas. Una solución al problema es un subconjunto de características  $\phi \subseteq \Phi$  que maximiza  $M(\phi, A(T, X \downarrow \phi))$ , donde  $X \downarrow \phi$  es la proyección de los datos  $X$  sobre las características pertenecientes a  $\phi$ .

En estos términos, el problema de tipificación de ADN utilizando RFLP se traduce a encontrar un  $\phi \subseteq \Phi$  que maximice  $M$ ; donde  $M$  está compuesta por 2 factores, la clasificación correcta de todos los tipos de virus y la cantidad de enzimas necesarias para la clasificación. El espacio de características  $\Phi$  estará definido por la concatenación de espacios de 4 características por enzima, formando un vector de la forma

$$\Phi = \{\{R_1\}, \{R_2\}, \dots, \{R_k\}\} \quad (1)$$

donde  $R_k$  es el espacio reservado  $R$  para las distancias de los fragmentos de los patrones de restricción generados por la enzima  $k$ . Se escogió un tamaño de 4 para cada espacio  $R_k$  debido a que el número de fragmentos en los patrones de restricción generados por una enzima se encuentra entre 1 y 4 en los datos con los que contamos. Así, cada  $R_k$  tendrá 4 datos, el primero será la distancia recorrida por el primer fragmento, el segundo por el segundo fragmento y así sucesivamente. En caso de que una enzima genere menos que 4 fragmentos para cierto genotipo se asignará una distancia de 0 a los espacios de  $R_k$  sobrantes. Escoger un número menor a 4 espacios provocaría pérdida de información. Tampoco es aconsejable utilizar un tamaño superior pues ninguno de los 48 genotipos disponibles genera más de 4 fragmentos para ninguna de las 205 enzimas con las que contamos. De esta manera tenemos que un patrón de entrada  $x_i \in X$  estará formado de la siguiente manera:

$$x_i = \{\{r_{i1}\}, \{r_{i2}\}, \dots, \{r_{ik}\}\} \quad (2)$$

donde  $x_i$  es el patrón para el genotipo viral  $i$  formado por la concatenación de todos  $r_{ik}$  que son las distancias recorridas por los fragmentos del patrón de restricción  $r$  generado por la enzima  $k$  para el genotipo viral  $i$ .  $T$  es la relación uno a uno que existe entre las muestras contenidas en  $X$  y los genotipos que se desean clasificar.

Una vez que el algoritmo de selección escoja ciertas características bastará con verificar a cual de los  $R_k$  espacios pertenecen estas características para saber a que enzima corresponden. Tendremos un total de 48 patrones  $x_i$  en  $X$ , correspondientes a los 48 genotipos virales con los que trabajaremos. Cada  $x_i \in X$  estará formado por 820 características resultado de concatenar los espacios  $R_k$  de tamaño 4 para las 205 enzimas que usaremos.

El algoritmo de clasificación  $A$  será el Análisis de Componentes Principales utilizado como técnica de selección de características.

## 5. Análisis de Componentes Principales

También es conocido como la transformada Karhunen-Loève [2]. El Análisis de Componentes Principales es una de las técnicas de análisis multivariante más sencillas [10]. Su objetivo es el de, dado un conjunto de variables  $Y_1, Y_2, Y_3, \dots, Y_p$ , encontrar combinaciones lineales que produzcan índices  $Z_1, Z_2, Z_3, \dots, Z_p$ , llamados componentes principales, tal que éstos estén incorrelacionados y la varianza esté maximizada en cada uno de ellos. La falta de correlación entre los índices indica que éstos miden dimensiones diferentes de los datos. Además, los nuevos índices están ordenados en relación a la varianza de éstos.

Esta técnica es especialmente útil cuando hay una alta correlación entre las variables originales [10]. En estos casos es común poder representar el total de las variables originales utilizando solamente una fracción de los componentes principales generados. Debido a esta propiedad se ha hecho un procedimiento común para reducir la dimensionalidad de los datos de entrada en una gran variedad de aplicaciones.

### 5.1. Selección de Características utilizando PCA

Aunque no es su función principal, en [10], [5], [7] se menciona el uso de esta técnica para evaluar la importancia de las variables en un conjunto de datos, de lo que se intuye su capacidad como técnica de selección. En [11], y [3] se hace referencia al PCA explícitamente como técnica de selección de características.

Cuando se aplica el PCA sobre un conjunto de datos, como nuestro grupo de patrones virales  $X$ , se obtiene una matriz de transformación  $W$  donde cada renglón  $m$  contiene los coeficientes de la ecuación que permite calcular el  $m$  componente principal. Así, el cálculo de cada componente principal se obtiene de la forma:

$$Z_m = C_{m1}Y_1 + C_{m2}Y_2 + \dots + C_{mp}Y_p \quad (3)$$

donde  $Z_m$  es el  $m$ -ésimo componente principal y  $Y_p$  es la  $p$ -ésima característica o variable. En nuestro caso, las características  $Y_p$  serán los elementos del patrón viral  $x_i$  del que queramos obtener su  $m$ -ésimo componente principal.

Los coeficientes  $C_{mp}$  resultantes del PCA pueden verse como la *importancia* que tiene la variable o característica  $p$  en el cálculo del componente  $m$ . Podemos suponer que aquellas características o variables  $p$  cuyos coeficientes  $C_{mp}$  sean elevados para los primeros componentes principales serán aquellas características que contengan la mayor información. Las características que se escojan de esta manera serán los elementos de  $\phi$ .

Es importante aclarar que la selección de dichas características sigue teniendo un grado de subjetividad [7], pues aunque se tienen métricas para comparar la varianza de los diferentes componentes principales, escoger cuantos de éstos son suficientes para representar adecuadamente las características originales no es un procedimiento algorítmico. De igual manera, no existe una regla fija que nos diga cuando un coeficiente  $C_{mp}$  es lo suficientemente pequeño como para descartarlo.

Para ayudar a eliminar la incertidumbre del método de selección proponemos utilizar el desempeño de redes neuronales (RNA) como métrica de evaluación  $M$  de las características escogidas  $\phi$ . La selección se hace por etapas, primeramente se escoge el número mínimo de componentes principales que permitan la clasificación correcta de todos los genotipos de VPH, la clasificación se hace con una red neuronal. Posteriormente escogemos grupos de enzimas mínimas basándonos en los coeficientes  $C_{mp}$  de la ecuación de cálculo de estos componentes. Se vuelve a evaluar el desempeño con una red utilizando no los componentes principales escogidos sino solamente las características contenidas en  $\phi$ . Si el desempeño es aceptable se procede a utilizar los patrones de restricción generados por las enzimas asociadas a las características en  $\phi$  para evaluar el desempeño del clasificador.

En todas las etapas de la selección se usan RNAs alimentadas hacia delante y entrenadas por retro-propagación. Cada red permite evaluar el desempeño del grupo de características seleccionado en la etapa correspondiente. Una evaluación aceptable de un grupo de características resulta de una clasificación correcta de los 48 genotipos después de realizado el entrenamiento; una evaluación no aceptable resulta de la clasificación incorrecta de al menos un genotipo. Estas redes son de una sola capa que tiene un número de nodos  $j = 48$  correspondientes a los 48

genotipos virales con los que contamos y que deseamos clasificar. El número de entradas  $i$  depende del grupo de características  $\phi$  que se pretenda evaluar; si se está en la primera etapa,  $i$  corresponde al número de componentes principales seleccionados; si se está en la segunda etapa  $i$  corresponde al número de enzimas relacionadas con los componentes principales seleccionados en la primera etapa multiplicado por 4, pues cada enzima se relaciona a un  $R_k$  que es de tamaño 4 (ver 4).

## 6. Conclusiones y trabajo a futuro

Mostramos que el problema de geno-tipificación de ADN a través de RFLP puede ser atacado como un problema de selección de características. Proponemos el uso del PCA combinado con una red neuronal alimentada hacia adelante como técnica de selección de características. Después de evaluar el desempeño de la técnica descrita compararemos con otras técnicas de selección de características como el Análisis de Características Principales (Principal Feature Analysis, PFA) [3] y Algoritmos Secuenciales de Selección de Características [1], específicamente Selección Secuencial hacia Delante (Forward Sequential Selection, FSS) y Selección Secuencial hacia Atrás (Backward Sequential Selection, BSS).

## Referencias

1. D. W. Aha and R. L. Bankert. A comparative evaluation of sequential feature selection algorithms, 1995.
2. M. Á. Carreira-Perpiñán. A review of dimension reduction techniques. Technical Report CS-96-09, Dept. of Computer Science, University of Sheffield, January 1997.
3. I. Cohen, Q. Tian, X. Zhou, and T. Huang. Feature selection using principal feature analysis, 2002.
4. R. Fernández. Desarrollo de algoritmos para la clasificación de secuencias. Master's thesis, Universidad de las Américas, Puebla, 2002.
5. B. Flury and H. Riedwyl. *Multivariate Statistics: A practical approach*. Chapman and Hall, 1988.
6. R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273-324, 1997.
7. W. J. Krzanowski. *Principles of Multivariate Analysis*. Oxford statistical science series. Oxford University Press, 1988.
8. J. Lozano-Yécora, J. Garcés-Eisele, M. Osorio-Galindo, C. Castañeda-Roldán, and P. Gómez. Genotyping of human papilloma virus (hvp): a combinatorial optimization family of np-hard problems. Documento interno de trabajo. Universidad de las Américas-Puebla, 2004.
9. J. M. Lozano-Yécora. La genotipificación del virus del papiloma humano: una familia de problemas de optimización combinatoria. Master's thesis, Universidad de las Américas, Puebla, 2004.
10. B. F. J. Manly. *Multivariate Statistical Methods: A Primer*. Academic Press, 1986.
11. K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Probability and Mathematical Statistics. Chapman and Hall, 1995.
12. I. Tsamardinos and C. Aliferis. Towards principled feature selection: Relevancy, 2003.

# Using Graph-Based Concept Learning to the Predictive Toxicology Domain

Ivan Olmos and Jesus A. Gonzalez

Instituto Nacional de Astrofísica, Óptica y Electrónica  
Luis Enrique Erro #1, Sta. Maria Tonantzintla; Puebla; México  
{iolmos, jagonzalez}@ccc.inaoep.mx

**Abstract.** The increasing amount of stored data has created the necessity of tools that are able to get patterns that cannot be found with standard tools like statistics or spreadsheets. The predictive toxicology domain is an example where the database can be so large that the task of finding patterns in it is not easy. This is where data mining tools are appropriate to analyze this kind of data. In this paper we use a graph-based structured concept learning system called Subdue for finding concepts in the predictive toxicology database. We show our experiments with five representations to make clear the kind of concepts that can be found with each of them.

## 1. Introduction

For concept learning systems, data representation is crucial. A good representation might make possible the learning of a concept that was not learnable using other representations. For example, in the case of the earthquake domain (earthquake data such as its location, epicenter, intensity, depth, type, etc.) we can just use an attribute-value representation. However, we are missing important information such as the distance between the earthquakes' epicenters and the difference in time between the earthquakes. This information is lost in an attribute-value representation; information that could make possible the learning of an important concept.

Graphs [6] are a powerful and flexible knowledge representation used to model simple and complex structured domains. For example, in chemical compounds, vertices and edges can represent not only atoms and bonds but also functional groups that are formed by a set of atoms [3]. In the Web domain, vertices can represent web pages and edges represent the links between them [4]. In general, vertices describe objects or parts of objects in the domain, while edges represent relations between objects. The representation power and flexibility is the reason of why a graph-based representation model has been adopted by researchers in different areas [1 and 2].

In this paper we show the effectiveness of a graph-based concept learning system applied to the predictive toxicology domain. We use Subdue, which is a graph-based concept learning system where the concept to learn is represented by graphs. Subdue uses graph operations to verify if an example belongs to a concept or not, and to find new concepts.

## 2. Graph-Based Discovery

Subdue [1] is a relational learning system used to find substructures (subgraphs) that appear repetitively in the graph representation of databases. Subdue starts by looking for the substructure that best compresses the graph using the Minimum Description Length (MDL) principle [5], which states that the best description of a data set is the one that minimizes the description length of the entire data set. In relation to Subdue, the best description of the data set is the one that minimizes  $I(S) + I(G|S)$ , where  $S$  is the substructure discovered in the input graph  $G$ ,  $I(S)$  is the length (number of bits) required to encode  $S$ , and  $I(G|S)$  is the length of the encoding of graph  $G$  after being compressed using substructure  $S$ , i.e., every instance of  $S$  in  $G$  is replaced by a single vertex.

We used the concept learning option of Subdue known as SubdueCL [7]. SubdueCL is able to work with positive and negative examples in a supervised way using a set covering algorithm (as opposed to Subdue that uses an unsupervised algorithm and does not distinguish between positive and negative graphs).

### 2.1. Substructure Evaluation and the SubdueCL Algorithm

SubdueCL uses an evaluation formula to give a value to all the generated substructures. This formula assigns a value to a substructure according to how well it describes the positive examples (or a subset of the positive examples) without describing the negative examples. Then, positive examples covered by the substructure increase the substructure value while negative examples decrease its value. In this formula the positive examples that are not covered and the negative examples covered by the substructure are considered errors, because the ideal substructure would be one covering all the positive examples without covering any negative example. Then, the substructure value is calculated as follows:

$$value = 1 - Error$$

where the error is calculated with respect to the positive and negative examples covered by the substructure using the following formula:

$$Error = \frac{\#PosEgsNotCovered + \#NegEgsCovered}{\#PosEgs + \#NegEgs}$$

Using this formula, SubdueCL chooses rules that maximize the substructure's value, and in this way it minimizes the number of errors made by the substructures used to form the concept. The positive examples not covered by the substructure and the negative examples covered by the substructure are considered errors.  $\#PosEgsNotCovered$  is the number of positive examples not covered by the substructure, and  $\#NegEgsCovered$  is the number of negative examples covered by the substructure.  $\#PosEgs$  is the number of positive examples remaining in the training set (positive examples already covered in a previous iteration were removed from the training set), and  $\#NegEgs$  is the total number of negative examples. This

number does not change, because negative examples are not removed from the training set. For hypotheses with the same error, those covering more positive examples are preferred.

```

SubdueCL(Gp, Gn, Limit, Beam)
  ParentList = (All substructures of one vertex in Gp) mod Beam
  repeat
    BestList = {}
    Exhausted = TRUE
    i = Limit
    while ( (i > 0) and (ParentList ≠ {}))
      ChildList = {}
      foreach substructure in ParentList
        C = Expand(Substructure)
        if CoversOnePos(C,Gp)
          then BestList = BestList ∪ {C}
        ChildList = ( ChildList ∪ C ) mod Beam
        i = i - 1
      endfor
      ParentList = ChildList mod Beam
    endwhile
    if BestList = {} and ParentList ≠ {}
      then Exhausted = FALSE
      Limit = Limit * 1.2
    until ( Exhausted = TRUE )
  return first(BestList)
end

```

**Fig. 1. SubdueCL Algorithm.**

The function takes as parameters the positive examples  $G_p$ , the negative examples  $G_n$ , the Beam length (since SubdueCL's search algorithm is a beam search), and a Limit on the number of substructures to include in its search. SubdueCL starts building a ParentList creating a substructure for each vertex in the graph with a different label, but keeping only as many substructures as the length of the Beam. The "*mod* Beam" qualifier means that the lists keep only as many substructures as the Beam size. Each of those substructures in the parent list is then expanded by one edge or one vertex and an edge in all possible ways and evaluated according to equation presented earlier. Those substructures that cover at least one positive example are kept in the BestList, but limited to the Beam size. A ChildList keeps all the substructures that were obtained from the expansion of the substructures in the ParentList and is also limited by the Beam size. According to the evaluation formula, only the best substructure is chosen in each iteration.

### 3. Predictive Toxicology Domain

In this section we describe how SubdueCL was employed to find patterns in the predictive toxicology domain. The experiments use data from the "Predictive Toxicology Challenge 2000 - 2001".



The data contains information about experiments in four classes of animals (male rats MR, female rats FR, male mice MM, or female mice FM). The information is divided in positive and negative examples, where the characteristics of the compounds are described. The information that we considered for each compound was the name (i.e., tr000), class\_MR (classification of this compound for Male Rats), class\_FR, class\_MM, class\_FM, nAmine (number of Amine groups of this compound), nP\_amine, nS\_amine, nT\_amine, nAmine\_salt, nImine, nNitrile, nIso\_cyanate, nNitro, nCarbonyl2, nester, nCarbonic\_acid, nAnhydrid, nAmide, nP\_amide, nS\_amide, nT\_amide, nAldehyde, nAcetale, nKetone, nAlcohol, nFenol, nether, nThio\_ether, nSulfonic\_acid, nSulfonic\_amide, nPhosphate\_ester, nPhosphor\_amidate, nMehyl, nAlkane, nAlkene, nRingstruct, nAromatic, nAro2n2, and nAro2n.

### 3.1. Graph-Based Representation for Predictive Toxicology Domain

Choosing the right knowledge representation is fundamental in any concept learning process. A good representation will favor the identification of patterns that would not be found with a different one. In our research we explore five different representations with non-directed labeled graphs.

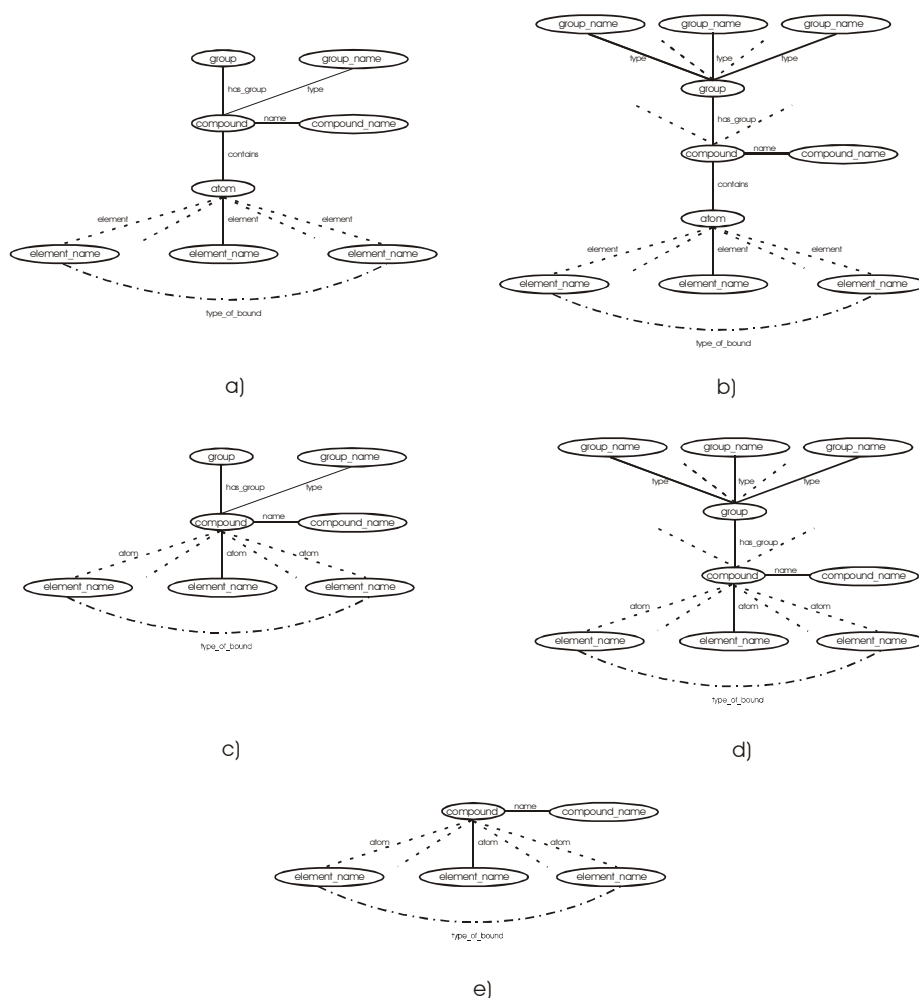
We use one vertex to represent the central node of each compound as we can see in figure 5. The compound's name and other attributes are represented each with a vertex and are linked to the central node with edges in a star topology. We also include information about the compound atoms and its characteristics in different ways (depending on the representation). The bonds linking atoms were represented by edges, one for each bond, where the edge label represents the bond type. The functional groups of the compounds were considered in some of the representations (see figures 5a, 5b, 5c and 5d).

## 4. Experimental Results

In our experiments, we created ten testing files for each class of animals (MR, FR, MM and FM), where positive and negative examples are present. We executed SubdueCL with these files and the five proposed representations.

In figure 3 we show four substructures found by SubdueCL. As we can see, the substructures in figures 3a and 3b have a star topology while the substructures shown in figures 3c and 3d are more structured. Figure 3a shows a substructure saying that a compound that contains two halide groups of type "br" and that also has an alkane group may cause cancer (the rest of the substructures are read in a similar way).

The figure 3c shows a compound with 2 carbon atoms, one hydrogen atom and one bromine atom. This substructure covered 9 positive examples (this compound has a high probability to cause cancer). Moreover, SubdueCL found substructures that covered not only positive examples but also some negative examples. For example, the substructure shown in figure 5d covered 12 positive examples and 1 negative example.

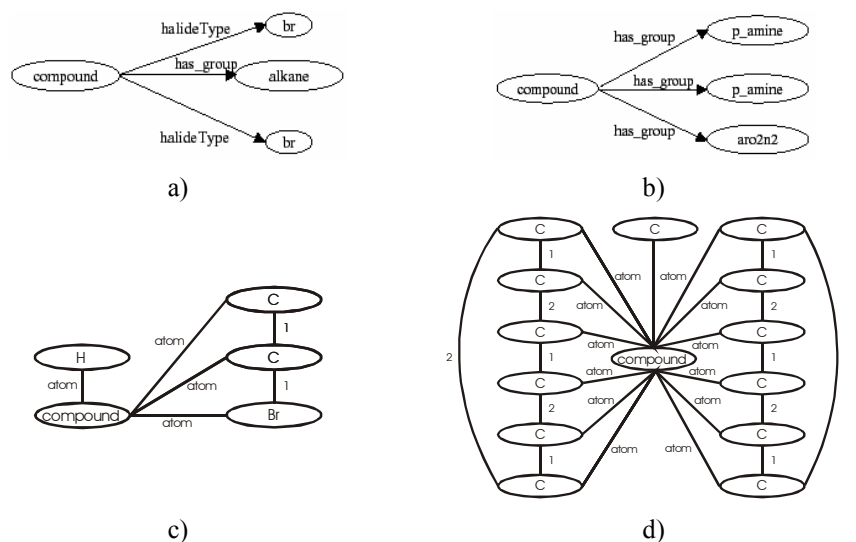


**Fig. 2. Five Different Representations for the Predictive Toxicology Domain.**

## 5. Conclusions

Our results show the variety of patterns that can be found with graph-based representations and the Subdue knowledge discovery system. We explored five types of representations based on star and tree graph topologies. Each of these representations generated interesting results that can be applied to different compounds in the Predictive Toxicology Domain. We want to stress that choosing the right representation will lead to find good quality patterns; this means that while some representations may help to find a pattern others will not allow finding it.

We will continue using graph-based representations because of their flexibility to represent flat and structural domains and at the same time its easiness to be understood by end users. In order to get quality patterns we need to be creative while choosing the right representation. We find graph-based concept learning for data mining as a promising research area that can be used in complex structured domains.



**Fig. 3. Different Types of Compounds Finding in our Experiments.**

## References

1. Cook, D. J.; Holder, L. B. Substructure discovery using minimum description length and background knowledge. *Journal of Artificial Intelligence Research*, 1:231-255, 1994.
2. Cook, D. J.; Holder, L. B.; and Djoko, S. Knowledge Discovery from Structural Data. *Journal of Intelligence and Information Sciences*, Vol. 5, Number 3. (1995). 229-245.
3. R. Chittimoori, J. Gonzalez, and L. B. Holder. Structural Knowledge Discovery in Chemical and Spatio-Temporal Domains. Sixteenth National Conference on Artificial Intelligence, 1999.
4. Srivastava, J. Web Usage Mining: Discovery and Applications of Usage Patterns form Web Data. *SIGKDD Explorations*, Vol. 1, Issue 2 (2000).
5. Rissanen, J. Stochastic Complexity in Statistical Inquiry. World Scientific Publishing Company, 1989.
6. Diestel, Reinhard. *Graph Theory*. Springer – Verlag. Second Edition (2000).
7. J. Gonzalez. Empirical and Theoretical Analysis of Relational Concept Learning Using a Graph-Based Representation. Doctoral Dissertation. University of Texas at Arlington. 2001.
8. J. Gonzalez, Lawrence Holder, Diane Cook. Application of Graph-Based Concept Learning to the Predictive Toxicology Domain. Predictive Toxicology Challenge 2000 – 2001.

# Stereo-sensitive Cortical Regions Identified Using Shape Discrimination from Stereopsis: an fMRI study

Héctor-Gabriel Acosta-Mesa<sup>1</sup>, Nicandro Cruz-Ramírez<sup>1</sup>, John Frisby<sup>2</sup>,  
Ying Zheng<sup>2</sup>, David Buckley<sup>3</sup>, Janet Morris<sup>3</sup>, and John Mayhew<sup>2</sup>

<sup>1</sup> Faculty of Physics and Artificial Intelligence, Department of Artificial Intelligence,  
University of Veracruz, Sebastián Camacho # 5, 91000, Xalapa, Ver. México.  
*{heacosta, nacruz}@uv.mx*

<sup>2</sup> Artificial Intelligence Vision Research Unit, Department of Psychology,  
University of Sheffield, Western Bank, S10 2TP, Sheffield, UK.  
*{j.e.mayhew, j.p.frisby, ying.zheng}@sheffield.ac.uk*

<sup>3</sup> Section of Academic Radiology, Royal Hallamshire Hospital,  
University of Sheffield, Glossop Road, S10 2JF, Sheffield, UK.  
*d.buckley@sheffield.ac.uk, janet.morris@sth.nhs.uk*

**Abstract:** Functional magnetic resonance images (fMRI) were analyzed using digital image processing techniques to investigate the cortical regions involved in stereoscopic vision using red/green anaglyphs to present random dot stereograms. The methodology of fMRI data analysis is introduced. In the experiment the subjects were instructed to follow the path of a square defined by depth and moving in the horizontal plane contrasted with a similar sized square defined by a slight difference in luminance. Three main regions were identified V3A, V3B and BA7. In agreement with other fMRI studies, our results suggest that areas V3A and BA7 (precuneus) are involved in stereo disparity processing. We hypothesise that the activation of the V3B region was produced by the second order motion component induced by the spatio-temporal changes in disparity (stereoscopic motion).

**Key words:** neuroimaging, physiological signals, digital image processing, functional magnetic resonance imaging, stereo vision.

**Resumen:** Imágenes de resonancia magnética funcional (fMRI) fueron analizadas usando técnicas de procesamiento digital de imágenes para investigar las regiones del cerebro involucradas en visión estereoscópica. La metodología de análisis de datos es presentada. Los estímulos estereoscópicos fueron generados utilizando estereogramas anáglifos rojo/verde de puntos aleatorios. En el experimento los sujetos fueron instruidos para seguir visualmente la ruta de un cuadrado definido por profundidad desplazándose en el plano horizontal. Como contraste se utilizó una condición en la que el cuadrado fue definido por luminosidad. Tres regiones fueron identificadas: V3A, V3B y BA7. En concordancia con otros estudios similares, nuestros resultados sugieren que las regiones V3A y BA7 (precuneus) están involucradas en el procesamiento de disparidades estereoscópicas. Nuestra hipótesis es que la activación de la región V3B fue producida por la componente de movimiento de segundo orden inducida por los cambios espacio-temporales en las disparidades (movimiento estereoscópico).

## 1. Introduction

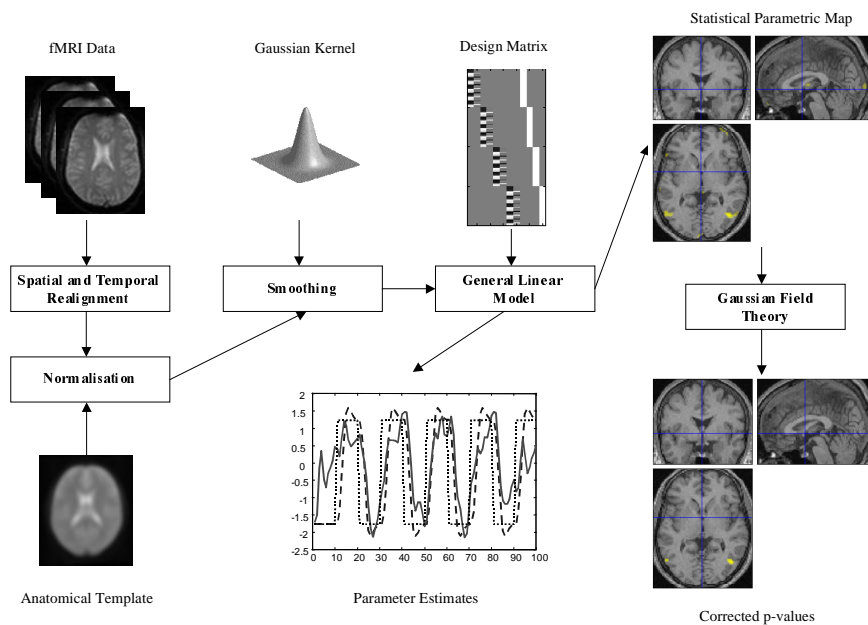
Although many psychophysical studies have investigated how the human brain computes stereoscopic information, it is uncertain which cortical areas are involved in its implementation. Some electrophysiological studies in monkeys report the sensitivity of V1 to absolute disparities, suggesting that this area could be a preliminary stage of processing for stereo information [1]. MT/V5 in monkeys shows a columnar organisation tuned for disparity [2]. MT/V5 in human brains has been widely reported as a motion sensitive area [3]. Given the similarity between the visual system of the monkey and the human, it is possible that V5 in human brains is involved in the processing of stereo information. Recently, modern non-invasive neuroimaging techniques, eg PET and fMRI, have been used to explore the functional anatomy of stereoscopic vision in humans. The results of these studies suggest that stereo disparity processing may be widespread over a network of cortical regions in the occipital and parietal lobes, including V1, V2, V3, V3A, V3B, [4]. However, there is no general agreement about the cortical regions selective to stereo disparities or the specific role that each of these has in the perception of depth [6]. The main goal of the present paper is to show the methodology of fMRI data analysis to investigate the cortical regions sensitive to stereo disparities. Stereo stimulus that maximises the attentional demands was used to avoid adaptation. Under this principle, one experiment was developed using functional magnetic resonance imaging to identify stereo sensitive regions stimulated by random dot anaglyph stereograms.

## 2. Materials and methods

Ten healthy subjects, nine right-handed and one left-handed volunteers (7 female, 3 male) aged from 20 to 30 years participated in the experiment. All subjects gave informed written consent. The stereo acuity of the subjects was measured using a stereo vision test (RANDOT SO-002), all of them were below 40 sec of arc. The subjects were given a preliminary practice session outside the magnet to become familiar with the visual stimulation. *Stimulus presentation:* Subjects lay on their backs in the magnet. They wore red/green anaglyph glasses and looked via a mirror angled at  $\sim 45^\circ$  from their visual axes at a back illuminated screen located just outside the magnet. The viewing distance was 2.4 m. Stimuli were projected on to the screen using an EPSON (EMP-7300) projector driven by a 3G Mac running Psychophysics Tool Box ver. 2.44 under MATLAB ver. 5.3. Although the stimuli were displayed at a video frame rate of 60 Hz, the image was only updated on every 10th frame, producing an effective frame rate of 6 Hz. *Data acquisition and analysis:* Subjects were scanned in a 1.5 T whole-body MRI scanner (Eclips Marconi) with BOLD contrast echo planar imaging (TR=3s, TE=40 ms, 128 x128 voxel, voxel size 1.875 x 1.875 x 4 mm.). Thirty two slices covering the whole brain were acquired. The data was pre-processed and analysed using SPM99 (Wellcome Department of Cognitive Neurology). The first five scans of each run were discarded to exclude magnet saturation artefacts.

## 2.1 Algorithmic analysis for statistical parametric mapping.

Statistical Parametric Mapping (SPM) is a technique for fMRI data processing [7]. Its main objective is to generate brain maps of cortical activations produced by controlled experimental stimulation. The SPM data analysis can be divided in two general procedures, the pre-processing of the images and its statistical analysis. Each of these processes is equally important and both are detailed in the following sections. A graphical representation of the data analysis is shown in the figure 1. It is important to clarify that the purpose of the present section is only to introduce the concepts of SPM analysis. For a detailed explanation on the mathematical formalism and the use of the software consult [5].



**Figure 1** Statistical parametric mapping analysis of fMRI data.

*Spatial alignment:* One of the basic assumptions in the analysis of brain volumes is the fact that the location of a voxel refers to the same anatomical location over time. The violation of this assumption usually introduces artefactual signals. Head movements are the most common source of this type of noise. Despite the minimising of the effects by the use of head holders or bite bars, it is necessary to correct the images for small head movement. This spatial correction is usually made by aligning all the volumes with respect to a reference volume (e.g. the first volume).

*Temporal alignment:* Because the multi-slice acquisition is made in sequential order, slices of the same volume are acquired at slightly different times. This correction is usually made by interpolating the data in the slices (of the same volume) with respect to a reference slice. This procedure is also called slice timing correction.

*Normalization:* An important feature in the analysis of results is to compare the activations between subjects. Although the human brain differs in size and folding structure, it is possible to approximate it into a standard template. This normalisation ensures the consistency between space and anatomical structures. SPM normalises to a MNI (Montreal Neurological Institute) space.

*Smoothing:* The smoothing procedure has the effect to reduce noise in the data. It is performed by applying a low-pass filter on the images, i.e. the images are convolved with a gaussian kernel. Smoothing also helps to create a normal distribution on the noise, characteristic that is important in the p-value correction procedure described in a previous section.

*Statistical analysis:* SPM statistical analysis is performed in a voxel-based fashion, i.e. the statistical tests are developed for every single voxel. The statistical significance of the activations is made using multiple linear regression which is a special case of a General Linear Model (Tabachnik and Fidell 2001).

*General linear model:* The General Linear Model (GLM) is a framework to develop multivariate regression. In fMRI time series, the fundamental idea is to use it to represent the time series of each voxel ( $y_i$ ) as a linear combination of the input sequences used in the experimental stimulation ( $x_n$ ), plus an error disturbance, for example:  $y_i = \theta_{1i} x_1 + \theta_{2i} x_2 + \dots + \theta_{ni} x_n + e_i$ . In matrix notation:  $\mathbf{y} = \theta \mathbf{x} + \mathbf{e}$ .

Each column of matrix  $\mathbf{x}$  (called design matrix) is an explanatory variable (predictor or regressor) of  $\mathbf{y}$ , i.e. each of these columns represents a stimulus effect of interest. In order to have a better approximation of the BOLD activations, the design matrix is constructed by convolving the input stimulation sequences (e.g. box car) with a canonical hemodynamic response function. It is important to clarify that the design matrix can contain other explanatory variables that could be present in the data. For example, heart beat pulsation, breathing, or signal trends introduced by magnetic fluctuations. The parameters  $\theta$  associated with each column of the design matrix represents the relatively strength of each predictor on the explanation of the dependent variable. These parameters are usually assessed using standard least squares in the following way:  $\theta = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y}$ . The significance of each parameter is assessed in terms of the Student's t-test statistics.

*Gaussian field theory:* The basis idea of GFT correction is the estimation of the number of false positive blobs (activated regions) in the volume. This estimation is made using the Euler characteristic (EC). Although it is not exactly the same, the EC can be thought of as the number of false positive blobs expected after the volume has been thresholded at given z-score.

### 3. Global stereo tracking experiment

The experiment was designed to activate stereo sensitive regions by requiring the subjects to perform a task of global stereo tracking (GST). Random dot stereograms were used to define a square region moving horizontally in front of the background from left to right and vice versa. The subjects were instructed to perform pursuit eye movement to follow the path of the square with their eyes. The performance of the task depended on the maintenance of the perception of depth defined by stereo disparities.

Experiment design: Subjects were given a sequence of 4 scans (sequences) each lasting 5.15 min. (10 epoch) with a 5 min. interscan interval to permit subjects to rest. One hundred image volumes were obtained in each sequence. Each condition lasted 30s. giving 10 multislice volumes per condition (TR=3s.). Each scan consisted of alternating epochs in a boxcar configuration. A dummy condition of a blank screen was presented during the first 15s. (5 scans) of each run which were excluded to control for magnetic saturation effects. The display contained 1,024 dots (with radius 0.1 deg. and zero disparity) distributed over the screen (mean dot density 1.5 dot deg<sup>-2</sup>). The subjects were instructed to fixate on the right superior corner of a square (5.23 deg. side long) moving laterally across the screen (13 deg. field of view). The square was moved from left to right and vice versa at a constant speed (2.19 deg.sec<sup>-1</sup>), each time that the square reached one edge of the screen it changed its direction. Two modalities were used to define the square, each one representing an experimental condition:

a). Two dimensional tracking (2D): The square was luminance defined, its luminance (8.56 cd/m<sup>2</sup>) was lower than the background (18 cd/m<sup>2</sup>). This condition was used as a base line. b). Three dimensional tracking (3D): The square was depth defined (red/green anaglyph stereogram), presented 0.3 deg at the front of the background (zero disparity). The square moved horizontally in the plane X/Y, not in plane Z (motion in depth). Stereo-sensitive regions were identified by comparing the activation during the 2D control condition with the neural activity during the 3D stereo tracking condition. The paper which explains in detail this experiment was submitted for publications in the "Computación Científica e Informática Médica" WorkShop of the Mexican International Conference in Computer Science (ENC-04).

### 4 Results

Stereo-sensitive regions were identified using a contrast that compared the activation produced by the moving square defined by depth with that produced by the moving square defined by luminance. The results show consistent activation in areas V3A, V3B and parietal cortex. Area V3A in the right hemisphere was activated by depth in 7 of the 10 subjects. In two of these subjects there was bilateral activation. Area V3B in the right hemisphere was activated in 8 of the 10 subjects. In five of these there was bilateral activation. One subject showed activation only in left V3B. Area BA7 (precuneus) in the right hemisphere was activated in 3 of the 10 subjects. Two subjects had left hemisphere activation. There was no bilateral activation. Activation in the right superior parietal area was found in 2 other subjects.



## 5. Conclusions and Future Work

The results of our experiment reveal three main regions sensitive to stereoscopic depth: V3A, V3B and precuneus. In contrast to the evidence from physiological experiments in monkeys [2] we did not find evidence that V5 was involved in the processing of stereoscopic information. Consistent with the results of other studies, we find evidence that V3A and precuneus (BA7) showed sensitivity to stereo disparities [4,6]. It is unlikely that these activations are due either to different patterns of eye movements or to stronger attentional engagement because the control condition provided identical requirements for these parameters and similar areas of activation were found in a second experiment (not yet published) in which no eye movements were required to perform the task. Our results suggest a network involving the cortical areas V3A, V3B and precuneus in the processing of stereo disparities. The high proportion of activations located in the right hemisphere supports the notion of right cerebral dominance in stereo vision [9]. Contrary to the results reported in studies with monkeys, our experiments did not reveal any evidence of the sensitivity of V5 to stereo disparity processing. The present work shows the methodology to analyze fMRI data and the potentiality of this technique to explore cognitive process. As the objective of the presented experiment was to investigate the cortical areas selective to stereoscopic information, an experiment designed to test the hypothesis of the selectivity of the V3B region to stereoscopic motion is for future work.

## 6. References

- [1] B. G. Cumming and A. J. Parker, "Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity," *J Neurosci*, vol. 19, pp. 5602-18., 1999.
- [2] G. C. DeAngelis and W. T. Newsome, "Organization of disparity-selective neurons in macaque area MT," *J Neurosci*, vol. 19, pp. 1398-415., 1999.
- [3] R. B. Tootell, J. D. Mendola, N. K. Hadjikhani, P. J. Ledden, A. K. Liu, J. B. Reppas, M. I. Sereno, and A. M. Dale, "Functional analysis of V3A and related areas in human visual cortex," *J Neurosci*, vol. 17, pp. 7060-78., 1997.
- [4] B. Gulyas and P. E. Roland, "Binocular disparity discrimination in human cerebral cortex: Functional anatomy by positron emission tomography," *Proc. Natl. Acad. Sci. USA (Neurobiology)*, vol. 91, pp. 1239-1243, 1993.
- [5] Veltman, D. and C. Hutton (2001). "SPM99 Manual." Wellcome Department of Cognitive Neurology.
- [6] R. D. Freeman, "Stereoscopic vision: Which parts of the brain are involved?," *Curr Biol*, vol. 9, pp. R610-3., 1999.
- [7] K. J. Friston, H. A. P., W. K. J., P. J. P., F. C. D., and F. R. S. J., "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, pp. 189-210, 1995.
- [8] Tabachnik, B. G. and L. S. Fidell (2001). "Using multivariate statistics." Allyn and Bacon Fourth Edition: 764-771.
- [9] I. P. Howard and B. Rogers, "Binocular vision and stereopsis," Oxford University press., 1995.

# **Computación clínica e informática médica**

Claudia Feregrino Uribe  
Oscar Yáñez Suárez  
(Editores)



## **Introducción**

La computación clínica y la informática médica conciernen a la creación y validación de modelos de aprovechamiento de los conocimientos e información disponibles dentro de la medicina y la ingeniería biomédica. Los investigadores en estas áreas estudian y diseñan métodos para adquirir, representar, procesar y manejar dichos elementos con el fin de promover el bienestar asociado al cuidado de la salud.

La permanente demanda de aplicaciones prácticas de las tecnologías computacionales para la atención integral del paciente, el apoyo al quehacer del médico y la gestión de servicios clínicos convoca la participación de grupos de trabajo interdisciplinarios. Para que dichos grupos se conformen, es necesario que se identifiquen problemas y áreas de interés compartido. Por ello pretendemos, con la realización de este Taller, construir un foro para que los grupos nacionales e individuos interesados en los ámbitos de la computación clínica y la informática médica, den a conocer el trabajo que realizan dentro de sus áreas de especialidad y presenten las oportunidades de formación, desarrollo y colaboración asociadas a su labor.

Esta primera edición del Taller agrupa trabajos que tocan, como era de esperarse, temáticas diversas: sistemas de gestión de servicios clínicos, comunicación y análisis de imágenes médicas, realidad virtual, y procesamiento de señales. Esperamos que esta aproximación plural a nuestro campo de referencia enriquezca a todos los participantes y propicie un futuro de desarrollo conjunto y colaboración.

## **Comité Científico**

Copresidentes:

Claudia Feregrino Uribe, Gr. Cómputo Reconfigurable, INAOE, [cferegrino@inaoep.mx](mailto:cferegrino@inaoep.mx)

Oscar Yáñez Suárez, Labo. de Neuroimagenología, UAM, [yaso@xanum.uam.mx](mailto:yaso@xanum.uam.mx)

Mercedes Jatziri Gaitan Gonzalez, L. Fisiología Humana, UAM, [mjgg@xanum.uam.mx](mailto:mjgg@xanum.uam.mx)

Gerardo Gabriel Gentiletti Faenze, L. Neuroimagenología, UAM, [ggentiletti@ieee.org](mailto:ggentiletti@ieee.org)

# Aplicación de una herramienta de realidad virtual colaborativa y adaptación de modelos gráficos en 3D para el apoyo del diagnóstico médico de lesiones óseas

*Lenin Alejandro Cervantes Medina*  
Facultad de Telemática  
Universidad de Colima  
Colima, Col., México  
[lenin@ucol.mx](mailto:lenin@ucol.mx)

*Miguel Ángel García Ruiz*  
Laboratorio de Realidad Virtual  
Universidad de Colima  
Colima, Col., México  
[mgarcia@ucol.mx](mailto:mgarcia@ucol.mx)

**Abstract:** The purpose of this paper is to show an exploratory study about the application of collaborative virtual reality to support physicians' analysis of bone injuries, where the main objective is to facilitate the physicians' communications and analysis about this problem. When a person suffers a bone fracture or another similar injury, the first step that a doctor do is to take a X-ray and then give a diagnosis. However, sometimes the doctor do not have the necessary knowledge and time, therefore his/her diagnosis is not the better. In addition, sophisticated technology like computerized tomography (CT) is used to help giving a better diagnosis. We propose the use of a collaborative virtual environment where 3D images of a bone fracture taken with a CT could be manipulated and analyzed by some doctors, to reach an optimal diagnosis.

**Resumen:** Esta ponencia muestra un estudio exploratorio de la aplicación de realidad virtual colaborativa como soporte para los doctores en el análisis de lesiones óseas, en donde el objetivo principal sea facilitar la comunicación y análisis de dicho problema. Cuando una persona sufre una lesión ósea como puede ser una fractura, el médico obtiene una radiografía de rayos X y en base a ella da un diagnóstico; pero en algunas ocasiones el doctor no cuenta con los conocimientos o tiempo necesarios, por lo que su diagnóstico puede no ser el correcto. Es así como en esta investigación se hace uso de tecnología sofisticada como es la tomografía axial computarizada (TAC) para ofrecer un mejor diagnóstico. Proponemos la aplicación de un ambiente virtual colaborativo con gráficas 3D de una fractura tomadas con un TAC, para su manipulación y análisis por varios doctores donde cada uno exprese su opinión para ofrecer un diagnóstico más óptimo.

**Palabras clave:** Realidad virtual, trabajo colaborativo soportado por computadora, lesiones óseas, tomografía axial computarizada, red de computadoras.

## 1. Introducción

Actualmente cuando una persona tiene un problema de salud y va a visitar al médico, un solo doctor es el que hace el diagnóstico; pero que sucede cuando el percance es más serio, en donde una buena alternativa sea tener la opinión de varios doctores para

dar la mejor solución. Entonces la pregunta es ¿de qué les sirve a los pacientes el que varios doctores y no uno sólo de un diagnóstico, si todos saben lo que se tiene que hacer?.

Tratando de dar solución a esta incógnita podemos cuestionarnos, es verdaderamente factible que varios doctores con diversos conocimientos analicen un mismo problema y aporten algo al caso cuando así se requiera requiera.

La respuesta es sí pero ahora surge una nueva pregunta ¿cómo?. En busca de dar solución a este problema, en los últimos años se ha incrementado el interés en una importante rama de la computación llamada "Realidad Virtual", o "RV", y mas aún si a esta herramienta le agregamos otro tópico como es el Trabajo Colaborativo Asistido por Computadora [1, 2, 3].

Se puede definir a la realidad virtual como la ilusión de interactuar en un ambiente en tercera dimensión simulado por computadora, donde los objetos que se encuentran dentro de este entorno pueden ser manipulados por el usuario [4].

Alan Wexelblat comenta en su libro: “El Trabajo Colaborativo Soportado en Computadora (CSCW por sus siglas en inglés) es relativamente una disciplina nueva” [5].

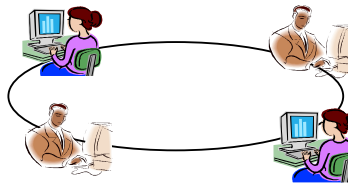


Fig. 1. CSCW.

Pensando en ello se llega a lo que se conoce como “RV Colaborativa” que ahora en nuestros días esta cobrando mucha fuerza. De tal forma que el enfoque que se dará a la RV Colaborativa en este caso será hacia la medicina, específicamente en lesiones óseas.

Sherman y Judkins comentan que la realidad virtual se ha empezado a usar en las escuelas de medicina, principalmente en anatomía y cirugía reconstructiva. Algunas instituciones como la Universidad de Washington en U.S.A. usan la RV como herramienta de apoyo en sus clases. En esta universidad se desarrolló un cadáver virtual donde los estudiantes pueden practicar cirugías [6].

## 1.1 Problemática

En ciertos casos las decisiones que toman los doctores al momento de diagnosticar a un paciente con una fractura podrían ser mejores si se consultara a otro médico y se tuvieran dos o más opiniones.

El problema radica en que los doctores a quien se acudiría para que dieran una segunda opinión se encuentran en diferentes lugares debido a que son especialistas y por lo general en una institución médica solo hay uno.

## 2. Estado del Arte

De acuerdo a [7], existen varios proyectos en donde se ha utilizado realidad virtual con CSCW entre los que se encuentra: GRACILE (Japanese GRAMmar Colaborative Intelligent agents LEarning). El propósito es desarrollar agentes inteligentes que propicien la colaboración para el intercambio de conocimientos entre los participantes. Los grupos que soporta son de dos a cuatro estudiantes, en una red LAN a través de Ethernet. Los usuarios interactúan escribiendo puntos de vista desde diferentes lugares remotos.

Suárez Quirós, García Díaz, et al. [8] desarrollaron una herramienta que permite visualizar estructuras anatómicas de imágenes bidimensionales de TAC así como de implantes en un entorno colaborativo. Esto permite que un grupo de usuarios analicen la imagen médica desde diferentes lugares remotos conectados a través de Internet. El entorno cuenta con una interfaz sencilla que facilita la navegación en el entorno.

El Ohio Supercomputer Center está desarrollando un ambiente virtual realista para el análisis del hueso temporal. Este proyecto utiliza modelos gráficos de tomografías y de resonancia magnética [9]. Este proyecto de investigación contempla la visualización en estéreo del modelo del hueso.

## 3. Metodología

Se implementó un entorno virtual donde se visualizaron imágenes tridimensionales de fracturas obtenidas de un TAC a través de una herramienta de RV colaborativa ya desarrollada y se realizó una prueba piloto de usabilidad para conocer el impacto que tenía entre los participantes (ver Fig. 4).

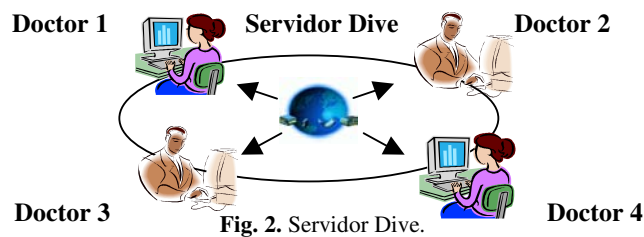
En el caso de las imágenes, algunos aparatos de tomografía manejan la tecnología Direct3D la cual genera archivos de imágenes en 3D con extensión .x los cuales están constituidas por una cabecera y coordenadas en x,y,z., éstos archivos pueden ser grabados en un CD desde el mismo TAC. Una vez que se cuenta con los archivos, tienen que ser convertidos a formato VRML 1.0 lo cual se realiza con el software Deep Exploration de Right Hemisphere.

Cabe mencionar que para la prueba piloto que se realizó se hizo uso de un modelo desarrollado en Maya 5.0 de un hueso, debido a que por diversas circunstancias no se pudieron conseguir imágenes reales de un TAC. Pero lo expuesto en el párrafo anterior esta debidamente fundamentado para que se pueda realizar.



El desarrollo del ambiente virtual se realizó en DIVE (Distributed Interactive Virtual Environment) el cual es un software libre desarrollado en el Instituto Sueco de Computación. Este software permite:

- o Desarrollar e interactuar con entornos virtuales.
- o Hacer conexiones remotas mediante un servidor proxy; esto es, poder diagnosticar desde cualquier parte del mundo a través de Internet.
- o Cargar y manipular gráficas 3D en formato VRML 1.0 que sería el formato al cual se exportaría la imagen de la lesión ósea obtenida del TAC.
- o Que los usuarios intercambien sus puntos de vista mediante un chat o su voz.



La prueba piloto de usabilidad se realizó con un grupo de 5 estudiantes de medicina de onceavo semestre de la Universidad de Colima que se encuentran realizando su servicio social. Se consideró que tuvieran este perfil debido a que están próximos a egresar de su carrera y ya contaban con un cierto grado de conocimientos acerca del tema.

El material que se utilizó fueron computadoras:

- Pentium 4 de 2.66 GHz  
512 MB en RAM y tarjeta de video Trident CyberALADDIN con 32 MB
- AMD Athlon 2.2 GHz  
1 GB en RAM y tarjeta de video ATI RADEON 9700 PRO con 128 MB
- Audífonos (diademas) con micrófono

La prueba se llevó a cabo en las Instalaciones del Laboratorio de Realidad Virtual de la misma Universidad con el grupo de personas que arriba se describe, para que hicieran uso del entorno. Los participantes estuvieron interactuando en el ambiente virtual durante aproximadamente 30 minutos durante los cuales intercambiaron puntos de vista mediante su voz y un chat, de una lesión ósea que se les presentó. Al término de la prueba se les aplicó un cuestionario de usabilidad para recabar información.

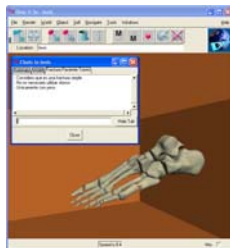


Fig. 3. Imágenes de la lesión ósea analizada por los participantes, en este caso una fractura en el dedo meñique del pie izquierdo.

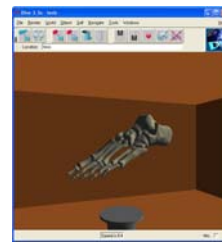
Esta prueba resultó muy interesante ya que los futuros doctores comentaron que no tuvieron ningún problema en interactuar dentro del ambiente virtual además de que consideraron que el entorno resultaría de gran ayuda debido a que estarían de acuerdo en intercambiar puntos de vista para llegar a un mejor diagnóstico y ofrecer un tratamiento mejor.

El proceso de analizar la fractura se realizó de dos maneras:

- ❖ Mediante un grupo de chat llamado *piloto* donde los participantes ingresaron y escribieron sus puntos de vista acerca del problema que se les presentó.
- ❖ Uso de audífonos (diademas) con micrófono, esto es, intercambiando sus opiniones a través de su propia voz.



**Fig. 4.** Ventana de chat donde interactuaron los doctores.



**Fig. 5.** Interacción por medio de audio por parte de los participantes.

#### 4. Resultados preliminares

De acuerdo a los datos recabados en la prueba con los doctores se pudo observar que al principio se mostraron un poco impacientes dentro del entorno virtual pero conforme se les fue explicando cual era la finalidad y lo que tenían que hacer se fueron entusiasmando. También se pudo observar que al iniciar la prueba los participantes hicieron mas uso del chat que del audio, aunque cabe aclarar que ya para finalizar ésta, casi todos los participantes ya se encontraban tanto hablando como haciendo uso del chat por igual.

Según los resultados de los cuestionarios de usabilidad aplicados, la mayoría de los participantes consideró que la comunicación por voz fue la mejor manera de intercambiar información de la lesión ósea. Todos los participantes concordaron en que el ambiente virtual ayudo mucho a mejorar el diagnóstico del problema que se les presentó. Algo a considerar es que en su totalidad los participantes alguna vez en su vida habían jugado videojuegos donde se involucraban gráficas en 3D lo que posiblemente ayudó a que no tuvieran problemas en interactuar dentro del entorno virtual.

#### 5. Conclusiones y perspectivas

Considerando todo el proceso que se realizó a lo largo de esta investigación desde investigar que se había ya realizado acerca de este tema hasta el desarrollo del ambiente virtual, en donde los doctores en una de las pruebas analizaron una lesión

ósea y donde se pudo constatar mediante sus comentarios que el entorno virtual de diagnóstico se les hizo algo muy novedoso y muy útil, se concluye que se logró el propósito planteado, el cual era poner a disposición de un grupo de doctores un ambiente virtual colaborativo donde se pueda analizar una fisura o fractura a distancia sin la necesidad de salir de sus lugar de trabajo. También se puede concluir que la realidad virtual puede ser una herramienta muy útil en el área médica debido a la amplia gama de usos que se le pueden dar todo esto en base a la información recabada por lo cuestionarios aplicados. Debido a que la prueba que se realizó fue piloto, en un futuro se tiene pensado realizar una prueba con más participantes para obtener datos más completos.

## Referencias

- [1] García, Miguel: “*Aplicaciones de la Realidad Virtual en la Educación. Breve panorama general* “ <<http://www.cogs.susx.ac.uk/users/miguelga/realid.htm>> (Fecha de publicación: 9 de Noviembre de 1998)
- [2] Schneiderman, B. *Designing the User Interface*. Addison-Wesley, Reading, 1998.
- [3] Mike Fraser, Tony Glover, Ivan Vaghi, Steve Benford, Chris Greenhalgh, Jon Hindmarsh, Christian Heath. *Revealing the realities of collaborative virtual reality*. Proceedings of the third international conference on Collaborative virtual environments, 2000.
- [4] Sherman R. W., Craig, A. B. *Understanding Virtual Reality: Interface, Application and Design*. Morgan Kaufmann Publishers, San Francisco, CA, 2003.
- [5] Wexelblat, Alan., MIT Media Lab. Cambridge, Massachussets, USA. [wex@media.mit.edu](mailto:wex@media.mit.edu)
- [6] Sherman B., Judkins P. *Glimpses of heaven, visions of hell: virtual reality and its applications*, Hodder & Stoughton, London, 1994.
- [7] Santacruz Valencia, Liliana Patricia (1998). *CSCW y enseñanza*. [Consultada el 15 de julio de 2004]. Disponible en: <http://www.it.uc3m.es/liliana/paginas/masinfo/cscw.htm>
- [8] Suárez Quirós, Javier; García Díaz, Rafael Pedro (2003). *Desarrollo de un entorno gráfico colaborativo para el diseño y fabricación de implantes quirúrgicos a medida*. [Consultada el 15 de julio de 2004]. Disponible en: <http://www.adm-ingegraf.unina.it/Italiano/Sommari/ADM043-SuarezQuiros.pdf>
- [9] Bryan, J., Don Stredney, Greg Wiet, Dennis Sessanna. *Virtual Temporal Bone Dissection, a Case Study*. 12th IEEE Visualization 2001 Conference (VIS 2001) October 24 - 26, 2001.

# **Apoyo a las actividades médicas a través de Servicios Web basados en HL7/CDA**

Pedro César Santana Mancilla, Leonardo Galicia Jiménez,  
Ana Isabel Martínez García, José Antonio García Macías

Centro de Investigación Científica y de Educación Superior de Ensenada, CICESE  
e-mail : { psantana, lgalicia, martinea, jagm }@cicese.mx  
Km. 107 Carretera Tijuana-Ensenada Apdo Postal 2732, CP 22860  
Ensenada, B.C, México.

**Abstract:** The introduction of new Information Technology, and standards such as those proposed by the Health Level Seven (HL7), enable the interoperability between software applications from different health organizations. This paper introduces an architecture for supporting medical activities, based on a framework that facilitates the use of standards and technology available to the health sector. This work uses HL7 to develop a central architecture and test it with a client Web developed for a Tablet PC and a light client for a Personal Digital Assistant (PDA).

**Resumen:** La aparición de nuevas tecnologías de la información, así como estándares y acuerdos, como el propuesto por la organización *Health Level Seven* (HL7), permiten la interoperabilidad entre aplicaciones de sistemas de salud en distintas partes del mundo. El presente artículo introduce una arquitectura de soporte a las actividades que pretende facilitar el uso de estándares y tecnologías disponibles hacia el sector salud, especialmente hacia instituciones públicas. El trabajo parte del uso del estándar HL7 para generar una arquitectura central y probar tal arquitectura en un cliente Web desarrollado para *Tablet PC* y en un cliente ligero para un asistente digital personal (PDA por sus siglas en inglés *Personal Digital Assistant*).

**Palabras Clave:** Expediente, Médico, Electrónico, Movilidad, HL7

## **Introducción**

En el entorno clínico actual es clave el soporte a las actividades médicas por medio de sistemas de información clínica bien diseñados, compuestos por aplicaciones que deben estar construidas con base a estándares, de manera que aseguren que los usuarios del sistema informático clínico puedan compartir fácilmente la información a través de toda su organización y en un momento determinado compartirla con otras organizaciones. De esta manera los hospitales, los centros de salud, clínicas privadas y cualquier tipo de organización de asistencia médica pueden capturar información vital sobre los pacientes a través de aplicaciones que permitan validar la información para

tener mayor precisión, y guardarla en un archivo que cumpla con las características del estándar propuesto por la arquitectura de documentos clínicos (CDA por sus siglas en inglés - *Clinical Documents Architecture*), que se puede leer y utilizar en otros sistemas compatibles con este último.

Los médicos de hoy en día suelen encontrarse en la situación de tener que introducir los mismos datos sobre sus pacientes desde varias fuentes. Es un proceso pesado, tanto para los médicos como para los pacientes a los que atienden, e incluso a veces existen omisiones en los historiales de estos últimos. Al utilizar un sistema que les permita administrar la información de los pacientes, los médicos estarán capacitados para simplificar el proceso de recuperar la información que requieran rápidamente, por lo que pueden enfocarse en proveer un mejor servicio y confiar en la disponibilidad de la información en el momento apropiado.

*Health Level Seven (HL7)* [1] es una organización desarrolladora de estándares dirigidos a permitir la interoperabilidad entre aplicaciones heterogéneas en el ámbito de la salud. Esta desarrolló CDA basado en el lenguaje de marcado extendido (XML por sus siglas en inglés - *eXtended Markup Language*) [3] la cual anteriormente era conocida como arquitectura de expedientes de pacientes (PAR por sus siglas en inglés - *Patient Record Architecture*) [2][3], CDA provee un modelo de intercambio de documentos y facilita a las industrias de la salud la realización de un expediente médico electrónico [4].

Por otro lado los servicios Web son una tecnología emergente que permite la utilización del software como servicios a través de Internet. Proveen servicios a sus consumidores a través de un protocolo basado en XML llamado protocolo de acceso simple (SOAP por sus siglas en inglés - *Simple Access Protocol*), el cual provee un mecanismo simple y ligero para intercambiar información estructurada en un ambiente distribuido. La interfaz de los servicios Web es descrita en un documento XML llamado lenguaje de descripción de servicios Web (WSDL por sus siglas en inglés *Web Service Description Language*), para que se puedan construir aplicaciones clientes usando los servicios proveídos por los servicios Web.

El trabajo presentado en este artículo consiste en una arquitectura de soporte a las actividades médicas basada en un marco de clases que utiliza los estándares HL7 y servicios Web, ambos usando XML y particularmente dirigirlos hacia instituciones públicas, de manera que, aprovechando el estándar HL7, las tecnologías y poca infraestructura tecnológica disponibles, se pueda ofrecer un entorno médico electrónico que de soporte a las diversas actividades que tales instituciones de salud requieren en su haber diario mejorando de esta forma la calidad de sus servicios.

Este artículo se encuentra organizado en cinco secciones. La primera presenta la motivación para el desarrollo del presente trabajo. Enseguida se define la arquitectura del sistema abordando la implementación y su funcionalidad, así como la tecnología utilizada. La tercera sección muestra escenarios de uso de cómo la arquitectura desarrollada ofrece una solución a un problema determinado y como esta podría extenderse a otros casos. Finalmente se presentan las conclusiones y trabajo futuro.

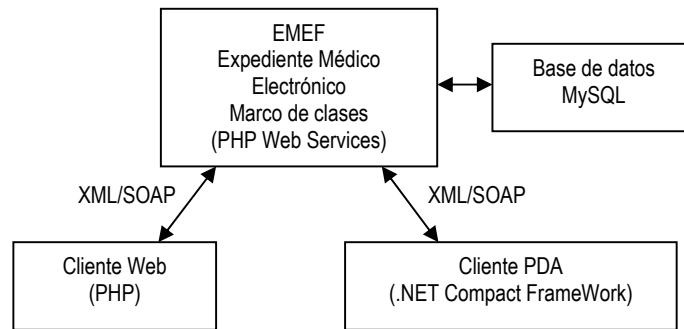
## 1. Motivación

La motivación de este trabajo surge a partir del estudio de los procesos de dos áreas de trabajo de una institución pública de salud, específicamente sobre el área de medicina física y rehabilitación y la de radiología e imágenes. Estos trabajos se realizaron utilizando técnicas de la ingeniería de procesos; mediante la metodología para análisis y diseño de procesos (PADM por sus siglas en inglés *Process Analysis and Design Methodology*) [6].

Durante el estudio de los procesos en estas áreas se pudieron detectar que algunos de ellos se realizan de manera similar, y que carecen de una infraestructura de tecnología de información en común. Estos procesos hacen uso de expedientes clínicos los cuales son entregados y administrados de manera centralizada por una entidad llamada Archivo Clínico. Una vez que los expedientes clínicos de los pacientes son entregados a las áreas solicitantes, ninguna otra área tiene acceso a los expedientes mientras estos se encuentran fuera de archivo clínico, no es sino hasta que estos expedientes son entregados por las áreas que los solicitaron y regresen nuevamente al control de archivo clínico, que el resto de las áreas pueden consultarlos. Lo anterior conlleva que la información no se encuentre siempre disponible para los solicitantes, lo que ocasiona que las áreas estudiadas creen duplicados de la información concerniente a ellas, creando de esta forma sub expedientes clínicos para poder consultar y realizar anotaciones de los pacientes que están bajo su responsabilidad, ocasionando a su vez inconsistencia de la información entre ambos expedientes. El escenario expuesto ofrece un área de oportunidad rica para introducir un soporte electrónico de expediente médico como el propuesto por HL7, de forma tal, que los estándares para lograr esto se encuentran disponibles. De acuerdo al problema planteado se diseñó la arquitectura de servicios y las aplicaciones que dan soporte a los estándares antes descritos.

## 2. Arquitectura del sistema

La arquitectura desarrollada se basa en diversos estándares tecnológicos y está centrada en un marco de clases con el paradigma cliente-servidor utilizando servicios Web. La idea principal de este trabajo fue desarrollar un servidor central a manera de núcleo del sistema, en el cual se depositaron todos los servicios Web necesarios (*Back-End*) de manera que las aplicaciones cliente (*Front-End*) independientemente de la plataforma y el dispositivo pudieran utilizar los servicios que el núcleo ofrece (Figura 1). Tales aplicaciones cliente dependiendo de su propósito pueden o no hacer uso de todos servicios implementados en el núcleo.



**Figura 1.-** Arquitectura del marco de clases EMEF

Para lograr tal propósito se realizó la implementación del núcleo del sistema con PHP utilizando servicios Web mediante SOAP logrando con esto heterogeneidad entre los clientes al utilizar el protocolo de transferencia de hipertexto (HTTP por sus siglas en inglés *Hyper Text Transfer Protocol*) como medio de transporte y XML como medio de intercambio de información entre aplicaciones. De esta manera, el núcleo ofrece diversos servicios que las aplicaciones pueden consumir [5].

Dentro de la estructura del núcleo, uno de los servicios más importantes que se desarrolló fue el *parser* HL7, este *parser* en forma de servicio Web, ofrece a las aplicaciones cliente una comunicación de dos vías, por un lado toma toda la información almacenada en el núcleo de manera tradicional en una base de datos, la codifica en un documento HL7 en formato XML y lo envía a las aplicaciones cliente. Por otro lado, cuando las aplicaciones cliente realizan cambios sobre la información solicitada o simplemente generan nueva información clínica, esta es enviada al núcleo en formato HL7 a través de un servicio Web que el núcleo tomará, decodificará y almacenará de manera tradicional en su base de datos.

Adicionalmente al núcleo del sistema se desarrollaron dos tipos diferentes de clientes móviles que consumen los servicios Web ofrecidos por el marco de clases. Por un lado se desarrollo un cliente basado en una interfaz Web diseñada para usarse en una *Tablet PC*; y por otro lado, se desarrollo una aplicación para PDA como segundo cliente. Ambos clientes pensados para usarse en ambientes inalámbricos. Estos clientes consultan la información almacenada en el servidor a través de las interfaces provistas por los servicios Web. Una vez obtenida la información cada cliente la procesa y visualiza de acuerdo al propósito del diseño del cliente. Por un lado, el cliente Web ofrece una interacción más robusta, debido a que hace uso de la mayor parte de los servicios Web del sistema, permitiéndole una mayor funcionalidad. En contraste el cliente PDA hace uso de aquellos servicios que sólo le permitan visualizar la información al usuario. Aunado a esto, el estándar HL7 provee de un mecanismo para la inclusión de imágenes dentro de un expediente médico mediante la definición de etiquetas para tal propósito; etiquetas que el *parser* del lado de los clientes pueden reconocer y obtener las imágenes desde el servidor y mostrarlas. Los clientes, así como

los servicios que provee la arquitectura se introducen en la siguiente sección a través de posibles escenarios de uso.

### 3. Escenarios de uso

El uso de escenarios nos permite ilustrar de manera más tangible el uso de la tecnología propuesta sobre un problema en particular. En este caso se ofrecen dos posibles escenarios de uso que se muestran a continuación.

Mientras el médico de turno realiza sus rondas de rutina en los cuartos de los pacientes, llega a la cama 12 y desea verificar la evolución del paciente, entonces haciendo uso de un dispositivo manual (*Tablet PC*) con conexión inalámbrica ingresa al cliente Web (Figura 2) del expediente electrónico e introduce el identificador del paciente, inmediatamente obtiene en su pantalla la información del paciente, donde puede revisar sus síntomas, verificar su evolución e incluso realizar nuevas anotaciones al expediente.



Fig. 2- Cliente *Tablet PC*.

Mientras el médico de turno, esta en la sala de urgencias, llega la ambulancia con un paciente el cual presenta un cuadro clínico grave, por lo que se necesita obtener su expediente médico para conocer su historial clínico, sus reacciones a los medicamento y las alertas clínicas que se tengan registradas en el expediente, como el tiempo es crítico en estos casos, el médico toma su PDA con conexión inalámbrica, ingresa al cliente ligero del expediente electrónico e introduce el número de registro del paciente, inmediatamente obtiene en su pantalla la información médica, donde puede revisar de manera oportuna la información del paciente.



#### 4. Conclusiones y trabajo futuro

La funcionalidad ofrecida por el sistema ofrece algunas ventajas significativas observadas, que son una consecuencia directa de los estándares utilizados. Los clientes construidos, cada uno diferente entre sí, comparten la misma información: el expediente clínico electrónico. Cada uno de los clientes ofrece funcionalidad extra, tal funcionalidad se puede ir aumentando con el tiempo e ir personalizando de acuerdo a las necesidades actuales y futuras de la institución donde se implemente. Con el desarrollo de nuevos estándares computacionales que permiten la comunicación entre aplicaciones y la creación de estándares médicos que hagan uso de las tecnologías de la información, la brecha de intercomunicación entre organizaciones en cada vez menor, independientemente de la plataforma y los dispositivos que se utilicen para generar y consumir la información.

#### 5. Agradecimientos

Este trabajo fue apoyado por el proyecto C01-40799 y con la beca 179371 del segundo autor, ambos del CONACYT.

#### 6. Referencias

- [1] Health Level Seven [ <http://www.hl7.org/> ]
- [2] Patient Record Architecture  
[ [http://www.hl7.org/Special/committees/sgml/PRA/PRA\\_TSCPackageCover.htm](http://www.hl7.org/Special/committees/sgml/PRA/PRA_TSCPackageCover.htm) ]
- [3] ***HL7 Document Patient Record Architecture: An XML Document Architecture Based on a Shared Information Model*** Robert H. Dolin; Liora Alschuler; Fred Behlen; Paul V. Biron; Sandy Boyer; Dan Essin; Lloyd Harding; Tom Lincoln; John E. Mattison; Wes Rishel; Rachael Sokolowski; John Spinosa; Jason P. Williams, MS Kaiser Permanente; The Word Electric; U. of Chicago; Information Assembly Automation Inc.; Rand Corp.; U. of Illinois at Chicago; Wes Rishel Consulting; Magnolia Technologies; Scripps Health; Oceania Inc.
- [4] ***An Update on HL7's XML-based Document Representation Standards*** Robert H. Dolin, MD; Liora Alschuler; Sandy Boyer, BSP; Calvin Beebe for the Kona Editorial Group Kaiser Permanente (Robert.H.Dolin@kp.org); The Word Electric; Mayo Clinic
- [5] WSDL del marco de clases del expediente médico electrónico.  
[ [http://www.pecesama.net/emep/ws/emep\\_server.php?wsdl](http://www.pecesama.net/emep/ws/emep_server.php?wsdl) ]
- [6] Wastell, D.; White, P. y Kawalek, P., "A methodology for business process redesign: experiences and issues", Journal of Strategies information Systems, 1994.

# Clasificación no Paramétrica de la Distribución de Bordes Espectrales para la Valoración del Estado de Profundidad Anestésica

Brenda Deni Escamilla Alonso, Oscar Yáñez Suárez

Laboratorio de Neuroimagenología  
Departamento de Ingeniería Eléctrica  
Universidad Autónoma Metropolitana – Iztapalapa  
Av. San Rafael Atlixco 186, Col. Vicentina, México D.F. 09340  
email: yaso@xanum.uam.mx

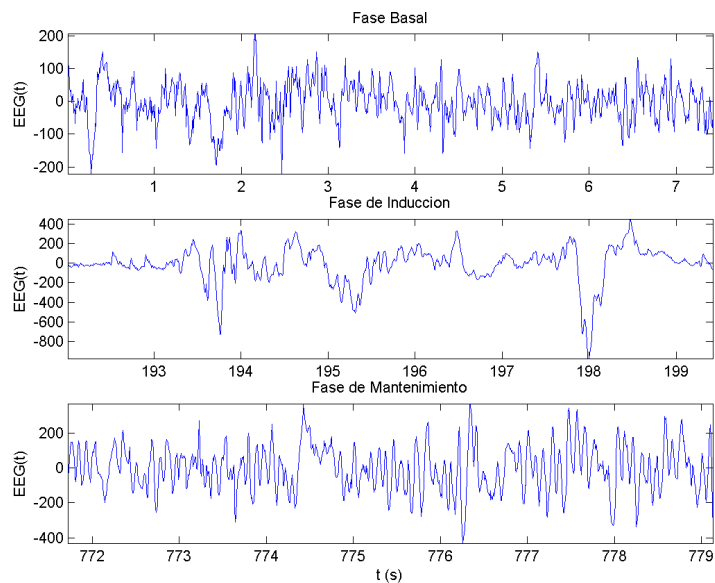
**Resumen.** Los anesthesiólogos utilizan una serie de observaciones de variables fisiológicas para estimar la profundidad anestésica en un paciente quirúrgico. De entre éstas, las mediciones espectrales del electroencefalograma se utilizan para establecer las diferentes etapas de sensibilidad por las que atraviesa un paciente durante una intervención. La técnica de medición es no invasiva y puede ser realizada fácilmente en la sala de operaciones. Este trabajo describe el diseño de un clasificador no paramétrico del tipo red neuronal probabilística para la identificación de las tres etapas del curso anestésico a partir de vectores de características espectrales simples.

**Abstract.** Anaesthesiologists depend on a series of physiological observations in order to estimate the depth of anaesthesia for a surgical patient. Among these, spectral measurements of the electroencephalogram are used to establish the different stages of sensibility that a patient undergoes throughout an intervention. The recording technique is non-invasive and can be easily carried out within the operating room. This paper describes the design of a non-parametric classifier – a probabilistic neural network – for the identification of three stages of the anaesthetic course from feature vectors composed of simple spectral properties.

**Palabras clave:** borde espectral, red neuronal probabilística, anestesia

## 1. Introducción

La señal electroencefalográfica (EEG), utilizada clínicamente para el estudio de la actividad eléctrica de la corteza cerebral (medida indirectamente y en forma no invasiva mediante electrodos colocados sobre el cuero cabelludo), se ha aprovechado también para el monitoreo trans-operatorio de la *profundidad anestésica*. Aún sin tener una definición precisa, ésta profundidad se asocia típicamente con la capacidad de un paciente para percibir y responder a estímulos nocivos, como los provocados durante una intervención quirúrgica. A medida que se aplica la terapia anestésica y ésta va surtiendo efecto, pueden percibirse cambios característicos de la señal de EEG. La experiencia clínica apunta hacia una categorización informal del curso anestésico, en el cual el EEG pasa de un estado basal a uno de inducción (mientras se administran los



**Fig. 1. Diferentes fases del curso anestésico tienen manifestación diferente en los patrones del EEG. De arriba a abajo: estado basal, etapa de inducción y fase de mantenimiento, observadas en la derivación estándar Fpz.**

medicamentos iniciales) y de ahí a un estado de mantenimiento de la insensibilidad (ver Figura 1) [1,2].

El análisis espectral realizado sobre el EEG permite la generación de índices asociados al efecto de agentes anestésicos sobre los ritmos cerebrales. Entre estos índices se encuentran el borde espectral 95 (la frecuencia a la cual la potencia integrada corresponde al 95% de la potencia total), la frecuencia mediana (el borde espectral 50) [3], el índice biespectral (acoplamiento de fases en la matriz de correlación cruzada espectral) [4], o la latencia de las respuestas evocadas auditivas [5]. Recientemente, el análisis de tiempo-frecuencia del EEG trans-operatorio (con la distribución Choi-Williams) se ha usado para estimar índices de frecuencia instantánea, borde espectral instantáneo y medidas de no linealidad como el componente fractal [6].

Cualesquiera los índices obtenidos a partir del análisis de la señal de EEG, éstos pueden ser utilizados para la clasificación de las etapas de anestesia. En los reportes de investigación relacionados, es común que solamente una de las medidas mencionadas haya sido usada, desaprovechando con ello la posible complementariedad de la información. A manera de ejemplo, de la observación del espectrograma del EEG puede inferirse que no solo hay cambios de ancho de banda, sino de magnitud de la densidad espectral; es poco probable que un solo índice pueda capturar toda esta descripción.

Por lo anterior, en este trabajo se propone la utilización de vectores de características espectrales para describir el curso anestésico y clasificar sus etapas. Las características relevantes son elegidas como los bordes espectrales 25, 50, 75 y 95, además de la magnitud máxima de la densidad espectral, para ventanas temporales de EEG. Los vectores de características son usados para construir un clasificador no paramétrico basado en el modelo de red neuronal probabilística, el cual implementa un estimador Bayesiano de la probabilidad *a posteriori* utilizando un modelo de ventanas de Parzen con kernel Gaussiano [7].

## 2. Metodología

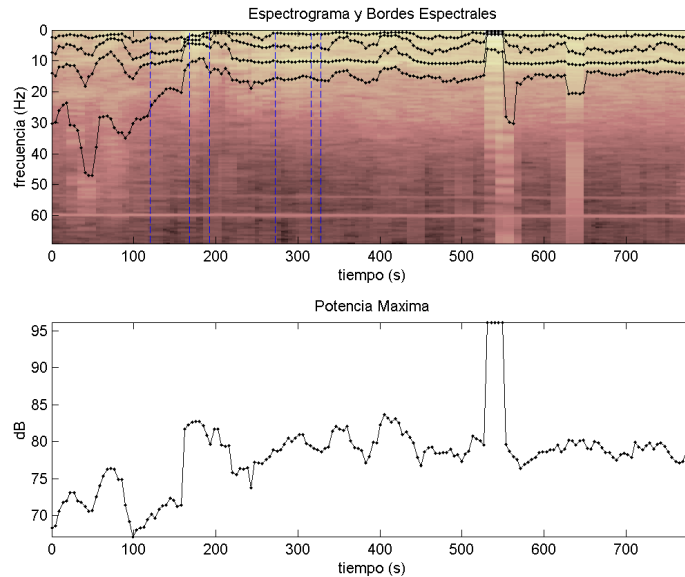
*Datos.* Cuatro sujetos sometidos a cirugías que no comprometieron la actividad cerebral, y a quienes se les aplicó anestesia general, fueron preparados para registro estándar (10-20) de EEG de 16 canales [6], previo consentimiento escrito. Todas las señales fueron digitalizadas a 140 Hz con resolución de 16 bits. Simultáneamente al registro, el anestesiólogo responsable anotó los tiempos en los cuales se iniciaron los siguientes eventos terapéuticos durante el curso de la anestesia: (i) administración de fentanyl, (ii) tiopental, (iii) pancuronio, y (iv) lidocaína, (v) intubación, y (vi) administración continua de isoflurano por vía aérea. La duración total de cada registro varió entre los 13 y 15 minutos. Para efectos del presente trabajo, sólo la derivación Fpz (frontal central) fue procesada y analizada.

*Pre-procesamiento.* Cada registro fue dividido en ventanas temporales de cinco segundos de duración, traslapadas un 10% de su longitud. Para la generación del espectrograma, la transformada discreta de Fourier de cada ventana fue calculada con 1024 puntos, previa multiplicación por una ventana de Hamming [8]. La distribución tiempo-frecuencia así obtenida fue suavizada mediante la aplicación de un filtro bidimensional con ventana promedio de 5 x 5 muestras. Para cada ventana, la potencia total fue calculada mediante integración trapezoidal acumulativa, y de aquí las frecuencias asociadas al 25, 50, 75 y 95% de la potencia total (bordes espectrales) fueron identificadas. Asimismo, se determinó la máxima densidad de potencia (en dB) para cada ventana. Todos los cálculos fueron realizados fuera de línea utilizando el programa Matlab ver. 6.5.

*Vectores de Características.* Cada ventana del espectrograma quedó representada por un vector de cinco características: los cuatro bordes espectrales y la densidad de potencia máxima. Los vectores fueron ampliados con un componente unitario, en preparación para el clasificador según dictan los usos recomendados [7]. Cada vector fue etiquetado, según el tiempo de ocurrencia de su ventana asociada y para efectos de la clasificación supervisada, como miembro de una de tres clases: (i) basal, previo a la administración de fentanyl, (ii) inducción, desde la administración del fentanyl hasta la intubación, y (iii) mantenimiento, durante la administración de isoflurano.

*Clasificador.* La totalidad de los vectores de características fueron usados para entrenar y validar una red neuronal probabilística mediante un esquema de validación cruzada de cinco vías, con prevalencias constantes para cada clase. El ancho del

kernel,  $\sigma$ , se eligió dentro del rango que minimizó el error de clasificación estimado. La implementación utilizada de la red probabilística provino del Classification Toolbox para Matlab [9].



**Fig. 2. (a)** Espectrograma típico del EEG trans-anestésico. Las líneas verticales punteadas indican los instantes de administración de fármacos para la inducción. Los trazos superpuestos al espectrograma indican (de arriba hacia abajo): borde espectral al 25, 50, 75, y 95% de la potencia total. **(b)** Máxima densidad de potencia espectral por ventana

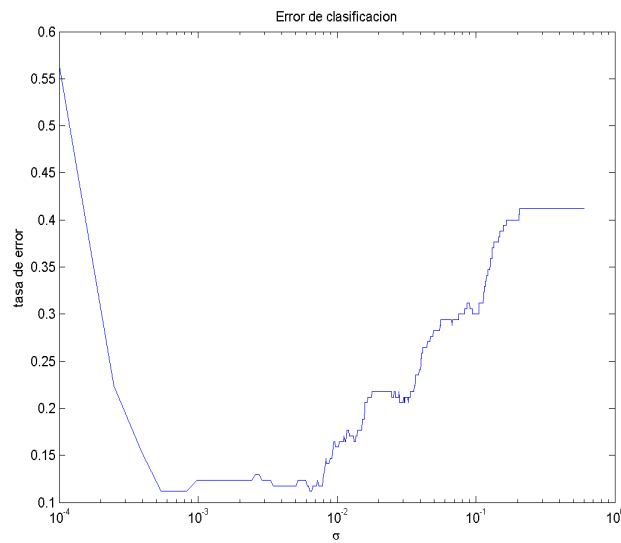
### 3. Resultados

La Figura 2(a) muestra un espectrograma típico del EEG trans-anestésico para un paciente. Los tonos claros corresponden a densidades de potencia más elevadas, así que se observa que la señal tiene primordialmente componentes de baja frecuencia (nótense, sin embargo, tanto la interferencia constante a 60 Hz como algunos artefactos alrededor de los 600 s). Sin embargo, es evidente que durante la fase basal el espectro del EEG mantiene un ancho de banda más amplio en contraste con las siguientes dos fases. Otra observación práctica sugiere que los bordes espectrales se mantienen más estables durante la porción final de la fase de inducción que durante las otras dos. Más aún, como se observa en la Figura 2(b), la máxima densidad de potencia es típicamente menor en la etapa basal.

La Figura 3 muestra la curva de entrenamiento de la red neuronal probabilística, en términos del error de clasificación estimado por la validación cruzada de cinco vías, en función del parámetro de ancho del kernel Gaussiano,  $\sigma$ . Como era de

esperarse, dicho comportamiento presenta un rango de valores para el cual el error es mínimo, y dentro del cual puede elegirse un valor de parámetro que provoque la mayor capacidad de generalización de la red. Para el estudio reportado en este trabajo, la elección de parámetro fue  $\sigma_{opt} = 0.0065$ .

El desempeño del clasificador diseñado se resume en las Tablas 1 y 2, donde se muestran las matrices de confusión promedio para el caso del conjunto de



**Fig. 3.** Diseño de la red neuronal probabilística. El error, estimado mediante validación cruzada de cinco vías, es sensible al ancho del kernel Gaussiano,  $\sigma$ .

**Tabla 1.** Matriz de confusión para el conjunto de validación

	basal	inducción	mantenimiento
basal	1.0	0.0	0.0
inducción	0.0	0.89	0.11
mantenimiento	0.1	0.1	0.8

**Tabla 2.** Matriz de confusión para el total de observaciones

	basal	inducción	mantenimiento
basal	1.0	0.0	0.0
inducción	0.0	0.91	0.09
mantenimiento	0.02	0.03	0.95

validación y para el total de los datos (entrenamiento más validación), respectivamente. Se resalta el hecho de que la identificación de la fase basal se realiza sin errores, mientras que las fases de inducción y mantenimiento se confunden en aproximadamente el 10% de los casos.

#### 4. Conclusiones

Se ha presentado una alternativa de solución al problema de la valoración de la profundidad anestésica mediante la clasificación no paramétrica de características espectrales derivadas del espectrograma del EEG frontal monocanal. A diferencia de las alternativas existentes actualmente, que soportan la decisión en un solo índice como el borde espectral 95, el índice biespectral o el componente fractal instantáneo, la propuesta considera vectores de características para representar la densidad espectral de ventanas temporales consecutivas y traslapadas.

La inclusión de cuatro bordes espectrales, así como del valor de la máxima densidad de potencia en cada ventana, permite describir con mayor detalle el contenido espectral del EEG durante el curso anestésico, usando sin embargo un número bajo de parámetros. En consecuencia, se logra un elevado porcentaje de acierto en la identificación de las etapas de profundidad mediante una red neuronal probabilística. La posibilidad de la aplicación directa de este clasificador sobre una nueva ventana de datos no vistos anteriormente, y su eficiencia de clasificación en registros de un solo canal, hace atractivo al procedimiento para una implementación en cuasi-tiempo real.

#### Referencias

- [1] Warren L.J., Shapiro-Harvey M., Maruchak G., Meathe E., "Automated EEG Processing for Intraoperative Monitoring", *Anesthesiology*, 53:223–236, 1980
- [2] Rampil I.J., "A Primer for EEG Signal Processing in Anesthesia", *Anesthesiology*, 89:980-1002, 1998
- [3] Schwender D., Dauberer M., Mulzer S., Klasing S., Finsterer U., Peter K., "Spectral Edge Frequency of the Electroencephalogram to Monitor Depth of Anaesthesia with Isoflurane or Propofol", *Br. J. Anaesthesia*, 77:179–184, 1996
- [4] Schraag S., Bothner U., "The Performance of Electroencephalogram Bispectral Index and Auditory Evoked Potential Index to Predict Loss of Consciousness during Propofol Infusion", *Anaesth. Analg.*, 89:1311-1315, 1999
- [5] Huang J.W., Lu Y., Nayak A., Roy R.J., "Depth of Anesthesia Estimation and Control", *IEEE Trans. BME*, 46:71-81, 1999
- [6] Andaluz-Suárez M., *Evaluación de la Profundidad Anestésica mediante el Análisis Tiempo-frecuencia del Electroencefalograma*, tesis de Maestría, Posgrado en Ingeniería Biomédica, Junio 2004.
- [7] Duda R.O., Hart P.E., Stork D.G., *Pattern Classification*, 2nd ed., Wiley, 2000
- [8] Oppenheim A., Schafer R., *Discrete Time Signal Processing*, 3rd ed., Morgan Kaufman, 1996
- [9] Stork D.G., Yom-Tov E., *Computer Manual in MATLAB to accompany Pattern Classification*, Wiley, 2004

# Digital Image Processing of Functional Magnetic Resonance Images to Identify Stereo-sensitive Cortical Regions Using a Global Stereo Stimulus

Héctor-Gabriel Acosta-Mesa<sup>1</sup>, Nicandro Cruz-Ramírez<sup>1</sup>, John Frisby<sup>2</sup>,  
Ying Zheng<sup>2</sup>, David Buckley<sup>3</sup>, Janet Morris<sup>3</sup>, and John Mayhew<sup>2</sup>

<sup>1</sup> Faculty of Physics and Artificial Intelligence, Department of Artificial Intelligence,  
University of Veracruz, Sebastián Camacho # 5, 91000, Xalapa, Ver. México.  
*{heacosta, ncruz}@uv.mx*

<sup>2</sup> Artificial Intelligence Vision Research Unit, Department of Psychology,  
University of Sheffield, Western Bank, S10 2TP, Sheffield, UK.  
*{j.e.mayhew, j.p.frisby, ying.zheng}@sheffield.ac.uk*

<sup>3</sup> Section of Academic Radiology, Royal Hallamshire Hospital,  
University of Sheffield, Glossop Road, S10 2JF, Sheffield, UK.  
*d.buckley@sheffield.ac.uk, janet.morris@sth.nhs.uk*

**Abstract:** Functional magnetic resonance images (fMRI) were analyzed using digital image processing techniques to investigate the cortical regions involved in stereoscopic vision using red/green anaglyphs to present random dot stereograms. In the experiment the subjects were instructed to follow the path of a square defined by depth and moving in the horizontal plane contrasted with a similar sized square defined by a slight difference in luminance. Three main regions were identified V3A, V3B and BA7. In agreement with other fMRI studies, our results suggests that areas V3A and BA7 (precuneus) are involved in stereo disparity processing. We hypothesise that the activation of the V3B region was produced by the second order motion component induced by the spatio-temporal changes in disparity (stereoscopic motion). We found no evidence for the involvement of the V5 area in the processing of stereoscopic stimuli.

**Key words:** neuroimaging, physiological signals, digital image processing, functional magnetic resonance imaging, stereo vision.

**Resumen:** Imágenes de resonancia magnética funcional (fMRI) fueron analizadas usando técnicas de procesamiento digital de imágenes para investigar las regiones del cerebro involucradas en visión estereoscópica. Los estímulos estereoscópicos fueron generados utilizando estereogramas anáglifos rojo/verde de puntos aleatorios. En el experimento los sujetos fueron instruidos para seguir visualmente la ruta de un cuadrado definido por profundidad desplazándose en el plano horizontal. Como contraste se utilizó una condición en la que el cuadrado fue definido por luminosidad. Tres regiones fueron identificadas: V3A, V3B y BA7. En concordancia con otros estudios similares, nuestros resultados sugieren que las regiones V3A y BA7 (precuneus) están involucradas en el procesamiento de disparidades estereoscópicas. Nuestra hipótesis es que la activación de la región V3B fue producida por la componente de movimiento de segundo orden inducida por los cambios espacio-temporales en las disparidades (movimiento estereoscópico). No se encontró evidencia de que el área V5 reaccione ante el estímulo estereoscópico.



## 1. Introduction

Although many psychophysical studies have investigated how the human brain computes stereoscopic information, it is uncertain which cortical areas are involved in its implementation. Some electrophysiological studies in monkeys report the sensitivity of V1 to absolute disparities, suggesting that this area could be a preliminary stage of processing for stereo information [1]. MT/V5 in monkeys shows a columnar organisation tuned for disparity [2]. MT/V5 in human brains has been widely reported as a motion sensitive area [3]. Given the similarity between the visual system of the monkey and the human, it is possible that V5 in human brains is involved in the processing of stereo information. Recently, modern non-invasive neuroimaging techniques, eg PET and fMRI, have been used to explore the functional anatomy of stereoscopic vision in humans. The results of these studies suggest that stereo disparity processing maybe widespread over a network of cortical regions in the occipital and parietal lobes, including V1, V2, V3, V3A, V3B, [4]. However there is no general agreement about the cortical regions selective to stereo disparities or the specific role that each of these has in the perception of depth [6]. The main goal of the present study was to process functional magnetic resonance images to investigate the cortical regions sensitive to stereo disparities using a stereo stimulus that avoids adaptation and at the same time maximises the attentional demands. Under this principle, one experiment was developed using functional magnetic resonance imaging to identify stereo sensitive regions stimulated by random dot anaglyph stereograms.

## 2. Materials and methods

Ten healthy subjects, nine right-handed and one left-handed volunteers (7 female, 3 male) aged from 20 to 30 years participated in the experiment. All subjects gave informed written consent. The stereo acuity of the subjects was measured using a stereo vision test (RANDOT SO-002), all of them were below 40 sec of arc. The subjects were given a preliminary practice session outside the magnet to become familiar with the visual stimulation. Stimulus presentation: Subjects lay on their backs in the magnet. They wore red/green anaglyph glasses and looked via a mirror angled at  $\sim 45^\circ$  from their visual axes at a back illuminated screen located just outside the magnet. The viewing distance was 2.4 m. Stimuli were projected on to the screen using an EPSON (EMP-7300) projector driven by a 3G Mac running Psychophysics Tool Box ver. 2.44 under MATLAB ver. 5.3. Although the stimuli were displayed at a video frame rate of 60 Hz, the image was only updated on every 10th frame, producing an effective frame rate of 6 Hz. Data acquisition and analysis: Subjects were scanned in a 1.5 T whole-body MRI scanner (Eclips Marconi) with BOLD contrast echo planar imaging (TR=3s, TE=40 ms, 128 x128 voxel, voxel size 1.875 x 1.875 x 4 mm.). Thirty two slices covering the whole brain were acquired. The data was pre-processed and analysed using SPM99 (Wellcome Department of Cognitive Neurology). The first five scans of each run were discarded to exclude magnet saturation artefacts.

All volumes were slice timed, motion corrected and normalised in the MNI (Montreal Neurological Institute) stereotaxic space. The data were smoothed using a 6 mm FWHM (full width at half maximum) isotropic Gaussian kernel. Data analysis was performed using a boxcar design matrix of the different conditions convolved with the hemodynamic response function. Specific effects were tested by applying the corresponding linear contrast to the parameters obtained applying General Linear Model (GLM) using the design matrix [7].

### 3. Global stereo tracking experiment

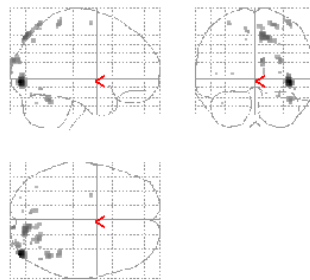
This experiment was designed to activate stereo sensitive regions by requiring the subjects to perform a task of global stereo tracking (GST). Random dot stereograms were used to define a square region moving horizontally in front of the background from left to right and vice versa. The subjects were instructed to perform pursuit eye movement to follow the path of the square with their eyes. The performance of the task depended on the maintenance of the perception of depth defined by stereo disparities.

Experiment design: Subjects were given a sequence of 4 scans (sequences) each lasting 5.15 min. (10 epoch) with a 5 min. interscan interval to permit subjects to rest. One hundred image volumes were obtained in each sequence. Each condition lasted 30s. giving 10 multislice volumes per condition (TR=3s.). Each scan consisted in alternating epochs in a boxcar configuration. A dummy condition of a blank screen was presented during the first 15s. (5 scans) of each run which were excluded to control for magnetic saturation effects. The display contained 1,024 dots (with radius 0.1 deg. and zero disparity) distributed over the screen (mean dot density 1.5 dot deg<sup>-2</sup>). The subjects were instructed to fixate on the right superior corner of a square (5.23 deg. side long) moving laterally across the screen (13 deg. field of view). The square was moved from left to right and vice versa at a constant speed (2.19 deg.sec<sup>-1</sup>), each time that the square reached one edge of the screen it changed its direction. Dynamic random noise was used in order to remove any motion cues introduced by the change in disparity [5]. Two modalities were used to define the square, each one representing an experimental condition:

a). Two dimensional tracking (2D): The square was luminance defined, its luminance (8.56 cd/m<sup>2</sup>) was lower than the background (18 cd/m<sup>2</sup>). This condition was used as a base line. b). Three dimensional tracking (3D): The square was depth defined (red/green anaglyph stereogram), presented 0.3 deg at the front of the background (zero disparity). The square moved horizontally in the plane X/Y, not in plane Z (motion in depth). Stereo-sensitive regions were identified by comparing the activation during the 2D control condition with the neural activity during the 3D stereo tracking condition. Although we could not track the eye movements to monitor the subject's performance of the task, we have no reason to believe that there was any difference between them in the two experimental conditions.

## 4 Results

Stereo-sensitive regions were identified using a contrast that compared the activation produced by the moving square defined by depth with that produced by the moving square defined by luminance. The results show consistent activation in areas V3A, V3B and parietal cortex. Area V3A in the right hemisphere was activated by depth in 7 of the 10 subjects. In two of these subjects there was bilateral activation. Area V3B in the right hemisphere was activated in 8 of the 10 subjects. In five of these there was bilateral activation. One subject showed activation only in left V3B. Area BA7 (precuneus) in the right hemisphere was activated in 3 of the 10 subjects. Two subjects had left hemisphere activation. There was no bilateral activation. Activation in the right superior parietal area was found in 2 other subjects. Figure 1 shows results from a representative subject, Tables I and II show the Montreal Neurological Institute (MNI) coordinates of regions of activation for all the subjects and the associated confidence levels. We did not use a group analysis (random effects) due to the small number of subjects. It is important to clarify that reversing the contrast revealed no across subject consistent regions in which the activations of the 2D tracking condition were greater than the 3D stereo tracking (data not shown).



**Figure 1** Depth against Luminance. The statistical map shows the areas sensitive to stereoscopic information. The activation includes V3A, V3, precuneus (BA7) and superior parietal regions.

Results of previous fMRI studies of eye movements have suggested the involvement of parietal areas in the control of pursuit and saccadic eye movements [8]. None report the involvement of V3A and V3B regions. In this study, despite the control in which the eye-tracking component was identical in the two conditions, we find an increased activation in the parietal cortex in the stereo tracking task in several of the subjects. This suggests some involvement of the parietal region in stereo-processing per se independent of eye movements. This was examined improved in a second experiment (not reported here).

## 5. Conclusions and Future Work

The results of our experiment reveal three main regions sensitive to stereoscopic depth: V3A, V3B and precuneus. In contrast to the evidence from physiological experiments in monkey [2] we did not find evidence that V5 was involved in the processing of stereoscopic information.

<i>Subject</i>	<i>Location</i>	<i>Z-Score</i>	<i>P corrected</i>	<i>Cluster size</i>
1R	32, -96, 16	5.65	0.002	21
2R	30, -92, 24	6.20	0.000	16
3R	28, -88, 16	5.09	0.029	3
4 R	34, -94, 16	5.62	0.003	4
7R	34, -92, 22	6.25	0.000	8
8R	36, -86, 16	(Inf)	0.000	45
8L	-28, -96, 16	(Inf)	0.000	62
10R	32, -94, 16	(Inf)	0.000	116
10L	-28, -98, 14	(Inf)	0.000	546

**Table I** Stereo sensitive region: V3A. The alphanumeric symbols in the first column refer to right and left hemispheres of the subject, for example, the left and right hemispheres of subject 1 are given as 1R and 1L respectively. Subjects or hemispheres not included in a particular table means that no significant activation was found.

<i>Subject</i>	<i>Location</i>	<i>Z-Score</i>	<i>P corrected</i>	<i>Cluster size</i>
1R	34, -84, -6	6.25	0.000	36
1L	-26, -94, 2	5.80	0.001	25
2R	40, -88, -4	6.98	0.000	38
3R	38, -88, -4	5.29	0.011	6
3L	-28, -96, -10	5.54	0.003	4
4 R	40, -88, 0	5.81	0.001	16
6 R	28, -104, 2	7.03	0.000	93
7 R	36, -92, 0	(Inf)	0.000	105
7L	-28, -102, 2	(Inf)	0.000	37
8R	36, -92, 4	(Inf)	0.000	35
8L	-26, -100, 6	(Inf)	0.000	176
9 L	-28, -100, -2	4.54	0.000	51
10R	38, -90, 2	5.88	0.000	3
10L	-28, -100, 2	7.84	0.000	7

**Table II** Stereo sensitive region: V3B.

Consistent with the results of other studies, we find evidence that V3A and precuneus (BA7) showed sensitivity to stereo disparities [4,6,13]. It is unlikely that these activations are due either to different patterns of eye movements or to stronger attentional engagement because the control condition provided identical requirements for these parameters and similar areas of activation were found in a second experiment (not yet published) in which no eye movements were required to perform the task. Our results suggest a network involving the cortical areas V3A, V3B and precuneus in the processing of stereo disparities. The high proportion of activations located in the right hemisphere supports the notion of right cerebral dominance in stereo vision [9]. Contrary to the results reported in studies with monkeys, our experiments did not reveal any evidence of the sensitivity of V5 to stereo disparity processing. Our results also showed a region in with functional profile and anatomical location which matched the V3B region. We have suggested that this region was activated by stereoscopic motion component of the dynamic tasks used. Although the task used in the experiment produced motion of spatio-temporal changes of shapes defined by disparity, this was not designed specifically to optimally produce disparity based second order motion. The objective of our study was to investigate the cortical areas selective to stereoscopic information, and an experiment designed to test the hypothesis of the selectivity of the V3B region to stereoscopic motion is for future work.

## 6. References

- [1] B. G. Cumming and A. J. Parker, "Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity," *J Neurosci*, vol. 19, pp. 5602-18., 1999.
- [2] G. C. DeAngelis and W. T. Newsome, "Organization of disparity-selective neurons in macaque area MT," *J Neurosci*, vol. 19, pp. 1398-415., 1999.
- [3] R. B. Tootell, J. D. Mendola, N. K. Hadjikhani, P. J. Ledden, A. K. Liu, J. B. Reppas, M. I. Sereno, and A. M. Dale, "Functional analysis of V3A and related areas in human visual cortex," *J Neurosci*, vol. 17, pp. 7060-78., 1997.
- [4] B. Gulyas and P. E. Roland, "Binocular disparity discrimination in human cerebral cortex: Functional anatomy by positron emission tomography," *Proc. Natl. Acad. Sci. USA (Neurobiology)*, vol. 91, pp. 1239-1243, 1993.
- [5] A. Hanazawa, R. Kawashima, K. Nakamura, Y. Sato, M. Surgiura, J. Watanabe, K. Sato, Y. Maeda, K. Matsue, and H. Fukuda, "The human posterior parietal cortex participates in stereoscopic depth perception. A fMRI study.," *NeuroImage (poster)*, vol. 11, 2000.
- [6] R. D. Freeman, "Stereoscopic vision: Which parts of the brain are involved?," *Curr Biol*, vol. 9, pp. R610-3., 1999.
- [7] K. J. Friston, H. A. P., W. K. J., P. J. P., F. C. D., and F. R. S. J., "Statistical parametric maps in functional imaging: A general linear approach.," *Human Brain Mapping.*, vol. 2, pp. 189-210, 1995.
- [8] M. Corbetta, E. Akbudak, T. E. Conturo, A. Z. Snyder, J. M. Ollinger, H. A. Drury, M. R. Linenweber, S. E. Petersen, M. E. Raichle, D. C. Van Essen, and G. L. Shulman, "A common network of functional areas for attention and eye movements," *Neuron*, vol. 21, pp. 761-73., 1998.
- [9] I. P. Howard and B. Rogers, "Binocular vision and stereopsis.," Oxford University press., 1995.

# Diseño de una Sala Virtual de Rayos X con Aplicaciones en Seguridad Radiológica

Bustos Mendoza Cesar Rogelio, García Ruiz Miguel Ángel, López Rosas Denisse Margarita, Hernández García Leopoldo

Universidad de Colima, Laboratorio de Realidad Virtual (CEUPROMED). Av. Universidad # 333, C.P. 28040, Colima, Colima, México. Tel./fax (312)3161093  
cbustos@ucol.mx, mgarcia@ucol.mx,  
dessed@hotmai.com,diabl170\_net@hotmail.com

**Abstract.** This paper describes the design of a virtual model of an X-ray room, created with the modeling software AC3D, with the purpose to offer a virtual learning environment about radiology safety teaching in hospitals, to benefit medical teams and other health personal who are exposed to radiation in a X-ray room. This virtual model is programmed using standard language VRML to simulate radiation fields, and warns people not to enter the room when X-rays are taken. We plan to use this model in teaching programs at the University of Colima's Virtual Reality laboratory, with the objective to get a better immersion effect and comprehension of the virtual model.

**Resumen.** El presente trabajo describe el diseño de un modelo virtual de una sala de rayos X, desarrollado con el programa de modelado AC3D, para proponer un espacio virtual de enseñanza sobre temas de seguridad radiológica en clínicas y hospitales locales para personal médico y otro personal del área de la salud que sea expuesto a radiación dentro de una sala. El modelo virtual se programa en VRML para simular la presencia de campos de radiación y evitar que personal no autorizado entre a la sala. Se propone a futuro incluir el uso del modelo en programas de enseñanza de radiología como medio didáctico, utilizando los recursos del Laboratorio de Realidad Virtual de la Universidad de Colima con el objetivo de alcanzar un grado mayor de inmersión y comprensión.

**Palabras clave:** Ambientes virtuales, Modelado en 3D, Sala de Rayos-X, Radiología.

## 1. Introducción

Las salas de rayos-X de los grandes hospitales o de gabinetes pequeños son lugares especializados donde acude personal médico, técnico y público en general con el propósito de obtener placas auxiliares en el diagnóstico de pacientes con diversas patologías como neumonía, tumores, artritis y principalmente en el caso de lesiones óseas como las fracturas es decir, en general, este tipo de salas permiten la visualización indirecta de estructuras internas [1].

Dado las características de estas salas se requiere infraestructura específica y además deberán estar provistas de blindaje especial, áreas específicas para revelado y

diagnostico y equipo médico sofisticado, lo que hace de estos lugares, sitios riesgosos, donde participan diferentes especialistas que constantemente se exponen a la radiación por lo cual son necesarias medidas preventivas basadas en regulaciones oficiales según la normatividad vigente en nuestro país y que necesitan ser difundidas y constantemente reforzadas entre el personal hospitalario.

En el Hospital Regional Universitario de Colima, el personal labora en 3 turnos diferentes, la protección se realiza con mandil y guantes especiales que se utilizan durante la aplicación de los rayos-X, y principalmente durante la fluoroscopia, que se define como una técnica que combina una fuente de rayos-X y una pantalla fluorescente para facilitar la observación directa de órganos o estructuras [1], la distancia a la que se localizan los usuarios cuando aplican radiación, es variable según el procedimiento y el tiempo de exposición a que se somete personal, también es variable pero se incrementa con el uso de la fluoroscopia y es cuando la protección es mas necesaria. El acceso a la sala es restringido de manera parcial y eso produce riesgos en el personal, además en varias ocasiones no se indica con la luz roja que se están haciendo estudios utilizando radiación y que por consiguiente no se puede pasar al interior de la sala.

La situación descrita en el Hospital Regional, nos permite inferir que es necesario reforzar los programas de seguridad radiológica de una manera más efectiva empleando recursos tecnológicos quizás no tan convencionales pero más efectivos.

## **2. Estado del arte**

La simulación de mundos virtuales ha permitido el entrenamiento virtual de diferentes situaciones peligrosas, donde es posible explicar los riesgos existentes y sus posibles consecuencias, además de que facilitan la visualización de información de dicho problema [2, 3].

En los hospitales se emplean los modelos tridimensionales con imágenes de órganos en equipos de imagenología, su uso es ya es muy extendido y ha permitido por una parte la comprensión de patologías en el aspecto medico y por el otro, en el área informática el desarrollo de nuevos algoritmos y metodologías para procesamiento digital de imágenes médicas [5,6].

La simulación en hospitales se enfoca principalmente en los entrenadores médicos que permiten el desarrollo de habilidades para realizar procedimientos endoscópicos que permiten mejorar el desempeño profesional de los médicos utilizando los beneficios de la aplicaciones de los ambientes virtuales en el campo de la medicina, como sucede en los entrenadores de laparoscopia y de endoscopia [4].

Sin embargo, se ha hecho muy poca investigación y desarrollo en cuanto a la simulación y modelado con ambientes virtuales acerca de aplicaciones de seguridad radiológica para su uso en programas de enseñanza y capacitación.

En base a lo anterior podemos hipotetizar que la aplicación de un modelo virtual de una sala de rayos-x ayudará a los técnicos y radiólogos a ubicar espacialmente (y visualmente en general) los riesgos que pueda haber en la exposición de radiación de los aparatos de radiología, y al manejo de los instrumentos de protección de una manera y en el tiempo adecuados.

### 3. Metodología

El desarrollo de este proyecto se lleva a cabo en 4 etapas:

1. Investigación teórica
2. Diseño del modelo en 3D (ambiente virtual)
3. Programación del ambiente virtual.
4. Pruebas del ambiente virtual.

Durante la etapa de investigación teórica se buscaron conceptos y definiciones de los elementos que integran una sala de rayos-X [7], además de recopilar las normas vigentes en seguridad radiológica para nuestro país, las cuales están establecidas en las Normas Oficiales Mexicanas NOM-156-SSA1-1996 Y NOM-157-SSA1-1996. En el estudio se resumen las características comunes a la mayoría de las salas de rayos X, además de hacer consideraciones básicas en el diseño y operación de la misma, para facilitar la representación del modelo virtual. Se buscó que la “sala ideal” estuviera provista de las características más comunes a la mayoría de las salas de rayos-X. Para lograrlo además, se realizaron visitas a salas de rayos-X reales, se formularon cuestionarios al personal y se obtuvieron imágenes reales representativas como base para el modelado. Además, en la etapa de investigación se encontró que las principales áreas que se involucran en la seguridad radiológica son:

- Sala de Exposición ;
- Cuarto de control;
- Vestidor;
- Cuarto Oscuro;
- Área de Diagnostico;
- Sala de Espera.

La elección del software de modelado AC3D ([www.ac3d.org](http://www.ac3d.org)) se decidió luego de realizar una tabla comparativa con 3 diferentes tipos de software de modelado (AC3D, 3D MAX y RHINO) y analizar ventajas y desventajas. El modelo se realizó aplicando técnicas de diseño en 3D del ambiente virtual.

En la programación de la simulación de la radiación se utilizó el lenguaje estándar VRML 2 para programación de mundos virtuales, se convirtió el modelo realizado en AC3D a VRML y se programó el efecto de radiación y señalización, por el exterior, se hizo mediante una luz indicadora que cambiaba de verde a rojo con un tiempo aleatorio para prohibir el acceso a la sala desde la sala de espera y en el interior



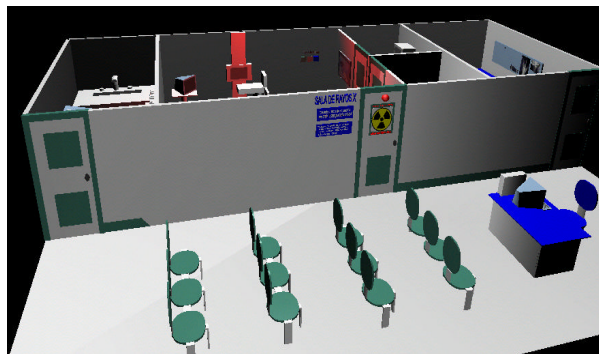
mediante una luz intensa de color rojo (sincronizada con la luz exterior) que aumenta de manera gradual para representar la radiación en el interior de la sala, que es el momento justo cuando el usuario real debe utilizar la protección radiológica (guantes, lentes, mandil, etc.)

La elección del AC3D para el modelado fue debido al costo económico en relación con 3D MAX y RHINO, y a que permitía una interfaz de usuario simple y facilidad de uso y aprendizaje, además de que se trata de un programa multiplataforma donde el formato del archivo es compatible a los objetos para el uso de archivos VRML que se programaron posteriormente para la simulación. En la sala del Hospital Regional, se puede observar la luz indicadora que impide el paso a la sala como se ilustra en la Figura 1.

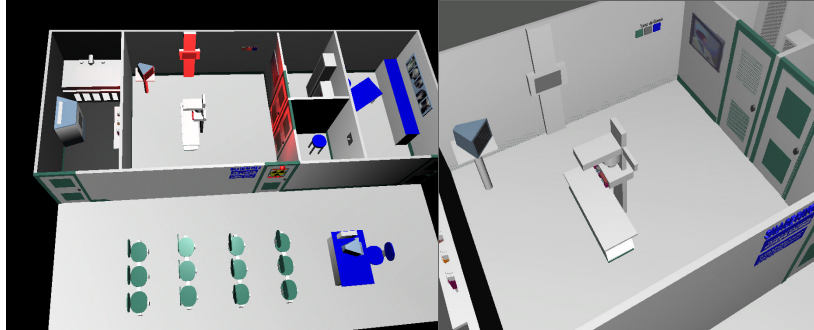


**Fig. 1.** Sala de Rayos X del Hospital Regional Universitario de Colima.

En el modelo resultante se representa este efecto de manera que el usuario pueda hacer conciencia de todos los elementos que intervienen cuando se está aplicando radiación en la sala, esto se puede realizar con la supervisión de personal capacitado en seguridad radiológica, de tal manera que pueda ser integrado a un programa de enseñanza, en las figuras 2, 3 y 4 podemos ver el modelo final.



**Fig. 2.** Vista General del Modelo en AC3D.



**Figs. 3 y 4.** Interior del modelo en 3D.

Con el objeto de hacer una revisión inicial del ambiente virtual en la tercera etapa del proyecto (pruebas del sistema), se realizó una evaluación heurística [8] del ambiente virtual. Dicha evaluación permite detectar problemas de uso, aplicación, funcionalidad, calidad y estética de una interfaz de computadora (en este caso el ambiente virtual) para su posterior corrección, siguiendo ciertos criterios de evaluación (heurísticas). Dos especialistas en interacción humano-computadora participaron en la evaluación, donde siguieron las 10 heurísticas básicas descritas por [8].

#### **4. Resultados preliminares**

Los resultados de la evaluación heurística permitió corregir ciertos detalles en el diseño e implementación del modelo en 3D, en especial la animación de la luz de advertencia que se encuentra sobre la puerta de entrada. A la fecha, en el proyecto se han completado las etapas 1, 2 y 3, que corresponden a la revisión de la literatura, diseño y programación del ambiente virtual, respectivamente. Se está continuando con la etapa de pruebas del sistema, ya que hacen falta más estudios de usabilidad del ambiente virtual, y pruebas con un grupo de radiólogos y técnicos para probar la efectividad del mismo.

#### **5 .Conclusiones y Perspectivas**

En el diseño de un ambiente virtual es necesario acercarse a los especialistas en su área de trabajo para determinar las necesidades reales y aplicar técnicas de modelado en 3D que nos permitan recrear situaciones peligrosas donde se puedan manipular los elementos de riesgo. En el área de radiología poco desarrollo existe en los modelos gráficos que permitan incluirlos en programas de capacitación para seguridad radiológica, este modelo generado también puede utilizarse para equipos interdisciplinarios que planeen el diseño o adaptación de una sala real, le da las

ventajas de discutir materiales, equipos, distancias, blindaje, instalaciones y demás factores que intervienen en su diseño.

Se planea incluir el modelo en un programa de capacitación del Hospital Regional, empleando los recursos del Laboratorio de Realidad Virtual de la Universidad de Colima, este programa deberá ser aplicado de manera periódica para reforzar este tema tan importante. En el programa se podría incluir el uso de documentales en video, y presentaciones de diapositivas, además de otros programas para cálculo de radiación para hacer un taller completo y abarcar una gran variedad de temas de seguridad radiológica.

## Referencias

- [1] J.D Bronzino, *Management of Medical Technology*, Butterworth-Heinemann, 1992.
- [2] J.C. Dürsteler, *Visualización de Información*. Barcelona: Ediciones Gestión 2000, 2003.
- [3] W.R. Sherman, *Understanding Virtual Reality. Interface, Application, and Design*, Morgan Kaufman Publishers, San Francisco California, 2003.
- [4] C. Krapicgler, M. Haubner, A. Losch, M. Schuhmann, K. Englmeier, “Physicians in virtual environments – multimodal human- computer interaction”. *Interacting with computers II*, 1999, 427 – 452.
- [5] N.Freud., P. Duvauchelle, S.A. Pistrui-Maximean, J-M Letang D. Babot. “Deterministic Simulation of first-order Scattering in Virtual X-ray Imaging”. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, Vol: 222, Issue: 1-2, July, 2004.
- [6] M.Rotter., A.Schneidewin., M.Doerr, M.Loewenhaupt., A.M. El Massalami, C.Detlefs, “Interpreting Magnetic X-ray Scattering on Gd-compounds using the Mcphase Simulation Program”, *Proceedings of the Conference on Polarised Neutron and Synchrotron X-rays for magnetism*, (345), 2003, Pp 231-234.
- [7] P.C. O’Kane, “A Simulation Model of a Diagnostic Radiology Department”. *European Journal of Operational Research*, 1981, (6), pp. 38-45.
- [8] J. Nielsen, How to Conduct a Heuristic Evaluation,  
[http://www.useit.com/papers/heuristic/heuristic\\_evaluation.html](http://www.useit.com/papers/heuristic/heuristic_evaluation.html), Obtenido de la red mundial el 10 de junio de 2004.

# Experiencias de Implantación del Estándar DICOM

Jesús A. Treviño y Claudia Feregrino  
Instituto Nacional de Astrofísica, Óptica y Electrónica  
Luis Enrique Erro #1 Tonantzintla,  
Puebla, México  
cferegrino@inaoep.mx

## Abstract

The need for communication with modalities from different vendors made possible the standardization of communication within the medical area. An example of such standards is DICOM (Digital Imaging and Communications in Medicine). Currently, the modalities follow this standard, however, there are diverse hospitals or clinics in our country whose modalities do not follow the standard and this is required if a Picture Archiving and Communication System (PACS) is to be implemented. This paper describes the experiences gained during the implementations of the DICOM standard for a hospital radiology area.

## Resumen

La necesidad de conectar equipos radiológicos de diferentes vendedores hizo posible el surgimiento de estándares de comunicación especializados para medicina, como lo es DICOM. Actualmente, las modalidades que se venden, siguen este estándar, sin embargo, existen en nuestro país diversos hospitales cuyos equipos no lo siguen, y para los cuales se requiere durante una implementación de un sistema de comunicación y almacenamiento de imágenes médicas. Este artículo describe las experiencias obtenidas durante la implantación del estándar DICOM para el área de radiología de un hospital.

Palabras clave: DICOM, SCU, SCP, PACS, Implantación.

## 1. Introducción

La necesidad de conectar equipos radiológicos de diferentes vendedores hizo posible el surgimiento de estándares de comunicación especializados para medicina. Por ejemplo, conectar un equipo de tomografía de marca SIEMENS con una impresora de marca KODAK hubiera sido casi imposible hace una década, sin embargo, con la estandarización de la conectividad ahora es posible. Anteriormente, un hospital quedaba totalmente cautivo con un determinado proveedor, pues ello casi garantizaba el funcionamiento entre equipos comprados subsecuentemente. Aunque esto implicara que otros proveedores, posiblemente de equipo con mejor desempeño o más barato, quedarán sin posibilidad de vender a dicho hospital. Para evitar este problema, NEMA [1] creó en 1993 el primer estándar donde las compañías o desarrolladores independientes asociados deberían seguirlo. Este estándar llamado DICOM [1], es el estándar para imagenología digital y comunicaciones en medicina.

Desde que el estándar fue publicado, ha sido aceptado ampliamente por vendedores de equipos radiológicos (modalidades) y se ha aplicado exitosamente para la comunicación de datos de imagenología entre equipos y Sistemas de Comunicación y Almacenamiento de Imágenes (PACS, por sus siglas en inglés) [2]. Hasta ahora, la capacidad de comunicación del estándar DICOM ha sido tal que se ha incluido en la mayoría de productos PACS, como sistemas de almacenamiento y estaciones de trabajo.

El estándar DICOM define protocolos, objetos, servicios y requerimientos de conformidad. El soportar el estándar es un prerequisite para una comunicación efectiva entre dispositivos de imagenología digital. Los objetos de imágenes DICOM incluyen a aquellos designados para varios equipos y están definidos por IODs (Information Object Definitions), es decir, cada IOD está asociado a una modalidad. Los IODs incluyen información codificada en atributos, junto con las imágenes. Estos atributos describen la información sobre la identificación, administración y técnicas de adquisición de las imágenes. Los IODs se combinan con servicios, tales como “storage service class” para producir “service-object pair classes”, los cuales son la unidad de conformidad con DICOM.

Un dispositivo o equipo radiológico puede actuar como un “usuario de la clase de servicio” o SCU (Service Class User), por ejemplo, uno que envía imágenes, o como un “proveedor de la clase de servicios” o SCP (Service Class Provider), por ejemplo, uno que recibe imágenes. Para establecer la comunicación es necesario contar con SCU cliente y un SCP servidor. Al asociar un IOD con un elemento de servicio de mensajes DIMSE (DICOM Message Service Element) recibe el nombre de SOP (Service-Object Pair). Algunos de los mensajes que contiene el estándar son C-GET, C-MOVE, C-FIND y C-STORE.

DICOM estandariza la comunicación entre dispositivos e incluye servicios que a) tratan con el flujo de trabajo, tanto dentro de un cuarto de exámenes o dentro de una compañía, b) soporta una organización eficiente de imágenes para su despliegue y c) soporta diferentes apariencias de imágenes. La implementación de este estándar no es simple y es precisamente lo que tratará este artículo.

Este artículo está organizado como sigue: la Sección 2 menciona el estado del arte, donde se identifica el proceso de desarrollo del estándar DICOM, la Sección 3 muestra la metodología que se siguió para llevar a cabo este trabajo, la Sección 4 describe la manera en que se implementó el estándar, la Sección 5 explica los resultados obtenidos y finalmente la Sección 6 concluye.

## **2. Estado del Arte**

Existen varias publicaciones sobre el estándar DICOM aparte del propio estándar. Entre los trabajos encontrados están [3, 4, 5 y 6]. En [3] se describen los requerimientos y diseño de la familia DX (rayos X digital) de almacenamiento de archivos DICOM (imágenes con formato de acuerdo al estándar) y los mecanismos por los cuales el equipo puede reunir eficientemente la información necesaria. Se da una recomendación a usuarios en forma de guía de compra, la cual enfatiza los factores principales en el despliegue de equipo de radiografía digital.

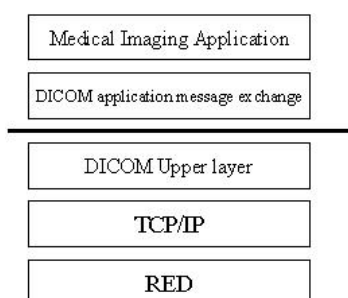
En [4] se habla también de la integración de equipo de imagenología de varios vendedores con DICOM para un Sistema de Comunicación y Almacenamiento de Imágenes (PACS, por sus siglas en inglés). Y en [5] se menciona cómo expandir el estándar DICOM a Internet mediante un esquema DICOM-QR URL (Petición-Envío basado en un Localizador de Recursos Uniforme).

Desde que DICOM fue creado se ha definido un conjunto de servicios que garantizan la consistencia de imágenes en escala de grises para mostrar tanto en pantalla como en impresión [6]. También han surgido integraciones de meta-lenguajes como XML con el estándar DICOM para formar plataformas de e-Salud que permiten una mejor integración de DICOM en una aplicación basada en Web [7].

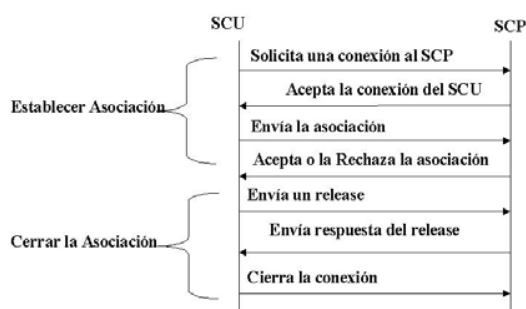
Los servicios DICOM más ampliamente conocidos son aquellos relacionados con la transferencia de imágenes y objetos relacionados a través de la red, mencionados en el estándar como “almacenamiento”. Las imágenes son transferidas (almacenadas) en formato específico de la modalidad, que incluye información común a varias modalidades, así como información sobre la técnica de adquisición específica. Los datos de los píxeles son almacenados en la forma específica a la modalidad [2].

El estándar indica que existen dos tipos de comunicación: 1) punto a punto, ya en desuso, y 2) bajo el ambiente de red. En la siguiente figura 1 se muestra cuáles son las capas en las cuales trabaja DICOM. Por basarse en TCP/IP se puede usar en cualquier ambiente o tipo de red. Cabe mencionar que el tiempo que puede tardar en transmitir una imagen puede variar dependiendo de las características de la red y del tamaño de la imagen.

Existen varias herramientas en el mercado para el desarrollo de aplicaciones a nivel del DICOM Upper layer, las cuales pueden ser adquiridas pagando la licencia o simplemente son gratuitas. Por citar algunas DCMTK, DICOMscop (gratuita), ezDICOM.



**Fig. 1 Capas de Red en la Comunicación DICOM**



**Fig. 2 Comunicación DICOM**

Una vez que se ha establecido la comunicación, hay que establecer la asociación. La asociación está compuesta por algunos de estos elementos *Presentation Id*, *Abstract Syntax* y *Transfer Syntax*. *Presentation Id* es un número no consecutivo empezando por el número uno, el *abstract* es el IOD, y el *Transfer Syntax* indica cómo se va a codificar. La asociación es enviada al SCP para que sea evaluada por éste. El SCP solo aceptará un *Transfer Syntax* por *Abstract Syntax*. La figura 2 muestra los pasos que son necesarios para establecer y cerrar una asociación.

Una vez establecida la asociación es factible enviar los mensajes. Para efectos prácticos, en la figura 3 sólo se muestra la asociación, solicitud de un C-Store y cierre de la asociación de un SCU a un SCP.

### 3. Metodología

Inicialmente se elaboró un documento base, como lo es para cualquier proyecto, el Documento de Requerimientos. Éste se llenó con el mayor cuidado, para que fuera conciso, no ambiguo y claro. Este documento ha sido también la base para realizar las pruebas o *test cases*.

Como todo comienzo en un proyecto de software, se seleccionó el mejor modelo de desarrollo. De los modelos existentes se evaluaron dos, el de cascada clásico y prototipado evolutivo. Se decidió por el modelo prototipado evolutivo por no ser cerrado, a diferencia del ciclo de vida clásico. Éste permite hacer el análisis, diseño, codificación y pruebas en cada etapa del proyecto. La ventaja de este modelo es que de encontrar un error en cualquier etapa, el esfuerzo para corregirlo es mínimo, a diferencia del modelo de cascada donde el esfuerzo es mayor. Otro motivo por el cual se seleccionó fue la falta de información y experiencia sobre el desarrollo de proyectos semejantes.



Fig. 3 Comunicación DICOM (C-Store)

Un objetivo claro fue el separar la comunicación entre cualquier aplicación que se desarrolle y su SCU, ya que el mantenimiento será más fácil. Para cumplir con esto se realizó un control ActiveX que tiene la funcionalidad antes mencionada. Cuando es necesario agregar funcionalidad o realizar modificaciones, sólo hay que insertar el control en la aplicación desarrollada.

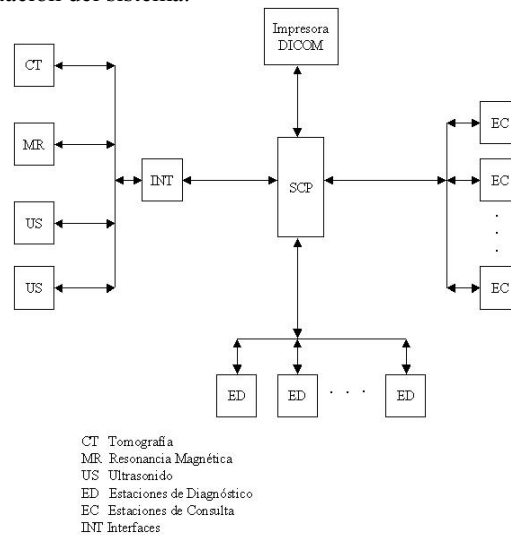
Los diagramas que fueron usados son los siguientes:

- 1) Diagrama de Clase: Permite visualizar de una forma clara los atributos y operaciones de una clase así como su relación con otras.
- 2) Diagrama de Secuencia: Permite ver qué mensajes se envían entre diferentes clases y el orden en que se mandan y se reciben las respuestas. Este diagrama es uno de los más importantes ya que nos indica los pasos a seguir para no estar fuera del estándar.

La codificación fue realizada en VC++ versión 6. Las pruebas de la implementación del estándar fueron realizadas con la ayuda de *casos de prueba*, que permiten realizar una prueba sin que exista una variación al repetirla. Las pruebas fueron basadas en el documento de requerimientos y su objetivo es verificar que los requerimientos que forman del documento fueran cumplidos en el desarrollo. Para lo anterior, se montó un laboratorio con 3 máquinas, una como SCP y dos como SCU. Las computadoras donde se probó el SCU y el SCP tienen las siguientes características: Procesador Pentium IV a 2.4 GHz, Sistema Operativo Windows 2000, Memoria 1 GB en RAM.

#### 4. Implantación

Se instaló un red en el área de radiología, posteriormente se colocó el SCP en el servidor, al cual se le conectó un RAID para el almacenamiento de las imágenes, el tamaño promedio por imagen es de 1 MB con el formato jpeg lossless. El SCP atiende por el momento tres estaciones de trabajo, varias interfaces, teniendo un total de 8 equipos conectados, los cuales pueden enviar mensajes simultáneamente. La figura 4 muestra la implantación del sistema.



**Fig. 4 Equipos y SW bajo comunicación DICOM**

Esta fase es la más importante porque debe existir una comunicación con las diferentes interfaces y una impresora que soporta el estándar DICOM.

La comunicación con equipos de proveedores que ya cuentan con el estándar DICOM (como equipos de impresión o de Radiografía Computada) está a punto de probarse, aunque se considera, por las pruebas realizadas internamente, que no habrá mayores problemas.



## 5. Resultados

El tener el SCU en un control ActiveX fue una buena práctica ya que existieron cambios que no afectaron a otro grupo que desarrolló aplicaciones por lo que fueron transparentes las modificaciones realizadas al control.

Las pruebas se desarrollaron con éxito, salvo pequeños problemas que se mencionan adelante. La comunicación se estableció sin dificultades mayores así como el envío del mensaje C-FIND, C-STORE y C-GET de un SCU a un SCP. Las imágenes enviadas por el SCP al SCU fueron enviadas sin mayores contratiempos. Entre los problemas encontrados están: a) la falta de atención durante el cambio de plataforma de software, al olvidar algunas bibliotecas, b) errores al insertar la imagen en el archivo DICOM durante el proceso de conversión a este formato y c) manejo de diferentes formatos de codificación de los datos (*big-endian* o *little-endian*) en los *Transfer-Syntax* entre el cliente y el servidor. El envío de imágenes a través de la red es de aproximadamente de menos de un segundo para enviar un archivos DICOM de 1 MB aproximadamente.

## 6. Conclusiones

El desarrollo de aplicaciones que deben comunicarse con modalidades es complicado pero no imposible, se necesita invertir mucho tiempo en el análisis y diseño para que el tiempo sea menor en las fases de codificación y pruebas.

Muchas veces no es posible realizar pruebas con equipos o dispositivos externos a adquirir, ya que los proveedores no cuentan con software de simulación de las modalidades y/o impresoras. Lo que puede aumentar el tiempo en el análisis y diseño.

La pérdida de bytes de las imágenes es otro aspecto en el cual hay que invertir un tiempo considerable para verificar que los *transfer syntax* que se manejan correspondan con los *transfer syntax* de las modalidades a los cuales se va a establecer la asociación ya sea de SCU o SCP.

## 7. Referencias

- [1] Página oficial del estándar DICOM, <http://medical.nema.org/>. Verificada por última vez 31 de Mayo del 2004.
- [2] Yongjian Bao, Jurgen Reyinger and Renas Rechid, An Integrated Modality, PACS and RIS Environment using the DICOM Standard, Proceedings of SPIE, Medical Imaging 1998, PACS Design and Evaluation: Engineering and Clinical Issues, Volume 3339, February 1998, San Diego, CA. págs. 303-315.
- [3] David A. Clunie, DICOM Implementations for Digital Radiography, Advances in Digital Radiography: RSNA Categorical Course in Diagnostic Radiology Physics 2003, págs. 163-172.
- [4] Albert W.K. Wong and H. K. Huang, Integration of multi-vendor imaging equipment with DICOM into PACS, Proceedings of SPIE, Medical Imaging 1998, PACS Design and Evaluation: Engineering and Clinical Issues, Volume 3339, February 1998, San Diego, CA. págs. 294-302.
- [5] Takaya Sakusabe, Baatar Shirchin and Michio Kimura, Expand the Internet Standard to the DICOM Standard: DICOM-QR URL Scheme, Proceedings of SPIE, Medical Imaging 1998, PACS Design and Evaluation: Engineering and Clinical Issues, Volume 3339, February 1998, San Diego, CA. págs. 316-321.
- [6] Jorg Riesmeier, Marco Eichelberg, Klaus Kleber, Dietrich H. W. Gronemeyer, Herman Oosterwijk, Peter Jensch, DICOM image display consistency: a test environment, Progress in Biomedical Optics and Imaging, Medical Imaging 2001, PACS and Integrated Medical Information Systems: Design and Evaluation, Vol. 2, No. 28, ISSN 1605-7422, Volume 4323. San Diego, CA., February 2001, págs. 47-56.
- [7] O. Baujard, J.C. Staub, D. Blot, D. Bandon, Y. Ligier, DICOM and XML: union makes strength, International Congress Series 1230, Elsevier Science B.V., 2001, págs. 820-823.

# SAPeR: un sistema para la administración de pacientes en rehabilitación

Lizbeth O. Escobedo Bravo, Mónica E. Tentori Espinosa  
Ana Isabel Martínez García, Leonardo Galicia Jiménez, J. Antonio García Macías

Centro de Investigación Científica y de Educación Superior de Ensenada, CICESE  
e-mail : { escobedo, mtentori, martinea, lgalicia, jagm }@cicese.mx  
Km. 107 Carretera Tijuana-Ensenada Apdo. Postal 2732, CP 22860, Ensenada, B. C.,  
México.

**Abstract:** Nowadays the organizations have been incorporating the technology in their workspaces to improve their process. The medical field isn't the exception, they are incorporating the Information Technology (IT) to their activities to make an effort to get better health services to the patients. In this paper, we present the development of a case study in the area of physical medicine and rehabilitation in a hospital and the development of technological support according to the needs and of the process nature, that enables through the use of mobile devices the enactment of the process with the IT.

**Resumen:** Con el fin de automatizar y optimizar sus procesos, hoy en día las organizaciones están optando por incorporar la tecnología en sus áreas de trabajo. El campo médico no es la excepción, en sus esfuerzos por mejorar los servicios de salud a sus pacientes, están integrando la Tecnología de la Información (TI) a sus procesos. En este artículo, se presenta un caso de estudio desarrollado en un hospital en el área de medicina física y rehabilitación y el desarrollo de un soporte tecnológico de acuerdo a las necesidades y naturaleza del proceso, que permite una sana convivencia entre el desarrollo de las actividades realizadas y la utilización de TI, a través del uso de dispositivos móviles.

**Palabras Clave:** Procesos, sector salud, hospital

## 1. Introducción

Actualmente las organizaciones enfrentan los efectos de los cambios tecnológicos y de una mayor dinámica en la realización de sus actividades [4]. Por esta razón, surge la necesidad de llevar a cabo estudios que les permitan comprender la forma en la que llevan a cabo sus actividades. Con la finalidad de lograr un entendimiento común de los procesos existentes, se requiere del uso de metodologías y técnicas para su análisis. La *Ingeniería de Procesos (IP)* nos provee de herramientas para su estudio. En este contexto definimos un proceso como “*un conjunto de pasos parcialmente ordenados con la finalidad de alcanzar un objetivo*” [3], y la IP como la colección de técnicas para el análisis, diseño y evaluación de los procesos organizacionales, basados en el uso del modelado de procesos [1]. En donde los modelos nos permiten tener una representación abstracta de la realidad. El propósito de estos modelos es reducir la

complejidad en su entendimiento, eliminando los detalles que no influyen el comportamiento de un fenómeno [2].

Con estos estudios, es posible capturar y entender el comportamiento de los procesos, y así buscar opciones para mejorar la forma en la que se lleva a cabo el trabajo, en caso de ser necesario. En este sentido las organizaciones en el área de salud no son una excepción. Debido a que el trabajo en los hospitales ha sido descrito normalmente como un ambiente en el que el personal médico desarrolla su trabajo en constante movilidad realizando múltiples actividades. Realizamos un caso de estudio en un hospital público en el área de medicina física y rehabilitación (MFyR), donde se analizaron sus procesos y se propuso soporte tecnológico como una propuesta a la mejora a sus procesos.

En este trabajo se presenta la realización de este estudio junto con el desarrollo tecnológico propuesto. El artículo inicialmente describe la metodología utilizada para el análisis de los procesos en el área. Enseguida se presenta el caso de estudio realizado y las especificaciones y aspectos de diseño del sistema de soporte propuesto, SAPeR. Posteriormente se presenta la funcionalidad del sistema a través de un escenario de uso.

## **2. Metodología**

Para la realización del caso de estudio, utilizamos la metodología para el análisis y modelado de procesos, conocida como PADM (Process Analysis and Design Methodology) [1]. Esta metodología consta de cuatro etapas: *definición del proceso*, involucra el establecimiento de los objetivos del proceso dado, sus principales entradas y salidas; *captura y representación del proceso base*, que consiste en modelar el proceso detalladamente (una representación gráfica); *evaluación del proceso*, el análisis, la identificación de problemas y debilidades en el proceso; y *el diseño del proceso*, que involucra el diseño de un nuevo proceso para la organización, ya sea por mejoras incrementales o cambios radicales al proceso base incluyendo una propuesta de soporte tecnológico.

Una vez realizado el estudio del proceso, se obtuvo un entendimiento claro de las actividades que se llevan a cabo en el área de MFyR, y en base a esto se determinaron las necesidades de soporte tecnológico para el proceso y los requerimientos para su desarrollo. En la siguiente sección se presentan los aspectos generales del caso de estudio y del proceso.

## **3. Caso de estudio: Proceso de Medicina Física y Rehabilitación.**

El área MFyR brinda soporte y atención al derechohabiente que ha sufrido algún trauma físico. Durante la realización del caso de estudio primeramente identificamos a las personas involucradas en este proceso (médicos, terapeutas y asistentes médicos) y nos familiarizamos con las actividades que desempeñan. Mediante técnicas de recopilación de información, se analizaron los procesos, se crearon modelos de ellos, y se propusieron mejoras y soporte a los mismos, enfatizando el uso de la TI apropiada.

El objetivo principal fue conocer cuales son las actividades que llevan a cabo los participantes en el proceso, como las ejecutan, que limitantes le impiden realizar eficientemente su trabajo, y cuales son las propuestas de mejora a sus procesos. El proceso que se lleva a cabo en el área de MFyR, puede ser visualizado en tres etapas: administración del área (a cargo de la asistente), valoración a pacientes (a cargo del médico) y aplicación de terapias (a cargo del terapeuta). Los cuales se describen en tres escenarios:

*Escenario 1.* El paciente llega por primera vez al área y se dirige con la asistente. Esta crea una cita de valoración con el médico. En la fecha de la cita, el paciente ingresa al área, confirmando su cita con la asistente. Una vez que el paciente ingresa al consultorio médico, el médico le realiza una valoración. En base a esta evaluación, genera un diagnóstico de la salud del paciente e indica el tratamiento a seguir para su rehabilitación. Esta información es capturada por el médico en un expediente clínico (utilizado en todas las áreas del hospital) y en otro expediente interno (que se maneja internamente en el área de MFyR) de forma manual en papel mediante notas médicas y de terapias. Cuando terminan las valoraciones, de acuerdo a los diagnósticos del médico la asistente asigna pacientes a los terapeutas para las posteriores terapias que deberán ser aplicadas.

*Escenario 2.* El paciente llega a su cita de terapias, se dirige directamente con su terapeuta, el cual aplica el tratamiento indicado por el médico de consultorio. El terapeuta registra sus observaciones de seguimiento de terapias en el expediente interno mediante notas de terapias.

*Escenario 3.* Se presenta en caso que sea necesario brindar atención a algún paciente fuera del área de MFyR. El médico y el terapeuta deben trasladarse al lugar solicitado para proporcionar sus servicios. El médico registra sus observaciones en el expediente clínico y en el expediente interno mediante notas médicas. El terapeuta registra sus observaciones en el expediente interno mediante notas de terapias.

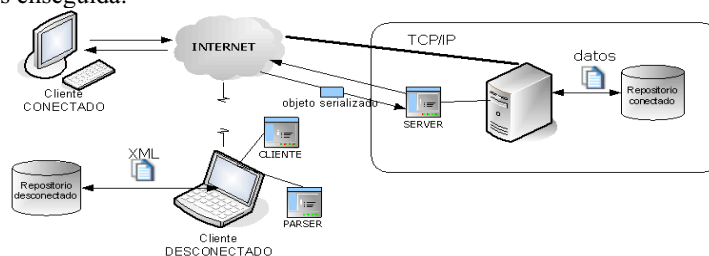
Con lo anterior, podemos observar que el trabajo de los médicos y terapeutas requiere cierta movilidad, debido a que la atención al paciente es indistintamente generada dentro o fuera de MFyR es por ello que se requiere contar con la información necesaria en todo momento. De la definición del proceso y los aspectos encontrados durante el caso de estudio, se identificaron las especificaciones y aspectos de diseño, que se presentan en la siguiente sección.

#### **4. Especificaciones y aspectos de diseño de SAPeR**

El uso de escenarios es una técnica que permite identificar especificaciones de diseño de sistemas de software [5]. La información recopilada del caso de estudio y los diferentes escenarios identificados, permitieron hacer una recolección de las especificaciones básicas que deben ser contemplados para desarrollar una herramienta que proporciona soporte a los procesos del área de MFyR, optimizando las tareas realizadas por el personal médico. Entre estas especificaciones se encuentran: (1) Soporte guiado en la administración y configuración de las actividades realizadas, para llevar a cabo su trabajo diario. (2) Apoyo en la captura y recuperación de información para realizar actividades dentro de un espacio físico por un período de tiempo

prolongado en movilidad entre diferentes espacios físicos. (3) Apoyo en la sincronización de información manejada en movilidad y en un ambiente local.

Con esto en mente lo siguiente fue encontrar los aspectos de diseño que debe contener la arquitectura para brindar soporte a los requerimientos mencionados. Chandra [6] especificó que una operación desconectada, en la cual un cliente accede a un servicio sin conexión a una red, es crucial para mejorar la disponibilidad, soportar movilidad y proveer ejecución independiente. De acuerdo a esta definición y las condiciones de movilidad dentro de MFyR, se diseñó una arquitectura en la cual los usuarios interactúan con dispositivos móviles desconectados a la red y dispositivos estacionarios con conexión a la red. Permitiendo también, establecer la sincronización entre los dispositivos. La arquitectura se basa en los principios del paradigma cliente-servidor [7]. En esta arquitectura, los principales componentes son: los clientes, el servidor y la infraestructura de comunicaciones, los cuales ilustramos en la figura 1 y describimos enseguida.



**Fig.1** Arquitectura del sistema inalámbrico

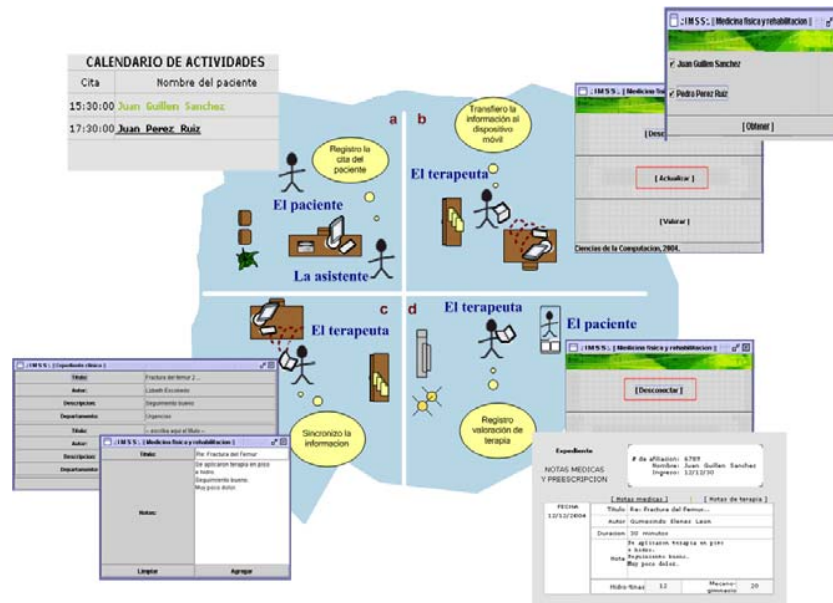
1. Cliente conectado. Posee una conexión continua a Internet. Su repositorio de información puede ser una base de datos, archivos, entre otros.
2. Cliente desconectado. Este toma dos roles: (a) *Con conexión*. En este rol actúa para sincronizar los dispositivos y debe tener conexión a Internet, (b) *Sin conexión*. Este no tendrá conexión a Internet, pero podrá manipular y/o modificar la información de su repositorio.
3. Servidor. Esta aplicación mantiene las conexiones de los clientes al servidor, administra los paquetes enviados por los clientes y realiza las llamadas a las funciones requeridas por el cliente mediante peticiones.

A continuación se muestra un escenario que ejemplifica el uso del sistema SAPER ilustrando el proceso de MFyR.

## 5. Sistema de administración de pacientes en rehabilitación

Como se mencionó el sistema desarrollado consta de dos subsistemas, uno que se encuentra permanentemente conectado a la red mediante una computadora de escritorio y el otro que permite la desconexión-conexión de un dispositivo móvil para ejecutar las actividades en movilidad. El sistema de escritorio proporciona soporte para: el ingreso de pacientes al área de MFyR, actualización de valoraciones y consulta del seguimiento de los pacientes. Mientras que el sistema móvil, permite realizar la desconexión y

actualización de información entre dispositivos, además de la actualización de valoraciones y consulta del seguimiento de los pacientes. De esta manera el sistema brinda soporte a los tres escenarios descritos.



**Fig.2** Soporte tecnológico para el área MFyR. (a) Configuración de citas. (b) Desconexión de información. (c) Valoración en dispositivo móvil (d) Actualización de información.

A continuación se describe el uso del sistema en el escenario del terapeuta, ilustrado en la figura 2. El paciente llega con la asistente para realizar su cita de valoración. Esta a través del sistema solicita los datos del paciente y registra su cita de valoración médica o terapia (fig. 2a). Posteriormente, el paciente llega a su cita al área de MFyR y pasa a consulta. Mientras tanto, el terapeuta revisa en su sistema estacionario los pacientes que deberá de atender. En ese momento, transfiere la información de los pacientes a ser atendidos a su dispositivo móvil (fig. 2b). Una vez realizada la transferencia, el terapeuta mediante su sistema móvil consulta y almacena la información relevante, mientras aplica la terapia al paciente (fig. 2c). Finalmente, el terapeuta actualiza la información entre el sistema móvil y el sistema de escritorio (fig. 2d).

## 6. Conclusiones y perspectivas

En el presente artículo, hemos propuesto un sistema que brinda soporte a las actividades realizadas por el personal de MFyR, donde estos últimos requieren realizar su trabajo entre diferentes áreas y manejar cierta información en movilidad. Debido a la naturaleza de los escenarios, donde la infraestructura de conexión a una red no es

constante, se propuso la alternativa de desconexión como medio para dar soporte a trabajadores móviles, mejorando y optimizando el desarrollo de sus actividades.

El caso de estudio, así como la propuesta de soporte tecnológico, fueron presentados a los médicos del hospital, donde se obtuvo retroalimentación del proceso. Los médicos se mostraron satisfechos con el trabajo realizado, y externaron su interés por la utilización de la propuesta de la tecnología presentada. En este sentido al momento la directiva del hospital se encuentra haciendo los trámites necesarios para la adquisición del equipo y la implantación del sistema. Con esto se lograría eficiencia en el proceso que actualmente se lleva de forma manual y con mucha duplicación de información y trabajo, permitiendo a los médicos y terapeutas enfocarse más a la actividad clínica-terapéutica al apoyar la captura, manejo y procesamiento de la información generada dentro de las actividades del área de MFyR por medio del sistema móvil.

Aunque la aplicación desarrollada, es extensible y simple, no es independiente. Es decir, es necesario que el usuario indique la información que requiere ser transportada al dispositivo móvil. Probablemente sería interesante proveer agentes inteligentes, que identifiquen de manera transparente una posible desconexión y sincronicen la información cuando la conexión sea recuperada de manera automática. De esta manera el usuario no deberá identificar cuándo y qué información debe ser desconectada.

## **Agradecimientos**

Este trabajo es apoyado por CONACYT, con las becas 179370, 179381 y 179371 de los dos primeros y cuarto autores respectivamente, y con el proyecto C01-40799. Agradecemos también el apoyo de los médicos Simitrio Rojas y Julia Mora, y a los médicos del área de Medicina Física y Rehabilitación del hospital.

## **Referencias**

- [1] Wastell, D.; White, P. y Kawalek, P., "A methodology for business process redesign: experiences and issues". *Journal of Strategies Information Systems*, 1994.
- [2] Curtis, B.; Kellner, M. I. y Over, J., "Process Modeling". *Communications of the ACM*, Vol. 35, No. 9, 1992.
- [3] Humphrey, W. S. y Feiler, P. H., "Software process development and enactment: concepts and definitions". *Proceedings of the second international conference on the software process*. 1992.
- [4] Silver, B., "Domino Workflow white paper. Automating Real-World Business Processes". Bruce Silver Associates. *Industry Trend Reports*, 1999.
- [5] Carroll, J. y Rosson, M., "Getting Around the Tast Artifact Cycle: How to Make Claims and Design by Scenario". *ADM Transactions on Information Systems*, 1992.
- [6] Chandra, B., et. al., "Resource management for scalable disconnected access to web services". *ACM*, 2001.
- [7] Ford, N., "Web Developer.Com Guide to Building Intelligent Web Sites with Java Script". Wiley, New York, 1998.

# **SNIAC - Sistema Nutricional Inteligente Asistido por Computadora**

Ramírez Sánchez Jorge Luis, Cruz Zamora José Antonio

Instituto Tecnológico Superior de San Martín Texmelucan  
Centro de Cómputo  
Calle Miguel Nájera No. 22 Col. Centro C.P. 74060.  
Tel. 012484873909

**Palabras clave:** Nutrición, Sistema de Inferencia Difuso, Análisis Clínico, Mamdani, Sugeno, Tsukamoto.

## **1.-Abstract**

One of the main problems that affect our country is malnutrition, which is shown regularly in children among 0 to 12 years old, infant mortality in most of the cases is originated by bad feeding habits. The present paper describes the structure of an expert fuzzy system that has the purpose of taking information generated by the para-clinic study of seric Albumin to determine the health state of the infant using a system of fuzzy inference and produce messages similar to those of the experts in nutrition using Fuzzy Logic.

## **1.-Resumen**

El principal problema que afecta a nuestro país es la desnutrición, la cual se manifiesta regularmente en niños entre los 0 meses y los 12 años de edad [1,5], la mortandad en infantes en la mayoría de los casos es originada por una alimentación inadecuada [2]. El presente trabajo propone el uso de Lógica Difusa (LD) en el área de diagnóstico clínico así como una simple demostración del desempeño de los modelos propuestos por Mamdani, Sugeno y Tsukamoto. Además describe la estructura de un sistema experto difuso que tiene el objetivo de tomar los datos generados por el Estudio Paraclínico de la Albúmina Sérica [2,3], determinar el estado nutricional del infante y mediante un sistema de Inferencia Difusa generar mensajes similares a los del experto en nutrición utilizando LD.



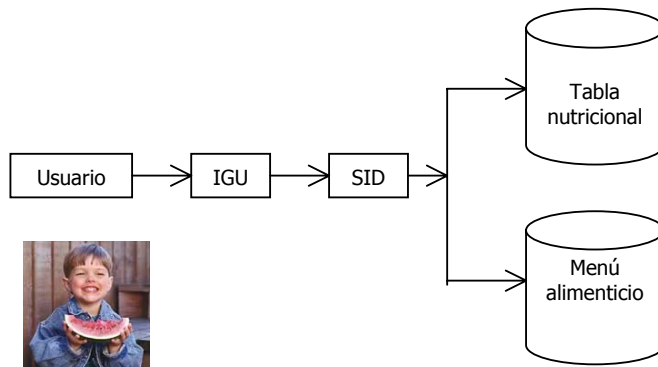
## 2.-Estado del arte.

Actualmente la fase del presente artículo muestra el desarrollo del Sistema Inteligente y su consecuente demostración mediante un lenguaje de investigación. Este es Matlab Ver 5.0.

El objetivo que perseguimos es demostrar la viabilidad de aplicar LD [4] en diagnóstico Clínico, así como elegir al modelo (Mamdani, Sujeno o Tsukamoto) con mejor rendimiento diagnosticando desnutrición en infantes.

El proyecto se encuentra en un avance del 50% ya que el resto contempla la implementación del sistema descrito, en un lenguaje de programación comercial como Visual Fox Pro, este permitirá el manejo de tablas nutricionales que en función de los datos generados por el módulo de inferencia difuso sea capaz de generar dietas balanceadas.

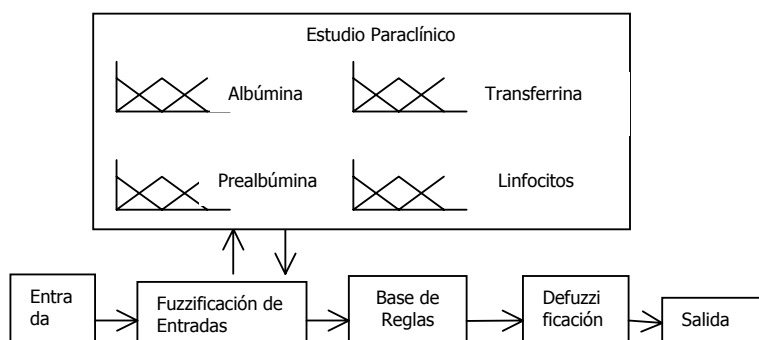
## 3.-Metodología



**Fig(1).** Diagrama general del sistema

El sistema propuesto (**Fig 1**) consta de una Interfase Grafica de Usuario (IGU) la cual captura la información del infante [4] (Albumina Sérica y parámetros antropométricos) para enviarla al Sistema de Inferencia Difuso (SID), éste procesa la información, determina el grado de desnutrición del infante en base a los parámetros de la Albumina Sérica y finalmente genera un menú recomendado.

El SID (**Fig 2**) es el Núcleo de nuestro Sistema, éste se apoya en la función triangular típica utilizándola para definir los conjuntos difusos, estos son cada uno de los parámetros arrojados por el estudio de Albúmina Sérica (albúmina, transferrina, prealbúmina, linfocitos **ver tabla 1**). Para cada uno de ellos utilizaremos cuatro variables difusas [5] que reflejan el estado nutricional del infante: Nula, Moderada, Severa y Extrema, las cuales permiten a través de la base de reglas difusas [6,7] generar un diagnóstico similar al del médico en nutrición con un dato porcentual adicional que refleje el grado de desnutrición del infante.



**Fig(2).** Sistema de inferencia difuso con funciones típicas triangulares.

Indicador	Normal	Desnutrición Leve	Desnutrición moderada	Desnutrición Severa
Proteínas Plasmáticas Albúmina (mg/dL)	3.8-5.0	2.8 - 3.5	2.1 -2.7	<2.1
Transferrina (mg/dL)	200 – 400	150 – 200	100 -150	<100
Prealbúmina (mg/dL)	20 – 36	10 -15	5 -10	<5
Linfocitos (mm <sup>3</sup> )	2000 – 7000	1200 – 2000	800 – 1200	<800

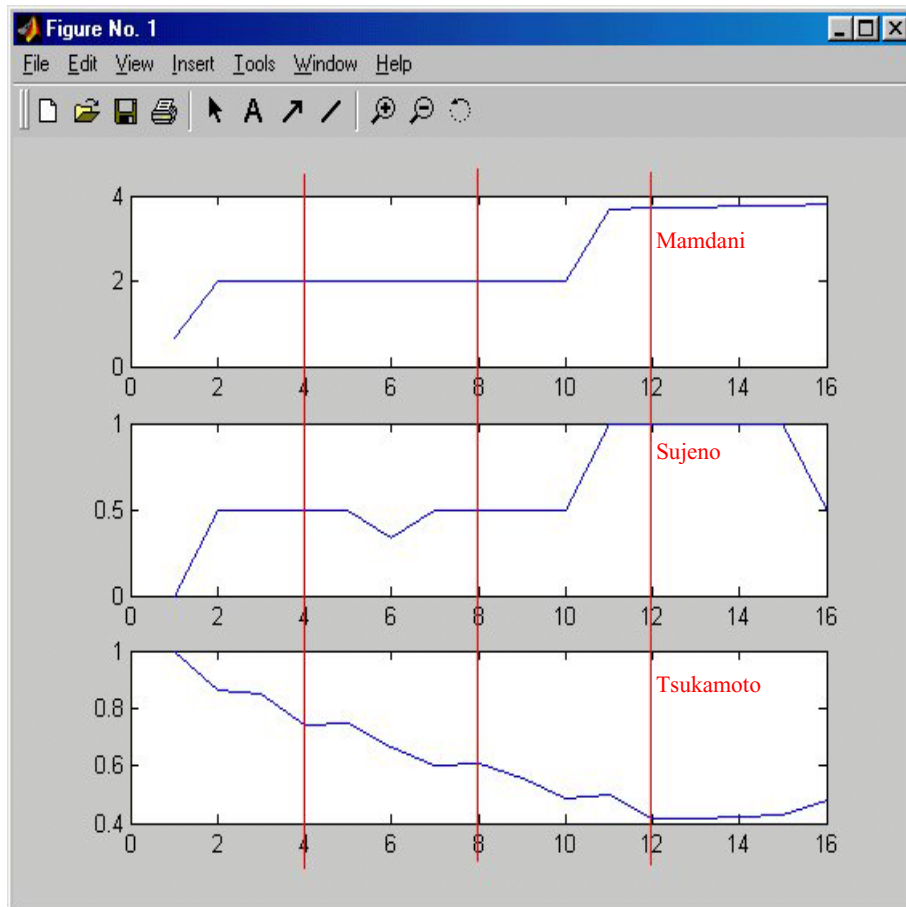
**Tabla (1).** Rangos esperados en el Estudio Paraclínico Albúmina Sérica.

De esta manera se valora el rendimiento de los modelos propuestos por Mamdani, Sujeno y Tsukamoto [8] modelando cada uno de ellos con el mismo problema, es decir se definen los mismos conjuntos de entrada, reglas difusas y los conjuntos de salida para finalmente medir los resultados en función de los rangos esperados.

## 4. Resultados obtenidos

Para demostrar la viabilidad de nuestro sistema hemos dividido a cada uno de los rangos del estudio Albúmina Sérica en 16 partes para ser procesadas por nuestro sistema con la finalidad de obtener 16 diagnósticos.

Por lo tanto se espera generar, 4 casos de desnutrición extrema, 4 de severa, 4 de moderada y 4 de Nula. Fig 3.



**Fig 3.** Diagnóstico de Mamdani Sujeno Tsukamoto.

En el eje X se refleja un rango entre 0 y 16 el cual a sido dividido en cuatro partes iguales: el primer rango que inicia en 0 y termina en 4 corresponde a 4 casos en los que se espera que los sistemas determinen desnutrición “Extrema”, el segundo que va de 4 a 8 corresponde a los casos esperados de desnutrición “Severa”, el tercero que va de 8 a 12 es para los casos esperados de desnutrición “Moderada” y el último que corresponde a los casos de desnutrición “Nula” que va desde 12 a 16 unidades.

El eje Y es el rango esperado de respuesta de nuestros tres sistemas, este comienza en 0 y termina en 4, por lo que si obtenemos un valor difuso entre 0 y 1 (para Mamdani y Sujeno), [1,2], [2,3], [3,4]el diagnóstico es desnutrición “Extrema”, “Severa”, “Moderada” y “Nula respectivamente”, por otro lado Tsukamoto por su naturaleza (Función monotonica) solo responderá con un rango entre 0 y 1.

## 5. Conclusiones

La figura 4 muestra gráficamente los resultados obtenidos, 16 casos que han sido sometidos a evaluación de los tres modelos (Mamdani, Sujeno, Tsukamoto) y para los cuales cada uno de ellos a dado un diagnóstico.

La gráfica (**Fig 3**) muestra claramente como Mamdani y Sujeno tienen un comportamiento similar, cumpliendo con los rangos esperados y un tanto rígidos. Por otro lado Tsukamoto únicamente responde con valores entre 0 y 1 quedando desplazado por Mamdani y Sujeno, frente a esta desventaja el presenta un comportamiento más suave.

Queda demostrada la viabilidad de implementación de nuestro sistema teniendo como primer alternativa el modelo Mamdani junto con el de Tsukamoto y en segundo plano queda Sujeno ya que presenta un comportamiento similar al de Mamdani.

Lógica Difusa es un paradigma que permite desarrollar Sistemas Inteligentes Aplicados al Diagnóstico Clínico con una economía considerable de recursos y fácil implementación.

## 6. Referencias

- [1] <http://www.fondonestlenutricion.org.mx>
- [2] Diagnóstico del estado nutricional en la infancia EDGAR M. VÁSQUEZ GARIBAY. ACTA PEDIATR MEX 1998; 19(S): 1-11
- [3] Estrategias y acciones para corregir deficiencias nutricias *Dr. Juan Angel Rivera-Dommarco\** Instituto Nacional de Salud Pública, Cuernavaca, Morelos, México.
- [4] Orientación Nutricional Especializada Lic. Fabrizio Delissalde Av. Convento de Sta Mónica # 79, Jard. de Santa Mónica, Tlalnepantla. Edo. Mex. 397\_1574 397\_7302
- [4] Fuzzy and Applications. MJ Patyra, Dm Miynek. Willey & Teubner
- [5] <http://www.inegi.gob.mx/inegi/default.asp>
- [6] <http://delta.cs.cinvestav.mx/~gmorales/ldifl/ldifl.html>
- [7] [http://answermath.com/logica\\_difusa\\_conjuntos\\_nebulosos.htm](http://answermath.com/logica_difusa_conjuntos_nebulosos.htm)
- [8] Redes Neuronales y Sistemas Difusos. Segunda Edición Ampliada y Mejorada. Bonifacio Martín del Brío, Alfredo Sanz Molina. Alfaomega&Ra-Ma



## Índice de autores

- Aceves, Rita M. 281  
Acosta, Ricardo 96  
Acosta Mesa, Héctor Gabriel 641, 669  
Aguilar Cisneros, Jorge 47, 347  
Agus, Marco 188  
Alejandro Castro, Noé 289  
Alfredo Sánchez, J. 56, 571  
Álvarez, Omar 62  
Álvarez R., Francisco 338, 393  
Andrade, María 96  
Angel Vera, José 235, 289  
Arámbula Miranda, Luis Enrique 503  
Arenas Valdés, Ángeles 405  
Arias-Torres, Dante 553  
Arriola-Arreola, Edgar 595  
Avetisyan, Arutyun 481, 489  
Ayala Ramírez, Víctor 540
- Balladares Ocaña, Leandro 356  
Barreto, Gerardo 62  
Barreto Palacios, Gerardo 5  
Bayona, Sofía 178  
Bolshakov, Igor A. 289  
Botello, Alejandro 439  
Bravo-Hernández, Silvana 559  
Briones-García, J. Alfonso 559  
Brizuela, Carlos A. 623  
Buckley, David 641, 669  
Bustos Mendoza, C. R. 675
- Cardona S., Juan P. 338, 393  
Carreto Arellano, Chadwick 565  
Castañeda-Roldán, C. 629  
Castellanos, Lucía 617  
Castellanos, Nohema 571  
Castillo Díaz, Víctor D. 577  
Cervantes López, Teresa 425  
Cervantes Medina, L. A. 651  
César Aguilar, Julio 446  
Contreras-Castillo, Juan 5  
Cortés Verdín, Karen 13  
Cristóbal-Salas, Alfredo 330  
Cruz Ramírez, Nicandro 641, 669  
Cruz Zamora, J. A. 693
- Díaz, Marissa 170  
Díaz, Pablo D. 251  
Dinora Orantes, Sandra 439  
Drogoul, Alexis 522  
Duran-Limon, Hector A. 583
- Escamilla Alonso, B. D. 663  
Escobedo Bravo, L. O. 687  
Espadero, Jose M. 178
- Fairén, Marta 159  
Favela, Jesus 79, 607  
Feregrino Uribe, C. 681  
Fernández, Jose M. 178  
Fernández M., Luis Felipe 399  
Fernández Peña, Juan Manuel 405, 433  
Fernández Ramírez, Lourdes 56  
Fraguela, Liset 243  
Frisby, John 641, 669
- Gaissaryan, Sergey 481  
Galicia Jiménez, L. 657, 687  
Galindo-Rodríguez, Angel 595  
Gallardo, Román 96  
Gama-Moreno, Luis A. 601  
Gamboa, Fernando 86  
Garcés-Eisele, J. 629  
García, Alberto 56  
García, Marcos 178  
García Macías, José Antonio 522, 589, 657, 687  
García Ruiz, M. A. 651, 675  
García Tamayo, Cutberto 356  
García-Martínez, Christian P. 589  
García-Valdez, José Mario 330  
Garduño, Gabriel 219  
Gasca, Christopher 617  
Gelbukh, Alexander 311  
Giachetti, Andrea 188  
Gobbetti, Enrico 188  
Gómez-Gil, P. 629  
Gonzalez, Jesus A. 635  
González, Victor M. 79  
González Calleros, Juan M. 47

González-Gurrola, Luis C. 623  
 Grushin, Dmitry 497  
 Guerra, Noemí 96  
 Gutiérrez Almaraz, Beatriz 446  
 Gutiérrez Tornes, Agustín 565  
  
 Hernández, Benjamín 121  
 Hernández García, L. 675  
 Hernández Rodríguez, María de Lourdes 433, 446  
 Herrera, Oscar 617  
  
 Ivannikov, Victor 481  
 Iwasaki, Yohei 129  
  
 Jiménez, Héctor 203, 212, 243, 267, 296, 304  
  
 Kaneko, Toyohisa 129  
 Kolčárek, Pavel 149  
 Kuri, Angel 617  
 Kvaratskhelia, Alexander 474  
  
 Lazarev, Alexander 465, 474  
 Leontyev, Hennadiy 481  
 Lepe-Aldama, Oscar-Iván 595  
 López, Aurelio 227, 274  
 López, Gilberto 19  
 López Mendoza, Salvador 366, 373  
 López Rosas, D. M. 675  
 López-Morteo, Gabriel 19  
 Lozano-Yécora, J. 629  
 Lucero, Cupertino 203  
 Lugo, Tania 311  
  
 Martínez, Edgar A. 607  
 Martínez García, A. I. 657, 687  
 Martínez-Muñoz, Jorge 601  
 Mayhew, John 641, 669  
 Mayora Ibarra, Oscar 28, 72, 512  
 Medina, Alfonso 111, 219  
 Medina Hernández, Misael 323  
 Menchaca García, Rolando 565  
 Menchaca-Méndez, Rolando 565, 577  
 Millán, Erik 139  
 Miramontes Rivera, Adriana 19  
 Miranda-Palma, Carlos 72  
  
 Mizuno, Isamu 129  
 Montes, Manuel 227, 251, 281  
 Montes de Oca, Carlos 399  
 Mora Lumbreras, Marva Angélica 323  
 Morales, Rafael 366, 373  
 Moreno, Marlene 111  
 Moreno Rocha, Mario Alberto 105, 387  
 Morris, Janet 641, 669  
 Moyotl, Edgar 212  
 Muñoz Arteaga, Jaime 47, 347, 405, 418, 425  
 Muñoz Meléndez, Angélica 530, 545  
  
 Navazo, Isabel 159  
 Nolazco, Juan A. 251  
  
 Olmedo-Aguirre, J. Oscar 559  
 Olmos, Ivan 635  
 Ortiz, Martha 617  
 Ortiz Sánchez, Edgar 323  
 Osorio-Galindo, M. 629  
  
 Pacheco-Sánchez, José Antonio 512  
 Padaryan, Vartan 481  
 Padilla, Gerardo 399  
 Paniagua, Sonia 304  
 Peralta Calvo, Rosario 105  
 Peredo Valderrama, Rubén 356  
 Pérez, M. 227  
 Pérez y Pérez, Rafael 86  
 Pérez-Navidad, Frankli-Augusto 595  
 Pinto, David 203, 267, 296, 304, 522  
 Pomares Hernández, Saúl 347  
 Ponce, Carlos 617  
  
 Raffin, Bruno 457  
 Ramírez Sánchez, J. L. 693  
 Rodríguez, Miguel 296  
 Rodríguez Urrea, Marcela 79, 607  
 Rodríguez-Muro, M. 629  
 Romo Contreras, Maribel 418  
 Rudomin, Isaac 170, 121, 139  
 Ruiz Rodríguez, Ricardo 105, 387  
 Ryzhov, Andrey 497  
  
 Sadykov, Ruslan 465  
 Salazar, Hilario 267

Salgado, Luis R. 251  
Samovarov, Oleg 497  
Sánchez, Abraham 522  
Sánchez Arias, Víctor Germán 366  
Sánchez Herrera, Héctor 405  
Sánchez Martínez, Fernando 28  
Sánchez Yañez, Raúl E. 522  
Sandoval, Sara 62  
Santana Mancilla, P. C. 657  
Sidorov, Grigori 235, 289, 311  
Sierra, Gerardo 111, 219, 243  
Sochor, Jiří 149  
Solares, Eduardo 111  
Solorio, Thamar 274  
Sumano López, María de los Ángeles  
13, 405, 433  
  
Tchernykh, Andrei 474  
Téllez, Alberto 259  
Tentori Espinosa, M. E. 687  
Theoktisto, Víctor 159  
Torres Tovar, José Guillermo 418  
  
Treviño, J. A. 681  
  
Uraga-Serratos, Esmeralda 72  
  
Valdez Batista, Ironelis 5, 62  
Valdivia-Beutelspacher, Roberto 512  
Vargas Cerdán, Dolores 13  
Vázquez Valerdi, Claudia 37  
Velázquez Navarro, Tirso 411  
Verduzco, Jesus 457  
Villaseñor Pineda, Luis 37, 251, 281  
Vizcarra-Corral, Luis-Enrique 595  
  
Weitzenfeld R., Alfredo 393  
  
Yañez Suárez, O. 663  
Zanetti, Gianluigi 188  
Zapata, René 522  
Zheng, Ying 641, 669  
Zorcolo, Antonio 188