Editorial

THIS issue of *Polibits* includes a thematic selection of papers on Information Retrieval and Natural Language Processing. Information Retrieval comprises the technologies that help searching for documents in large text collections or Internet. Natural Language Processing, also known as Computational Linguistics, offers improvements to many applications that involve computers and human language, such as information retrieval, spelling correction, and machine translation, among others.

The **special section on Information Retrieval and Natural Language Processing** includes the first seven papers. First three of them are related with **Information Retrieval** and **Question Answering**. Question answering is a specific kind of information retrieval: instead of searching for whole documents, it goes a step further by providing the user with the answer to a specific question. For example, if the user wants to know *who is the president of Mexico*, an information retrieval application will present them all the documents that mention the words *president of Mexico* and the user themselves can manually look for the name in those documents; a question answering application will do it automatically and return the name: *Felipe Calderon*.

The paper "*TrainQA: a Training Corpus for Corpus-Based Question Answering Systems*" describes the development of lexical resource useful for the question answering task: a corpus of questions and their corresponding answers marked in the text. This corpus can be used for both supervised training of question answering systems and as a benchmark for such system.

The paper "*Cross Language Information Retrieval using Multilingual Ontology as Translation and Query Expansion Base*" evauates an information retrieval system of a special type: cross language information retrieval system, i.e., a system that accepts a search query in one language (English in this case) but searches for relevant documents written in another language (Arabic). The paper demonstrates that using an ontology yields significantly better results than using a simple dictionary for translation.

The paper "*English-to-Japanese Cross-Language Question-Answering System using Weighted Adding with Multiple Answers*" is devoted to a topic that combines the ideas of question answering and cross-language information retrieval: now, not only the question can be formulated in a language different from that of the documents, but the answer can be looked in texts in different languages.

The next two papers are devoted to the internal tasks of natural language processing and information retrieval, namely, to **word sense disambiguation** and **named entity extraction**.

The paper "*Using Sense Clustering for the Disambiguation of Words*" is devoted to disambiguation of homonymous or polysemous words in a specific context. For example, the word *bank* means different things in the contexts *bank account* and *West bank of Jordan*. The paper presents a method for automatically determining the correct sense of the word in context by determining groups of similar senses.

The paper "*Improving Named Entity Extraction Accuracy using Unlabeled Data and Several Extractors*" shows how several methods can be combined to improve the quality of identifying named entities in natural language documents. A named entity is a sequence of words that refer to a single concept, such as *Ministry of Foreign Affairs* or *John Smith*.

The last two papers of the special section address **spelling correction** in information retrieval context and **lexical resources acquisition**.

The paper "*Revised N-Gram based Automatic Spelling Correction Tool to Improve Retrieval Effectiveness*" presents an improved method for correcting spelling errors in words via statistics of letter combinations. Besides the usual application of the spelling correction in improving human writing, the paper shows that spelling correction is useful for information retrieval by correcting spelling errors in the query and in the documents being searched and thus improving chances for correct matching between the query and the documents.

Finally, the paper "*Bilingual Lexical Data Contributed by Language Teachers via a Web Service: Quality vs. Quantity*" discussed the issue of lexical resources acquisition. Many of the natural language techniques depend crucially on dictionaries and other sources of knowledge about language. Constructing such dictionaries is a major concern for the research community. The paper discusses the authors' experience in involving language teachers in a collaborative project aimed to develop a bilingual dictionary.

This paper concludes the special section on information retrieval and natural language processing. The last five papers are **regular papers**.

The paper "*Tecnología RFID Aplicada al Control de Accesos*" presents a short introduction to the modern radio frequency identification technology that permits remotely identify objects or resources via special tags attached to them. The paper also shows a practical application of this technology in access control.

The paper "*An Extended Payment Model for M-Commerce with Fair Non-Repudiation Protocols*" suggests an improved model for mobile commerce using fair non-repudiation protocols widely used in the context of electronic commerce but not studied in enough detail in the context of mobile commerce. Non-repudiation protocols ensure that the buyer or

seller will not be able to refute the validity of the transaction after it has been concluded.

The paper "*Análisis Numérico de Pérdidas de Inserción de Conmutadores Diseñados con Diodos p-i-n*" gives a detailed introduction in the theory and applications of the *p-i-n* diodes and microwave switches used in high-frequency circuits. The paper then presents numeric analysis of the insertion loss in microwave switches designed with the *p-i-n* diodes.

The paper "*Restricción del Uso de Teléfonos Celulares en Ambientes Controlados*" presents a Bluetooth-based tool that can disable mobile phones at a certain place or block some of their functions such as transmitting of video signal. This allows to enforce the restrictions on use of mobile phones at some public places such as cultural events, banks, or airplanes.

Finally, the paper "*Evaluation of E-Learning Readiness: A Study of Informational Behavior of University Students*" studies the way students in Thailand use information sharing and search in the computer-assisted learning process. The results of this study can be used for elaboration of the part of e-learning curriculum that deals with personal information management and can be relevant for countries with similar cultural and economic situation, such as most of Latin American countries.

Alexander Gelbukh
Head of the Natural Language Processing Laboratory,
Center for Computing Research, National
Polytechnic Institute, Mexico