Editorial

This issue of *Polibits* presents a thematic selection of papers on natural language processing (NLP) and related areas of research and application. NLP, also referred to as computational linguistics or human language technologies, is a relatively new and fast growing area of artificial intelligence, studying how computers can make reasonable use of texts written in human languages such as English or Spanish or naturally interact with humans in their own language.

NLP is closely interrelated with other areas of artificial intelligence and computer science as well as with many areas of linguistics. Many of the results presented in this issue will be useful not only for those interested in this specific area, but for experts and students working in many other areas of computer science, knowledge representation, computer-aided education, human-computer interaction, or linguistics.

The first five papers in this issue discuss various aspects of **knowledge management**, **information retrieval**, and **information extraction**.

Namely, the paper "*Disentangling the Wikipedia category graph for corpus extraction*" proposes BorderFlow, a novel approach to graph clustering, and presents its application to construction of domain-specific corpora in an information retrieval context. The algorithm will be useful for many other natural language processing tasks, such as ontology learning and terminology extraction. I recommend this paper to a wider audience interested in other areas of computer science where this novel graph clustering algorithm can prove to be useful.

The paper "*Semantic enterprise search (but no Web 2.0)*" discusses practical issues arising in managing corporate knowledge or a body of information generated by the researchers of a university. It shows how semantic search can be incorporated in the knowledge management scenario within an organization. The author explains the differences of his proposal from the widely known ideas underlying Web 2.0.

The paper "*Semantic web framework for development of very large ontologies*" addresses some practical issue of interoperability of various information sources, such as formal ontologies and dictionaries (lexical ontologies) within the Semantic Web paradigm. The experiments reported in it were carried out with Russian WordNet, which illustrates handling non-English data and even non-Latin alphabet.

The paper "*FlexIR: A Domain-Specific Information Retrieval System*" presents an architecture of an information retrieval system with flexible configuration for domain-specific queries, considering as case study the medical domain. The user interface and the integration of information sources in a multilingual setting are described.

Finally, the paper "*Mining Reviews for Product Comparison and Recommendation*" presents an automated recommending system capable of recommending commercial products basing on sentiment analysis of the text of customers' reviews on the company website. The system is currently in practical use.

The next two papers in this issue are devoted to the internal art and craft of natural language processing: **morphological analysis** and **bilingual text alignment**.

The paper "*SMM: Detailed, Structured Morphological Analysis for Spanish*" presents an implementation of a morphological analyzer for Spanish language, showing, in particular, implementation decisions for some interesting morphological phenomena of Spanish basing on the formalism of Left-Associative Grammar. The paper will be particularly useful for linguists interested in Romance morphology. On the other hand, it shows a non-trivial application of left-associative grammars, which will be useful for a wide audience specializing in computer science.

The paper "*CLAU – A Service-Oriented System for Complex Language Alignment: Architectural Aspects*" proposes a service-oriented interactive application allowing users to import, evaluate, correct, and share XML-based annotations in parallel texts. The system is intended to be used by linguists to study the issues arising when importing the annotations from one language to another.

The last three papers show different applications of natural language processing techniques in **machine translation**, **distance learning**, and **human-computer interaction**.

The paper "*Application of Pronominal Divergence and Anaphora Resolution in English-Hindi Machine Translation*" addresses the role of anaphoric expressions (such as pronouns) in machine translation, considering as case study the world's second and third (after Chinese) most spoken languages.

The paper "*E-Learning Content Designs and Implementation based on Learners' Levels*" describes a web-based system that evaluates the student's current level of knowledge basing on their answers and helps the teacher to compose learning material customized to the student's level.

Finally, the paper "*Modeling Multimodal Multitasking in a Smart House*" shows how natural language dialogue can be incorporated into multimodal home automation system interacting with the user in controlling a smart house.

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