

Editorial

With a great pleasure I present to the reader this issue of *Computación y Sistemas*, featuring a thematic section on computational linguistics and natural language processing, a topic that steadily gains increasing interest from both research community and industry.

The thematic section includes papers devoted to intelligent natural language human-computer interfaces, sociological findings from the analysis of the Internet users' behavior, the formal logic of natural language, building lexical resources, logical reasoning over natural language texts, machine translation, and natural language generation.

Bayan AbuShawar and Eric Atwell from Jordan and UK in their paper "ALICE Chatbot: Trials and Outputs" describe their experiments on building very large knowledge base for a natural language conversational agent. The knowledge base they built automatically from text corpora contains over a million categories. In the paper they also explain in detail the structure of the popular ALICE chatbot engine. Their main conclusion is that very simple techniques in combination with learning from very large text corpora can produce industrial-strength conversational agent capable of answering questions in natural language.

Nikolai Buzikashvili from Russia in his paper "Query Topic Classification and Sociology of Web Query Logs" present a methodology for sociological research based on the analysis of the search queries of Internet users. He shows that such analysis can contribute to a portrait of a nation, a generation, or a community by revealing what people search for, who those people are, and how they use the search engines.

Marie Duží and Martina Číhalová from Czech Republic in their paper "Questions, answers and presuppositions" analyze from the standpoint of formal logic the presuppositions containing in questions. A presupposition is a statement that must be true no matter what answer is given to the question. For example, the question "Have you quitted smoking?" presupposes that you smoked before, no matter whether you answer "yes" or "no"; similarly, the question "Does your

mother know that you smoke?" presupposes that you do smoke, whether you answer "yes" or "no". Such phenomena are important for our understanding of how the language works in general and specifically for building programs that understand natural language.

Ilia Markov et al. from Mexico and Portugal in their paper "A Rule-Based Meronymy Extraction Module for Portuguese" describe their system capable of automatic extraction of semantic relations of a specific type from texts in Portuguese. Namely, they target the whole-part relation (called meronymy, together with the set-element relation), and specifically, those relations that involve human body: *index finger* is a part of *hand*, which is a part of *human body*.

Rohini Basak et al. from India and Mexico in their paper "Recognizing Textual Entailment by Soft Dependency Tree Matching" show that very simple rule-based approximate matching of dependency trees, even without the use of any other tools or lexical resources, gives state-of-the-art accuracy on the recognizing textual entailment task. This task consists in automatically deciding whether some text logically implies another text, such as "the police has captured John's assassin" logically implies that "John is dead", but not vice versa. This task is crucial in many applications of natural language processing.

Rahma Sellami et al. from Tunisia and Canada in their paper "Improved Statistical Machine Translation by Cross-Linguistic Projection of Named Entities Recognition and Translation" propose a novel method for the treatment of named entities, such as personal names or names of organizations, countries, etc., in machine translation. Such named entities are often translated erroneously by machine translation systems: for example, systems often attempt to translate such words as *Windows*, *Apple*, or *Bush* by the corresponding words for these objects in the target language (*ventanas*, *manzana*, and *arbusto* in Spanish) even when they represent names of a product, company or a person, correspondingly, and thus should be

transliterated or translated by the names traditionally used in the target language.

Belém Priego Sánchez and David Pinto from France and Mexico in their paper "Identification of Verbal Phraseological Units in Mexican News Stories" continue the discussion of automatic semantic disambiguation, specifically targeting verbal multiword expressions, such as "to read between the lines." The meaning of the latter expression in a document is most probably not literally "reading something in the space between the lines of a text" but "understanding what was not said explicitly". The authors present a machine learning technique for detecting the cases where the interpretation of such multiword expression should not be literal. Using their technique, they have built a large dictionary that lists such multiword expressions in Mexican Spanish along with the probability for the expression to have interpretation different from the literal one.

Eduardo Vázquez-Santacruz et al. from Spain in their paper "La generación de lenguaje natural: análisis del estado actual" ("Natural Language Generation: Revision of the State of the Art"), written in Spanish but supplied with an English abstract, present a detailed tutorial on the state-of-the-art techniques for natural language generation. The task of natural language generation consists in producing fluent, natural, linguistically correct text from some formal description of the facts that are to be expressed in this text. Such formal representation can range from a simple time series (such as a sequence of currency rates in a given period) or a set of parameters (such as weather report) to a logical entity-predicate form describing a complex situation. The task is, then, to describe such data in fluent natural language. An obvious application is symbolic machine translation, where the task consists in "understanding" the text in the source language, i.e., automatically transforming it into a formal logical representation, and then automatically generating a text in the target language from such formal representation. Another important application of natural language generation is automatic generation of a text (such as a user manual for a gadget) in a large number of languages.

In addition, the issue includes four regular papers on topics ranging from artificial intelligence to automatic control.

Alejandro Cervantes-Herrera et al. from Mexico in their paper "Output Regulation and Consensus of a Class of Multi-agent Systems under Switching Communication Topologies" address the problem of reaching a consensus of agents in a multi-agent system when the dynamics of each agent is represented by a switched linear system. The authors provide the corresponding theory and illustrate it with a specific numerical example. Multi-agent systems represent an artificial intelligence technique that combines the advantages of complex logical reasoning and decision-making with the advantages of population-based optimization methods.

David Gómez-Gutiérrez et al. from Mexico and Italy in their paper "Observability and Observer Design for Continuous-Time Perturbed Switched Linear Systems under Unknown Switchings" continue the topic of switched linear systems. The authors derive necessary and sufficient conditions for a new observability notion for perturbed switched linear systems, under less restrictive conditions than those reported in the current state of the art. Their results are meaningful for practical applications.

Fabián López et al. from Mexico and USA in their paper "Hybrid Heuristic for Dynamic Location-Allocation on Micro-Credit Territory Design" addresses the problem of optimal geographical distribution of branch offices of a small financial institution, taking into account numerous considerations such as geographical distribution of potential clients, transportation expenses, logistic overhead, security considerations, etc. The authors suggest a hybrid heuristic approach to the solution of the corresponding large-scale optimization problem, provide an algorithm for its solution, give comprehensive statistical analysis, and extensively illustrate their methodology on computational simulation results.

Gordana Jovanovic Dolecek and Alfonso Fernandez-Vazquez from Mexico in their paper "Sharpening Minimum-Phase Interpolated Finite Impulse Response Filters" present a simple procedure for direct design of low-pass minimum-

phase finite impulse response filters in the context of signal processing, with potential applications in communications, speech processing, predictive coding and other areas. A minimum-phase filter does not contain zeros outside the unit circle, which leads to important properties illustrated in the paper. The authors used MATLAB's implementation of the Remez algorithm in their experiments.

This issue will be useful for all those interested in computational linguistics, natural language processing, human language technologies, and

more generally in artificial intelligence and its numerous applications, as well as computer science in general.

Alexander Gelbukh
Guest Editor

Research Professor and Head,
Natural Language Processing Laboratory, CIC-IPN;
Member, Mexican Academy of Sciences