Editorial

It is my pleasure to present to the readers a new issue of IJCLA. This issue presents papers on four topics: lexical resources and specifically WordNet; grammar, semantics, and dialogue; information extraction; and sentiment analysis and social networks.

Lexical resources are the heart of most natural language processing technologies. Specifically, WordNet has traditionally been the most widely used lexical resource. It groups words with the same meaning together (such groups are called synsets and represent specific lexical meanings that exist in a given language) and separates different senses of the same word into different synsets. In addition, it specifies a wide variety of relationships between such lexical meanings, such as genus–species, part–whole, etc.

R. Amaro *et al.* (Portugal and Spain) report new developments in building a Portuguese WordNet. The original WordNet was built for English; development of WordNet-like dictionaries for other languages is a priority task for the corresponding communities. In addition, such development sheds light on the commonalities between languages and differences that require adjustments in the structure of WordNet. For Portuguese recently a number of rich high-quality lexical resources have been recently developed (a comparative analysis of some of them is given in another paper in this volume), which makes Portuguese an attractive alternative to English for language-independent and multilingual natural language processing experiments. Amaro et al. describe their efforts on increasing the density of relationships represented in Portuguese WordNet.

A. A. Freihat *et al.* (Italy and India) address the phenomenon of specialization polysemy and study it on the material found in WordNet. Specialization polysemy is a phenomenon observed when a word has two senses, one of which can be considered in a certain way more specific than another, that is, included in the other. Detection of this phenomenon is important in natural language processing applications such as machine translation or information retrieval. Freihat *et al.*

describe and classify different situations in which specialization polysemy appears in WordNet.

The next section is devoted to classical problems of natural language processing: grammar, semantics, and dialogue.

T. Seraku (UK) presents a solution to a class of problems related with incremental syntactic parsing. In some languages the syntactic structure of a sentence is defined very late in the analysis process, practically only when the whole sentence has been read by the parser. This poses efficiency and complexity problems to the parsing algorithm. Seraku proposes a solution based on dynamic addressing of the parsing tree nodes in the parser's internal memory. Examples are given for the Japanese language.

Y. Haralambous and **V. Klyuev** (France, Japan) consider the task of semantic analysis of text, which can also be called text understanding. Text understanding is in a way the philosopher's stone, the ultimate goal of natural language processing with which, when it is achieved, all other tasks would be easily solved. In its turn semantic analysis requires deep and broad knowledge about language and about the world and human life. Probably the wider available single source of such knowledge is Wikipedia. Haralambous and Klyuev improve a particular technique of semantic analysis, known as Explicit Semantic Analysis, using knowledge that can be extracted from the structure of Wikipedia.

N. Vadász *et al*. (Hungary) continues the topic of understanding text, addressing the problem of understanding the intentions of the speakers in monologs or dialogs. They present a formal framework for representing people's beliefs, desires, and intentions in a logical form. Their framework allows for analysis of joking, lying, fibbing, bluffing, expressing polarity and opinions, etc. They illustrate their formal ideas with numerous examples.

Z. Wei *et al*. (HK, UK, Quatar, China) show how to identify exploratory dialogs in text. Exploratory dialog is a type of communication between persons with deep understanding of each other's ideas, which implies proactive, positive, and creative participation in the communication. Such dialogs are especially important in learning environments and can occur between students or between the teacher and a student. It is important to detect and encourage this kind of dialog, and discourage non-explorative dialogs in learning and academic environments. Wei *et al*. use machine learning techniques to classify dialogs into explorative and non-explorative. I guess the authors themselves have mastered well the important art of

explorative communication, showing an impressive example of successful academic cooperation between teams from four different countries!

The next paper is devoted to the field of information extraction: identifying specific facts or relations in a given thematic domain expressed in the text.

R. Nawaz *et al*. (UK) present their system for determining descriptions of new biological events in biomedical scientific papers. The amount of published scientific literature nowadays does not allow the researcher to read or even look through all published literature on the topic of their research. Instead, automatic or semi-automatic methods have to be used to locate relevant pieces of information. This problem is especially observed in biomedical literature with its huge and rapidly growing body of published experimental data. Reports about newly observed events are intermixed in the texts with mentions of already known events; however, it is important to identify the novel contents of a scientific paper and the new biological events communicated in this paper. Nawaz *et al*. report more than 99% accuracy of their system in classifying the mentions of bio-events into new and previously known.

Finally, the last three papers are devoted to sentiment analysis, opinion mining, and analysis of the phenomena in the blogosphere and social networks. This is a very hot topic nowadays, with a lot of attention from private companies and governmental bodies drawn to it.

L.A. de Freitas and **R. Vieira** (Brazil) compare a number of lexical resources available for opinion mining and sentiment analysis in Portuguese language. As I have mentioned above, Portuguese natural language analysis community has developed good infrastructure with rich and high-quality lexical resources, which are not only useful for development of accurate applications for this language, but also for testing language-independent or otherwise non-English-oriented methods. Freitas and Vieira give an overview of such lexical resources available for Portuguese in the area of emotion, sentiment, and opinion analysis.

B.-K.H. Vo and N. Collier (Japan) analyze the emotions that Twitter users expressed during the tragic 2011 Great East Japan earthquake and tsunami and its aftermath that included leak of radioactivity from the Fukushima nuclear reactors and uncertainty about possible nation-wide nuclear catastrophe. They argue that

automatic analysis of emotions expressed in social networks in critical situations can help the government to quickly make correct decisions on social help and overall control of the situation. Specifically for the earthquake situations, Vo and Collier present a selection of the corresponding emotions to be tracked and two classification methods for these emotions to be automatically identified in massive Twitter flows.

L. Jia *et al*. (USA) address the task of blog retrieval with an additional requirement: the retrieved blog posts should not only correspond to the user query but also be of a specified facet: opinionated or factual, personal or official, and in-depth or shallow. Obviously, for this the blog posts are to be classified along these dimensions. Jia *et al*. propose the corresponding classifiers and show experimental results that confirm the effectiveness of their proposed methods.

This issue of IJCLA will be useful for researchers, students, and general public interested in various aspects of natural language processing.

GUEST EDITOR:

AJITH ABRAHAM
DIRECTOR,
MACHINE INTELLIGENCE RESEARCH LABS
(MIR LABS), USA
E-MAIL: <AJITH.ABRAHAM@IEEE.ORG>