Editorial

Computational Linguistics and its Applications

This special issue contains several interesting papers related to computational linguistics and its applications. The papers were carefully selected by the guest editors on the basis of peer reviews. We are happy that authors from various countries chose this forum for presenting their work: USA, Spain, Mexico, China, Germany, Hungary, India, Japan, Lithuania; submitting the high quality research results.

The first paper "Recent Advances in Computational Linguistics" by Yulia Ledeneva and Grigori Sidorov (Mexico), presents the view of the authors concerning some aspects of the current state of computational linguistics and its applications.

The paraphrase ability can be considered one of the important characteristics of the usage of natural language that proves the correct understanding of phrases. This interesting phenomenon is discussed in the paper "Paraphrase Identification using Weighted Dependencies and Word Semantics" by Mihai C. Lintean and Vasile Rus (USA). The authors analyze the paraphrase using syntactic and semantic (lexical) information. Evaluation data is presented.

Some important issues of automatic summarization are discussed in the paper "Challenging Issues of Automatic Summarization: Relevance Detection and Quality-based Evaluation" by Elena Lloret and Manuel Palomar (Spain). Namely, two related ideas are presented. First, it is shown that the code quantity principle (most important information) is applicable to automatic summarization. Second, an evaluation of quality of summaries is discussed.

Axel-Cyrille Ngonga Ngomo (Germany) presents in his paper "Low-Bias Extraction of Domain-Specific Concepts" a new approach for extraction of the domain specific terms that tries to be independent from the specific domain. The approach is based on graph clustering.

In the work of Alberto Téllez-Valero, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda and Anselmo Peñas-Padilla (Mexico, Spain) "Towards Multi-Stream Question Answering using Answer Validation" a method of determining the correctness of automatically generated answers is discussed. The method uses the combination of several answering systems.

The paper by Asif Ekbal and Sivaji Bandyopadhyay (India) "Named Entity Recognition using Appropriate Unlabeled Data, Post-processing and Voting" evaluates the improvement of NER by using unlabeled data, post processing and weighted voting of various models.

In the paper "Assigning Library of Congress Classification Codes to Books Based Only on their Titles" by Ricardo Ávila-Argüelles, Hiram Calvo, Alexander Gelbukh, and Salvador Godoy-Calderón (Mexico and Japan), experiments are presented for book classification using very small amount of information, namely, the book title. Several measures of comparison are explored and encouraging results are reported.

An important problem of what is collocation and how it can be detected automatically using modern machine learning techniques is discussed in the paper "Automatic Identification of Lexical Units" by Vidas Daudaravicius (Lithuania).

The paper "Finding Maximal Sequential Patterns in Text Document Collections and Single Documents" whose authors are René García, J. Fco. Martínez-Trinidad, and J. Ariel Carrasco-Ochoa (Mexico), presents two efficient algorithms that allow for detection of all maximal sequential patterns in a text collection without the necessity of recomputing if a new document is added. Experiments that show that the algorithms are better than the state of the art methods are presented.

The paper "Grammar of ReALIS and the Implementation of its Dynamic Interpretation" written by Gábor Alberti and Judit Kleiber (Hungary) presents a grammar for formal discourse analysis and discusses its implementation using Hungarian as an example language.

The idea of using various classifiers simultaneously is explored in "Using Bagging and Boosting Techniques for Improving Coreference Resolution" by Smita Vemulapalli, Xiaoqiang Luo, John F. Pitrelli, and Imed Zitouni (USA). The area of application of these classifiers is coreference resolution. It is shown that the technique that uses a combination of classifiers outperforms single classifiers.

An interesting topic of temporal multi-document summarization, whose goal is to analyze the information according to its temporal perspective in various documents, is discussed in "Cascaded Regression Analysis based Temporal Multi-document Summarization" by Ruifang He, Bing Qin, Ting Liu, and Sheng Li (China). Macro and micro importance discriminative models are combined to form a cascaded regression analysis approach that is verified using several available dataset.

Yulia Ledeneva
Autonomous University of the State of Mexico
Santiago Tianguistenco, Mexico
yledeneva@yahoo.com

Grigori Sidorov
Center for Computing Research, National Polytechnic Institute
Mexico City, Mexico
www.cic.ipn.mx/~sidorov