Editorial

It is my great pleasure to present to the readers of the journal the current trends in Computational Linguistics and Artificial Intelligence. This issue contains papers related to philosophic issues of Artificial Intelligence, natural language semantics, methods and methodologies of calculation of similarity between objects in vector spaces, various applications of computational linguistics such as machine translation, computational morphology and syntax, automatic text summarization, named entity recognition, paraphrase detection, as well as music recommendation systems.

John F. Sowa from USA in his paper "Why Has Artificial Intelligence Failed? And How Can it Succeed?" analyzes the current situation of Artificial Intelligence and proposes a general direction for future research. He explains that the expectations of Artificial Intelligence were very high, but they never came true because each of the applied methods is too narrow and specific. The author considers that the principal future direction of Artificial Intelligence is the intelligent combination of the exiting Artificial Intelligence methods.

Marie Duží from Czech Republic in her paper "Structural Isomorphism of Meaning and Synonymy" deals with the phenomenon of synonymy. She considers the description of meaning in frame of the Transparent Intensional Logic theory, where the sense of an expression is represented in algorithmic form. The author introduces the notion of structural isomorphism of such algorithmic description: if two expressions have structurally isomorphic algorithmic representations, then they are synonymous. The proposed solution is thoroughly elaborated and justified from the philosophical point of view.

Lionel Ramadier *et al*. from France in their paper "Inferring Relations and Annotations in Semantic Network: Application to Radiology" present a new semantic lexical resource for radiology applications in French, which they developed based on an existing general semantic lexical network JeuxDeMots. The resource includes weighted information such as frequencies, which unfortunately is not often found in general ontologies. The authored explain in detail the advantages of this new resource in medical applications.

Huayi Li *et al*. from USA and China in their paper "Spotting Fake Reviews using Positive-Unlabeled Learning" address a very important problem of decided whether a review of a product or service. As the applications of text mining and sentiment analysis begin crucially affect the way people and companies make their economic decisions, the problem of false and fake information in user-generated content become more and more painful. The authors suggest a technique for detecting such fake information.

Zvi Ben-Ami *et al*. from Israel and USA in their paper "Using Multi-View Learning to Improve Detection of Investor Sentiments on Twitter" continue the topic of sentiment analysis. A common problem in sentiment analysis is that affective expressions are related with something not easy to measure objectively: human emotions. In this paper the authors show how to learn affective expressions by comparing them with known objective facts, thanks to their observation that in a particular domain, namely, financial one, the objective reality related to Twitter messages is often known.

Grigori Sidorov *et al*. from Mexico in their paper "Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model" show how similarity between features can be accounted for in the widely used Vector Space Model used, in particular, in machine learning. They propose a novel concept of similarity, which they call *soft similarity*. In the area of the Natural Language Processing, where feature are words or n-grams or syntactic n-grams, they calculate the

ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2049

similarity of features using the Levenshtein distance between them; also WordNet similarity can be used, etc. In particular, they generalize the cosine measure, suggesting what they call *soft cosine measure*. The soft cosine measure outperforms the best scores in the CLEF 2014 entrance exams competition.

Iria da Cunha *et al*. from Spain, France, Canada, and Mexico in their paper "SIMTEX: An Approach for Detecting and Measuring Textual Similarity based on Discourse and Semantics" presents a way of measuring similarity between texts using deep linguistic information—not only semantic but even discursive. They use the Rhetorical Structure Theory (RST) for interpretation of discourse and lexical-semantic relations included in EuroWordNet. While they test their algorithm, called SIMTEX, on Spanish texts, the presented methodology is language-independent.

Hiram Calvo *et al*. from Mexico in their paper "Dependency vs. Constituent Based Syntactic N-Grams in Text Similarity Measures for Paraphrase Recognition" address the important problem of paraphrase recognition. The authors apply the recently introduce notion of syntactic n-grams as features. They use a special type of n-grams: continuous syntactic n-grams, i.e., the n-grams that are obtained from paths in syntactic trees without bifurcations; such n-grams do not have elements of the same level in the parse tree. They analyze various types of syntactic representations: dependency vs. constituent. They report experiments using various sets of features. Surprisingly, in their experiments slightly better results are obtained using constituent based syntactic n-grams.

Zuzana Nevěřilová from Czech Republic in her paper "Paraphrase and Textual Entailment Generation in Czech" presents a system capable of generation of paraphrases (expressions bearing the same meaning but with different wording) and logical implications (textual entailment) for the given sentence. She developed an annotation game to crowd-source human annotations on the correctness of paraphrases and entailments generated by her system. The system has been developed for the Czech language, but the same methodology can be generalized to other languages.

Rabeb Mbarek *et al*. from Tunisia and Saudi Arabia in their paper "Vector Space Basis Change in Information Retrieval" apply an idea similar to that presented by Sidorov et al. (this issue) to the task of information retrieval. They show that a particular transformation of the basis of the vector space model improves the efficiency of relevance feedback in information retrieval. They incorporate their vector space basis change strategy to the Rocchio algorithm and show that the algorithm benefits from this change.

Marina Litvak and **Natalia Vanetik** from Israel in their paper "Multi-document Summarization using Tensor Decomposition" use tensors instead of traditional vectors to represent natural language texts. They apply their representation to the task of text summarization: compiling a short text that preserves as much of important information from the original longer document as possible. The authors use the Tensor Decomposition technique to rank the topics of the document by their importance, and then extract the sentences related to the most important topics of the document.

Utpal Kumar Sikdar *et al*. from India in their paper "Entity Extraction in Biochemical Text using Multiobjective Optimization" address the problem of extracting entity names from chemical and biomedical texts. This is a very important problem given the enormous, and fast growing, body of published medical and biological information. The authors propose a two-stage technique for entity extraction. First, a state-of-the-art multi-objective classifier determines a set of relevant features. Then a subset of the solutions found with these features is used to construct an ensemble of classifiers. The authors report a high recall and precision in their experiments.

Jordi Centelles *et al*. from Spain, Singapore, and Mexico in their paper "On-line and Off-line Chinese-Portuguese translation service for Mobile Applications" present the architecture of a machine translation system that takes advantage of Internet connectivity but is capable of working in offline mode, too. This is an important technological problem that combines the advantages of cloud-based or server-based resource-intensive machine translation technology with the real-life challenge of limited or insufficient connectivity of mobile devices. The system is implemented in Android and uses a set of state-of-the-art technologies such as optical character recognition and speech recognition.

Hammo Bassam *et al*. from Jordan in their paper "Formal Description of Arabic Syntactic Structure in the Framework of the Government and Binding Theory" give a detailed formal analysis of the syntax of Arabic language in frame of a solid linguistic theory, Government and Binding. Deep understanding of syntactic rules of a given language is important for the development of high-quality parsers for this language and for general advance of parsing technology. Formal analysis of a language is especially valuable in this context because it allows for direct software implementation of the rules that describe syntactic structure of sentences.

Banu Yergesh *et al*. from Kazakhstan in their paper "Semantic Hyper-graph Based Representation of Nouns in the Kazakh Language" present an approach to morphological analysis and generation of the Kazakh language based on construction of a semantic hyper-graph. The vertices of such a hyper-graph represent morphological features of words, such as grammatical gender, number, case, animacy, etc., and the edges represent relationships between these features: the particular order in which these features should be expressed in words, especially when one deals with agglutinative languages (cf. English "ungrammaticalitynessless" but not "grammaticunlessalnessity", which is composed of the same morphs). A very large dictionary of Kazakh noun word-forms generated with this approach is available for download for research purposes.

Félix Castro Espinoza *et al*. from Mexico in their paper "Towards the Automatic Recommendation of Musical Parameters based on Algorithm for Extraction of Linguistic Rules" apply linguistic rules to the analysis of data on people's emotional responses to automatically generated music in the context of computer-assisted creativity. The rules allow for automatic or computer-assisted generation of musical pieces that invoke a given combination of emotions. The authors give an introduction into fractal-based music generation.

This issue will be useful for all those interested in Computational Linguistics, Natural Language Processing, and Human Language Technologies and more generally in Artificial Intelligence and its numerous applications.

Alexander Gelbukh
Guest Editor

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

ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2049