International Journal of Computational Linguistics and Applications, Vol. 6, No. 1, 2015, pp. 7–9
ISSN 0976-0962, http://ijcla.bahripublications.com

Editorial

This issue of IJCLA presents papers on social linguistics, stylometry, analysis or literary texts, sentiment analysis, text summarization, automatic essay scoring, automatic speech recognition, word sense disambiguation, semantic text similarity, and text representation.

R. Cotterill et al. (UK) present a corpus of dialogs in which participants have perceived difference in social power. They show that such dialogs can be automatically classified, with above-chance precision, by the relationship between the participants of a particular dialog, which demonstrates that the corpus presents important features that reflect such relationship. The corpus will be very useful for sociolinguistic research.

E. Davoodi & **L. Kosseim** (Canada) investigate the relationship between complexity of text, i.e., its readability level, and its discourse-level properties. On the material of Simple English Wikipedia, they show that simple text contains the same discourse relations as normal, or complex, text; however, the lexical choices for the discourse markers are affected by the desirable readability level of the text.

M.-A. Boukhaled et al. (France) propose an objective interestingness measure that allows extracting meaningful syntactic patterns for computational stylistic research without any prior knowledge. They apply their approach to classic French literature texts. An important property of their measure is that it can be applied to both long and short texts; this allows its application to literature works that do not belong to any larger collection of texts.

C. Martínez et al. (Chile) present a study of emotional charge of the texts of Chilean school textbooks from first to eighth year. They automatically determine the degree of expression of the six basic emotions (anger, sadness, fear, disgust, surprise, happiness) in the text; the performance of their

automatic procedure is evaluated by human experts. They show that happiness is by far a predominant emotion expressed in the analyzed textbooks, followed by sadness and fear. They observed that each emotion is expressed with approximately the same degree in the texts of different genres, except that anger was not observed in songs.

H. Zidoum et al. (Oman) describe the use of lexical cohesion measured with help of lexical chains for extractive text summarization of Arabic documents. Arabic language is underrepresented in computational linguistics literature in general and in the literature on text summarization in particular. The authors give a detailed step-by-step account of their algorithm and compare the obtained results with human judgements.

D. Aguirre et al. (USA) present a method for automatic evaluation of text summaries written by elementary school students, with the aim of facilitating the work of schoolteachers and improve the feedback time. They show that the use of semantic similarity measures and pre-processing such as spelling correction improve the precision of their automatic grader. Their method achieves 98% of precision at 9-point grading scale, which is comparable with the agreement between human graders.

T. Nadungodage et al. (Sri Lanka and UK) show how to reduce the number of samples necessary for speaker adaptation in automatic speech recognition. The method allows building general speaker adaptation models, which outperform speaker-independent models based on the same amount of training data. The task is important in the situation when the available training corpora are too small for training the recognition system in a traditional way. The authors apply their method to Sinhala, a low-resource language spoken by the majority of population of Sri Lanka.

M. A. Sobrevilla Cabezudo & **T. A. Salgueiro Pardo** (Brazil) address the problem of word sense disambiguation for verbs in Brazilian Portuguese. Verbs are more difficult to disambiguate than nouns, and, while well studied for English, word sense disambiguation in Portuguese has received relatively little attention in literature. The authors use the Portuguese

WordNet, WordNet-Pr, as the sense inventory. Their experiments show poor performance of existing methods, which implies that the task needs more attention from the research community.

D. Cantone et al. (Italy) propose a corpus-based statistical measure of closeness between two sets of words. The measure is based on co-occurrence statistics of the words from the two sets, calculated over a large enough text corpus. The problem is important in a great number of text processing and computational linguistics-related tasks where semantic text similarity is used, ranging from information retrieval to plagiarism detection. The authors present computationally efficient algorithms for computing the proposed closeness measure.

M. Mouriño-García *et al*. (Spain) introduce the bag-of-concepts representation scheme for text processing and show its advantages of the traditional bag-of-words representation scheme. In particular, the use of concepts instead of words alleviates the problems related to synonymy and polysemy of words. Their experiments confirm this intuition; however, the results heavily depend on the quality of concept extraction method used to build the bag-of-concepts representation of the documents.

This issue of IJCLA will be useful for researchers, students, software engineers, and general public interested in natural language processing and its applications.

ALEXANDER GELBUKH
EDITOR IN CHIEF