

Special Issue on Web Intelligence, Mining and Semantics

Preface

This special section consists of substantially extended and revised versions of the papers selected from the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS), held during June 25-27 2018, in Novi Sad, Serbia. Five best papers closely related to the Journals scope were initially selected from the 44 accepted papers, from 31 countries, which were presented at the conference. Three of them were finally included in this special section, after passing through a rigorous reviewing procedure. WIMS is a series of peer-reviewed International Computer Science conferences. It is a forum for researchers and practitioners to present their state-of-the-art results in building Intelligent Web, to examine performance characteristics of various approaches in Web-based intelligent information technology, and to cross-fertilize their ideas on the development of Web-based intelligent information management solutions across different domains. Accepted papers bring to readers interesting research results and we hope that they will enjoy reading them and find presented achievements inspiring and useful.

The first paper Improvement of automated essay grading by grouping similar graders by Kaja Zupanc and Zoran Bosnic, is generally in the area of clustering text documents. In the paper, the authors concentrated on automated essay evaluation as practical solution for replacing time consuming manual grading of student essays. Automated systems are used in combination with human graders in different high-stake assessments, where grading models are learned on essays datasets scored by different graders. As human graders can unintentionally introduce subjective bias into scores, a grading model has to learn from data that represents a noisy relationship between essay attributes and its grade. Authors offered an approach for partitioning a set of essays into subsets that represent similar graders, which uses an explanation methodology and clustering. The obtained results confirm the assumption that learning from the ensemble of separated models can significantly improve the average prediction accuracy on artificial and real-world datasets.

The second paper Improving Short Text Classification using Information from DBpedia Ontology by Jernej Flisar, Vili Podgorelec, is generally in the area of text classification. The authors detected that a huge amount of short textual documents are generated on a daily basis, for which effective tools for organization and classification are needed. They proposed a new approach to identify relevant concepts in short text documents with the use of the DBpedia Spotlight framework. Additionally, they enriched the text with information derived from DBpedia ontology, which reduces the sparseness. Six variants of text enrichment methods have been developed and tested on four short text datasets using

seven classification algorithms. The obtained results show that the proposed text enrichment approach significantly improves classification of short texts. Also, it is robust with respect to different input sources, domains, and sizes of available training data. Finally, it has been confirmed that the proposed methods are beneficial for classification of short text documents, especially when only a small amount of documents is available for training.

The last paper Time and Space Efficient Large Scale Link Discovery using String Similarities by Andreas Karampelas and George A. Vouros, is in area of Link Discovery and Named Entity Linking. In the paper, the authors proposed and evaluated time and space efficient methods for matching entities in large data sets based on effectively pruning the candidate pairs to be matched. For this purpose, they used edit distance as a string similarity metric. Three filtering methods have been proposed and compared, which build on a basic blocking technique to organize the target data set, facilitating efficient pruning of dissimilar pairs. The first method clusters entities and exploits the triangle inequality using the string similarity metric. The second method uses only the substring matching rule, whereas the third method uses the substring matching rule in conjunction to the character frequency matching filtering rule. Obtained results showed the pruning power of the different filtering methods used, also in comparison to the string matching functionality provided in the state of the art frameworks LIMES and SILK.

Acknowledgement

We are grateful to the Editor-in-Chief of Fundamenta Informaticae Journal, Prof. Damian Niwiski, for giving us the opportunity to guest edit this special section and for his support and guidance. We would also like to thank the reviewers for helping authors to improve their papers, through their timely and valuable feedback.

Special issue editors:

Costin Badica
University of Craiova, Romania
costin.badica@ucv.ro

Mirjana Ivanovic
University of Novi Sad, Serbia
mira@dmi.uns.ac.rs

Yannis Manolopoulos
Open University of Cyprus, Cyprus
yannis.manolopoulos@ouc.ac.cy

Riccardo Rosati
Sapienza Università di Roma, Italy
rosati.riccardo@uniroma2.it

Paolo Torroni
University of Bologna, Italy
p.torroni@unibo.it