
Editorial

Krzysztof Jassem*

Faculty of Mathematics and Computer Science,
Adam Mickiewicz University,
Ul. Umultowska 87, 61-614 Poznań, Poland
E-mail: jassem@amu.edu.pl

Maciej Piasecki

Wroclaw University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wroclaw, Poland
E-mail: maciej.piasecki@pwr.wroc.pl

Biographical notes: Krzysztof Jassem is a Professor at the Faculty of Mathematics and Computer Science, Adam Mickiewicz University in Poznań. His main field of interest is machine translation. Since 1998, he has been the leader of a research team working on a system translating from and into Polish. The system is currently published as a commercial product, Translatica. Since 2004, he is the CEO of an IT company Poleng. Poleng is a member of the PWN Group, a leading publisher of scientific books, manuals and electronic media in Poland. His 'professional' hobby is bridge. He won the silver medal in the European Championship in 2009.

Maciej Piasecki is an Assistant Professor at the Institute of Informatics, Wrocław University of Technology, where he has been working since 1994. He is the Leader of plWordNet project – the Polish WordNet – and G4.19 Language Technology Research Group. He received his PhD and MSc in Computer Science from Wrocław University of Technology. He worked on formal semantics of the natural language, machine translation and human-computer interaction. His current research areas include extraction of linguistic knowledge from large text corpora, lexical resources, information extraction, text mining and human language technology with especial emphasis to the basic language technology for Polish.

The problem of data mining concerns both structured and unstructured data. Structured data may be contained in fixed fields (e.g., in spreadsheets), organised as well-defined elements of a structure (e.g., in relational databases) or tagged with special marks (e.g., in .xml documents). Unstructured data, either textual (e.g., documents, presentations, SMS messages, e-mails) or non-textual (e.g., graphical, audio or video files) miss such organisation.

This issue of *IJDMMM* is devoted to the problem of data extraction from unstructured textual resources. The simplest methods for mining texts do not require any specific linguistic knowledge: indexing algorithms allow for storing and searching huge amounts of text data. These methods, however, are not aimed at 'understanding the text', thus significantly limiting the possibilities for information extraction.

Computational linguistics (CL) is the domain of science that attempts to take a deeper look into the meaning of a text from the computer point of view. CL methods make it possible to extract structured relations from unstructured data.

Computational linguistics-application (CL-A) is a new forum for the dissemination of research in CL. It started in 2007 as a track of AAIA Conference¹. Since then it has evolved into a workshop (in 2008) and finally into an independent conference (in 2010). In 2010, each paper submitted to CL-A was evaluated by four reviewers and rated on a scale from 1 to 5. The authors of papers that concerned data mining and got average grade above 4.7 were requested to publish extended versions of their papers in this special issue of *IJDMMM*.

Waszczuk et al. describe the tool for text annotation (which is often a preliminary step for creating text mining algorithms) and present the results of a classifying algorithm trained on thus annotated data. Maciołek and Dobrowolski present two novel algorithms for automatic document classification. Kisilevich et al. suggest a new approach to opinion and sentiment analysis of photo comments. Seretan and Wehrli present a tool for syntactic concordancing and multi-word expression detecting. The tool is intended to help lexicographers, linguists and translators in their everyday's work.

Although the papers contained in this issue concern different aspects of CL, they are all aimed at the same target: finding the proverbial needle in the haystack of unstructured texts.

Notes

- 1 AAIA – Advances in Artificial Intelligence and Applications Conference was first held in 2006 within the International Multiconference on Computer Science and Information Technology in Wisla, Poland.