



**University of Vienna**  
**Faculty of Computer Science**  
**Research Group for Scientific Computing**  
**Univ.-Prof. Dipl.-Ing. Dr. Peter Brezany**

Währinger Straße 29  
A-1090 Vienna  
Austria

Phone: +43 664 539 74 22  
Fax: +43 1 4277 878402  
E-mail: [peter.brezany@univie.ac.at](mailto:peter.brezany@univie.ac.at)

**Fakulta elektrotechniky a komunikačních technologií VUT v Brně**  
**Vědecké oddělení**  
**Technická 10, 616 00 Brno**  
**Czech Republic**

Vienna, November 22, 2018

**Review of the Dissertation Thesis *Dolovanie znalostí z textových dát použitím metód umelej inteligencie* (Text Mining Based on Artificial Intelligence Methods) submitted by Ing. Lukáš Povoda**

### **Thesis Assessment**

This doctoral thesis addresses selected challenges associated with modern text data mining applications that include security applications, biomedical applications, business and marketing applications, sentiment analysis, scientific literature mining, digital humanities and computational sociology, online media applications, etc. One of the main objectives of text mining processes is to derive high-quality information from text (this typically refers to some combination of precision, relevance, novelty and interesting). Besides optimizing traditional methods, like statistical pattern learning, investigation of novel methods based on the state-of-the-art in deep learning and considering improvement in hardware computational and storage power must be considered.

The author of this thesis, Ing. Lukas Povoda conducted research in the above outlined context and the subjects that he addressed are **extremely relevant to the need of scientific community and aligned with the current world-wide top-level research.**

The thesis text was nicely formatted with the Latex system, with carefully elaborated graphics and is very well structured and readable.

The introductory part briefly outlines the motivation and context for this research and characterizes the contents of the 5 involved chapters.

Chapter 1 briefly lists the topics of current text analysis and mining research and provides an overview of the state of the art. Besides traditional methods, the focus is on methods enabled by the advent of deep learning techniques—the underlying theory is explained on image data.

Chapter 2 Characterizes the main goal of the thesis: propose and validate a method for machine understanding non-structured text data – it should be sufficiently universal (for text data having any features; no specific precautions concerning language and its grammar are needed), reusable for various languages, and provide sufficiently precise results. Seven subgoals driving the research conducted in the thesis were specified. **These objectives are ambitious.**

Chapters 3 and 4 represent the kernel parts of the thesis and its key contributions.

Chapter 3 presents the approach and solution of the specified objectives. New methods are investigated and developed in three subchapters: the method for classification of emotions based on the SVM classifier, in 3.1; extensions considering a “Big Data” scenario (an input database with more than 2.4 million text patterns was created, in 3.2; and a solution based on deep learning partially modelling cognitive behaviour of a person reading a text – seven different text databases were used for training and testing during the method development and the proposal of the structure of the deep neural network realizing the method; here the key focus was on the Recurring Kernel.

Chapter 4 deals with evaluation of the proposed and realized methods and comparison of their performance (precision) with performance of most successful state-of-the-art methods. Further, the hardware infrastructure used for experiments is also briefly discussed. This analysis was conducted in a very systematic way; relevant performance improvements are reported. A reduced computer memory demand was also achieved. The proposed methods were experimentally verified by classification of text data expressed in five different languages.

Chapter 5 summarizes the research work done and provides discussion about the results achieved. It does not try to outline the future direction (or, some vision) of the research conducted in the thesis.

### **Publication activity and scientific competence of the author**

The author has an impressive publication record (2 publications in journals with impact factors – one of them is in a review process, 3 in journals without impact factor, 10 in conference proceedings, 3 software products available for downloads).

To sum up, it is possible to state that the Ph.D. candidate conducted excellent scientific and technical work during his Ph. D. study and brought original contributions to the artificial intelligence domain and data science.

### **Questions to the Ph.D. candidate**

1. Did you consider applying the proposed methods to text stream processing (result precision vs. processing speed and memory limitation)?
2. Do you consider working on some form of research results transfer to real applications, like those mentioned in the 1<sup>st</sup> paragraph of this review (important for driving further research), e.g. to upgrade them to a prototype or a full-fledged framework available for distribution?
3. Do you see any rationales to port the results to a concrete tele-medicine sphere?
4. How did you manage experimenting with text data expressed in Chinese language?

### **Conclusion**

The author of the thesis proved to have an excellent ability to perform research and to achieve relevant scientific results. I do recommend the thesis for presentation at the defence with the aim of receiving a Ph.D. degree. Due to my opinion, it fulfils recognized requirements for awarding of doctoral degree.



Peter Brezany