"Towards Semantic Measures of Information Content - Integrating LinkGraph with Seminstructured Textual Information"

Mieczysław Kłopotek
Institute of Computer Science, Warsaw, mieczyslaw.klopotek@ipipan.waw.pl

The aim of the project is to carry out research in the following areas:

- learning to rank –learning proper document ranking rules based on human computer interaction;
- sentiment analysis;
- extraction of concept hierarchies from document collection (hypernyms, hiponyms, synonyms etc.) and their application in fact mining and factsearch in search engines;
- spam detection –documents, websites etc.

Within each of these areas the goal is to:

- evaluate state of the art models;
- model improvement;
- creation of resources for the tasks;
- creation of methods and resources for model evaluation;•integration of selected models/methods with semantic search engine;
- search for synergic effects between these various areas and their exploration for (1) improvement of search engine performance(applied research) (2) better understanding of the nature of human approach to information value (fundamental research).

Understanding the value of information from the human point of iew is one of the most important concepts of the modern IT dealing with the processing of information on the Internet. It has not only significant from theoretical but also from an engineering point of view, since the present interpretation of information content, based on Shennon entropy, seems to be completely inadequate. In particular, the pressing issue is to provide information that corresponds to human understanding and allow for an automatic assessment of information in a manner that satisfies a human person.
The elements of such assessment include, among others, research in such areas as: above-mentioned learning to rank, sentiment analysis, concept hierarchy extraction and spam detection.

Recommended readings:

[1] Zhou, Lina, 2007: „Ontology learning: state of the art and open issues”, Information Technology and Management”, 2007/09 pp. 241- 252 UR - https://doi.org/10.1007/s10799-007-0019-5DO - 10.1007/s10799-007-0019-5

[2] Kaity, Mohammed and Balakrishnan, Vimala, 2019: "An automatic non-English sentiment lexicon builder using unannotated corpus", The Journal of Supercomputing. 2019/75, pp. 2243—2268

[3] Roffo, Giorgio, 2017: “Ranking to Learn and Learning to Rank: On the Role of Ranking in Pattern Recognition Applications.”. arXiv http://arxiv.org/abs/1706.05933