SEARCHING AND SUMMARIZING IN A MULTILINGUAL ENVIRONMENT

MICHAL TOMAN[1]; JOSEF STEINBERGER[1]; KAREL JEZEK[1]

[1] Faculty of Applied Sciences, University of West Bohemia,
Univerzitni 22, Plzen, Czech Republic
e-mail: mtoman@kiv.zcu.cz; jstein@kiv.zcu.cz; jezek_ka@kiv.zcu.cz

Multilingual aspects have been gaining more and more attention in recent years. This trend has been accentuated by the global integration of European states and the vanishing cultural and social boundaries. The ever increasing use of foreign languages is due to the information boom caused by the emergence of easy internet access. Multilingual text processing has become an important field bringing a lot of new and interesting problems. Their possible solutions are proposed in this paper. Its first part is devoted to methods for multilingual searching, the second part deals with the summarization of retrieved texts. We tested several novel processing techniques: a language-independent storage format, semantic-based indexing, query expansion or text summarization leading to faster and easier retrieval and understanding of documents. We implemented a prototype system named MUSE (Multilingual Searching and Extraction) and compared its qualities with the state-of-the-art search engine – Google. The results seem to be promising; MUSE shows high correlation with the market-leading products. Although for our experiments we used Czech and English articles, the main principle applies to other languages as well.

Keywords: multilingual text processing; searching; summarization; EuroWordNet