

Search term recommendation and non-textual ranking evaluated

Philipp Mayr, Peter Mutschke, Philipp Schaer, York Sure
GESIS – Leibniz Institute for the Social Sciences, Bonn, Germany
philipp.mayr@gesis.org

In typical metadata-driven Digital Libraries three major difficulties arise: (1) the vagueness between search and indexing terms, (2) the information overload from different information retrieval (IR) systems, and (3) the problem that pure term frequency based rankings, such as term frequency – inverse document frequency (tf-idf), provide results that often do not meet user needs (Mayr et al. 2008).

In our paper three science model driven retrieval services are presented and evaluated: query expansion based on co-word analysis (our search term recommendation service), re-ranking via Bradfordizing and author centrality. The most important question is how effective and helpful the services are in an actual search. In a search portal like sowiport.de¹ which provides unified access to a variety of databases providing high quality content such as bibliographic metadata, full texts and data sets, the question becomes crucial whether and how retrieval services can improve search.

The services in our project are evaluated with relevance assessments from which two important implications emerge: (1) precision values of the retrieval service are the same or better than the tf-idf retrieval baseline and (2) each service retrieved a disjoint set of documents (compare Mayr et al., submitted).

In this paper we want to show pros and cons of the proposed retrieval services²: Do central authors (core journals) provide more relevant hits? Do highly associated co-words have any positive effects? Do the services provide other relevant hits than typical approaches?

All services are implemented in our prototype. See <http://www.gesis.org/beta/prototypen/irm/>

References

Mayr, Philipp; Mutschke, Peter; Schaer, Philipp; Sure, York (submitted): Science Models as Value-Added Services for Scholarly Information Systems.

<http://www.ib.hu-berlin.de/~mayr/arbeiten/sciencemodels-2010.pdf>

Mayr, Philipp; Mutschke, Peter; Petras, Vivien (2008): Reducing semantic complexity in distributed digital libraries: Treatment of term vagueness and document re-ranking. In: Library Review 57, No. 3, pp. 213-224. URL: http://www.ib.hu-berlin.de/~mayr/arbeiten/mayr-et-al_LR08.pdf

¹ <http://sowiport.de/>

² Our current research on value-added service for digital libraries is funded by DFG, grant no. INST 658/6-1. <http://www.gesis.org/forschung-lehre/programme-projekte/informationwissenschaften/projektuebersicht/irm/>