

Creation of custom KOS-based recommendation systems

Thomas Lüke¹, Wilko van Hoek, Philipp Schaer, Philipp Mayr

Abstract

With the growing size and complexity of modern digital libraries (DL) it has become difficult to express an information need in precise domain-specific terminology. While the problem used to be to describe the search with the right words, large current DL like Europeana deliver results for almost any search term. The challenge today lies within narrowing down a query with the intention to lead the user to the results which are most relevant to his information need. Search term recommendation systems (STR) have been proven to help users fulfill this process of query formulation and refinement. These systems try to find terms that are related to the original query and are closely linked to the semantic backbone of the DL. Therefore terms from knowledge organization systems (KOS) like thesauri or classifications are used to build STRs [1]. Through the enrichment of a query with such terms and concepts a search becomes more specific and delivers better results in terms of document relevance.

These effects of STR-based query expansion have been proven in real life search [2] as well as in a typical information retrieval (IR) setting [3] using a predefined set of topics and corresponding result sets. While the real life study showed the benefit of search term recommendation for users of a digital library the second type proved that expansion of a query in an automatic setup leads to better precision and recall values.

One aspect in our recent research is the discussion of the effectiveness of general vs. specific recommenders. In our study we can show that recommendations based on sub-discipline specific vocabulary (e.g. a specific discipline in the social sciences) is better suited to improve retrieval results than query expansion with general and broader terms from a large domain like social sciences [3]. Based on these findings this paper outlines the technical infrastructure IRSA² which allows researchers to create their own specific STRs. The infrastructure can be accessed through a web service. Inputs to create these STRs could be individual repositories or custom sub-samples of a repository which are indexed with a KOS. The approach supports well known XML or OAI formats [4]. It is also possible to set up a copy of the infrastructure within the researchers working environment.

We will give a short introduction how the system works and what steps researchers need to take to create their own custom STRs. The technical requirements that have to be met for a successful creation of custom STRs will also be addressed. Beside this introduction to the web-service hosted on our systems a demonstration of an installation of the system to a fresh environment is also possible.

¹ Corresponding author: thomas.lueke@gesis.org

² <http://sourceforge.net/projects/irsa/>

References

- [1] P., Schaer, P., Sure, Y. (2009). User Interface Design for Search-Term Recommendation and Interactive Query Expansion Services. 8th European NKOS Workshop 2009
- [2] Mutschke, P., Mayr, P., Schaer, P., Sure, Y. (2011). Science models as value-added services for scholarly information systems. *Scientometrics*. 89, 349–364.
- [3] Lüke, T., Schaer, P., & Mayr, P. (2012 to appear). Improving Retrieval Results with discipline-specific Query Expansion. *Proceedings of Theory and Practice of Digital Libraries 2012*
- [4] Schaer, P. et al. (2012) Building Custom Term Suggestion Web Services with OAI-Harvested Open Data. *Proceedings of the 2. DGI-Conference 2012*