

Towards Evaluating the Impact of Ontologies on the Quality of a Digital Library Alerting System

Alfons Huhn, Peter Höfner, and Werner Kießling

Institut für Informatik, Universität Augsburg,
86159 Augsburg, Germany
{huhn, hoefner, kiessling}@informatik.uni-augsburg.de

Abstract. Advanced personalization techniques are required to cope with novel challenges posed by attribute-rich digital libraries. At the heart of our deeply personalized alerting system is one extensible preference model that serves all purposes in conjunction with our search technology Preference XPath and XML-based semantic annotations of digital library objects. In this paper we focus on the impact of automatic query expansion by ontologies. First results indicate that use of ontologies improves the quality of the result set and generates further results of higher quality.

1 Introduction

P-News [4] is an experimental system that alerts users, when user-relevant documents newly arrive at the digital library. Its main feature consists in a deep personalization which is achieved by a highly flexible preference methodology [3] with powerful query capabilities. This methodology enables a consistent description framework for the users' wishes, the defaults assigned to user groups, and the domains of discourse modeled by ontologies. We evaluate the impact of query expansion by ontologies on the quality of the result set by analyzing the set of best matching objects (BMO-set) [3], when ontologies are used or are ignored. The full version of this poster is available at [1].

2 Impact of Ontologies on the Quality of the Result

In ontologies any agent easily finds *synonyms*, *hyponyms*, *hypernyms* as well as other complex semantic contexts. Depending on the involved attributes of the query, P-News decides which ontologies have to be used for the query expansion. This attribute-aware procedure sharpens the focus by just applying those ontologies which semantically model the domain of the attribute. Not surprisingly an increase of quality arises only in those test cases, where we did not have perfect matches and where the user's terms are also be found in the involved ontology. As stated in [2] recall for the ontology-based query expansion tops recall for keyword-based techniques. To reduce loss of focus, we exploit the semantic context for sense disambiguation.

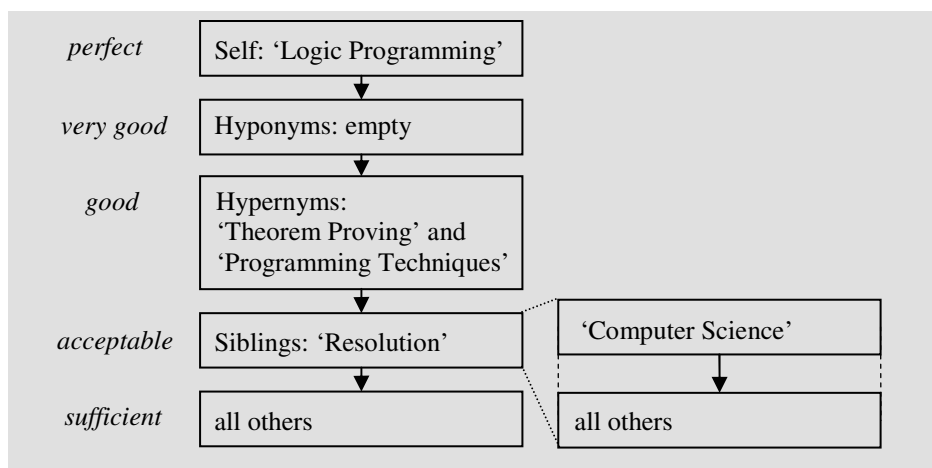


Fig. 1. Quality refinement reducing a loss of focus

Let's look at Fig.1: Suppose that a user's query for 'Logic programming' (*self*) achieves no perfect hit and no hyponyms exist. Assuming that a *sibling* (e.g. 'Resolution') has multiple hypernyms (e.g. 'Programming Techniques', 'Screen'), loss of focus may arise. To cope with this challenge, we exploit the context. We identify the common hypernyms of self and this sibling. All the hypernyms of the common hypernyms (e.g. 'Computer Science') are used as the POS-set of a POS preference [3] applied to the BMO-set of the sibling, i.e., they are preferred in comparison with all others (e.g. especially 'Screen'). The common hypernyms (e.g. 'Theorem Proving') must not be considered, because they already have a better quality (see Fig. 1).

With our case studies performed so far we could find strong evidence for the following conjectures: Use of ontologies can improve the quality of the BMO query result, and it can be controlled to reduce a loss of focus during query expansion.

Acknowledgement. P-News is funded within the German Research Foundation's strategic research initiative 'Distributed Processing and Delivery of Digital Documents (V^3D^2)'.

References

1. A. Huhn, P. Höfner, W. Kießling: Towards Evaluating the Impact of Ontologies on the Quality of a Digital Library Alerting System. Technical Report 2005-07, Institute of Computer Science, University of Augsburg
2. L. Khan, D. McLeod, E. Hovy: Retrieval effectiveness of an ontology-based model for information selection. In *The VDLB Journal* (2004) 13-1, 71-85, Springer New York
3. W. Kießling: Foundations of Preferences in Database Systems. In *Proc. Int. Conf. on Very Large Databases (VLDB 2002)*, Hong Kong, China, 311-322
4. Q. Wang, W.-T. Balke, W. Kießling, A. Huhn: P-News: Deeply Personalized News Dissemination for MPEG-7 based Digital Libraries. In R. Heevy, L. Lyon (eds.): *Research and Advanced Technology for Digital Libraries (ECDL2004)*, Bath, UK, 256-268