Fine-Grained Relevance Feedback for XML Retrieval

Hanglin Pan Ralf Schenkel Gerhard Weikum
Max-Planck-Institut für Informatik
Saarbrücken, Germany
{pan, schenkel, weikum}@mpi-inf.mpg.de

ABSTRACT

This demonstration presents an XML IR system that allows users to give feedback of different granularities and types, using Dempster-Shafer theory of evidence to compute expanded and reweighted queries.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Relevance feedback

General Terms

Algorithms

Keywords

relevance feedback, XML IR, dempster shafer theory

Introduction

Relevance Feedback is an important way to enhance retrieval quality by integrating relevance information provided by a user. In XML retrieval, existing feedback engines usually generate an expanded keyword query from the content of elements marked as relevant or nonrelevant. This approach that is inspired by text-based IR completely ignores the semistructured nature of XML.

The system presented in this demonstration exploits advanced user feedback beyond simple per-result feedback. Instead of supplying a single relevance value for a complete result element, the user may give feedback with different granularities and types. For example, the user may like a section in an article for content, but dislike the entire article for structure and also dislike a specific paragraph for content within the good section. We maintain a pool of possible query refinements (i.e., reweighing, adding or removing terms, structural constraints, or ontological expansions). Following earlier work by Ruthven and Lalmas [1] for text retrieval, we combine for each possible refinement candidate the relevance value for the result as delivered by the search engine, the initial weight of the candidate and the user feedback (with tunable weights according to granularity and type) using the Dempster-Shafer theory of evidence [2]. We then use the Transferable Belief Model [3] to compute, for each refinement candidate, a probability that it can identify relevant results.

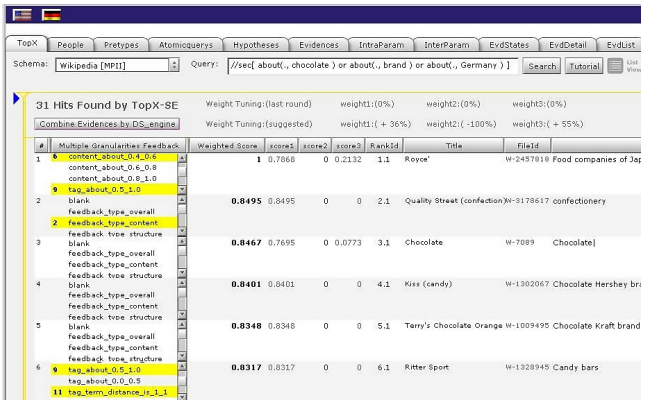

Figure 1: Screenshot of the user interface

Demonstration

The demonstration presents a full implementation of the relevance feedback techniques on top of the TopX Search Engine [4], using the INEX Wikipedia documents. Visitors can formulate their own queries, submit feedback of different types on the results, and examine the modified result lists. The system is implemented as a suite of Web services in Java, running on a Tomcat server.

1. REFERENCES

[1] I. Ruthven and M. Lalmas. Using dempster-shafer's theory of evidence to combine aspects of information use. *Journal of Intelligent Information Systems*, 19(3):267–302, 2002.

[2] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, NJ, 1976.

[3] P. Shets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–243, 1994.

[4] M. Theobald, H. Bast, D. Majumdar, R. Schenkel, and G. Weikum. Topx: efficient and versatile top- k query processing for semistructured data. *VLDB J.*, 17(1):81–115, 2008.

SIGIR'08, July 20–24, 2008, Singapore.
ACM 978-1-60558-164-4/08/07.