How does the EconBiz Relevance Ranking work?

General Thoughts

Relevance lies in the eye of the beholder. Students might consider a textbook or an MA or BA thesis to be highly relevant, while a researcher might need the latest working paper or an article from a peer-reviewed journal. The automatic relevance ranking in EconBiz is based on a number of assumptions about basic user needs. These may be helpful in one context and unsuitable in another. We want to be open about our criteria, so that users can understand how our result lists are built.

We continue to work on improvements and welcome your questions and comments:
info@zbw-online.eu

Our criteria

The relevance ranking is based on a simple text matching approach (TF-IDF; term frequency-inverse document frequency). Special syntactic features such as proximity of search words (especially phrases) or exact matching of the content of metadata fields are favoured. Matches in the title or in subject field are considered most important, but matches in other fields such as author, abstract, table of contents etc. also influence the ranking.

Other criteria are also used for *text matching*:

- Hits in the title and in the keyword field are more important than other hits.
- When several search words appear close together (e.g. as a phrase), this is more important than when the search words appear by themselves.
- Exact hits in a field are more important than hits where a search word or phrase only occurs.
- Exact hits are especially important in the keyword field. In the title field, exact matches and phrase matches of the entire search query are especially important.

In addition, other features of the documents may influence the ranking:

- Newer documents are preferred.
- Open Access documents are preferred, especially over versions of the same document that are not directly accessible.
- Some special document types are ranked down (e.g. there are some bachelor theses in the database BASE; if there is other relevant material, in most cases the bachelor theses should not appear on top of the list.)

Ranking and sorting options

If the relevance ranking is not suitable for a particular search, there are several sorting options that can help you find relevant results. In addition to the sorting options "newest" or "oldest" results first, there is a sorting option "articles prioritized". With this option, publications with the publication type "article" receive a boost and thus appear higher up in the result list, which is otherwise sorted according to the relevance criteria described above. (The value for the article boost is currently: 3).

You can also manually influence the weight of individual search. By using "^" you can give more weight to a search term and influence the sorting of the results. Example: unemployment AND (finland^20 OR scandinavia^1)

Some publications on this topic:

Flimm, O. (2007). Die Open-Source-Software OpenBib an der USB Köln – Überblick und Entwicklungen in Richtung OPAC 2.0. Bibliothek Forschung und Praxis, 31(2), 2-20.

Langenstein, A. & Maylein, L. (2010): Relevanz-Ranking im OPAC der Universitätsbibliothek Heidelberg. B.I.T. Online 12(4), 408-413, http://archiv.ub.uni-heidelberg.de/volltextserver/volltexte/2010/10343/pdf/Langenstein_Maylein_aus_BIT_4_09_kpl_kl.pdf (new URL: https://archiv.ub.uni-heidelberg.de/volltextserver/10343/).

Lewandowski, D. (2009). Ranking library materials. Library Hi Tech, 27(4), 584-593.

Lewandowski, D. (2010a). Der OPAC als Suchmaschine. In P. Danowski & J. Bergmann (Eds.), Handbuch Bibliothek 2.0. München: Saur / de Gruyter, 87-107.

Lewandowski, D. (2010b). Using search engine technology to improve library catalogs. In A. Woodswoth (Eds.), Advances in Librarianship, Vol. 32. Bingley: Emerald, 35-54.

Dellit, A. & Boston, T. (2007). Relevance ranking of results from MARC-based catalogues: from guidelines to implementation exploiting structured metadata. *Library*, (February), 1-14, http://www.nla.gov.au/openpublish/index.php/nlasp/article/viewArticle/1052 (new URL, no fulltext https://api.semanticscholar.org/CorpusID:60634362).

A first version of this document was created on July 06, 2012, additions from April 02, 2024.