

## Wie funktioniert das Relevanzranking bei EconBiz?

### Vorbemerkung

Relevanz liegt letztlich natürlich im Auge der Betrachtenden. Für Studierende kann ein Lehrbuch oder eine Bachelor- oder Masterarbeit zu einem Thema sehr hilfreich und damit hoch relevant sein, für Forschende ist ein ganz aktuelles Working-Paper oder ein Artikel aus einer speziellen Zeitschrift eher von Interesse. Die automatische Relevanzbewertung muss (solange wir keine individuellen Lösungen anbieten können) einige Grundannahmen treffen. Diese können in einem Zusammenhang gut, in einem anderen Zusammenhang wiederum unpassend sein. Mit dem folgenden Einblick in die Relevanzkriterien bei EconBiz wollen wir zunächst einmal Transparenz schaffen.

Über Fragen und Rückmeldungen zu diesem Thema freuen wir uns.

Sprechen Sie uns gern an: [info@zbw-online.eu](mailto:info@zbw-online.eu)

### Die Relevanzkriterien in EconBiz

Die Relevanz eines Dokumentes kann bezüglich verschiedener Kategorien bewertet werden:

- Bzgl. der konkreten Suchanfrage (*keyword matching* oder *text matching*)
- Bzgl. allgemeiner Eigenschaften des Dokumentes, wie z.B. Popularität (etwa in Form von Ausleihzahlen, Downloads, Zitationshäufigkeit, etc.)
- Bzgl. spezieller Präferenzen der Nutzer:innen, wie z.B. Aktualität oder Verfügbarkeit

In EconBiz werden derzeit nur Kriterien aus der ersten und der letzten Kategorie benutzt. In der letzten Kategorie müssen entsprechend Annahmen über die Bedürfnisse getroffen werden.

Das *text matching* basiert auf einem Standard-Verfahren (**TF-IDF**; term frequency-inverse document frequency). Dabei wird ein Dokument als relevanter erachtet, je häufiger ein Suchwort in dem Dokument vorkommt. Besteht die Suchanfrage aus mehreren Wörtern, hat ein selteneres Wort mehr Einfluss auf die Relevanzbestimmung. Je mehr Suchwörter in einem Dokument vorkommen, desto relevanter ist das Dokument (bei OR-Verknüpfungen).

Zusätzlich werden noch weitere Kriterien beim *text matching* benutzt:

- Treffer im Titel und im Schlagwortfeld sind wichtiger als sonstige Treffer.
- Treten mehrere Suchwörter nah beieinander auf (z.B. als Phrase), ist dies wichtiger, als wenn die Suchwörter allein stehend auftreten.
- Exakte Treffer in einem Feld sind wichtiger, als Treffer in denen ein Suchwort bzw. eine Phrase nur vorkommt.
- Im Schlagwortfeld sind exakte Treffer besonders wichtig. Im Titel sind exakte Treffer und Phrasentreffer der gesamten Suchanfrage besonders wichtig.

Neben den *text matching* Kriterien werden die folgenden Präferenzen der Nutzer:innen angenommen:

- Aktuellere Dokumente werden bevorzugt. Wenn die Dokumente nicht aus den letzten 10 Jahren sind, nimmt diese Präferenz stark ab.
- Direkt verfügbare bzw. frei zugängliche Dokumente haben ein höheres Gewicht. (Ist eine Publikation z.B. mehrfach – gedruckt, online lizenzpflichtig und online frei verfügbar – nachgewiesen, so sollte die frei zugängliche Publikation als erste erscheinen.)
- Einige spezielle Dokumententypen sind i.A. nachrangig. (In der Datenquelle BASE sind z.B. auch Bachelorarbeiten nachgewiesen. Wenn diese als solche in den Metadaten ausgezeichnet sind, werden sie als weniger wichtig bewertet.)

## Ranking und Sortieroptionen

Wenn das Relevanzranking für einzelne Anliegen unpassend ist, können verschiedene Sortieroptionen beim Auffinden passender Treffer helfen. Neben den Sortieroptionen „neueste“ bzw. „älteste“ Treffer zuerst, gibt es auch eine Sortieroption „Artikel priorisiert“. Bei dieser Option erhalten Publikationen mit dem Publikationstyp „Artikel“ einen Boost und landen somit in der ansonsten nach den oben beschriebenen Relevanzkriterien sortierten Trefferliste weiter oben. (Der Wert für den Artikel-Boost beträgt aktuell: 3.)

Darüber hinaus kann die Gewichtung einzelner Suchwörter manuell beeinflusst werden. Durch Verwendung von „^“ kann ein Suchbegriff stärker gewichtet und so die Sortierung der Treffer beeinflusst werden. Beispiel: *unemployment AND (finland^20 OR scandinavia^1)*

### Weiterführende Literatur zum Thema:

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](http://archiv.ub.uni-heidelberg.de/volltextserver/volltexte/2010/10343/pdf/Langenstein_Maylein_aus_BIT_4_09_kpl_kl.pdf) (neue 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 (Hrsg.), *Handbuch Bibliothek 2.0*. München: Saur / de Gruyter, 87-107.

Lewandowski, D. (2010b). Using search engine technology to improve library catalogs. In A. Woodsworth (Hrsg.), *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> (neue URL, ohne Volltext: <https://api.semanticscholar.org/CorpusID:60634362>).

Eine erste Version dieses Dokuments wurde erstellt am 06.07.2012, Ergänzungen vom 02.04.2024.