Only the first results count: user-feedback-modified relevance ranking in E-LIB Bremen

Since the introduction of new search engine-based library catalogues and resources discovery systems, users have been able to find millions of online documents that are available in parallel to local holdings. Users of these new, continually expanding catalogue systems need to be able to find titles in a convenient way. The quality of recommendation systems and especially relevance ranking methods is very important for the success of discovery catalogues. Up to now, library catalogue ranking methods have largely ignored user feedback. However, in 2011, the State and University Library Bremen (SuUB) developed modified ranking methods for its own discovery system, Electronic Library Bremen (E-LIB Bremen). These make use of popularity factors such as click statistics and information about the number of editions published as well as how many copies of a given title are available. This customized development has enabled fast and flexible new retrieval functions to meet the needs of local users.

Since the introduction of new search engine-based library catalogues and resource discovery systems, users have been able to find millions of online documents that are available in parallel to local holdings. Users of these new, continually expanding catalogue systems need to be able to find titles in a convenient way. The quality of recommendation systems and especially relevance ranking methods is very important for the success of discovery catalogues in academic libraries. However, current methods of relevance ranking do not provide user-friendly result sets in these systems. This is particularly the case for simple queries, where these systems tend to produce high numbers of hits with an identical value for the calculated relevance. The identification of relevant titles in these result sets is still difficult for the library user. In this situation it is necessary to modify the method of title display sorting (ranking) in a user-friendly way by refined differentiation of the ranking order.

In contrast to ranking methods in current web search engines, library catalogues and discovery systems do not analyse properties that express the popularity of a particular title with the user. Ways of addressing the appropriate additional ranking parameters have been described in the literature.[1,2]

At the State and University Library Bremen (SuUB), the catalogue search engine E-LIB Bremen has used the analysis of such properties for the refinement of the purely text-statistical ranking method since 2011. Users' click statistics, the edition number of copies and the total number of copies of each title held in the library are used in this modified method of ranking.

MARTIN BLENKLE

Head of Digital Services
State and University Library Bremen (SuUB), Germany

RACHEL ELLIS

Electronic Resources Librarian
SuUB, Germany

ELMAR HAAKE

Search Engine Developer and Subject Librarian
SuUB, Germany

Relevance ranking of search engines and library catalogues

There is currently a major change taking place in libraries, whereby older catalogues (OPACs), whose search techniques date back to the 1990s, are being replaced by resource discovery systems. These new 'catalogues' not only offer enlarged search index metadata pools, but also a fundamentally different retrieval concept.

In general, traditional OPACs use the method of 'exact match' retrieval, which delivers titles that contain only the exact search term used. In case of multiple search terms, the result

list can be modified by the rules of Boolean algebra. Metadata objects may be considered either appropriate or not to the relating search term. A ranking of results or a differentiated evaluation of the suitability of objects in terms of the search question is not provided in this model. The standard sorting method for such systems is chronological as there are no other appropriate alternatives for sorting.

For untrained users, a number of shortcomings make searching in traditional library catalogues difficult and may prevent successful retrieval. The use of Boolean search logic is unfamiliar and inconvenient for many users.[3] In a very large index it is increasingly difficult to find records accurately when using this retrieval model, particularly in the case of more general or simple search terms. The chronological sorting in OPACs today contradicts the user's anticipation of common methods of sorting in descending relevance. Relevance ranking in this context is defined as the relevance calculated according to system algorithms, though the user expected relevance is subjective and can differ from this. The aim should be to ensure that the relevance calculated by the system is compliant with the so-called objective relevance of a bigger user group within an academic university.

Since all hits resulting from an 'exact match' routine are basically valid in relation to the search terms, the requested title may be found on the last page of the result set. In contrast, however, most of the user's attention is linked to the first page of results, with second and all later result pages now largely ignored by users.[4,5,6] The successful positioning of the user's expected titles in the top ten search results is therefore an increasingly important factor for the acceptance of a catalogue search engine.

'most of the user's attention is linked to the first page of results'

Today's library customers use search services such as Google intuitively, providing them with results using a 'best match' retrieval model, which is popular with users. Current discovery catalogues work in a similar way. The matching and result sorting process of these systems is largely based on a calculated relevance analysis, which analyses the statistical text similarity between search terms and the metadata object. Documents and queries are represented as vectors in a multidimensional space, and then compared with each other in terms of their similarity. The more similar the query and the document vector are, the higher the calculated relevance of each document will be.[7] At the same time, the calculated relevance value of a document determines its position in the search results. The rules for determining the statistical text similarity are based on known methods such as term weighting (tf-idf),[8] e.g. the analysis of the frequency of terms in the document as compared to the overall incidence in the whole index database.

In order to improve the recall, natural language processing is then applied. For example, several morphological variants of an input term can be reduced to a common root form (stemming). The presentation of results is subsequently generated in order of decreasing similarity or relevance between query and hits, and should be in compliance with the user's expectation. In contrast to the 'exact match' process, these methods can produce relevant hits even in cases of suboptimally worded or incorrect requests. In subject retrieval queries, the transition from high to medium to little relevant hits can be expressed much better using 'best match' methods. By showing most relevant hits at the top of the results list, these new catalogues meet the current user expectation in a significantly better way.

'these methods can produce relevant hits even in cases of suboptimally worded or incorrect requests'

The needle in the haystack: problems of current ranking algorithms in library catalogues

The techniques described above, which are used in all current implementations of best-match-based catalogue systems (e.g. Lucene-SOLR/Vufind or Elasticsearch), are able to deliver user-friendly results in significantly expanded indexes and can clearly improve library search for untrained users.[9] However, the current state of ranking technology in library systems has potential for even further improvement.

The algorithms of the best match method are based on rules which are much more complex than a simple yes/no logic of an 'exact-match'-based catalogue. Therefore, the interpretation of result sets often appears to be less easy to understand for users and librarians. In addition to this, the exact rule sets are generally not published in detail by library system vendors and cannot be fully customized by a library systems team.

Another problem is caused by the relatively small amount of text in library metadata records. The vector space retrieval is well suited to deliver differentiated results in the statistical analysis of large amounts of text, such as full text of web pages or electronic full-text documents. However, bibliographic metadata, in contrast, have very different attributes regarding their text properties: for example, the amount of text that can be analysed is much less. Therefore, statistical analysis of pure bibliographic metadata can lead to a dissatisfactory differentiation of ranking results.[10]

'statistical analysis of pure bibliographic metadata can lead to a dissatisfactory differentiation of ranking results'

The significant problem of ranking methods was recognized at SuUB Bremen in connection with continuously expanding databases, leading to a growing number of objects indexed in the catalogue. Subsequently, important titles were increasingly difficult to locate using the discovery catalogue.

Simple queries produced many hits (often hundreds) showing nearly identical results in the relevance analysis within the large set of data held by the current discovery systems. In this instance a chronological sorting and presentation of hits with equal relevance is required. The main target of the catalogue ranking methods – the presentation of a user-oriented result list – cannot be achieved in such cases. Despite the use of advanced search algorithms and ranking techniques, users are not able to locate requested titles in the huge number of equally weighted results.

This is especially the case in common, repeatedly requested standard literature and basic titles, leading to a critical problem for users and a lack of acceptance of new catalogues as a result. For this reason, the ranking methods of E-LIB Bremen were modified in 2011 to achieve a user-oriented, location- and usage-dependent differentiation of the search engine's result lists in such cases.

Electronic Library Bremen (E-LIB Bremen) indexing and retrieval methods

In-house development of the library's discovery catalogue offers options to implement adjustments and settings tailored to the needs of the campus for SuUB Bremen on a large scale. This refers both to a complete control of the index as well as the mode of the result presentation and the adjustment of relevance ranking algorithms.

The discovery search interface of E-LIB Bremen was originally introduced in 2004. The system holds e-media and print media metadata in a single local index and presents all titles in an integrated result list. This system replaced the previously used OPAC in 2010 and handles approximately 13,000-18,000 daily searches.[11,12]

The main search and indexing methods of E-LIB Bremen (CiXbase) are based on a development begun in 1993 at the University Library and the University of Osnabrück, Germany. From 1999, this system was developed (among others) and used by the State and University Library Bremen. Various types of metadata are imported and converted to a homogenous XML flat file structure in order to build the local knowledge base. The resulting consistent metadata pool is indexed via CiXbase technology based on statistical balanced B*-trees algorithms serving as a basis for the search engine.[13] In addition to conventional methods of processing natural language queries, the system has a number of local methods for the analysis of semantic and linguistic units of a particular query.[14]

How can sorting of the result lists be further improved in discovery systems?

In order to achieve a user-oriented adjustment of the search engine's ranking, which aims to shift frequently requested titles to a prominent position in the result list, additional parameters have been introduced to modify the ranking of every single title. Parameters calculated purely on text statistics are not sufficient.

'additional parameters have been introduced to modify the ranking of every single title'

Current web search engines show that a search in large text domains can be successfully designed that includes both similarity analysis of the search terms and the popularity of the indexed objects. For example, Google expands its relevance ranking algorithms by adding parameters that reflect user feedback. The relevance of the results is weighed not only by textual characteristics, but also by their popularity with the user. To measure a link's popularity, Google obtained a patent for the 'Page Rank' method[15] in 1997. Hits are rated higher if a lot of other internet websites link to this particular host.

Within the E-LIB Bremen discovery system, we assume that through incorporation of further title characteristics, the relevance ranking of large result lists in library catalogues can be much more user oriented. Parameters for measuring title popularity are used in the same way as by the large internet search engines. These aim to favour highly requested titles in a result list where many titles are of similar relevance.

'Parameters for measuring title popularity are used in the same way as by the large internet search engines'

As the algorithms for relevance ranking are already very complex, a rather simple set of rules was developed. At the same time, modification of the ranking should not result in a dynamic increase in the effect for popular titles.

The following specifications have been used as suitable parameters by E-LIB Bremen since 2011:

1. Number of copies held locally: the acquisition policy of the library reflects the local request situation.
2. Edition information of the title: this property reflects a more general user request. Titles that have reached a high number of editions are of a general high interest.
3. Internal click statistics showing user interaction within the E-LIB system.

Highly requested titles are selected more often in the system. Based on an internal title-based transaction log analysis (TLA), these titles can be identified accurately. In each one of the three above categories, all the index-included metadata get a small increase in the relevance impact factor if a certain defined, significantly elevated level has been reached. These titles are then placed higher in the displayed result list. The exact parametrization is determined empirically. It is very important that the newly defined relevance impacts are tuned in a subtle way so the text statistic relevance factors do not get overcompensated. Table 1 shows the simple set of rules followed by the E-LIB Bremen search engine.

Popularity parameter	Relevance boost in % (added only once per title)
Acquisition policy (number of copies)	3
Media properties (edition)	3
User clicks	5

Table 1. Rules followed by E-LIB Bremen search engine

The relevance boost is only assigned once in each category for each title in order to eliminate dynamic self-increase of titles. Hits that already have a 100% relevance due to the text statistical relevance are favoured this way. Titles held at Bremen that match all three

criteria – that is, a high number of copies, several editions and most frequently selected titles within the E-LIB system – get a one-time boost in all three categories and can reach a maximum relevance of 111%.

E-LIB Bremen title rank: lessons learned and summary

The method described above was implemented within E-LIB Bremen and tested with different settings in the additional relevance parameter. User requests at our search engine can be monitored on an anonymous basis, so it is possible to analyse typical user requests, review these based on the new method and compare the results to the previous standard method.

The impact of 5% for click activities, 3% for editions and 3% for the number of copies have shown the best results for Bremen:

- as expected, the relevant hits that only have a very general or highly used title phrase were previously ranked lower in the result list and are now better positioned
- recent classic titles are displayed more prominently – locally requested, popular textbooks show up more quickly
- highly specific journal articles that do not get a popularity impact boost in general are now positioned lower in the result list for a very general search approach.

'locally requested, popular textbooks show up more quickly'

Evaluation of the ranking parameters is continuing by observing user searches; however, using a rather simple set of rules already shows impressive improvements in the ranking impact. New releases may have a slight disadvantage due to missing click activities with this method, but this effect is only short lived as the project evaluation shows that this is compensated very quickly by user interaction.

'shows impressive improvements in the ranking impact'

After implementation four years ago, these modified ranking algorithms are now part of the routine operation at the State and University Library Bremen. Looking forward, there is potential for further development of popularity factors for ranking in library catalogues, such as the LIBRANK[16] research project (a Hamburg University of Applied Sciences and German National Library of Economics project), which started in 2014 and is currently investigating new forms of relevance ranking and also evaluation via systematic user research.

Competing interests: The authors have declared no competing interests.

References

1. Mi, J and Weng, C, Revitalizing the library OPAC: interface, searching, and display challenges, *Information Technology and Libraries*, 2008, 27(1), 5–22 http://dx.doi.org/10.6017/ital.v27i1.3259
2. Lewandowski, D, Ranking library materials, *Library Hi Tech*, 2009, 27(4), 584–593: http://dx.doi.org/10.1108/07378830911007682
3. Holman, L, Millennial students' mental models of search: implications for academic librarians and database developers, *Journal of Academic Librarianship*, 2011, 37(1), 19–27. http://dx.doi.org/10.1016/j.acalib.2010.10.003
4. Huang, Y, Zhang, M and Ge, X, A user behavior based study on search engine ranking, *Computer Engineering and Intelligent Systems*, 2014,.5(12), 94–105 http://www.iiste.org/Journals/index.php/CEIS/article/view/18139 (accessed 6 February 2015).
5. Hochstotter, N and Koch, M, Standard parameters for searching behaviour in search engines and their empirical evaluation, *Journal of Information Science*, 2009, 35(1), 45–65: http://dx.doi.org/10.1177/0165551508091311
6. Bar-Ilan, J, Levene, M and Mat-Hassan, M, Methods for evaluating dynamic changes in search engine rankings: a case study, *Journal of Documentation*, 2006, 62(6), 708–729: http://dx.doi.org/10.1108/00220410610714930
7. Manning, C D, Raghavan, P and Schütze, H, *Introduction to information retrieval*, 2008, Cambridge, Cambridge University Press, 100–123.
8. Zhai, C, *A brief review of information retrieval models*, 2007: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.5325 (accessed 6 February 2015).

9. Holman, L, ref 3.
10. Schneider, K G, 25 July 2007, Relevance ranking and OPAC records, Free Range Librarian: KG Schneider's blog on librarianship, writing, and everything else: http://freerangelibrarian.com/2007/07/25/relevance-ranking-and-opac-records/ (accessed 6 February 2015).
11. Staats- und Universitätsbibliothek (SuUB) Bremen: http://www.suub.uni-bremen.de/home-english/ (accessed 13 May 2015).
12. Blenkle, M, Ellis, R and Haake, E, Next-generation library catalogues: review of E-LIB Bremen, *Serials*, 2009, 22(2), 178–181; DOI: http://dx.doi.org/10.1629/2268 (accessed 28 May 2015).
13. Blenkle, M, Ellis, R and Haake, E, ref 12.
14. Ostermann, T, et al, Linguistic processing and classification of semi structured bibliographic data on complementary medicine, *Cancer Informatics*, 2009, 7, 159–169: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2730176/ (accessed 6 February 2015).
15. Page, L et al, *The PageRank citation ranking: bringing order to the web*, Technical Report, 1999, Stanford InfoLab: http://ilpubs.stanford.edu:8090/422/ (accessed 6 February 2015).
16. LIBRANK: http://www.searchstudies.org/de/librank.html (accessed 6 February 2015).

Corresponding author: Rachel Ellis

Electronic Resources Librarian
State and University Library Bremen, Germany
E-mail: ellis@suub.uni-bremen.de
ORCID ID: http://orcid.org/0000-0002-3517-8111

Martin Blenkle
ORCID ID: http://orcid.org/0000-0003-0898-3753

Elmar Haake
ORCID ID: http://orcid.org/0000-0002-8115-6044

To cite this article:
Blenkle, M, Ellis, R and Haake, E, Only the first results count: user-feedback-modified relevance ranking in E-LIB Bremen, *Insights*, 2015, 28(2), 75–80; DOI: http://dx.doi.org/10.1629/uksg.235.

Published by UKSG in association with Ubiquity Press on 07 July 2015