Optimizing Relevance Ranking to Enhance the User's Discovery Experience*

Tamar Sadeh

Ex Libris Group, Jerusalem, Israel
tamar.sadeh@exlibrisgroup.com

Abstract. With the introduction of library discovery systems, the display of results according to relevance, as determined by the system, has become a norm. To investigate how relevance ranking could be optimized, a provider of a widely used discovery system developed methods of evaluating the system's relevance ranking. As a result, new factors were added to the calculation of search results' relevance—information about the individual user and the user's information needs, and an indicator representing the academic significance of materials. Methods of monitoring the impact of changes were also established.

Keywords: relevance ranking, user experience, discovery systems, personalized ranking.

1 Introduction

Developed at the turn of the millennium to enable users with little information literacy to reach general information without the help of mediators (such as travel agents, salespeople, or librarians), Web search engines have shaped the way in which students and researchers seek scholarly information today. However, the ease of use, immediacy of results, and heterogeneous nature of information provided by Web search engines such as Google trigger expectations that libraries have only recently started to meet with "new-generation" library discovery systems (e.g., [1], [2], [3]).

These index-based discovery systems, available from 2009, aim to provide a single entry point to the scholarly information landscape and enable users to search in a Google-like way in local, global, and regional collections that libraries offer their users. Such global and regional collections, provided by primary and secondary publishers and aggregators, include journal articles, e-books, conference proceedings, newspaper articles, theses, patents, and other types of materials. The collections differ in many ways, such as in the type of content (metadata, abstract, or full text), the format and depth of the available metadata, and the licensing options.

Although new-generation discovery systems may match Web search engines in ease of use and speed, the success of a search process is measured not just by the amount of time that elapses until a result list is displayed. Much more important is the

* To protect Ex Libris Ltd.'s intellectual property, some details are not included in this paper.

T. Catarci, N. Ferro, and A. Poggi (Eds.): IRCDL 2013, CCIS 385, pp. 153–164, 2014.

amount of time that an information system takes to satisfy a user's need and furnish the desired outcome. A key challenge of library discovery systems is how to provide users with the most relevant items from the immense landscape of available content. To meet this challenge, developers have enriched systems with new features. For example, faceted navigation helps users quickly refine their result list and focus on its subsets (see [4] and [5]), and recommendations based on other users' prior selections draws searchers' attention to items related to the topic of their search, even when such articles do not match the query terms that the searchers entered.

However, because of their familiarity with Google and other search engines, users of discovery systems tend to scan only the topmost results; hence, items that are most relevant for a particular search can easily remain unnoticed if they are not displayed near the top of the list. Whereas past discussions of relevance in information retrieval distinguished between what is relevant and what is not, the current focus is on the *degree* of relevance, with the understanding that "in the most fundamental sense, relevance has to do with effectiveness of communication" [6]. Relevance ranking, whose purpose is to highlight materials that the system deems the most pertinent for the particular query, has become a major factor in satisfying user needs. Together with the immediate delivery of retrieved items, relevance ranking has had a huge impact on increasing the value of library services for users and institutions.

2 Relevance in the Scholarly Domain

Although "relevance is a, if not even *the*, key notion in information science in general and information retrieval in particular" [7], the application of relevance in library information systems developed in the past was (and in some of those systems, continues to be) limited to a binary approach: a document was considered either relevant or not relevant to a specific query. The determination was based on the textual similarity between the item and the query; the user's specific information need, among other things, was not considered, and results were not arranged according to the degree of relevance to the specific user. Rather, results were displayed according to unambiguous criteria, such as the date or the alphabetic order of the title or author name.

Over the past decade, library information systems entered a new realm and adopted relevance ranking so that result lists would address user expectations. Harnessing traditional relevance-ranking methods used by other information systems, library systems adjusted the methods for scholarly materials.

The theoretical aspects of relevance ranking have been studied for many years [e.g., 8], but literature on practical implementations has been lacking. In an attempt to develop an automated calculation of relevance that includes information about the user and the user's information need and leverages well-structured metadata and other data available about scholarly materials, one discovery system vendor, Ex Libris, conducted a research program whose results transformed the discovery system's relevance ranking.

3 Direct and Inferred Assessment of Relevance

One of the major challenges in developing relevance ranking is assessing its success. Because relevance is subjective, a qualitative evaluation of success depends on the characteristics of each individual who tests the information system. One's specific information needs, expectations, and expertise in both the area of research and searching techniques come into play when one assesses the relevance of displayed results.

Another approach, less dependent on the engagement of individuals in testing, is to compare the order in which a system displays the results with the order of results in another system that is known to have excellent relevance ranking. This approach, while easy to implement, has major shortcomings. First, no scholarly information system has a relevance-ranking method that is reliable enough to serve as a model. Second, the content available in every such system is different: in a discovery system, the content depends on the institution's subscriptions and policies; when it comes to Google Scholar or Microsoft Academic Search, the exact search scope is not documented. Third, the format, degree of completeness, and depth of the available metadata (basic or comprehensive metadata) and the presence or absence of full text in addition to metadata differ from one system to another.

An alternative approach is to quantitatively test how modifications in relevance-ranking technology affect searchers' behavior. An analysis of usage data from a large number of users can shed light on their overall satisfaction; measures such as an increase in the number of sessions that culminate in the selection of an item and a higher average position of selected items in the result list are likely to indicate that the relevance ranking has improved.

4 Relevance-Ranking Project

4.1 Aim and Objectives of the Project

In 2010, a centralized, cloud-based index of hundreds of millions of global and regional scholarly materials such as journal articles, e-books, conference proceedings, and newspaper articles was added to the Ex Libris Primo® discovery system. As a result, users at Primo libraries can launch a single search that covers both their local library holdings and an information landscape that transcends the traditional boundaries of the library's offering. This landscape includes subscribed and open-access materials and, if the library so desires, materials that the library or the user can purchase on demand. Because of the amount of data available, the heterogeneous nature of this virtual collection, and users' tendency to implement "simple" search queries [9], [10], the Primo system's relevance ranking had to be enhanced to effectively support the needs and expectations of the users.

To investigate the ways in which Primo relevance ranking should be improved, Ex Libris assembled a team of information retrieval experts, research and development staff, and product managers. In March 2011, the team began a long-term project to support the continued development of the Primo relevance-ranking technology. The team defined the following project objectives:

- Examine search logs and usage statistics to understand search trends and the ways in which Primo accommodates the various search practices of end users
- Set a baseline for the assessment of Primo relevance ranking by obtaining researchers' evaluations of the current sorting order of the results
- Define metrics to evaluate the effectiveness of Primo relevance ranking
- Build a test environment in which the defined metrics are used to test and monitor changes to the technology
- Optimize the calculations underlying Primo relevance ranking and incorporate additional information about an item (the item's "value score"), the user, and the user's needs at the time of the search
- Monitor real-world Primo users' search behavior over time to evaluate the impact of the improvements and suggest further improvements

4.2 Test Environment and Testing Tools

To set a baseline that would later serve as a reference point and to test the changes that would be introduced to the relevance-ranking technology, the project team set up a test environment at the Ex Libris research and development lab and developed testing tools. The methods for evaluating relevance ranking in the test environment were based on both a qualitative evaluation by users and a comparison with Google Scholar.[1] This comparison was not limited to an examination of the differences between the result lists of Primo and Google Scholar; rather, the examination included input from researchers regarding both lists.

The test environment consisted of a Primo implementation with a search scope of 40 million records, a representative sample of the content indexed in the Primo Central Index at the time of the testing. In addition, the team developed two testing tools:

- A tool that makes ad hoc changes to parameters affecting the ranking method but does not require reindexing of the data or changes in the software. With this tool, the team could further boost materials published in the current year.
- A tool that runs test queries and calculates two common measures of quality[2]:
  - Mean average precision (MAP), which indicates how successful the system was in ranking "best" items above other items. A "best" item is one that should appear on the first page of results.
  - Mean reciprocal rank (MRR), which indicates how high in the result list the first "best" item appears

[1] Both Primo and Google Scholar index a large number of academic publications of various types and from many information providers, and enable users to search via a single, simple interface. Despite differences between these systems (primarily, libraries' control over the search scope, user interface, and integrated services), the search experience is similar.

[2] See `http://en.wikipedia.org/wiki/Mean_average_precision` and `http://en.wikipedia.org/wiki/Mean_reciprocal_rank`.

MAP and MRR metrics typically distinguish between relevant items and non-relevant items in an entire result list. However, for the purposes of the project's relevance-ranking tests, a distinction was made between "best" results—those that should appear on the first page—and "other" results (which may be relevant to the query but are not necessarily relevant enough to be on the first page). Furthermore, the MAP and MRR values were calculated for only the top 20 results.

4.3 Setting a Baseline

The project began with a user evaluation of the Primo relevance ranking that was available in March 2011. To enable the team to obtain a more objective assessment, the evaluation included a comparison between Primo relevance ranking and that of Google Scholar (the team kept in mind the caveats regarding such a comparison, as previously mentioned). The data from the evaluation formed a baseline to help the team assess the impact of future improvements.

A group of senior researchers from several institutions agreed to evaluate the results of the Primo relevance ranking. These evaluators work in various disciplines: agriculture, anthropology, biology, medicine, military and security studies, philosophy, and physics.[3] Each evaluator created search queries related to his or her area of expertise, including broad-topic queries, narrow-topic queries, known-item queries, and other types (Fig. 1). In addition, the project team conducted in-depth interviews with three evaluators to learn what they expect from a scholarly information system.

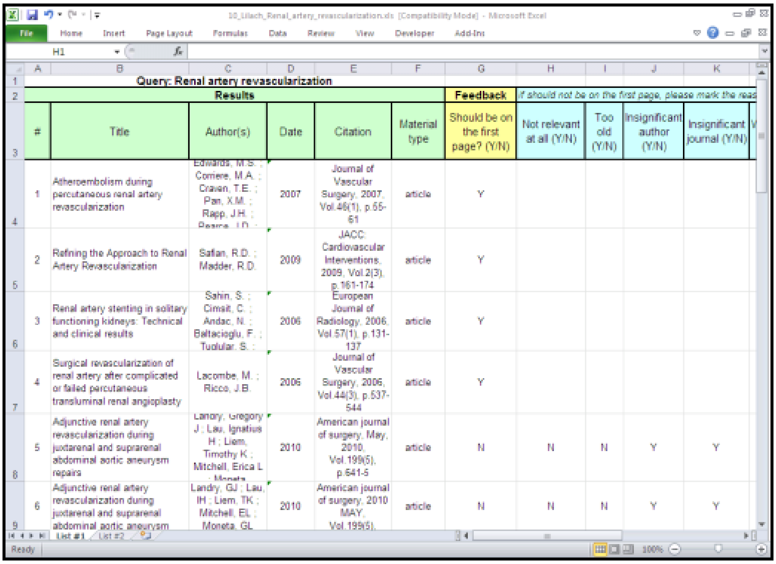
Flavor in soft leptogenesis
gated super conductivity
halperin lee read
labor process & structures of feeling
Particle cosmology
marxist anthropology
anthropology of class israel
ranciere nights of labor
chloroplast
photoinhibition
the role of proteases in repair from photoinhibition
Twin earth thought experiment
Frank Jackson what mary didn't know the journal of philosophy
Renovascular hypertension
Controversies in renal artery stenosis: a review by the American Society of Nephrology Advisory Group on Hypertension

Fig. 1. A sample of the experts' queries

The team ran the queries in Primo and Google Scholar and sent each evaluator a spreadsheet that listed, for each query, the first 20 results returned by Primo and the first 20 results returned by Google Scholar. To exclude bias as a factor in the evaluation, the team did not inform the evaluators of the results' origin. The evaluators' task was to indicate whether each item should be displayed on the first page of results. Also, for each item that the evaluators said should not be on the first page, they were

[3] The decision to set a baseline with the help of experts was based on the assumption that these experts know which results to expect when submitting their queries.

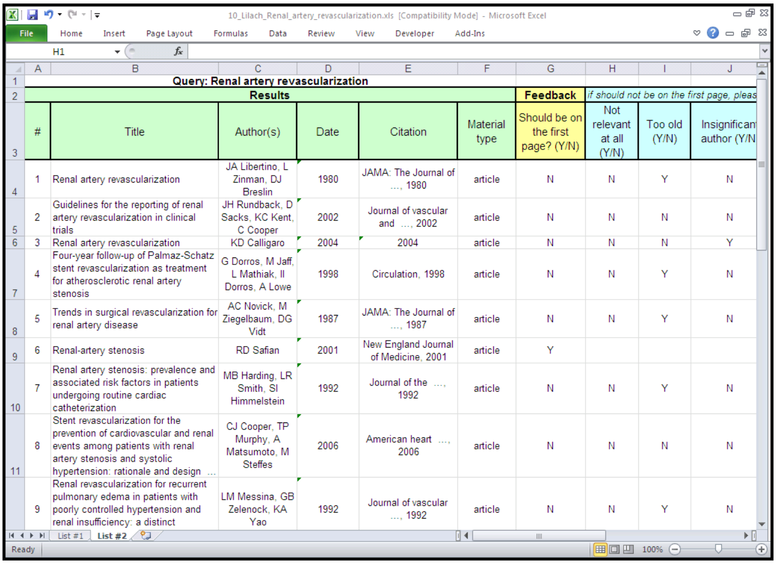
asked to explain why, by selecting at least one reason from a list or writing their own reasons (Fig. 2 and Fig. 3). The spreadsheet offered the following reasons: not relevant at all, too old, insignificant author, insignificant journal, wrong material type, too specific, too broad, and not the specific article the researcher was looking for.



Query: Renal artery revascularization

#	Title	Author(s)	Date	Citation	Material type	Should be on the first page? (Y/N)	Not relevant at all (Y/N)	Too old (Y/N)	Insignificant author (Y/N)	Insignificant journal (Y/N)
1	Atheroembolism during percutaneous renal artery revascularization	Edwards, M.S. ; Corriere, M.A. ; Craven, T.E. ; Pan, X.M. ; Rapp, J.H. ; Pearce, J.D. ;	2007	Journal of Vascular Surgery, 2007, Vol.46(1), p.55-61	article	Y				
2	Refining the Approach to Renal Artery Revascularization	Safian, R.D. ; Madder, R.D.	2009	JACC: Cardiovascular Interventions, 2009, Vol.2(3), p.161-174	article	Y				
3	Renal artery stenting in solitary functioning kidneys: Technical and clinical results	Sahin, S. ; Cimsit, C. ; Andac, N. ; Baltacioglu, F. ; Tuglular, S. ;	2006	European Journal of Radiology, 2006, Vol.57(1), p.131-137	article	Y				
4	Surgical revascularization of renal artery after complicated or failed percutaneous transluminal renal angioplasty	Lacombe, M. ; Ricco, J.B.	2006	Journal of Vascular Surgery, 2006, Vol.44(3), p.537-544	article	Y				
5	Adjunctive renal artery revascularization during juxtarenal and suprarenal abdominal aortic aneurysm repairs	Landry, Gregory J ; Lau, Ignatius H ; Liem, Timothy K ; Mitchell, Erica L ; Moneta	2010	American journal of surgery, May, 2010, Vol.199(5), p.641-5	article	N	N	N	Y	Y
6	Adjunctive renal artery revascularization during juxtarenal and suprarenal abdominal aortic aneurysm	Landry, GJ ; Lau, IH ; Liem, TK ; Mitchell, EL ; Moneta, GL	2010	American journal of surgery, 2010 MAY, Vol.199(5),	article	N	N	N	Y	Y

Fig. 2. Results returned by Primo and evaluated by a medical researcher



Query: Renal artery revascularization

#	Title	Author(s)	Date	Citation	Material type	Should be on the first page? (Y/N)	Not relevant at all (Y/N)	Too old (Y/N)	Insignificant author (Y/N
1	Renal artery revascularization	JA Libertino, L Zinman, DJ Breslin	1980	JAMA: The Journal of ..., 1980	article	N	N	Y	N
2	Guidelines for the reporting of renal artery revascularization in clinical trials	JH Rundback, D Sacks, KC Kent, C Cooper	2002	Journal of vascular and ..., 2002	article	N	N	N	N
3	Renal artery revascularization	KD Calligaro	2004	2004	article	N	N	N	Y
4	Four-year follow-up of Palmaz-Schatz stent revascularization as treatment for atherosclerotic renal artery stenosis	G Dorros, M Jaff, L Mathiak, II Dorros, A Lowe	1998	Circulation, 1998	article	N	N	Y	N
5	Trends in surgical revascularization for renal artery disease	AC Novick, M Ziegelbaum, DG Vidt	1987	JAMA: The Journal of ..., 1987	article	N	N	Y	N
6	Renal-artery stenosis	RD Safian	2001	New England Journal of Medicine, 2001	article	Y			
7	Renal artery stenosis: prevalence and associated risk factors in patients undergoing routine cardiac catheterization	MB Harding, LR Smith, SI Himmelstein	1992	Journal of the ..., 1992	article	N	N	Y	N
8	Stent revascularization for the prevention of cardiovascular and renal events among patients with renal artery stenosis and systolic hypertension: rationale and design ...	CJ Cooper, TP Murphy, A Matsumoto, M Steffes	2006	American heart ..., 2006	article	N	N	N	N
9	Renal revascularization for recurrent pulmonary edema in patients with poorly controlled hypertension and renal insufficiency: a distinct	LM Messina, GB Zelenock, KA Yao	1992	Journal of vascular ..., 1992	article	N	N	Y	N

Fig. 3. Results returned by Google Scholar and evaluated by the same medical researcher who evaluated the results in Fig. 2

The project team received the evaluations and calculated the lists' MAP and MRR. Despite the fact that the test environment covered only 40 million Primo Central Index records, the results clearly showed that as a starting point for the project, the Primo ranking was respectable. For narrow-topic searches, the MAP of Primo was better than that of Google Scholar; for known-item searches, the Primo MAP was surprisingly high in light of the fact that the data searched was much smaller than Google Scholar's; and for broad-topic searches and "other" searches (not one of these three types), Google Scholar's MAP was better (Table 1).

Table 1. Baseline mean average precision scores for Primo and Google Scholar results

Query type	Primo MAP	Google Scholar MAP
Broad-topic	0.31	0.72
Narrow-topic	0.52	0.37
Known-item	0.52	0.67
Other	0.36	0.70

An analysis of evaluators' reasons for disqualifying results from inclusion on the first result page showed that in 29% of the cases, the journal in which the paper was published was deemed insignificant; in 20% of the cases, the evaluator considered the author insignificant; and in 15% of the cases, the material (typically an article) was too specific. These reasons do not apply to known-item searches, in which users want a specific item, regardless of the prominence of the author or journal. In exploratory searches, which can yield a very large number of results, evaluators expected results of a general nature, such as the most important review articles on a topic from top journals and leading researchers. Evaluators disqualified 7% of the items for being too old.

The queries from the initial tests were stored with the researchers' evaluations and later served as queries for testing new enhancements to the Primo relevance ranking.

4.4 Optimizing the Basic Relevance-Ranking Technology

Before introducing new factors, the project team set out to improve the relevance-ranking technology available in March 2011. This technology used traditional ranking methods that had been adapted to the scholarly domain. One method, for example, was the weighting of metadata fields according to the significance of the bibliographic detail stored in them; hence, query terms in a subject field, for instance, contributed more value to a specific item's ranking than the terms' occurrence in the full text.

On the basis of the evaluators' input and customer feedback on the system, the team leveraged the built-in boosting mechanism of Primo: using a tool developed for the testing, the team changed the boosting factors and then monitored the impact. For example, although journal articles were already more prominent than other material types, the team decided to boost these articles even more. The resulting MAP and MRR values peaked when the boosting factor was roughly 1.5 (Fig. 4), indicating that 1.5 is the optimal boosting factor for this material type.

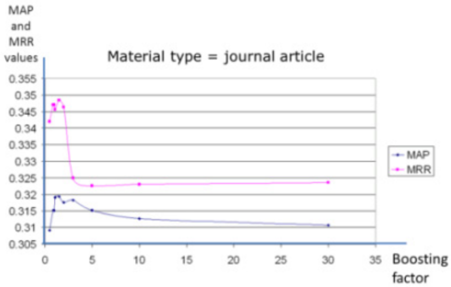


Fig. 4. MAP and MRR values after each change in the boosting factor for the journal-article material type

4.5 Adding a Value Score

After optimizing the technology, the team decided to introduce the first new factor. As a result of the interviews and the evaluators' responses, the team realized that the scholarly significance of an item should play a role in determining its position in the result list. This new factor, the item's *value score,* can be compared to Google Page-Rank in that both are attributes of an "item" (a Web site in Google searches) and are independent of a given query.

The team determined that value scores would be based on various types of information, including the number of citations and usage data from the Ex Libris bX article recommender database.[4] After incorporating value scores in the relevance ranking, the team recalculated the MAP and MRR values for the test queries and found that, indeed, the values went up, indicating that the addition of a value score improved the relevance ranking and that the "best" items were now located farther up the result list.

In the future, the team plans to enrich value scores with other types of information, such as the scholarly impact of a journal or an author.

4.6 Monitoring the Changes

Primo 3.1, released in June 2011, incorporated these changes in the relevance-ranking method. The project team monitored the impact of the changes on users' information-seeking patterns, examining the following factors in particular:

- The mean number of times per session that a user moved to the next page of results. The team assumed that users would typically go to the next page only when their information need was not satisfied by the items already displayed.
- The percentage of sessions that culminated in the selection of an item[5]
- The mean number of seconds that elapsed from the beginning of a session until an item was selected
- The mean position of the selected items in the result list

[4] See `http://www.exlibrisgroup.com/category/bXRecommender`.

[5] A selection, in this context, is the user's explicit request to obtain an item.

4.7 Gathering More Data

Before moving on to further improvements, the project team extended the corpus of test queries by engaging more users—graduate students recruited from several universities and various academic fields. This time, the test environment enabled the evaluators to access the hundreds of millions of records covered by the Primo Central Index (as opposed to the 40 million records available for the first evaluators). In addition, the Primo interface in the test environment was enhanced with thumbs-up and thumbs-down buttons next to each of the first 20 results (Fig. 5). The testers used the buttons to indicate whether an item should or should not be located on the first result page. The system logged each query, its first 20 results, and the users' feedback.

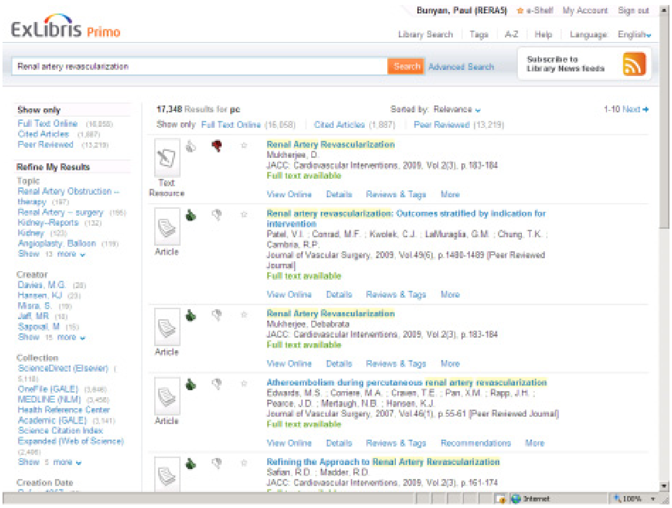


Fig. 5. New test environment for providing feedback on the relevance of materials displayed on the first page of results

This test ran for nine weeks beginning in August 2011, and 560 valid queries with evaluations were logged. The mean MAP and MRR values of all the queries (0.69 and 0.79, respectively) were much higher than the values from the initial testing phase (Fig. 6). The improvement was attributed primarily to the relevance-ranking technology's enhancements and the availability of the full scope of the index for searching. However, the fact that the evaluators were graduate students rather than senior researchers is likely to have played a role as well: graduate students' information needs probably differ from those of senior researchers, and the students are perhaps less familiar with the available content than more experienced researchers are. In addition, graduate students might be less confident in their understanding of what should and should not be on the first page of results. Indeed, Anderson points out that "the evaluation of relevance is thereby linked to the measure of uncertainty…Previous knowledge (or lack of knowledge) is indispensable in assessments of relevance" [11].

121	CO2 dynamics	0.816667	1	Graduate	Physics, Astronomy, ar
122	silicon fractionation	0.494946	0.5	Graduate	Physics, Astronomy, ar
123	silicon isotopes fractionation	0.717248	1	Graduate	Physics, Astronomy, ar
124	phreeqc	0.52243	1	Graduate	Physics, Astronomy, ar
125	quartz kinetics	0.545673	1	Graduate	Physics, Astronomy, ar
126	albite dissolution kinetics	0.889727	1	Graduate	Physics, Astronomy, ar
127	kaolinite precipitation kinetics	0.630305	1	Graduate	Physics, Astronomy, ar
128	Gypsum Nucleation	0.756605	1	Graduate	Geology
129	Gypsum Nucleation Dead sea	0.601651	1	Graduate	Geology
130	Gypsum crystal grows	0.71622	1	Graduate	Geology
131	Gypsum Kinetics	0.633892	0.5	Graduate	Geology
132	red sea dead sea conduit	0.700735	1	Graduate	Geology
133	gypsum preciptation potential	0.833333	1	Graduate	Geology
134	Microbial ecology Dead Sea	1	1	Graduate	Geology
135	rate law for gypsum nucleation	0.868034	1	Graduate	Geology
136	radium	1	1	Graduate	Biology, Life Sciences,
137	radium AND barium	0.67006	1	Graduate	Biology, Life Sciences,
138	radium AND barium co-precipitation	0.721398	1	Graduate	Biology, Life Sciences,
139	activity coefficient Pitzer	0.874559	1	Graduate	Biology, Life Sciences,
140	activity coefficient Pitzer AND radium	1	1	Graduate	Biology, Life Sciences,
141	activity coefficient Pitzer AND barite	0.738095	1	Graduate	Biology, Life Sciences,
142	activity coefficient Pitzer AND gypsum	0.421775	0.5	Graduate	Biology, Life Sciences,
143	solid solution	0.227273	0.25	Graduate	Biology, Life Sciences,
144	solid solution AND barite	0.826356	1	Graduate	Biology, Life Sciences,

Fig. 6. A sample of MAP (second column) and MRR (third column) values calculated for the results displayed on the first page, as evaluated by graduate students

4.8 Adding the User to the Equation

The next phase focused on adjusting the relevance-ranking calculations to suit certain characteristics of the individual and his or her specific information need as evidenced in the query. Primo 4.0 (released in June 2012) incorporated these adjustments.

Broad-Topic Queries. According to an analysis of search logs, 25% to 30% of the queries submitted by Primo users are general in nature or applicable to many areas. Search logs have yielded many examples of broad-topic queries, such as *korean poetry*; *buddhist art*; *adventure education*; *stereoisomerism*; *social mobility*; *mining engineering*; *operator theory*; *sensitivity*; and *tetrodotoxin*.

When users enter a broad-topic query, they do not have a specific document in mind and usually receive a large result set. The project team's assumption was that users invoke a broad-topic query when they want to explore an area with which they are not very familiar—for example, undergraduates who are seeking material for a class assignment about a general topic or researchers who are looking for information related to a discipline with which they are not familiar.[6]

The team also assumed that for broad-topic queries, users would prefer results that provide general information. For example, when a user submits the query *operator theory*, the most appropriate results would be documents that explain what operator theory is, such as reference materials or review articles, rather than articles addressing specific issues related to operator theory.

To improve the ranking of results from broad-topic queries, the team made several adaptations to the relevance-ranking calculations, including offline processing, ways of identifying broad-topic queries in real time, and the boosting of materials that are more likely to address the specific needs of users who formulate this type of query.

[6] According to a survey by University of Minnesota Libraries, the types of queries that undergraduates submit differ significantly from queries that graduate students submit: over 70% of undergraduate searches are exploratory, but as researchers become more knowledgeable in their area, they tend to search more for specific items [12].

Author-Related Queries. According to search logs, about 10% of Primo queries consist of author names (sometimes with other words) in the default search box (as opposed to names that were typed in the author field). Search logs have yielded many examples of what appear to be author-related queries, such as *wescott, D and oil and australia*; *Peterson, T (2004)*; *walt whitman leaves of grass*; *"Nam Soon Huh"*; *Text types adults Nunan and Lamb*; and *Nemes & Coss Effective Legal Research*.

As with broad-topic queries, the team adapted the relevance-ranking calculations to improve the handling of author-related queries. These adaptations include offline processing, methods of identifying author names in real time, and the rephrasing of queries to better address the presumed intention of the user.

Personalized Ranking. Personalized ranking shows much promise and can be addressed on several levels. By deploying such ranking, a discovery system may be able to minimize the number of irrelevant results stemming from topic ambiguity; furthermore, the system can adjust the type of materials to the user's level of expertise. In interdisciplinary research specifically, any division of the search scope by formal disciplinary boundaries may prevent successful discovery; however, an all-encompassing system that tailors the results to the particular user will improve the likelihood of a successful outcome.

Today, the Primo technology takes into account a user's academic degree and discipline (without identifying the person). Users decide whether they wish to provide these details, and if so, the system keeps the information in the users' personal profile, where they can modify it.

Identifying a user's academic degree enables Primo to provide the most suitable results for a topic search. Undergraduates probably need general materials, whereas experienced researchers are likely to seek more in-depth publications. Therefore, on the basis of a user's degree, Primo boosts specific material types for topic searches.

With information about a user's discipline, Primo can boost items in that area—a particularly important feature when search terms apply to several areas. For example, unless the user's discipline is factored into the ranking calculation, the query *memory efficiency* yields results in engineering, medicine, psychology, and other areas. Similarly, when an author-related query contains a common name, the relevance-ranking technology can boost items from authors who publish in the user's field.

5 Conclusions

The relevance-ranking project has demonstrated the feasibility of a methodic approach to developing and testing methods of relevance ranking in a scholarly information system. Although the relevance of a document is always a function of a specific user at a certain point in time, some aspects of relevance can be generalized and deployed in an automated system. The project described here shows that when relevance ranking incorporates the scholarly significance of items, assumptions about a user's type of information need as inferred from the query, and a correlation between a user's academic area and the topics of available materials, the probability that the most relevant items will be displayed at the top of the result list increases.

Nevertheless, the calculation of relevance remains an area that requires constant rethinking, monitoring, and tuning. We can only agree with Saracevic's remark that "there were, still are, and always will be many problems with relevance. This is not surprising. Relevance is a human—not a systems—notion and human notions are complex, even messy. Oh well, they are human" [7].

References

1. Centre for Information Behaviour and the Evaluation of Research (CIBER): Information Behaviour of the Researcher of the Future (2008)
2. OCLC Online Computer Library Center: Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. OCLC, Dublin, Ohio (2005)
3. OCLC Online Computer Library Center: College Students' Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. OCLC, Dublin, Ohio (2006)
4. Sadeh, T.: Multiple Dimensions of Search Results. Paper given at Analogous Spaces Interdisciplinary Conference. Ghent University, Belgium (2008)
5. Hearst, M.A.: Clustering versus Faceted Categories for Information Exploration. Communications of the ACM 49(4), 59–61 (2006)
6. Saracevic, T.: Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. J. Am. Soc. Inf. Sci. Technol. 26(6), 321–343 (1975)
7. Saracevic, T.: Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part II: Nature and Manifestations of Relevance. J. Am. Soc. Inf. Sci. Technol. 58(13), 1915–1933 (2007)
8. Maglaughlin, K.L., Sonnenwald, D.H.: User Perspectives on Relevance Criteria: A Comparison among Relevant, Partially Relevant, and Not-Relevant Judgments. J. Am. Soc. Inf. Sci. Technol. 53(5), 327–342 (2002)
9. Markey, K.: The Online Library Catalog: Paradise Lost and Paradise Regained? D-Lib Magazine 13(1/2) (2007)
10. Markey, K.: Twenty-five Years of End-User Searching, Part 2: Future Research Directions. J. Am. Soc. Inf. Sci. Technol. 58(8), 1123–1130 (2007)
11. Anderson, T.D.: Uncertainty in Action: Observing Information Seeking within the Creative Processes of Scholarly Research. IR Information Research 12(1) (2006)
12. Hanson, C., Hessel, H., Boudewyns, D., Fransen, J., Friedman-Shedlov, L., Hearn, S., Herther, N., Theis-Mahon, N., Morris, D., Traill, S., West, A.: Discoverability: Phase 2. Final Report, University of Minnesota Libraries (2010)