



Article

The Usefulness of Related Functions in Web of Science and Scopus

Dorothy C.P. Char
Public Works and Government Services Canada
Gatineau, Quebec, Canada
Email: Dorothy.Char@tpsgc-pwgsc.gc.ca

Isola Ajiferuke
Faculty of Information and Media Studies, University of Western Ontario
London, Ontario, Canada
Email: iajiferu@uwo.ca

Received: 18 Nov. 2010

Accepted: 19 Feb. 2011

© 2011 Char and Ajiferuke. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 2.5 Canada (<http://creativecommons.org/licenses/by-nc-sa/2.5/ca/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Objective – This study evaluates the effectiveness of the related search functions in Web of Science and Scopus. Web of Science has one related search function (searching by references) whereas Scopus has three related search functions (searching by references, authors, or keywords).

Methods – Thirty queries were searched in both Web of Science and Scopus. For each query, the most relevant document was retrieved and its first thirty related documents were assessed for relevancy to the original query. Results for both databases were compared using the median values of precision. For Scopus the three different methods of relevance were compared using median precision values.

Results – The median precision value for the related documents retrieved from Web of Science was 0.63, while the median for those retrieved from Scopus using the related by references function was 0.62. A Wilcoxon test showed no significant difference in the two medians. In the comparison of the three related functions in Scopus, the median precision values were 0.62, 0.42, and 0.43 for the related search functions by references, authors, and keywords respectively. A Friedman's test showed that the median precision value for relatedness by references was significantly higher than the median values for the other two related functions. In Scopus, the

effectiveness of the related search function using all keywords was not as effective when compared to the effectiveness using select keywords. The median precision value with select keywords was 0.17.

Conclusions – The related search functions by references for both Web of Science and Scopus were moderately effective in retrieving additional relevant documents on a given topic, and there was no significant difference in their performance. When comparing the three methods available in Scopus, the related search function by references was found to be more effective than the system's related functions by authors and keywords.

Introduction

In online searching, the size of the initial search output may be too few, just adequate, or too much. To address the quantity of output problems, searchers often adopt tactics to narrow or broaden search results. However, even when the quantity of output is acceptable, the examination of the retrieved documents may indicate deficiency in terms of quality, i.e., the retrieved set may contain too many irrelevant documents. To address this, various retrieval refinement methods have been suggested (Salton & McGill, 1983). These include relevance feedback, clustering, and citation searching. Over the years, online retrieval systems have provided practical ways of conducting these refinements. In the case of relevance feedback, the systems allow users to view the terms (whether keywords, identifiers, or descriptors) that have been used to index a document, and by clicking on a particular term for a relevant document, other documents that have been indexed by that term can be retrieved.

Clustering of search output is common with clustering search engines, such as Yippy, though traditional retrieval systems such as Scopus and Web of Science now provide categorization of search output by subject areas. In the case of citation searching, many of the retrieval systems allow searchers to view either the list of references or citations referencing a selected document. However, due to the limitations of citing practices, such as honorific or perfunctory citations, citing methodological procedures or sources of data, or negative citations (Libmann, 2007), some of

these references or citations may not be relevant to the original query. Others have suggested that the use of bibliographic coupling or co-citation might produce better results (Bichteler & Eaton, 1980; Badran, 1984). Bibliographic coupling is the act of relating two documents together based on having a certain number of common references, while co-citation is based on having a certain number of common citations (Wissmann, 1993). While no retrieval system has provided a means of retrieving documents which have a certain number of common citations within a selected document, Web of Science and Scopus allow users to view documents related to a selected document by common references. In the case of Scopus, documents can also be related by authors or by common keywords (similar to functions provided by Google or PubMed).

Yet, the question remains, how relevant to the original query are additional documents produced through bibliographic coupling? The objective of this study is to answer this question by determining (and comparing) the effectiveness of the related functions of Web of Science and Scopus. Given that Scopus provides three different functions for relating documents (common references, common authors, and keywords), these methods will be compared against each other. From a practical standpoint, this information will provide searchers with knowledge of the most useful means of obtaining additional relevant documents on a given topic.

Literature Review

Studies comparing Web of Science and Scopus can be categorized into two groups, namely those comparing their searching features and those comparing their capabilities for conducting informetric studies. In the first category, LaGuardia (2005) asserted that Scopus is more suited for subject searching than Web of Science. In terms of subject coverage, Scopus covered the sciences better than Web of Science, while Web of Science covered arts and humanities and social sciences better than Scopus.

In his review of the two systems Jacso (2005) showed that 77% of the Web of Science records were in the Science Citation database, 14% in the Social Sciences Citation database, and 9% in the Arts and Humanities Citation database. In Scopus, 60% of the records were in the health and life science subject areas, 25% in the chemistry, physics, math and engineering subject areas, 13 % in the biological, agricultural, earth and environmental science subject areas, and 2% in social science, psychology and economics subject areas. So while Scopus covers science better than Web of Science, Web of Science's coverage of science is better than its coverage of arts and humanities. In their study of the coverage overlap between the two systems, Gavel and Iselid (2008) found that Web of Science included 8,901 journal titles in its databases, while Scopus had 13,690 journal titles. There were 7,434 titles indexed by both databases.

In their citation analysis of medical documents, a number of studies have found Scopus retrieved more citations than Web of Science. For example, Falagas, Pitsourni, Malietzis, and Pappas (2008) found that Scopus listed 20% more citations than Web of Science for some biomedical articles. Torres-Salinas, Lopez-Cozar, and Jimenez-Contreras (2009) found that articles in the area of the health sciences authored by researchers at the University of Navarra (Spain) received 14.7% more citations in Scopus than in the Web of Science. Kulkarni, Aziz, Shams, and Busse

(2009) found that for articles published in general medical journals, Scopus retrieved more citations per article (median=149) than did Web of Science (median =122).

In their comparison of the suitability of Scopus and Web of Science for bibliometric analyses, Gorraiz and Schloegl (2008) found that of the top 100 pharmacy journals listed by ISI's *Journal Citation Reports -Science*, the impact factor was higher for 82 titles, and the immediacy index was greater for 78 journals in Scopus in 2005. These results are not totally surprising given Scopus' wider coverage of the health and life science subject areas.

Similar results have been obtained in other subject areas. In a study of the impact of 22 top human-computer interaction (HCI) researchers from EQUATOR, a large British Interdisciplinary Research Collaboration project, Meho and Rogers (2008) found that the study sample had been cited in 6,919 papers indexed in Scopus and 4,011 papers indexed in Web of Science. In a similar study of the scholarly impact of 80 randomly selected full-time, information studies faculty members from North America, Meho and Sugimoto (2009) found that the study sample had been cited in 7,103 papers in Scopus and only 5,269 papers in Web of Science.

In a citation analysis of the 25 most accessed articles in 163 social sciences journals, Levine-Clark and Gil (2009) found that the average number of citations received by these articles in Web of Science was 7.95, while the average citation count for the same articles in Scopus was 9.27. The higher number of citations found in Scopus may not necessarily be due to better coverage of a particular subject area, but it may be because Scopus covers more source titles, as discussed by Gavel and Iselid (2008).

In summary, this review shows that Web of Science and Scopus have been examined and compared regarding their subject coverage, reflecting their capabilities as tools for informetric studies. However, no studies have compared them in terms of the effectiveness of their related search functions. This study will

fill that gap in the research literature with the intention of providing useful guidance for users of these databases.

Methods

A total of 30 queries (Appendix A) were searched in both Web of Science and Scopus. The subjects reflected the broad scope of the databases' content. The queries were selected in such a way that 10 of them could be searched in each of the three indexes of Web of Science (i.e., arts and humanities, social sciences, and science), and five in each of the six subject areas of Scopus (i.e., life sciences, physical sciences, health sciences, social sciences, and arts and humanities).

The queries were collected from assignments given in an information retrieval course as well as from the Text Retrieval Conference (TREC) Web site (<http://trec.nist.gov/>). They were reformatted into queries suitable for the selected databases and comprehensible to the researchers. The first author of this paper identified the key concepts of each query and formulated the appropriate search strings for each database. Unless otherwise stated in the query, the search output was not limited by any restrictions (e.g., language, year of publication, or document type).

The results of each query searched in Web of Science were sorted according to relevance. For each query the highest ranked document had the top 30 related articles of each query's top-related references assessed for relevancy to the original query. The relevance ranking of Web of Science (as well as that of Scopus) is based on the number of search terms found in each document record. It is not unusual for a user to find that the topmost ranked document is not the most relevant, or on occasion, not even relevant to his or her query. In this study and for each query, the goal of this research was to find whether the topmost ranked document was relevant to the query.

The number of related documents was limited to 30, as previous studies have shown that most users usually view no more than the top

30 documents retrieved in response to a Web query (Spink & Wolfram, 2001). Researchers recorded both the total number of references, as well as the number of shared references between the related and original document, were recorded. The percentage of common references for each document in each of the 30 queries was computed according to the formula: (the number of common references) divided by [(the number of references in the original document plus the number of references in the related document) minus the number of common references].

The first author of this paper assessed the relevancy of each reference using as either "relevant" or "not relevant." For the queries collected from assignments in an information retrieval course, the relevance judgment was based on familiarity with the queries (from previous searching experience). Guidelines for judging relevance were used for those taken from TREC (Text Retrieval Conference, 2000). For example, for Query #20 (Appendix A), the instruction for judging relevance of a retrieved document stated

"a relevant document will contain an argument for or against requiring students to wear uniforms in schools. Advertisements for uniforms and simple statements that particular school districts require uniforms are not relevant" (Text Retrieval Conference, 2001, Number 533)

The two most commonly used measures of effectiveness of an information retrieval system are recall and precision. Recall is defined as the proportion of relevant items retrieved (i.e., the number of relevant items retrieved divided by the number of relevant items in the database). Precision is defined as the proportion of retrieved items that are relevant (i.e., number of relevant items retrieved divided by the number of retrieved items). However, due to the inherent difficulty in determining the number of relevant items in a large database, precision was the only practicable measure for this study. Hence, the precision value for each of the 30 queries was

computed as the number of relevant related documents divided by the total number of related documents examined. This calculation attempted to determine the precision of the databases in identifying related documents.

Table 1
Precision Values of the Related Documents by Common References Retrieved from Web of Science and Scopus

Query #	Precision for Web of Science	Precision for Scopus
1	.7667	Not available
2	.7333	Not available
3	.8667	.8000
4	.7333	.5000
5	.4667	.3000
6	.9667	1.000
7	.6667	Not available
8	.7667	.6000
9	Not available	Not available
10	.7000	.7667
11	.9333	Not available
12	Not available	.1667
13	.8000	.7000
14	.6000	.6667
15	.3000	Not available
16	.7333	.9000
17	.8333	.7333
18	.3667	.4333
19	.2000	.5000
20	.4333	Not available
21	.1000	.6333
22	.6333	.9000
23	Not available	.5667
24	.2667	.5000
25	.3667	Not available
26	.8000	.3333
27	.6000	Not available
28	.5000	.6667
29	.3667	.1667
30	.0667	Not available

The results of each query searched in Scopus were also sorted by relevance, and the first 30 related articles of the topmost relevant document (either all references, all authors, or all keywords) were assessed for relevancy to the original query. The total number of references, authors, and keywords, as well as

the number of such shared elements, were recorded, in the same manner as the Web of Science data was evaluated. The percentage of common references, common authors, and keywords were also recorded. The relevancy assessment of the related documents was performed using the same method, and the precision values were obtained using the formula referenced above. The option to search for related documents to the topmost ranked document by using some, instead of all, of its keywords was explored; for each topmost ranked document, three to five of its keywords were used to search for related documents. Descriptive statistical methods described the effectiveness of each related function, and inferential statistical tests compared the effectiveness of the functions.

Results

Comparison of Relatedness by Common References in Web of Science and Scopus

The study determined the effectiveness of the related search function for 27 of the queries in Web of Science and 20 queries in Scopus. For the 27 queries in Web of Science that produced related documents, the mean and median precision values for the related documents retrieved were 0.58 and 0.63 respectively. For Scopus, the mean and median precision values were 0.59 and 0.62 respectively (Table 1). The distributions of precision values shown in Table 1 are skewed, which implies that the median is the more appropriate measure of central tendency to be used in comparing the two sets of values. Hence, the Wilcoxon test, a nonparametric statistical test used to compare the median values of two groups, was used to determine if the median precision values were the same. The test yielded a z-score, of -0.218 with the corresponding p-value of 0.828. The p-value indicated no statistical significant difference between the effectiveness of the related search function by references for Web of Science when compared to that of Scopus.

Table 2
 Details of the Related Documents Retrieved for the Topmost Document (With 14 References) for
 Query 1 in Web of Science

Related Document #	Number of References	Number of common references with the topmost document	Percentage of references in common with the topmost document	Relevancy of the related document
1	132	8	5.80	Yes
2	23	7	23.33	Yes
3	110	7	5.98	Yes
4	21	4	12.90	Yes
5	32	3	6.98	Yes
6	997	3	0.30	Yes
7	221	3	1.29	Yes
8	109	3	2.50	Yes
9	16	3	11.11	Yes
10	43	3	5.56	Yes
11	35	3	6.52	Yes
12	70	2	2.44	No
13	53	2	3.08	No
14	48	2	3.33	No
15	29	2	4.88	Yes
16	36	2	4.17	No
17	23	2	5.71	No
18	41	2	3.77	Yes
19	81	2	2.15	Yes
20	36	2	4.17	Yes
21	44	2	3.57	Yes
22	35	2	4.26	No
23	21	2	6.06	Yes
24	33	2	4.44	No
25	51	2	3.17	Yes
26	6	2	11.11	Yes
27	13	2	8.00	Yes
28	15	2	7.41	Yes
29	135	2	1.36	Yes
30	20	2	6.25	Yes

The study attempted to determine the minimum number of references (or percentage of references) that a related document needed to have (relative to the topmost document) for it to be judged relevant. However, as evidenced by the sixth related document in Table 2, for one of the queries searched in Web of Science, there was no discernible pattern. Hence, it was not possible to determine such a number (or percentage) for all the queries in either database.

Comparison of Relatedness by Common References, Keywords and Authors in Scopus

Given that Scopus can obtain related documents through three different search functions, the study compared the effectiveness of these three related functions against each other. The precision values for relatedness by all references were the same as described above. The mean and median values for relatedness by all authors were .41 and .42 respectively, while for relatedness by

keywords, the mean and median values were .44 and .43 respectively (Table 3). In order to statistically compare the three related search functions, the Friedman test, a nonparametric statistical test used to compare the medians of more than two groups, was used to determine if the median precision values for all three functions were the same. The test statistic for this test is the chi-square, and we obtained a value of 6.125 with 2 degrees of freedom and a p-value of 0.047. Hence, we rejected the hypothesis that the three medians were the same. A follow-up multiple comparison test showed that the related search function by all references was more effective than the other two related functions.

Comparison of Relatedness by All Keywords and Select Keywords in Scopus

Each of the three related search functions in Scopus has two options of using all (e.g., all references) or a select few. It might be difficult for a user to determine which references or authors to select, but relatively easy to choose a few keywords related to his or her original query. As a result, the study compared the effectiveness of using all keywords and select keywords. The performance values for "all keywords" were as described above (i.e., mean and median precision values of .41 and .42 respectively), while the mean and median precision values for relatedness by "select keywords" were .25 and .17 respectively (Table 4).

Table 3
Precision Values of Related Documents by References, Authors and All Keywords Retrieved from Scopus

Query #	Precision for all references	Precision for all authors	Precision for all keywords
1	Not available	Not available	.0667
2	Not available	.2333	.3667
3	.8000	.3333	.7667
4	.5000	.4000	.1667
5	.3000	Not available	.1667
6	1.000	.4333	.8667
7	Not available	.3000	.2000
8	.6000	.5667	.3333
9	Not available	Not available	Not available
10	.7667	.4333	.9000
11	Not available	.5333	.1667
12	.1667	.0909	.4783
13	.7000	.1667	.6000
14	.6667	.5667	.4333
15	Not available	.2000	.3000
16	.9000	.3000	.7333
17	.7333	.5333	.5000
18	.4333	.0800	.6000
19	.5000	Not available	Not available
20	Not available	Not available	Not available
21	.6333	.6667	.3333
22	.9000	.5667	.4000
23	.5667	1.000	.6000
24	.5000	Not available	.4333
25	Not available	Not available	Not available
26	.3333	Not available	.6333
27	Not available	Not available	Not available
28	.6667	Not available	.1000
29	.1667	Not available	Not available
30	Not available	Not available	Not available

The distributions in Table 4 were skewed. In order to accurately assess whether the median precision values for the two keywords options were the same, the Wilcoxon test for paired samples provided the needed statistical analysis. The z-score of the test was -3.761 with a p-value of 0.000, which implied a significant statistical difference between the effectiveness of the two related search functions. Hence, the research concluded that the effectiveness of the "all keywords" related function was better than that of "select keywords" in Scopus.

Discussion

The results demonstrate the usefulness of the related search functions in both Web of Science and Scopus. Researchers often need to examine a moderate amount of relevant materials on a topic, especially if they are compiling a bibliography or writing a literature review. This may be equally true for professional librarians, who may be engaged in their own research projects, compiling course bibliographies, or conducting searches on behalf of their patrons.

Table 4
Precision Values of Related Documents by All Keywords and Select Keywords Retrieved from Scopus.

Query #	Precision for "all keywords"	Precision for "select keywords"
1	.0667	.0000
2	.3667	.5000
3	.7667	.4000
4	.1667	.1000
5	.1667	.0000
6	.8667	.7667
7	.2000	.1333
8	.3333	.1000
9	Not available	Not available
10	.9000	.6667
11	.1667	.0000
12	.4783	.0000
13	.6000	.3667
14	.4333	.2333
15	.3000	.4000
16	.7333	.5000
17	.5000	.0000
18	.6000	.1667
19	Not available	Not available
20	Not available	Not available
21	.3333	.1333
22	.4000	.3667
23	.6000	.2333
24	.4333	.1667
25	Not available	Not available
26	.6333	.4333
27	Not available	Not available
28	.1000	.0667
29	Not available	Not available
30	Not available	Not available

The following example demonstrates how the functions work in practice: Search for relevant materials on the applications of tagging in online public catalogs in Web of Science, using “tag* and (OPAC* or online public catalog*)” in the topic field. This search yielded twelve documents, of which three were relevant. The most relevant of those was, “Enhancing library services with Web 2.0 functionalities” by Gavrilis, Kakali, and Papatheodorou (2008). A review of the related records for this article retrieved 371 documents with at least one reference in common. Further analysis identified four relevant documents with three references in common with the target article. Of the 47 documents with two references in common with the target article, 17 were identified as relevant to the original search question. A review of the records of related references to the article by Gavrilis, et al. (2008) retrieved 21 additional relevant documents. The search of the same terms in Scopus yielded five documents, of which only three were included in the list of those found in Web of Science. Of the five documents from Scopus, three were found to be relevant, with the most relevant again being the article by Gavrilis et al. (2008).

The similarity in the effectiveness of the two related search functions by references may be due to the considerable overlap in the subject coverage of the two databases or the similarity in the operationalization of the methods used in selecting the related documents. Even though Scopus does not explicitly state the criterion used in its ranking of related documents, they are likely to be similar to Web of Science’s ranking methods, that is to place, in decreasing order, the number of common references the related documents have with the document already identified.

In Scopus, there are three types of related functions, while only one is available in Web of Science. For Scopus there are related search functions by references, authors, and keywords. Under each related function type, the user can select the option of using all attributes from the article (e.g., all references) or the user can choose to use only a selected

number of attributes (e.g., a selected number of references). This study investigated the effectiveness of using all attributes for all three types, but used only a selected number of attributes for keywords. The reason for this was that while there is often not a clear criterion to use in selecting from the list of reference or authors, it would be possible to select keywords related to the original topic of inquiry. The study determined that the effectiveness of using select keywords was much worse than when using all keywords. The recommendation would be to discourage users from making use of the select keywords for the related search function.

Of the three related search functions, the most effective proved to be the search by all references. This may be due to the fact that while two documents may have the same author, the subject areas or topics that they cover may be different. Also, two documents may be indexed by the same keyword but with varying depth of treatment of the keyword in the two documents.

For Scopus the result of this research indicates that the user should try the related search function by all references first, and should only explore the other two types afterwards if there is still a need for more relevant documents. Documents obtained using the related search function based on the Gavrilis et al., (2008) article and using all three search functions (i.e., all references, all authors, and all keywords) demonstrated this. Results yielded 710 documents that shared at least one common reference, 42 documents that shared at least one author, and 4,452,046 documents that shared at least one keyword with the target article (Gavrilis et al., 2008). The 710 documents retrieved by all references were sorted by relevance (and not by the number of common references as with Web of Science). Only 20 of the first 60 documents were determined to be relevant. Of the 42 documents retrieved by all authors, only one was found relevant. Finally, when the 4,452,046 documents retrieved by all keywords were ranked by relevance, only 10 of the first 60 documents were determined to be relevant.

There were several limitations of this study. First, the queries were not obtained from users with real information needs, and users were not involved in the assessment of the relevance of the related documents retrieved. Second, some top-ranked documents did not produce any related documents, because they either had either too few references, or too few authors, or too few keywords. Third, a few of the top-ranked documents did not have detailed records in Scopus, so users were redirected to another database, CSA Illumina. For such documents, it was not possible to obtain their related documents from Scopus. Due to the last two limitations, it was not possible to determine the minimum number of references that a related document needed to share with an already identified relevant document for it also to be judged relevant.

Conclusions

The main objective of this study was to determine the usefulness of the related search functions in Web of Science and Scopus with regard to finding additional relevant documents on a given topic. The related search functions by references for both Web of Science and Scopus proved to be moderately effective in retrieving additional relevant documents, and there was no significant difference in their effectiveness. Users who want to obtain additional documents relevant to a particular query, should do so by retrieving documents related by common references to an already identified relevant document in either Web of Science or Scopus. When comparing the related search functions within Scopus, the related by references search function was found to be more effective than either the related function by author or the related function by keyword. If using Scopus alone, the relevant by references function was more useful than relevant by either author or keyword. Future research will use a higher number of queries to further explore the possibility of determining the minimum number of references that a related document needs to share with an already identified relevant document to a query for it also to be judged relevant.

References

- Badrn, O. M. (1984). Alternative search strategy to improve information retrieval. *Proceedings of the ASIS Annual Meeting, 21*, 137-140.
- Bichteler, J., & Eaton III, E. A. (July 1980). The combined use of bibliographic coupling and cocitation for document retrieval. *Journal of the American Society for Information Science, 31*(4), 278-282.
- Falagas, M. E., Pitsouni E. I., Malietzis, G. A., & Pappas, G. (Feb. 2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB Journal, 22*(2), 338-342. doi: 10.1096/fj.07-9492LSF
- Gavel, Y., & Iselid, L. (Feb. 2008). Web of Science and Scopus: a journal title overlap study. *Online Information Review, 32*(1), 8-21. doi: 10.1108/14684520810865958
- Gavrilis, D., Kakali, C., & Papatheodorou, C. (2008). Enhancing library services with Web 2.0 functionalities. *Lecture Notes in Computer Science, 5173/2008*: 148-159. Research and Advanced Technology for Digital Libraries. ECDL '08 Proceedings of the 12th European conference on Research and Advanced Technology for Digital Libraries, Aarhus, Denmark. Sept. 14-19 2008. doi: 10.1007/978-3-540-87599-4_16
- Gorraiz, J., & Schloegl, C. (Oct. 2008). A bibliometric analysis of pharmacology and pharmacy journals: Scopus versus Web of Science. *Journal of Information Science, 34*(5), 715-725. doi: 10.1177/016555150708669991
- Jacso, P. (Nov./Dec. 2004). ISI Web of Science, Scopus, and SPORTDiscus. *Online, 28*(6), 51-54.
- Jacso, P. (10 Nov. 2005). As we may search – comparison of major features of the

- Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537-1547.
- Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (9 Sept. 2009). *JAMA*, 302(10), 1092-1096. doi: 10.1001/jama.2009.1307
- Kumar, M. J. (2009). Evaluating scientists: citations, impact factor, h-index, one page hits and what else? *IETE Technical Review*, 26(3), 165-168.
- LaGuardia, C. (Jan. 2005). E-Views and Reviews: Scopus vs. Web of Science. *Library Journal*, 130(1), 40-42.
- Levine-Clark, M., & Gil, E. (2009). A comparative analysis of social science citation tools. *Online Information Review*, 33(5), 986-996. doi: 10.1108/14684520911001954
- Libmann, F. (May/June 2007). Web of Science, Scopus, and classical online: philosophies of searching. *Online*, 31(3), 36-40.
- López-Illescas, C., de Moya Anegón, F., & Moed, H. F. (April 2009). Comparing bibliometric country-by-country rankings derived from the Web of Science and Scopus: the effect of poorly cited journals in oncology. *Journal of Information Science*, 35(2), 244-256. doi: 10.1177.165551508098603
- Meho, L. I., & Rogers, Y. (Sept. 2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11), 1711-1726.
- Meho, L. I., & Sugimoto, C. R. (Dec. 2009). Assessing the scholarly impact of information studies: a tale of two citation databases – Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 60(12), 2499-2508. doi: 10.1002/asi.21165
- Meho, L. I. & Yang, K. (Nov. 2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125. doi: 10.1002/1si.20677
- Morris, S. A. (2005). Manifestation of emerging specialties in journal literature: a growth model of papers, references, exemplars, bibliographic coupling, cocitation, and clustering coefficient distribution. *Journal of the American Society for Information Science and Technology*, 56(12), 1250-1273.
- Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill, Inc.
- Spink, A., & Wolfram, D. (Feb. 2001). Searching the Web: the public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226-234.
- Text Retrieval Conference. (2000). Text Retrieval Conference (TREC). Retrieved from <http://trec.nist.gov/>
- Text Retrieval Conference. (2001). TREC 2001 web ad hoc topics. In *Data – English Test Questions (Topics) Files List*. Retrieved from http://trec.nist.gov/data/topics_eng/topics.501-550.txt
- Torres-Salinas, D., Lopez-Cozar, E. D., & Jimenez-Contreras, E. (2009). Ranking of departments and researchers within a university using two different databases: Web of Science versus Scopus. *Scientometrics*, 80(3), 761-774. doi: 10.1007/S11192.008.2113.9

Wissmann, C. (1993). Techniques of data retrieval for scientometric research in the ISI citation indexes. *Journal of Information Science*, 19(5), 363-376.
doi:10.1177/016555159301900504

Zhao D., & Strotmann A. (2008). Evolution of research activities and intellectual influences in information science 1996-2005: introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, 59(13), 2070-2086.

Appendix A: List of Queries

Query no.	Query
1	What is the eating behaviour of ladybugs in North America? Please find documents in English.
2	I am looking for English documents that discuss the way in which radiation treats Burkitt Lymphoma.
3	Documents since 2005 that discuss photosynthesis for evergreen trees.
4	I am looking for English documents that discuss the evolutionary history of sharks.
5	I am looking for English documents that discuss the differences between Darwin's theory of evolution with theories of creationism.
6	What is the impact of farm raising salmon on ecosystem health?
7	What is the drainage basin of the Danube, and through which countries does it flow?
8	I am interested in the vesuvius: what were the causes of its past eruptions, and how was it formed?
9	I am interested in documents that discuss the predictability of astrology on human relationships.
10	What do recent documents, since 2000, state regarding the composition of the earth's crust?
11	What are sub-atomic particles, and how do they impact energy?
12	Find documents in English that discuss Rutherford's contribution to physics and his work with Niels Bohr.
13	What are the pros and cons of stretching versus not stretching before exercise. I am looking for documents since 2000 in English.
14	What is the impact of smoking on the prevalence of heart attacks in Canadian women?
15	I am interested in documents that discuss how to prevent osteoporosis.
16	I would like to know the effects of decaffeinated and regular coffee on diabetes.
17	What is the relationship between pain, psychological distress, and demographic variables at the end of life?
18	I am looking for documents that discuss the overall dental health of female adolescents, since 2005.
19	I am looking for documents that discuss the advantages or disadvantages of placing both male and female youths in the same residential units in correctional centres.
20	I am interested in documents that discuss the pros and cons of students wearing a school uniform or adhering to a dress code. The documents should be in English.

21	Find English documents that discuss the barriers to women's participation in sporting activities in Muslim countries.
22	I am interested in English documents that compare the reasons why men have eating disorders with the reasons why women have such disorders.
23	What factors impede the integration of immigrants in Canada? Documents from last 5 years.
24	Find English documents that discuss government social programs for immigrants in only Canada. The documents must have been published since 2000.
25	I am looking for information on Lacan and his writings on literary theory, specifically, those that mention Freud.
26	I am interested in the origins of Hinduism and what current research states about the religion. I would like documents in English.
27	Find documents that discuss the history of the English language.
28	I am looking for documents in gender theory that discuss performativity in English literature.
29	I am interested in English documents that discuss Wordsworth's use of pastoral imagery in his Romantic poems.
30	Find documents in English that discuss women in both Muslim and Christian faith.