



Date : 27/07/2006

Multi-faceted Approach to Citation-based Quality Assessment for Knowledge Management

**Lokman I. Meho and
Kiduk Yang**

School of Library and Information Science
Indiana University
Bloomington, Indiana, USA

Meeting:	146 Knowledge Management with Statistics and Evaluation
Simultaneous Interpretation:	Yes

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL

20-24 August 2006, Seoul, Korea

<http://www.ifla.org/IV/ifla72/index.htm>

Abstract:

One of the key tasks of knowledge management (KM) is to assess the quality of information. Before we transform information to knowledge by knowledge representation and organization, we must first identify quality information in a given knowledge domain. To assess the influence and quality of a scholarly publication, an author, or a journal, for example, citation-based evaluation methods are often employed. The typical citation analysis, however, suffers from two fundamental shortcomings. First, conventional citation analysis methods yield one-dimensional and sometimes misleading evaluation as a result of not taking into account differences in citation quality, not filtering out citation noise such as self-citations, and not considering non-numeric aspects of citations such as language, culture, and time. Second, the coverage of citations in citation databases of today is disjoint and incomplete, which can result in conflicting quality assessment outcome across different data sources. To address these limitations, we are developing a multi-faceted approach to information quality assessment that employs a range of citation-based methods to analyze data from multiple sources. The paper gives a brief overview of a work-in-progress prototype system called CiteSearch, which analyzes combined data from multiple citation databases to produce citation-based quality evaluation measures, and discusses a citation analysis pilot study, which measures the impact of scholarly publications based on the data mined from Scopus and Google Scholar.

1. Introduction

One of the key tasks of knowledge management is to assess the quality of information. Before we transform information to knowledge by knowledge representation and organization, we must first identify quality information in a given knowledge domain. To assess the influence and quality of a scholarly publication, an author, or a journal, for example, citation-based evaluation methods are often employed. The typical citation analysis, however, suffers from two fundamental shortcomings. First, conventional citation analysis methods yield one-dimensional and sometimes misleading evaluation as a result of not taking into account differences in citation quality, not filtering out citation noise such as self-citations, and not considering non-numeric aspects of citations such as language, culture, and time. Second, the coverage of citations in citation databases of today is disjoint and incomplete, which can result in conflicting quality assessment outcome across different data sources.

To address these limitations, we are developing a multi-faceted approach to information quality assessment that employs a range of citation-based methods to analyze data from multiple sources. The paper briefly describes a work-in-progress prototype system called *CiteSearch*, which will analyze combined data from multiple citation databases to produce citation-based quality evaluation measures such as CiteRank, H-Index, and Mentor-Index, and discusses a citation analysis pilot study, which measures the impact of scholarly publications based on the data mined from Scopus and Google Scholar.

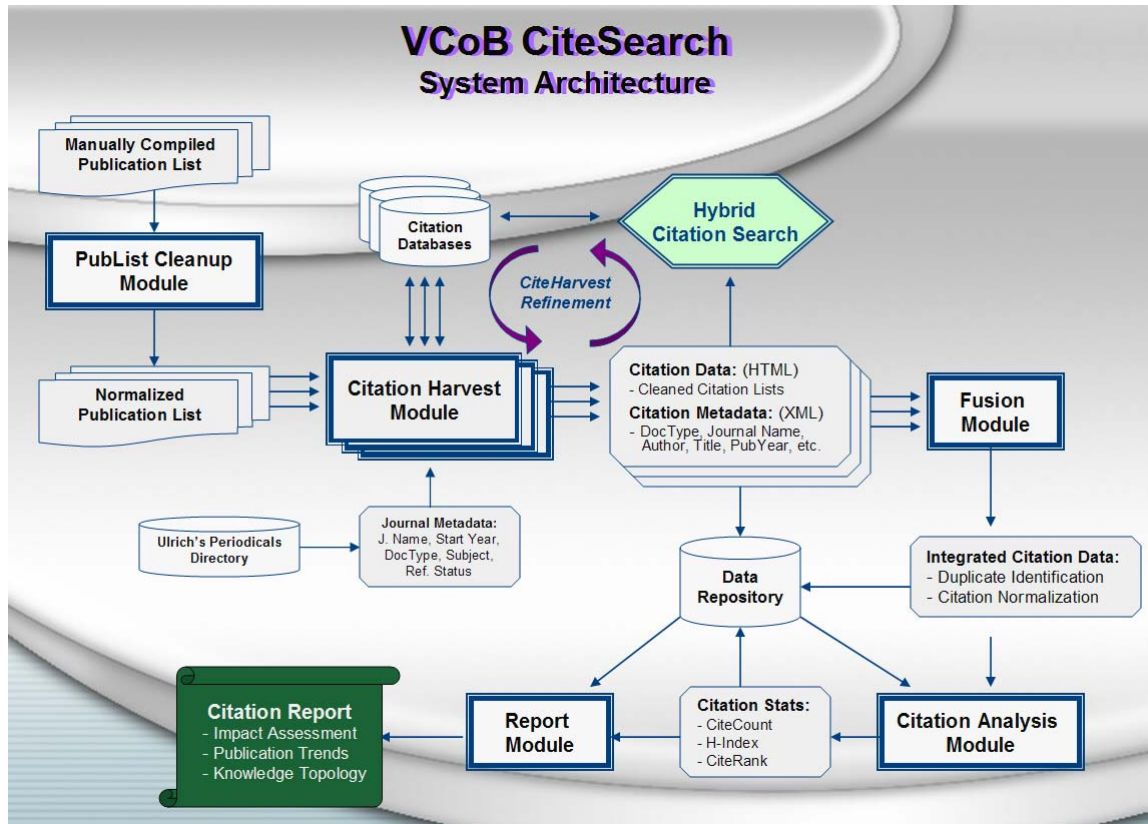
2. CiteSearch System

The CiteSearch system, which is being developed by the VCoB project¹, is a Web-based citation search and analysis system that facilitates the citation-based assessment of information by extracting and analyzing citation metadata from multiple citation databases. The implementation of CiteSearch prototype is currently under way with the target completion date of January 2007, so what follows is a general description and brief overview of the system design.

Given a publication title, for example, the *CiteSearch* system will automatically search multiple Web-based citation databases such as *Google Scholar* and *Google Book Search* and analyze the search results to produce bibliographical metadata of all citations and compute various citation-based quality evaluation measures such as CiteRank, which is a citation propagation measure similar to PageRank, and weighted CiteRank, which is CiteRank weighted by source, author, or time of citations. As a pilot study, we are using complete publication lists of full-time faculty members randomly selected from American Library Association-accredited library and information science programs to generate citation metadata with CiteSearch system. The initial citation metadata will then be aggregated and analyzed to produce meta-level citation measures for authors, publications, and schools. In addition to CiteRank, the meta-level citation analysis will compute the H-Index, an index developed by Hirsch to quantify an individual's scientific research output, as well as the Mentor-Index, an index that measures the mentoring impact by the research impact or performance of students produced. Figure 1 displays the overview of the CiteSearch system architecture.

¹ The Virtual Collection Builder (VCoB) is one of the research project undertaken by the Web Information Discovery Integrated Tool (WIDIT) Laboratory (<http://elvis.slis.indiana.edu/>) at Indiana University School of Library and Information Science. The aim of the VCoB project is to develop an adaptive, interactive agent for building and maintaining a virtual collection of Web documents.

Figure 1. CiteSearch Prototype System Architecture



3. CiteSearch Study

In order to explore the existing citation analysis environment for the CiteSearch system, we conducted a pilot study to examine the impact of using *Scopus* and *Google Scholar* on the citation count and citation ranking of LIS faculty members as measured by *Web of Science*. *Web of Science*, which comprises the three ISI citation databases (*Arts & Humanities Citation Index*, *Science Citation Index*, and *Social Sciences Citation Index*), has been the standard tool for a significant portion of citation studies worldwide. A simple keyword search in *Web of Science* and other databases (e.g., *Pascal*, *Medline*, *EMBASE*, *Biosis Previews*, and *INSPEC*) indicates that ISI databases have been used, or referred to, in several thousand journal articles, conference papers, and chapters in books in the last three decades.

Web of Science's website provides substantial factual information about the database, including the number of records and the list of titles indexed. It also offers powerful features for browsing, searching, sorting and saving functions, as well as exporting to citation management software (e.g., *EndNote* and *RefWorks*). Coverage in *Web of Science* goes back to 1945 for *Science Citation Index*, 1956 for *Social Sciences Citation Index*, and 1975 for *Arts & Humanities Citation Index*. As of February 2006, there were over 35 million records in the database from approximately "8,700" scholarly titles, including approximately 900 conference proceedings and several hundred trade publications and open access journals (Thomson, 2006).² Subjects covered in *Web of Science* include disciplines found in the curricula of most universities in arts,

²The number of conference proceedings was generated by identifying the number of indexed titles that included the keyword conference, proceedings, symposium, or workshop. The titles were browsed to exclude journals from the count. The number of trade publications and open access journals are author estimates.

humanities, sciences, and social sciences. For more details on *Web of Science*, see Goodman and Deis (2005) and Jacso (2005a).

Similar to ISI, Elsevier, the producer of *Scopus*, provides substantial factual information about the database, including the number of records and the list of titles indexed. It also offers powerful features for browsing, searching, sorting, and saving functions, as well as exporting to citation management software. Coverage in *Scopus* goes back to 1966 for bibliographic records and abstracts and 1996 for citations. As of February 2006, there were over 27 million records in the database from over 15,000 “peer-reviewed” titles including 535 Open Access journals, 750 conference proceedings, and 600 trade publications (Elsevier, 2006). Subject areas covered in *Scopus* include: Chemistry, Physics, Mathematics, and Engineering (4,500 titles), Life and Health Sciences (5,900 titles and 100% Medline coverage), Social Sciences, Psychology, and Economics (2,700 titles), Biological, Agricultural, and Environmental Sciences (2,500 titles), and General Sciences (50 titles). For more details on *Scopus*, see Goodman and Deis (2005) and Jacso (2005a).

In contrast to ISI and Elsevier, Google does not offer a publisher list, title list, document type identification, or any information about the time-span or the refereed status of records in *Google Scholar*. This and other studies, however, have found that *Google Scholar* covers print and electronic journals, conference proceedings, books, theses, dissertations, preprints, abstracts, and technical reports available from major academic publishers, distributors, aggregators, professional societies, government agencies, and preprint/reprint repositories at universities, as well as those available across the web (Bauer & Bakkalbasi, 2005; Gardner & Eng, 2005; Jacso, 2005b; Noruzi, 2005; Wleklinski, 2005). Examples of these sources include: The American Physical Society, Annual Reviews, arXiv.org, Association for Computing Machinery (ACM), Blackwell, Cambridge Scientific Abstracts (CSA), HighWire Press, Ingenta, Institute of Electrical and Electronics Engineers (IEEE), Macmillan, Meta Press, NASA Astrophysics Data System (ADS), National Institute of Health (NIH), National Oceanic and Atmospheric Administration (NOAA), Nature Publishing Group, Project MUSE, PubMed, RePEc (Research Papers in Economics), Sage, Springer, Taylor & Francis, University of Chicago Press, and Wiley, among others. Although *Google Scholar* does not cover material from all major publishers (e.g., American Chemical Society and Elsevier), it contains citations to articles from the American Chemical Society and Elsevier when documents from other sources cite these articles.

3.1 Citation Searching Methods

All data were manually collected by one of the authors (LIM) in February 2006. The “Cited Author” feature was used in *Web of Science* to identify citations to each individual item published by 22 LIS faculty members who constitute the study sample. Details of the same are discussed further below. The “Cited Author” search feature in *Web of Science* displays all the cited items of an author. The searcher then goes through all the entries and selects the relevant ones based on the information displayed for each entry (e.g., cited author, cited source, publication year of the cited item, and page numbers relevant to the cited item). Citations to items in which the faculty members were not first authors were included in the study.

Unlike *Web of Science*, *Scopus* does not provide the ability to browse the cited authors, cited works, or cited journals indexes or fields. Consequently, all available methods had to be used to accurately locate all potentially relevant citations in the database for each individual faculty member. The three methods used were:

- *Author Searches*: This feature allowed us to retrieve all articles in the database for each individual faculty member and, subsequently, identify all the records in the database that have cited these articles. Although the majority of citations found through this method

overlapped with those found in the Exact Match Searches method described above, this method identified a few unique citations.

- *Exact Match Searches*: This method used the title of an item as a search statement to locate an exact match of it in the “References” field. This method allowed us to identify most of the documents in the database that have cited items published by the study sample. In cases where the title was too short or ambiguous to refer to the item in question, we used additional information as keywords (e.g., the author’s last name, journal name, book or conference title, and/or the publisher name) ANDed with the title of the item. In cases where the title was too long, we used the first few words of the title. When in doubt, we manually examined all retrieved records to make sure that they cite the items in question. The Exact Match Searches identified the largest number of relevant citations for our study sample.
- *Advanced Searches*: This method was particularly useful for faculty members with unique last names, such as Mostafa and Nisonger. It was used only for double-checking rather than as a main method for locating citations. However, similar to the Author Searches method, the Advanced Searches method identified a few unique citations.

To make sure that citations were not missed by a database due to searching errors, we looked for the bibliographic records of all citations that were found in one database but not the other. When the bibliographic record of these items (1,300 or 44.7% of all 1996-2006 citations) were found in a database, we examined their cited references field to determine why they were not retrieved through the citation search methods described above and whether or not they should be counted as citations. Items that were missed due to searching errors (n=39) were counted as citations toward their respective databases. These errors were primarily caused by the use of long search statements or very short ones. In the case of long search statements, the search failed because some bibliographic references included automatically hyphenated words that prohibited the system from identifying an exact match with the search statement used. In the case of short search statements, items were missed because too many additional keywords were used (e.g., authors’ last names and journal name). Items that were missed due to database errors (e.g., lack of cited references information, incomplete lists of citations, citing errors, and misspellings) were tallied but were not counted as citations. Approximately 150 citations were missed from both databases due to these types of database errors.

Google Scholar was searched for citations in two different ways:

- *Author Searches*: This type of searches retrieves items published by an author and ranks them by relevance. In most cases, highly cited items appear first as *Google Scholar* uses Google’s crawler to index the content of research materials and automatically extracts and adds citation counts to retrieved documents to raise or lower individual articles in the rankings of a result set. Once the items are retrieved, the searcher will need to click on the “Cited by . . .” link to view the documents that cite each item. In cases where an author’s name is very common, additional keywords (e.g., journal name or keywords in title) are necessary to use to increase precision. Also needed is searching under variations of the author name to account for all name changes and/or citing styles, such as last-name first-name, first-name last-name, and first-name middle-initial last-name. All these variations of the author name can be ORed in the same search statement with each phrase placed between quotation marks. In cases where an accurate author search is not possible or impractical, exact match search strategy is recommended (albeit being much more tedious than author searches).

- *Exact Match Searches*: This type of searches uses the title of each item (e.g., journal article, book, book chapter, or conference paper) to determine whether or not it was cited. To ensure high precision and recall, the title should be searched for, or used, as an “exact phrase.” The result will be a list of documents that cite the item. In cases where the title is too long, it is recommended that only the first few words of the title phrase are used (enough string of words to make it a unique phrase). In cases where the title is too short or ambiguous to refer to only the item in question, the searcher has to use additional information as keywords (e.g., the author’s last name, journal name, book or conference title, the publisher name, or a combination of these keywords) ANDed with the phrase search string to narrow the result set to relevant documents.

A major disadvantage of *Google Scholar* is that its records are retrieved in a way that is very impractical for use with large sets or large numbers of study participants, requiring a very tedious process of manually extracting, cleaning, organizing, classifying, and saving the information into meaningful and useable formats. Unlike *Web of Science* and *Scopus*, *Google Scholar* does not allow re-sorting of the retrieved sets in any way (such as by date, author name, or data source); as mentioned earlier, retrieved records are rank ordered by how relevant they are to a query (taking into consideration the title and the full text of each article as well as the publication in which the article appeared and the number of citations). The result sets show short entries, displaying the title of the cited article and the name of the author(s); entries which include the link [Cited by . . .] indicate the number of times the article has been cited. Clicking on the link will take users to the list of citing articles. Users will be able to view the full-text of only those items that are available for free and those that their libraries subscribe to.

Other major disadvantages of *Google Scholar* include duplicate citations (i.e., counting a citation published in two different forms, such as preprint and journal article, as two citations) as well as the lack of any information about document type, document length, and the refereed status of the retrieved citations. In many cases, the item for which citations are sought for is retrieved and considered a citation by *Google Scholar*. The presence of all these problems in *Google Scholar* requires investigators to manually visit the retrieved citations for an author to assess and determine their relevance and their detailed information (e.g., document type, document length, refereed status, and even confirm whether or not it cites the author in question); otherwise investigators will generate skewed data and make inaccurate conclusions. Unless a system is developed that will automatically and accurately parse retrieved sets into meaningful, useable, and comparable data, all of these problems make *Google Scholar* prohibitive for large-scale citation database comparative studies.

3.2 Sample and Units of Analysis

In order to analyze the impact of using additional sources besides *Web of Science* on the citation count and citation ranking of LIS faculty members and be able to generalize the findings to the field, this study explored the difference in citation count and citation ranking *Scopus* and *Google Scholar* make to results from *Web of Science* for all 22 faculty members of the School of Library and Information Science at Indiana University (SLIS). These faculty members not only cover most if not all of the mainstream LIS research areas as identified by the Association of Library and Information Science Education (ALISE, 2006), they also belong to one of the largest and most published and cited American Library Association accredited LIS programs in North America (Budd, 2000). As of December 2005, the 22 SLIS faculty members have published over 1,118 scholarly work, including: 452 refereed journal articles, 260 conference papers (mostly refereed), 179 book chapters (some refereed), 46 books, and 21 edited works, among others (see Table 1). These faculty members have also been cited in 3,963 documents when citations to

individual faculty members are counted and 3,640 documents when citations to the school as a whole are counted (see Tables 2 and 3).

Table 1. SLIS Publication Data*

Document Type	Count**
Journal articles	462 (452)
Conference papers	272 (260)
Chapters	183 (179)
Technical reports / Working papers	72 (71)
Non-refereed journal and magazine articles	68 (67)
Books	46 (46)
Edited volumes	21 (21)
Encyclopedia articles	12 (12)
Bibliographies (monographs)	10 (10)
Total	1,145 (1,118)

*Book reviews, abstracts, editorial materials, letters to editors, panels, presentations, and so on are excluded from this table.

**Figures in parentheses refer to unique records (i.e., after removing duplicates due to co-authorship among SLIS faculty members).

Table 2. SLIS Citation Count By Year and Data Source

Years*	<i>Scopus</i>		<i>Web of Science (WoS)</i>		Union of <i>Scopus & WoS</i>	
	Total*	Unique**	Total*	Unique**	Total*	Unique**
1971-1975	0	0	1	1	1	1
1976-1980	0	0	16	16	16	16
1981-1985	0	0	137	133	137	133
1986-1990	0	0	222	217	222	217
1991-1995	4	3	388	367	388	367
Subtotal	4	3	764	734	764	734
1996	108	102	127	119	150	142
1997	130	124	137	129	162	154
1998	132	125	162	151	188	176
1999	152	137	150	137	187	172
2000	185	177	207	188	257	238
2001	276	249	241	213	318	288
2002	256	225	240	206	315	278
2003	337	300	295	261	400	360
2004	520	468	374	326	574	522
2005	485	430	393	341	556	498
2006 (February)	83	70	83	70	92	78
Subtotal	2,664	2,407	2,409	2,141	3,199	2,906
Grand Total**	2,668	2,410	3,173	2,875	3,963	3,640

*Sum of citations received by each faculty member.

**Citation counts of the school as a whole.

Table 3. Citation Count by Research Area and Time Period

Research areas of individual faculty members*	Overall	1996-2006
Citation analysis, informetrics, scholarly communication, and strategic intelligence	1,002	591
Human computer interaction	934	872
Computer-mediated communication, gender and information technology, and discourse analysis	379	370
E-commerce, information architecture, information policy and electronic networking	261	194
Information seeking and use and design and impact of electronic information sources	200	134
Bibliometrics, Collection development and management, evaluation of library sources and services, and serials	189	138
Community of practice and social informatics	178	178
Information visualization, data mining, and data modeling	174	172
Intelligent interfaces for information retrieval and filtering, knowledge discovery, and user modeling	160	157
Classification and categorization, ontologies, metadata, and information architecture	95	93
Information policy, social and organizational informatics, and research methods	78	35
Computational linguistics, computer-mediated communication, and sociolinguistics and language acquisition	47	46
Critical theory and documentation	43	43
Information retrieval	40	40
Citation analysis, bibliometrics, and data retrieval and integration	33	33
Faculty Members Total	3,963	3,199
School Total	3,640	2,906

*Two faculty members have not been cited yet.

The inclusion of high and diverse number of faculty members and publications in this study provided us with valuable framework to make citation comparisons between *Web of Science*, *Scopus*, and *Google Scholar*. It should be noted here that the analysis of data from *Web of Science* and *Scopus* was based on the publications and citations of all 22 faculty members, whereas the analysis of data from *Google Scholar* was based on the publications and citations of two faculty members. The use of data for only two faculty members in the case of *Google Scholar* was inevitable due to the labor-intensive nature of collecting data from this database. In this study, it took more time collecting and examining data from *Google Scholar* for two faculty members than it took collecting and examining data for 22 faculty members from *Web of Science* and *Scopus* combined.

To generate and analyze valuable data from *Google Scholar*, we selected two faculty members with extensive research background as well as those who had published in a variety of research areas and document types (e.g., journal articles, chapters, conference papers, books, technical reports, and so on). The research areas in this case included: bibliometrics, citation analysis, collection development and management, information retrieval and filtering, knowledge discovery, personalized delivery of information, serials, and user modeling.

All data were entered into an Excel file where items were coded by citing source (e.g., journal name, conference proceeding, chapter, and so on), document type (e.g., journal article, review article, conference paper, and so on), refereed status of the citing item, and source used to identify the citation. The refereed status of the citations was determined through *Ulrich's International Periodicals Directory* and the domain knowledge of the researchers.

3.3 Results And Discussion

The results of this study are presented and discussed in three sections: (1) The impact of using Scopus on the citation count and citation ranking of LIS faculty members as measured by Web of Science; (2) The sources of citations found in both databases; and (3) The impact and characteristics of citations from *Google Scholar* on the citation count and citation ranking of LIS faculty members as measured by Web of Science and Scopus combined. As mentioned earlier, the analysis of results from both Web of Science and Scopus is based on data for all 22 faculty members whereas the analysis of results from Google Scholar is based on data for two faculty members.

3.3.1 Impact of Scopus on Citation Count and Citation Ranking of LIS Faculty Members

To show the difference Scopus makes on the citation counts and citation ranking of LIS faculty members as measured by Web of Science, we decided to compare the number of citations retrieved by both databases, show the percentage of increase Scopus contributes towards the total number of citations to the school as a whole and to individual faculty members, examine the influence Scopus has on altering the relative citation ranking of faculty members, and explore the amount of overlap between the two databases. The refereed status of citations found in Web of Science and Scopus is not discussed because virtually all citations from these two databases came from refereed journals and conference proceedings.

As shown in Table 2, Web of Science retrieves 465 (or 19.3%) more citations for SLIS than Scopus does (2,875 in comparison to 2,410, respectively). This, however, is influenced by the fact that Web of Science provides citation coverage from 1945 to the present whereas Scopus provides citation coverage from 1996 to the present. Therefore, to make accurate assessments of the impact of Scopus on results from Web of Science and to make correct comparisons between the two databases, we decided to limit the analysis to citations from 1996 to 2006 only (all discussion hereafter is based on citation data from this period). When doing this, data show that Scopus retrieves 266 (or 12.4%) more citations than Web of Science. This may have been a result of the fact that Scopus indexes many more titles than Web of Science (over 15,000 in comparison to 8,700, respectively), but it also shows that Scopus provides more comprehensive coverage of LIS literature than Web of Science.

The data show that the addition of citations from Scopus to those from Web of Science increases the number of unique citations of the 22 SLIS faculty members by an average of 35.7% (or from 2,141 to 2,906 citations). In other words, if only Web of Science is used to locate citations for LIS faculty members and schools, on average, they would miss more than one-third of their citations due to deficiency in coverage. The percentage of increase, or loss of citations, among individual LIS faculty members, however, varies considerably depending on their research areas (see Table 4). For example, faculty members with research strengths in such areas as computer-mediated communication, data mining, data modeling, human computer interaction, information retrieval, information visualization, and social informatics will find their number of citations increase considerably more than those faculty members with research strengths in other areas. While this finding implies that certain faculty members will benefit more than others from using both databases to identify citations, it also suggests that to generate an accurate citation count for an LIS school and its faculty members and compare them to one another if needed, one has to use both Web of Science and Scopus.

The importance of using Scopus in addition to Web of Science is further evidenced by the fact that: (1) the relative ranking of faculty members changes (in some cases considerably) when citations from both databases are counted (see results for faculty members E, F, and I in Table 5); and (2) that the overlap between the two databases varies significantly from one faculty member or research area to another, ranging from a low 42.5% to a high 79.2% (see Table 6).

Table 4. Impact of Scopus on Web of Science Citation Count (1996-2006)

Research areas of individual faculty members*	Web of Science (WoS)	Union of WoS & Scopus	% Increase
Human computer interaction	559	872	56.0
Citation analysis, informetrics, scholarly communication, and strategic intelligence	533	591	10.9
Computer-mediated communication, gender and information technology, and discourse analysis	264	370	40.2
E-commerce, information architecture, information policy and electronic networking	167	194	16.2
Bibliometrics, Collection development and management, evaluation of library sources and services, and serials	123	138	12.2
Information seeking and use, design and impact of electronic information sources, and informetrics	121	134	10.7
Information visualization, data mining, and data modeling	119	172	44.5
Intelligent interfaces for information retrieval and filtering, knowledge discovery, and user modeling	118	157	33.1
Community of practice and social informatics	93	178	91.4
Classification and categorization, ontologies, metadata, and information architecture	84	93	10.7
Critical theory and documentation	35	43	22.9
Computational linguistics, computer-mediated communication, and sociolinguistics and language acquisition	34	46	35.3
Citation analysis, bibliometrics, and data retrieval and integration	31	33	6.5
Information retrieval	28	40	42.9
Information policy, social and organizational informatics, and research methods	27	35	29.6
Faculty Members Total	2,409	3,199	32.8
School Total	2,141	2,906	35.7

*Two faculty members have not been cited yet.

Table 5. Impact of Adding Unique Citations from Scopus on the Ranking of SLIS Faculty Members

Research Areas of Individual Faculty Members*	Web of Science		Union of Web of Science & Scopus	
	Count	Rank	Count	Rank
A	559	1	872	1
B	533	2	591	2
C	264	3	370	3
D	167	4	194	4
E	123	5	138	8
F	121	6	134	9
G	119	7	172	6
H	118	8	157	7
I	93	9	178	5
J	84	10	93	10
K	35	11	43	12
L	34	12	46	11
M	31	13	33	15
N	28	14	40	13
O	27	15	35	14

*Two faculty members have not been cited yet.

Table 6. Overlap Between Scopus and Web of Science

Research areas of individual faculty members*	Web of Science	Scopus	Union	Overlap
Human computer interaction	559	741	872	430 (49.3%)
Citation analysis, informetrics, scholarly communication, and strategic intelligence	533	467	591	409 (69.2%)
Computer-mediated communication, gender and information technology, and discourse analysis	264	314	370	209 (56.5%)
E-commerce, information architecture, information policy and electronic networking	167	165	194	139 (71.6%)
Bibliometrics, Collection development and management, evaluation of library sources and services, and serials	123	106	138	91 (65.9%)
Information seeking and use, design and impact of electronic information sources, and informetrics	121	114	134	101 (75.4%)
Information visualization, data mining, and data modeling	119	138	172	85 (49.4%)
Intelligent interfaces for information retrieval and filtering, knowledge discovery, and user modeling	118	131	157	92 (58.6%)
Community of practice and social informatics	93	162	178	77 (43.3%)
Classification and categorization, ontologies, metadata, and information architecture	84	78	93	69 (74.2%)
Critical theory and documentation	35	38	43	30 (69.8%)
Computational linguistics, computer-mediated communication, and sociolinguistics and language acquisition	34	38	46	26 (56.5%)
Citation analysis, bibliometrics, and data retrieval and integration	31	22	33	20 (60.6%)
Information retrieval	28	29	40	17 (42.5%)
Information policy, social and organizational informatics, and research methods	27	31	35	24 (68.6%)
Faculty Members Total	2,409	2,664	3,199	1,878 (58.7%)
School Total	2,141	2,407	2,906	1,645 (56.6%)

*Two faculty members have not been cited yet.

**Total after removing duplicate citations.

It should be emphasized here that if only one database is available to use to identify citations to an author's work, faculty members with research strengths in computer-mediated communication, data mining, data modeling, human computer interaction, information retrieval, information visualization, and social informatics, among others are ought to use Scopus instead of Web of Science, whereas faculty members with research interests in bibliometrics, citation analysis, classification and categorization, collection development and management, evaluation of library sources and services, information access, information architecture, informetrics, metadata, ontologies, scholarly communication, serials, and strategic intelligence are better off with using Web of Science as it retrieves more citations in their research areas than Scopus (see Table 7). As far as the type of documents in which the citations were found is concerned, no major differences were found between the two databases. Both Web of Science and Scopus retrieve most of the citations to LIS faculty members from journal articles followed by conference papers and review articles (see Table 8).

In conclusion, the findings here suggest that most if not all of the previous studies that exclusively used Web of Science to generate citation data to evaluate and/or rank scholars, journals, programs, and so on have been based on skewed and incomplete data and may have, consequently, resulted in making inaccurate assessments and developing imprecise rankings. One

should not forget, however, that previously, Web of Science was the only source available to use and conduct major citation analysis projects. The point here is that future studies will have to rely on both Web of Science and Scopus to generate an accurate citation account of LIS authors, journals, programs, and so on. This is very likely to be the case for other fields as well, but further investigation is required to verify this claim.

Table 7. Citation Count by Data Source for Individual Faculty Members

Research areas of individual faculty members*	Web of Science	Scopus	Difference
Human computer interaction	559	741	+182 (+32.6%)
Citation analysis, informetrics, scholarly communication, and strategic intelligence	533	467	-66 (-14.1%)
Computer-mediated communication, gender and information technology, and discourse analysis	264	314	+50 (+18.9%)
E-commerce, information architecture, information policy and electronic networking	167	165	-2 (-1.2%)
Bibliometrics, Collection development and management, evaluation of library sources and services, and serials	123	106	-17 (-16.0%)
Information seeking and use, design and impact of electronic information sources, and informetrics	121	114	-7 (-6.1%)
Information visualization, data mining, and data modeling	119	138	+19 (+16.0%)
Intelligent interfaces for information retrieval and filtering, knowledge discovery, and user modeling	118	131	+13 (+11.0%)
Community of practice and social informatics	93	162	+69 (+74.2%)
Classification and categorization, ontologies, metadata, and information architecture	84	78	-6 (-7.7%)
Critical theory and documentation	35	38	+3 (+8.6%)
Computational linguistics, computer-mediated communication, and sociolinguistics and language acquisition	34	38	+4 (+11.8%)
Citation analysis and data retrieval and integration	31	22	-9 (-40.9%)
Information retrieval	28	29	+1 (3.6%)
Information policy, social and organizational informatics, and research methods	27	31	+4 (+14.8%)
Faculty Members Total	2,409	2,664	+255 (+10.6%)
School Total	2,141	2,407	+266 (+12.4%)

*Two faculty members have not been cited yet.

Table 8. Citations by Document Type

Document Type	Web of Science		Scopus		Combined	
	Count	%	Count	%	Count	%
Journal articles	1,612	75.3	1,862	77.4	2,108	72.5
Conference papers	225	10.5	338	14.0	486	16.7
Review papers	190	8.9	165	6.9	195	6.7
Editorial materials	68	3.2	37	1.5	68	2.3
Book reviews	23	1.1			23	0.8
Book chapters	13	0.6			13	0.4
Other	10	0.5	5	0.2	12	0.4
Total	2,141	100.0	2,407	100.0	2,906	100.0

3.3.2 Type and Sources of Citations in Web of Science and Scopus

As mentioned earlier, only 56.6% of all the citations were duplicated in both databases, raising an important question which is “where did all the 1,261 unique citations come from?” Answering this question is important because it will identify coverage strengths and weaknesses in both databases. Data show that the 2,141 citations from Web of Science come from 528 different

journals and conference proceedings whereas the 2,407 citations from Scopus come from 699 different titles. The 2,906 unique citations from both databases come from 816 different journals and conference proceedings. Data, however, show that 95 (or 11.6%) of these 816 titles account for 62.1% of all citations (see Table 9). Data also show that of the top 95 sources of citations, 23 are not indexed by Web of Science whereas only three of them are not indexed by Scopus.

Further analysis show that when a journal or conference proceeding is indexed by both databases, Web of Science tend to provide in some cases significantly better coverage of the titles than Scopus. For example, of the top 95 sources of citations, 69 of them are indexed by both Web of Science and Scopus. These 69 titles generate a total of 1,509 citations of which 1,430 are found through Web of Science but only 1,210 through Scopus. So, the question is why these two databases are missing 79 and 299 citations, respectively. In the case of Web of Science, most of the 79 citations were missed due to database errors

3.3.3 Conclusions And Implications

This study provides direct and meaningful implications for faculty members who need assistance in compiling their own citation records and also for use as a general reference tool (e.g., for locating citations to a particular paper or book). The study informs reference and other information specialists of novel ways of identifying citations to an author, paper, or journal. Until very recently, ISI citation databases were essentially the only practical sources for locating these references and citations. This study showed that other practical methods and sources, such as *Scopus* and *Google Scholar*, can be used to locate citations not covered by ISI. Significantly, this study showed that:

- 1) *Web of Science* should not be used alone for locating citations to an author or title.
- 2) *Google Scholar* is evidently multi-disciplinary and thus can be useful in any field for citation searching purposes.
- 3) *Scopus* and *Google Scholar* can help identify a considerable number of potentially valuable citations not found in *Web of Science*;
- 4) *Scopus* and *Google Scholar* can help identify a considerable number of citations in document types not covered by ISI citation databases (conference proceedings in the case of both *Scopus* and *Google Scholar* and additionally in *Google Scholar* preprints, technical reports, research reports, theses, dissertations, and so on;
- 5) *Scopus* and *Google Scholar* may assist in providing a more comprehensive picture of the extent of international and interdisciplinary nature of scholarly communication of and among researchers; and
- 6) All three databases complement each other rather than replace each other, particularly as shown in Mostafa's case.

This study, furthermore, has significant implications on the wider scholarly community as researchers start to adopt the search method used here and *CiteSearch* that was developed as part of the study to identify citation sources in such fields as business, economics, history, law, medicine, political science, psychology, and sociology.

Given the continuous advances in information technology and improvement in online access to tens of millions of records through databases and services that provide citation information, future studies should explore:

1. Other sources and searching methods that can and should be used to locate citations not covered by ISI citation databases, *Scopus*, or *Google Scholar*.
2. Differences that these sources could make in citation counts and citation traits for authors, papers, and journals.

3. Whether broader sourcing of citations can alter one's relative ranking vis-à-vis others and, if so, how.
4. Which sources of citations provide better coverage of certain subject disciplines than others.

We hope that other researchers can use this study as a model for exploring the impact of broadening the sources of citations in other fields. In short, while all three databases provide considerable coverage of literatures in all fields, *Google Scholar* stands out in its coverage of international, non-English language journals, among others. *Google Scholar* also indexes a wide variety of document types, some of which may be of significant value to researchers. An important finding is the ability of using one database to identify errors in the other database.

The increasing availability of online information resources and open access journals will make *Google Scholar* very popular among scholars as they try to find citations to their work or to items they are using for research. With the use of CiteSearch, *Google Scholar* will eventually become an indispensable data source for citation analysis and other bibliometric analyses.

REFERENCES

- Bauer, K., & Bakkalbasi, N. (2005). An Examination of Citation Counts in a New Scholarly Communication Environment. *D-Lib Magazine*, 11(9). Retrieved March 25, 2006, from <http://www.dlib.org/dlib/september05/bauer/09bauer.html>.
- Gardner, S., & Eng, S. (2005). Gaga over Google? Scholar in the Social Sciences. *Library Hi Tech News*, 22(8), 42-45.
- Goodman, D. & Deis, L. (2005). *Web of Science* (2004 version) and *Scopus*. *The Charleston Advisor*, 6(3). Retrieved March 25, 2006, from <http://www.charlestonco.com/dnloads/v6n3.pdf>.
- Jacsó, P. (2005a). 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. Retrieved March 15, 2006, from <http://www.ias.ac.in/currsci/nov102005/1537.pdf>.
- Jacsó, P. (2005b). *Google Scholar*: the pros and the cons. *Online Information Review*, 29(2), 208-214.
- Noruzi, A. (2005). *Google Scholar*: The new generation of citation indexes. *Libri*, 55(4), 170-180.
- Thomson Corporation. (2006). *Web of Science 7.0*. Retrieved June 15, 2005, from <http://scientific.thomson.com/support/products/wos7/>.
- Wleklinski, J.M. (2005). Studying *Google Scholar*: wall to wall coverage? *Online*, 29(3), 22-26.