Multi-faceted Citation Analysis for Quality Assessment of Scholarly Publications

학술논문 품질평가를 위한 다방면 인용분석방식

Kiduk Yang*
Lokman Meho**

ABSTRACT

Despite the widespread use, critics claim that citation analysis has serious limitations in evaluating the research performance of scholars. 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 citation database coverage of today is disjoint and incomplete, which can result in conflicting quality assessment outcomes across different data sources. This paper discuss the findings from a citation analysis study that measured the impact of scholarly publications based on the data mined from *Web of Science*, *Scopus*, and *Google Scholar*, and briefly describes a work-in-progress prototype system called CiteSearch, which is designed to overcome the weaknesses of existing citation analysis methods with a robust citation-based quality assessment approach.

초 록

인용분석은 학자들의 연구실적 평가에 가장 많이 사용되는 방법 중 하나이지만 비평가들은 오늘날의 인용분석 자료와 방법론에 근본적인 문제가 있다고 주장한다. 전통적 인용분석 방식은 인용품질과 인용소음뿐만 아니라 언어, 시간, 문화와 같은 비수치적인 요소들을 고려하지 않아 단순하고 그릇된 평가를 가져올 수 있으며, 적용범위가 각각 다르고 불완전한 인용 데이터베이스들은 충돌적인 인용분석결과를 초래하기 쉬울 수 있다. 이러한 문제들을 해결하려면 포괄적인 인용데이터를 다 방면과 다 방식으로 분석하는 새로운 인용분석연구가 필요하다. 본 논문은 *Web of Science*, *Scopus*와 *Google Scholar*를 비교 분석한 연구의 결과를 논의하며 기존의 인용분석 방법의 약점을 극복하기 위해 설계한 CiteSearch라는 프로토타입 시스템을 간략하게 설명한다.

Keywords: citation analysis, quality assessment, scholarly publication, fusion method, citation database
인용분석, 품질평가, 학술논문, 통합방식, 인용데이터베이스

---

* Keimyung University. Department of Library and Information Science. Daegu, Korea (The First Author, kiyang@kmu.ac.kr)
** American University of Beirut. Department of Political Science and Public Administration. Beirut, Lebanon(lmeho@aub.edu.lb)

▪ Received: 15 May 2011 ▪ Revised: 24 May 2011 ▪ Accepted: 3 June 2011
▪ Journal of the Korean Society for Information Management, 28(2): 79-96, 2011.
〔DOI:10.3743/KOSIM.2011.28.2.079〕

1. Introduction

Citation analysis is one of the most widely used methods in evaluating the research performance of scholars (Holden, Rosenberg, & Barker 2005; Moed 2005; Lewison 2001). The basic assumption underlying citation analysis is that citations are a way of giving credit to and recognizing the value, quality, or significance of an author's work (Cronin 1984; van Raan 1996). While the proponents have reported the validity of using citation counts for research assessments (Aksnes & Taxt 2004; Holmes & Oppenheim 2001; Martin 1996), critics claim that citation analysis has serious limitations in both data and methodology (MacRoberts & MacRoberts 1996; Seglen 1998). The problems reported in literature point to two fundamental shortcomings with the typical citation analysis approach. 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 outcomes across different data sources.

One of the ways to address these limitations is to develop a multi-faceted citation analysis approach that employs a range of quality assessment methods to analyze comprehensive citation data from multiple sources. This paper discuss the findings from a citation analysis study that measured the impact of scholarly publications based on the data mined from *Web of Science*, *Scopus*, and *Google Scholar*, and briefly describes a work-in-progress prototype system called CiteSearch, which is designed to overcome the weaknesses of existing citation analysis methods with a robust citation-based quality assessment approach. The paper is organized as follows: Section 2 summarizes the related research and section 3 discusses the findings from the citation analysis study, which is followed by a description of the CiteSearch prototype in section 4.

2. Related Research

Nisonger (2004), who conducted a self-study to show how ISI coverage compared to citation data he collected, found that ISI coverage was only a fraction of his own. ISI captured only 29% of his total citations with non-US citation coverage at 20% and non-English citations at 2%. Nisonger concluded that assessment of faculty productivity should not be based on ISI citation counts alone, especially when demonstration of international impact is important. He also suggested that rankings based on ISI data of a discipline's most-cited authors or academic departments might be significantly different if non-ISI citation data were included.

Bauer and Bakkalbasi (2005) compared citation counts provided by *Scopus*, *Google Scholar*, and *Web of Science* for articles from the *Journal of the American Society for Information Science and Technology* published in 1985 and in 2000. The results for 1985 articles were inconclusive, but for 2000 articles, *Google Scholar* provided statistically significant higher citation counts than either *Scopus* or *Web of Science*. The authors concluded that researchers should consult *Google Scholar* in addition to *Scopus* or *Web of Science*, especially for relatively recent publications. Jacsó (2005), who conducted several tests comparing *Google Scholar*, *Scopus*, and *Web of Science*, found many unique documents that were relevant and substantial in each source. Noruzi (2005) studied the citation counts of 36 webometrics papers in *Google Scholar* and *Web of Science*; in most cases, he found that *Google Scholar* provided higher citation counts than *Web of Science*. These findings were corroborated by the results of Vaughan and Shaw (2008) for information science.

Bakkalbasi, Bauer, Glover, and Wang (2006) compared citation counts for articles from 11 oncology journals and 11 condensed matter physics journals published in 1993 and 2003. Their data showed a significant difference in the mean citation rates between all pairs of resources except between *Google Scholar* and *Scopus* for condensed matter physics in 2003. For articles published in 2003, *Web of Science* returned the largest amount of unique citing material for condensed matter physics and *Google Scholar* returned the most for oncology. The authors concluded that all three data sources returned some unique material and that the question of which provided the most complete set of citing literature might depend on the subject and publication year of a given article. In four science disciplines, Kousha and Thelwall (2006) found that the overlap of citing documents between *Google Scholar* and *Web of Science* varies from one field to another and, in some cases, such as chemistry, it is relatively low (33%).

Norris and Oppenheim (2007) used all but 720 of the journal articles submitted for the purpose of the 2001 Research Assessment Exercise in the social sciences, as well as the list of 2,800 journals indexed in the *International Bibliography of the Social Sciences*, to assess the coverage of four data sources (CSA Illumina, *Google Scholar*, *Scopus*, and *Web of Science*). They found that *Scopus* provides the best coverage of social science literature from among these data sources and concluded that *Scopus* could be used as an alternative to *Web of Science* as a tool to evaluate research impact in the social sciences. Bar-Ilan (2006) carried out an ego-centric citation and reference analysis of the works of the mathematician and computer scientist, Michael O. Rabin, utilizing and comparing Citeseer, *Google Scholar*, and *Web of Science*. She found that the different collection and indexing policies of the different data sources lead to considerably different results. In another study, Bar-Ilan, Levene, and Lin (2007) compared the rankings of the publications of 22 highly-cited Israeli researchers as measured by the citation counts in *Google Scholar*, *Scopus*, and *Web of Science*. The results showed high similarity between *Scopus* and *Web of Science* and lower similarities between *Google Scholar* and the other databases.

More recently, a citation study carried out by the current investigators further demonstrated the necessity of using multiple citation sources (Meho & Yang 2007). The study used citations to more than 1,400 works by 25 library and information science faculty to examine the effects of additionally using *Scopus* and *Google Scholar* on the citation counts and rankings of these faculty members as measured by *Web of Science*. The study found that the addition of *Scopus* citations to those of *Web of Science* significantly altered the relative ranking of faculty in the middle of the rankings (Spearman Rank Order correlation coefficient = -0.45 at 0.01 level). The study also found that *Google Scholar* stands out in its coverage of conference proceedings as well as international, non-English language journals. According to the authors, the use of *Scopus* and *Google Scholar*, in addition to *Web of Science*, reveals a more comprehensive and complete picture of the extent of the scholarly relationship between library and information science and other fields. Most recently, Bar-Ilan (2008) compared the *h* index of a list of 40 highly-cited Israeli researchers based on citation counts from *Google Scholar*, *Scopus*, and *Web of Science*. In several cases, she found that the results obtained through *Google Scholar* were considerably different from those in *Scopus* and *Web of Science*, primarily due to citations covered in non-journal items.

3. Comparison of *Web of Science*, *Scopus*, and *Google Scholar*

Quality assessment of scholarly publications faces some major challenges. As the studies indicate, citation databases today are far from comprehensive. Furthermore, they contain fragmented, duplicate, and even erroneous citations at times. On top of incomplete and inaccurate data, conventional citation analysis utilizes count-based assessment methods, which lead to one-dimensional and simplistic evaluation of scholarly work.

One way to address these limitations of the existing citation databases and citation analysis methods is to develop a multi-faceted and fusion-based citation analysis approach that applies a spectrum of quality assessment methods to a combined data from multiple citation databases. As a first step towards this goal, authors investigated the existing citation analysis environment by comparing *Web of Science*, *Scopus* and *Google Scholar*. In order to collect the citation data, *Web of Science*, *Scopus*, and *Google Scholar* were searched for citations to about 1,100 publi-cations by fifteen faculty members at Indiana Uni- versity School of Library and Information Science[1]. This section presents a summary of findings and conclusions from this investigation.

3.1 Citation Database Coverage

As Table 1 indicates, there is a marked difference in coverage between *Google Scholar* and the other two databases. 500 million plus records in *Google Scholar*, estimated as of 2006 (Giustini 2006), is over 10 times the size of either *Web of Science* or *Scopus*, which means the coverage *Google Scholar* is likely to dwarf other citation databases today.

1) Over 10,000 citations were examined for the study.

<Table 1> Comparison of Citation Database Coverage

	Web of Science	*Scopus*	*Google Scholar*
Breadth of Coverage	46M records 11,000 + titles Journals & Conference Papers	41M records 18,000 titles Journals & Conference Papers	500M + records ? titles 30 + Document Types
Coverage Years	A&HCI: 1975-present SCI: 1900- present SSCI: 1956- present	1996-present (with cited references) 1966-present (without cited references)	Unknown

Data as of 2010.

Powered by Google's army of Web crawlers, *Google Scholar* will continue to grow at a much faster rate than any other citation databases in existence.

The massive coverage of *Google Scholar*, however, is offset by the quality of both its data and service. The citations that are harvested with little human intervention (i.e., no quality control) contain high level of noise that ranges from incomplete and inconsistent citations to duplicate and erroneous citations. In addition, *Google Scholar* 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 its database, nor does it offer much in terms of service other than a count of citations that are hyperlinked to citing publications.

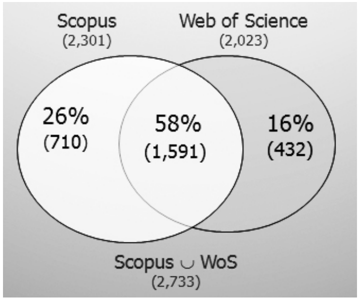
Web of Science and *Scopus*, on the other hand, offers powerful features for browsing, searching, sorting and saving functions, as well as exporting to citation management software (e.g., *EndNote* and *RefWorks*). Although their database is much smaller than that of *Google Scholar*, the quality of data in *Web of Science* and *Scopus* is thought to be much higher due to data validation and normalization efforts employed by their parent companies Thompson and Elsevier. In addition, both Thompson and Elsevier are at forefront of citation analysis research by supporting new study efforts with sharing of their data and fostering the exchange of study findings in sponsored conferences. Consequently, *Web of Science* and *Scopus* now offer a multitude of new citation analysis measures such as Eigenfactor that go beyond simple counting of citations.

3.2 *Scopus* and *Web of Science*

In our study sample, *Scopus* contained 278 more citations than *Web of Science* (*WoS*). When combined, *Scopus* increases *WoS* citations by 35% (710/2023), whereas *WoS* increases *Scopus* citations by only 19% (432/2301). These patterns reflect more comprehensive coverage by *Scopus* (18,000 vs. 11,000 titles), while the Venn diagram in Figure 1 illustrates the relative low overlap (58%) and high uniqueness (42%) between two databases.

The impact of adding *Scopus* citations to *WoS* varies greatly between research areas. As shown in Table 2, increase in citations ranges from 5% to 99% when both databases are combined. When faculty members are ranked by the total number of citation counts to their publications, the impact of database coverage is even more meaningful. According to the study data, adding *Scopus* citations to *WoS* significantly altered the relative rankings of faculty members at middle ranks[2] (Table 3).

Examining citation counts by document type reveals an interesting difference between *Scopus* and *WoS*. Although both databases contain similar ratio of journal articles to conference papers (Table 4), *Scopus* has a much better coverage of conference proceedings than *WoS*. Figure 2, which shows only conference papers in both databases, illustrates a very small overlap (18%) and the unique citations by *Scopus* (54%) twice as large as that of *WoS* (28%).



〈Figure 1〉 *Scopus* vs. *Web of Science*

〈Table 2〉 Impact of *Scopus* by Research Area

Research areas of individual faculty members*	*WoS*	*Scopus*	Union of *WoS* and *Scopus*	% Increase
Human-computer interaction	544	740	853	56.8%
Citation analysis, informetrics, scholarly communication, and strategic intelligence	508	459	564	11.0%
Computer-mediated communication, gender and information technology, and discourse analysis	273	313	365	33.7%
E-commerce, information architecture, information policy and electronic networking	162	168	188	16.0%
Bibliometrics, Collection development and management, evaluation of library sources and services, and serials	123	108	137	11.4%
Information seeking and use, design and impact of electronic information sources, and informetrics	122	111	128	4.9%
Intelligent interfaces for information retrieval and filtering, knowledge discovery, and user modeling	118	129	154	30.5%
Information visualization, data mining, and data modeling	115	133	165	43.5%
Communities of practice	88	159	175	98.9%
Classification and categorization, ontologies, metadata, and information architecture	83	80	93	12.0%
Critical theory and documentation	35	37	42	20.0%
Computational linguistics, computer-mediated communication, and sociolinguistics and language acquisition	32	38	44	37.5%
Citation analysis, bibliometrics, and data retrieval and integration	29	21	31	6.9%
Information retrieval and data integration	28	32	40	42.9%
Information policy, social and organizational informatics, and research methods	28	31	34	21.4%
Faculty Members Total	**2,288**	**2,559**	**3,013**	31.7%
School Total**	**2,023**	**2,301**	**2,733**	35.1%

*Each row in the table represents a single faculty member and the main research topics covered by him/her. It would have been practically impossible to classify citations by individual topics rather than individual faculty members.
**Excludes duplicate citations.

2) Spearman Rank Order correlation coefficient was -0.45 at 0.01 level for the middle third of the ranking changes.

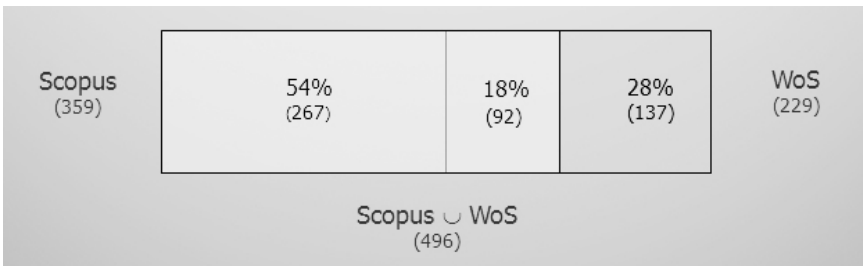
<Table 3> Impact of *Scopus* on Faculty Ranking

Faculty Member	*WoS*		Union of *WoS* and *Scopus*	
	Count	Rank	Count	Rank
A	544	**1**	853	**1**
B	508	**2**	564	**2**
C	273	**3**	365	**3**
D	162	**4**	188	**4**
E	123	**5**	137	**8**
F	122	**6**	128	**9**
G	118	**7**	154	**7**
H	115	**8**	165	**6**
I	88	**9**	175	**5**
J	83	**10**	93	**10**
K	35	**11**	42	**12**
L	32	**12**	44	**11**
M	29	**13**	31	**15**
N	28	**14T**	40	**13**
O	28	**14T**	34	**14**

<Table 4> *Scopus* and *Web of Science* by Document Type

Document Type	*Web of Science*		*Scopus*		Union	
	Count*	%	Count*	%	Count*	%
Journal articles	1,529	75.6%	1,754	76.2%	1,968	72.0%
Conference papers	229	11.3%	359	15.6%	496	18.1%
Review articles	172	8.5%	147	6.4%	175	6.4%
Editorial materials	63	3.1%	36	1.6%	64	2.3%
Book reviews	17	0.8%	0	0.0%	17	0.6%
Letters to editors	9	0.4%	2	0.1%	9	0.3%
Bibliographic essays	2	0.1%	2	0.1%	2	0.1%
Biographical item	2	0.1%	1	0.0%	2	0.1%
Total	**2,023**	**100.0%**	**2,301**	**100.0%**	**2,733**	**100.0%**
Total from Journals	1,794	88.7%	1,942	84.4%	2,237	81.9%
Total from Conference Papers	229	11.3%	359	15.6%	496	18.1%
Total	**2,023**	**100.0%**	**2,301**	**100.0%**	**2,733**	**100.0%**

*Excludes duplicate citations.



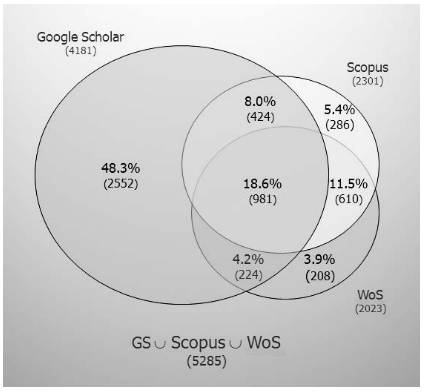
<Figure 2> *Scopus* and *Web of Science*: Conference Papers Only

3.3 *Google Scholar*

Google Scholar differs from *Scopus* and *WoS* not only in size of its database but also in document types covered. In contrast to *Scopus* and *WoS* whose data consist mostly of journal articles and conference papers at a rough ratio of 9 to 1, *Google Scholar* contains some 30-plus document types with much smaller proportion of journal articles (42%) and conference papers (34%).

Another major difference of *Google Scholar* from other citation databases lies in its international coverage. As Table 5 shows, *Google Scholar* provides markedly better coverage of non-English materials (7%) than either *Scopus* (1%) or *WoS* (1%).

The overall impact of *Google Scholar* is much more pronounced. Adding citations from *Google Scholar* to the union of *Scopus* and *WoS* increases their total count by 93%, whereas the union of *Scopus* and *WoS* data increases *Google Scholar* counts by only 26% (Figure 3). Much of this increase in citations by *Google Scholar* is likely due to its even broader coverage of conference proceedings than *Scopus* (1,849 by *Google Scholar* vs. 496 by the union of *Scopus* and *WoS*). In fact, *Google Scholar* has over twice as many unique citations as *Scopus* and *WoS* combined (2,552 vs. 1,104, respectively).



〈Figure 3〉 *Google Scholar* vs. *Scopus*∩*WoS*

Google Scholar citation counts by research area

〈Table 5〉 Citations by Language

	Web of Science		*Scopus*		*Google Scholar*		Total	
	Count	%	Count	%	Count*	%	Count*	%
English	2,000	98.86%	2,285	99.30%	3,891	93.06%	4,972	94.08%
Portuguese					92	2.20%	92	1.74%
Spanish	4	0.20%	3	0.13%	63	1.51%	68	1.29%
German	13	0.64%	9	0.39%	38	0.91%	50	0.95%
Chinese					44	1.05%	44	0.83%
French	3	0.15%	1	0.04%	32	0.77%	35	0.66%
Italian					8	0.19%	8	0.15%
Japanese	3	0.15%	3	0.13%	1	0.02%	4	0.08%
Swedish					3	0.07%	3	0.06%
Czech					2	0.05%	2	0.04%
Dutch					2	0.05%	2	0.04%
Finnish					2	0.05%	2	0.04%
Croatian					1	0.02%	1	0.02%
Hungarian					1	0.02%	1	0.02%
Polish					1	0.02%	1	0.02%
Total	**2,023**	**100.00%**	**2,301**	**100.00%**	**4,181**	**100.00%**	**5,285**	**100.00%**
Non-English	**23**	**1.14%**	**16**	**0.70%**	**290**	**6.94%**	**313**	**5.92%**

show patterns similar to *Scopus* and *WoS* in that the coverage varies greatly between research areas. In comparison to 5% to 98% increase in citation counts when *Scopus* data is added to *WoS*, adding *Google Scholar* data to the union of *Scopus* and *WoS* increases the counts from 24% to 144% (Table 6). Closer examination suggests *Google Scholar* to have a strong coverage in Computer Science and Information Science areas such as human computer interaction, computational linguistics, and social informatics, whereas *Scopus* and *WoS* seem to have stronger coverage in Library Science fields such as bibliometrics, collection development, and information policy.

Despite its obvious advantage in size and coverage of conference proceedings, addition of *Google Scholar* data to *Scopus* and *WoS* does not significantly change the relative rankings of the faculty members in the study data, which is in contrast to significant ranking changes caused by addition of *Scopus* data to *WoS*. One possible explanation of this phenomena is that the increased coverage of *Google Scholar* does not significantly alter the citation patterns present in the combined database of *Scopus* and *WoS* but just brings to the table more of the same type

<Table 6> Impact of *Google Scholar* (GS) by Research Area

Research areas of individual faculty members*	Union of *Web of Science* and *Scopus*	GS	Union of Three Sources	% Increase
Human-computer interaction	853	1,786	2,078	143.6%
Citation analysis, informetrics, scholarly communication, and strategic intelligence	564	517	802	42.2%
Computer-mediated communication, gender and information technology, and discourse analysis	365	671	797	118.4%
E-commerce, information architecture, information policy and electronic networking	188	164	244	29.8%
Bibliometrics, Collection development and management, evaluation of library sources and services, and serials	137	94	169	23.4%
Information seeking and use, design and impact of electronic information sources, and informetrics	128	114	171	33.6%
Intelligent interfaces for information retrieval and filtering, knowledge discovery, and user modeling	154	260	291	89.0%
Information visualization, data mining, and data modeling	165	187	249	50.9%
Communities of practice and social informatics	175	342	403	130.3%
Classification and categorization, ontologies, metadata, and information architecture	93	76	125	34.4%
Critical theory and documentation	42	46	60	42.9%
Computational linguistics, computer-mediated communication, and sociolinguistics and language acquisition	44	73	92	109.1%
Citation analysis, bibliometrics, and data retrieval and integration	31	29	39	25.8%
Information retrieval and data integration	40	46	59	47.5%
Information policy, social and organizational informatics, and research methods	34	20	42	23.5%
Faculty Members Total	**3,013**	**4,425**	**5,621**	**86.6%**
School Total**	**2,733**	**4,181**	**5,285**	**93.4%**

of data that can be found in *Scopus* and *WoS*. Another possible explanation is that the study sample, which is largely library science-centric, favors *Scopus* and *WoS*, thus masking the impact of *Google Scholar* on the relative rankings of scholars.

3.4 Study Findings

One of the most important findings from the study is that *Scopus*, *WoS*, and *Google Scholar* complement, rather than replace, one another. *Google Scholar* can be useful in showing evidence of broader international impact than could possibly done utilizing *Scopus* and *WoS* data alone, while *Scopus*, together with *WoS*, can play a vital role in assessing the relative rankings of scholars.

The broad coverage of *Google Scholar* may be useful for citation searching purposes, but it is not yet conducive for a large-scale comparative citation analysis due to data noise and lack of services. In fact, data collection from *Google Scholar* took over 3,000 hours, compared to 100 hours for *WoS* and 200 hours for *Scopus*. Citation analysis study, even in a small scale, requires enormous efforts to refine the search strategy, parse search results, eliminate data noise, and extract and normalize the resulting citation metadata.

Small overlap in coverage between citation databases suggests that combined sources may significantly influence the outcome of citation analysis. The fact that citation coverage varies widely across research area, document type, and language suggests that simplistic frequency based citation analysis will be an inadequate measure for assessing the quality of scholarly works that is inherently multi-dimensional and often contextual.

In recent years, a variety of new citation analysis measures has come about. The popular *h*-index (Hirsh 2005) leverages the distribution of citations to a scholar's publications to quantify his or her scientific research output with a single number (*h* papers receiving at least *h* citations each), while *g*-index (Egghe 2006) uses a similar approach (*g* papers receiving at least g^2 citations together). Though these measures are improvements over straight citation counts, they are still one-dimensional measures that can be unduly influenced by specific data patterns and thus produce misleading assessments of scholarly impact.

The most promising of recent citation analysis measures is the eigenfactor (Bergstrom 2007; West, Bergstrom, & Bergstrom 2010). Eigenfactor, which is based on the Google's PageRank algorithm, estimates the importance of a body of scholarly work by a propagation of citation importance, thus effectively addressing the problem of variance in citation quality. As is the case with PageRank, however, the computation of eigenfactor requires a recursive propagation across citation links, whose effectiveness is heavily influenced by the scope and quality of harvested citation topology. Consequently, neither the data collection nor the computation requirements for eigenfactor can be easily met by everyday research of today.

The inadequacy of using single measures is illustrated in Table 7, where all individuals in the expanded dataset[1] of the study are ranked by three different

<Table 7> Ranking of Scholars by Publication Count, Citation Count, and Citation Log-sum

Author	#pub	cn	cn-log	Author	#pub	cn	cn-log	Author	#pub	cn	cn-log
20976	30	88	10.553	11991	1	17552	9.773	421	26	8038	105.745
421	26	8038	105.745	11813	1	16665	9.721	32	18	5045	85.073
3717	26	26	0.000	1	1	15076	9.621	35	16	5038	73.069
2376	22	1173	70.522	3	8	13302	50.881	2376	22	1173	70.522
1017	21	230	47.593	6547	7	10294	40.460	456	15	3278	66.247
23657	20	100	23.475	421	26	8038	105.745	422	11	5753	56.595
32	18	5045	85.073	20258	1	6975	8.850	791	12	1148	53.157
2378	18	771	47.137	8559	1	6887	8.837	418	12	1800	52.428
1265	17	1487	48.008	5	5	6755	32.713	3	8	13302	50.881
35	16	5038	73.069	11269	5	6126	35.064	982	12	1192	49.459
17359	16	336	36.132	11262	1	5759	8.659	1265	17	1487	48.008
15818	16	171	32.894	422	11	5753	56.595	1017	21	230	47.593
456	15	3278	66.247	14308	1	5210	8.558	17332	11	1349	47.222
8193	15	467	45.655	11281	1	5130	8.543	2378	18	771	47.137
4892	15	161	24.924	32	18	5045	85.073	8193	15	467	45.655
1137	15	117	16.134	35	16	5038	73.069	30	9	2151	45.211
8796	14	507	44.214	1422	1	5010	8.519	473	12	1619	44.507
11756	14	127	22.035	2	2	5006	12.847	8796	14	507	44.214
16018	14	62	16.760	12017	1	4507	8.413	827	10	1178	40.795
418	12	1800	52.428	56	1	4495	8.411	6547	7	10294	40.460
473	12	1619	44.507	13151	2	4437	15.250	54	7	2695	38.506
477	12	1577	37.201	4676	1	4378	8.384	1459	9	934	37.942
982	12	1192	49.459	15788	1	4355	8.379	1596	8	1326	37.475
791	12	1148	53.157	12519	1	4314	8.370	477	12	1577	37.201

measures, namely the number of publications (*#pub*), total number of citations to a person's publications (*cn*), and the sum of log of citations to each publication (*cn-log*). Three rankings by three measures differ radically from one another. The leftmost table, which uses publication counts for ranking, ignores differences in publication quality, while the middle table rankings using total citation counts are unduly influenced by a small number of publications with inordinately large number of citations. The rightmost table, which uses the sum of log citation counts to publications seems to balance the publication and citation counts by applying log transformation to citation counts for normalization. *cn-log* is a promising measure in that it is simple to calculate yet accommodates the log-like distribution of citations.

4. CiteSearch System

In order to realize the goal of developing a robust citation-based quality assessment method that overcomes the weaknesses of existing citation analysis approaches, we are developing 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 ongoing, so what follows is a general description and brief overview of the system design.

4.1 Multi-faceted Citation Analysis

The CiteSearch system focuses on two key objectives: integration of citation data from multiple sources to ensure that quality assessment data is both complete and clean (i.e., without noise), and multi-faceted citation analysis to leverage not only multiple sources but also multiple aspects of evidence in the citation data landscape. Integration of citation data from multiple sources involves data mining, filtering, metadata extraction, data normalization and data fusion, while multi-faceted citation analysis involves derivation and fusion of multiple citation-based quality evaluation measures such as *CiteRank* (a citation propagation measure similar to Google's PageRank), *Mentor-Index* (an index to measure mentoring impact by aggregation of students' research impact), and *CiteAuthority*, which is based on Kleinberg's (1998) HITS (Hyperlink Induced Topic Search) algorithm.

To transform citation data into an effective measure of publication quality, one must go beyond the mere frequency of citations and consider multiple facets of citation evidence. More specifically, the key objective of multi-faceted citation analysis is to generate quality/impact assessment measures that account for variance in citation quality (e.g., weighted citation counts, CiteRank), consider various facets of the evaluation metric (e.g., document type, language), and accommodate different aspects of quality assessment (e.g., *h*-index, Mentor-index). Using the integrated citation database, document-level quality assessment measures, such as citation count, CiteRank, and CiteAuthority, will be computed for each item in the LIS publication list, and author-level measures, such as publication counts, *h*-index, Mentor-index, will be computed for each LIS faculty. The document-level scores will then be aggregated and propagated by entity levels such as author, program, and institution to incorporate the impact factors of those entities into the quality assessment process, while author-level scores will be used to weight the quality/impact of citations by the authors.

4.2 Document-level Measures

The simplest of the document-level quality evaluation measures is the count of citations to a publication. Although the citation count in itself is not the most reliable measure of publication quality, it is nevertheless the basic building block of all other citation-based measures. The citation count can be weighted to reflect the importance of citation sources, aggregated to estimate the research quality of a person, project, or organization, and propagated to compute the publication quality in a recursive manner.

One such proposed measure using a recursive algorithm is CiteRank, which is modeled after Google's PageRank (Page et al. 1998). As is the case with PageRank, CiteRank computes a global measure of a publication based on the aggregate of human-judged importance implied in each citation by employing a recursive formula that propagates the CiteRank scores forward through in-citation links of the entire citation link graph. CiteRank considers not only the count of citations but also the quality of each citing document, where publication quality is measured by the sum of the quality of its in-citations. This idea is captured in the CiteRank formula as follows:

$$R(p) = \sum_{i=1}^{k} \frac{R(p_i)}{C(p_i)}, \qquad (1)$$

where $C(p)$ is the number of out-citations of p, and p_i denotes the in-citations of p. $R(p)$ can be calculated iteratively, starting with all $R(p_i)$ values equal to 1 and repeating computations until the values converge.

The underlying assumption of CiteRank is the notion that a citation from publication p_i to publication p signifies the recommendation of p by the author of p_i. By aggregating all such recommendations recursively over the entirety of the citation link graph, where each recommendation is weighted by its importance and normalized by its outdegree, CiteRank arrives at an objective measure of quality from subjective determinations of quality scattered over the citation graph. By the same token, CiteRank can be said to measure a collective notion of quality.

Another proposed measure modeled after a link analysis method is CiteAuthority. CiteAuthority is based on Kleinberg's (1998) HITS (Hyperlink Induced Topic Search) algorithm, which considers both inlinks and outlinks to identify mutually reinforcing communities of "authority" and "hub" pages. Like HITS, CiteAuthority defines "authority" as a publication that is cited by many good hubs and defines "hub" as a publication that cites many good authorities. The CiteAuthority formula, which is identical to the HITS formula is shown below:

$$a(p) = \sum_{q \to p} h(q), \qquad (2)$$

$$h(p) = \sum_{p \to q} a(q). \qquad (3)$$

These equations define the authority weight $a(p)$ and the hub weight $h(p)$ for each publication p, where $p \to q$ denotes "publication p cites publication q". CiteAuthority is similar to CiteRank in that it estimates the quality of publication p based on the aggregate values of publications that cite p. CiteAuthority, however, differs from CiteRank in two major regards. First, it takes into account the contributions from both in-citations and out-citations to compute two mutually reinforcing measures of publications. Second, CiteAuthority is computed from a relatively small subset rather than the totality of the citation graph.

Unique to CiteAuthority is the premise that the universe of scholarly publications contains mutually reinforcing communities (i.e., hubs and authorities). To identify these communities, CiteAuthority will start with a root set S (e.g., the initial list of LIS publications), expand S to a base set T with the in-citations and out-citations of S a fixed number of times, eliminate self-citations in T to define the graph G, and run the iterative algorithm (equations 4 and 5) on G until convergence to compute $a(p)$ weights. The iterative algorithm works as follows: Starting with all weights initialized to 1, each step of the iterative algorithm computes $h(p)$ and $a(p)$ for every publication p in T, normalizes each of them so that the sum of the squares adds up to 1, and repeats until the weights stabilize. In fact it can be shown that the authority weights at convergence correspond to the principal eigenvalues of A^TA and hub weights correspond to those of AA^T, where A is the link matrix of the base set *T*.[3)]

4.3 Author-level Measures

The simplest of author-level evaluation measures is the count of publications by an author (i.e, publication count). As is the case with the citation count, the simple publication count does not differentiate among the quality of publications and thus reflect more of the productivity than the overall quality of an author's scholarly work. One simple solution to this problem is to make use of document-level citation scores when aggregating author's work. For instance, total citation count for an author, albeit simplistic, may be a better measure of a person's overall research quality than the publication count. By the same token, the sum of CiteRank (i.e., $\text{CiteRank}_{\text{author}}$), CiteAuthority (i.e., $\text{CithAuthority}_{\text{author}}$), or combination of both, may prove to be even better author-level measures.

Although the aggregation of document-level scores for an author takes into consideration the quality/impact of each publication, it may not necessarily capture the broad impact of an author's work as a whole. For example, the aggregation method will produce a higher score for an author with a single publication with 100 citations and nine publications with zero citations than another with 10 publications each of which are cited 9 times. To account for such scenarios and to add to the multi-faceted analysis, two additional author-level measures, named *h*-index and Mentor-index, will be introduced to the mix.

h-index, which is a relatively new measure designed to quantify the impact of individual scientists' research output (Hirsch 2005), is computed by ranking the publications of an author by citation counts and selecting the lowest publication rank where the rank is greater than or equal to the citation count. In other words, an *h*-index of 9 (the second author in the example above) means that the author has 9 publications with at least 9 citations each, whereas an H-index of 1 (the first author in the example) implies that the author in question has only one publication of any consequence. *h*-index, therefore, looks to the body of an author's work rather than a handful of high-impact publications to assess the overall impact of an author. Furthermore, *h*-index leverages the aspect of time in its computation since the *h*-index of an author will increase linearly with time as the author's research matures and he/she produces more quality publications.

Mentor-index is an index modeled after MPACT, which measures the mentoring impact of a faculty based on the mentoring activity data such as number of dissertations advised and number of dissertation committees served (Marchionini et al. 2006). Mentor-in-

3) The $(i,j)^{th}$ entry of A is 1 if there exists a link from page *i* to page *j*, and is 0 otherwise. In A^T, the transpose of the link matrix A, the $(i,j)^{th}$ entry of A corresponds to the link from page *j* to page *i*. The $(i,j)^{th}$ entry of AA^T gives the number of pages pointed to by both page *i* and page *j* (bibliometric coupling), while the $(i,j)^{th}$ entry of A^TA gives the number of pages that point to both page *i* and page *j* (co-citation).

dex applies the idea behind MPACT to citation analysis by aggregating students' research impact to measure the mentoring impact of a faculty. In the proposed project, Mentor-index of a faculty will be computed by summing up the author-level scores of the students that the faculty advised.

4.4 Propagation of Quality Measures

Document- and author-level quality measures are valuable in themselves since they estimate the impact/quality of a publication and author's work respectively. The mutually reinforcing relationship between publications and authors, however, is only leveraged once in the one-way computation of author-level measures. To fully propagate the quality assessment of authors in computing the document quality and vice versa, the investigators propose the following recursive propagation algorithm:

1. Compute citation-based quality scores (CQS) for each publication.
2. Compute CQS for authors using publication CQS.
3. Compute CQS for each publication weighted by author scores.
4. Compute CQS for authors using weighted publication CQS.
5. Repeat steps 3 and 4 until convergence.

The CQS propagation algorithm, which generates various weighted scores for each of the document-and author-level measures (e.g., weighted CiteRank, weighted $\text{CiteRank}_{\text{author}}$, etc.), can be extended beyond the author-level to include other entity levels such as school and publisher. Furthermore, the recursive propagation can be tempered by fixing the contribution of certain entity levels (e.g. publisher) with manual assessment scores to give more weights to direct and explicit human judgments for a given set of entities.

4.5 System Architecture

Construction of a comprehensive citation database via data integration and development of multi-faceted citation-based quality assessment measures, although immensely valuable on their own, do not necessarily guarantee the timely assessment of scholarly work. In fact, both the data integration and multi-faceted citation analysis are more or less insurmountable tasks without the aid of automated processes. The main objective of the *CiteSearch* system is to automate as much of the data integration and citation analysis as possible to facilitate efficient and effective assessment of scholarly work.

Given a publication title, for example, the *CiteSearch* system will automatically search multiple Web-based citation databases such as *EBSCO*, *Proquest* and *Google Scholar*, and analyze the search results to produce bibliographical metadata of all citations and compute various citation-based quality evaluation measures such as CiteRank, weighted CiteRank, which is CiteRank weighted by source, author, or time of citations. The initial citation metadata will then be aggregated and analyzed to produce meta-level citation measures for authors, publications, and schools.

5. Concluding Remarks

Despite the recent advances in citation analysis research, an effective and robust method for assessing the quality of scholarly work is yet to be developed. *Google Scholar*, despite its massive coverage and formidable resources, appears to be falling behind the likes of Thompson and Elsevier, who are proactive in their pursuit of the next-generation citation analysis approach. In the meantime, scholars are left more or less on their own under the whims of those who evaluate their work using inadequate measures based on incomplete data. CiteSearch approach is but an example of a viable alternative to vendor-based solutions to the citation analysis problem.

References

Aksnes, D. W. and R. E. Taxt. 2004. “Peer reviews and bibliometric indicators: A comparative study at a Norwegian university.” *Research Evaluation*, 13(1): 33-41.

Bakkalbasi, N., K. Bauer, J. Glover, and L. Wang. 2006. “Three options for citation tracking: Google Scholar, Scopus and Web of Science.” *Biomedical Digital Libraries*, 7.

Bar-Ilan, J. 2006. “An ego-centric citation analysis of the works of Michael O. Rabin based on multiple citation indexes.” *Information Processing and Management*, 42(6): 1553-1566.

Bar-Ilan, J. 2008. “Which h-index? - A comparison of WoS, Scopus and Google Scholar.” *Scientometrics*, 74(2): 257-271.

Bar-Ilan, J., M. Levene, and A. Lin. 2007. “Some measures for comparing citation databases.” *Journal of Informetrics*, 1(1): 26-34.

Bauer, K. and N. Bakkalbasi. 2005. “An examination of citation counts in a new scholarly communication environment.” *D-Lib Magazine*, 11(9).

Bergstrom, C. T. 2007. “Eigenfactor: Measuring the value and prestige of scholarly journals.” *College & Research Libraries News*, 68(5).

Cronin, B. 1984. *The Citation Process: The Role and Significance of Citations in Scientific Communication*. London: Taylor Graham.

Egghe, L. 2006. “Theory and practise of the g-index.” *Scientometrics*, 69(1): 131-152.

Giustini, D. 2006. “Blog posting at UBC Academic Search-Google Scholar Blog.” [cited 2011. 05.05]. <http://weblogs.elearning.ubc.ca/googlescholar/archives/025964.html>.

Hirsch, J. E. 2005. “An index to quantify an individual's scientific research output." *Proceedings of the National Academy of Sciences of the United States of America*, 102(46): 16569-16572.

Holden, G., G. Rosenberg, and K. Barker. 2005. "Bibliometrics: A potential decision making aid in hiring, reappointment, tenure and promotion decisions." *Social Work in Health Care*, 41(3-4): 67-92.

Holmes, A., and C. Oppenheim. 2001. "Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: Library and Information Management." *Information Research*, 6(2).

Jacsó, P. 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.

Kleinberg, J. 1998. "Authoritative sources in a hyperlinked environment." *Proceeding of the 9th ACM-SIAM Symposium on Discrete Algorithms*, 668-677.

Kousha, K., and M. Thelwall. 2006. "Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines." In *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium: 72-73.

Lewison, G. 2001. "Evaluation of books as research outputs in history of medicine." *Research Evaluation*, 10(2): 89-95.

MacRoberts, M. H., and B. R. MacRoberts. 1996. "Problems of citation analysis." *Scientometrics*, 36(3): 435-444.

Marchionini, G., P. Solomon, C. Davis, and T. Russell. 2006. "Information and library science MPACT: A preliminary analysis." *Library & Information Science Research*, 28(4).

Martin, B. R. 1996. "The use of multiple indicators in the assessment of basic research." *Scientometrics*, 36(3): 343-362.

Meho, L. I., and K. Yang. 2007. "Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar." *Journal of the American Society for Information Science and Technology*, 58(13): 2105-2125.

Moed, H. F. 2005. *Citation Analysis in Research Evaluation*. Springer, Berlin.

Nisonger, T. E. 2004. "Citation autobiography: An investigation of ISI database coverage in determining author citedness." *College & Research Libraries*, 65(2): 152-163.

Norris, M., and C. Oppenheim. 2007. "Comparing alternatives to the Web of Science for coverage of the social sciences' literature." *Journal of Informetrics*, 1(2): 161-169.

Noruzi, A. 2005. "Google Scholar: The New Generation of Citation Indexes." *Libri*, 55(4): 170-180.

Seglen, P. O. 1998. "Citation rates and journal impact factors are not suitable for evaluation of research." *Acta Orthopaedica Scandinavica*, 69(3): 224-229.

van Raan, A. F. J. 1996. "Advanced bibliometric methods as quantitative core of peer-review based evaluation and foresight exercises." *Scientometrics*, 36(3): 397-420.

Vaughan, L., and D. Shaw. 2008. "A new look at evidence of scholarly citation in citation indexes and from web sources." *Scientometrics*, 74(2): 317-330.

West, J. D., C. T. Bergstrom, and T. C. Bergstrom. 2010. "Big Macs and Eigenfactor Scores: Don't Let Correlation Coefficients Fool You." [online]. [cited 2011.05.05]. <http://arxiv.org/abs/0911.1807>.