

Citations to the “Introduction to informetrics” indexed by WOS, Scopus and Google Scholar

Judit Bar-Ilan

Received: 11 May 2009 / Published online: 17 February 2010
© Akadémiai Kiadó, Budapest, Hungary 2010

Abstract Google Scholar and Scopus are recent rivals to Web of Science. In this paper we examined these three citation databases through the citations of the book “Introduction to informetrics” by Leo Egghe and Ronald Rousseau. Scopus citations are comparable to Web of Science citations when limiting the citation period to 1996 and onwards (the citation coverage of Scopus)—each database covered about 90% of the citations located by the other. Google Scholar missed about 30% of the citations covered by Scopus and Web of Science (90 citations), but another 108 citations located by Google Scholar were not covered either by Scopus or by Web of Science. Google Scholar performed considerably better than reported in previous studies, however Google Scholar is not very “user-friendly” as a bibliometric data collection tool at this point in time. Such “microscopic” analysis of the citing documents retrieved by each of the citation databases allows us a deeper understanding of the similarities and the differences between the databases.

Keywords Citations · Web of Science · Scopus · Google Scholar · Introduction to informetrics

Introduction

Until the fall of 2004 there was only a single, comprehensive citation data base: ISI’s (Thomson Reuters) Web of Science (WOS), the Web based version of the Citation Indexes established by Eugene Garfield. In the fall of 2004 two new citation databases appeared: Elsevier’s Scopus and the freely available Google Scholar (GS). Several previous studies compared and evaluated these citation databases, but most of them concentrated on citation counts and coverage. One exception is Meho and Yang (2007) where the citations were carefully cleansed to remove duplicates and citations from non peer-reviewed material retrieved by Google Scholar. In this study we also analyze citations and not only citation counts. The chosen source item was “Introduction to Informetrics: Quantitative methods in

J. Bar-Ilan (✉)

Department of Information Science, Bar-Ilan University, Ramat Gan 52900, Israel
e-mail: barilaj@mail.biu.ac.il

Library, Documentation and Information Science” by Leo Egghe and Ronald Rousseau, and the citation data were retrieved from WOS, Scopus and Google Scholar. The aim of the study was to analyze the quality and accuracy of citation data produced by Google Scholar and to examine the overlap between the three databases in terms of citations to a book that is not indexed either by WOS or by Scopus.

Literature review

Although Google Scholar exists for 4 years only, there are already a large number of evaluations of it. At first it was received with mixed feelings. “Although there are no detailed studies, many librarians report that faculty members and students are beginning to use the search engine; some suspect that Scholar will replace more established, and more costly, search tools. It already directs more online traffic to Nature websites than any other multidisciplinary science search engine.” (Giles 2005, p. 554). Librarians seem to be less enthusiastic than their clients: in the summer of 2005 only a minority (24%) of the 113 ARL library sites linked to Google Scholar (Mullen and Hartman 2006). By 2007 the situation changed considerably, and 68% of the ARL libraries listed Google Scholar on the alphabetical list of indexes and databases. In 2005, 43 libraries had link resolvers, providing direct access to their resources through Google Scholar, and this number had more than doubled by 2007 (93 out of the 113 libraries, 82%) (Hartman and Mullen 2008).

A number of studies assessed citation counts on Google Scholar versus the Web of Science. Bauer and Bakalbassi (2005) in an early study compared citation counts reported by WOS and GS for JASIST articles published in 1985 and 2000. The results were inconclusive for 1985, but for 2000, GS counts were considerably higher than WOS counts. Jacso (2006) criticised the fact that the media only highlighted the findings for 2000, ignoring the inconclusive results for 1985.

Jacso emphasized and demonstrated the pros and cons of Google Scholar in a series of articles (e.g. Jacso 2006; Jacso 2008). Contrary to his opinion, Harzing and Van Der Wal (2008, p. 61) see Google Scholar “as a new source for citation analysis”. Anne-Wil Harzing together with Tarma Software research developed a very useful software utility, called Publish-or-Perish (<http://www.harzing.com/pop.htm>) for analyzing Google Scholar data. In an early study Neuhaus et al. (2006) studied the coverage of Google Scholar versus 47 different databases as of 2005. Since Google Scholar is much more dynamic than the more traditional databases, it is unclear whether the findings are still applicable as of 2008.

Noruzi (2005) compared the Web of Science and Google Scholar for most cited papers in Webometrics and concluded that Google Scholar is “a free alternative or supplement to other citation indices. Meho and Yang (2007) also compared WOS, Scopus and Scholar on citation in the field of information science, they looked for the citations to works of 15 faculty members of the School of Library and Information of the Indiana University at Bloomington. They found that Scopus data significantly alters rankings based on WOS only, and both Scopus and GS compliment WOS. They also noted the difficulties in working with GS. Bar-Ilan (2006a) calculated the h-indices of Price-medalists when the computation is based on WOS and GS and found that the h-indices based on GS are higher.

Bakalbassi et al. (2006) compared the citation counts of WOS, Scopus and Google Scholar for several journals in oncology and in condensed matter physics. They were not able to single out the “best” citation database, and concluded that the choice of the “best” tool was dependent on the discipline and on the year of publication. Kousha and Thelwall (2007) found significant correlations between Google Scholar and WOS citation counts in

all disciplines studied (biology, chemistry, physics, computing, sociology, economics, psychology and education). The correlations were between 0.551 and 0.825. For computer science and the social sciences the citation counts of GS were considerable higher than the WOS citation counts. The study was based on Open Access journals. Kousha and Thelwall (2008) further studied the unique citing sources indentified by GS of 882 open access articles from 39 open access journals in four disciplines: biology, chemistry, physics and computer science. For biology and chemistry the majority of the unique citation sources were journals, while for physics the dominant type was eprints, and for computer science proceedings; showing large disciplinary differences between types of citing documents.

There are already studies that use Google Scholar as the only source of citation data. For example Rahm and Thor (2005) used GS to compare the impact for two major proceedings and three journals on databases in the field of computer science. It should be noted, that because WOS had almost no coverage of proceedings and proceedings are the primary publication venue in computer science (see for example Moed and Visser 2007 or Bar-Ilan 2006b), WOS is a less useful source for computer science. Although Scopus has a better coverage of proceedings, still for researchers in computer science, Google Scholar had considerably higher citation counts, at least in 2006 (Bar-Ilan 2008). Note that as of October 2008, when searching the Web of Science, data from Conference Proceedings Citation Indexes of Thomson Reuters are incorporated provided the user has subscriptions to both databases.

Norris and Oppenheim (2007) considered several alternatives to Web of Science for coverage in the social sciences. The compared databases included Scopus and Google Scholar. They found that Scopus was comparable to WOS in terms of coverage from 1996 and onwards, but Google Scholar "cannot be seriously thought of as a database from which metrics could be used to measure scholarly activity" (p. 168).

In a recent study MEHO and ROGERS (2008) compared WOS and Scopus in the area of human-computer interaction and found that Scopus provides significantly better coverage due to indexing ACM and IEEE proceedings. Results were also compared with data retrieved through Google Scholar, and very high correlations were found between GS and combined WOS-Scopus rankings.

Gavel and Iselid (2008) studied the journal title overlap of Scopus and Web of Science and found that 84% of WOS titles are indexed by Scopus and 54% of Scopus titles are indexed by WOS. Note that these results were correct at the date of the study, WOS and Scopus continuously change (mainly extend) their journal coverage. Moed and Visser (2008) also compared the coverage of WOS and Scopus but on a paper-by-paper basis. They found that 89% of 1996 publications indexed by WOS are also indexed by Scopus, and for 2005 the percentage reaches 95%. Scopus has larger coverage than WOS for recent publications, but the additionally covered journals tend to have lower impact at least in oncology, as shown in (López-Illescas et al. 2008).

Moed and Visser (2008) compared WOS and Scopus on a paper-by-paper basis; here we compare WOS, Scopus and GS on a citation-by-citation basis. We not only compare citation counts, but study the overlap between the citations and characterize the unique citations appearing in each database.

Data collection

Citation data were collected from the three databases on March 22, 2008. Note that neither WOS nor Scopus index "Introduction to Informetrics: Quantitative methods in Library,

Documentation and Information Science” as a source item, because these databases do not index books.

Citations to non-source items can be found on WOS only through cited reference search. Non-source items are only indexed by the first author of the publication, and there is no canonical form of the title of the publication. Thus on WOS, data were collected through “cited reference search” of the publications of Egghe L and 259 citations were identified under 27 variations of the book’s title (see Fig. 1 for some of these variations).

On Scopus we searched for documents by the author Egghe, L and on the search results page we clicked on “More” (this option displays citations to items not indexed by Scopus). Here we found 13 title variations and 218 citations to these variations. The same results are obtained when searching for non-source item citations of Rousseau, R i.e., Scopus non-source items are not attributed to the first author only. The Scopus search is illustrated in Fig. 2.

The search method on Google Scholar is much more fuzzy, there are no clear rules or instructions. We were trying to get exhaustive answers, thus we experimented with a number of variations of the book title. These were the most frequently occurring variations on WOS and Scopus. We searched for three variations of the title (with and without the

<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS	1990		2
<input type="checkbox"/>	EGGHE L	INTRO INFOMETRICS QU	1990		5
<input type="checkbox"/>	EGGHE L	INTRO INFORM	1990		4
<input type="checkbox"/>	EGGHE L	INTRO INFORM QUANTIT	1990		14
<input type="checkbox"/>	EGGHE L	INTRO INFORMATICS	1990		1
<input type="checkbox"/>	EGGHE L	INTRO INFORMATICS QU	1990		1
<input type="checkbox"/>	EGGHE L	INTRO INFORMATICS	1990		1
<input type="checkbox"/>	EGGHE L	INTRO INFORMATICS Q	1990		1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS QU	1990		2
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRIC QU	1990		1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS	1990		95
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS	1990	263	3
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS	1990	357	1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS	1990	380	1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS	1990	450	1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS	1990	1990	1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS A	1990		1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS Q	1990		119
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS Q	1990	211	2
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS Q	1990	220	1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS Q	1990	222	1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICS Q	1990	374	1
<input type="checkbox"/>	EGGHE L	INTRO INFORMETRICZXS	1990		1

Fig. 1 Cited reference search on the Web of Science

The screenshot shows a Scopus search interface. At the top, it says 'Scopus - More Search Results: AUTHOR-NAME(Egghe I) - Mozilla Firefox'. The search query is 'AUTHOR-NAME(Egghe I)'. There are tabs for 'Scopus (124)', 'More: 123', 'Web (0)', and 'Patents (0)'. Below this, it says 'Results from references for your query: AUTHOR-NAME(Egghe I)'. A 'Refine Results' section contains a table with columns 'Source Title', 'Author Name', and 'Year'. The 'Source Title' column lists 'Journal of the American Society for Information Science (14)', 'Scientometrics (8)', and 'Journal of the American Society for Information Science and Technology (6)'. The 'Author Name' column lists 'Egghe, L. (121)', 'Rousseau, R. (53)', and 'Ravichandra Rao, I.K. (4)'. The 'Year' column lists '2008 (3)', '2007 (6)', and '2006 (11)'. Below the 'Refine Results' section, it says 'Results: 123'. There are options to 'Output' and 'Cited by', and a 'Select' dropdown set to 'All'. A table of results is shown with columns 'Document (sort by relevance)', 'Author(s)', 'Date', 'Source Title', and 'Cited By'. The first three results are:

Document (sort by relevance)	Author(s)	Date	Source Title	Cited By
1. [No title available]	Egghe, L., Rousseau, R.	1990	Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science	132
2. [No title available]	Egghe, L., Rousseau, R.	1990	Introduction to Informetrics	74
3. [No title available]	Egghe, L.	2005	Power Laws in the Information Production Process: Lotkian Informetrics	53

Fig. 2 Scopus search for non-source items

authors): “Introduction to informetrics”; “Introduction to informetrics: Quantitative methods in library, documentation and information science” and “Quantitative methods in library, documentation and information science”, browsed the search results to identify relevant records and collected 358 items that supposedly referred to the book. Each item was downloaded to ascertain that there is a reference to the book and in order to eliminate duplicates. Figure 3 displays the top results when searching for “Introduction to informetrics”. We will use this specific example to describe the basic data cleansing that had to be carried out.

- 1) The first entry on the Google Scholar list is the book with its full title and it is said to have been cited 227 times. When clicking on the “cited by” link, only 219 items were retrieved (see Fig. 4).
- 2) Among the 219 citing items identified for this source item, there were two URL duplicates; exactly the same item residing at the same URL was listed twice as citing the book. This is a clear over count of the number of citations, that but for all cases in this study, these kinds of duplicates occurred very rarely.
- 3) GS listed the short title as a different source item with 134 additional citations. In this case it displayed 129 citations. There were 4 items with exactly the same URL that were listed both as citing items of the full title of the book and citing items of the short title. This is again an over count of citations, but this happened in only a very few cases. It should be noted that GS reported another 6 citations under the title variation “Introduction to Informetrics: Quantitative Methods in Library” (this was the fifth

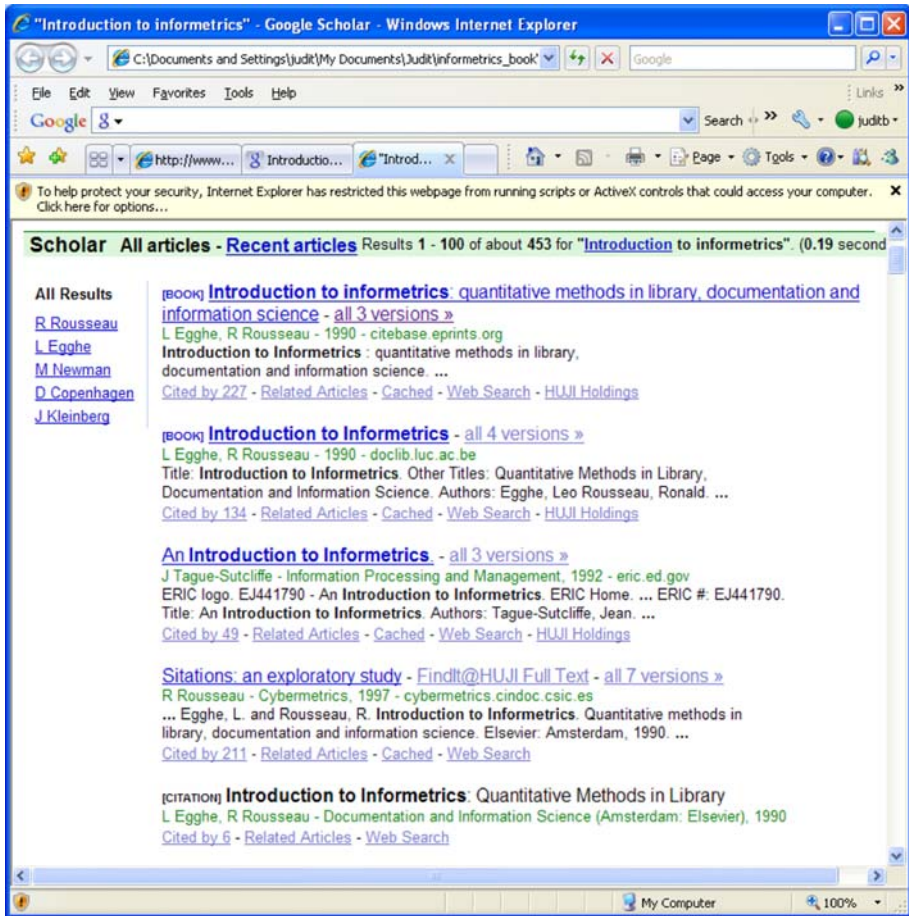


Fig. 3 Google Scholar search results for “Introduction to informetrics” as of March 21, 2008

result for the query displayed in Fig. 3). In this case, however, 4 out of the 6 citing items were already listed (same URLs) as citing documents of the first item (i.e. among the 219 citations retrieved for the full title).

- 4) A third source of over count of citation is caused by content duplicates, i.e. the same citing item listed under different names or residing at different URLs (GS makes efforts to avoid this and applies extensive unification algorithms). In the data collection process we identified 358 citing items (after the removal of the URL duplicates). Out of these 24 items were (6.7%) were content duplicates. A closer examination shows that there were only two cases where both citing items had exactly the same title and author list (two copies of the publication residing at two different URLs); there were 17 cases where GS gave different titles to the same publication (in at least one of the cases the title was defined incorrectly, or the subtitle was missing), and in 5 cases the author lists were wrong in at least one of the instances. See for example Fig. 5, the authors listed here are J Wiley and I Sons, where in reality the paper was written by Leo Egghe. Mistakes of this kind were noticed

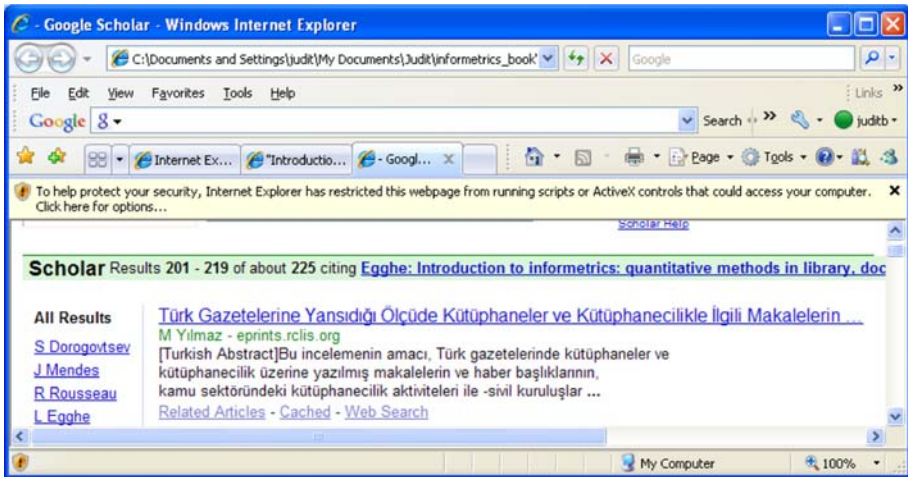


Fig. 4 The last result page when clicking on the citations of the first item in Fig. 3

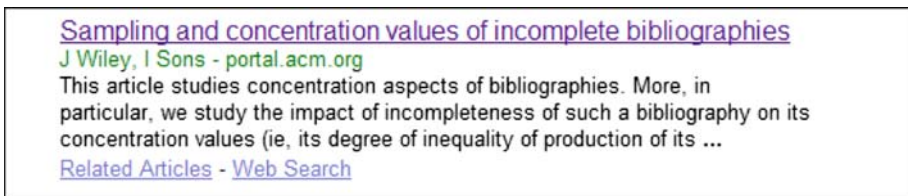


Fig. 5 An example illustrating incorrect author and source identification

earlier by Jacso (2006). As of November 2008, “J Wiley” is still a very prominent author on GS with 6,490 publications.

- 5) In addition to duplicate removal, all citing documents were checked. There were 11 inaccessible documents, 5 items that were not scientific publications (a project proposal, a course proposal, a course handout, a lecture abstract and a citation format example).

To sum up out of the total 358 potential citing items that were identified after the removal of URL duplicates; 307 (85.8%) were distinct scientific publication that cited the book. This relatively high accuracy for the specific case is rather surprising in light of the “complaints” of Jacso (2006) and Meho and Yang (2007). We of course cannot make conclusion about “deflated, inflated and phantom citation counts” (Jacso 2006, p. 297) in general, but for the specific case, the accuracy of the list of the citing documents retrieved by GS was quite satisfactory.

An interesting point to note about Google Scholar is that citation counts are not necessarily monotonically increasing over time. This may be due to corrections made to the database or to other unknown factors. As an example, consider “Sititions: an exploratory study” by Ronald Rousseau. GS reported 211 citations on March 21, 2008 (see the fourth item on Fig. 3). However on August 23, 2008, GS reported only 204 citations (see Fig. 6).

In addition to the basic data collection method, we tried to locate each item that was retrieved by one database but not by the other(s) in each of the databases. In most cases we

[PDF] ► [Sitations: an exploratory study - FindIt@HUJI Full Text](#)
 R. Rousseau - *Cybermetrics*, 1997 - [dialnet.unirioja.es](#)
 ... & Egghe, eds.), (1997) 71- 79. Egghe, L. and Rousseau, R. **Introduction to Informetrics**. Quantitative 7 one with seven self-citations ...
[Cited by 204](#) - [Related articles](#) - [View as HTML](#) - [Web Search](#) - [All 11 versions](#)

Fig. 6 Decrease in the number of citations between March and August 2008

searched using the titles of the citing item. All unique items found through Google Scholar were downloaded (with the exception of the 11 inaccessible items) and were checked to ascertain that the item referenced the “Introduction to informetrics” and that it was not identical to any of the other items retrieved by Google Scholar.

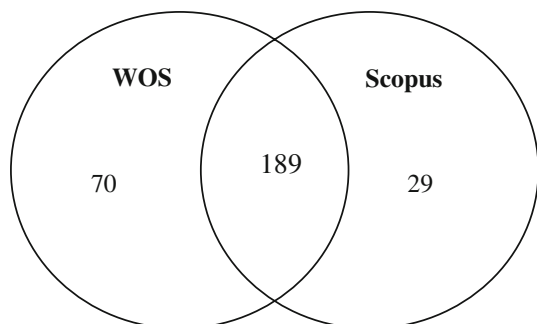
Results and discussion

First we provide a detailed comparison of WOS versus Scopus citations. The Web of Science and Scopus together located 288 distinct citations of “Introduction to informetrics”. The Web of Science listed 259 out of these 288 citations (89.0%) and Scopus listed 218 of the 288 citations (75.7%). Figure 7 displays the distribution of the citing items for the Web of Science and Scopus.

Next we examine the unique citations in WOS. Note that the book “Introduction to informetrics” was published in 1990 and Scopus provides consistent citation data only from 1996 and onwards.

- 46 out of the 70 unique items (65.7%) were published before 1996.
- Out of the 24 remaining unique items, 8 (33.3%) appeared in journals and for some reason were not indexed by Scopus.
- Two items were from JASIST, published in 2008, these items have been indexed by now. A third item from JASIST was published in 1998—it is a rejoinder, which might explain why Scopus did not index it. One item was an article published in 2005 in *Scientometrics*—this is obviously a mistake in the indexing process. Another item was from *Libri* 1996—Scopus indexes *Libri*, but for some reason it only indexed the first two issues in 1996, while this item appeared in the fourth issue that year. *Journal of Statistical Mechanics Theory and Experiment* is indexed by Scopus from 2005, and the specific item was published in 2004. One item was published in the *ACM Computing*

Fig. 7 The distribution of the citing items for the Web of Science and Scopus



Surveys in 1999. This source title is indexed by Scopus, but the specific item appeared in the supplement which might explain why the item was not covered by Scopus.

- Three items appeared in series not indexed by Scopus (4.3%),
- Nine items were indexed by Scopus, but were indexed without the list of references, i.e., in the indexing process the references were left out by mistake (12.9%).
- Four items were indexed with the references by Scopus, but did not appear among the citations of the book when using the "More" option of Scopus for publications of Leo Egghe.

The 29 unique items found by Scopus can be characterized as follows:

- 18 out of the 29 unique items were published in journals that were not indexed by WOS as of March 2008 (62.1%). The journals are: *Journal of Informetrics* (it is now indexed by WOS); *Information-Wissenschaft und Praxis* (indexed by WOS until 2003, items from 2006); *Information Research* (indexed by WOS from 2003, the specific item was published in 2001), *Research Evaluation* (indexed from 2000, the specific item was published in 1999), *Asia Pacific Journal of Management* (indexed by WOS from 2008), *BMC Medical Research Methodology* (indexed by WOS from 2008), *Education for Information* (indexed by WOS until 1997, the specific item was from 1998), *Advances in Library Administration and Organization*, *Archival Science and Libres*.
- Six items were published in proceedings (20.7%). One of these appeared in a proceedings series also indexed by WOS: *Lecture Notes in Computer Science*, but for some reason the specific item was not indexed.
- Three items are indexed by WOS as well, but without their reference list, and thus these items do not show up when searching for citations of "Introduction to informetrics" (10.3%).
- Two additional items were indexed by WOS, but there were mistakes in indexing the reference list and thus the citations were not attributed to the book (in one case the reference was attributed to Eggle instead of Egghe).

To sum up, the overlap between the two databases is very high, especially if we only consider the 242 unique citations that appeared in items published after 1996. We also see some minor errors in the indexing processes of both databases. Both databases constantly update their source lists, thus the results presented here reflect the situation as of March 2008.

When adding the citing documents identified by Google Scholar as well, the total number of unique citations becomes 397. The overlap between the three citation databases is displayed in Fig. 8.

First let us take a closer look at the 90 items that were retrieved by WOS and/or Scopus, but were missed by Google Scholar.

- More than half of the 90 items (46, 51.1%) were indexed by GS, and GS had access to the full text of the items, thus should have extracted the references to the book and these should have appeared as citations to the book.
- Another 29 items (32.2%) were indexed by Google Scholar, but there was no apparent access to the full text, and thus GS was not necessarily aware of the references appearing in these items (examples of such cases are items indexed by *cat.inist* (<http://www.cat.inist.fr/>), a free bibliographic service by INIST/CNRS, or items indexed by the ERIC database (<http://www.eric.ed.gov/>).
- Only 15 out of the 90 items (16.7%) had no traces in Google Scholar.

Fig. 8 The distribution of the citing items for the WOS, Scopus and GS

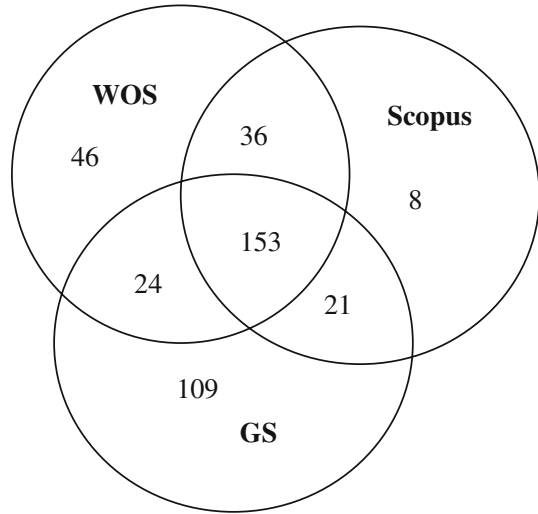


Table 1 Document types of the unique items retrieved by Google Scholar

Document type	No. items	% out of total (109)
Journal	28	25.7%
Proceedings	25	22.9%
Thesis	15	13.8%
Book chapter	13	11.9%
Report	10	9.2%
Manuscript	7	6.4%
In Chinese	4	3.7%
Book	3	2.8%
Newsletter	2	1.8%
Encyclopedia entry	1	0.9%
Patent	1	0.9%

Figure 8 shows that Google Scholar covered 68.8% of the items identified by WOS and/or Scopus, but if we add the additional items that GS should have been able to retrieve, the percentage increases to 84.7% (244 items). Its coverage of the items published before 1996 is rather low, it found only 8 out of the 46 items indexed by WOS as citations of the book, and had access to the full text of 14 others, but for some reason were not retrieved as citing items.

Next we analyze the 109 items uniquely found by Google Scholar. The basic results are displayed in Table 1.

Let us concentrate on the 28 journal publications indexed by GS but not covered by WOS or Scopus. The journal *Ciência da Informação* was the source of seven of the citing items, and *Cybermetrics* and *Glottometrics* were the sources of two citing items each. Out of the 20 different journals eight are being indexed by Scopus (only one for the period in which the citing item was published) and one is also indexed by WOS (it was not indexing the journal at the time of data collection). Based on the homepages of the journals, 18 of them are peer reviewed.

Among the 25 conference papers, 4 were published in the proceedings of the Cuban Congreso Internacional de Información, 3 in the Turkish Proceedings Değişen Dünyada Bilgi Yönetimi Sempozyumu, 2 each from Collnet and from the ACM SIGMOD-SIGACT-SIGART Conference series.

Even if we discount all 43 citations from books, theses, reports, manuscripts, newsletters, encyclopedias, patents, and from non-peer-reviewed journals that Google Scholar identified, still there are 66 scholarly items that cited “Introduction to informetrics” that were not indexed by either Scopus or WOS. Notable is the wider coverage of Google Scholar of publications in non-English languages. We located publications in Spanish, German, French, Italian, Turkish, Portuguese and Chinese and Croatian.

To sum up, if we discard the 43 citations mentioned in the previous paragraph from GS, then the three databases altogether retrieved 354 unique citations. Based on the number of unique citations, we can rank the databases as follows: Google Scholar 264 citations (74.6%), WOS—259 citations (73.2%) and Scopus—218 citations (61.6%).

Another possible calculation is to consider only the publications from 1996 and onwards, because Scopus provides citation data only for this period. In this case we have a total of 307 citations. Google Scholar indexes 278 items (90.6%), Scopus 218 items (70.0%) and WOS 213 (69.4%).

If we limit ourselves to the 242 citing items published from 1996 and onwards that are indexed by WOS and/or Scopus (because these two apply some kind of quality control, whereas there are no guidelines as to what is indexed by Google Scholar); then Scopus indexes 218 (90.1%), WOS indexes 213 of them (86.6%), and GS indexes 190 (78.5%). However if we take into account also the items that Google Scholar indexed, and had access to their fulltext, but for some reason were not retrieved as citing items, then there are 222 items identified by Google Scholar, which gives it a coverage of 91.7%.

Limiting the citation sources to journal papers only, results in 303 citing items, out of which 255 (84.2%) were identified by WOS, 215 (71.0%) by Google Scholar and 211 (69.6%) by Scopus. If we consider the citations coming from journal articles only that were published after 1995, then the total number of these citations is 257. Scopus retrieved 211 citations (82.1%), the Web of Science retrieved 209 (81.3%) citations and Google Scholar 208 citations (80.9%).

We see that different methods result in different rankings. Overall all three databases had comparable performance for the specific case studied in this paper, but it was considerably more difficult to extract the information from Google Scholar than from the other two databases.

Conclusion

The results clearly show that the databases supplement each other. Our study supports previous findings (e.g. Bakalbassi et al. 2006 or Meho and Yang 2007) that currently there is no single citation database that can replace all the others. This point is further illustrated when we consider the different ranking methods demonstrated above, when considering the total number of citations, GS is the “best”; when looking at journal citations only, the Web of Science is “best” and when considering citations from 1996 and onwards that are indexed by either WOS or Scopus, Scopus becomes the “leader”.

The coverage of WOS and Scopus is quite comparable, at least for the citations of the “Introduction to informetrics”, when limiting the publication year to 1996 and onwards. Google Scholar’s coverage was surprisingly good, and its accuracy was also better than

expected. Of course one cannot generalize based on a case study, however the study allowed us to investigate the subtleties of data collection and data cleansing in each of the three citation databases: the Web of Science, Scopus and Google Scholar.

An additional finding is that one has to be rather resourceful and ready to invest time while searching for a comprehensive set of citations to a given publication.

References

- Bakalbassi, N., Bauer, K., Glover, J. & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Research Libraries* 3(7).
- Bar-Ilan, J. (2006a). H-index for Price medalists revisited. *ISSI Newsletter*, 2(1), 3–5.
- Bar-Ilan, J. (2006b). 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.
- Bauer, K., Bakalbassi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9). <http://www.dlib.org/dlib/september05/bauer/09bauer.html>.
- Gavel, Y., & Iselid, L. (2008). Web of Science and Scopus: A journal title overlap study. *Online Information Review*, 32(1), 8–21.
- Giles, J. (2005). Start your engines. *Nature News*, 438, 554–555.
- Hartman, K. A., & Mullen, L. B. (2008). Google Scholar and academic libraries: An update. *New Library World*, 109(5/6), 2122–2222.
- Harzing, A. K., & Van Der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, 6, 61–73.
- Jacso, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review*, 30(3), 297–309.
- Jacso, P. (2008). Google Scholar revisited. *Online Information Review*, 32(1), 102–114.
- Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055–1065.
- Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics*, 74(2), 273–294.
- López-Illescas, C., Moya-Anegón, F., & Moed, H. F. (2008). Coverage and citation impact of oncological journals in the Web of Science and Scopus. *Journal of Informetrics*, 2, 304–316.
- Meho, L. I., & Yang, K. (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.
- Meho, L. & Rogers, Y. (2008). Citation counting, citation ranking and h-index of human-computer interaction researchers: A comparison between Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11), 1711–1726.
- Moed, H. F. & Visser, M. (2007). Developing bibliometric indicators of research performance in computer science: An exploratory study. CWTS Report 2007-01.
- Moed, H. F. & Visser, M. (2008). Comparing Web of science and Scopus on a paper-by-paper basis. In *Book of Abstracts of the 10th International Conference on Science and Technology Indicators*.
- Mullen, L. B., & Hartman, K. A. (2006). Google Scholar and the Library Web Site: The early response by ARL Libraries. *College & Research Libraries*, 67(2), 106–122.
- Neuhaus, C., Neuhaus, E., Asher, A., & Wrede, C. (2006). The depth and breadth of Google Scholar: An empirical study. *Portal: Libraries and the Academy*, 6(2), 127–141.
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, 1, 161–169.
- Noruzi, A. (2005). Google Scholar: A new generation of citation indexes. *Libri*, 55(4), 170–180.
- Rahm, E., & Thor, A. (2005). Citation analysis of database publications. *ACM SIGMOD Record*, 34(4), 48–53.