International Journal of Scientometrics, Informetrics and Bibliometrics
ISSN 1137-5019

VOLUME 13 (2009): ISSUE 1. PAPER 4

A Closer Look at the Sources of Informetric Research

Judit Bar-Ilan[1]

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

Abstract

Currently existing data sources for informetric research are far from being perfect. Some of the imperfections are caused by uneven coverage, errors or changes in indexing policies that are often not retroactive or by mistaken or ineffective retrieval strategies employed by the users. Being aware of the limitations of the data sources and a closer inspection of the data we work with can improve the validity and interpretation of our findings. In this paper I discuss current limitations of several data sources, emphasize the ever-changing nature of these sources and recommend trying to understand the specific problems and limitations at the time the study is conducted instead of relying on methods recommended in previous studies.

Keywords

informetrics, citation databases, WOS, Scopus, Google Scholar

Introduction

We are living in the "information age": incredible amounts of information are available to us through the Internet. The Web has existed for twenty years only, yet the large majority of the data sources for informetric research are available through the Web. ISI's Web of Science (now a Thomson Reuters company) was launched in 1997 (Thomson, 2007), before that ISI data were only available through commercial providers (e.g. Dialog and STN), on tapes or CDs (from 1989 and onwards), or in the "ancient times" in print. In November 2004 two additional major citation databases appeared on the Web: Elsevier's Scopus (2004) and Google Scholar (Acharaya, 2004).

Not only the citation databases are online, but all major scientific journals appear now in electronic format beside the traditional printed version. There are already well-established journals that appear in electronic format only. This trend has begun in the late 1990's (Elsevier, 2009), and by now the publishers have digitized many volumes that originally appeared in print only. And of course, one cannot ignore the astronomical amounts of "digitally-born" data on the Web, which also include valuable information for informetric research in general and specifically for webometrics. Thus electronic access to data has become the norm. The computing power and the storage capabilities have also increased by several orders of magnitude over the last two decades, and there are easily accessible and often open-source software tools that enable to collect and analyze large quantities of data even on a personal computer. It has become easy to conduct "desktop or poor-man's bibiliometrics" (Moed, 2009). The data for informetric research have never been perfect, but now that informetric analysis can be conducted with much greater ease than before, it is even more important to understand the limitations and problems of data sources and methods and to assess the validity of the results. In the following sections I discuss some limitations of the existing sources. Often there are no easy solutions to overcome the problems, but by being aware of their existence one can provide better interpretations of the research findings.

The citation databases

The citation databases are major sources of informetric research. Each has specific limitations, and because the indexing and retrieval policies of the databases change from time to time, and the changes are not necessarily retroactive, these changes may cause internal inconsistencies in the databases. Retroactive changes in the indexing policies are often not feasible, and thus new features are often applied from a certain point in time and onwards. In addition, when using multiple databases, either for more comprehensive data collection or for comparison, one must be aware of the differences in the applied algorithms and policies. In the following I give a few examples of these problems.

The Web of Science and the Journal Citation Reports

A lot has been written on the ISI Citation Indexes and the way ISI computes the journal impact factor. In this paper we will discuss the web-based product, the Web of Science (WOS). In previous works, coverage is one of the main reasons of criticism of the ISI Citation Databases: poor coverage of non-English publications, insufficient coverage of the social sciences and poor coverage of the arts and humanities (see Moed, 2005, chapter 7 for an extensive discussion of coverage by discipline).

An additional issue related to coverage is the date the database started to cover certain publications. In the third quarter of 2008, ISI integrated the Proceedings Indexes into WOS, but proceedings are indexed only from 1990 and onwards, whereas journals in the Science Citation Index are indexed from 1900, in the Social Science Citation index from 1956, and in the Arts & Humanities Citation Index from 1975 and onwards (Thomson-Reuters, 2009). Thus the coverage of publications of active researchers is non-uniform � their works before 1990 are only covered if they appeared in journals. This of course is given and cannot be changed, but this must be taken into account, especially in areas where proceedings are an important publication venue, for example in computer science (Bar-Ilan 2006 and 2009).

As an example, let us consider the ISSI conferences (International Conference on Scientometrics and Informetrics). Proceedings of four conferences are indexed: 1999 (Colima), 2001 (Sydney), 2005 (Stockholm) and 2007 (Madrid). The conferences that took place before 1999 and the 2003 conference in Beijing are not indexed. There is no uniform name for the conference series, three of them can be found when looking for ISSI in the publication name, but the 1999 proceedings is not, but a search for "scientometrics and informetrics" when searching in the publication list page works (see Figure 1). Of course, this can change in the future, but at the time of writing the coverage of this conference series was not uniform. Providing uniform names to all the proceedings in a conference series is highly desirable.

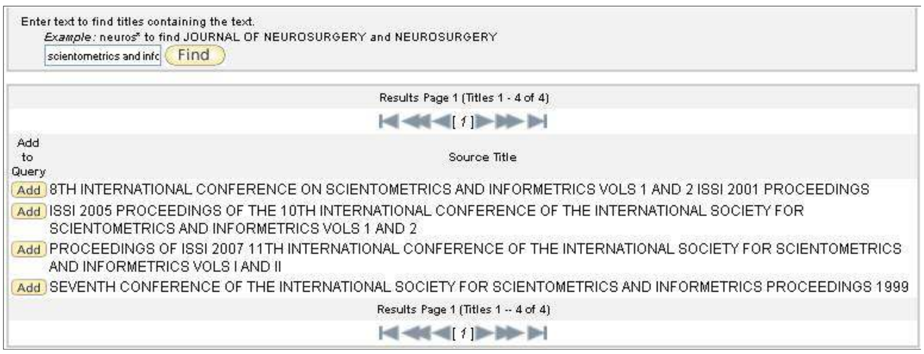
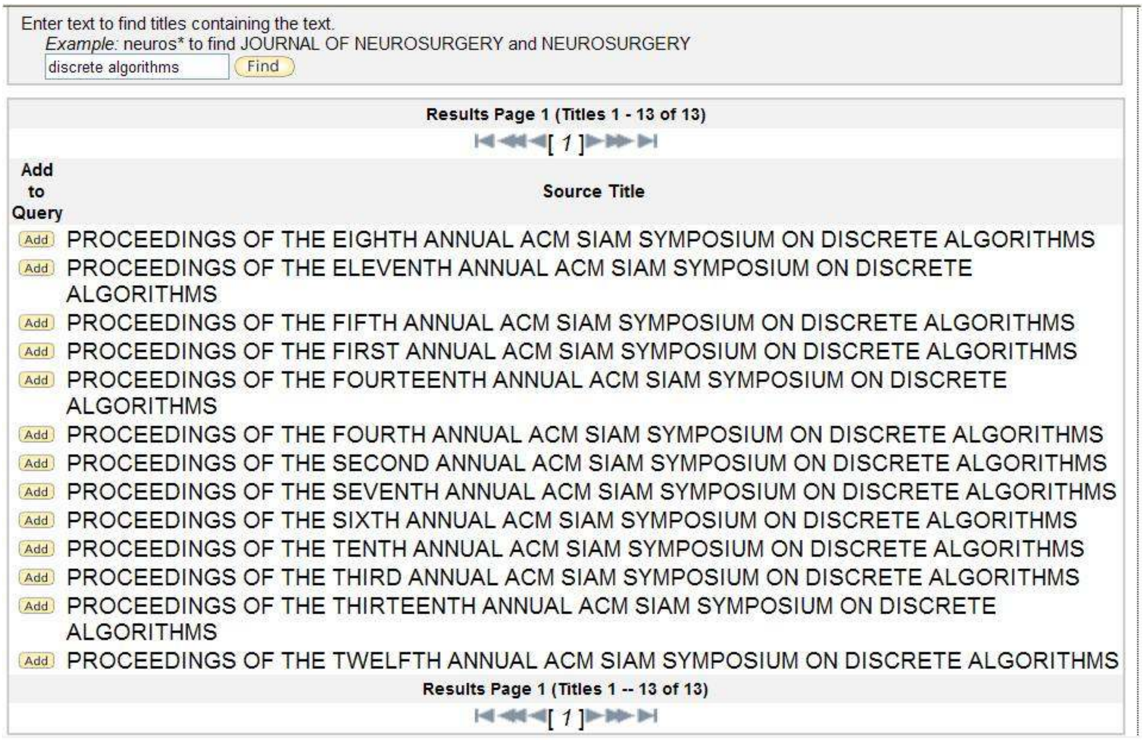


Fig 1. The titles of the different ISSI conferences as indexed by WOS as of July 2009.

An additional interesting example is the computer science conference series, the ACM Symposium on Discrete Algorithms (SODA). Web of Science with Conference Proceedings indexed in July 2009 the first fourteen proceedings (with the exception of the 9th conference), but for some reason, the indexing has been discontinued, even though the conference is being held every year, and in 2009 the 19th conference in the series took place (ACM, 2009). The list of the proceedings indexed in the series can be seen in Figure 2. One possible solution to the problem of uneven indexing is for the scientific societies holding these conferences to inform ISI of new/missing proceedings. Of course we do not know if ISI is interested in such involvement of the societies.



Enter text to find titles containing the text.
Example: neuros* to find JOURNAL OF NEUROSURGERY and NEUROSURGERY
discrete algorithms Find

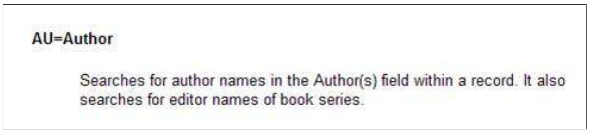
Results Page 1 (Titles 1 - 13 of 13)
[1]

Add to Query	Source Title
Add	PROCEEDINGS OF THE EIGHTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE ELEVENTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE FIFTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE FIRST ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE FOURTEENTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE FOURTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE SECOND ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE SEVENTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE SIXTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE TENTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE THIRD ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE THIRTEENTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS
Add	PROCEEDINGS OF THE TWELFTH ANNUAL ACM SIAM SYMPOSIUM ON DISCRETE ALGORITHMS

Results Page 1 (Titles 1 -- 13 of 13)
[1]

Fig 2. The list of the SODA Proceedings indexed by WOS as of July 2009

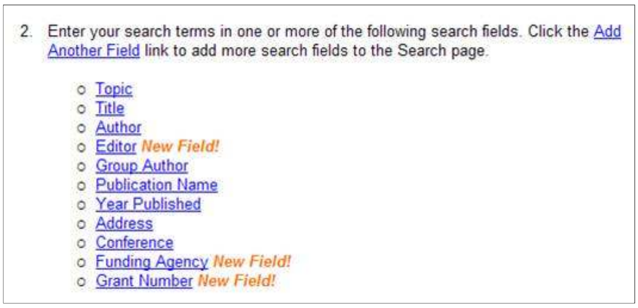
As an example of changes that occur in databases, consider how WOS interpreted author in January 2009 (see Figure 3a), and how this was changed by April 2009 (see Figure 4). In January 2009, when searching for an author all items that were either published by the author or edited by him/her were retrieved. This was not clear on the search form, but was explained in the help files, as retrieved in January 2009 (see Figure 3b). This indeed was a somewhat illogical feature, and by April 2009 a new field "Editor" was added (see Figure 3), allowing to search for edited and authored publications separately. This change in policy is an excellent example of the point I am trying to make in this paper: one must not rely on what has been said in the past about database features, but one must check in depth the situation as it is at the time of data collection.



AU=Author

Searches for author names in the Author(s) field within a record. It also searches for editor names of book series.

Fig 3a. Excerpt from the WOS help file as of January 13, 2009.



2. Enter your search terms in one or more of the following search fields. Click the Add Another Field link to add more search fields to the Search page.

- Topic
- Title
- Author
- Editor *New Field!*
- Group Author
- Publication Name
- Year Published
- Address
- Conference
- Funding Agency *New Field!*
- Grant Number *New Field!*

Fig 3b. Excerpt from the WOS help file as of April 13, 2009.

Changes over time are an important issue, because usually changes are not retroactive. A few examples for the ISI databases are: inclusion of abstracts since 1991; indexing all the authors of a publication versus indexing items by the first author only. Updating all the existing records to include this new information involves enormous amounts of work, and thus changes in the indexing policy are usually applied from a given point in time and onwards. This has to be taken into account when analyzing records retrieved from different periods of time.

Note that as of July 2009, for non-source items only the first author is indexed. As an example, consider the recently published book by Manning, Raghavan and Sch�tze "Introduction to Information Retrieval" (a non-source item, published in 2008) with 54 listed references when searching for cited references of Christopher Manning, but no citations are attributed to Prabhakar Raghavan or to Hinrich Sch�tze. The same is true for the "Introduction to Informetrics" � the 299 citations to this book (as of September 2009) are attributed to Leo Egghe, but not to Ronald Rousseau. Again, indexing all authors of non-source items may not be practical for the database, but this has to be taken into account when analyzing data based on cited references, especially in disciplines where the first author is not necessarily the primary author, like in computer science, where authors usually appear in alphabetical order.

Matching author affiliations is an extremely important issue for scientometric purposes. Until 2008, WOS listed all the affiliations of the authors, but if for example an article had four authors, A, B, C, D, that worked at two different institutions, it was not clear whether A, B and C worked at the first institution and only D at the second, or A worked at the first institution and B, C and D at the second (and numerous additional configurations are possible). As of January 2008, WOS matches the authors to their affiliations. This is a very important improvement, but again it is not retroactive, and has to be taken into account.

Another recurring issue discussed in the literature is the way ISI computes the impact factor. One problem is with ISI's definition of citable documents: when counting the number of publications only citable documents are taken into account, while for citations, citations to "non-citable" items are also counted (Moed & van Leeuwen, 1995). The question is how are "citable" documents defined? Moed and van Leeuwen found strong evidence that in 1995, "citable documents" meant articles, notes and reviews, although they were unable to find an "official definition". There is no clear definition as of now either, but David Pendlebury (2008) from the Thomson Research Services Group writes: "Although all primary research articles and reviews (whether published in front-matter or anywhere else in the journal) are included, a citable item also includes substantive pieces published in the journal that are, bibliographically and bibliometrically, part of the scholarly contribution of the journal to the literature. Research at Thomson has shown that, across all journals, more than 98% of the citations in the numerator of the Impact Factor are to items considered "citable" and counted in the denominator". Thus it is to be assumed that the term "citable" is journal dependent, which is quite reasonable, but it would be nice to know how the decision reached on of what is citable and what is not. Pendelbury's explanations are a reaction is to an article by Rossner, Van Epps and Hill (2007) complaining about the lack of integrity and transparency in the way ISI computes journal impact factors.

Scopus and SCImago

The major complaint in the literature against Scopus is that it has systematic coverage, including citation data from 1996 and onwards only. Over time this problem will become less and less serious, because informetric research usually studies recent activities. Scopus (like WOS) has been working hard on author identification, the results are getting better over time, but there is still more work to do. As an example consider an author search on Blaise Cronin in Scopus. The results as of April 2009 are displayed in Figure 4. Clearly all six identities are the same and should be grouped together, and the most recent affiliations are wrong. If we extend the search to Cronin B, we get 70 results, which include several additional publications of "our" Blaise Cronin, under Cronin, B (3 publications) and under Cronin Biaise from Indiana University (3 publications). By September 2009, the affiliation of Blaise Cronin, with the largest group of publications has been corrected, see Figure 5.

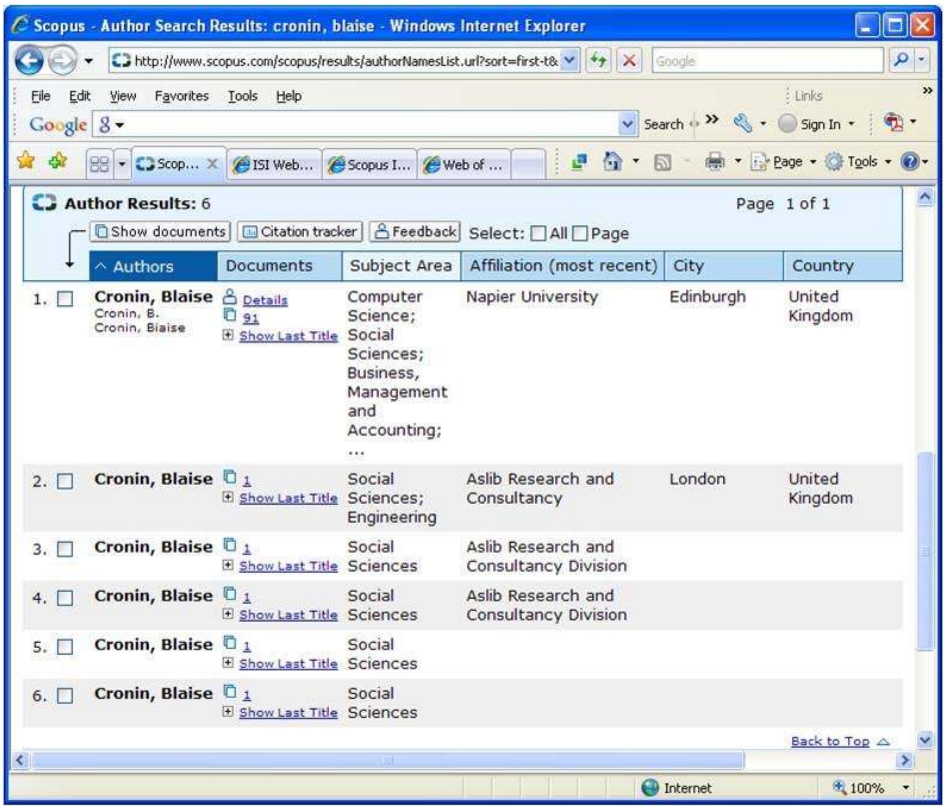


Fig 4. An example of author identification on Scopus as of April 13, 2009.

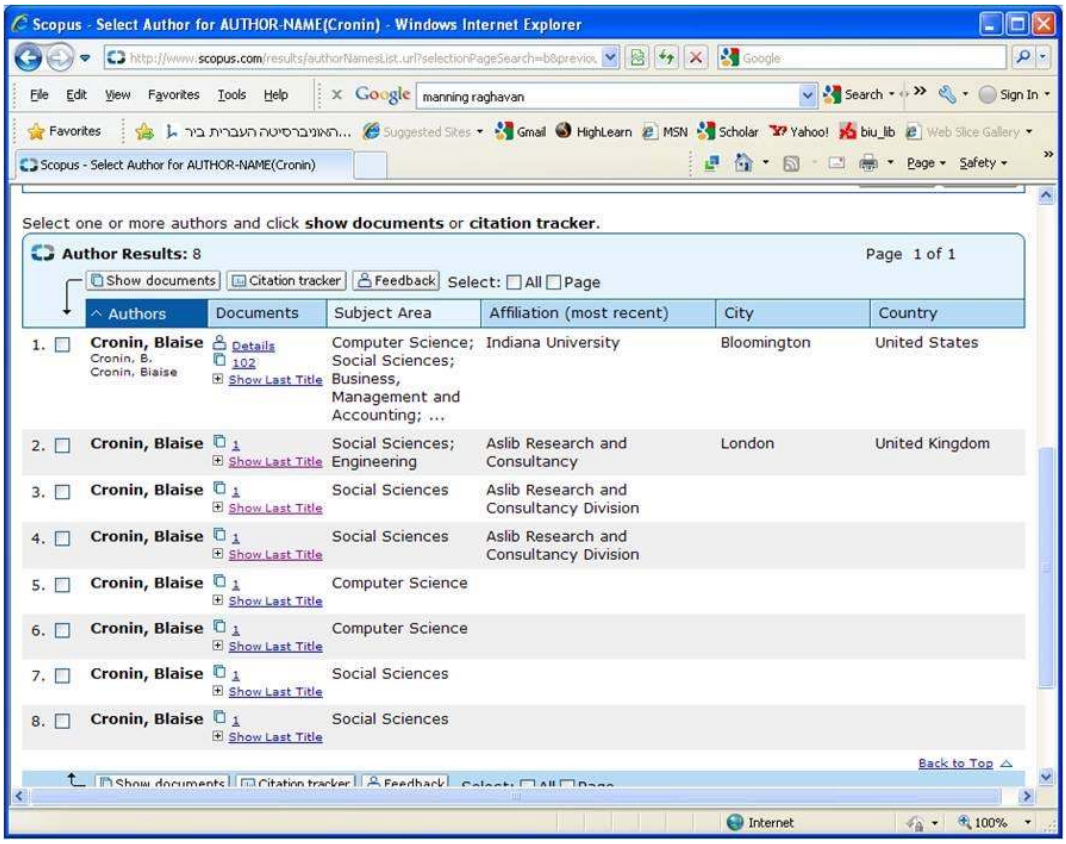


Fig 5. The same example as of September 6, 2009.

Scopus also indexes proceedings; it covers 3.6 million proceedings papers (Scopus, 2008), which is less than the 5.2 million records reported by WOS (Thomson-Reuters, 2009b), but WOS covers proceedings from 1990 and onwards, while Scopus covers them only from 1996 and onwards. Scopus does not index any of the ISSI conference proceedings. It indexes SODA from 1997 and onwards. For some reason the 2007 volume is missing, and the 2009 volume has not been indexed as of September 2009, although the 2009 conference took place in January 2009 (see Figure 6). Thus it seems that Scopus' coverage of proceedings series is not complete either, and it seems that Scopus also has some problems with assigning uniform names to conferences in the same series; see for example the different names assigned to the proceedings of the International World Wide Web conference series (see Figure 7).

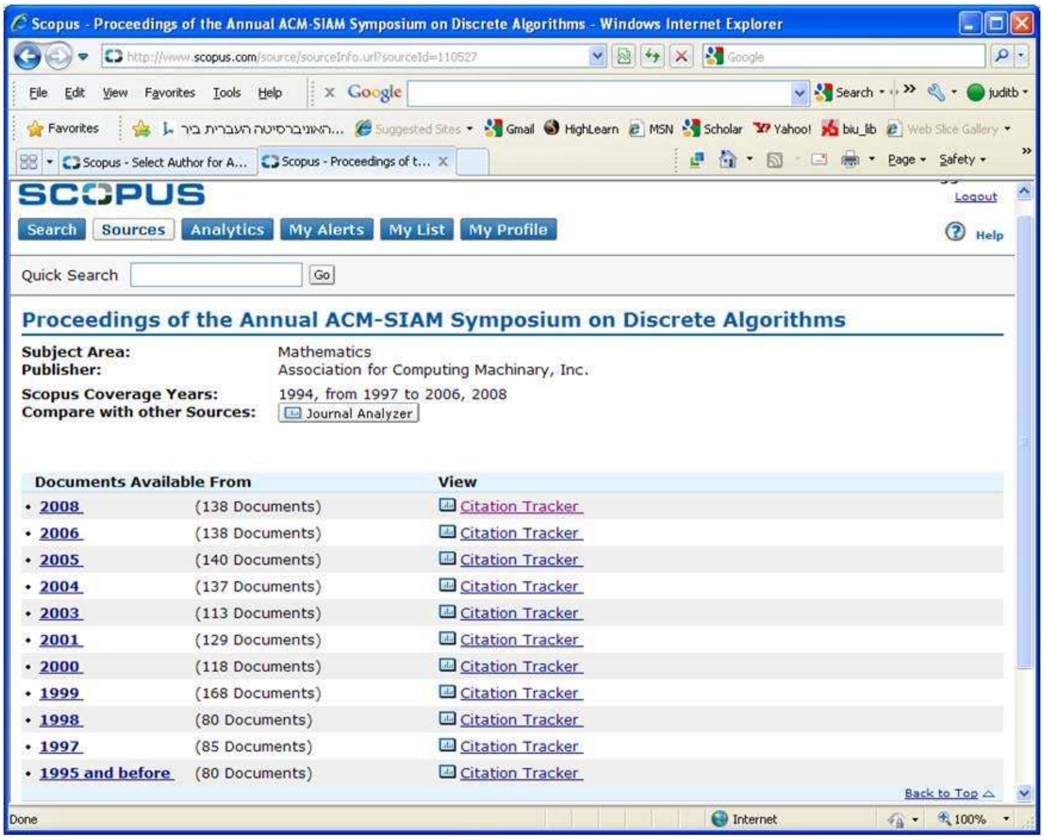


Fig 6. The list of the SODA Proceedings indexed by Scopus as of September 2009.

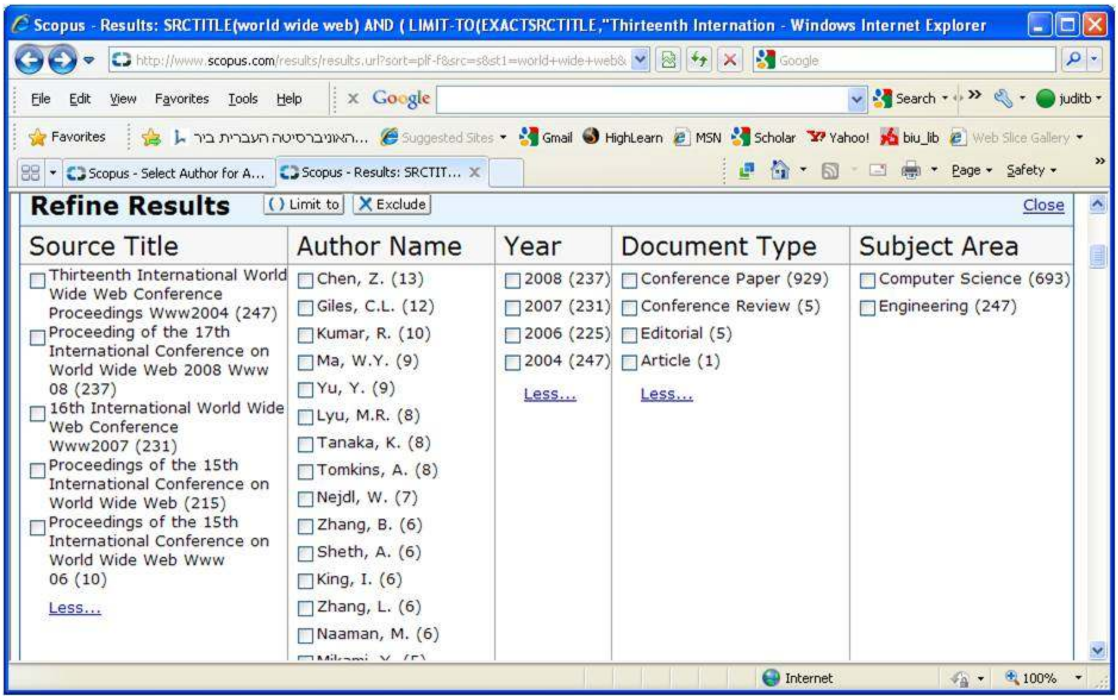


Fig 7. The titles of the different WWW conferences as indexed by Scopus as of September 2009.

Scopus indexed only the first 99 authors of a publication before 2006. Since 2006 it seems to index all authors. Note that in specific areas, like high-energy physics there are often many more than 100 authors to a paper. Author attribution is not uniform on Scopus, especially for conference proceedings, where sometimes only the affiliation of the first author is given and sometimes all addresses are given and attributed. For papers in the life and biomedical sciences, currently, for papers from 1997 and before, seemingly only the affiliation of the first author is given, This may be the result of data transferred from PubMed, because in PubMed even as of today only the affiliation of the first author is given. It seems that some or most of the records have already been corrected, because for recent articles all affiliations are given. It is not known to us whether Scopus intends to go back and to add all affiliations to older records as well.

As said before, attribution of all authors is extremely important for evaluation purposes, in addition if only the affiliation of the first author is given, then papers of a given institution will not be picked up through an affiliation search on Scopus (in case authors from the given institution were not first authors). On the other hand, for non-source items, seemingly not only the first author is given credit.

Another interesting issue is journal categorization. This is especially important when we are interested in journal rankings within categories (e.g. JCR categories). ISI's 2007 JCR category for Information and Library Science contains 56 journals. Scopus defines only 30 subject areas, but by downloading the complete journal list from the Scopus site (**http://info.scopus.com/detail/documents/title_list.xls**) one can see more refined classifications into multiple categories. Lists of journals in the smaller subject categories can also be retrieved from SCImago (2007) (**http://www.scimagojr.com/**). The SCImago portal aims to provide journal and country specific indicators derived from Scopus data. It turns out that the Scopus and SCImago categories for Library and Information Sciences are far from being identical. The SCImago list contains 92 journals and the Scopus list contains 133 journals. A possible reason could be that the Scopus list represents the journals indexed by Scopus as of March 2009 and the SCImago site presents data as of 2007 (similarly to the current JCR which provides citation data from 2007), but the differences are considerable, so there must be additional reasons for the differences. For example the Journal of Documentation and ARIST are missing from the SCImago list.

When comparing the Scopus list with the JCR list, all 56 journals are indexed by Scopus, but seven of them do not belong to the Library and Information Sciences category: three are classified as business/information systems journals, two as health journals and two as communication journals. The Scopus list includes 84 additional journals that do not appear in the JCR list, for example Cybermetrics and D-Lib Magazine, but rather interestingly the Journal of Informetrics is not among the journals in this category; it is primarily classified under Decision Sciences.

The differences between the JCR and the SCImago lists are further emphasized when we consider the ranked lists. We ranked the SCImago list according to cites per document (2 years), which is supposed to be the equivalent of the impact factor (see SCImago, 2009). Now JASIST is ranked fifth as opposed to 13th on the JCR list for 2007, but the question is fifth or thirteenth out of what? Journals rankings are often used as proxies for journal quality by decision makers and we have to make sure that they are aware of the meaning of such rankings. Of course citation counts are also dependent on the citation database, as an example, JCR reports a 2007 impact factor of 1.436 for JASIST and 1.472 for Scientometrics (ranked 12th) as opposed to 1.77 for JASIST and 1.76 for Scientometrics (ranked 7th). The issue of the sources of data used for computing the h-index was discussed in (Bar-Ilan, 2008).

Google Scholar

A lot has been written on Google Scholar, some write rather negatively about it and emphasize its weaknesses (e.g. Jacs�, 2008a & 2008b), while others praise it (e.g. Harzing & Wal, 2008 & 2009) and some emphasize the great amounts of time needed in cleansing the data (e.g. Meho & Yang, 2007, Bar-Ilan, 2006). Harzing developed "Publish or Perish" (**http://www.harzing.com/pop.htm**), a very useful tool for retrieving data from Google Scholar.

The major weaknesses of Google Scholar besides the need for extensive data cleansing are that:

1) It is not clear whether Google is committed to continue to maintain and develop Google Scholar � it is still in beta four and a half years after is was launched.

2) It does not disclose its data sources, and there is no clear list of journals and proceedings that are covered.

On the positive side:

1) It is free and quite heavily used by students and academics.

2) Google Scholar is less sensitive to typing/spelling errors than WOS or Scopus and manages to group together some misspelled items.

3) In my experience it indexes new material relatively fast.

4) It also covers areas not well-covered by WOS or Scopus (Walters, 2007).

As an example of a problematic retrieval that could be probably easily corrected by Google, consider the query 'journal citation "impact factor"' (impact factor as a phrase), limited to the year 2000. Google Scholar (GS) reported 4,820 results for this query, whereas WOS retrieved only 21 results and Scopus 25 results for the same query in April 2009. Is Google Scholar's coverage so much greater? Let us take a closer look at the first result page of Google Scholar (see Figure 8):

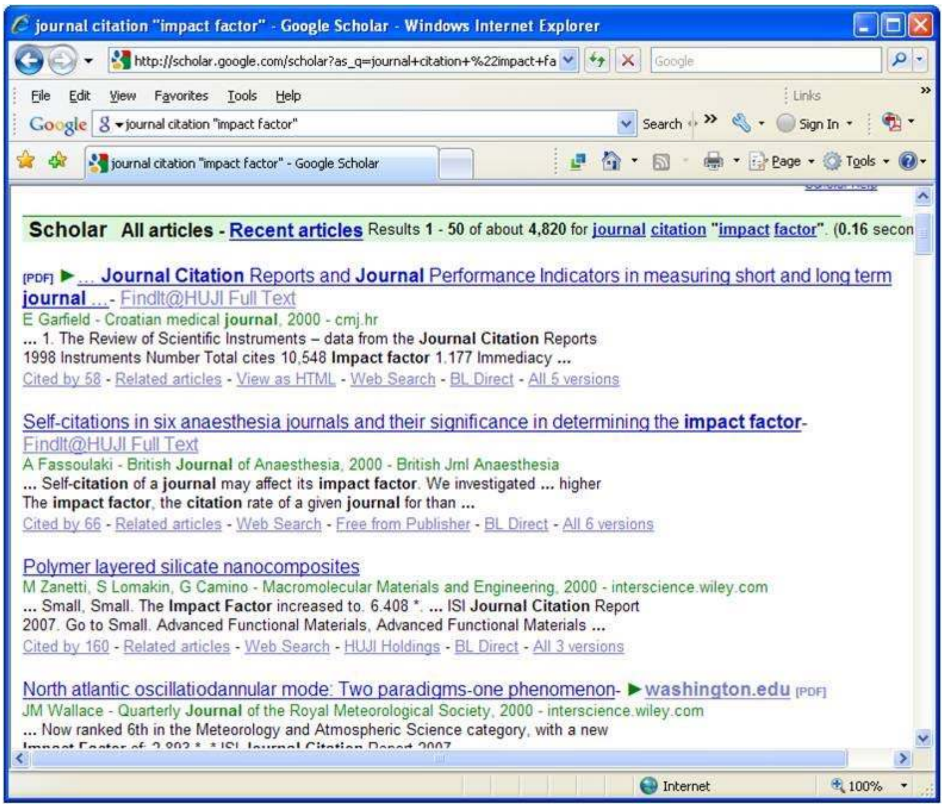


Fig 8. Top results of 'journal citation "impact factor"' from GS on April 13, 2009 with publication year limited to 2000.

The third result looks very interesting, it has been cited 160 times � it must be highly relevant to the topic! It is not easy to see how 'polymer layered silicate nanocomposites' are related to impact factors, but it is worth to try. After clicking on the result and searching for "impact factor", the mystery is solved: the search terms appear on the bottom right of the page on the side bar (see Figure 9) of every journal on the Wiley Interscience Platform announcing the impact factor of some of the Wiley journals in the current year, and has nothing to do with the specific article or the year it was published.

Fig 9. The sidebar of the page displaying the abstract of 'Polymer layered silicate nanocomposites'.

The query was rerun in September 2009, the number of results increased considerably (recall that the query is limited to publications that appeared in 2000), and Google Scholar reported about 15,500 results this time (see Figure 10).

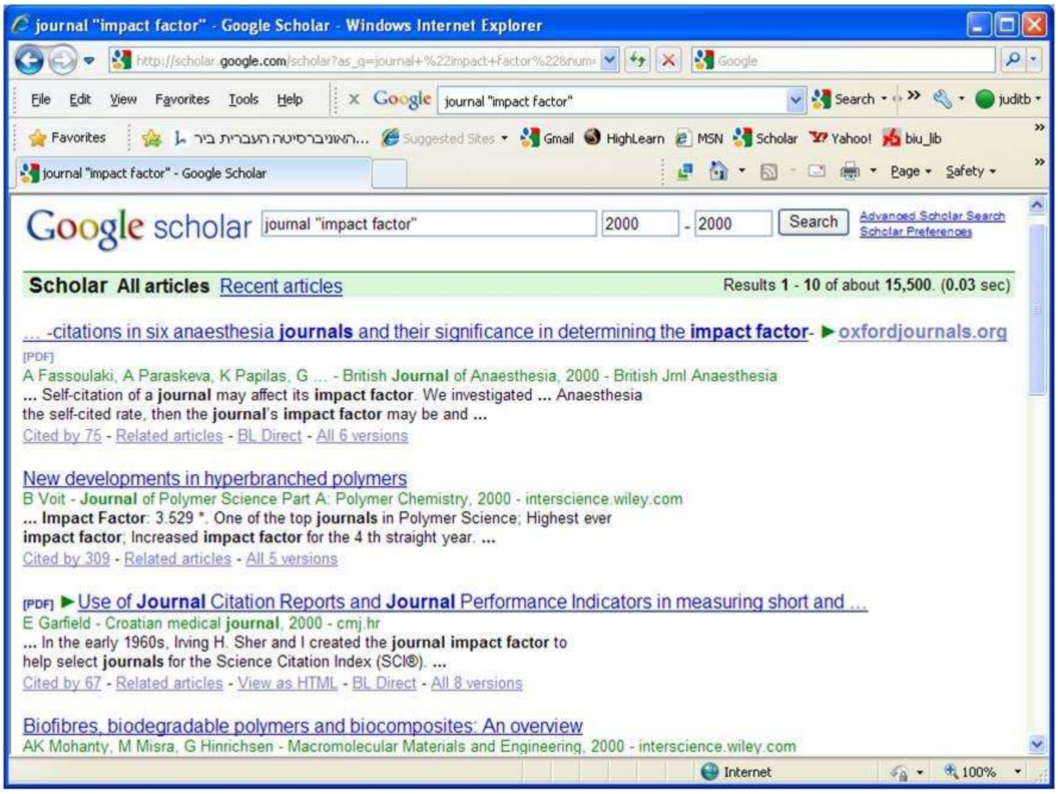


Fig 10. Top results of 'journal citation "impact factor"' from GS on September 12, 2009 with publication year limited to 2000.

Again, the second result looks suspicious. What is more interesting is that when clicking on the result, the phrase "impact factor" does not appear on the page anymore. In the snippet we see that the impact factor of the journal is 3.529. This is indeed the journal's 2007 impact factor; its 2008 impact factor is 3.821. Wiley removed the sidebar displaying the 2007 impact factor and has not replaced it by a new banner with the 2008 impact factor as of September 2008. It seems that GS has indexed the page some time during 2008 or early 2009 and has not revisited the page since.

On the other hand, in a previous paper (Bar-Ilan, 2008) I mentioned that Almind & Ingwersen's highly cited paper on "Informetric analyses on the World Wide Web", was incorrectly attributed to D. Copenhagen. This problem has been corrected since, showing again that the data sources are dynamic and are changing over time.

A major limitation of Google Scholar for informetric data collection is that it does not retrieve more than 1,000 results even if it reports to have found say 4,820 results like in the above case. In informetric research we often use large datasets, thus if we want to consider using Google Scholar as a data source, this problem has to be solved.

When criticizing Google Scholar, one has to take into account that it is not meant to be an informetric data collection tool. Its goal is to provide "a simple way to broadly search for scholarly literature", while aiming "to sort articles the way researchers do". Google Scholar is very confident and announces: "[t]he most relevant results will always appear on the first page." (Google, 2009).

Conclusions

In this paper I tried to demonstrate some limitations and shortcomings of frequently used informetric data sources. The data and the data collection tools change all the time, and the examples in this paper might not be valid in the future. The examples are not important; the major point is that when conducting an informetric study, we should thoroughly check whether the data collection process works as planned and whether the collected data are valid for the purposes of the research.

It should also be emphasized that some systems in some cases do retrospective conversion, but this is not always the case. Thus it is not enough to check current records, but if the older records are also part of the dataset under consideration, one should be aware that there may be changes in the features/indexing within the dataset.

Note

[1] A preliminary version of this paper was a keynote presentation at the 12th International Conference on Scientometrics and Informetrics, in Rio de Janeiro in July 2009.

References

Acharaya, A. (2004). Scholarly Pursuits. Retrieved April 12, 2009, from **http://googleblog.blogspot.com/2004/10/scholarly-pursuits.html**

ACM (2009). SODA: Archive. Retrieved September 5, 2009, from **http://portal.acm.org/toc.cfm?id=SERIES422&idx=SERIES422&type=series&coll=portal&dl=ACM&part=series&WantType=Proceedings&title=SODA**

Angus, E., Thelwall, M., & Stuart D. (2008). General patterns of tag usage among university groups in Flickr, **Online Information Review**, 32(1), 89-101.

Bar-Ilan, J. (2005). Expectations versus reality – Search engine features needed for Web research at mid 2005. **Cybermetrics**, 9, issue 1, paper 2. Retrieved April 12, 2009, from **http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html**

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, 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. (2009). Web of Science with the Conference Proceedings Citation Indexes – The case of computer science. In B. Larsen and J. Leta (Eds.) *Proceedings of the 12th International Conference on Scientometrics and Informetrics.*

Bar-Ilan, J. & Peritz, B. C. (2009). The lifespan of "informetrics" on the Web: An eight year study (1998-2006). **Scientometrics**, 79(1), 7-25.

Bing (2009). *Advanced search keywords.* Retrieved Spetember 12, 2009, from **http://help.live.com/help.aspx?project=wl_searchv1&market=en-US&querytype=keyword&query=egapemoh&domain=www.bing.com:80**

Elsevier (2009). E-journals at Elsevier. Retrieved April 12, 2009, from **http://www.elsevier.com/wps/find/authored_newsitem.cws_home/companynews05_00021**

Google (2009). About Google Scholar. Retrieved September 12, 2009, from **http://scholar.google.com/intl/en/scholar/about.html**

Harzing, A. W. K. & Wal, R. van der (2008). Google Scholar as a new source for citation analysis. **Ethics in Science and Environmental Politics**, 6, 61-73.

Harzing, A.W.K. & Wal, R. van der (2009). A Google Scholar h-Index for journals: An alternative metric to measure journal impact in economics & business? **Journal of the American Society for Information Science and Technology**, 60(1), 41-46.

Jacsó, P. (2008a). Google Scholar revisited. **Online Information Review**, 32(1), 102-114.

Jacsó, P. (2008b). The pros and cons of computing the h-index using Google Scholar. **Online Information Review**, 32(3), 437-452.

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.

Microsoft (2009). Microsoft Yahoo! change search landscape. News Press Release. Retrieved September 12, 2009, from **http://www.microsoft.com/Presspass/press/2009/jul09/07-29release.mspx**

Moed, H. F. (2005). Citation Analysis in Research Evaluation. Dortrecht, Netherlands: Springer.

Moed, H. F. (2009). Presentation at the SSH Bibliometric Database Stakeholder Workshop, Brighton, March 18, 2009. Retrieved April 12, 2009, from **http://www.sussex.ac.uk/Units/spru/esf/stakeholder.php**

Moed, H.F. & van Leeuwen, Th. N. (1995). Improving the accuracy of Institute for Scientific Information's Journal Impact Factors. **Journal of the Society for Information Science**, 46(6), 461-467.

Pendelbury, D. (2008). Thomson Scientific Corrects Inaccuracies in Editorial. Retrieved April 12, 2009, from **http://forums.thomsonscientific.com/t5/blogs/blogprintpage/blog-id/citation/article-id/4**

Rossner, M., Van Epps, H. & Hill, E. (2007). Show me the data! **Journal of Cell Biology**, 179 (6), 1091-1092.

SCImago (2007). SJR — SCImago Journal & Country Rank. Retrieved April 13, 2009, from **http://www.scimagojr.com**

SCImago (2009). SJR – Help. Retrieved April 13, 2009, from **http://www.scimagojr.com/help.php**

Scopus (2005). Elsevier is Proud to Announce Today's Launch of the World's Largest Scientific Abstract Database. Retrieved April 12, 2009, from **http://www.info.scopus.com/news/press/pr_041103.asp**

Scopus (2008). Scopus in detail – What does it cover? Retrieved September 6, 2009, from **http://info.scopus.com/detail/what/**

Thelwall, M. (2004). Link analysis: An Information Science Approach. Amsterdam: Elsevier.

Thelwall, M. (2007). Blog searching: The first general-purpose source of retrospective public opinion in the social sciences? **Online Information Review**, 31(3), 277-289.

Thelwall, M. (2008). Social networks, gender and friending: An analysis of MySpace member profiles, **Journal of the American Society for Information Science and Technology**, 59(8), 1321-1330.

Thomson-Reuters (2009a). Web of Science help. Retrieved September 5, 2009, from **http://images.isiknowledge.com/WOK46/help/WOS/h_database.html**

Thomson-Reuters (2009b). Web of Science databases. Retrieved September 5, 2009, from **http://images.isiknowledge.com/WOK46/help/WOS/h_database.html**

Thomson Scientific (2007). Company Timeline. Retrieved April 12, 2009, from **http://web.archive.org/web/20071012002902rn_2/scientific.thomson.com/isi/timeline/**

Walters, W. H. (2007). Google Scholar's coverage of a multidisciplinary field. **Information Processing and Management**, 43, 1121-1132.

Received 12/September/2009
Accepted 7/December/2009