NORDIC JOURNAL OF INFORMATION LITERACY IN HIGHER EDUCATION
2009, vol. 1, issue 1, 41-51 noril.uib.no

Google Scholar Compared to Web of Science: A Literature Review

Susanne Mikki *

University of Bergen

Abstract

The scope of the article is to give a literature review over comparison of the two services. To obtain insight into Google Scholar, it is tested against Web of Science (WoS), the most recognized proprietary database for peer reviewed journal content. Both databases are multidisciplinary, provide links to library holdings and offer opportunities for export of references. In addition they have the powerful feature of tracking citing items. Comparisons are based on database content, recall and research impact measures. The article touches library teaching issues at higher education institutions, and argues for which reasons Google Scholar along with WoS is worthwhile to be included in the library programs for information literacy teaching. Google Scholar is popular among faculty staff and students, but has been met with scepticism by library professionals and therefore not yet established as subject for teaching.

Keywords: information literacy; teaching; library and information science

***Contact:**
S. Mikki, dr.scient. and Senior Academic Librarian
Bergen University Library, Science Library, Bergen, Norway
E-mail: susanne.mikki@uib.no

Received: 22 May 2008; Revised: 16 September 2008; Accepted: 23 February 2009

Nordic Journal of Information Literacy in Higher Education, 2009. ©2009 Susanne Mikki

Background

One of the main tasks at research libraries is to teach information literacy. As defined by the Association of College and Research Libraries (2000), information literacy consists of a set of abilities requiring individuals to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information". In that regard, database content is crucial for locating information. It is essential, that the services offer features to make it easy to evaluate the search results, for example by how often an item has been cited. In order to efficiently use the information it is important to get fast access and have the possibility to export bibliographic data to common reference management tools.

For locating literature, traditionally Web of Science (No Date) is regarded as the most useful and trustful source and therefore, at least for subjects where publications mainly appear in journals, main subject for teaching. WoS has a thorough journal selection process based on publication standards, expert judgements, regular appearances and quality of citation data (Garfield, 1990). Its richly structured data is a premise for advanced, controlled searches.

Google Scholar (No Date) covers a wider variety of publications than Web of Science (WoS). It is based on agreements of use with the journal publishers, database vendors or scholarly societies. However, content lacks important sources and the amount of noise makes the service less useful for thoroughly literature searching. Its search algorithm is developed to return best matches, including items apparently not matching the search expression. Compared to WoS, less degree of control is offered for performing systematically searches. Google Scholar, therefore, has been met with scepticism and not yet been really established as subject for teaching. As summarized by Drewry (2007), the criticism is related to issues as inaccurate notification of content and inefficient use of metadata.

However, a user study among students at Uppsala University in Sweden measuring the effect of training for Google Scholar showed that students may be enabled to retrieve full text peer-reviewed documents, relevant for their assignment (Haya, Nygren, & Widmark, 2007). Using Google Scholar had a positive effect and increased their degree of information literacy according to the aspects of locating and using information. Drewry (2007) even refers to Google Scholar as a new paradigm in academic research.

Also advanced researchers extensively use Google for searching. It offers easier access to full text than many library provided portals (Haglund & Olsson, 2008; Haya et al., 2007; Webb, Gannon-Leary, & Bent, 2007, pp. 18-20). As discussed by Booth (2007), academic researchers use cited reference searching or known author searching rather than a keyword approach to cover their information need. This way of information handling matches well with services offered by for example Google Scholar. Haglund and Olsson (2008) claim that researchers preferably use their networks and perform simple, aimless and unstructured searches to access information. As the saying goes "It is only librarians that love to search, everyone else wants to find", they suggest that libraries should take the behaviour of researchers into consideration when designing their services. However, unorganized online searching, following hyperlinks, narrows the range of findings and ideas research is build upon (Evans, 2008). Evans found that as more publications are published digitally, the articles cited tend to be fewer and more recent. He fears that this trend, accelerate consensus, while alternative ideas that do not become consensus quickly, may be forgotten before their useful impact is recognized.

This short introduction shows that there is a discrepancy between the aims of librarians and researchers regarding the use of information search tools. As pointed out, being information literate consists of many aspects. Teaching information literacy is wider than instructing search techniques for locating information. For library and information personnel, this involves a deeper awareness about the services and how they affect science.

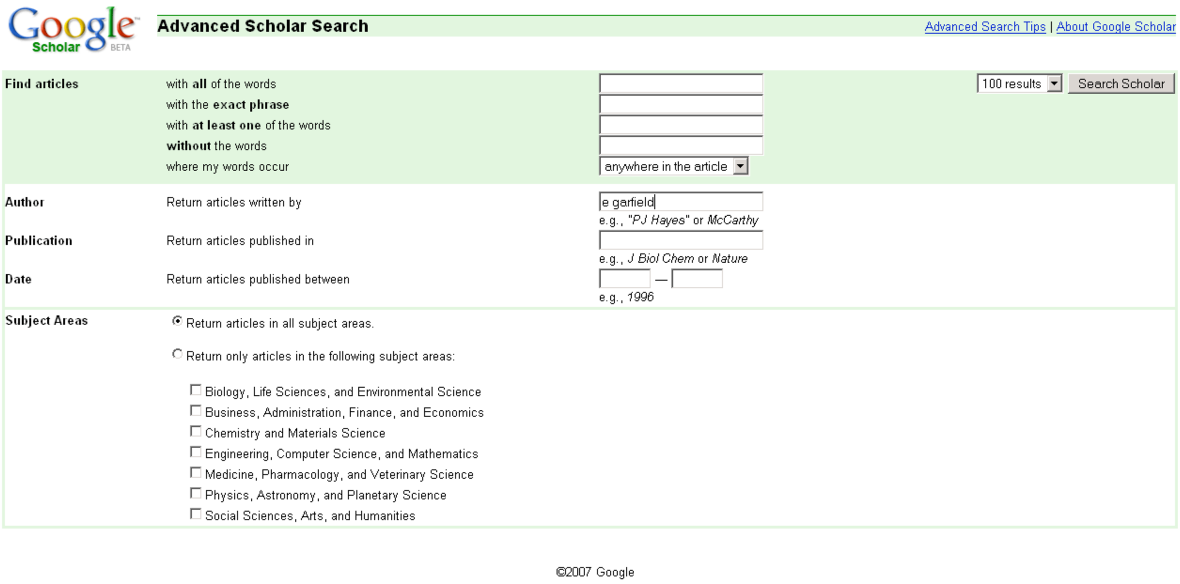
Aims and Objectives

The current study aims at constructing a deeper understanding of Google Scholar. It is based on a simple search test to exemplify database features and a literature review for assessing content and citation metrics. Further, the impact on teaching is discussed.

Database Features

Cited reference searching is, as already pinpointed by Garfield in 1955, a recognized method for searching. Alike WoS, Google Scholar keeps track of the citation data, and offers an efficient tool for finding relevant sources. Utilizing the fact that citing and cited documents are associated to each other; information can be retrieved independent of language and descriptors as subject headings or classification codes. Eugene Garfield's idea of citation searching lead to the powerful Science Citation Index (WoS). The Science Citation Index remained as the unique online citation service until 2004, when Elsevier's Scopus and Google Scholar were launched. It is natural to compare Google Scholar with WoS, both in relation to coverage and ranking because they both are multidisciplinary and include citation data.

In WoS citations are controlled partly manually. Google Scholar extracts citation automatically from reference lists of recognized scientific documents. As reported by Jacsó (2005; 2008) automatically indexing can lead to misinterpretations and noise, an annoying feature which might have been avoided if metadata had been used more extensively. Usually, scholarly documents are richly structured and tagged, searchable by their descriptors, and sorting them is possible in various ways. This is the case for subscription-based scholarly databases, such as WoS. However, Google Scholar's use of metadata is insufficient. Although the *Advanced Scholar Search* (Figure 1) offers options for searching Author, Publication, Date and Subject Area, results do not match precisely the search expression, and even simple boolean searches seem to be misinterpreted. It is a best match system, based on Google Scholar's algorithm for detecting, filtering and ranking documents.



Advanced Scholar Search

Find articles
with **all** of the words
with the **exact phrase**
with **at least one** of the words
without the words
where my words occur: anywhere in the article
100 results

Author Return articles written by: e garfield
e.g., *"PJ Hayes"* or *McCarthy*

Publication Return articles published in
e.g., *J Biol Chem* or *Nature*

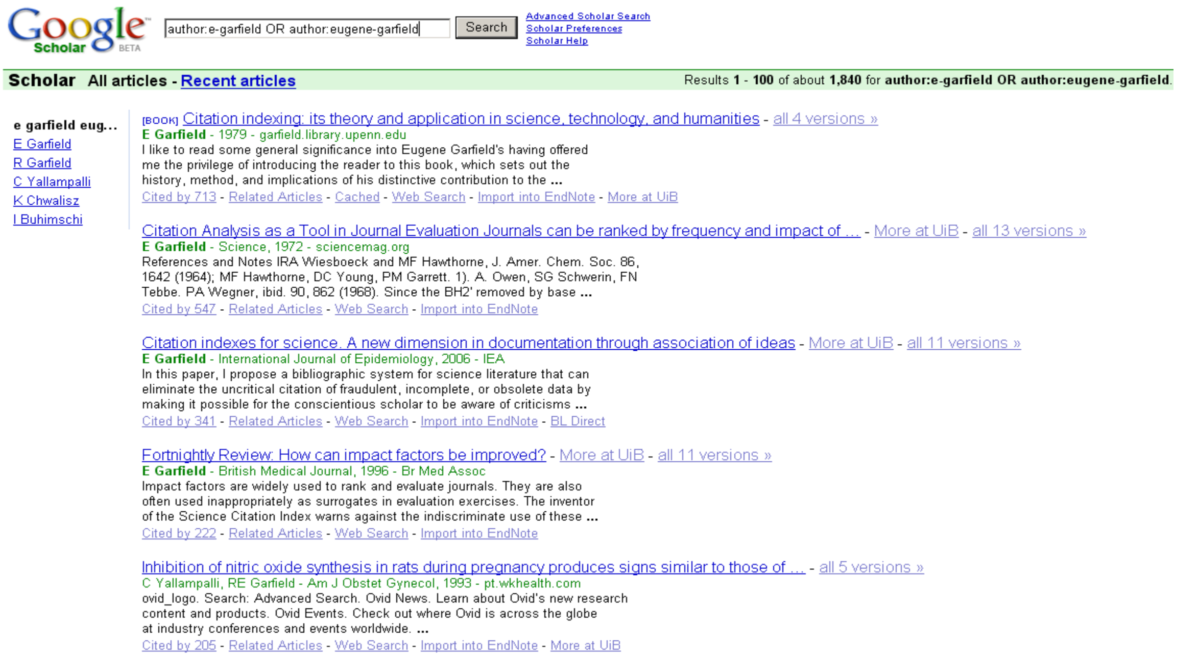
Date Return articles published between —
e.g., *1996*

Subject Areas
● Return articles in all subject areas.
○ Return only articles in the following subject areas:
□ Biology, Life Sciences, and Environmental Science
□ Business, Administration, Finance, and Economics
□ Chemistry and Materials Science
□ Engineering, Computer Science, and Mathematics
□ Medicine, Pharmacology, and Veterinary Science
□ Physics, Astronomy, and Planetary Science
□ Social Sciences, Arts, and Humanities

©2007 Google

Figure 1. Google Scholar – Advanced search features

As reported by Jacsó (2008; 2006), most of the negative aspects of Google Scholar are related to its software features, such as insufficient grouping of identical citations, resulting in duplicates, inflated citation counts, and the inability of properly identifying authors.



author:e-garfield OR author:eugene-garfield

Results **1** - **100** of about **1,840** for **author:e-garfield OR author:eugene-garfield**.

e garfield eug...
E Garfield
R Garfield
C Yallampalli
K Chwalisz
I Buhimschi

[BOOK] Citation indexing: its theory and application in science, technology, and humanities - all 4 versions »
E Garfield - 1979 - garfield.library.upenn.edu
I like to read some general significance into Eugene Garfield's having offered me the privilege of introducing the reader to this book, which sets out the history, method, and implications of his distinctive contribution to the **...**
Cited by 713 - Related Articles - Cached - Web Search - Import into EndNote - More at UiB

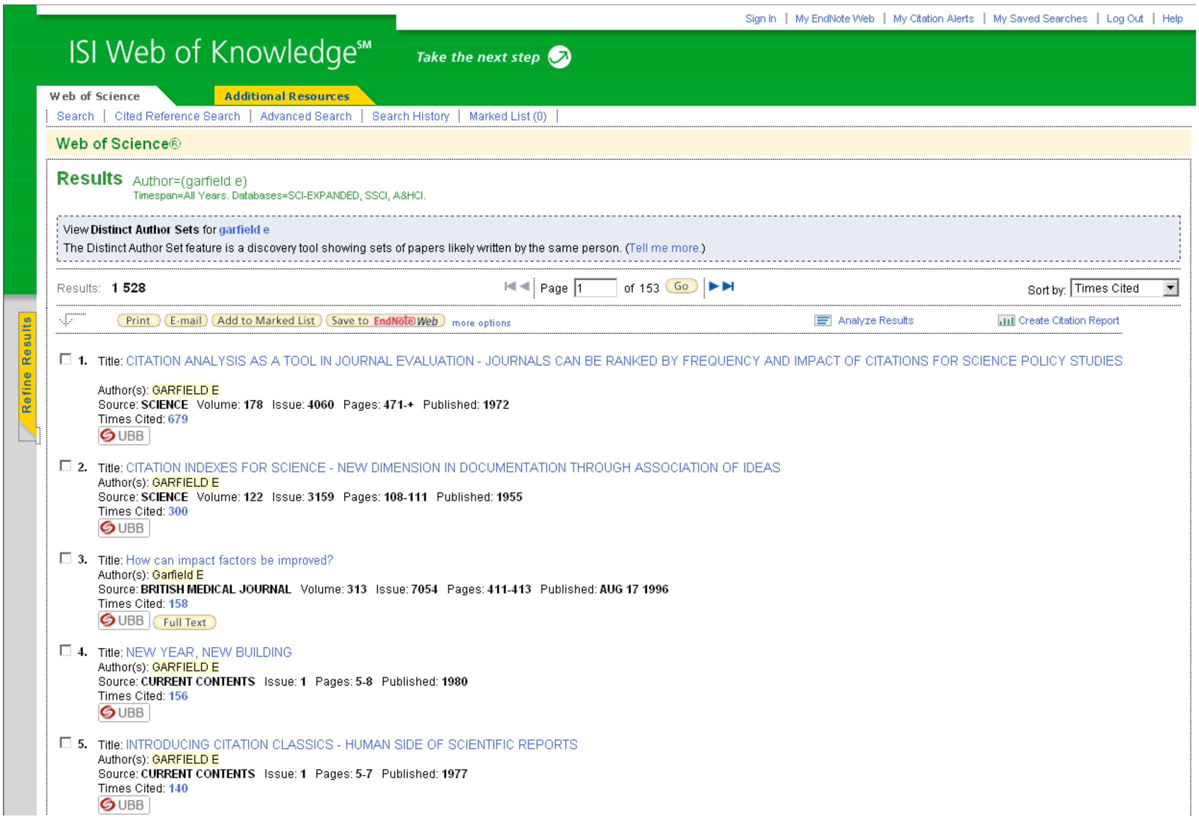
Citation Analysis as a Tool in Journal Evaluation Journals can be ranked by frequency and impact of ... - More at UiB - all 13 versions »
E Garfield - Science, 1972 - sciencemag.org
References and Notes IRA Wiesboeck and MF Hawthorne, J. Amer. Chem. Soc. 86, 1642 (1964); MF Hawthorne, DC Young, PM Garrett. 1). A. Owen, SG Schwerin, FN Tebbe. PA Wegner, ibid. 90, 862 (1968). Since the BH2' removed by base **...**
Cited by 547 - Related Articles - Web Search - Import into EndNote

Citation indexes for science. A new dimension in documentation through association of ideas - More at UiB - all 11 versions »
E Garfield - International Journal of Epidemiology, 2006 - IEA
In this paper, I propose a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms **...**
Cited by 341 - Related Articles - Web Search - Import into EndNote - BL Direct

Fortnightly Review: How can impact factors be improved? - More at UiB - all 11 versions »
E Garfield - British Medical Journal, 1996 - Br Med Assoc
Impact factors are widely used to rank and evaluate journals. They are also often used inappropriately as surrogates in evaluation exercises. The inventor of the Science Citation Index warns against the indiscriminate use of these **...**
Cited by 222 - Related Articles - Web Search - Import into EndNote

Inhibition of nitric oxide synthesis in rats during pregnancy produces signs similar to those of ... - all 5 versions »
C Yallampalli, RE Garfield - Am J Obstet Gynecol, 1993 - pt.wkhealth.com
ovid_logo. Search: Advanced Search. Ovid News. Learn about Ovid's new research content and products. Ovid Events. Check out where Ovid is across the globe at industry conferences and events worldwide. **...**
Cited by 205 - Related Articles - Web Search - Import into EndNote - More at UiB

Figure 2. Google Scholar - Presentation of search results



Web of Science®

Results Author=(garfield e)
Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI.

View **Distinct Author Sets** for garfield e
The Distinct Author Set feature is a discovery tool showing sets of papers likely written by the same person. (Tell me more.)

Results: **1 528** Page 1 of 153 Sort by: Times Cited

□ **1.** Title: CITATION ANALYSIS AS A TOOL IN JOURNAL EVALUATION - JOURNALS CAN BE RANKED BY FREQUENCY AND IMPACT OF CITATIONS FOR SCIENCE POLICY STUDIES
Author(s): GARFIELD E
Source: **SCIENCE** Volume: **178** Issue: **4060** Pages: **471-+** Published: **1972**
Times Cited: 679

□ **2.** Title: CITATION INDEXES FOR SCIENCE - NEW DIMENSION IN DOCUMENTATION THROUGH ASSOCIATION OF IDEAS
Author(s): GARFIELD E
Source: **SCIENCE** Volume: **122** Issue: **3159** Pages: **108-111** Published: **1955**
Times Cited: 300

□ **3.** Title: How can impact factors be improved?
Author(s): Garfield E
Source: **BRITISH MEDICAL JOURNAL** Volume: **313** Issue: **7054** Pages: **411-413** Published: **AUG 17 1996**
Times Cited: 158

□ **4.** Title: NEW YEAR, NEW BUILDING
Author(s): GARFIELD E
Source: **CURRENT CONTENTS** Issue: **1** Pages: **5-8** Published: **1980**
Times Cited: 156

□ **5.** Title: INTRODUCING CITATION CLASSICS - HUMAN SIDE OF SCIENTIFIC REPORTS
Author(s): GARFIELD E
Source: **CURRENT CONTENTS** Issue: **1** Pages: **5-7** Published: **1977**
Times Cited: 140

Figure 3. WoS – Presentation of search results

Still, grouping of identical citations is a step in right direction to reduce noise, see Figure 2. The fact that Google Scholar indexes the same documents from several sources, such as commercial databases, open access archives and homepages of institutions or researchers, can be considered a strength, since it facilitates free access. Also provision of local library holdings, by open URL-linking is a value-added feature. This is exemplified in Figure 2: "More at UiB", for holdings at the University of Bergen Library. It is enabled by the libraries, in order to offer access to their collection such as subscription based journals.

The last line of the record of Figure 2, shows the feature "Import to EndNote". This possibility exemplifies one of several options for exporting references offered by Google Scholar. Both this and the full-text linkage are useful and efficient tools for information management, worthwhile to include in library courses. Considering Google Scholar still being in its beta versions, further improvements are expected to come.

Google Scholar – Content

The notification of content by Google Scholar itself is rather vague, see Box 1. However, some valuable information can be extracted from numerous studies which have been carried out over the last years for assessing content and coverage. As outlined by O'Leary (2005) content in Google Scholar is provided by

- Proprietary databases
- Publishers
- Intranets of research institutions

What is Google Scholar?
Google Scholar provides a simple way to broadly search for scholarly literature. From one place, you can search across many disciplines and sources: peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations. Google Scholar helps you identify the most relevant research across the world of scholarly research.

Features of Google Scholar
Search diverse sources from one convenient place
Find papers, abstracts and citations
Locate the complete paper through your library or on the web
Learn about key papers in any area of research

How are articles ranked?
Google Scholar aims to sort articles the way researchers do, weighing the full text of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature. The most relevant results will always appear on the first page

Box 1: About Google Scholar
(retrieved 03.01.2008 from http://scholar.google.no/intl/en/scholar/about.html)

Content has grown significantly; and Google Scholar has done an outstanding job entering into partnership with academic publishers and institutions, indexing huge amounts of their scholarly content, including books, proceedings and journals.

However, content is incomplete, still missing important publishers and top ranking journals, whose digital collections are only partly indexed (Jacsó, 2008). Furthermore, Jacsó (2005) and Neuhaus et al. (2006) report a delay for newly published items, finding that Google Scholar is not as regularly updated as other databases assessed in their research. Still the time lag is a problem for bibliographic databases in general where recently published research may not be available for a certain period of time.

Some records are labelled by type of document for example [BOOK] (compare Figure 2, first record on the list) or [CITATION]. These labels are added to records which are automatically recognized as either a book or a citation. Citations are references extracted from documents, and in general do not provide a link to the full text.

Searching for E Garfield – An Example

The different database features can be illustrated closer through a search for *E Garfield*, the inventor of the science citation index. Figure 2 and Figure 3 list results searching in Google Scholar and WoS. The number of items retrieved is highest in Google Scholar with about 1840, while the number retrieved in WoS is exactly 1528. Google Scholar does not list more than the first 1000 hits. Results ranked lower than 1000 can therefore not be controlled, which is problematic for citation studies. For WoS the maximum number of displayed items is

10 000. Applying the *Cited Reference Search* in WoS would have lead to a higher number of search results mainly including duplicates, but also items not indexed by the service itself. Therefore, the number of search results and their citations in WoS is in general under-reported. For Google Scholar, however, the number is inflated. This is due to several factors, like insufficient grouping (not shown in figure), inclusion of authors with alike spellings without exactly match to the query. This can be seen by conferring to hit 5 which shows result for the author RE Garfield. This author actually contributes considerably to the result list, providing more than a fourth of the hits. The item with highest citation count (713) belongs to a book in Google Scholar. Eugene Garfield's landmark paper from 1955 receives 341 citations in Google Scholar (see item 3 in Figure 2) and 300 in WoS (compare to item 2 in Figure 3). Note that the citation in Google Scholar is not his original work published in *Science*, but a reprint in *International Journal of Epidemiology* from 2006.

To sum up, this single example reveals some characteristic features for Google Scholar:

- Results are over-reported
- Books receive high citation counts (Bar-Ilan, Levene, & Lin, 2007)
- Citation counts are similar for Google Scholar and WoS
- Elderly articles not posted on the web, are not likely to be indexed (Neuhaus et al., 2006; Pauly & Stergiou, 2005; Walters, 2007)
- Lack of content from certain publishers, here *Science* (Jacsó, 2008; Neuhaus et al., 2006)

Books and conference proceedings, in general, are not indexed by WoS, but still are valuable and highly cited sources. Especially for certain disciplines (e.g. physics, computer science, and technology) proceedings may be the main and only source for publishing. For these sources Google Scholar has proved its usefulness (J. Bar-Ilan, 2008; Meho & Yang, 2007). Also, documents written in a non-English language are better covered by Google Scholar than in WoS (Jacsó, 2006; Meho & Yang, 2007).

Comparative Assessment of Content

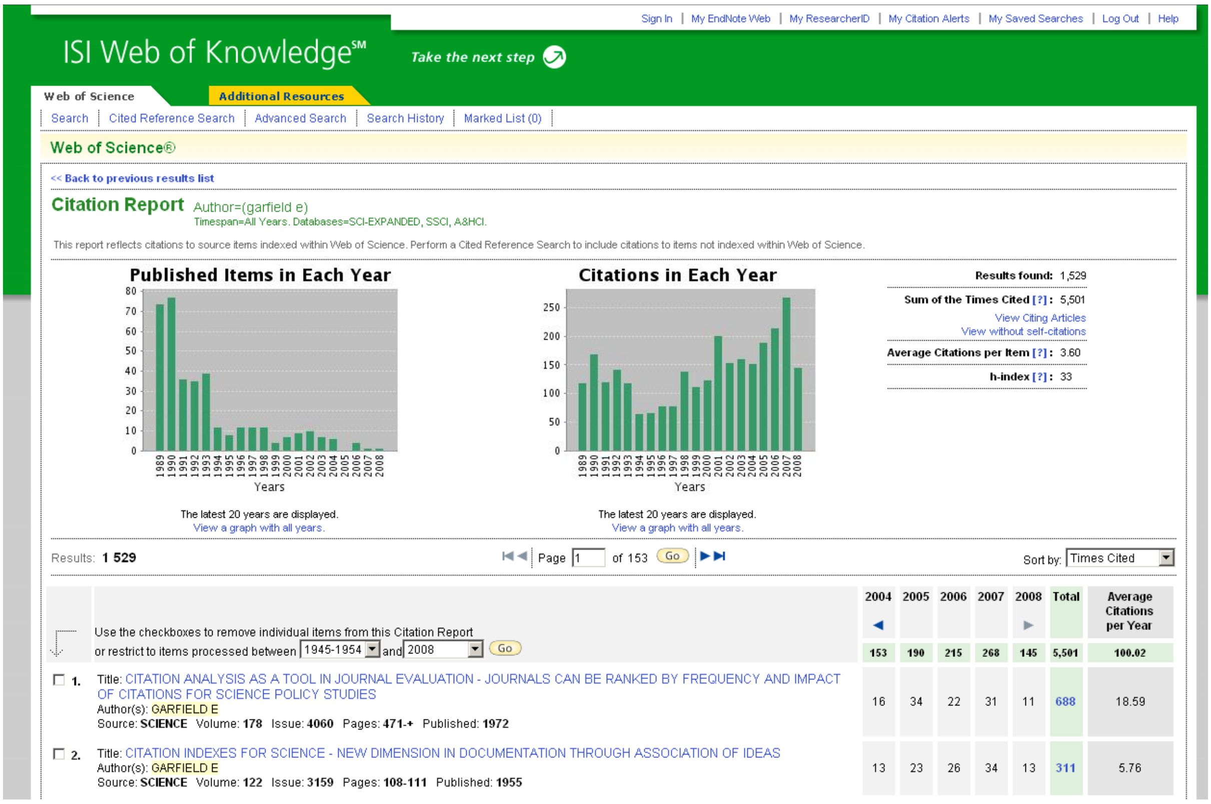
Neuhaus et al. (2006) compared content of 47 databases from various fields of disciplines with Google Scholar. 50 randomly selected documents from each database were tried recalled in Google Scholar. For science and medicine 76% of documents were covered by Google Scholar. Degree of coverage for other subjects decreased to 41% in education, 39% in social sciences and 10% in humanities. The coverage of multidisciplinary databases was related to the databases as follows; Synergy (Blackwell) with 94% coverage, Science Direct and Wiley InterScience with 90%, Ingenta with 82% and SpringerLink with 68%.

The study by Walters (2007) assessing coverage of a specific subject in social science (later-life-migration) showed that Google Scholar indexed the greatest number of core articles (93%), even though citations could be incomplete without linkage to full text or without abstracts. His findings differ from the study by Neuhaus et al. performed one year earlier, finding only 39% coverage of social sciences in Google Scholar. The difference may be explained by the method of sampling. Neuhaus et al. sampled articles from selected databases through their entire time range, while Walters only considered articles published from 1990 to 2000. This may partly explain Google Scholar's low coverage of GeoRef (26%) found by Neuhaus et al. Elderly items indexed in GeoRef are seldom posted on the web and therefore not retrievable by Google. In addition GeoRef includes records to publications written in non-English languages, whereas Google Scholar has a pronounced bias towards English language (Neuhaus et al., 2006; Noruzi, 2005). Mikki (forthcoming) compared WoS and Google Scholar for earth science content. She found that 85% of content in WoS was recalled by Google Scholar. Results were based on 29 author searches. Although citation counts and order of displayed results by the two services were similar, citation counts were significantly higher in WoS for articles indexed by both services, confirming WoS' position as a leading citation index.

Dependent on discipline, Google Scholar does compete with WoS in regard to locating information. Google Scholar is fast and familiar for many users and is therefore their first choice when in need of information. However, content in Google Scholar is not exhaustive, neither is the information provided in WoS. The comparative studies of the content in the databases indicates that in information literacy teaching a thorough literature research needs to be performed in several services.

Research Impact Studies

Research impact studies are in this paper understood as measures based on the number of publications and their citations. WoS offers different measures of impact, naming the well known *Impact Factor* for journals and the detailed *Citation Report* (Figure 4) for particular search results, displaying total, mean citation counts and the h-index. The latter has only recently got attention, and is understood as a robust measure for scientific performance. As defined by Hirsch (2005), "a scientist has index h if *h* of his or her N_P papers have at least *h* citations each and the other ($N_P - h$) papers have ≤ h citations each".



Citation Report Author=(garfield e)
Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI.

This report reflects citations to source items indexed within Web of Science. Perform a Cited Reference Search to include citations to items not indexed within Web of Science.

The latest 20 years are displayed.

The latest 20 years are displayed.

Results found: 1,529

Sum of the Times Cited [?]: 5,501

Average Citations per Item [?]: 3.60

h-index [?]: 33

Results: 1 529

Page 1 of 153

Sort by: Times Cited

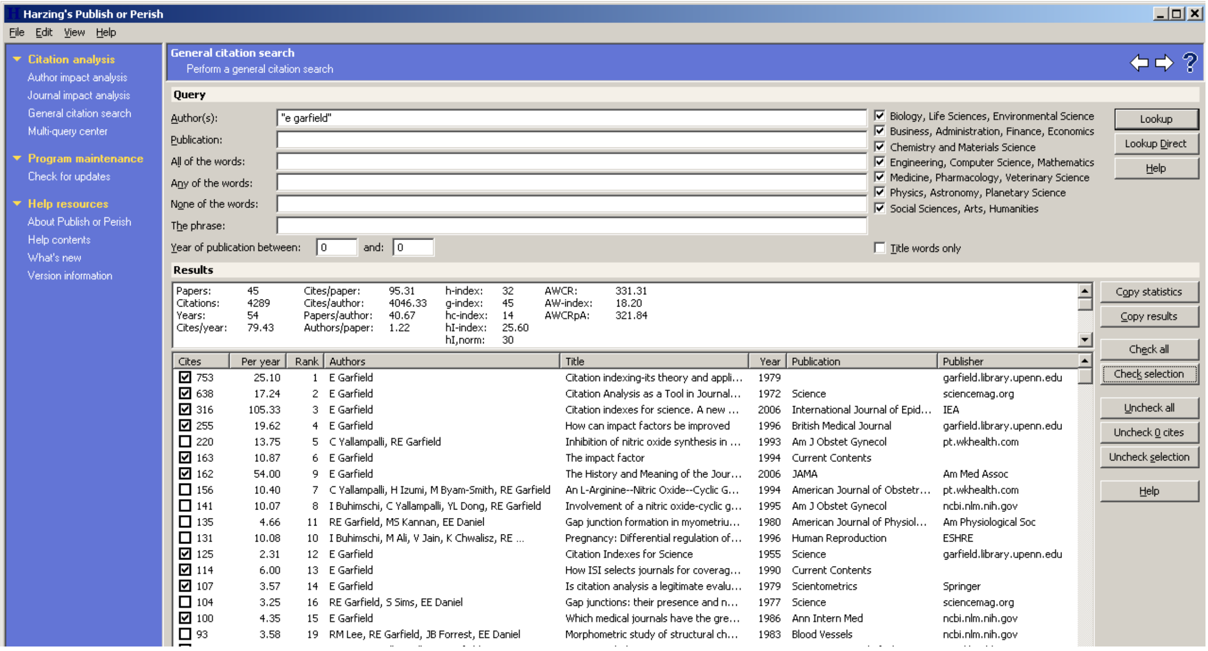
		2004 ◀	2005	2006	2007	2008 ▶	Total	Average Citations per Year
	Use the checkboxes to remove individual items from this Citation Report or restrict to items processed between 1945-1954 and 2008	153	190	215	268	145	5,501	100.02
□ 1.	Title: CITATION ANALYSIS AS A TOOL IN JOURNAL EVALUATION - JOURNALS CAN BE RANKED BY FREQUENCY AND IMPACT OF CITATIONS FOR SCIENCE POLICY STUDIES Author(s): GARFIELD E Source: SCIENCE Volume: 178 Issue: 4060 Pages: 471-+ Published: 1972	16	34	22	31	11	688	18.59
□ 2.	Title: CITATION INDEXES FOR SCIENCE - NEW DIMENSION IN DOCUMENTATION THROUGH ASSOCIATION OF IDEAS Author(s): GARFIELD E Source: SCIENCE Volume: 122 Issue: 3159 Pages: 108-111 Published: 1955	13	23	26	34	13	311	5.76

Figure 4: WoS – Citation Report for Garfield E. The h-index is equal 33.

Google Scholar on the other side lists the results by relevance, mainly sorted by times cited (see Box 1). Both services keep track of data useable for scientometric evaluation and ranking. In this section some results of comparative studies are summarized.

Studies by Belew (2005), Pauly and Stergiou (2005) and Meho and Yang (2007) document similar citation counts for articles indexed in both WoS and Google Scholar. For ranking institutions or scientists, Meho and Yang (2007) suggest that data retrieved by WoS alone is insufficient for giving an accurate picture of the impact of scientists and depends on the particular database policy and discipline. Bar-Ilan et al (2007) examined rankings by WoS, Scopus, and Google Scholar of 22 highly cited scientists by terms of *Spearman's footrule*. *Spearman's footrule* is a measure for the relative ranking of overlapping items. Differences in rankings are summarized and normalized so that the value 1 reflects complete equal ranking, and the value 0 complete opposite ranking of two result lists. *Spearman's footrule,* was calculated to 0.884 comparing WoS and Scopus, 0.830 comparing WoS and Google Scholar and 0.78 comparing Scopus and Google Scholar. Similar results are obtained by Mikki (forthcoming). By applying *Spearman's footrule*, she found a value slightly over 0.8 when comparing WoS and Google Scholar. The results prove good agreement in ranking between the databases when using this method.

For calculating impact, Hirsch (2005) suggests the h-index to be a more significant and robust measure than the mean citation count. The method cuts off a long tail rarely cited documents, and it reduces the impact of inflated citation counts of single documents. Vanclay (2007), Bar-Ilan (2008) and Mikki (forthcoming) investigated data retrieved by ISI WoS and Google Scholar and found similar values for the two services. Publish and Perish (Harzing, No Date) is a program which analyzes results by Google Scholar. It allows discarding erroneous records, and to some extent handling the data set. One test using the programme "Publish or Perish" (see Figure 5) was carried out searching for the author E. Garfield. This lead to an h-index of 32 after cleansing the data, i.e. 32 of E. Garfield's publications received more than 32 citations. The corresponding value in WoS was 33 (Figure 4). This single example indicates that the h-index is a value which returns similar results derived by the two services. Although, the averaged measures in the example returned similar results, for single scientists they could be considerable different. One reason for this is the publication practice of the subject discipline, as reported by Meho and Yang (2007) and Bar-Ilan (2008).



Author(s): "e garfield"
Publication:
All of the words:
Any of the words:
None of the words:
The phrase:
Year of publication between: 0 and: 0

☑ Biology, Life Sciences, Environmental Science
☑ Business, Administration, Finance, Economics
☑ Chemistry and Materials Science
☑ Engineering, Computer Science, Mathematics
☑ Medicine, Pharmacology, Veterinary Science
☑ Physics, Astronomy, Planetary Science
☑ Social Sciences, Arts, Humanities
☐ Title words only

```
Papers:      45       Cites/paper:     95.31     h-index:   32      AWCR:      331.31
Citations:   4289     Cites/author:    4046.33   g-index:   45      AW-index:  18.20
Years:       54       Papers/author:   40.67     hc-index:  14      AWCRpA:    321.84
Cites/year:  79.43    Authors/paper:   1.22      hI-index:  25.60
                                                 hI,norm:   30
```

Cites	Per year	Rank	Authors	Title	Year	Publication	Publisher
☑ 753	25.10	1	E Garfield	Citation indexing-its theory and appli…	1979		garfield.library.upenn.edu
☑ 638	17.24	2	E Garfield	Citation Analysis as a Tool in Journal…	1972	Science	sciencemag.org
☑ 316	105.33	3	E Garfield	Citation indexes for science. A new …	2006	International Journal of Epid…	IEA
☑ 255	19.62	4	E Garfield	How can impact factors be improved	1996	British Medical Journal	garfield.library.upenn.edu
☐ 220	13.75	5	C Yallampalli, RE Garfield	Inhibition of nitric oxide synthesis in …	1993	Am J Obstet Gynecol	pt.wkhealth.com
☑ 163	10.87	6	E Garfield	The impact factor	1994	Current Contents	
☑ 162	54.00	9	E Garfield	The History and Meaning of the Jour…	2006	JAMA	Am Med Assoc
☐ 156	10.40	7	C Yallampalli, H Izumi, M Byam-Smith, RE Garfield	An L-Arginine--Nitric Oxide--Cyclic G…	1994	American Journal of Obstetr…	pt.wkhealth.com
☐ 141	10.07	8	I Buhimschi, C Yallampalli, YL Dong, RE Garfield	Involvement of a nitric oxide-cyclic g…	1995	Am J Obstet Gynecol	ncbi.nlm.nih.gov
☐ 135	4.66	11	RE Garfield, MS Kannan, EE Daniel	Gap junction formation in myometriu…	1980	American Journal of Physiol…	Am Physiological Soc
☐ 131	10.08	10	I Buhimschi, M Ali, V Jain, K Chwalisz, RE …	Pregnancy: Differential regulation of…	1996	Human Reproduction	ESHRE
☑ 125	2.31	12	E Garfield	Citation Indexes for Science	1955	Science	garfield.library.upenn.edu
☑ 114	6.00	13	E Garfield	How ISI selects journals for coverag…	1990	Current Contents	
☑ 107	3.57	14	E Garfield	Is citation analysis a legitimate evalu…	1979	Scientometrics	Springer
☐ 104	3.25	16	RE Garfield, S Sims, EE Daniel	Gap junctions: their presence and n…	1977	Science	sciencemag.org
☑ 100	4.35	15	E Garfield	Which medical journals have the gre…	1986	Ann Intern Med	ncbi.nlm.nih.gov
☐ 93	3.58	19	RM Lee, RE Garfield, JB Forrest, EE Daniel	Morphometric study of structural ch…	1983	Blood Vessels	ncbi.nlm.nih.gov

Figure 5: Harzing's Publish or Perish – Citation metrics for E Garfield. For the highest ranked records, records belonging to RE Garfield are discarded (not checked). The h-index is equal 32.

Although WoS remains an indispensable service, it may be necessary to additionally use Google Scholar for a more complete analysis of impact measures. Google Scholar samples a wider range of publications. Although mainly of lower quality, it contains books and proceedings of importance which may alter considerable citation metrics.

Citation measures, as all statistics, must be handled carefully. They may be useful for evaluating sources, but have their limitations. Brilliancy of science is not always reflected by these measures. However, citation data is easily available and may be used for other purposes for example for performance measures of scientists, and for recruitment decisions. Awareness about the data stored in the databases will contribute to a deeper understanding of different features provided in the services. This can also contribute to the wider understanding of different information literacy aspects. Especially post graduate students and advanced researchers may benefit from this competence. They may be enabled to take decisions in relation to choosing publishing channels, where visibility of own research and getting cited are important aspects. Both highly ranked journals and open access journals or institutional repositories are alternatives to be aware of.

Summary

Due to database vendors, journal publishers and scholarly societies who provide their content to Google, the amount of qualified scholarly content has increased considerably in Google Scholar since it was launched in 2004. It offers easy, free and fast access to literature. Together with enhanced features such as exporting references, displaying citing articles and full text linking, Google Scholar is becoming an important service in literature research. However, subscription-based scholarly databases, such as WoS offer a richer tool for advanced information retrieval, containing richly structured documents, which are searchable by their descriptors.

Performance measures based on citation data, seem to be quite similar for the two services compared. However, citation data are in general under-reported in WoS and over-reported in Google Scholar, therefore citation data have to be handled with caution. For more thorough analysis, it will be wise to apply different services, WoS for its guaranteed proofed scientific content and controlled citation data, and Google Scholar for its wider collection including books and proceedings.

Citation searching and ranking by times cited are powerful features only provided by a few services. Both WoS and Google Scholar offer these features and both are multidisciplinary. Many subject fields, as for example earth sciences, reveal a high coverage with WoS, and qualify Google Scholar to be an alternative source. It offers a supplementary tool for searching and locating.

Premises for providing a thorough program for teaching information literacy for advanced scholars are to be currently aware of the different database policies and their changing features. To make use of and to be critical about citation data and their powerful potential for assessing scholarly outcome is crucial. Evaluating information by ranking may be one useful approach to get acquainted with a subject. In particular undergraduate students may benefit from this. Still it should not remain the only method. As discussed by Evans (2008), it puts researchers in touch with prevailing opinions, which may accelerate consensus and narrow the range of findings. Competencies for developing strategies for literature searching remain therefore important for all members of the academic community. Advanced searching presumes documents to be richly structured in order to keep control and methodically explore a subject. It also presumes that metadata is made searchable. Google Scholar and WoS have very different policies according scholarly searching. To be aware of the differences and limitations of the services is part of being information literate, and to be aware of how implemented features influence scholarly behaviour is an ethical aspect to be discussed and taught in information literacy courses.

References

Association of College and Research Libraries. (2000). Information Literacy Competency Standards for Higher Education. Retrieved 15 May 2008, from http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.cfm

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

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

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century - A review. *Journal of Informetrics, 2*(1), 1-52.

Belew, R. K. (2005). Scientific impact quantity and quality: Analysis of two sources of bibliographic data. Retrieved 15 May 2008, from http://arxiv.org/abs/cs/0504036

Booth, A. (2007). Researchers require tailored information literacy training focusing on information management, not simply information retrieval. *Research Information Network*. Retrieved 19 May 2008, from http://www.rin.ac.uk/files/Information%20Literacy%20Training%20-%20A%20Booth.doc

Drewry, J. M. (2007). *Google Scholar, Windows Live Academic Search and beyond: A study of new tools and changing habits in ARL libraries.* Retrieved 11 July 2008, from http://hdl.handle.net/1901/429

Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science, 321*(5887), 395-399.

Garfield, E. (1955). Citation Indexes for Science - New dimension in documentation through association of ideas *Science, 122*(3159), 108-111.

Garfield, E. (1990). How ISI selects journals for coverage - Quantitative and qualitative considerations. *Current Contents, 22*, 5-13.

Google Scholar. (No Date). About Google Scholar. Retrieved 11 July 2008, from http://scholar.google.com/intl/en/scholar/about.html

Haglund, L., & Olsson, P. (2008). The Impact on University Libraries of Changes in Information Behavior Among Academic Researchers: A Multiple Case Study. *The Journal of Academic Librarianship, 34*(1), 52-59.

Harzing, A. W. (No Date). Harzing's Publish or Perish. Retrieved 14 August 2008, from http://www.harzing.com

Haya, G., Nygren, E., & Widmark, W. (2007). Metalib and Google Scholar: a user study. *Online Information Review, 31*(3), 365-375.

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

ISI WoS. Thomson Scientific. (No Date). Web of Science. Retrieved 11 July 2008, from http://scientific.thomsonreuters.com/products/wos/

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.

Jacsó, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review, 30*(3), 297-309.

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

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.

Mikki, S. (forthcoming). Comparing Google Scholar and ISI WoS for Earth Literature. *Accepted by Scientometrics*.

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.

Noruzi, A. (2005). Google Scholar: The new generation of citation indexes. *Libri, 55*, 170-180.

O'Leary, M. (2005). Google Scholar: What's in it for You? *Information Today, 22*(7), 35-39.

Pauly, D., & Stergiou, K. I. (2005). Equivalence of results from two citation analyses: Thomson ISI's citation index and Google's scholar service. *Ethics in Science and Environmental Politics,* 33-35.

Vanclay, J. K. (2007). On the robustness of the h-index. *Journal of the American Society for Information Science and Technology, 58*(10), 1547-1550.

Walters, W. H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing and Management, 43*(4), 1121-1132.

Webb, J., Gannon-Leary, P., & Bent, M. (2007). *Providing effective library services for research*. London: Facet Publ.