

Google Scholar: The New Generation of Citation Indexes

ALIREZA NORUZI

Department of Library and Information Science, University of Tehran, Tehran, Iran

Google Scholar (<http://scholar.google.com>) provides a new method of locating potentially relevant articles on a given subject by identifying subsequent articles that cite a previously published article. An important feature of Google Scholar is that researchers can use it to trace interconnections among authors citing articles on the same topic and to determine the frequency with which others cite a specific article, as it has a "cited by" feature. This study begins with

an overview of how to use Google Scholar for citation analysis and identifies advanced search techniques not well documented by Google Scholar. This study also compares the citation counts provided by *Web of Science* and Google Scholar for articles in the field of "Webometrics." It makes several suggestions for improving Google Scholar. Finally, it concludes that Google Scholar provides a free alternative or complement to other citation indexes.

Background of the study

Eugene Garfield first outlined the idea of a unified citation index to the literature of science in 1955. "Citation indexes resolve semantic problems associated with traditional subject indexes by using citation symbology rather than words to describe the content of a document" (Weinstock 1971). Eugene Garfield's main purpose in proposing the construction of a citation index for science, in which the references in scientific articles are used as index terms, was for the citation index to function as an information retrieval tool for scientific information (Garfield 1955). The rationale behind this kind of indexing is to exploit what Garfield calls the "association-of-ideas" or "Citations are the formal, explicit linkages between papers that have particular points in common" (Garfield 1979, 1).

Soon after the beginning of the World Wide Web, the literature available on the Web increased very rapidly. The growing amount of literature on the Web and the need for multidisciplinary information retrieval accentuated the need for improved retrieval methods because while the documents were readily available, locating them

and relating them to each other was difficult. The proposed retrieval solution for the Web has been called a "Web Citation Index" (Eysenbach and Diepgen 1998). In effect, Google Scholar builds something similar to the *Science Citation Index (SCI)*, which was proposed 50 years ago for paper publishing, and provides the first Web citation index.

Citations link articles on a specific topic, and Google Scholar is built on the basis of this internal structure of subject literatures. However, as noted at the start of this article, the citation index is not a recent idea. In fact, "the first practical application of a citation index was Shepard's Citations, a legal reference tool that has been in use since 1873" (Weinstock 1971). Moreover, citation analysis is not a new idea. For instance, since the appearance of Islam in a branch of Islamic theology called the Science of Hadith, researchers have identified the accuracy and legitimacy of documents (sources) based on citations alone (Horri 1983). For more information about the history and role of citation indexing, see the works published by Dr. Eugene Garfield who has opened many doors for research and applications in infometrics, scientometrics and bibliometrics.

The principal rationale and advantage for Google Scholar is that it will democratize access to the intellectual resources of elite institutions (Banks 2005). Google Scholar enables researchers to navigate the scholarly literature on the Web in unique ways. Researchers are able to locate related articles, independent of title words, language, nomenclature or author-supplied keywords. This automated citation index is a multidisciplinary index covering virtually all sciences and disciplines and not limited to a single language, country, field or discipline; it also covers all types of published source items. However, Google Scholar is not fully comprehensive.

The purpose of this study is to answer the following questions:

- What is the purpose of Google Scholar as a free citation index?
- What are the advantages and disadvantages of Google Scholar?

Introduction to Google Scholar

Google Scholar is the scholarly search tool of the world's largest and most powerful search engine, Google. Google Scholar was developed by Anurag Acharya, an Indian-born computer scientist. It is an incredible tool allowing researchers to locate a wide array of scholarly literature on the Web, including scholarly journals, abstracts, peer-reviewed articles, theses, dissertations, books, preprints, PowerPoint presentations and technical reports from universities, academic institutions, professional societies, research groups, and preprint repositories around the world. As such, it has become a gateway to accessing scholarly information on the Web. Every day more scholarly information is available online and we continue to discover new reasons to need access to this information. If Google Scholar makes more open-access scholarly material accessible, the price of academic journals and databases may decrease or stabilize as they strive to compete. Thus the greater the accessibility of scholarly material, the greater is the value for researchers.

What makes Google Scholar most useful is its citation index feature. Google Scholar consists of articles, with a sub-list under each article of the subsequently published resources that cite the article; Google Scholar shows who cited a given ar-

ticle at a later point in time. In Google Scholar, "papers with many citations are generally ranked highest, and they get a further boost if they are referenced by highly cited articles" (Butler 2004). Google Scholar ranks search results by how relevant they are to a query, considering the title and the full text of each article as well as the publication in which the article appeared and how often it has been cited in other scholarly literature (Google Scholar 2005). So the most related documents should appear at the top of the retrieved results. Furthermore, Google Scholar automatically extracts and analyzes citations and presents them as separate results, even if the documents they refer to are not available on the Web. So it analyzes the popularity of a document according to the number of times it has been cited by other documents, and generally displays the retrieved results showing the most-cited references first.

In the future, Google Scholar may be used for citation analysis, through bibliometric techniques, which measure the impact factor of an individual publication as a function of the number of citations it receives from subsequent authors. In addition, any author may legitimately wish to determine whether his/her own work has been criticized or used by others on the Web. Authors are interested in knowing whether anyone has cited their works and/or whether other researchers in their fields have commented on them. Google Scholar facilitates this type of feedback in the scholarly communication cycle on the Web. Regardless of the year that the article was published, Google Scholar permits researchers to identify where that article was cited. Researchers can locate recent articles that have cited the particular article. A further use of Google Scholar is to identify scientists currently working in specific branches of science in order to suggest collaboration, to enter into correspondence, etc. Moreover, Google Scholar provides remote access to the indexed resources.

Comparing Google Scholar and Web of Science

A commonly used technique of conducting a literature search is to begin the search with a relevant article and look up the references cited in this article as well as the articles citing it. For ex-

ample, in 1997 Almind and Ingwersen published a paper in the *Journal of Documentation* entitled "Informetric analyses on the World Wide Web: methodological approaches to webometrics." In this paper, they established the word "webometrics" as a synonym for the concept of "bibliometric studies on the World Wide Web". This paper is among the first in the literature of webometrics. Customarily, when other authors use the term "webometrics" in subsequent articles, they will give credit to Almind and Ingwersen as the originators of the term, by citing their original article. As a result, in Google Scholar, the new articles would automatically be grouped together as the citations of the abovementioned work.

If the researcher is familiar with the term "webometrics," Google Scholar will enable him/her to find Almind and Ingwersen's article and the subsequent articles that specifically mention "webometrics." The researcher will find the original article plus all subsequent citing articles, whether or not they specifically mention "webometrics." This is especially useful to a researcher who is not familiar with the jargon of a different discipline.

A most important feature of Google Scholar is the ability to bring the researcher forward in time from an earlier known reference. As soon as the researcher locates a starting "cited" item, s/he is brought forward to items that are currently citing the original. The researcher can browse Google

Scholar and go backward and then forward again into related articles via cited references (see Table 1 and 2).

Table 1. Citation counts from Google Scholar and Web of Science (WoS) for Almind & Ingwersen (20 September 2005)

Times Cited on G. S.	Times Cited on WoS	Citations only on G. S.	Citations only on WoS	Citations on both
98	81	64	47	34

The analysis of citations shows that Google Scholar is good in finding additional citations. However, there is overlap (n=34). Google Scholar sometimes uniquely finds citations which are in journals and conference proceedings not indexed on Web of Science (WoS), especially in European languages apart from English, e.g. French, Danish, Spanish, Portuguese. So, there are 64 Google Scholar citations and 47 unique WoS citations. While it would be most useful to analyze these differences further, that goes beyond the scope of this current study.

Table 2 compares citations for articles searched with the search argument "Webometrics OR Webometric" on Google Scholar and WoS. Note that WoS results are fairly close in number as compared to Google Scholar. While such results may not occur for all searches, Tables 1 and 2 indicate the utility of Google Scholar for current topics.

Table 2. Most-cited Authors in the field of Webometrics on Google Scholar and WoS

Author(s) name	Cited Work	Times Cited on Google Scholar	Times Cited on WoS
Almind, T.C., & Ingwersen, P.	Informetric analyses on the World Wide Web: Methodological approaches to webometrics	98	81
Borgman, C.L., & Furner, J.	Scholarly communication and bibliometrics	75	40
Björneborn, L., & Ingwersen, P.	Perspectives of webometrics	67	52
Thelwall, M.	Extracting macroscopic information from Web links	59	54
Cronin, B.	Bibliometrics and beyond: Some thoughts on web-based citation analysis	53	49
Bar-Ilan, J.	Data collection methods on the Web for informetric purposes: A review and analysis	45	38
Thelwall, M.	Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university Web sites	42	33
Cronin, B., Snyder, H.W., Rosenbaum, H., Martinson, A., & Callahan, E.	Invoked on the Web	39	51
Wilkinson, D., Harries, G., Thelwall, M. & Price, L.	Motivations for academic web site interlinking	35	8

Choo, C.W., Detlor, B., & Turbull, D.	Web work: Information seeking and knowledge work on the World Wide Web	35	0
Vaughan, L., & Thelwall, M.	Scholarly use of the Web: What are the key inducers of links to journal Web sites?	34	30
Kim, H.J.	Motivations for hyperlinking in scholarly electronic articles: A qualitative study	34	26
Smith, A., & Thelwall, M.	Web impact factors for Australasian universities	34	32
Thomas, O., & Willett, P.	Webometric analysis of departments of librarianship and information science	32	31
Egghe, L.	New informetric aspects of the Internet: Some reflections-many problems	32	37
Boerner, K., Chen, C., Boyack, K., & Hamming, R.W.	Visualizing knowledge domains	31	17
Hernandez-Borges, A.A., Macias-Cervi, P., & Gaspar, M.A.	Can examination of WWW usage statistics and other indirect quality indicators help to distinguish the relative quality of medical Web sites?	31	0
Harter, S.P. & Ford, C.E.	Web-based analyses of e-journal impact: approaches, problems, and issues	30	32
Chu, H., He, S. & Thelwall, M.	Library and information science schools in Canada and USA: A webometric perspective	26	0
Thelwall, M.	Evidence for the existence of geographic trends in university web site interlinking	26	25
Thelwall, M.	A comparison of sources of links for academic web impact factor calculations	23	4
Smith, A., & Thelwall, M.	Interlinking between Asia-Pacific university Web sites	22	10
Thelwall, M., & Harries, G.	The connection between the research of a university and counts of links to its web pages: An investigation based upon a classification of the relationships of pages to the research of the host university	20	14
Thelwall, M.	What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation	20	1
Björneborn, L.	Small-world linkage and co-linkage	17	0
Thelwall, M., & Wilkinson, D.	Three target document range metrics for university Web sites	16	1
Leydesdorff, L.	Indicators of innovation in a knowledge-based economy	16	0
Thelwall, M., & Tang, R.	Disciplinary and linguistic considerations for academic web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan	15	1
Thelwall, M., Tang, R., & Price, L.	Linguistic patterns of academic web use in Western Europe	15	12
Bar-Ilan, J.	The Web as an information source on informetrics? A content analysis	14	13
Thelwall, M.	A research and institutional size based model for national university web site interlinking	14	9
Prime, C., Bassecouard, E., & Zitt, M.	Co-citations and co-sitations: A cautionary view on an analogy	13	8
Leydesdorff, L.	The mutual information of university-industry-government relations: An indicator of the Triple Helix	12	5
Koehler, W.	Digital libraries and World Wide Web sites and page persistence	12	0
Thelwall, M.	An initial exploration of the link relationship	12	7
Vaughan, L., & Shaw, D.	Bibliographic and web citations: What is the difference?	11	8

Key advantages and capabilities of Google Scholar

Google Scholar provides most of the advantages of other citation indexes. The primary advantage in using Google Scholar is that it leads the researcher to the latest articles; that is, it goes for-

ward in time rather than solely backward; it identifies relationships between articles, breaking through disciplinary and geographic boundaries. So a researcher can go forward to determine who has cited an earlier work. By starting with a single article, s/he can identify additional articles that have referred to it. And each retrieved article

may provide a new list of references with which to continue the citation search on the Web. Google Scholar allows researchers to trace what articles are cited by a particular article and where the article has been cited elsewhere. This can be useful for developing a bibliography or tracing the development of a topic or issue on the Web. Citation searching helps in identifying authors and key works, which can lead to finding new resources.

Google Scholar has a number of important advantages when compared with other databases. It locates documents posted on the Web. Since several authors post preprints to their Web sites much earlier than the articles appear in printed journals, researchers may find more current information than they would through commercial databases. The autonomous nature of Google Scholar keeps the cost of maintaining the index much lower than other citation indexes, which are often manually created, and thus provides a free alternative or complement to other citation indexes. It can also give up-to-date impact measures of particular articles.

Other advantages of Google Scholar include the following:

- It provides international coverage of journals and scholarly resources.
- It allows researchers to conduct broad-based, comprehensive, and multidisciplinary searches to discover hidden subject relationships on the Web.
- There is no bias due to subjective selection of journals; however, it may have a language bias, as we conducted a search with the following query (site:edu.cn filetype:pdf) to find how many Chinese-language articles are indexed. We did not find any Chinese articles. Currently Google Scholar indexes documents in English, French, German, Spanish, Italian, and Portuguese.
- Google Scholar is not restricted to articles – preprints, technical reports, theses, dissertations, and conference proceedings are also indexed.
- It is able to recognize variant forms of citations. However, in some cases it has problems with the name of authors that have diacritical marks (ö, ø, é). For instance, Björneborn
- Users can combine searches of words from the article title, keywords, and authors and domain name.
- Google Scholar is available on the Web, it contains full text of many articles and users can search all years simultaneously.

The advantages of citation indexing have been discussed in considerable detail by Eugene Garfield. Briefly these advantages include the ability to rank and evaluate literature by understanding how it is used (i.e. cited) and who is using it, automating analysis of citations to eliminate the bias that human analysis can introduce and observing that collections of citations can form a highly accurate view of the key literature in a field. An article on the history of citation indexing summarizes his contributions as follows:

Garfield's achievement lay in establishing the utility and objectivity of a citation index in pulling up related papers in published literature that at first glance might not have seemed pertinent to the researcher's inquiry. Today, it is considered to be one of the most reliable of resources in tracing the development of an idea across the multitude of disciplines that are part of our body of scientific knowledge. [Thomson ISI 2005]

Disadvantages

Google Scholar is, however, not without its disadvantages. Sometimes, Google Scholar includes administrative notes, library tours, student handbooks, etc., which are not exactly scholarly material from the point of view of the traditional definition of scholarly information. Sources of publications may not be universally recognized as scholarly. Moreover, "what it does not include is important. If we understand correctly what it does index, it is time to get on with the much larger job of identifying more trusted scholarly sources" (Hamaker and Spry 2005). Unfortunately, Google Scholar's algorithm cannot distinguish between articles, editorial notes or library guides. Google Scholar is a beta version and an experiment that has some limitations:

- It currently has a language bias. We conducted two searches with the following queries (site:edu.cn filetype:pdf) and (site:ac.ir filetype:pdf) to find articles in Chinese and Persian, but there are no articles available in these languages. Google Scholar does not index complex script languages, such as Persian, Arabic, Chinese, and Japanese. It indexes only European languages. Researchers should consider this inherent limit.
- There is inconsistency in citation styles (i.e. spelling variations, incomplete citations).

- It uses author initials, so several different authors with the same last name and initials cannot be differentiated.
- Many scholarly periodicals and magazines are not indexed.
- There is no subject indexing and/or classification access - searching is by keyword in the journal title, article title, abstract, or text.

Moreover, Peter Jacso (2004) criticized Google Scholar in several papers. A basis for his concerns is the incorrect counting of citations because, in a quote given to *The Scientist*, "Google Scholar (GS) does a really horrible job matching cited and citing references." This quote and the background for it are amply explored in a posting on Jacso's own Web site. [Jacso 2005]

Problems for university libraries

There are also some special problems for university scholars and students. Google Scholar consists of citations and links to journal articles that are not free (not even their abstracts). Librarians' main concern is that some students and faculty members may pay for articles on Google Scholar that are already available from subscribed databases in university libraries. Google Scholar also sometimes links back to databases available in university libraries, so faculty members and students may get frustrated using it. But there is a lot of potential for this new scholarly search engine, especially in the area of open access journals, conference proceedings and e-theses.

Table 3. Useful search tips often overlooked by searchers

Command	Function	Example
+	Searches stop words	+to +be +or not +to +be
-	Removes a word or phrase	eyes diseases -animal
OR	Boolean operator to expand search. Must be capitalized.	bibliometrics OR informetrics
"Quotation marks"	Phrase searching	"Persian Gulf War"
intitle:	Returns results that include the search term in the title of the document/page.	intitle:competitive intelligence
allintitle:	Searches for all the words in the title of the document/page	allintitle:competitive intelligence
site:	Searches for the word in the site/domain name. Limits searches to a special domain or site.	site:ac.uk "digital libraries" site:edu
inurl:	Searches for the word in the URL	inurl:webdex
allinurl:	Searches for all the words in the URL	allinurl:semantic web
author:	Searches for the word in the author's name	author:Berners-lee
filetype:	Limits file type and retrieves a special file format	metadata filetype:pdf allintitle:metadata filetype:ps
*	Searches the phrase (enclosed in quotation marks) and * replaced by any single word. This operator can be used for proximity searching to retrieve a compound name or a phrase that appear a specified number of words in the middle of it.	"web * analysis" "citation ** analysis" "web *** ontology" "Anglo American * Rules"
..	Number range	"digital camera" "5..5000 megapixel"

Search strategies and operators

Quick and easy to search, Google Scholar's interface search screen is one simple search-box. Google Scholar search strategies can be combined with a variety of keywords, article title, author name, and domain searches. This combination of actions allows the researcher to search for articles by author name, article title, keywords, or journal title, find scholarly documents that cite a particular article, and look at the context of citations made within and to a particular article. Table 3 summarizes and illustrates some of the powerful search strategies often overlooked by searchers.

By using Boolean operators and special characters, such as AND (+), OR (|), NOT (-), and phrase searching using "quotation marks", researchers can fine-tune their search queries and increase the accuracy and relevancy of retrieved results. Pressing ENTER after typing a search query will default to search. When the article is available, the first link gives the URL for the original article.

Overall, the best way to learn how to search more effectively is to understand how Google Scholar works and how it interprets search requests. Based on our experience with Google Scholar and Google, we offer the following basic tips for better searching:

1. Keyword Search Rules

- Keywords must be exact and specific, searching for foot will not return feet or football
- Keyword searches are not case sensitive – New Delhi is the same as new delhi
- Keyword order is important especially in phrase searching, searching for “technology watch” will produce more results than searching for “watch technol-ogy”.

2. The AND Assumption

Google Scholar always assumes there is an AND between any keywords. For example, if a searcher enters:

Web intelligence

Google Scholar will search for documents containing both the word web and the word intelli-gence.

To force Google Scholar to search for a particular term, a + (plus) sign should be put in front of the word in the query. Note that there should not be a space between the + and the word, as illustrated in the following example:

+What +is P2P

The + sign is typically used in front of stop words (e.g. of, the, in, an, I, it, and, as, where, how) that Google Scholar would otherwise ignore or when a searcher wants Google Scholar to return only those documents that match his/her search terms exactly. However, the + sign can be used on any term. For example, in the query World War I, “I” is a stop word and is not included in a search unless the searcher precede it with a + sign; for example:

World War +I

Google Scholar excludes common words in English and in other languages, such as “de” (which means “of” in French, Spanish, Italian, and Portuguese) and “la” (which means “the” in these languages). So if Google Scholar ignores a term critical to the search, a + sign should be put in front of it. For example:

+La Bibliothèque Nationale +de France

3. The NOT Command

To find documents *without* a particular term, a - sign should be put in front of the word in the query. The - sign indicates that the searcher wants to exclude documents that contain a specific term. Note that there should not be a space between the - and the word, for example

Persepolis – football

which searches for the city Persepolis in Iran, but does not return documents relating to the Persepolis football team. No documents containing the word “football” will be returned by this query.

4. Searching for an Exact Phrase

Putting quotation marks around keywords will make it possible to search for an exact phrase. For example, if a searcher enters:

“Competitive intelligence”

Google Scholar will only search for documents containing the entire phrase competitive intelli-gence.

To search for a phrase, a proper name, or a set of words in a specific order, the searcher should put them in quotation marks. A query with terms in quotes finds documents containing the exact quoted phrase. For example, [“Eugene Garfield”] finds documents containing exactly the phrase “Eugene Garfield.” So this query would find documents mentioning Eugene Garfield, but not documents containing “E. Garfield” or “Garfield.” The query [Eugene Garfield] (without quotes) would find documents containing any of “Eugene Garfield,” “Garfield,” or “Eugene.”

Google Scholar will search for common words (stop words) included in quotation marks, which it would otherwise ignore. For example:

“How to write a scientific paper”
“to ask or not to ask”

It is possible to include more than one quoted string in a query. All quoted query phrases must appear on a result page; the implied AND works on both individual words and quoted phrases.

For example:

“Link motivations” “web links”
allintitle: “citation indexing”
“information retrieval”

5. Common words are Automatically Excluded

Google Scholar automatically excludes most single letters, single digits and common words such as “where,” “what,” “how,” “to,” and so on. Google Scholar will always display which words were excluded just above the search results. If a searcher wants Google Scholar to include a common word, the + sign should be put in front of it. For example:

+what +is intelligence
Persian Gulf War +I

It is also possible to search for an exact phrase which includes the common words. For example:

“how to use citation analysis”

6. The OR Command or | (vertical bar)

If a searcher would prefer to search for one word or another, s/he should enter OR in capital letters between the keywords. Any (or all) of the search terms separated by the operator must appear in the record. It is also possible to use | (vertical bar). For example:

scientometrics | scientometric
webometrics OR webometric

Google Scholar will search for documents containing *either* the word webometrics *or* the word webometric.

7. Domain Search

It is possible to search within a specific web site or top-level domain by entering the search terms, followed by the word “site” and a colon followed by the domain name. This command searches only a specific web site, or excludes that site from

the search. It can be used in combination with other search operators. For example:

“citation indexing” site:edu
“citation indexing” -site:edu
allintitle: “citation analysis” site:edu
allintitle: “co-citation analysis” site:ac.uk

Moreover, it is possible to search for scholarly documents published only in one country. For example, if a searcher would like to access all documents about “digital libraries” which are available online on the Web sites of Indian universities, the following queries can be used:

“digital libraries” site:ac.in
“digital library” site:ac.in
“digital libraries” OR “digital library” site:ac.in
digital library site:ac.in

8. Author Search

Google Scholar will make it possible to search for author names by entering the first name (initial) first, followed by a space and last name. It is also possible to enter last names without initials. First names and initials are searchable when using “author:” command, but results are often erratic. It is possible to enter a surname without any initials to find all authors with that surname. For example:

author: Garfield

author: Garfield alone finds the same names. It may also find additional records where Garfield appears without any initials. author:Almind AND author:Ingwersen finds articles authored by both Almind and Ingwersen. author:Gorman -author:Crawford finds articles in which Gorman appears but not Crawford. It is possible to separate two or more names by the Boolean operator OR. If the author has a common last name, the searcher should combine variations with OR. author:Smith OR author:Thelwall finds articles authored by either Smith or Thelwall (or both Smith and Thelwall). It is also possible to use quotation marks (“) around the words. For example:

author:“Z Parsa”
author:“Zohreh Parsa”

Figure 1. Google Scholar's Advanced Scholar Search

If the surname contains an apostrophe, or hyphen, the searcher should enter the name both with and without the punctuation mark. Likewise, s/he should enter a surname with embedded spaces with and without the spaces. It is also possible to join two versions of the name by the Boolean operator OR. For example:

author:"R O'Brien" OR author:"R O'Brien"
 author:"M O'Hara" OR author:"M OHara"
 author: « Y Deville » OR author: « Y De Ville »
 author:"Lopez-Gonzalez" OR
 author:Lopezgonzalez

The * sign can be used as the proximity search operator to search for authors that have middle names. *Proximity searching* is efficient when the searcher knows the first name and the last name of an author, but being unsure of his/her middle name. For example:

"Shiyali * Ranganathan"
 "John * Harvey"
 author:"J * Harvey"

"Proximity searching can be useful when you want to find documents that include someone's name in any of the following orders: first middle last, last first middle, first last, last first" (Blachman 2005). Moreover, diacritical marks are searchable. Variations of names that contain diacritical marks in the original can be searched. The name Björneborn may appear in the database as Björneborn or Bjoerneborn:

author:Bjoerneborn OR author:Björneborn

Generally, author names appear in Google Scholar exactly as they do in the source document. It is advisable to search for all variations of names.

9. Cited Reference Search

Search strategies are not limited to keywords or phrases, but can be directed to include the connections created through the references made by the authors of a given article as well as the citations to that article by authors of subsequent articles. Cited reference searching in Google Scholar is a fast efficient way to find online articles that answer researcher questions or provide essential background literature for research purposes. This feature enables researchers to find articles that have cited a previously published work and to identify more recent articles on the same topic. For example, it is possible to find all works that reference the book *Citation indexing: Its theory and application in science, technology, and humanities* by Eugene Garfield.

This type of cited reference searching often locates relevant articles because the underlying assumption is that there is a relationship between the citing and the cited document. Thus, researchers can discover how an idea or technique has been extended, improved, confirmed, and applied. The question is how to do a cited reference search on Google Scholar:

- a. Enter the name of an author and the title or keywords of his/her work. Then press ENTER or click on *Search* button. It is also possible to use the "author:" command, but you may retrieve hundreds of hits, many of them not relevant.
- b. If you retrieve too many hits, return to the form and add the title of the work, and enter the name of the first author of a multi-authored article or book.
- c. Click on "cited by #" link and see references on Google Scholar that contain the cited author/work data you entered and then you have the list of citing documents. The number following the "cited by #" indicates the number of times each document has been cited on the Web.

10. Date Search

Using the "Advanced Scholar Search" (see Figure 1) makes it possible to limit results to a date range or specific publication or journal.

Suggestions for improvements

Google Scholar has many great features, but there are ways it can be improved. As mentioned previously, an author wants to get cited and wants to

know how often s/he is cited and who is citing his/her work. Therefore, Google Scholar should provide a *citation alert* system to notify authors by e-mail whenever a new article has cited one of his or her articles. Alternately, researchers could use this feature to keep track of citations to their favourite or most useful articles by others.

Google Scholar would be more useful if it allowed searching by journal titles as cited works, which would be interesting for journal editors and even authors, to know how often a particular journal is cited (i.e. impact factor) and to find references to articles published in a journal. Such data can be used for bibliometric analysis. It should be noted that it is now possible to search a journal title to find articles and see the number of times individual articles have been cited.

Other areas are suggested for improvement in Google Scholar:

1. To cover significant journals more comprehensively (less important over time as more journals become available online).
2. To distinguish subfields more accurately (e.g. Google Scholar will not disambiguate two authors with the same name).
3. To add a collection of checkpoints to search by language, field, country, and other advanced searching techniques.
4. To distinguish between full-text documents and abstract-only documents.
5. To distinguish between document formats (e.g. journals, theses, dissertations, course notes, presentations, books, etc).
6. To add "wildcard" searches (e.g. Librar*).
7. To enhance and improve the "similar-document" retrieval system.
8. To improve and expand the 'cited by' feature and offer a more comprehensive database in the final version.

Conclusion

Regardless of its limitations, the two unique advantages of Google Scholar are its use of citation indexing and its multidisciplinary coverage. Comparing Google Scholar to other databases is difficult given the differences in formats and coverage indexed in the resources. However, to stress the value of this free search tool, the author conducted a quick search on 'Webometrics' (see Tables 1 and 2). Since Google Scholar is a citation

index, it seemed reasonable to compare results to a commercial citation index. Google Scholar retrieves several documents that do not appear in scholarly journals but are part of the growing collection of scholarly information on the Web. Thus, Google Scholar again serves as a good complement to commercial databases. Ultimately, despite some disadvantages and the need for improvements, Google Scholar offers another resource for locating quality information. In comparison to commercial databases, it complements the researcher's needs by providing access to resources not covered by traditional citation indexes.

The increasing availability of online information resources and open access journals will place Google Scholar at the fingertips of most working scholars. It may also become an extremely important database for citation analysis. Improvements in the Google Scholar system will increase its use by those already familiar with it and gain it new users. In the future, the data available on Google Scholar may enable us to study the epidemiology of knowledge on the Web and may be the basis for bibliometric studies.

Acknowledgments

The author would like to thank Mrs. Marjorie Sweetko for her useful suggestions.

References

- Almind, T.C., and Ingwersen, P. 1997. Informetric analyses on the World Wide Web: methodological approaches to Webometrics. *Journal of Documentation* 53(4): 404-26.
- Banks, M. A. 2005. The excitement of Google Scholar, the worry of Google Print. *Biomedical Digital Libraries* 2(2 March). URL: <http://www.bio-diglib.com/content/2/1/2> [viewed September 20, 2005].
- Blachman, N. 2005. *Google guide: Making searching even easier*. URL: <http://www.googleguide.com/> [viewed September 20, 2005].
- Butler, D. 2004. Science searches shift up a gear as Google starts Scholar engine. *news@Nature*. URL: http://www.nature.com/news/2004/041122/pf/432423a_pf.html [viewed September 20, 2005].
- Eysenbach, G., & Diepgen, T. L. 1998. Towards quality management of medical information on the Internet: Evaluation, labelling, and filtering of information. *British Medical Journal* 317: 1496-1500.
- Garfield, E. 1955. Citation indexes for sciences: A new dimension in documentation through association of ideas. *Science* 122 (3159): 108-111.
- Garfield, E. 1979. *Citation indexing: Its theory and applications in science, technology and the humanities*. New York: Wiley Interscience.
- Google Scholar. 2005. About Google Scholar. URL: <http://www.scholar.google.com/scholar/about.html> [viewed September 20, 2005].
- Hamaker, C. & Spry, B. 2005. Google Scholar. *Serials* 18(1): 70-72.
- Horri, A. 1983. Citation analysis. *Nashr-e Danesh* 4(winter).
- Kennedy, S., & Price, G. 2004. Web search - Google big news: "Google Scholar" is born. *ResourceShelf*. URL: <http://www.resourceshelf.com/2004/11/wow-its-google-scholar.html> [viewed September 20, 2005].
- Jasco, P. 2004. Google Scholar Beta. *Peter's Digital Reference Shelf*. URL: <http://www.galegroup.com/reference/archive/200412/googlescholar.html> [viewed September 20, 2005].
- Jasco, P. 2005. Peter Jasco: Google Scholar and The Scientist (October 2005). URL: <http://www2.hawaii.edu/~jacso/extra/gs/> [viewed November 4, 2005]
- Thomson ISI. 2005. History of citation indexing. URL: <http://scientific.thomson.com/knowtrend/essays/citationindexing/history> [viewed November 4, 2005]
- Web of Science (WoS). URL: <http://www.isiknowledge.com/>
- Weinstock, M. 1971. Citation indexes. In: Kent, A. (ed.), *Encyclopedia of Library and Information Science*. New York: Marcel Dekker, Vol. 5: 16-41.

Editorial history:

paper received 23 September 2005;

final version received 8 November 2005

accepted 11 November