GOOGLE SCHOLAR: A TOOL TO SEARCH SCHOLARLY INFORMATION ON THE WEB

Bulu Maharana ***Sasmita Mahapatra***

Abstract

Google Scholar is the scholarly search tool of the world's largest and most powerful search engine, Google. It enables the users to search for scholarly literature including peer-reviewed papers, theses, books, preprints, abstracts, and technical reports from all broad areas of research. Google Scholar gets its information directly from publishers and by crawling the Web for scholarly content. Following a descriptive method of study, the paper provides a detail history and development timeline of Google and Google Scholar. Discusses in detail the features, search strategies, citation indexing values of Google Scholar.

Keywords: Google; Google Scholar; Citation Indexing; Search strategies; Search engines

1. Introduction

Google is one of the largest and most popular Web search engines. It is a tool for finding resources on the World Wide Web. Google can scan web pages to find instances of the keywords that the searcher has entered in the search box. The mission of Google is to "organise the world's information and make it universally accessible and useful." It is the largest search engine on the Web that receives over 200 million queries each day through its various services. Google is inspired from the term "Google", which is a common spelling of 'Googol' meaning 10100 (Brin and Page; 2006).

The Internet users today prefer Google as the one-stop searching point. They are imagining their library catalogue and their favourite journals searchable through Google's search box, which have been declared the winner for its simplicity. Surveys after surveys are telling that users are going away from libraries to Google (Sathyanarayana; 2004; p.1). Google has already been established as world's # 1 search engine (http://www.searchenginewatch.com). Since its inception, Google is in a process of continuous development of varieties of tools and services for easy retrieval of World's information. A few such tools and services include Google Directory, Google Image Search, Google News, Google email (Gmail), Google Catalogue Search, etc.

The extent to which the use of the Google search engine has permeated academic culture is illustrated by a single statistics from a recent survey of academic authors, which found 72% of the authors are using the Google search engine to search for scholarly articles (Swan and Brown; 2005). An important limitation on the use of the general search engine Google is the growth of content on the World Wide

Web (Friend; 2006). This has created real problems for users seeking the information most relevant to their needs. A reader with very little knowledge of a subject is easily led down paths into an information dilemma when so many Web sites are available. Hence, the development of specialist search services, such as Google Scholar, is an inevitable and a welcome development.

2. Brief Histroy and Development

Google, the most popular search engine was not developed in a single day. It was developed in 1995 as a part of a long term research project of two graduate students, Lawrence Page and Sergey Brin at Stanford University (Koller, 2004). They took up the challenge to solve one of the biggest challenges of 'retrieving relevant information from a massive set of data on the Internet'. Since 1995, Google has developed more rapidly till yet. The growth and development of different tools, services and achievements of Google has been chronologically arranged in the table below:

Table-1: Growth and development of Google tools and services

(Source: http://www.en.wikipedia.org/wiki/google)

1995	Lawrence Page & Sergey Brin started working on the project of retrieving relevant information from a massive set of data.
1996	Begun collaboration on a search engine called "BackRub".
1997	Link analysis made 'BackRub' a growing reputation.
1998	In September 1998, Google Inc. opened its door in Menlo Park, California. • PC Magazine named Google one of its Top 100 websites and the search engines for 1998.
1999	Google's appearance on Time Magazine's Top Ten Best Cybernetic list for 1999.
2000	Google Directory • Webby Award • People's Voice Award • Google Toolbar • 100 million search queries per day.
2001	Google Image Search • Google Catalog Search • Google Search Index reached 3 billion searchable web documents
2002	Adwords • Froogle • Google News • Search Engine Watch Awards for Outstanding Search Service
2003	Version 2.0 of the Google Toolbar was released
2004	Brand channel named Google as "Brand of the year" for 2003 • Local Search • Gmail • Google Desktop Search • Google SMS • Google Scholar
2005	Google Video • Google Maps • Google Web Accelerator • Google Earth • Google Book Search • Music Search • Google Reader.
2006	Google Video Store • Google Pack, a free collection of safe and useful software for improving the web experience

3. 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 in the year 2005 (Banks; 2005). 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 (Noruzi; 2005) . Basically, Google Scholar includes Web pages that either look like an article or other scholarly document. As such, it has become a gateway for accessing scholarly information on the Web. Every day, more and more scholarly information is available online and the users continue to discover new reasons to need access to this information.

The scope of Google Scholar includes journal articles from more than 30 publishers except Elsevier and American Chemical Society (Giustini; 2005), abstracts from bibliographic databases, and data from e-print servers. Some prominent collections include ACM, Annual Reviews, arXiv, Blackwell, IEEE, Ingenta, Institute of Physics, NASA Astrophysics Data System, PubMed, Nature Publishing Group, RePEc (Research Papers in Economics), Springer, and Wiley Interscience, although not all in their entirety. Many Web sites from universities and nonprofit organizations are included but only documents that seem like scholarly journal articles (Notess; 2005). From all these sources, Google Scholar displays several types of records including Web documents, article citations (only records), and book citation s (only records).

4. Components of A Google Scholar Search Result

Each Google Scholar search result represents a body of scholarly work. This may include one or more related articles, or even multiple versions of one article. Each search result contains bibliographic information, such as the title, author names, and source of publication. One set of bibliographic data is associated with the entire group of related articles and is the representative article for the group. This bibliographic data is based on information from the articles in the group, as well as on citations to these articles from other scholarly works.

Fig-1: Google Advance Scholar Search (Source: http://scholar.google.com)

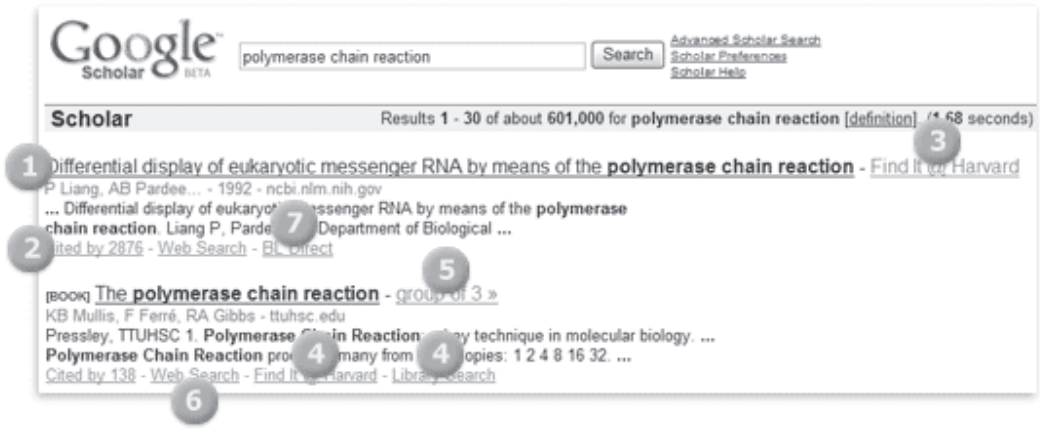


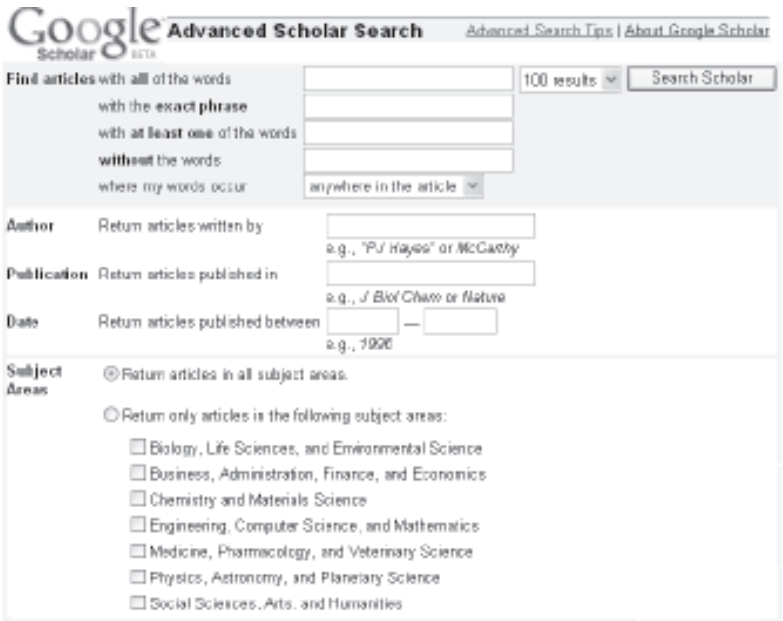
Table-2: Explanation of Links in a Google Scholar hit

1.	Title – Links to the abstract of the article, or when available on the web, the complete article.
2.	Cited By – Identifies other papers that have cited articles in the group.
3.	Library Links (online) – Locates an electronic version of the work through the affiliated library's resources. These links appear automatically if the searcher is on campus.
4.	Library Links (offline) – Locates libraries which have a physical copy of the work.
5.	Group of – Finds other articles included in this group of scholarly works. Examples include preprints, abstracts, conference papers or other adaptations.
6.	Web Search – Searches for information about this work on Google.
7.	BL Direct – Purchase the full text of the article through the British Library. Google receives no compensation from this service.

5. Search Strategies and Operators

Google Scholar has a quick and easy to search interface, which search in one simple search-box. The basic search involves 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. Besides, the Advance Scholar Search allows the user to restrict the search by publication, date and subject area (See Fig-2).

Fig-2: Screenshot of Google Scholar's Advanced Scholar Search



As with Google, the entire search tips, such as Boolean Operators (AND/OR/NOT), Phrase searching, synonyms (~), word order and wild cards are used for effectively searching Google Scholar. Besides, Google Scholar has its own set of 'operators' to help with fine-tuning a search. The following table enlists the operators used for basic search and advance search of Google Scholar.

Table-3: Basic and Advanced Search Operators of Google Scholar

Search Type	**Function**	**Operators**
Basic Search Operators		
Author Search	Searches for the word in the author's name.	author:"a ainstein"
Title Search	Returns results that include the some/all search in the title of the document/page.	Intitle:on the origin of species Allintitle: "on the origin of species"
Phrase Search	Returns results by a search phrase.	"Persian Gulf War"
Domain/Site Search	Searches for the word in the site/domain name. Limits searches to a special domain or site.	site:ac.uk "digital libraries" site:edu
URL Search	Searches for the some/all word(s) in the URL.	Inurl:information Allinurl:information literacy
+	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
File Type Search	Limits file type and retrieves a special file format	metadata filetype:pdf
Advanced Search Operators		
Publication Restrict Search	Limits the search by name of the journal.	American Journal of Clinical Nutrition
Date Restrict Search	Returns results published with in a specific date range.	2005-2006
Subject Restrict	Limits the search to a particular subject area by selecting a discipline from Google Scholar's subject area list.	"Engineering"

6. Google Scholar and Citation Indexing

Google Scholar provides most of the advantages of other citation indexes. 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 article 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). Alireza Noruzi (2005) reports that Google Scholar ranks search results by their degree of relevance 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. 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. 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.

7. Pros and Cons of Google Scholar

- Google Scholar has a number of important advantages when compared with other databases. Guistini (2005) and Noruzi (2005) have categorically discussed about the following key advantages and disadvantages of Google Scholar separately.
- It allows researchers to conduct broad-based, comprehensive, and multidisciplinary searches to discover hidden subject relationships on the Web.
- Google Scholar is not restricted to articles – preprints, technical reports, theses, dissertations, and conference proceedings are also indexed.
- Google Scholar is available on the Web, it contains full text of many articles and users can search all years simultaneously.
- Provides federated searching to scholarly resources and plays as a citation tracking tool.
- Provides single interface to access library catalogues, indexes and websites.
- Provides access to interdisciplinary, disparate and lots of full text content on the web.
- Fast, easy to use and easy to navigate.

Disadvantages

- Coverage is incomplete and unclear, not always scholarly.
- Delays in updating.
- Lacks advanced search features common in academic indexes.
- It does not include publications of Elsevier or American Chemical Society, two huge science and technology publishers.
- It claims to include all citations from Pubmed and other databases but independent tests show that it is not true yet.

8. Conclusion

Google Scholar has raised much interest amongst information community. Inspite of the limitations as discussed by several authors the unique advantages of Google Scholar are its use of citation indexing, multidisciplinary coverage and scholarly search features. Noruzi (2005) found comparing

Google Scholar to other databases is difficult given the differences in formats and coverage indexed in the resources. 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.

References

1. Brin, Sergy and Page, Lawrence (2003). The anatomy of a large scale hyper textual web search engine (Online article) Available at http://www-db.stanford.edu/~backrub/google.html. Retrieved on 20/07/06
2. Sathyanarayana, N V. (2004). De-branding the e-content: Outsmarting Google. Informatics Newsletter; September/October 2004. p. 1
3. Swan, Alma and Brown, Sheridan. Open access self-archiving: an author study May 2005 http://www.keyperspectives.co.uk/OpenAccessArchive/2005_Open_Access_Report.pdf. Retrieved on 01/07/06
4. Friend, Frederick J (January 2006). Google Scholar potentially good for users of academic information. Journal of Electronic Publishing. 9 (1).
5. David Koller (2004). http://www.stanford.edu
6. Banks, Marcus A (2005). The Excitement of Google Scholar and the worry of Google Print. Biomedical digital Libraries. 2 (2). Available at http://www.bio-diglib.com/content/2/1/2. Retrieved on 01/07/2006
7. Noruzi, Alireza (2005). Google Scholar: The new generation of citation indexes. LIBRI. 55(4), p. 170-180.
8. Guistini, Dean. Google Scholar ….and the rise of findability in (re)search (PowerPoint Presentations). Available at http://www.ubc.ca retrieved on 28/07/06
9. Butler, D. 2004. Science searches shift up a gear as Google starts Scholar engine. Nature, (November 2004). Available at http://www.nature.com/news/2004/041122/pf/432423a_pf.html. Retrieved on 20/07/2006
10. Google Scholar. 2005. About Google Scholar. Available at http://www.scholar.google.com/scholar/about.html. Retrieved on 20/07/2005
11. Mulliken, Adina. 2005. "What is Google Scholar?" Available at http://www.library.syr.edu/. Retrieved on 15/07/2006
12 Websites Referred
    - http://www.searchenginewatch.com
    - http://www.en.wikipedia.org/wiki/google
    - http://scholar.google.com
    - http://www.google.stanford.edu/