HARVESTING CAPABILITY OF GOOGLE SCHOLAR: A COMPARATIVE STUDY OF THREE MAJOR JOURNAL LISTS-DOAJ, HIGH WIRE AND BIOMED CENTRAL

Tahir Ahmed Batt
Instructor/Library Assistant at SKUAST-K, Shalimar-191201
E-mail: tahirbatt2015@gmail.com

Naasir Latief Hashia
E-mail: nasir4231@gmail.com

ABSTRACT

Google scholar a subset of the larger Google search engine, an incredible tool allowing researchers to search a wide array of scholarly literature across the Web, including scholarly journals, abstracts, peer reviewed articles, theses, dissertations, books, preprints, Power Point presentations and technical reports from universities. The paper seeks to examine the awareness and usage of Harvesting Capability of Google Scholar: A Comparative Study of Three Major Journal Lists-DOAJ, High wire and Biomed Central The present study was carried out by conducting a survey using questionnaire method.

Key words: DOAJ, High wire, Biomed Central, Google Scholar.


1. INTRODUCTION

Google scholar is a subset of the larger Google search engine. It is a freely-accessible Web search engine that indexes the full text of scholarly literature across an array of publishing formats and disciplines. Still in Beta version, Google scholar had its debut in November 2004 (Google, 2008).

Google Scholar enables a simple way to search specifically for scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research. It offers us to find articles from a wide variety of academic
Harvesting Capability of Google Scholar: A Comparative Study of Three Major Journal Lists-DOAJ, High Wire and Biomed Central

publishers, professional societies, preprint repositories and universities, as well as scholarly articles available across the web (Mayr and Kathrin, 2007).

Google scholar not indexes scholarly articles but in fact search results also lead to results for a large number of book titles and/or edited collections and anthologies. Although Google Scholar covers a great range of topical areas, it appears to be strongest in the sciences, particularly medicine, and secondarily in the social sciences (Vine, 2006).

Google Scholar searches for terms from the full text of scholarly sources, not just abstracts or titles, and so provides a high level of discoverability for many interdisciplinary topics. Google Scholar can be labeled a “blended” resource because it does not easily fit into a single resource category. It can function as a web-based scholarly search engine, a citation analysis tool, a portal to open access materials on the open web and in repositories, a connection to library journal subscriptions as well as book collections, and an adequate alternative in some cases to native subscribed databases or commercial federated search products (Hartman and Mullen, 2008).

Google scholar boosts the worldwide visibility and accessibility of the contents of the publishers of scholarly information. Much of Google Scholar's index derives from a crawl of full-text journal content provided by both commercial and open source publishers. Specialized bibliographic databases like OCLC's Open WorldCat and the National Library of Medicine's PubMed are also crawled. Since 2003, Google has entered into numerous individual agreements with publishers to index full-text content not otherwise accessible via the open Web (Vine, 2006)

Although Google does not divulge the sources that it covers, the number or names of publishers that have entered into crawling or indexing agreements with the company, it is easy to see why publishers would be eager to boost their content's visibility through a powerhouse like Google. As for the publisher partner program, more scholarly, commercial, and open access publishers have come on board. Publishers, such as Elsevier, who had exposed their content to Google Scholar, have brought a tremendously enhanced level of linked content to the search capability. Publishers recognize that use of their products will increase as more means of web discovery are provided (Hartman and Mullen, 2008).


Like the larger Google search engine index, the search interface of Google Scholar is simple and easy to use. Search options include some limiting criteria such as author, article title, journal title, publication year and subject area. Results are returned in a relevance-ranked order, which relies primarily on the full text of each document and its citation count. Thus, results emphasize documents that are cited more often, creating a bias toward older literature. In this regard, some sort options would be helpful (Neuhaus and Daniel, 2008).

Because so much of the content of Google Scholar's index comes from licensed commercial journal content, most users will discover that clicking on a link in Google Scholar's search results may reveal only an abstract—not full text—accompanied by a pay-per-view option. Institutions can configure Open URL link resolvers, to authenticate users to provide access to full-text content that is available through institutional subscriptions (Vine, 2006).
Still Google Scholar is no substitute for the various subject specific databases and indexing and abstracting services, which can provide more focused subject area coverage and that allows one to fine tune the search and build a highly targeted search strategy. However, there is more variety in Google Scholar and a higher number of results, but they are not necessarily as scholarly or relevant to the discipline (Gardner and Susanna, 2005).

1.1. Problem

Googlism is emerged as synonymous terms for searching the web. Scholarly community relay on Google particularly Google Scholar, which is intended and claimed to searching exclusively scholarly documents. The study was undertaken to understand and investigate the coverage of scholarly literature by Google Scholar and explored the deficiencies in the coverage. The study was intended to point up which web server is most important data provider and which information sources are highly representing and to explore its open access content accessibility.

1.2. Scope

The study undertaken has been confined to a freely accessible web search engine Google’s scholarly index i.e. Google Scholar. Three major open access Journal Lists i.e. Biomed Central, DOAJ j and High wire were used as a platform for accessing the journal titles to ascertain the harvesting capability of Google Scholar. The field of Microbiology is chosen from all the three repositories.

1.3. Objectives

- To determine the harvesting capability of Google Scholar from three major journal lists.
- To find out the number of journal titles found in the first ten or above ten results found in Google Scholar.
- To find out the document type of the record.
- To analyze the web server of the record whether it is a commercial or non-commercial one.

1.4. Methodology

Google Scholar offers a great way to index the contents of a large number of journal lists. Since it was impossible to query all the available journal lists. So the selection was made from three (3) major journal lists i.e.

- Directory of Open Access Journals (DOAJ)
- Biomed Central
- High wire

Precision was ensured by the use of Phrase search as a base for extracting the titles from Google Scholar.

2. LITERATURE REVIEW

A comprehensive literature search was undertaken multidimensionally. The search reflected various facets of Google Scholar visualizing its various potentialities as a scholarly tool accompanied with some other deficiencies. But only those titles were incorporated in literature review which are having a greater coincidence to my topic as can be visualized under:

Myhill (2005) studied that Google Scholar is an internet-based search engine designed to locate scholarly literature (including peer-reviewed papers, theses, books, preprints, abstracts
and technical reports from all broad areas of research) available across the Web, results are returned in a relevance-ranked format. Noruzi (2005) revealed the study that Google Scholar provides a free alternative or complement to other citation indexes or databases. The study explained that Google Scholar has an advantage of multidisciplinary coverage. In this study several suggestions were made for improving Google Scholar. Pomerantz (2006) analyzed that Google Scholar is the latest tool in a long history of information-seeking technologies that increasingly realize the goal of achieving 100 percent availability of information. However Google Scholar does not provide access to 100 percent of information resources in existence; but rather enables discovery of information resources, and allows for the possibility that these resources will be discoverable by the user. Neuhaus, et al. (2006) conducted the study on Google Scholar content and exposed its characteristics. The study revealed that Google Scholar freely offers researchers and libraries a database with great depth and breadth. Robinson and Wusteman (2007) carried out a small scale quantitative evaluation of the scholarly information search engine, Google Scholar and revealed that it has the highest precision, relative recall and ability to retrieve top ranked pages for scientific queries when compared with three popular search engines: Ask. Com, Google and Yahoo. However, the results of this small-scale study show Google Scholar to be most useful for searching for scientific information in comparison to the literature of other non-scientific disciplines. Jasco (2008) studied the pros and cons of Google Scholar and observed that its software still has serious deficiencies with basic search operations. In spite of the deficiencies and shoddiness of its software, the free Google Scholar can act as a huge and diverse multidisciplinary search engine with good at scholarly resource search process. Hartman and Mullen (2008) carried out the study about the Google Scholar’s integration into ARL Libraries’ web sites and revealed the continued acceptance of Google Scholar and integration of this resource on the web pages of ARL libraries. Moreover, the mean number of paths to Google Scholar more than doubled from 2005 to 2007.

Walters (2007) analyzed that the Google Scholar indexed the greatest number of core articles(93%) and provided the most uniform publishers and data coverage by evaluating its content with seven other databases(Academic Search Elite, Ageline, Article First, GEObASE, POPLINE, Social Science Abstracts and Social Science Citation Index). Haya, et al. (2007) analyzed that Google Scholar performed better in almost all measures. Responses to Google Scholar were more positive than to Metalib. However, the students were not overwhelmingly enthusiastic about either of the tools. Norris, et.al (2008) demonstrated the relative effectiveness of the search tools in finding out open access versions of peer reviewed academic articles and found that the general search engines Google and Google Scholar ranked first in their retrieval as compared to Open DOER and OIAster.

Gardner & Susanna (2005) came to a conclusion that Google Scholar yields more results and a greater variety in its types of sources along with a higher rate of relevancy, but less currency. However, Scholar’s lack of quality control and inability to let the user manipulate data make it less effective at finding scholarly material in the social sciences when compared with the fee-based databases i.e. PsycINFO, Social Science Citation Index, and ERIC. Burright (2006) studied that Google Scholar offers various search options for free academic resources on the Internet but its lack of authority control for basic data elements such as author names and publication titles greatly limits its ability to sustain a serious scientific and technical research audience as an exclusive source of literature. As a free Internet search engine, Google Scholar falls short of another free search engine such as PubMed.
Mayr & Walter (2007) showed the deficiencies in the coverage and up-to-dateness of the Google Scholar index by putting the queries against five journal lists. The study pointed the most important web servers as the data providers for this search engine and which information sources are highly represented. The conclusion drew revealed weaknesses in the accessibility of open access content and lack of transparency and completeness. Wilson (2007) explored that the content by Google Scholar varies greatly from database to database and from discipline to discipline by analyzing its content with that of forty seven databases. The results of this study serve to alert researchers and information professionals that Google Scholar has poor coverage in certain areas. It shows weakness in providing accessibility to commercial databases. Shultz (2007) observed that Google Scholar does not appear to be a replacement for PubMed, though it may serve effectively as an adjunct resource to complement databases with more fully developed searching features by comparing test searches between Pubmed and Google Scholar. However the suggestion was made to repeat the study in one or two years to determine if further refinements have improved their performance. The study by Neuhaus and Daniel (2008) provided an overview of the citation enhanced databases – Chemical Abstract, Scopus and Google Scholar, when used as a data source for performing citation analysis and found Google Scholar not a useful tool for citation analysis in its beta version.

3. GOOGLE SCHOLAR: AN OVERVIEW

Google Scholar is the scholarly search tool of the world’s largest and most powerful search engine. Google as a Web search engine has undoubtedly had a great impact on all those who search for information on the Web. The instant response, huge repositories, sophisticated search mechanism and relevance-ranking feature have combined to make Google Scholar the most popular Web search engine. In late 2004, Google launched several exciting products, one of which is a beta version of Google Scholar aiming to provide a single repository for scholarly information (Sadeh, 2006). Undoubtedly, Google deserves credit for making the first step in providing a tool for discovering scholarly information (Jasco, 2005). It entered the high-stakes world of research databases. The hype and hubbub surrounding this event were tremendous but understandable, considering the player involved. Google, such a monolithic Internet power, is synonymous in the minds of so many with the Internet. Indeed, one does not find information, one "Googles" it. With this overwhelming name recognition, a large clientele, and a tradition of successful spin-offs such as Froogle and Google Image Search, Google should have little difficulty persuading many to try its new "scholarly paper" search engine (Neuhaus, Chris et al.2006). Google's launch of Scholar indicates the growing sophistication of Internet searchers. It addresses concerns about the quality of information found on the Internet and integrates previously inaccessible, high-quality commercial sites with more reliable sites available on the public Internet (Henderson, 2005).

Google Scholar was developed by Anurag Acharya, an Indian-born computer scientist. Still in Beta version, Google Scholar had its debut in November 2004 (Google,2009; Mayr and Walter,2007; Jasco, 2008). Acharya, the chief engineer of Google Scholar, aimed to make Google Scholar as the "best possible scholarly search" and a "single place to find scholarly materials" covering all research areas, all sources, all time (Sadeh, 2006). Google Scholar is also thought to provide increased access to gray literature, as it retrieves more than journal articles and includes preprint archives, conference proceedings, and institutional repositories (Shultz, 2007). For its first year in operation, Google Scholar offers a broad range of discipline coverage with substantial depth in some areas (Wilson, 2007; Noruzi, 2005). Google makes it very easy and free to find scholarly information about any topic – an important service for those who do not have access to the most appropriate fee-based
indexing/abstracting databases which traditionally have helped in information discovery (Jasco, 2005). Google Scholar stands out not just for the technology employed, but also for the efforts made to restrict searches to scientific information. It offers us to find articles from a wide variety of academic institutions, professional societies, research groups, and preprint repositories around the world (Mayr and Walter, 2007). Google Scholar represents itself as one of the endorsed research tools for students and faculty in academic institutions. Google scholar not only indexes scholarly articles but in fact search results also lead to results for a large number of book titles and/or edited collections and anthologies. It, being extremely fast, covers a broad, heterogeneous range of information sources. Although Google Scholar covers a great range of topical areas, it appears to be strongest in the sciences, particularly medicine, and secondarily in the social sciences (Vine, 2006; Neuhaus et al., 2006; Kesselman and Watstein, 2005).

The working approach of Google scholar is same as in the pilot project CrossRef Search (www.crossref.org/crossrefsearch.html), the chosen Google Scholar approach is to work in cooperation with academic publishers. But what stands out is that Google Scholar, delivers results restricted to exclusively scientific documents and this constraint has yet to be consistently implemented by any other search engine. Google Scholar is a freely available service with a familiar interface similar to Google Web Search. Much of the content indexed by Google Scholar is stored on publishers’ servers where full-text documents can be downloaded for a fee, but at least the abstracts of the documents found will be displayed at no cost. The Google approach does, however, provide documents from the open access and self-archiving areas.

Google Scholar helps us identify the most relevant research across the world of scholarly research (Google, 2009). This free service alone is roughly equivalent to what several of the traditional online indexing/abstracting databases have been providing for a hefty fee (Jasco, 2005). Certainly, it is an important service for those who do not have access to expensive multidisciplinary databases such as the Thomson Scientific citation indexes or Scopus. 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 (Noruzi, 2005).

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” (Noruzi, 2005). Google Scholar aims to rank articles by how relevant they are to a query, considering the title and 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. So the most relevant results will always appear at the top of the retrieved results (Google, 2009; Gardner and Susanna, 2005). 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 (Noruzi, 2005).
Google Scholar searches for terms from the full text of scholarly sources, not just abstracts or titles, and so provides a high level of discoverability for interdisciplinary topics. Google Scholar can be labeled a “blended” resource because it does not easily fit into a single resource category in today's research library. It can function as a web-based scholarly search engine, a citation analysis tool, a portal to open access materials on the open web and in repositories, a connection to library journal subscriptions as well as book collections, and an adequate alternative in some cases to native subscribed databases or commercial federated search products. Its main area of interest to librarians is the potential usefulness of Google Scholar for finding materials in institutional and subject repositories worldwide. Many of these open access materials, along with published conference proceedings, can benefit from more widespread indexing. The open access literature has matured and evolved, and Google Scholar is always mentioned as an effective portal to this vast array of valuable content. With major emphasis, especially in STEM (science, technology, and medicine) fields on open access, Google Scholar may become a major player in the search for these materials. The point has been made that as open access materials proliferate, a need has arisen for an adequate means of searching across OAI compliant repositories, integrating this content with other scholarly materials. Google Scholar still in beta testing, has the potential to improve significantly before it becomes fully operational.

3.1. Selected Journal Lists
The journal lists contain collections of scholarly full-text documents maintained by government agencies, associations, universities, professional volunteer groups, as well as new and traditional scholarly publishers (directly or indirectly through their digital facilitators). The selected Journal Lists include DOAJ, Biomed Central and High wire.

3.1. DOAJ
Lund University Libraries has launched the Directory of Open Access Journals supported by the Information Program of the Open Society Institute along with SPARC - The Scholarly Publishing and Academic Resources Coalition. It was initiated during the first Nordic Conference on Scholarly Communication in Lund/Copenhagen October 2002 and the project started in January 2003 and was launched in May, the Directory of Open Access Journals embodies the conference's idea of creating a comprehensive directory of open access journals. DOAJ is sponsored by the National Library of Sweden, INASP, Swedish Library Association, Lund university. The Directory of Open Access Journals is a wonderful research tool for users wishing to quickly search many open access journal articles. The vision is to contribute to the movement towards free unlimited access to scientific results – the Open Access movement - by providing a service that contributes to the visibility of Open Access journals. When launched the directory listed only 375 titles. DOAJ is the authoritative source for open access journals with over 8 million monthly hits, up to 4000 journals published in almost 100 countries (Doaj, 2009).

Project Background
The proliferation of freely accessible online journals, the development of subject specific pre- and e-print archives and collections of learning objects provides a very valuable supplement of scientific knowledge to the existing types of published scientific information (books, journals, databases etc.). However these valuable collections are difficult to overview and integrate in the library and information services provided by libraries for their user constituency. At the First Nordic Conference on Scholarly Communication in Lund/Copenhagen the idea of creating a comprehensive directory of Open Access Journals was discussed. The conclusion was that it would be a valuable service for the global research
and education community. Open Society Institute (OSI) supported the initial project work. Available technologies make it possible to collect and organize these resources in a way that allow libraries worldwide to integrate these resources in existing services thus offering added value both for the service providers of these resources and for the global research and education community (Doaj, 2009).

The composition of the DOAJ collection is characterized by a very large number of publishers, each publishing a very small number of journals on different platforms, in different formats and in more than 50 different languages. The Directory aims to be comprehensive and cover all open access scientific and scholarly journals that use a quality control system to guarantee the content. It tends to increase the visibility and ease of use of open access scientific and scholarly journals thereby promoting their increased usage and impact. DOAJ service supports the OAI protocol for metadata harvesting (OAI-PMH). Thus, any other service can obtain records from DOAJ for inclusion in their collection. OAI is well established, and easy to use (Doaj, 2009).

The **selection criteria** of DOAJ are very broad and comprehensive:

**Contract Provisions**

Full Open Access, with no embargo, and with the ability for the user to “read, download, copy, distribute, print, search, or link to the full texts of these articles” per the Budapest Open Access Initiative, is mandatory for inclusion in DOAJ. This is a significant strength of DOAJ. No permission is needed to hand out articles in class, include in e-reserves, or share with anyone, inside or outside your own institution. The search provision is very important, as this will facilitate data mining and research based on artificial intelligence.

**Authentication**

As an Open Access initiative, no authentication is needed, another significant benefit of DOAJ. No authentication means less work, and less troubleshooting.

**Coverage**

- Subject: all scientific and scholarly subjects are covered
- Types of resource: scientific and scholarly periodicals that publish research or review papers in full text.
- Acceptable sources: academic, government, commercial, non-profit private sources are all acceptable.
- Level: the target group for included journals should be primarily researchers.
- Content: a substantive part of the journal should consist of research papers. All content should be available in full text.
- All languages

**Access**

- All content freely available.
- Registration: Free user registration online is acceptable.
- Open Access without delay (e.g. no embargo period).

**Quality Control**

- For a journal to be included it should exercise quality control on submitted papers through an editor, editorial board and/or a peer-review system.

**Periodical**

- The journal should have an ISSN (International Standard Serial Number)
However, the following contents are **not included** in the directory:

- Newsletters that mainly inform about an institution.
- Journals that charge any access or subscription fees.
- Journals that have an embargo period.

The Directory of Open Access Journals is a significant resource. DOAJ has developed a well-deserved reputation for quality, and is the world’s most authoritative list of fully Open Access, peer-reviewed titles. More than 10 percent of the world’s peer-reviewed journals are now included in DOAJ, making DOAJ among the world’s largest collections of peer-reviewed scholarly journals. There are more peer-reviewed journals in DOAJ than Science Direct; more non embargoed, peer-reviewed journals in DOAJ than in EBSCO’s Academic Search Premiere or Gale’s OneFile.

### 3.2. Biomed Central

BioMed Central (BMC) is a UK-based open access publisher specializing in open access publication. Of the new publishers, BioMed Central stands out by offering open access to full-text documents. BMC publishes over 180 scientific journals, and describes itself as the first and largest open access science publisher. In most of the journals published or hosted by BMC all the research articles are open access, but there are significant exceptions for other types of publications which require an institutional or a personal subscription to view other content, such as reviews or paper reports. Some of these are non-BMC journals which switched to using the BMC platform, such as the *Acta Veterinaria Scandinavica*, and became fully open access. The Chairman is Vitek Tracz and the Publisher is Matt Cockeril. BioMed Central, through its Open Access Charter, is committed permanently to maintaining this policy and has created a Board of Trustees to ensure that it is maintained.

BioMed Central (BMC) is an independent publishing house committed to providing immediate free electronic access to peer-reviewed biomedical research. The BioMed Central database is one of the largest providers of open access full-text journal articles covering all areas of Biology and Medicine. BioMed itself has been a mightily powerful open access database with links to open access articles. Hence, the content is not hidden behind passwords and can be crawled by major search engines.

Articles published on BioMed Central are:

- Available to all -- globally, free and without barriers to access of any kind
- Subject to full, high-quality peer review
- Published rapidly, taking advantage of web technology to improve the efficiency of manuscript provision, peer review and page layout
- Indexed by PubMed
- Permanently available online in full text through PubMed Central

To start a new research journal on BioMed Central, researchers must commit to the company’s copyright policy, by which neither BioMed Central nor the journal owns copyright, only the authors. Journal officers determine the scope, content, and quality parameters before launching a journal. Each journal must have an advisory or editorial board with at least three members who are grantees of a major funding organization to act as official guarantors of the scientific and ethical standards of the journal. The journal officers set a peer-review policy, e.g., entry level, number of referees, acceptance/rejection policies. BioMed Central offers automated peer-review support. Journals contain only primary research material or secondary material in the form of editorials or reviews commissioned or approved by the

editors. With these protections in place, BioMed Central will accept and permanently archive full-text searchable articles from the journal. In-house journal editors at BioMed Central work with external editors to produce the final product. Though committed to the goal of immediate free access to peer-reviewed primary research in the biological and medical fields (i.e., no subscription charges to users), BioMed Central does allow charging for access to online review material. It also allows charges for print versions of journal material. Individual printing of articles is available for a fee, but Adobe PDF files of articles can be downloaded at no charge.

BioMed Central journals are archived in several ways: on the company’s own servers, on PubMed Central’s servers, through CD-ROM collections developed by national libraries, through the Scirus archive, etc. BioMed Central also offers annual print compilations for a fee. The company has worked with ISI to develop citability standards, enabling its articles to be reliably cited in both conventional and online journals. Since the material is available without registration or subscription, search engines such as Google also index much of the research material. When users register (again for free), they can customize the site and set up individual interest profiles.

BioMed Central and its sister sites Chemistry Central and PhysMath Central are committed to providing immediate and free access to all the research they publish. BioMed Central and its sister sites Chemistry Central and PhysMath Central now has 284 Members and Supporter Members in 33 countries. Researchers from Member organizations can publish research articles in its journals without directly paying any article processing charges. The Membership Program enables academic and research institutions, societies, groups, funders and corporations to actively support open access in scholarly publishing, and help ensure the most widespread dissemination of the research published by their investigators or Members.

3.3. Highwire
HighWire Press is a division of the Stanford University Libraries, which produces the online versions of high-impact, peer-reviewed journals and other scholarly content. Recipient of the ALPSP Award for "Service to Not-for-Profit Publishing", HighWire partners with influential scholarly societies, university presses and publishers to create a collection of the finest, fully searchable research and clinical literature online. Together, these partners produce 71 of the 200 most-frequently-cited journals publishing in science. As of May 2009, it hosts the largest repository of high impact, peer-reviewed content, with 1245 journals and 5,909,326 full text articles from over 140 scholarly publishers. HighWire-hosted publishers have collectively made 1,894,133 articles free. Since 1995, with the launch of the Journal of Biological Chemistry (JBC), to the continuous online production of hundreds of prestigious journals, such as Science Magazine, the New England Journal of Medicine, PNAS and JAMA, HighWire has established an outstanding reputation for helping to disseminate primary scientific information on the Web (Highwire, 2009).

Mission Statement
The mission is to ensure the continuing success of independent, society-based and other scientific and scholarly publishers in their efforts to disseminate high-quality content worldwide. An enterprise of Stanford University, HighWire builds both the community and the technological environment that publishers need to thrive within the challenging business of electronic publishing. HighWire and its publishing partners develop and explore new ideas and emerging technologies to innovate sustainable solutions that meet the ongoing challenges of research communication (Highwire, 2009).
HighWire Press provides the largest number of open access articles, about 1,370,000 articles as of the end of July 2006. This is close to 40 per cent of the 3.6 million full-text article collection of HighWire Press. Although Ingenta, MetaPress and Allen Press also host some open access journals for traditional publishers, Highwire Press is the largest and most outstanding digital facilitator.

HighWire hosts nearly 200 of the journals of Oxford University Press alone (25 per cent of them open access), among the nearly 1000 journals of dozens of publishers. It hosts one third of the 200 highest impact factor journals in the sciences and social sciences. The third and fifth ranked journals of the Reproductive Biology Section, Biology of Reproduction and Reproduction, are also offered in open access format by HighWire Press on behalf of their respective publishers. It is the host also of Proceedings of the National Academy of Sciences (PNAS) with more than 28,500 open access full-text documents from this journal along with a plethora of free powerful and appealing add-on services. PNAS is the third highest impact factor serial publication in the Multidisciplinary category of the ISI Journal Citation reports (Jasco, 2006).

HighWire Press, show the best examples of the professional attitude of enlightening users about their free information services. It illustrates the volume and quality of the open access journals hosted by Highwire Press.

4. CONCLUSIONS
All results and conclusions in this study are current and based on sample tests (100 hits per query). Like the widely used, familiar search service Google Web Search, Google Scholar offers fast searching with a simple, user-friendly interface. The pros of this are that the search is free of charge and is done across interdisciplinary full-text collections. The Google Scholar approach offers some potential for literature retrieval, for example, automatic citation analysis and the ranking built up from this, and oftentimes direct downloading of full-text which is sometimes also described as a subversive feature but the service cannot be seen as substitute for the use of special indexing and abstracting services and professional database as meager percentage.

The results show that the expanding sector of open access journals (DOAJ list) is underrepresented among the servers. Google Scholar claims to provide “scholarly articles across the web,” the claim is true according to study as the ratio of articles from open access journals or the full-text is comparably high. In comparison with many abstracting and indexing databases, Google Scholar does not offer the transparency and completeness to be expected from a scientific information resource. Google Scholar can be helpful as a supplement to retrieval in abstracting and indexing databases mainly because of its coverage of freely accessible materials.

REFERENCES
Harvesting Capability of Google Scholar: A Comparative Study of Three Major Journal Lists-DOAJ, High Wire and Biomed Central

http://www.emeraldinsight.com/Insight/ViewContentServlet?contentType=Article&Filename=Published/EmeraldFullTextArticle/Articles/2390220810.html


