

2-2009

Using Google Scholar to Search for Online Availability of a Cited Article in Engineering Disciplines

Virginia A. Baldwin

University of Nebraska-Lincoln, vbaldwin2@unl.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/libraryscience>



Part of the [Library and Information Science Commons](#)

Baldwin, Virginia A., "Using Google Scholar to Search for Online Availability of a Cited Article in Engineering Disciplines" (2009).
Faculty Publications, UNL Libraries. 182.
<http://digitalcommons.unl.edu/libraryscience/182>

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Using Google Scholar to Search for Online Availability of a Cited Article in Engineering Disciplines

Virginia A. Baldwin

Engineering Branch Librarian, Physics and Patent and Trademark Librarian
University of Nebraska
Lincoln, Nebraska
vbaldwin2@unl.edu

Copyright 2009, Virginia A. Baldwin. Used with permission.

Abstract

Many published studies examine the effectiveness of Google Scholar (Scholar) as an index for scholarly articles. This paper analyzes the value of Scholar in finding and labeling online full text of articles using titles from the citations of engineering faculty publications. For the fields of engineering and the engineering colleges in the study, Scholar identified online access for 25% of the chemical engineering and 13% of the mechanical engineering citations. During the study the format that Scholar (which is in beta version) used to present the result set changed. This change now makes discovery of online access to full text of an article readily apparent when it occurs.

Introduction

In a 2007 study of citations from core ecology journals, Google Scholar located over 70% of the citations ([Christianson 2007](#)). Of 840 citations studied, 38 or 9% were found to be full-text accessible from off campus without authentication. In a comparison of Google Scholar and Compendex Meier ([2008](#)) revealed that Scholar found nearly 90% of the items during the time period 1990-2007 for the search examples and retrieval set used in the study. Covered were examples in eight fields of engineering. The Meier study did not analyze the full-text coverage of the article citations. In a 2008 study of the needs of users without access to Chemical Abstracts Service, Lafferty noted that compared to Academic Search Premier, in her "cursory searches, Google Scholar did a better job of locating older literature...(i.e., published more than 15 years ago)" for chemical information ([Lafferty 2008](#)).

Burright (2006) reported several challenges in using Google Scholar in her review of Scholar as a search engine. "First, the order in which results appear is unclear." In searching for a particular title this challenge was not a factor, at least for the current study since the citations which led to the article itself (whether full text, to a publisher's web site, or to a third party that provides pay-per-view access) were grouped together by Scholar as the first entry in the result set. "Second, the user has no control over the types of documents returned, their level of full-text access...." For the same reason above, this was not a factor when using Scholar to search for a particular title from a citation.

Earlier studies gave varying percentages that are not comparable because of the continuing advancement of Scholar coverage at least for the engineering disciplines in this study. One in particular is Markland's article (2006) that examines the efficacy of using Scholar to retrieve repository items from United Kingdom Institutional Repositories (IRs). We can now predict that for any IR registered with the Registry of Open Access Repositories coverage would be virtually 100%, though, of course, retrieval of a given deposited article depends on the likeness of the entered search terms to those in the repository entry for the article.

In his discussion of the value of Google Scholar in approaching the goal of 100 percent availability of information Pomerantz (2006) noted that our library users are using Google Scholar and are familiar with its interface. Therefore "it is to the libraries' benefit to see that it is used well." Friend (2006) foretold the ability of Scholar to provide identification of the various versions in its search results. At the time of this writing Scholar does identify and provide links to HTML and PDF versions of full text. "The strength of the open access movement," Friend continues, "is that it has arisen from and developed within the academic community it exists to serve." The development of the institutional repository is an outgrowth of that movement. Now that the Scholar search engine provides access to repository content, the cooperation that exists between the Registry of Open Access Repositories (ROAR) and Google has come full circle in that ROAR uses the Google Custom Search Engine in its own content search.

In her article, "Googling DSpace," Robin Peek (2004) asked, "What does a pilot project between Google and DSpace mean for scholarly communication? Is this a tipping point or merely a blip on the radar?" Not a blip at all, today the Google Scholar Custom Search Engine is employed by the Registry of Open Access Repositories to search worldwide repositories no matter what software is used to create them.

Another open access breakthrough involving Scholar is open URL linking to library holdings as discussed in a Grogg (2005) article. While identifying some librarian concerns, such as the requirement to give Google electronic holdings information, Grogg concluded that celebration is in order as we explore partnerships that make it easier for our users to access needed publications. In the current study we are not discussing Scholar as the first choice to locate holdings but rather as a choice of last resort when a paid subscription is not available to the searcher.

Notess (2005) described three types of Google Scholar records. One of these is citation-only type in a Scholar result set. This type of record does not have a link to any additional information and the citation-only type occurs in current Scholar result sets as well. Other records in the result set

in a screen capture in the Notess article show an earlier version of the result display. Figure 1 is part of a screen for a Scholar result set from this current study. An article title was entered in the search box, enclosed in quotation marks. Because the article is available in full text through the University of Nebraska-Lincoln (UNL) institutional repository (IR), the first record in the result set is for that occurrence of the exact title. Furthermore, the record has an indicator at the end of the title that there is a link to the PDF version of the article through unl.edu. Clicking on unl.edu takes the searcher directly to the article in the UNL IR. This screen also shows an example of the citation-only type of Google Scholar record as identified by Notess above.

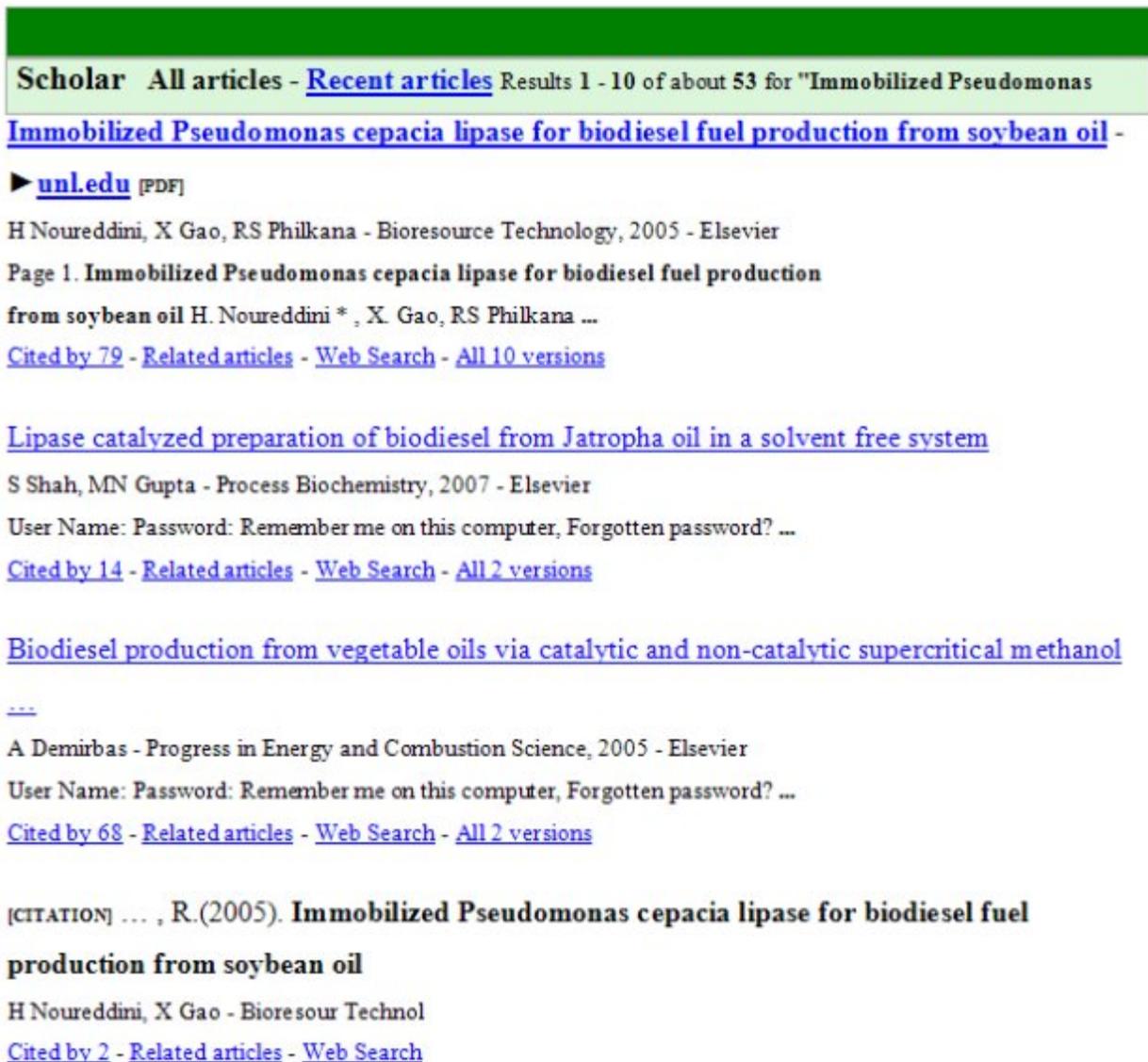


Figure 1 Google Scholar Screen Showing Result Set Current at the Time of This Study

There are many sources through which online access to full text is available. Google Scholar indexes publisher web sites, PubMed Central (PubMed), institutional repositories, preprint

archives, etc. It also locates full text that results from research groups posting articles online for their own use and failing to make access proprietary. In its current form this beta version of Google Scholar now labels and pulls to the top of its result set those items for which a PDF image or HTML version of an article can be freely accessed online. The fact that its coverage is so extensive means that a non-null result set will invariably appear from a title search. This was so universally true that whenever a search provided no results, the cause would be some quirk in the way the title was represented in the citation. This universality of coverage differs from other studies reported above. Among the factors that may explain this difference are:

1. Timeliness. Only citations of already published articles were examined, rather than the most recently published articles themselves.
2. Overall coverage. This is a 2008 study and Google Scholar coverage appears to expand through time to new sources.
3. Discipline coverage. There may be differences depending on the studied discipline. This study only covered engineering disciplines.

Several services are now operating on the World Wide Web that contribute to the universality of online representation of citations in science and engineering fields. Among them are:

1. Cat.inist.¹ <http://cat.inist.fr/>
2. Chemical Engineering Research Information Center.² <http://www.cheric.org/>
3. National Center for Biotechnology Information.³ <http://www.ncbi.nlm.nih.gov/>
4. Centre for Agricultural Bioscience International.⁴ <http://www.cabi.org/home.asp>
5. IngentaConnect.⁵ (<http://www.ingenta.com/>)

Methodology

The author used the Thomson Reuters database Web of Science (WOS) to collect the citations of engineering faculty publications from five engineering colleges.⁶ The study was done from outside the IP range of the author's institution so there was no confusion about whether access to full text resulted from affiliated subscriptions. The availability of each citation was determined using Google Scholar as well as the Registry of Open Access Repositories, and the OAI protocol, Oaister. Only the Scholar results are presented here and only those citations for which WOS included the title could be assessed. WOS does not include the titles of monographs, theses, and such document types, nor does it include the titles for incomplete or erroneous citations in its database unless the latter can be resolved and corrected (Marie McVeigh, personal communication, August 8, 2008).

For each citation, the full title as represented in WOS was entered surrounded by quotation marks to accommodate exact title searching in the Scholar basic interface search box. This has the same effect as entering the title without quotes in the "with the exact phrase" search box in the Advanced Scholar Search interface. If the response was "Your search ... did not match any articles" the search was repeated using an abbreviated title or alternative form of words in the title. With few exceptions such changes would result in at least one occurrence of the title in the result set. Failure of the search engine to produce a result is invariably a function of the

difference between the WOS representation of the title and the title as it is represented online. Christianson (2007) found similar "discrepancies" in her study. Differences include:

1. Use of the Greek character in the title vs. the name of the Greek character
2. The appearance of a part in the title which can be represented in several ways, such as .1. or Part 1, etc.
3. A single digit number vs. the word for the number.
4. The appearance of subscripts, superscripts, and parentheses in the title.
5. Words run together, such as invitro vs. in vitro

Often merely removing the portion of the title that contains these "anomalies" and everything before or after it will enable the database to bring up records for the title. Sometimes the title then becomes too short with the result that too many erroneous versions are retrieved to sort through. In this case adding unambiguous words from the title back into the search box with a second set of quotation marks usually produces record(s) for the desired article. If the correct article is not produced, Scholar often suggested an alternative word or spelling of a word in the title. The author's experience is that choosing the alternative rarely, though sometimes, produced the desired result.

Scholar now clearly labels the occurrence in its result set of a link that will take the searcher to online access if it occurs, regardless of affiliated subscription, to the title searched. It puts that title, again almost invariably, as the first entry in its result set. Occasionally a quick scan of the first screen of the result set will find another entry that goes to full text. Online access results from several sources. Among them are:

1. Institutional repository deposition
2. NIH funded articles in PubMed
3. Availability in disciplinary repositories
4. Availability in open access journals or as an open access article in a journal
5. Other forms of online availability such as research groups posting articles related to their research without limiting access to their group members.

For each title searched, the form of the open access was recorded. In some cases there was more than one for. For example, the article could be available through PubMed as well as through the publisher's web site. During the course of this study articles on PubMed could more frequently be found on the publisher's web site and this seems more and more to be common practice.

Results

Availability of article full text as determined by searching using Google Scholar was compiled for citations of faculty in mechanical and chemical engineering. Thirteen percent of the citations were found for mechanical engineering and nearly twice that amount for chemical engineering. The difference is largely explained by the greater occurrence of chemical engineering cited articles in PubMed Central and is likely indicative of a greater portion of research in this discipline being a result of NIH funding. [Table 1](#) identifies the percentage of citations that were

found for each subject area. Table 1 also shows the percentages of available articles that were found from each of the five sources:

1. Institutional repositories
2. PubMed Central
3. Disciplinary repositories (ArXiv, Citeseer, etc.)
4. Open access at the publisher's site
5. Other

Table 1
Percentages of Referenced Articles that were Found Online by Google Scholar

Engineering department/(% of citations found by Google Scholar)	% of references in an Institutional Repository	% of references in PubMed Central	% of references in a disciplinary repository (Citeseer, arXiv, etc.)	% of references in open access at publisher's site	% of references in other forms of open access
Mechanical/ (13%)	10	9	11	7	66
Chemical/ (25%)	13	48	4	27	50

These percentages add up to greater than 100% since many articles appeared in more than one form of 1-4 above. In the "other" category, however, are only those articles that were not available in 1-4 above. These percentages are depicted in the bar charts in Figures 2 and 3.

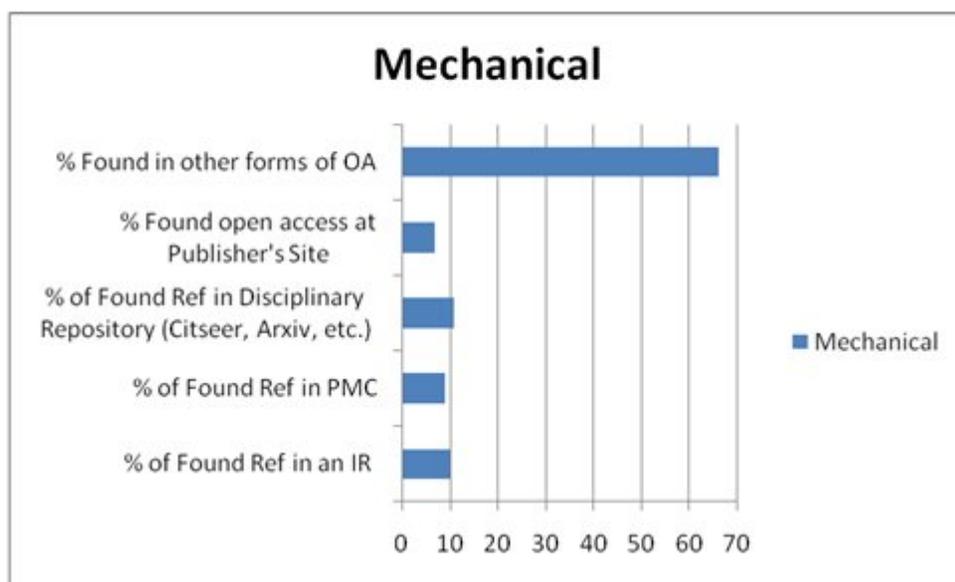


Figure 2. Percentages of Reference Articles of Mechanical Engineering Faculty that were found online by Google Scholar

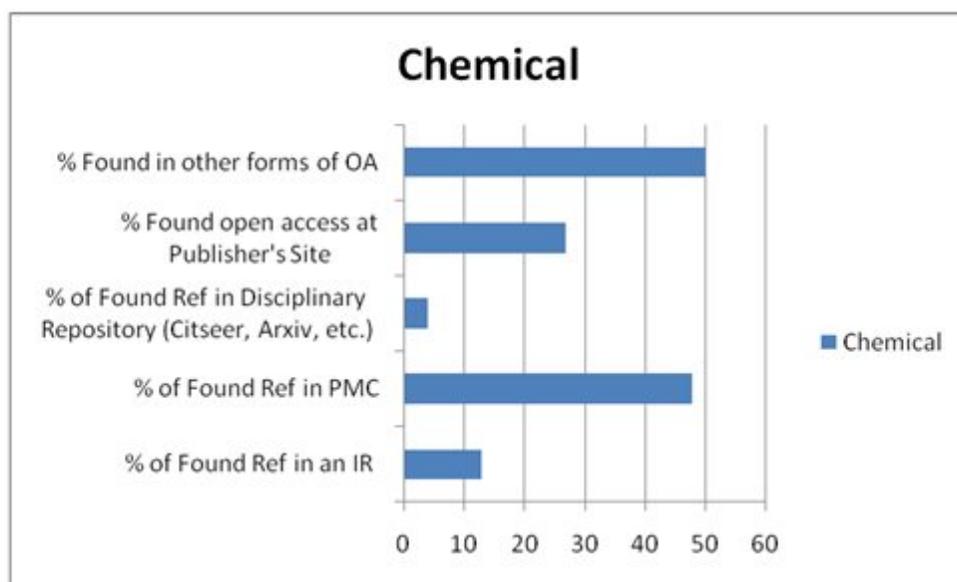


Figure 3. Percentages of Reference Articles of Chemical Engineering Faculty that were found online by Google Scholar

Conclusions

While still in beta version, Google Scholar has value in locating full text of articles cited by engineering faculty in their publications when not available through institutional subscription. In its current form Scholar search results simplify identification of the record that connects to the full text when it is available. Furthermore, if the user's institution has set up Open URL Linking with Google Scholar, those articles which are available through the institution's subscriptions will be linked from the Scholar result set as well. There are other improvements that would make Scholar more useful as a general search engine for content on a topic. An example would be the ability to clump records with duplicate titles or portions of titles that appear because the title is cited in several sources (as it does now with duplicate content) to reduce the size of the result set. Another example would be to offer functional ways to make search results more precise. Google has consistently shown responsiveness to user needs while maintaining the goal of its search engine creator, Anurag Acharya, of making Google Scholar "the one place to go for scholarly information across all languages and disciplines ([Giles 2005](#)).

The results presented here are based on a study of citations in recently published articles of engineering faculty. Further studies are needed to determine the timeliness of Scholar coverage in a search for very recent citations or recent publications. Similar studies in other disciplines can be done to assess coverage in them by Google Scholar.

How does this study affect the role of librarians? Many users are now using Scholar to locate articles. As our library budgets contract relative to the continual expansion of journal costs, would promotion of effective use of Scholar add value to library services, especially the function of providing access to publications scholars need in their research? In the course of library instruction or through online guides librarians can prompt users to examine title idiosyncrasies and drop out portions of titles in Scholar searching when a title search yields no results.

Perhaps as importantly, while encouraging their users to use Scholar, librarians can also encourage faculty and graduate students to post onto their own or co-author's IRs, and to post articles they want to share with their research groups in ways that comply with copyright law. The result would be more access to their own publications by other researchers. Ultimately these actions could produce that tipping point that brings on universal availability of published studies.

Notes

¹ Cat.inist provides "about 15 million bibliographic records (back to 1973) of documents held in the INIST/CNRS collections and covering all fields of worldwide research in Science, Technology, Medicine, Humanities and Social Sciences" and a direct link to their document copy ordering service (<http://cat.inist.fr/>). "For over a decade, the Institute for Scientific and Technical Information (INstitut de l'Information Scientifique et Technique -- INIST) at the French National Center for Scientific Research ([CNRS](#)) has been designing and developing services to facilitate access to and valorization of scientific information produced by international basic and applied research communities" (<http://international.inist.fr/>).

² Chemical Engineering Research Information Center (<http://www.cheric.org/>) is a search interface from Seoul, Korea.

³ The National Center for Biotechnology Information was "established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information -- all for the better understanding of molecular processes affecting human health and disease" (<http://www.ncbi.nlm.nih.gov/>).

⁴ Chemical Engineering Research Information Center, CABI, "is a not-for-profit organization specializing in scientific publishing, research and communication" (<http://www.cabi.org/home.asp>).

⁵ IngentaConnect is a search interface for "25,481,082 articles, chapters, reports and more..." (<http://www.ingenta.com/>)

⁶ The study on which this evaluation of Google Scholar is based examined citations of faculty in mechanical, civil, and chemical engineering departments of five universities. The universities selected were: University of Michigan, Ohio State University, Georgia Institute of Technology, University of Nebraska-Lincoln, and Massachusetts Institute of Technology. These five universities were selected based on the content in their institutional repositories (IRs) as reported on the web site <http://roar.eprints.org/> which ranks IRs by content. The results reported here are based on 4,500 citations in the fields of mechanical and chemical engineering.

References

Burright, Marian. 2006. Google Scholar -- science & technology. *Issues in Science & Technology Librarianship*. [Online]. Available: <http://www.istl.org/06-winter/databases2.html>. [Accessed August 3, 2008].

Christianson, Marilyn. 2007. Ecology articles in Google Scholar: levels of access to articles in core journals. *Issues in Science & Technology Librarianship* [Online]. Available: <http://www.istl.org/07-winter/refereed.html>. [Accessed December 17, 2008].

Friend, Frederick J. 2006. Google Scholar: potentially good for users of academic information. *Journal of Electronic Publishing*. 9(1). [Online]. Available: <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;cc=jep;q1=3336451.0009.1%2A;rgn=main;view=text;idno=3336451.0009.105>. [Accessed August 3, 2008].

Giles, Jim. 2005. Science in the web age: start your engines. *Nature* 438(7068): 554.

Grogg, J. E. and Ferguson, C. L. 2005. OpenURL linking with Google Scholar. *Searcher*. 13(9): 39-46.

Lafferty, Meghan. 2008. Does chemistry content in a state electronic library meet the needs of smaller academic institutions and companies? *Issues in Science & Technology Librarianship*. [Online]. Available: <http://www.istl.org/08-winter/refereed4.html>. [Accessed December 17, 2008].

Markland, M. 2006. Institutional repositories in the UK: what can the Google user find there? *Journal of Librarianship and Information Science*. 38(4): 221-228.

Meier, John J. and Conkling, Thomas W. 2008. Google Scholar's coverage of the engineering literature: an empirical study. *The Journal of Academic Librarianship* 34(3): 196-201.

Notess, Greg R. 2005. Scholarly web searching: Google Scholar and Scirus. *Online*. 29(4): 39-41.

Peek, Robin. 2004. Googling DSpace. *Information Today*. 21(6): 17-18.

Pomerantz, J. 2006. Google Scholar and 100 percent availability of information. *Information Technology and Libraries*. 25(2): 52-56.