

Literature searches with Google Scholar: Knowing what you are and are not getting

Jerome V. DeGraff, USDA Forest Service, 1600 Tollhouse Road, Clovis, California 93611, USA, jdegraff@fs.fed.us; **Nicholas DeGraff**, Digital Marketing Consultant, 834 Price Court, Sacramento, California 95815, USA; **H. Charles Romesburg**, Utah State University, Dept. of Environment & Society, Logan, Utah 84322, USA

ABSTRACT

Whether you are a student developing a senior thesis or a geoscientist preparing a research proposal, finding relevant concepts, data, and information produced by other geoscientists is a crucial step to eventual success. The quote, “If I have seen farther it is by standing on the shoulders of giants,” attributed to Sir Isaac Newton, acknowledges this fact. Our expanding knowledge base, increasing professional specialization, and greater involvement in interdisciplinary studies make it less and less likely we will know all the useful information important to any study we may consider undertaking. For this reason, we turn to scholarly literature to remedy any important deficiencies. For both students and researchers, this has increasingly meant employing Google Scholar.

SEARCHING LITERATURE IN THE DIGITAL AGE

Some of the same technological changes that enhance our ability to collect relevant data also facilitate our ability to search scholarly literature. Abstracting and indexing of published literature is now provided in computerized, searchable bibliographic databases. These include both discipline-specific databases, such as GeoRef, produced by the American Geological Institute, and multidisciplinary ones, such as Web of Science from Thomson Reuters and Scopus from Elsevier (Walters, 2011). Internet-based platforms are available for accessing these databases.

In November 2007, Google Scholar was introduced as another means for geoscientists to conduct a computer-based literature search. It is an Internet search engine rather than a computerized bibliographic database. Access is through the widely used Google Internet portal. Like all new methods and ideas, thoroughly examining the strengths and limitations of Google Scholar ensures we understand what it does or does not actually deliver.

Before looking closely at Google Scholar’s use in literature searches, it is important to understand how searching differs between Google and Google Scholar. Both are search engines

owned by Google Inc. and both use proprietary software to identify Web-based links relevant to the search terms entered by the user. The terms entered into a Google search initiate a hunt through all publically accessible files on Web servers connected to the Internet that match those words. The Google Scholar search engine utilizes a variant of this software that searches for the user’s terms within only scholarly publications as defined by the source servers; e.g., universities and scientific publishers (Walters, 2011).

How the results obtained can differ is demonstrated by a search we conducted on 27 Jan. 2013 using the search term “Indian ocean tsunami.” The Google search returned 6.8 million results with the ten listed on the first screen page including an entry about the 2004 event on Wikipedia, a news item on National Geographic’s website, and reports from six major national and international news organizations. Some news items related to the 2004 event and others to the tsunami watch that occurred after the 11 Apr. 2012 earthquake. The remaining two entries consisted of collected still images and videos about the tsunami.

In contrast, Google Scholar returned a comparatively modest 28,000 results. Except for four of the ten entries, the first page provided links to articles in scholarly journals ranging from *Nature* to the *International Journal of Hospitality Management* (Elsevier). The other entries were technical pages on a university website and technical reports on websites established by international donors for relief efforts and a government disaster response agency. This illustrates the very different search algorithms employed by Google and Google Scholar in terms of result numbers, content, and sources. It is worth noting that searches using this term over time returned widely differing result numbers for Google but not for Google Scholar. This reflects the more dynamic nature of Internet content as a whole compared to that part defined as scholarly content by Google Scholar.

MECHANICS OF BIBLIOGRAPHIC DATABASES AND GOOGLE SCHOLAR

To fully explore any advantages or disadvantages of Google Scholar requires understanding how a search engine differs from a computerized bibliographic database. The content of bibliographic databases is developed through indexing done by the organization producing them. Indexed entries are added to these databases by organizations’ employees based on a set of criteria related to specific sources and standards. The GeoRef thesaurus is

GSA Today, v. 23, no. 10, doi: 10.1130/GSAT175GW.1.

¹GSA Supplemental Data item 2013316, GeoRef and Google Scholar search results for “wildfire-related debris flows,” is available online at <http://www.geosociety.org/pubs/ft2013.htm>. You may also obtain a copy by writing to gsatoday@geosociety.org.

an example of an indexing standard used in compiling that particular database. This compilation approach ensures the scholarly content and quality of these databases (Gray et al., 2012). Available bibliographic databases with content in the geological sciences are offered via subscriptions. Many students and researchers access these databases via subscriptions paid for by their libraries or organizations.

GeoRef is a bibliographic database familiar to most geoscientists because it is specifically targeted to our professional needs. This traditional abstracting and indexing service assumes its audience is informed geoscientists familiar with the defined vocabulary used by GeoRef to describe the subject content of the database (Tahirkheli, 2009). Available through various interfaces, GeoRef searches can be limited by various parameters such as date, journal articles, source language, or recent database updates. As Tahirkheli (2009) points out, a searcher can examine indexes providing the author name, journal name, and publication type before choosing a specific entry. Authors found in a search may then be searched separately using associated live links. Similarly, citations found during the search may have links to the full-text article (Tahirkheli, 2009).

Google Scholar is designed for use by many different disciplines including the geosciences. It is accessed via the Google Internet portal. Retrieval via Google Scholar requires that the article be in digital format on the Internet. Gray et al. (2012) and the inclusion guidelines provided by Google Scholar (<http://scholar.google.com/intl/en/scholar/inclusion.html>) highlight that effectively finding documents depends partly on the quality of the metadata for these electronic documents. Users enter their search terms, such as article title, author, or key words in a manner similar to the familiar Google search (Tahirkheli, 2009). The search algorithm will return those links that most closely match the terms entered where the full text is available (Tahirkheli, 2009). Where there are many articles found, it will provide those having the most links to other Internet pages first and then others following in descending order. Thus, papers with similar key words or titles would be represented with the one most often cited being listed first. Because this may place more recent relevant articles farther down the list, a user interested in primarily recent articles can limit the search by a year range or publication after a particular year.

The search term “wildfire-related debris flows” was recently used to illustrate differences between GeoRef and Google Scholar (see GSA Supplemental Data¹ for more information) GeoRef returned 127 citations compared to 276 from Google Scholar. Google Scholar included 85% of the GeoRef citations with the missing ones being limited to conference proceedings, government reports, technical publications, foreign language journals, and theses. Both GeoRef and Google Scholar distinguished abstracts from full articles. Retrieving full-text articles for citations returned by GeoRef and Google Scholar may require payment to the publisher. However, free articles were available in PDF format for 88% of citations returned by Google Scholar. They were available from open-access journals or via links to organizational sites where authors had posted their publications.

Repeated evaluations of Google Scholar for both simple and advanced searches have demonstrated its ability to deliver results equivalent to those provided by traditional computerized bibliographic methods (Hightower and Caldwell, 2010; Walters, 2011; Gray et al., 2012). Given its generally high precision and recall

compared to other databases, it is a valuable tool for literature research (Walters, 2011).

DISCUSSION

Literature research is done by undergraduate and graduate students as part of their learning process and by academic or institutional researchers as part of their work. Applied geoscientists in government organizations and private industry conducting scholarly research find Google Scholar attractive because it is accessible outside academic institutions or research organizations holding subscriptions to traditional bibliographic services. Also, using Google Scholar is free to anyone with an Internet connection and can achieve useable results without knowledge of sophisticated search functions or familiarity with the vagaries of different interfaces (Gray et al., 2012).

Equally attractive is the ability to quickly access full-text articles via the link associated with the citation found through Google Scholar (Tahirkheli, 2009; Walters, 2011). Continued growth in the number of open-access journals and institutional repositories will increase the number of articles readily available for free via Google Scholar. This trend especially benefits geoscientists unaffiliated with organizations that provide access to journal subscriptions. It will also help students from other countries who became accustomed to ready access to journal articles while obtaining their degrees at universities and colleges in the United States. A number of publishers are digitizing past issues of journals, too.

Google Scholar users should recognize that this technology continues to change. Just as many negative reactions to the beta version initially released are no longer relevant to the current version, tomorrow's version will be different and may include new positive and negative elements. Recognizing that Google search retrieval remains based on software, it may return some material that is not vetted for the quality, accuracy, and authority expected from traditional bibliographic services (Gray et al., 2012). Consequently, geoscientists should stay informed on changes to ensure that the results they are getting conform to their standards and expectations. When possible, it will continue to be a good practice to conduct literature searches utilizing the specific advantages offered by both Google Scholar and traditional bibliographic databases.

REFERENCES CITED

- Gray, J.E., Hamilton, M.C., Hauser, A., Janz, M.M., Peters, J.P., and Taggart, F., 2012, Scholarish: Google Scholar and its value to the sciences: *Issues in Science and Technology Librarianship*, Summer 2012, doi: 10.5062/F4MK69T9.
- Hightower, C., and Caldwell, C., 2010, Shifting Sands: Science researchers on Google Scholar, Web of Science, and PubMed, with implications for library collections budgets: *Issues in Science and Technology Librarianship*, Fall 2012, doi: 10.5062/F4V40S4J.
- Tahirkheli, S.N., 2009, GeoRef and Google Scholar—similarities and differences: *Proceedings, Geoscience Information Society*, v. 38, p. 39–43.
- Walters, W.H., 2011, Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases: *Portal: Libraries and the Academy*, v. 11, no. 4, p. 971–1006.

Manuscript received 10 Mar. 2013; accepted 8 July 2013.