

Google Scholar and 100 Percent Availability of Information

Jeffrey Pomerantz

This paper discusses Google Scholar as an extension of Kilgour's goal to improve the availability of information. Kilgour was instrumental in the early development of the online library catalog, and he proposed passage retrieval to aid in information seeking. Google Scholar is a direct descendent of these technologies foreseen by Kilgour. Google Scholar holds promise as a means for libraries to expand their reach to new user communities, and to enable libraries to provide quality resources to users during their online search process.

Editor's Note: This article was submitted in honor of the fortieth anniversaries of LITA and ITAL.

Fred Kilgour would probably approve of Google Scholar. Kilgour wrote that the paramount goal of his professional career is "improving the availability of information."¹ He wrote about his goal of achieving this increase through shared electronic cataloging, and even argued that shared electronic cataloging will move libraries toward the goal of 100 percent availability of information.²

Throughout much of Kilgour's life, 100 percent availability of information meant that all of a library's books would be on the shelves when a user needed them. In proposing shared electronic cataloging—in other words, online union catalogs—Kilgour was proposing that users could identify libraries' holdings without having to travel to the library to use the card catalog. This would make the holdings of remote libraries as visible to users as the holdings of their local library.

Kilgour went further than this, however, and also proposed that the full text of books could be made available to users electronically.³ This would move libraries toward the goal of 100 percent availability of information even more than online union catalogs. An electronic resource, unlike physical items, is never checked out; it may, in theory, be simultaneously used by an unlimited number of users. Where there are restrictions on the number of users of an electronic resource—as with subscription services such as NetLibrary, for example—this is not a necessary limitation of the technology, but rather a limitation imposed by licensing and legal arrangements.

Kilgour understood that his goal of 100 percent availability of information would only be reached by

leveraging increasingly powerful technologies. The existence of effective search tools and the usability of those tools would be crucial so that the user would be able to locate available information without assistance.⁴ To achieve this goal, therefore, Kilgour proposed and was instrumental in the early development of much library automation: he was behind the first uses of punched cards for keeping circulation records, he was behind the development of the first online union catalog, and he called for passage retrieval for information seeking at a time when such systems were first being developed.⁵ This development and application of technology was all directed toward the goal of improving the availability of information. Kilgour stated that the goal of these proposed information-retrieval and other systems was "to supply the user with the information he requires, and only that information."⁶

Shared catalogs and electronically available text have the effect of removing both spatial and temporal barriers between the user and the material being used. When the user can access materials "from a personal microcomputer that may be located in a home, dormitory, office, or school," the user no longer has to physically go to the library.⁷ This is a spatial barrier when the library is located at some distance from the user, or if the user is physically constrained in some way. Even if the user is perfectly able-bodied, however, and located close to a library, electronic access still eliminates a temporal barrier: accessing materials online is frequently faster and more convenient than physically going to the library. Electronic access enables 100 percent availability of information in two ways: by ensuring that the material is available when the user wants it, and by lowering or removing any actual or perceived barriers to the user accessing the material.

Library automation

Weise writes that "for at least the last twenty to thirty years, we [librarians] have done our best to provide them [users] with services so they won't have to come to the library."⁸ The services that Weise is referring to are the ability for users to search for and gain access to the full text of materials online. Libraries of all types have widely adopted these services: for example, at the author's own institution, the University of North Carolina at Chapel Hill, the libraries have subscriptions to approximately seven hundred databases and provide access to more than 32,000 unique periodical titles; many of these subscriptions provide access to the full text of materials.⁹ Additionally, the State Library of North Carolina provides a set of more than one hundred database subscriptions to all academic and public libraries around the

Jeffrey Pomerantz (pomerantz@unc.edu) is Assistant Professor in the School of Information and Library Science, University of North Carolina at Chapel Hill.

state; any North Carolina resident with a library card may access these databases.¹⁰ Several other states have similar programs. By providing users with remote access to materials, libraries have created an environment in which it is possible for users to be remote from the library. Or rather, as Lipow points out, it is the library that is remote from the user, yet the user is able to seek and find information.¹¹

This adoption of technology by libraries has had the effect of enabling and empowering users to seek information for themselves, without either physically going to a library or seeking a librarian's assistance. The increasing sophistication of freely available tools for information seeking on the Web has accelerated this trend. In many cases, users may seek information for themselves online without making any use of a library's human-intermediated or other traditional services. (Certainly, providing access to electronic collections may be considered to be a service of the library, but this is a service that may not require the user either to be physically in the library or to communicate with a librarian.) Even technically unsophisticated users may use a search engine and locate information that is "good enough" to fulfill their information needs, even if it is not the ideal or most complete information for those purposes.¹² Thus, for better or worse, the physical library is no longer the primary focus for many information seekers.

Part of this movement by users toward self-sufficiency in information seeking is due to the success of the Web search engine, and to the success of Google in particular. Recent reports from the Pew Internet and American Life Project shed a great deal of light on users' use of these tools. Rainie and Horrigan found that "on a typical day at the end of 2004, some 70 million American adults logged onto the Internet."¹³ Fallows found that "on any given day, 56% of those online use search engines."¹⁴ Fallows, Rainie, and Mudd found that of their respondents, "47% say that Google is their top choice of search engine."¹⁵ From these figures, it can be roughly estimated that more than 39 million people use search engines, and more than 18 million use Google on any given day—and that is only within the United States.

This trend seems quite dark for libraries, but it actually has its bright side. It is important to make a distinction here between use of a search engine and use of a reference service or other library service. There is some evidence that users' questions to library reference services are becoming more complex.¹⁶ Why this is occurring is less clear, but it may be hypothesized that users are locating information that is good enough to answer their own simple questions using search engines or other Internet-based tools. The definition of "good enough" may differ considerably between a user and a librarian. Nevertheless, one function of the library is education, and as with all education, the ultimate goal is to make

the student self-sufficient in self-teaching. In the context of the library, this means that one goal is to make the user self-sufficient in finding, evaluating, and using information resources. If users are answering their own simple questions, and asking the more difficult questions, then it may be hypothesized that the widespread use of search engines has had a role in raising the level of debate, so to speak, in libraries. Rather than providing instruction to users on simply using search engines, librarians may now assume that some percentage of library users possess this skill, and may focus on teaching higher-level information-literacy skills to users (www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.htm).

Simple questions that users may answer for themselves using a search engine, and complex questions requiring a librarian's assistance to answer are not opposites, of course, but rather two ends of a spectrum of the complexity of questions. While the advance of online search tools may enable users to seek and find information for themselves at one end of this spectrum, it seems unlikely that such tools will enable users to do the same across the entire spectrum any time soon; perhaps ever. The author believes that there will continue to be a role for librarians in assisting users to find, evaluate, and use information.

It is also important to make another distinction here, between the discovery of resources, and access to those resources. Libraries have always provided mechanisms for users to both discover and access resources. Neither the card catalog nor the online catalog contains the full text of the materials cataloged; rather, these tools are means to enable the user to discover the existence of resources. The user may then access these resources by visiting the library. Search engines, similar to the card and online catalogs, are tools primarily for discovery of resources: search-engine databases may contain cached copies of Web pages, but the original (and most up-to-date) version of the Web page resides elsewhere on the Web. Thus, a search engine enables the user to discover the existence of Web pages, but the user must then access those Web pages elsewhere. The author believes that there will continue to be a role for libraries in providing access to resources—regardless of where the user has discovered those resources.

In order to ensure that libraries and librarians remain a critical part of the user's information-seeking process, however, libraries must reappropriate technologies for online information seeking. Search engines may exist separate from libraries, and users may use them without making use of any library service. However, libraries are already the venue through which users access much online content—newspapers, journals, and other periodicals; reference sources; genealogical materials—even if many users do not physically come to the library or consult a librarian when using them. It is possible for

libraries to add value to search technologies by providing a layer of service available to those using it.

Google Scholar

One such technology for online information seeking to which libraries are already adding value, and that could add value to libraries in turn, is Google Scholar (scholar.google.com). Google Scholar is a specialty search tool, obviously provided by Google, which enables the user to search for scholarly literature online. This literature may be on the free Web (as open-access publications become more common and as scholars increasingly post preprint or post-print copies of their work on their personal Web sites), or it may be in subscription databases.¹⁷ Users may access literature in subscription databases in one of two ways: (1) if the user is affiliated with an institution that subscribes to the database, the user may access it via whatever authentication method is in place at the institution (e.g., IP authentication, a proxy server), or (2) if the user is not affiliated with such an institution, the user may pay for access to individual resources on a pay-per-view basis. There is not sufficient space here to explore the details of Google Scholar's operation, and anyway that is not the point of this paper; for excellent discussions of the operation of Google Scholar, see Gardner and Eng, and Jacsó.¹⁸

Pace draws a distinction between federated searching and metasearching: federated search tools compile and index all resources proactively, prior to any user's actual search, in a just-in-case approach to users' searching.¹⁹ Metasearch tools, on the other hand, search all resources on the fly at the time of a user's search, in a just-in-time approach to users' searching. Google Scholar is a federated search tool—as, indeed, are all of Google's current services—in that the database that the user searches is compiled prior to the user's actual search.

In this, Google Scholar is a direct descendent of Kilgour's work to develop shared online library catalogs. A shared library catalog is a union catalog: it is a database of libraries' physical holdings, compiled prior to any actual user's search. Google Scholar is also a union catalog, though a catalog of publishers' electronic offerings provided by libraries, rather than of libraries' physical holdings. It should be noted, however, that while this difference is an important one for libraries and publishers, it might not be understood or even relevant for many users.

Many of the resources indexed in Google Scholar are also available in full text. This fact allows Google Scholar to also move in the direction of Kilgour's goal of making passage retrieval possible for scholarly work. By using Google's core technology—the search engine and the inverted index that is created when pages are indexed by

a search engine—Google Scholar enables full-text searching of scholarly work. As mentioned above, when users search Google Scholar, they retrieve a set of links to the scholarly literature retrieved by the search.

Google Scholar also makes use of Google's link-analysis algorithms to analyze the network of citations between publications—instead of the network of hyperlinks between Web pages, as Google's search engine more typically analyzes. A *Cited By* link is included with each retrieved link in Google Scholar, stating how many other publications cite the publication listed. Clicking on this *Cited By* link performs a preformulated search for those publications. This citation-analysis functionality resembles the functionality of one of the most common and widely used scholarly databases in the scholarly community: the ISI Web of Science (WoS) database (scientific.thomson.com/products/wos). WoS enables users to track citations between publications. This functionality has wide use in scholarly research, but until Google Scholar, it has been largely unknown outside of the scholarly community. With the advent of Google Scholar, however, this functionality may be employed by any user for any research.

Further, there is a plugin for the Firefox browser (www.mozilla.com/firefox) that displays an icon for every record on the page of retrieved results that links to the appropriate record in the library's OPAC (Google Scholar does not, however, currently provide this functionality natively²⁰). This provides a link from Google Scholar to the materials that the library holds in its collection. When the item is a book, for example, this link to the OPAC enables users to find the call number of the book in their local library. When the item is a journal, it enables them to find both the call number and any database subscriptions that index that journal title. Periodicals are often indexed in multiple databases, so libraries with multiple-database subscriptions often have multiple means of accessing electronic versions of journal titles. A library user may access a periodical via any or all of these individual subscriptions without using Google Scholar—but to do so, the user must know which database to use, which means knowing either the topical scope of a database or knowing which specific journals are indexed in a database. As a more centralized means of accessing this material, many users may prefer a link in Google Scholar to the library's OPAC.

Google Scholar thus fulfills, in large part, Kilgour's vision of shared electronic cataloging. In turn, shared cataloging goes a long way toward achieving Kilgour's vision of 100 percent availability of information by allowing a user to discover the existence of information resources. However, discovery of resources is only half of the equation: the other half is access to those resources. And it is here where libraries may position themselves as a critical part of the information-seeking process. Search engines

may enable users to discover information resources on their own, without making use of a library's services, but it is the library that provides the "last mile" of service, enabling users to gain access to many of those resources.

Conclusion

Google Scholar is the topic of a great deal of debate, both in the library arena and elsewhere.²¹ Unlike union catalogs and many other online resources used in libraries, it is unknown what materials are included in Google Scholar, since as of this writing Google has not released information about which publishers, titles, and dates are indexed.²² Google is known to engage in self-censorship—or self-filtering, depending on what coverage one reads—and so potentially conflicts with the American Library Association's Freedom to Read Statement (www.ala.org/ala/oif/statementspols/ftirstatement/freedomreadstatement.htm).²³ Google is a commercial entity and, as such, a primary motivation of Google must be profit, and only secondarily, meeting the information needs of library users. For all of these and other reasons, there is considerable debate among librarians about whether it is appropriate for libraries to provide access to Google Scholar.

Despite this debate, however, users are using Google Scholar. Google Scholar is simply the latest tool to enable users to seek information for themselves; it isn't the first and it won't be the last. Google Scholar holds a great deal of promise for libraries due to the combination of Google's popularity and ease of use, and the resources held by or subscribed to by libraries to which Google Scholar points. As Kesselman and Watstein suggest, "libraries and librarians need to have a voice" in how tools such as Google Scholar are used, given that "we are the ones most passionate about meeting the information needs of our users." Given that library users are using Google Scholar, it is to libraries' benefit to see that it is used well.

Google Scholar is the latest tool in a long history of information-seeking technologies that increasingly realize Kilgour's goal of achieving 100 percent availability of information. 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 100 percent of the time.

Google Scholar may be on the vanguard of a new way of integrating library services into users' everyday information-seeking habits. As Taylor tells us, people have their own individual sources to which they go to find information, and libraries—for many people—are not at the top of their lists.²⁵ Google, however, is at the top

of the list for a great many people.²⁶ Properly harnessed by libraries, therefore, Google Scholar has the potential to bring users to library resources when they are seeking information.

Google Scholar may not bring users physically to the library. Instead, what Google Scholar can do is bring users into contact with resources provided by the library. This is an important distinction, because it reinforces a change that libraries have been undergoing since the advent of the online database: that of providing access to materials that the library may not own. Ownership of materials potentially allows for a greater measure of control over the materials and their use. Ownership in the context of libraries has traditionally meant ownership of physical materials, and physical materials by nature restrict use, since the user must be physically collocated with the materials, and use of materials by one user precludes use of those materials by other users for the duration of the use. Providing access to materials, on the other hand, means that the library may have less control over materials and their use, but this potentially allows for wider use of these materials.

By enabling users to come into contact with library resources in the course of their ordinary Web searches, Google Scholar has the potential to ensure that libraries remain a critical part of the user's information-seeking process. It benefits Google when a library participates with Google Scholar, but it also benefits the library and the library's users: the library is able to provide users with a familiar and easy-to-use path to materials. This is (for lack of a better term) a "spoonful of sugar" approach to seeking and finding information resources: by using an interface that is familiar to users, libraries may provide quality information sources in response to users' information seeking.

Green wrote that "a librarian should be as unwilling to allow an inquirer to leave the library with his question unanswered as a shop-keeper is to have a customer go out of his store without making a purchase."²⁷ A modern version of this might be that a librarian should be as unwilling to allow an inquirer to abandon a search with his question unanswered. Google Scholar and online tools like it have the potential to draw users away from libraries; however, these tools also have the potential to usher in a new era of service for libraries: an expansion of the reach of libraries to new users and user communities; a closer integration with users' searches for information; and the provision of quality resources to all users, in response to all information needs. Google Scholar and online tools like it have the potential to enable libraries to realize Kilgour's goals of improving the availability of information, and to provide 100 percent availability of information. These are goals on which all libraries can agree.

Acknowledgements

Many thanks to Lisa Norberg, instruction librarian, and Timothy Shearer, systems librarian, both at the University of North Carolina at Chapel Hill, for many extensive conversations about Google Scholar, which approached coauthorship of this paper. This paper is dedicated to the memory of Kenneth D. Shearer.

References and notes

1. Frederick G. Kilgour, "Historical Note: A Personalized Prehistory of OCLC," *Journal of the American Society for Information Science* 38, no. 5 (1987): 381.
2. Frederick G. Kilgour, "Future of Library Computerization," in *Current Trends in Library Automation: Papers Presented at a Workshop Sponsored by the Urban Libraries Council in Cooperation with the Cleveland Public Library*, Alex Ladenson, ed. (Chicago: Urban Libraries Council, 1981), 99–106; Frederick G. Kilgour, "Toward 100 Percent Availability," *Library Journal* 114, no. 19 (1989): 50–53.
3. Kilgour, "Toward 100 Percent Availability."
4. Frederick G. Kilgour, "Lack of Indexes in Works on Information Science," *Journal of the American Society for Information Science* 44, no. 6 (1993): 364; Frederick G. Kilgour, "Implications for the Future of Reference/Information Service," in *Collected Papers of Frederick G. Kilgour: OCLC Years*, Lois L. Yoakam, ed. (Dublin, Ohio: OCLC Online Computer Library Center, Inc., 1984): 9–15.
5. Frederick G. Kilgour, "A New Punched Card for Circulation Records," *Library Journal* 64, no. 4 (1939): 131–33; Kilgour, "Historical Note"; Frederick G. Kilgour and Nancy L. Feder, "Quotations Referenced in Scholarly Monographs," *Journal of the American Society for Information Science* 43, no. 3 (1992): 266–70; Gerald Salton, J. Allan, and Chris Buckley, "Approaches to Passage Retrieval in Full-Text Information Systems," in *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York: ACM Pr., 1993), 49–58.
6. Kilgour, "Implications for the Future of Reference/Information Service," 95.
7. Kilgour, "Toward 100 Percent Availability," 50.
8. Frieda Weise, "Being There: The Library As Place," *Journal of the Medical Library Association* 92, no. 1 (2004): 10, www.pubmedcentral.nih.gov/articlerender.fcgi?artid=314099 (accessed Apr. 9, 2006).
9. It is difficult to determine precise figures, as there is considerable overlap in coverage; several vendors provide access to some of the same periodicals.
10. North Carolina's database subscriptions are via the NC LIVE service, www.nclive.org (accessed Apr. 9, 2006).
11. Anne G. Lipow, "Serving the Remote User: Reference Service in the Digital Environment," paper presented at the *Ninth Australasian Information Online and On Disc Conference and Exhibition, Sydney, Australia, 19–21 Jan. 1999*, www.csu.edu.au/special/online99/proceedings99/200.htm (accessed Apr. 9, 2006).
12. J. Janes, "Academic Reference: Playing to Our Strengths," *portal: Libraries and the Academy* 4, no. 4 (2004): 533–36, http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v004/4.4janes.html (accessed Apr. 9, 2006).
13. Lee Rainie and John Horrigan, *A Decade of Adoption: How the Internet Has Woven Itself into American Life* (Washington, D.C.: Pew Internet & American Life Project, 2005), 58, www.pewinternet.org/PPF/r/148/report_display.asp (accessed Apr. 9, 2006).
14. Deborah Fallows, *Search Engine Users* (Washington, D.C.: Pew Internet & American Life Project, 2005), i, www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf (accessed Apr. 9, 2006).
15. Deborah Fallows, Lee Rainie, and Graham Mudd, *Data Memo on Search Engines* (Washington, D.C.: Pew Internet & American Life Project, 2004), 3, www.pewinternet.org/PPF/r/132/report_display.asp (accessed Apr. 9, 2006).
16. Laura Bushallow-Wilber, Gemma DeVinney, and Fritz Whitcomb, "Electronic Mail Reference Service: A Study," *RQ* 35, no. 3 (1996): 359–69; Carol Tenopir and Lisa A. Ennis, "Reference Services in the New Millennium," *Online* 25, no. 4 (2001): 40–45.
17. Alma Swan and Sheridan Brown, *Open Access Self-Archiving: An Author Study* (Truro, England: Key Perspectives, 2005), www.jisc.ac.uk/uploaded_documents/Open%20Access%20Self%20Archiving-an%20author%20study.pdf (accessed Apr. 9, 2006).
18. Susan Gardner and Susanna Eng, "Gaga over Google? Scholar in the Social Sciences," *Library Hi Tech News* 8 (2005): 42–45; Péter Jacsó, "Google Scholar: The Pros and the Cons," *Online Information Review* 29, no. 2 (2005): 208–14.
19. Andrew Pace, "Introduction to Metasearch . . . and the NISO Metasearch Initiative," *Presentation to the OpenURL and Metasearch Workshop, Sept. 19–21, 2005*, www.niso.org/news/events_workshops/OpenURL-05-ppts/2-1-pace.ppt (accessed Apr. 9, 2006).
20. This plugin was developed by Peter Binkley, Digital Initiatives Technology Librarian at the University of Alberta. See www.ualberta.ca/~pbinkley/gso (accessed Apr. 9, 2006).
21. See, for example, Gardner and Eng, "Gaga over Google?"; Jacsó, "Google Scholar"; M. Kesselman and S. B. Watstein, "Google Scholar and Libraries: Point/Counterpoint," *Reference Services Review* 33, no. 4 (2005): 380–87.
22. Jacsó, "Google Scholar."
23. Anonymous, Google Censors Itself for China, BBC News, Jan. 25, 2006, <http://news.bbc.co.uk/2/hi/technology/4645596.stm> (accessed Apr. 9, 2006); A. McLaughlin, "Google in China," Google Blog, Jan. 27, 2006, <http://googleblog.blogspot.com/2006/01/google-in-china.html> (accessed Apr. 9, 2006).
24. Kesselman and S. B. Watstein, "Google Scholar and Libraries," 386.
25. Robert S. Taylor, "Question-Negotiation and Information Seeking in Libraries," *College & Research Libraries* 29, no. 3 (1968): 178–94.
26. Fallows, Rainie, and Mudd, *Data Memo on Search Engines*.
27. Samuel S. Green, "Personal Relations between Librarians and Readers," *American Library Journal* 1, no. 2–3 (1876): 79. 1–11.