Preprint – cite as:
Pomerantz, J. (2006). Google Scholar and 100% Availability of Information. *Information Technology and Libraries*, 25(1), 52-56.

Google Scholar and 100% Availability of Information

Jeffrey Pomerantz
School of Information and Library Science
University of North Carolina at Chapel Hill
CB 3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
pomerantz@unc.edu

Abstract

This paper discusses Google Scholar as an extension of Fred 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.

Introduction

Fred Kilgour would probably approve of Google Scholar.

Kilgour wrote that the paramount goal of his professional career is "improving the availability of information" (1987, p. 381). Kilgour wrote about his goal of achieving this increase through shared electronic cataloging (1981), and even argued that shared electronic cataloging will move libraries towards the goal of 100% availability of information (1989).

Throughout much of Kilgour's life, 100% 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 a user as the holdings of their "home" library.

Kilgour went further than this, however, and also proposed that the full text of books could be made available to users electronically (1989). This would move libraries towards the goal of 100% 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% availability of information would only be reached by leveraging increasingly powerful technologies. The existence of effective search tools (1993), and the usability of those tools (1967b) would be crucial so that the user would be able to locate available information on his own. To achieve this goal, therefore, Kilgour himself 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 (1939), he was behind the development of the first online union catalog (1987), he called for passage retrieval for information seeking (Kilgour and Feder, 1992) at a time when such systems were first being developed (Salton, Allan, & Buckley, 1993). This development and application of technology was all directed towards the goal of improving the availability of information. Kilgour stated that the goal of these proposed information retrieval and other systems is "to supply the user with the information he requires, and only that information" (1967, p. 95).

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" (1989, p. 50), the user no longer has to physically go to the library. 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% availability of information in two ways: by insuring 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 (2004) 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." 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 700 databases and provide access to over unique 32,000 periodical titles; many of these subscriptions provide access to the full text of materials[1]. Additionally, the State Library of North Carolina provides a set of over one hundred database subscriptions to all academic and public libraries around the state; any North Carolina resident with a library card may access these databases[2]. 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 (1999) points out, it is the library that is remote from the user, yet the user is able to seek and find information.

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 a technically unsophisticated user may use a search engine and locate information that is "good enough" to fulfill his information need, even if it is not the ideal or most complete information for that purpose (Janes, 2004). Thus, for better or worse, the physical library is no longer the primary focus for many information seekers.

Part of this movement by users towards 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 & American Life Project shed a great deal of light on users' use of these tools. Rainie & Horrigan (2005) found that "on a typical day at the end of 2004, some 70 million American adults logged onto the internet" (p. 58). Fallows (2005)

[1] It is difficult to determine precise figures, since there is considerable overlap in coverage: several vendors provide access to some of the same periodicals.
[2] North Carolina's database subscriptions are via the NC LIVE service: www.nclive.org.

found that “on any given day, 56% of those online use search engines” (p. i). Fallows, Rainie, & Mudd (2004) found that of their respondents, “47% say that Google is their top choice of search engine” (p. 3). From these figures, it can be roughly estimated that over 39 million users use search engines, and over 18 million use Google on any given day… and that is only within the United States.

This trend seems quite dark for libraries, but actually it 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 (Bushallow-Wilber, DeVinney, and Whitcomb, 1996; Tenopir and Ennis, 2001). 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. As mentioned above, a user’s and a librarian’s definitions of “good enough” may differ considerably. Nevertheless, one function of the library is education, and as with all education, the ultimate goal is to make the student self-sufficient in teaching herself. 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 a user may answer for herself using a search engine, and complex questions that require 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, and 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 important to make another distinction here as well: 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 catalogued; 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 catalog, are tools primarily for discovery of resources: search engine databases may contain cached copies of webpages, but the original (and most up to date) version of the webpage resides elsewhere on the web. Thus, a search engine enables the user to discover the existence of webpages, but the user must then access those webpages 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. Libraries are, however, already the venue through which users access much online content – newspapers, journals, and other periodicals; reference sources; genealogical materials; etc. – 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, however, by providing a layer of service available to the user 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 pre- or post-print copies of their work on their personal websites (Swan & Brown, 2005)), or it may be in subscription databases. 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 (IP authentication, a proxy server, etc.), 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 the 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 (2005) and Jacsó (2005).

Pace (2005) 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. 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 Fred 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 too is a union catalog, though a catalog of publishers' electronic offerings that are 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 may 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 a user searches Google Scholar, she retrieves 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 webpages, as Google's search engine more typically analyzes. A "Cited by" link is included with each retrieved link in Google Scholar; this link states how many other publications cite the publication listed, and clicking on this "Cited by" link performs a pre-formulated 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 database (scientific.thomson.com/products/wos/). Web of Science enables users to track citations between publications. This functionality has wide use in scholarly research, but until Google Scholar 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[3] (Google Scholar does not, however, currently provide this functionality natively.) 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 the user to find the call number of the book in their local library. When the item is a journal, for example, this link to the OPAC enables the user 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 of course 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, a link in Google Scholar to the library's OPAC may be preferred by many users.

Google Scholar thus fulfills, in large part, Kilgour's vision of shared electronic cataloging. Shared cataloging in turn goes a long way toward achieving Kilgour's vision of 100% availability of information, by allowing a user to discover the existence of information resources. As discussed above, however, discovery of resources is only half of the equation: the other half is access to those resources. And it is here where libraries

[3] This plugin was developed by Peter Binkley, Digital Initiatives Technology Librarian at the University of Alberta. See: www.ualberta.ca/~pbinkley/gso/

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 (see, for example, Gardner and Eng, 2005; Jacsó, 2005; Kesselman and Watstein, 2005; and many others). 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 (Jacsó, 2005). Google is known to engage in self-censorship – or self-filtering, depending on what coverage one reads (Anonymous, 2006; McLaughlin, 2006) – and so potentially conflicts with the American Library Association's Freedom to Read Statement (www.ala.org/ala/oif/statementspols/ftrstatement/freedomreadstatement.htm). 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 (2005) 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" (p. 386). Given that Google Scholar is being used by library users, 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 Fred Kilgour's goal of achieving 100% availability of information. Google Scholar does not provide access to 100% 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% 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 (1968) tells us, people have their own individual sources to which they go to find information, and for many people libraries are not at the top of their lists. Google, however, is at the top of the list for a great many people (Fallows, Rainie, & Mudd, 2004). 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. When a library participates with Google Scholar it benefits Google, 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 for it) 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 (1876) 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" (p. 79). 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% 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 co-authorship of this paper. This paper is dedicated to the memory of Kenneth D. Shearer.

References

Anonymous. (2006, 25 January) Google censors itself for China. BBC News. http://news.bbc.co.uk/2/hi/technology/4645596.stm.

Bushallow-Wilber, Laura, Gemma DeVinney, and Fritz Whitcomb. “Electronic Mail Reference Service: A Study.” RQ 35, no. 3 (1996): 359-69.

Fallows, Deborah. “Search Engine Users.” Washington, DC: Pew Internet & American Life Project, 2005. http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf.

Fallows, Deborah, Lee Rainie, and Graham Mudd. “Data Memo on Search Engines.” Washington, DC: Pew Internet & American Life Project, 2004. http://www.pewinternet.org/PPF/r/132/report_display.asp.

Gardner, Susan, and Susanna Eng. “Gaga over Google? Scholar in the Social Sciences.” Library Hi Tech News 8, (2005): 42-45.

Green, Samuel S. “Personal Relations between Librarians and Readers.” *American* Library Journal I, no. 2-3 (1876): 74-81.

Grogg, Jill E., and Christine L. Ferguson. “OpenURL Linking with Google Scholar.” Searcher 13, no. 9 (2005): 39-46.

Jacsó, Péter. “Google Scholar: the pros and the cons.” Online Information Review 29, no. 2 (2005): 208-214.

Janes, J. (2004). Academic Reference: Playing to Our Strengths. portal: Libraries and the Academy, 4(4), 533-536. http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v004/4.4janes.html.

Kesselman, M., & Watstein, S. B. (2005). Google Scholar™ and libraries: point/counterpoint. Reference Services Review, 33(4), 380-387.

Kilgour, Frederick G. “A New Punched Card for Circulation Records.” Library Journal 64, no. 4 (1939): 131-133.

Kilgour, Frederick G. “Information Retrieval.” In Mcgraw-Hill Yearbook of Science and Technology. New York: Mcgraw-Hill Book Company, 1967a.

Kilgour, Frederick G. “Implications for the Future of Reference/Information Service.” In The Present Status and Future Prospects of Reference/Information Service, edited by Winifred B. Linderman. Chicago: American Library Association, 1967b.

Kilgour, Frederick G. “Concept of an On-Line Computerized Library Catalog.” Journal of Library Automation 3, no. 1 (1970): 1-11.

Kilgour, Frederick G. “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, edited by Alex Ladenson, 99-106. Chicago: Urban Libraries Council, 1981.

Kilgour, Frederick G. "Historical Note: A Personalized Prehistory of OCLC." Journal of the American Society For Information Science 38, no. 5 (1987): 381-384.

Kilgour, Frederick G. "Toward 100 Percent Availability." Library Journal 114 , no. 19 (1989): 50-53.

Kilgour, Frederick G. "Lack of Indexes in Works on Information Science." Journal of the American Society For Information Science 44 no. 6 (1993): 364.

Kilgour, Frederick G. and Feder, Nancy L. "Quotations Referenced in Scholarly Monographs." Journal of the American Society For Information Science 43, no. 3 (1992): 266-270.

Lipow, Anne G. "Serving the Remote User: Reference Service in the Digital Environment." Paper presented at the Ninth Australasian Information Online & On Disc Conference and Exhibition, Sydney, Australia, 19-21 January 1999. http://www.csu.edu.au/special/online99/proceedings99/200.htm.

McLaughlin, A. (2006, 27 January). Google in China. Google Blog. http://googleblog.blogspot.com/2006/01/google-in-china.html.

Pace, Andrew. "Introduction to Metasearch …and the NISO Metasearch Initiative." Presentation to the OpenURL & Metasearch Workshop, September 19-21, 2005. http://www.niso.org/news/events_workshops/OpenURL-05-ppts/2-1-pace.ppt

Rainie, Lee, and John Horrigan. "A Decade of Adoption: How the Internet Has Woven Itself into American Life." Washington, DC: Pew Internet & American Life Project, 2005. http://www.pewinternet.org/PPF/r/148/report_display.asp.

Salton, Gerald, 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 Press, 1993.

Swan, Alma, and Sheridan Brown. "Open Access Self-Archiving: An Author Study." Truro, UK: Key Perspectives Limited, 2005. http://www.jisc.ac.uk/uploaded_documents/Open%20Access%20Self%20Archiving-an%20author%20study.pdf.

Taylor, Robert S. "Question-Negotiation and Information Seeking in Libraries." College & Research Libraries 29, no. 3 (1968): 178-94.

Tenopir, Carol, and Lisa A. Ennis. "Reference Services in the New Millennium." Online 25, no. 4 (2001): 40-45.

Weise, Frieda. “Being There: The Library as Place.” Journal of the Medical Library Association 92, no. 1 (2004): 6-13. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=314099.