



Where Did Google Get Its Value?

James Caufield

abstract: Google's extraordinary success is usually attributed to innovative technology and new business models. By contrast, this paper argues that Google's success is mostly due to its adoption of certain library values. First, Google has refused to adopt the standard practices of the search engine business, practices that compromised service to the user for the sake of immediate corporate profit. Instead, Google has implemented many policies and design principles that correlate directly to established library values. Second, Google has implemented systems that replicate (or substitute for) valuable library functions. With these steps Google has introduced some traditional library practices and values to the Internet environment, and there can be little doubt that they have contributed enormously to its success.

Introduction

The value of Google is everywhere evident. Before Yahoo ceased using its services in February 2004, Google provided approximately 80 percent of Web searches.¹ Even after that, Google still provides more than 200 million searches per day, close to half the searches conducted on the Web.² The company took in approximately one billion dollars in advertising revenue in 2003, with a profit margin of 30 percent.³ Its position was so encouraging that shortly before the August 2004 initial public offering (IPO), David Menlow, president of IPOfinancial.com, predicted "virtually any price" for the stock would be considered acceptable because "the company seems to be a printing press for money."⁴ The first round of public sales yielded an unprecedented return of 200 times the original venture capital investment,⁵ and the company was valued at almost \$30 billion.⁶ "'From original investment to I.P.O., I'd have to say Google offers the biggest payoff in venture history' said Jesse Reyes, Vice President of Venture Economics, a unit of Thomson Financial."⁷

But where did Google get its value? The story is told in several ways. Most often it is explained as a triumph of technology, attributed to a new business model, or as a combination of the two, as in the following explanation:

portal: Libraries and the Academy, Vol. 5, No. 4 (2005), pp. 555–572.

Copyright © 2005 by The Johns Hopkins University Press, Baltimore, MD 21218.

Google owes its massive success to two events. First, Messrs Brin and Page came up with what was for some time the best algorithm for searching web pages. Second, Eric Schmidt, whom they hired as a chief executive in 2001, figured out how to “monetize” Google’s popularity by selling small and unobtrusive advertisements on related topics, so-called “sponsored links,” alongside the search results.⁸

A second narrative places less emphasis on technology and business models, focusing instead on the company’s values and integrity. Google’s founders are among those who have made these claims prominent, posting a company philosophy that includes the claim that “you can make money without doing evil.”⁹ Others share this view: “In Silicon Valley the company is revered for putting its mission before immediate financial gain.”¹⁰

A third version of Google’s success is told by librarians, who usually focus on the opposition between Google and some library values. For although librarians respect and employ Google, they are also concerned about “Googlefication”¹¹ or “Googlization.”¹² In part, this refers to patrons’ preference for quicker and easier results than librarians and current library systems can provide:

Librarians, publishers, and aggregators alike often call Google their main competitor. Google, or similar Web search engines, is the information finding tool of first choice for many users—far ahead of proprietary online services or libraries and light-years ahead of print sources.¹³

There are fears that the public is coming to see Google not simply as a competitor to libraries but as a substitute for them. The popular thinking seems to be, “Why do we need a librarian [if] we’ve got Google?”¹⁴ Even at an academic library, the new rule seems to be “Google first, ask [reference] questions later.”¹⁵

Users obviously value the ease of access that Google and other search engines provide, but it comes at the price of quality. Recently the library community has begun a debate whether libraries should adopt systems that compromise on quality in order to provide the ease of access that users expect. There is also some discussion about how

Google and other private search services threaten library values other than quality. One prominent critic of Google, Rory Letwin, points out that Google’s digitization of the contents of libraries certainly does not imply that Google will work to advance such values as equity

Apart from the growing recognition of the value of Google’s simplified search process, most discourse in the library profession has emphasized the opposition between Google’s values and those of libraries.

of access, collective ownership, and privacy.¹⁶ Apart from the growing recognition of the value of Google’s simplified search process, most discourse in the library profession has emphasized the opposition between Google’s values and those of libraries.

A fourth account also considers libraries and their values, but it does so only to dismiss them. This story sees Google’s success as the decisive resolution to an academic



debate involving two competing visions of information. Joel Achenbach quotes Peter Lyman of the Berkeley School of Information Management and Systems: “‘There’s been a culture war between librarians and computer scientists,’ Lyman says. ‘And the war is over,’ he adds. ‘Google won.’”¹⁷

While this is only a sampling of the various narratives relating to Google’s success, it serves to illustrate some general argumentative tendencies. Most accounts stress the role of technology and business models, and they fail to consider libraries and their values. When libraries and library values are considered, the tendency has been to see Google’s success as antithetical to them. None of these narratives, however, sufficiently represents the positive correlations between Google and library values or examines their importance to the Google enterprise. The thesis of this paper is that Google has succeeded mostly because it has adopted many library values.

Scope and Method

The many ethical questions relating to Google and library values cannot all be addressed in a single article. This paper seeks to show that elements essential to Google’s success can be mapped directly to certain traditional library values. Because of this focus, most of the examples given here will indicate a positive correlation. This is, of course, an incomplete picture, but this should not be construed as a lack of critical perspective. I readily admit that there is much to be written about how Google fails to embody library values. Still, a consideration of how Google *does* embody some library values can serve as a preliminary to those more critical works.

Although this is not an experimental paper, a few clarifications are in order. Library values are usually understood to be the ethical principles expressed in professional codes or guidelines. But libraries are institutions that have been built with certain goals and values in mind, and their environment has been structured in such a way as to embody these values. This paper extends the traditional meaning of library values to include these institutional structures and the valuable functions that libraries and their systems perform.

That said, it is now possible to delineate two ways in which Google has brought library values to the Web environment. First, Google has adopted many of the precepts that guide librarians in their work. Second, Google has created systems that replicate (or at least are analogous to) some of the valuable functions that libraries provide.

In order to capture a sense of the developments that have led to Google’s pre-eminence, the following discussion will attempt to present the elements of this story in their approximate chronological order.

The Early Web Was a Broken Library

What needs to be understood is that the early Web and its search engines were ineffective or inefficient precisely because they were a travesty of traditional library values. An incomplete list of those traditional library values would include the following:

- The library should have a collection of quality materials. For instance, *The Freedom to Read* includes the provision that “it is the responsibility of publishers and



librarians to give full meaning to the freedom to read by providing books that enrich the quality and diversity of thought and expression."¹⁸

- This collection should be balanced, representing diverse views. For instance, the *Library Bill of Rights* states, "Libraries should provide materials and information presenting all points of view on current and historical issues."¹⁹ *Intellectual Freedom Principles for Academic Libraries* evinces a similar concern: "Preservation and replacement efforts should ensure that balance in library materials is maintained."²⁰
- The library should facilitate access to materials. For instance, *Libraries: An American Value* states, "We connect people and ideas by helping each person select from and effectively use the library's resources."²¹

Above all, the work of librarians is service, usually to the public, and private interest should not be allowed to interfere. Accordingly, two over-arching values are:

- The personal interests of librarians should not compromise service to the user. For instance, the *Code of Ethics of the American Library Association* notes, "We do not advance private interests at the expense of library users."²²
- The personal interests and political views of individuals in the community served should not be allowed to compromise library service to the community as a whole. This seems to be implicit, for instance, in the library value of opposing censorship. For instance, *The Freedom to Read* section 6 asserts, "It is the responsibility of publishers and librarians, as guardians of the people's freedom to read, to contest encroachments upon that freedom by individuals or groups seeking to impose their own standards and tastes upon the community at large."²³

The early Web environment (including its search engines) came up short in every way. Most basic was the failure to provide a collection of quality materials. In the traditional paper system, quality was ensured on many levels. Editors acted as an initial filter, limiting the total universe of published information. Selectivity was further heightened by the process of review and recommendation carried out by newspapers, magazines, and journals, and more generally by academic and professional institutions. Finally, given their limited budgets, each library imposed yet another layer of discrimination as it acquired only a subset of the available and recommended items. In short, the universe of materials made available by the traditional paper-based library was the product of many levels of quality control, where human judgment had, on multiple occasions, deemed each item worthy of inclusion. Statements of library values usually do not make this value explicit, probably because selection for quality has been so deeply embedded in traditional publishing and library practices.

The World Wide Web stood this arrangement on its head. Now anyone with a modicum of technological skill could publish and be included in the collection, bypassing the traditional filtering processes. Worse, because there was no obvious profit potential, traditional publishers—the group that had contributed so much to the creation of reliable resources—supplied little to the Web collection. Conversely, those who did contribute "free" content were often interested in selling their products or promoting their views. This created problems not only with quality of the collection but also with the balance of the views represented.



Eventually these failings were to some extent compensated by the growth of the Web, for even though only a small percentage of materials approached library standards for quality, the entire collection was becoming extremely large. As Seth Godin remarks, "We hit a critical mass of really valuable stuff that was online . . . about 2000."²⁴ Thus the Internet provided the collection, both in terms of quantity and quality, that was the sine qua non of any library-like information resource.²⁵ The major remaining problem was how to find these quality materials.

While the Web provides physical access to materials, search engines and directories provide intellectual access. To do this, any search engine must perform two functions that are essential to a library: it must index the materials in the collection, and it must provide a retrieval system for matching search queries to the index. For purposes of this paper, a rough outline of search engine technology will suffice.²⁶

The process of indexing begins with crawlers or spiders that visit Web pages, gather information, and then store it in an index. When a user submits a query, a search engine will attempt to match the search terms to items in the index. In the early days of search engines, this matching process was carried out on the basis of keywords, usually with results ranked according to the frequency of a keyword on a page. Yet keyword indexing is not an especially effective method of access as it tends to produce search results of low relevance. The work of finding useful materials then devolves upon the user, who must sort through the results. As the size of the database increases, this task becomes more onerous. The Web, of course, grew extremely large, and so this limitation became severe.

An even greater problem was that unscrupulous Web masters could easily manipulate the automated indexing. The early Web was a riot of private interests run wild, and nothing better illustrated this than the prominence of pornography in search engine results.

Simply having the relevant words included many times on your web site was enough [to be included in search results]. But that led to a miasma of pornography web sites that emptied dictionaries onto their web pages, in invisibly small fonts in white text against a white background, to attract the attention of search engine "crawlers." The pollution of results by the end of 1998 had the search companies tearing their hair out.²⁷

Matters were only made worse when private interest intruded at another level, as search engine companies accepted money for placing a given Web site in the search results. With this, the early search engines violated a library value so basic that it had remained unspoken. Statements of library values have never explicitly forbidden catalogers from accepting money for promoting items in the search results, probably because it was inconceivable given the service ethos of the library.

Thus the early Web contravened every value for which libraries and librarians stood. At first the resources were too few and of too low quality; later the Web grew into a very large collection of materials, but much of what it contained was created by profiteers and self-promoters and was, as a result, often of low quality or unbalanced. Also, this largest database of all was being searched by an indiscriminate method—a weak keyword indexing and matching system that produced many irrelevant results. These low-quality results were further corrupted by self-interested manipulation, both on the level

of rogue indexers (Web masters who manipulated the automatic indexing systems) and by the search engine companies (who allowed advertising revenues to bias the search results). Save for the fact that it offered unprecedented physical access, this early Web represented everything that librarians abhorred.

Google Brings a Library Value to the Web Environment: Improved Access through Better Indexing

Something had to be done—but what? The Web presented unparalleled opportunities for the creation and publication of materials, but as yet there was no corresponding system to filter and sort these materials. Yet information pertinent to these tasks was available on the Web itself, and search engines soon began to draw upon it.

Searching the Web involves more than looking for matching keywords in the text of Web pages. The Web is not a collection of unrelated documents. There are links between these documents that imply relationships. Search engines that exploit these relationships to provide more relevant results have emerged in recent years.²⁸

Google's PageRank algorithm was among the earliest systems to gather this information about the relationships between Web pages and use it to filter and sort results. Algorithms are the complex mathematical formulas used by search engines to process information. Entire areas of research are devoted to their design and improvement, but for purposes of the present discussion a highly simplified explanation of algorithms will suffice.²⁹ Librarians are perhaps too easily convinced that algorithms belong to the technological and mathematical sciences and are completely beyond the purview of libraries and their values. Also, search engine algorithms are closely guarded secrets, which might further stifle discussion in library circles. The PageRank algorithm, however, has been patented, and so it—unlike most algorithm elements—has been published, allowing us to know part of Google's system for calculating search results.

Google's PageRank system involves higher mathematics. Yet Google's explanation of where it gets its values (in terms of rankings) does not hide behind a technological narrative of eigenvector calculation and the like. Instead we are given a political description:

PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B.³⁰

Any claim that Google's system is democratic is of course open to criticism, but I will leave that for another occasion. The concern here is how this system improved Internet searching and how it is analogous to a traditional library value or practice.

When indexing was done by keywords, the result was less reliable because a keyword may not provide an appropriate denotation of the content of a page. By contrast, a link to another Web page is a humanly created indicator explicitly designed to point to content, and presumably the link contains text describing that content. Google gathers these links (as well as the surrounding text and other elements of the page) for its index. By indexing at the level of links, Google is mining a far more selective and semantically richer field than the simple keyword method of indexing.



This process—called link analysis—is also performed by other search engines. Google is unique in that it makes link analysis into a recursive process. Instead of basing the rank of a page on a simple count of the number of inbound links, Google assigns a relative weight to each of the recommending pages. To do this it must assign a weight to the pages that recommend the recommending pages, and so on—a potentially infinite process. By Google’s own description, the strength of this recursive process is that it better approximates the social relations between parts of the Web than does a simple back link count.

Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves “important” weigh more heavily and help to make other pages “important.”³¹

While we usually think of library values as abstract ethical ideals, this paper extends the concept of library values to include the valuable functions that libraries perform, that is, the tools and services that promote or embody these values. In this wider sense, the Google PageRank process at least partially replicates a very important library value. Just as librarians tend to believe they cannot and need not understand anything about algorithms, people who think about search engines do not seem to often consider library practices. And even though Google gives a social and political explanation of where its values (in terms of relevant results) come from, this explanation is seldom linked to traditional library values and practices. For example, an otherwise thorough review of search engine history attributes Google’s success to the “groundbreaking insight . . . that the Web is a giant popularity contest—and that the most-cited pages will probably be the most useful.”³² Only rarely is it recognized that this is not groundbreaking at all but rather the translation of a traditional library (or academic) value to the Web environment. “What has made Google special is that, in assessing the quality of sites, it takes note of how many other pages link to any given page. This is an old idea from academia, called citation analysis.”³³

That the current discourse on Google and other search engines often fails to make this observation would seem to be a kind of amnesia. Google’s founders were quite explicit about the role their PageRank process would fill. Even at the inception of the project, they acknowledged that their valuation process was borrowed from academia, noting that PageRank is in some ways analogous to “academic citation analysis,”³⁴ and that the analysis of “back links . . . provide[s] a kind of peer review.”³⁵

Bringing the library practice of citation analysis to the Web environment improves relevance immensely. Rather than having a Web page provide its own indexing (in the sense of presenting text that will be searched by keyword), PageRank relies on information provided by other pages and thus is more objective. Also, PageRank increases relevance by mining Web pages on the higher semantic level of links rather than keyword alone.

No more could rogue indexers boost their rankings by hiding a dictionary on their home page. And since porn sites and other crass self-promoters tend not to link to one another, they will tend to fall in the rankings, as Larry Page et al. noted in 1998.³⁶ Thus Google brought to the Web a functional (if abbreviated) analog of the process of judging, filtering, and recommending materials that has traditionally been carried out by libraries, publishers, and educational institutions. With the PageRank algorithm, Google

made important steps toward resolving two vexing problems: the low relevance provided by keyword indexing and searching and the further corruption of results by Web masters who manipulated the automated indexing systems. PageRank represents a remarkable step forward in bringing the library values of access (in this case intellectual access) and uncorrupted indexing to the Web environment.

Thus Google brought to the Web a functional (if abbreviated) analog of the process of judging, filtering, and recommending materials that has traditionally been carried out by libraries, publishers, and educational institutions.

**Google Brings a Library Value to the Internet:
Better Access through Simple and Disinterested User Interface**

It is frequently noted that Google's success depended upon a value that is never mentioned in documents such as the *ALA Code of Ethics* or the *Library Bill Of Rights*—money.³⁷ The basic outline of the

PageRank search method was created in early 1998, but the infusion of venture capital, particularly the \$25 million it received in June of 1999, gave Google the power to hire the best, to research, and to implement on a massive scale. By contrast, libraries are mostly public institutions with already defined missions and commitments to certain services. In terms of their ability to innovate, libraries are at a distinct disadvantage.

But while money gave Google the potential to innovate, what differentiated it from other search engines was its willingness to adopt at least some of the traditional values of libraries and other information services. For instance, Google's statement of company philosophy has 10 principles, the first of which is very similar to precepts found in statements of library values: "Focus on the user, and all else will follow."³⁸

The business model of existing search engines—the use of banner ads—had prevented such an exclusive focus. If the Internet is the information superhighway, then banner ads are its billboards. As with highway billboards, the advertising revenue of banner advertisements depends on the amount of traffic that views the site. It is not simply a question of how many pass by but also of how long each viewer remains to take in the advertising. Here arises a potential conflict of interest between the searcher, who wants results quickly, and the search site, which wants to increase viewing time. While it would probably be an exaggeration to say that most commercial search sites are interested in creating traffic jams, it is not inconceivable that they view longer download times as an opportunity for the extended display of advertising rather than as an impediment to access. After all, the banner ad appears first, followed by the page.

There is another potential conflict of interest. It is very much in the interest of search services to gather user information, but this can be done only so long as the user remains on the Web site. Yet it is in the user's interest to find the most relevant materials wherever they might be. Faced with this choice, the standard practice of search engine companies has been to make themselves "sticky," encouraging users to remain on their sites for purposes of advertising and data collection. As a result, most search services became increasingly complex portals.



By contrast, Google insisted on the simplicity of the user interface, a practice recognized in the library community as a means of facilitating access. Google did this even though the lack of content on their own site meant that it was recommending users to other sites. As Monika Henzinger, director of research at Google, explains, “Portals try to keep the users there as long as possible. We are trying to keep the users as short a time as possible.”³⁹ The same concern with access had ramifications for Google’s advertising model—“We don’t want to use simple banners. They slow down the search experience—users want to get on and off the search page quickly, and we let them.”⁴⁰ Thus Google’s site design and advertising decisions were made with a view to improving access rather than maximizing immediate profit.

Simplicity of interface improves access in several ways. First, physical access is improved because the time required to load a page is shortened. Second, an uncluttered interface simplifies the user’s experience and so facilitates intellectual access. Perhaps most important, Google did not allow the potential benefits of retaining users on its site to interfere with the user’s untrammelled access to relevant materials, wherever they might be. This interface design contributed much to the advancement of access in the Internet environment.

Google Brings the Library Value of (Relatively) Unbiased Selection to the Web Environment

In terms of traditional library values, selection refers to the process of deciding which materials should be physically present in the library and thus accessible to users. In the Web environment, however, all documents (at least all that are not sequestered behind passwords) are physically available. Given the enormous number of Internet resources, the challenge for the user is no longer physical access but intellectual access. Of these many documents, which are relevant? It is impossible to sort through all the results, and it is frequently noted that searchers rarely go beyond the first few pages of search results. So it could be said that the search engine ranking process performs a function analogous to the traditional library practice by “selecting” the resources that will be intellectually accessible.

In the traditional library environment it would be inconceivable for a librarian to accept money for including or promoting certain items. Yet what is unthinkable in the library environment has been standard operating procedure in the Web environment. Search engine companies accepted (and continue to accept) money, either for including a Web site in the search results (paid inclusion) or for guaranteeing a given ranking in the results (paid placement). Here was another conflict of interest, for while the searcher wanted relevant results, search engines produced results according to another criterion—advertising revenue. Thus, for instance, services provided by Overture and Inktomi have been sponsor-centered rather than user-centered, focusing on the private interest of the advertisers and the search engine companies, to the detriment of users.⁴¹

To their credit, Google’s founders have sometimes expressed a genuine concern for core library values. In one notable instance, “[Google co-founder] Mr. Page said that commercial exploitation was ‘bastardizing’ the search industry.”⁴² The practice of selling placement or inclusion in search results could, of course, be profitable. As early as

2000, Larry Page was saying, "If we signed a deal with a Web advertising network like DoubleClick, we would be profitable right now. But we are in this for the long-term, and we want to do the right thing for business. Being profitable immediately isn't the right thing."⁴³

Given this reluctance to adopt the standard business model of the search industry, it is understandable that "the rows between Google's two strong-willed founders and their venture capitalists are legendary."⁴⁴ Yet their arguments were not about how to implement a core library value but rather about how to make money. The question is strategic: whether a short-term or long-term strategy is "the right thing for business." Rather than accepting money for placing advertisements in the search results, Google eventually (2002) came to accept bids for advertising words, which are then used to determine the relevance of an advertisement to any given search.⁴⁵ This "targeted advertising" is then placed in separate a column and clearly marked as advertising. By separating advertising from search, Google is able to accept advertising revenue while maintaining uncorrupted search results.

This has been hailed as a great innovation; but as with other claims about Google's systems, it is not completely true. Google's solution to the conflict of interest inherent in advertising is certainly consistent with the traditional library value of providing an unbiased selection of resources. But this business model draws less from libraries than it does from the practices of another kind of information service—the journalistic tradition. While news organizations do accept paid advertising, they recognize the potential conflict of interest between advertising and the unbiased provision of information, and so they establish a "firewall," a strict division between the advertising and news departments. In the same way, Google has erected a barrier between advertising and search.⁴⁶

Other companies are at least acknowledging Google's principles, and, to some extent, they are practicing them. Microsoft, for instance, has at least temporarily abandoned the practice of paid inclusion.⁴⁷ After Yahoo dropped the Google search service, it began accepting payment for indexing of Web pages, and for this they have received much criticism. While this is not paid placement, it still means that Yahoo accepts money for providing an advantage in the indexing process.⁴⁸ So although Tim Cadogan, a vice president at Yahoo, maintains that "this is not a pay-for-performance product, where you can influence your site's ranking,"⁴⁹ Jim Lazone, AskJeeves product manager, says that AskJeeves has dropped this practice because they "couldn't figure out how to make fair comparisons" if they used the system.⁵⁰

Google Produces Better Access through Uncorrupted Indexing

Google refused to corrupt its search results by accepting payment for placement or inclusion. With this they introduced a standard library practice to the Internet environment. But what of the problem of rogue indexers, the Web masters who manipulate the automated indexing to boost their own pages' rankings? Google's PageRank algorithm terminated the crudest efforts, but the indexing system remains vulnerable to spurious input (also known as "spamming" search engine results). Because Google ranks sites (at least partly) on the basis of inbound links, creating more inbound links can boost a site's ranking.



One form of this, called Google-bombing, is usually a prank. For instance, at one time a Google search for the terms “miserable failure” yielded George W. Bush’s biography as the first result.⁵¹ A more pervasive form of this problem is search engine optimization, “a multimillion dollar industry [that] has grown up around spamming the Google search results.”⁵² As spammers attempt to bias the search results, Google tries to maintain its unbiased selection process. When confronted by deliberate efforts of this sort, Google now takes steps to punish these sites, reducing their ranking or barring the site completely.⁵³ Google also makes modifications to its ranking process, and it seems that these are also in part intended to defeat the most dubious aspects of search engine optimization.

It is highly debatable whether these policies produce “objective” results (as Google maintains) or a “tyranny,” as its critics claim.⁵⁴ Though this question is extremely important, it is beyond the scope of this paper. Here I only wish to point out that, regardless of the success of implementation, it seems that Google’s motivation is in keeping with basic library values. For instance, *Intellectual Freedom Principles for Academic Libraries* recommends that “there should be alertness to efforts by special interest groups to bias a collection through systematic theft or mutilation.”⁵⁵ With a few substitutions we can produce a principle that describes Google’s practice: “There should be alertness to efforts by special interest groups to bias *search results* through *systematic efforts to manipulate the automated indexing system.*”

Google (and Other Search Engines) Seek to Create an Internet Analog to Another Valuable Library Function—The Reference Interview

According to Google’s technology director, Craig Silverstein, “the ultimate goal is to have a computer that has the kind of semantic knowledge that a reference librarian has.”⁵⁶ While this goal seems distant, another valuable aspect of the reference librarian’s work is in fact being replicated (or at least partially substituted for) in the online environment, and it has the potential to improve search results in the near future.

I have already noted that some library values are not so much expressed in ethical precepts as they are embodied in library practices. One such practice is the reference interview. Frequently the patron asks a question that the librarian could answer simply and cleanly, but a reference interview reveals that the patron is really seeking something quite different. In this regard, the patron’s first question resembles the bare keyword query sent to the search engine. If no further inquiry is made and no context for the question is established, then both the reference librarian and the search engine are engaged in a simple process of matching terms to resources, without any attempt to verify that the items are appropriate to the user’s need. The reference interview, a traditional library procedure, is an ethical action designed to identify the user’s information need and so provide better intellectual access.

Google and other search engines also gather feedback from the searcher but in a different fashion. Exactly what information search engines gather and how they use it cannot be dealt with here. What is clear is that Google deposits on each computer a cookie that reports browser search history to Google.⁵⁷ This can include a record of search terms used and sites visited.

Generally, Google's desire to gather personal information has been thought to be motivated by an interest in targeting advertising. While this is certainly true, user profiling can also render searches more relevant by providing a context for an otherwise isolated query. At least one search engine, Mooter, gathers information during a session and uses it "to adjust the rankings based on the user's behavior,"⁵⁸ and it is reasonable to think that other search engines are developing similar (and more extensive) techniques.

One of the next big leaps for search engines [will be] finding meaning in the way a single person searches the Web. . . . Search engines will study the user's queries and Web habits and, over time, personalize all future searches. Right now Google and the other search engines don't really know their users."⁵⁹

All the major search engines have said recently that they see personalized search results as a key way to advance relevancy.⁶⁰

How Google Threatens to Undermine One Library Value—Privacy

For the most part, this paper has been a discussion of how Google has adopted certain library values. Because of this focus, almost all of the examples used in this paper show congruence between library values and Google's practices. This should not be taken to imply that there are no critical issues; Google's practices do in some ways stand in opposition to traditional library values. This paper cannot engage in an extended discussion of these critical issues, but a consideration of one—privacy—might provide some balance.

While there is no technical barrier to the development of personalized search, there are serious ethical questions, particularly concerning the traditional library value of protecting patron privacy. For example, *Intellectual Freedom Principles for Academic Libraries* states, "The privacy of library users is and must be inviolable. Policies should be in place that maintain confidentiality of library borrowing records and of other information relating to personal use of library information and services."⁶¹

As noted above, Google and other search engines are faced with a dilemma. Even if a search engine company were wholly devoted to the benefit of its users, there would still be motive to compromise privacy. It is possible to advance one library value (ac-

So there are tremendous incentives—both user-centered and profit-centered—for a search engine company not to uphold the crucial library value of privacy.

cess, in the form of relevant material) at the expense of another library value (privacy). Library users make a similar—though far better safeguarded—concession when they confide their information needs to a reference librarian. In any case, both access and privacy benefit

the public, the users of search services. But if a search engine company is devoted primarily to private profit, it is also true that marketers and others would pay enormous sums for access to the detailed user profiles that companies such as Google could com-



pile. So there are tremendous incentives—both user-centered and profit-centered—for a search engine company not to uphold the crucial library value of privacy.

Google's Gmail privacy policy states, "We will never rent, sell or share information that personally identifies you for marketing purposes without your express permission."⁶² Yet despite this robust promise, Matthew Goldberg points out that it has little weight.

The Terms of Use agreements governing the use of Gmail, which by its own terms is the controlling document in the event of any inconsistency with the Privacy Policy, eviscerates any protection purportedly offered by the promise not to disclose personal information. The Terms of Use agreement states that "Google may, in its sole discretion, modify or revise these terms and conditions and policies at any time." Similarly, the Terms of Service agreement for Google's search engine states that Google "reserves the right to modify these Terms of Service from time to time without notice."⁶³

In April 2004 Google said that it did not plan to correlate its Gmail user profiles with searcher profiles, but Goldberg points out that after the California *Online Privacy Protection Act (OPPA)* went into effect on July 1, 2004, Google "clarified" its position.

That Google clarified its stance with respect to correlating search and e-mail immediately following passage of the *OPPA* suggests some duplicity on the company's part. The Google representatives had stated in the past that the company has no intention of correlating personal data among services. But if this is so, "why does it need to explicitly reserve the right to do so?"⁶⁴

One answer is that the correlation can provide more effective search, which would be in the interest of users. Another answer is that Google intends to build user profiles it can sell. In either case, Google's legal strategy seems to be at odds with their public statements, which calls into question the high-minded assertions Google has made.

Some Speculations about the Future of Library Values in the For-Profit Search Engine Environment

In some professions, such as medicine and law, institutions are held to high standards. This has been true of libraries. Yet in their earliest incarnations, search engines embodied almost none of the traditional library values and practices. It is gratifying that the search engine company that eschewed its immediate self-interest is the company that has succeeded most. It is a validation of at least some library values in that they have been essential even in a for-profit setting. If we imagine a spectrum of ethical actions ranging from those that are most directly self-interested to those that serve only the public interest, Google represents a step away from the most immediate self-interest of earlier search engine practices. It is possible that private search engine companies will find it in their interest to continue to take such ethical steps. For instance, outrage over privacy violations could conceivably provoke a massive boycott of certain search engines in favor of others that are more scrupulous.

This paper has argued that the still-immature Internet environment has become more functional as search engines have adopted some library values. This historical tendency might seem to imply that eventually all library values will be adopted by

search engine services. But the profit motive of search engine companies imposes limitations on this trend. For example, Google does not create recommendations of relevance, it only gathers them. Compared to the labor-intensive work performed by librarians, this automated gathering is relatively inexpensive and so makes profit possible. But while an automated system is cheaper than one directly managed by human

It seems that Google's ethical advances, admirable though they are, should be stated only as a matter of degree.

judgment, it is far less reliable and far more open to manipulation. No matter how much Google tries to maintain unbiased rankings, it is hard to see how it could ever attain the high standards of libraries. Similar criticisms can be made of the attempts to replace the reference interview through "personalization." These are analogs to library functions, and they represent great advances over earlier search engine practices. However, they are not the ethical and functional equivalents of the corresponding values provided by

libraries. It seems that Google's ethical advances, admirable though they are, should be stated only as a matter of degree.

It is possible that search engines have already adopted most of the library values that can be "monetized." What will happen to other library values that offer little potential for profit? Helen Nissenbaum and Lucas Inrona have argued that the Internet should be considered part of the public sphere because it fulfills "some of the same functions of other traditional public spaces—museums, parks, beaches, and schools," and that, accordingly, search engines must meet high ethical as well as technical standards.⁶⁵ Libraries also provide a public good, and library values have developed in the context of service, usually public service. It is crucial to identify those library values that cannot be embodied by privately held search engines. Letwin has made a good preliminary statement in this regard:

We should take care to remember what librarianship means in contradistinction to commercialized information, to remember the difference between individuals-as-citizens and individuals-as-consumers, and to remember that as librarians we are public stewards of the information commons and have an obligation to protect it.⁶⁶

Conclusion

Google's technological achievements have been effusively praised in the popular press, and partly for that reason this paper has perhaps underemphasized their importance. For instance, Google's technological innovation goes far beyond the PageRank algorithm, and it is likely that advances here, though not well publicized, are unprecedented. So it would be overreaching to say that Google's adoption of certain traditional library values has caused its enormous success.

Still, the importance of these values has been vastly underestimated. Consider just a few library values as they relate to the most common explanations of Google's success. Some might seek to weigh the contribution made by values against the contributions made by new technology and new business models, but the categories do not



separate out so neatly. Google's PageRank technology is an instantiation of the library function of citation analysis. This search technology has strong ethical components, in that PageRank and other link analysis algorithms bring one library value to the Web (by providing relevant results) and prevent damage to another one (by thwarting the work of unscrupulous Web masters). In the same way, Google's new business model cannot be considered to be distinct from questions of values. Here the crucial element is an ethical practice borrowed from journalism, namely the firewall between the information and advertising. To a certain degree, this firewall instantiates another library value—that service to users should not be compromised by private interest.

Google has brought some important library values and functions to the Internet environment. The PageRank algorithm has increased access to quality materials, and it has limited the influence of private interest by checking efforts at self-promotion. Google's design has also increased access, partly by simplifying the interface and partly by encouraging users to seek materials off site despite the search engine company's private interest in keeping users on site. Google has advanced the library value of unbiased selection by refusing to allow advertising to corrupt its results; it has advanced unbiased indexing by resisting the efforts of those who would "optimize" their site's ranking. While the practice raises serious questions concerning privacy, Google and other search engines seek to create an analog to the reference interview, employing user profiles to identify information needs and thereby facilitate access. So although Google does not embody all library values, it has brought many of them to the Internet environment, and there is little doubt they have been essential to its success.

Acknowledgments

This paper was developed from a presentation given at the Minnesota Academic and Research Libraries Conference held in Minneapolis, MN, April 2004. I would like to thank Johannes Britz for reading an earlier version.

James Caufield is assistant professor, University Library, New Mexico State University, Las Cruces, NM; he may be contacted via e-mail at caufield@lib.nmsu.edu.

Notes

1. Jefferson Graham, "Tech Titans Emulate Google's Success," *USA Today*, April 30, 2004.
2. Ibid.
3. Dominic Rushe, "Google Search—and Destroy," Business Section, *London Times*, April 4, 2004.
4. Griff Witte, interview with David Menlow, "Google Fans Fill Web with Buzz over IPO," *Washington Post*, January 13, 2004.
5. Gary Rivlin, "Google Is One for the Books," *New York Times*, August 23, 2004.
6. John Shinal, "It's about the Company: Plain and Simple," *San Francisco Chronicle*, August 22, 2004.
7. Rivlin.
8. "The Weakness of Google: The Internet," *Economist*, May 1, 2004.
9. Google "Corporate Information: Ten things," <http://www.google.com/corporate/tenthings.html> (accessed June 17, 2005).

10. Amy Harmon, "Is a Do-Gooder Company a Good Thing?" *New York Times*, May 2, 2004.
11. Gary Price, "What Google Teaches Us That Has Nothing to Do with Searching," *Webmastery, Searcher: The Magazine for Database Professionals* 11, 10 (November/December 2003), <http://www.infotoday.com/searcher/nov03/price.shtml> (accessed July 18, 2005).
12. Stephen J. Bell, "The Info Diet: How Libraries Can Offer an Appetizing Alternative to Google," *Chronicle of Higher Education*, February 20, 2004, 15.
13. Carol Tenopir, "Is Google the Competition?" *Library Journal* 129, 6 (April 1, 2004): 30.
14. Price.
15. Jim Newsome, "A Tough Day at the Citation Bar (A Story of the Googlization of Knowledge)" (presentation, Academic and Research Libraries Division, Minnesota Library Association, Minneapolis, MN, April 30, 2004).
16. Rory Letwin, "On Google's Monetization of Libraries," *Library Juice* 7, 26 (December 17, 2004), http://libr.org/juice/issues/vol7/LJ_7.26.html (accessed June 17, 2005).
17. Joel Achenbach, interview with Peter Lyman, "Search for Tomorrow," *Washington Post*, February 15, 2004.
18. American Library Association, *The Freedom to Read*, American Library Association (June 30, 2004), <http://www.ala.org/ala/oif/statementspols/ftstatement/freedomreadstatement.htm> (accessed June 17, 2005).
19. ———, *Library Bill of Rights*, American Library Association (January 23, 1980). <http://www.ala.org/ala/oif/statementspols/statementsif/librarybillrights.htm> (accessed June 17, 2005).
20. ———, *Intellectual Freedom Principles for Academic Libraries: An Interpretation of the Library Bill of Rights*, American Library Association (July 12, 2000), <http://www.ala.org/Template.cfm?Section=interpretations&Template=/ContentManagement/ContentDisplay.cfm&ContentID=8551> (accessed June 17, 2005).
21. ———, *Libraries: An American Value*, American Library Association (February 3, 1999), <http://www.ala.org/ala/oif/statementspols/americanvalue/librariesamerican.htm> (accessed June 17, 2005).
22. ———, *Code of Ethics of the American Library Association*, American Library Association (June 28, 1995), <http://www.ala.org/ala/oif/statementspols/codeofethics/codeethics.htm> (accessed June 17, 2005).
23. ———, *The Freedom to Read*, American Library Association (June 30, 2004), <http://www.ala.org/ala/oif/statementspols/ftstatement/freedomreadstatement.htm> (accessed June 17, 2005).
24. Achenbach, interview with Seth Godin, "Search for Tomorrow," *Washington Post*, February 15, 2004.
25. Of course this Web "collection" is, in some ways, not like a library collection since many of the highest quality resources are not freely available online.
26. For further explanation, see for instance Thomas M. Parks, "Search Engines," in *Encyclopedia of Information Systems*, ed. Hossein Bidgoli (Academic Press: Amsterdam), 4: 23–9.
27. Charles Arthur, "The Hit Factory: It Was Launched by Two Students in a Garage Seven Years Ago Today," *Independent* (February 24, 2004).
28. Parks, 23.
29. For more on the subject, see for instance Orrin E. Taulbee, "Algorithms," in *Encyclopedia of Library and Information Science*, 2nd ed., ed. Marianne A. Drake (Marcel Dekker, Inc.: New York, 2003), 98–100.
30. Google, "Our Search: Google Technology," <http://www.google.com/technology> (accessed July 18, 2005).
31. *Ibid.*
32. Wade Roush, "Search Beyond Google," *Technology Review* 107, 2 (March 2004), <http://www.technologyreview.com/articles/04/03/roush0304.2asp> (accessed July 17, 2005).
33. Achenbach.



34. Larry Page et al., "The PageRank Citation Ranking: Bringing Order to the Web" (1998), 1, <http://dbpubs.stanford.edu:8090/pub/1999-66> (accessed June 17, 2005).
35. *Ibid.*, 15.
36. *Ibid.*
37. "'Why haven't library system and information database designers included as many user-friendly features as Amazon and Google?' This question, posed at the recent annual meeting of the American Society for Information Science & Technology (ASIST), was answered by UCLA professor Marcia Bates, who said that from 20 years of research we know what should be included, but we've never had the money to implement fancy search features. Now we have a lot of catching up to do." Carol Tenopir, "Running with the Amazons," *Library Journal* 128, 20 (December 15, 2003): 34.
38. Google, "Corporate Information: Ten things," <http://google.com/corporate/tenthings.html> (accessed June 27, 2005).
39. Greta De Groat, interview with Monika Henzinger, "Perspectives on the Web and Google: Monika Henzinger, Director of Research, Google," *Journal of Internet Cataloging* 5, 1 (2002): 25. For related observations see Fiona Harvey, "Creative Business," *Financial Times* (March 30, 2004).
40. Technology Review.com, "Search Us, Says Google" (November/December 2000), <http://www.technologyreview.com/articles/00/11/qa1100.asp> (accessed June 17, 2005).
41. Overture Services began the auction model of advertising in which whoever paid the most was listed first. When users disdained this service, Overture "offer[ed] to split revenue with sites that display its listings on their search results pages. Yahoo, MSN, America Online, and all the other major sites—except Google— . . . agreed." Saul Hansell, "Google's Toughest Search Is for a Business Model," *New York Times* (April 8, 2002).
42. Economist.com, "How Good Is Google?" (October 30, 2003), http://www.economist.com/displaystory.cfm?story_id=2173573 (accessed June 17, 2005).
43. Technology Review.com.
44. Gary Rivlin, "Google Is One For the Books, Leaving Some With Regrets," *New York Times* (August 23, 2004).
45. Saul Hansell, "Google's Toughest Search Is for a Business Model," *New York Times* (April 8, 2002).
46. As recognized by David A. Vise: "The search engine has come to resemble a free circulation newspaper that makes its profit by selling ads. In Google's case, the company gives away the search engine for free . . . and then, like the newspaper, makes its money on clearly marked ads." David A. Vise, "Advertisers Bid, You Click, They Pay; Keyword Auction Nets Google Most of Sales," *Washington Post* (May 13, 2004).
47. David Pogue, "Microsoft on the Trail of Google," *New York Times* (July 8, 2004).
48. Among other things, paid indexing at Yahoo "has the benefit of 48-hour re-indexing, which allows you to make changes quickly and thus test hypotheses" about how to improve your position in the rankings. Stefan Spencer, "Casing the Competition," *Catalog Age* 21, 4 (April 1, 2001): 36.
49. Leslie Walker, "Yahoo Move Rekindles Pay-for-Inclusion Debate," *Washington Post* (March 7, 2004).
50. *Ibid.*
51. Danny Sullivan, "Google's (and Inktomi's) Miserable Failure," Search Engine Watch (January 6, 2004), <http://searchenginewatch.com/sereport/article.php/3296101> (accessed June 17, 2005).
52. David Suess, "Ten Years into the Web: The Search Problem is Nowhere Near Solved" (presentation, Computers in Libraries Conference, Washington D.C., March 11, 2004), <http://www.infoday.com/cil2004/presentations/> (accessed June 17, 2005).
53. Google warns that "illicit practices that may lead to a site being removed entirely from the Google index." Google, "Webmaster Guidelines," <http://www.google.com/webmasters/guidelines.html#quality> (accessed June 17, 2005). See also John Markoff, "The Coming Search Wars," *New York Times*, sec. 3, February 1, 2004.

54. Farhad Manjoo, interview with Daniel Brandt, "Meet Mr. Anti-Google," Salon.com (August 29, 2002), http://www.salon.com/tech/feature/2002/08/29/google_watch/ (accessed June 17, 2005).
55. American Library Association, *Intellectual Freedom Principles for Academic Libraries*.
56. Rushe.
57. Lance Ulanoff, "I Search, Therefore I Google," *PC Magazine* (November 26, 2003), <http://www.pcmag.com/article2/0,4149,1397999,00.asp> (accessed June 17, 2005).
58. Roush.
59. Achenbach.
60. Danny Sullivan, "Google Loses Tabs in New Look, Gains Web Alerts and Personalized Search Results," Search Engine Watch (March 30, 2004), <http://www.searchenginewatch.com/searchday/article.php/3332511> (accessed June 17, 2005).
61. American Library Association, *Intellectual Freedom Principles for Academic Libraries*.
62. Google, "Google's Gmail Privacy Policy," <http://gmail.google.com/gmail/help/privacy.html> (accessed June 17, 2005).
63. Matthew A. Goldberg, "Comment: The Googling of Online Privacy: Gmail, Search-Engine Histories, and the New Frontier of Protecting Private Information on the Web," *Lewis & Clark Law Review* 9, 249 (Spring 2005): 255.
64. Goldberg, quoting the Electronic Privacy Information Center, <http://www.epic.org/privacy/gmail/faq.html> (accessed June 17, 2005).
65. Helen Nissenbaum and Lucas D. Introna, "Shaping the Web: Why the Politics of Search Engines Matter," in *The Internet in Public Life*, ed. Verna V. Gehring (Oxford: Rowman and Littlefield Publishers, Inc., 2004), 19.
66. Letwin.