

**The Advertiser-Dependent Search Engine and
Its Implications for Information Professionals**

David Gwynn
LIS 600-01
Dr. Julie Hersberger

Revision: 20 November 2007

INTRODUCTION

Many have speculated that the commercial search engine may be the technology that finally puts librarians out of business. Indeed, a 2005 OCLC study showed that more individuals are now users of search engine than of traditional library facilities, and particularly of traditional electronic databases. Users also report a higher satisfaction rate with search engines than with libraries in five of seven study categories, mostly related to speed and convenience (DeRosa, Cantrell, Cellentani, Hawk, Jenkins, & Wilson, 2005, p. 2-18). However, commercial search engine technology is not without its limitations, and despite high levels of user satisfaction, the information provided may not always be accurate, relevant, and without bias. Search engine optimization strategies practiced by website owners, and semi-cloaked advertising strategies such as paid inclusion and paid placement can call the validity and relevance of search results into question, and may have implications both for end users and for information professionals.

This paper will explore issues surrounding the quality of information as provided through commercial search engines, and will seek to determine:

- If there is a problem with the nature of the information provided by commercial search engines,
- The extent of the problem - if any,
- Its nature, whether information or user-based, and
- Its potential impact on the role of librarians and other information professionals.

SEARCH ENGINE BASICS

Search engines are distinct from web directories in that the former are largely driven by automated processes, while the latter involve significantly more human intervention (Green, 2000). Search engines, Google being an example, engage in a process of "spidering", in which text and sometimes graphic content from all pages visited is acquired, copied to the search engine's servers, and indexed. Results are

then provided on a per-query basis to users, through a series of algorithms that rank pages in some manner specific to the search engine being used. Generally, snippets of the relevant text (often text featuring the search terms used) are included in the results with a link to the page from which it is taken. With Google results, there is often an additional link to a cached version of the page stored on Google's own servers the last time the page was indexed (Google, 2007a).

Web directories, on the other hand, are an attempt to catalogue the web, using human reviewers who visit the sites in question and assign them to specific content categories within the directory's parameters (Green, 2000). Yahoo began in 1994 as a pure web directory, but has since moved to a portal model, adding both search engine functions and such additional features as news, free email accounts, and job listings alongside its directory. Microsoft's MSN operates in similar fashion.

While Google offers many of these same services, its primary focus remains on the search engine. Launched in 1998, Google is almost universally recognized as the most-used and most-trusted commercial search engine. Its proprietary search results algorithm, PageRank, incorporates a form of citation ranking where prominence in search results is a factor of how many other pages link to the page in question. In addition, links are weighted based on the prominence the page doing the linking, and so on. Google stresses the integrity of its search results and asserts that "complex automated methods make human tampering with our search results extremely difficult" (Google, 2007b). Regaldo (2007, p. 3) notes that this method of determining authority based on popularity is an entirely new method of information retrieval, based on user input rather than the traditional publishing chain or authority.

Prominent placement (being among the first links at the top of the first results page) is a major issue for online businesses and content producers, since an extremely high proportion of website traffic is a direct result of search engine links. Human nature being what it is, most users consistently stop browsing search results after the first one or two pages, and are looking for *an* answer, not necessarily *the* answer. Search engine results have therefore become a big business (Goldsborough, 2005).

PAID PLACEMENT AND PAID INCLUSION

In the early days of the internet, search engines operated under a more traditional advertiser-supported system, financed primarily through banner ads, the web equivalent of print display ads. Search engines generated revenue based on ad impressions (number of views) and by “click-throughs,” where customers clicked on the ad and visited the sponsor’s site. Targeted ad sales based on specific keyword searches were a natural outgrowth, as were partnerships with online retailers to provide products ads related to user searches (Moxley, 2004, p. 61). Over time, however, users became more and more adept at ignoring banner ads, and a more sophisticated approach was developed: paid inclusion and paid placement.

Paid *inclusion* is the practice of charging a fee for expedited indexing and inclusion of a website on the search engine or web directory. Yahoo was among the first to offer the service, charging sites \$199 to avoid a possible six-month review delay in 1999 (Hansell, 2001). Non-paying sites would also be indexed eventually, but the fee guaranteed a quicker and more frequent indexing for those who were willing -- and could afford -- to pay it.

Paid *placement* involves payment for enhanced prominence at some specific place within search engine results. In most cases, paid placement links are fairly clearly labeled as a sort of “separate” content from regular, unsponsored, results. Designations such as “sponsored links” (Google), “sponsor results” (Yahoo) or “sponsored sites” (MSN) are common; the results are sometimes further segregated through the use of different colored backgrounds and graphical separators.

Whether this editorial boundary is self-imposed or is the result of government pressure, such as the Federal Trade Commission’s 2002 consumer alert, is anyone’s guess, but paid placement results have not always been so easily identifiable as numerous anecdotal reports have shown. Woodward (2004) found that links to for-profit treatment centers were the most common results on a search for “troubled teens.” Another noted that a search done on two separate but commonly-owned sites, Excite and Overture, produced the same top links on both sites; Overture identified the material as sponsored content, but Excite did not (Nicholson, Sierra, Eseryel, Park,

Barkow, Pozo & Ward, 2006, pp. 450-451). Such anecdotal evidence led Nicholson et al. to complete a quantitative study of the onscreen “real estate” of paid placement results, which found that, overall, 40% of the first screen (the part of the results page seen before scrolling down) was devoted to sponsored results. This study did not, however, consider the relevance of the results.

KEYWORDS, METADATA, AND SEARCH ENGINE OPTIMIZATION

Another area of commercial manipulation of search engine results is the use of keywords, metadata, and other search engine optimization (SEO) techniques. Meta and keyword tags are text-based information embedded in the code of a web page, visible to search engine indexing but not to the end user of the page. Originally envisioned by some as a potential standardized cataloguing aid for websites, the implementation of meta tags and keywords has not played out this way (Brooks, 2004). They are now seem to be primarily viewed as an aid to enhanced search engine placement, rather than as a means of imposing order on the web.

While some other search engines have been less hostile to metadata, Google has always regarded its use as a “bad faith” technique due to its cloaked nature; text that is neither part of main document text nor visible to the reader is deemed suspect. All the same, a large industry has grown up around SEO strategies designed to enhance placement on Google (and other search engines), despite the fact that Google closely guards its patented algorithms and constantly alters them to minimize such schemes. In fact, its entire business model depends on providing its end users with high quality and relevant or “unaffected” content (Brooks, 2004).

IMPLICATIONS FOR INFORMATION PROFESSIONALS

Given the tremendous level of both user adoption and user satisfaction with search engine technology, it seems unlikely that its use will decline in the foreseeable future, nor that websites and search engines will drastically change their *modi operandi* (Moxley, 2004, p. 65). In fact, librarians and other information professionals already use commercial search engines on a daily basis, and even the American Library Association itself uses Google as the search engine for its own internal website.

Advertiser-sponsored search engine results are undeniably controversial; questions about any potential problem with them seem to be threefold:

1. Are commercial potentially “tainted” commercial search engine results really a problem at all?
2. Assuming there is a problem, does it rest with the information itself or with the end user
3. What does it all mean for information professionals and end users.

Is there really a problem at all?

In order to determine whether or not there is a problem with the quality of commercial search engine results, it must first be determined if the results provided by commercial search engines is of sufficiently low quality to be of concern. Does “clutter” from advertisement and paid placement significantly detract from relevant and unsponsored content, particularly if the ads are prominently identified as such? As will be discussed later, there has been some anecdotal evidence of this, but these situations are increasingly rare, and most major search engines now have stated ethical policies on the issue.

There has been little actual research on the quality (as opposed to quantity) of sponsored search results. Since users apparently continue to click on sponsored results (Marable, 2003, p.15), and continue to report high levels satisfaction with search engines in general (OCLC, 2005, p. 2.18), it is very possible that such research might conclude that commercial results are as useful to their audience as unsponsored ones. It also seems likely that search behavior would vary based on the purpose for accessing the information; the nature of research for a scholarly publication is different from research on which DVD player to purchase, and it is probably safe to assume that most search engine users are not doing academic research. Information *wants* vs. information *needs* would no doubt result in different perceptions as to the suitability of information *retrieved*.

Information problem or end user problem?

A larger issue may be the apparent inability of end users to decipher the difference between sponsored and unsponsored results despite seemingly clear labels (Marable, 2003). There is a nagging tendency among many students in particular to be satisfied with whatever results they can find and to value convenience over quality (Regaldo, 2007, p. 2). Data released by the Educational Testing Service in 2006 showed that only 52% of students correctly judged the objectivity of a selected set of websites, and only 35% were able to successfully modify a search query to minimize the number of irrelevant results (Katz, 2007). Presumably, this tendency would be seen across other age groups as well. As Luther (2003) noted, for “many searchers, the quality of the results matter less than the process - they just expect the process to be quick and easy” (p.36).

Another issue with internet-only searches - one encountered frequently in the preparation of this paper - is the potential presence multiple versions of cited resources (pre-prints, rough drafts, related Powerpoint presentations, and the like) and the need to determine which version is the authoritative, or “final” version, if one even exists. While multiple copies of a document may also be retrieved from a library database product, they generally tend to be the same published version, the one that was ultimately published. This suggests that while a user-friendly tool like a search engine may be a good place to begin seeking material, the library’s database collection may be the better option for finalizing the research and for actual document retrieval.

Obviously, web-based information in general presents other special challenges, most notably the lack of a traditional (or any) editorial structure and process, since the barriers to entry to publishing have been significantly lowered. Not surprisingly, this has created a “much broader continuum” (Regaldo, 2007, p.2) of what may be considered authoritative. Any form of information requires filtering and assessment before it is appropriate to most research applications; the fact that search engine results are internet-based and commercially-supported merely makes this step somewhat more important than it already was. As Moxley suggests, librarians “must

make clear the distinction between services to our patrons and digestible matter for consumers” (2004, p. 65) by operating within a strong ethical framework and offering critical evaluations rather than merely presenting information.

Perhaps the larger issue is the lack of initiative, analytical skills, and context among information users, and not the quality or relevance of the information itself. If so, the role of librarians and other professionals as gatekeepers and filterers of information, not to mention as coaches and trainers, may become considerably more - - not less -- important in the coming years.

A value-added service

How can information professionals remain relevant amid this new technology? A good start would be to rethink the whole notion of competition with search engines as an “us vs. them” proposition, and to embrace the technology and its potential as part of a value-added package of services. Many reference librarians are already doing this. Abram (2005) recommended several strategies for adding value, include specialization among librarians, the very visible (to patrons) usage of search tools other than search engines, and an understanding of search engine limitations and issues such as SEO and paid placement. Essentially, librarians must educate themselves about the strengths and weaknesses of their “opponent” and use this knowledge to provide the most effective and informed package of services possible to their consumers.

The OCLC study (cited in the introduction to this paper) that found patrons and potential patrons rated search engines higher than librarians in five out of seven categories, also found that librarians had the edge in two very important ones: credibility and accuracy (De Rosa et al., 2005, p. 2.18). End users need help finding their way through the massive amounts of information currently available, determining what is available, what is and isn’t useful, and learning “information literacy skills” (Regaldo, 2007, p. 2). Librarians and trained information professionals are perfectly positioned to provide the training and guidance required to help patrons make their own informed decisions about the types of information they need, not to mention its credibility, relevance, and accuracy.

Obviously, the basic training on critical thinking must come from teachers and other professionals in conjunction with librarians. Areas in which librarians in particular can add value include:

- Understanding the strengths and limitations of search engines and sharing that knowledge with users to present a balanced package of research options, which may include the use of multiple search engines, or perhaps none at all (Moxley, 2004; Abram, 2005).
- Reminding users regularly that information sources other than search engines are available, and most importantly, facilitating and modeling their use (Abram, 2005). Licensing databases for online, home use may be a good start where economically feasible.
- Training users on specific search strategies to maximize the proportion of irrelevant results, such as the use of advanced search features, Boolean structure, etc. Nicholson et al. (2006) also found that entering synonyms in the search query can minimize the appearance of paid placement results. Tools like the online tutorial, GoogleGuide.com, which has separate modules for “novices”, “experts”, and “teens” could be a useful start.
- Marketing and promoting available services, and making sure their own library websites are accessible, provide necessary service and other information (De Rosa et al., 2005; Ojala, 2004).

The last area mentioned is particularly important because of its relationship to a growing online presence. Content providers are information professionals as well, and increasingly, librarians and teachers will need to see themselves in that role. An understanding of the ways in which online information is accessed will be essential for them. Libraries may soon find themselves on the other side of this issue, as they strive to make certain their own websites are easily discoverable and navigable, not to mention accessible to all users. There is already some discussion of the necessity for SEO for library websites (Ojala, 2004) and of the value of user information that can be gained using visitor statistics measured through an analytical program developed, interestingly enough, by Google as a means of enhancing its search and advertising functions (Fang 2007).

SUMMARY AND CONCLUSION

Commercial search engines are a fact of life for today's information professionals. They are probably the most used resource material today in much of the developed world, despite certain limitations including the potential for commercial manipulation of results. Librarians and other information professionals are well represented in that global user base, both as end users and as facilitators of other end users.

Like any other information source, search engine results must be analyzed and filtered to be truly useful for research purposes. If librarians and other professionals can begin to understand the limitations of search engines, and learn ways of minimizing these limitations, they can begin to offer their clients a more valuable information resource than the search engine alone could ever provide. Based on the literature, this author suggests the following priorities for information professionals:

- Understand the basics of SEO and paid placement and their impact on commercial search engine results.
- Recognize the ways end users perform searches and how the information returned will be used, and coach them on more effective strategies when needed.
- Encourage and facilitate the use of all available resources, including search engines, and recognizing which are the best in any specific situation.

The search engine is an important tool for casual users and information professionals. As with any tool, the trick is knowing how and when to use it.

REFERENCES

- Abram, S. (2005). Competing with Google in a special library. *Information Outlook* 9(11). 46-47. Retrieved 7 Spetmber 2007 from ProQuest Research Library database.
- Brooks, T.A. (2004). The nature of meaning in the age of Google. *Information Research*, 9(3) paper 180. Retrieved 6 September 2007 from <http://InformationR.net/ir/9-3/paper180.html>
- DeRosa, C., Cantrell, J., Cellentani, D., Hawk, J., Jenkins, L., & Wilson, A. (2005). *Perceptions of libraries and information resources: a report to the OCLC membership* [Electronic version]. Dublin, OH: OCLC Online Computer Library Center, Inc. Retrieved 8 September 2007 from http://www.oclc.org/reports/pdfs/Percept_all.pdf
- Fang, W. (2007), Using Google Analytics for improving library website content and design: A case study. *Library Philosophy and Practice 2007 (June)*, LPP Special Issue on Libraries and Google. Retrieved 6 September 2007 from <http://libr.unl.edu:2000/LPP/fang.pdf>
- Goldsborough, R. (2005). Web content: knowing whether it's to sell you or inform you. Retrieved 7 September 2007 from the World Wide Web: <http://www.infoday.com/linkup/lud070105-goldsborough.shtml>
- Google (2007a). Google Corporate Information: Technology. Retrieved 7 September 2007 from the World Wide Web: <http://www.google.com/corporate/tech.html/>
- Google (2007b). Google technology. Retrieved 7 September 2007 from the World Wide Web: <http://www.google.com/technology/>
- Green, D. (2000) The evolution of web searching. *Online Information Review* 24(2), 124-137. Retrieved 24 September from Emerald Full-text database.
- Hansell, S. (2001, June 4). Clicks for sale: Paid placement is catching on in web searches. *The New York Times*. Retrieved 7 September 2007 from <http://query.nytimes.com/gst/fullpage.html?sec=technology&res=9D04E1D7143FF937A35755C0A9679C8B63>
- Katz, I.R. (2007). ETS research finds college students fall short in demonstrating ICT literacy: National Policy Council to create national standards. *C&RL News* 68(1). Retrieved 6 September 2007 from <http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2007/january07/ets.cfm>
- Luther, J. (2003). Trumping Google? Metasearching's promise. *Library Journal* 128(16), 36-39.

- Marable, M. (2003). False Oracles: Consumer reaction to learning the truth about how search engines work. Retrieved 8 September 2007 from <http://www.consumerwebwatch.org/pdfs/false-oracles.pdf>
- Moxley, D., Blake, J., & Maze, S. (2004). Web search engine advertising practices and their effect on library service. *The Bottom Line: Managing Library Finances* 17(2), 61-65. Retrieved 6 September 2007 from Emerald full-text database.
- Nicholson, S., Sierra, T., Eseryel, U.Y., Park, J., Barkow, P., Pozo, E., & Ward, J. (2006). How much of it is real? Analysis of paid placement in web search engine results. *Journal of the American Society for Information Science and Technology* 57(4). 448-461. Retrieved 7 September 2007 from Wiley Interscience database.
- Ojala, M. (2004). Why information professionals should care about web search optimization. Paper presented at INFORUM 2004. Retrieved 8 September 2007 from http://www.inforum.cz/inforum2004/pdf/Ojala_Marydee2.pdf
- Regaldo, M. (2007), Research authority in the age of Google: Equilibrium sought. *Library Philosophy and Practice 2007 (June), LPP Special Issue on Libraries and Google*. Retrieved 6 September 2007 from <http://libr.unl.edu:2000/LPP/regalado.pdf>
- Woodward, C. (2004). Search Engines: Clogged with commerce and begging for an upgrade. Retrieved 7 September 2007 from <http://www.askquestions.org/articles/searches/>