The Wayback Machine - https://web.archive.org/web/20060503022825/http://eastbound.info:80/journal/2006-1/f…

Bettina Fabos

posted by **Bettina Fabos** published in 2006-03-13

The Commercial Search Engine Industry and Alternatives to the Oligopoly

Abstract

This essay details the search engine industry's transformation into an advertising oligopoly. It discusses how librarians, educators, archivists, activists, and citizens, many of whom are the guardians of indispensable noncommercial websites and portals, can band together against a sea of advertising interests and powerful and increasingly overwhelming online marketing strategies.

Keywords:

search engine industry, Google, Yahoo, MSN, OAI-PMH

Download as PDF

FreeWheel, Ex. 1048

The Commercial Search Engine Industry and Alternatives to the Oligopoly

Bettina Fabos - *Interactive Media Studies and Journalism, Miami University of Ohio*

This essay details the search engine industry's transformation into an advertising oligopoly. It discusses how librarians, educators, archivists, activists, and citizens, many of whom are the guardians of indispensable noncommercial websites and portals, can band together against a sea of advertising interests and powerful and increasingly overwhelming online marketing strategies.

Keywords: search engine industry, Google, Yahoo, MSN, OAI-PMH

> ***Google's approach—to create a situation where ideas and material not on the Web cease to exist—may actually have the effect the company wants.***
>
> *--Chris Nolan, Eweek, August 18, 2005*

> ***Google is probably not going to do anything that doesn't have a profit return on it.***
>
> *--Steven J. Bell, library director, Philadelphia University, 2005*

Introduction

The search engine industry, widely regarded as helpful, user friendly, impartial, democratically-oriented (bringing all types of information, including entire books, to the web), and doing its best to scour the *entire* web to find the most relevant information for users, has an admirable public image. Users tend to see search engines as empowering, trust that search engines are acting in their best interests, and are grateful to have access. Moreover, because there seem to be so many search engines to choose from (e.g., Google, Yahoo, MSN, Lycos, AOL, Dogpile, etc) users also feel empowered in knowing that, in the "democracy" of free enterprise capitalism, they can abandon ship for another, much more relevant search engine if results aren't satisfactory.

There is indeed some truth to users' ideas about search engines: they *are* formidable navigation tools. And yet, the search engine industry's responsiveness to and prioritization of users is built up more by persuasive public relations than by reality. The philanthropic façade of search companies, made more robust through activities such as the digitization of research and public libraries (Google and Yahoo!), works successfully to hide an increasingly profitable information and advertising industry. Search engine companies, of which there are really only three, have morphed into advertising conglomerates and now serve advertisers, not users, in a mutual, rather delightful, relationship. The advertiser pays the search engine to be affiliated with certain key words; the search company provides the sponsored links, which users click on; and traffic is driven to the sites of advertisers. Indeed, users are now universally described as *consumers* in the marketplace rhetoric of search engine enterprise; they are the pawns who the business world seeks to manipulate into clicking those links that will ultimately lead to the most profits.

In this arrangement, helping users find relevant information is a priority only in that, like other commercial media systems (think radio or television), there has to be *some* decent content to create a perception that Internet users matter. In fact, users only matter to the extent that they participate in the commercial system by knowingly—or unknowingly—clicking on sponsored links.[1] Keyword advertising generated an unprecedented $3.9 billion in 2004. By 2005 Google's advertising profit alone grew nearly 500 percent, and continues to exceed the most optimistic expectations (Gaither, 2005). The commercial role of search engines only promises to grow.

Moreover, with hardly any public or news media scrutiny (and, in fact, a tendency in the news media towards search engine glorification), three companies have come to dominate the search universe: Google, Yahoo and MSN. These big three giants are thirsty for the profits that come from connecting consumers to anything and everything for sale. Indeed, the big three are skewing the nature of all online information in favor of commercial enterprise, and will enormously impact and power the direction of information access and, indeed, democratic discourse, in the years to come.

This topic should be one of key importance to librarians, educators, most academics, archivists, activists, and citizens, many of whom are the guardians of indispensable noncommercial websites, portals, and public information in general. These noncommercial efforts may certainly face marginalization if the search industry continues to consolidate power, if internet users continue to buy into the successful mythmaking about the impartiality, relevance, competition, and breadth of the search engines they regularly use, and if clear alternatives to commercial search engines are not put in place.

No Surprise: Search Engine Commercialization

[1] *A 2005 survey of 2000 internet users commissioned by iCrossing, the leading search engine marketing firm, determined that more than half (56 per cent) did not know the difference between sponsored and natural search engine results. (Morrissey, Brian, 2005, June 21.* MediaWeek. *[Online]. Available: http://www.mediaweek.com/mw/news/interactive/article_display.jsp?vnu_content_id=1000096570*

The web has been a commercial medium since 1995, when the U.S. government (which had backed the web's development) quietly sold the Internet's backbone to private enterprise. It was at this time, ten years ago, when, instead of positioning the Internet as a commercial medium, we saw a tremendous push--from the Clinton Administration, Bill Gates and the computer and telecommunication industries in general--to position the Internet as an educational tool. High-profile television advertising during the 1996 Olympics and various Super Bowls from technology heavyweights such as MCI, Microsoft, AT&T, Oracle promoted the Internet as a panacea for education and democracy. For example, MCI ("Is this a good time or what?") showed images of savvy children in front of computers enchanted by various medical, historical, and astronomy websites. The children floated above their school desks with the wonders of the educational possibilities of online learning (Fabos, 2004).

In this way, the web was characterized as an uplifting, if not resuscitating educational tool. The push, however, was not necessarily to bring the promised "universe of knowledge" (Clinton's words) to all young and "lifelong" learners alike. Instead, the push was a careful public relations strategy to build up a user base so that the web could become a viable commercial advertising medium. Indeed, the rhetoric and accompanying media campaign of the mid-1990s had been successful: in just five short years, the web (as part of the larger Internet) became a mass medium--faster than any communication medium before it (Fabos, 2004).

Keyword Advertising Comes of Age

If the Internet was primed for advertising, it wasn't until 1998 that search engines became the logical tool to connect advertisers to consumers. There were five main search engine providers by 1998: Google, Alta Vista, AlltheWeb, Teoma, and Inktomi. These companies had all designed complex algorithms for searching the web, and had assembled databases of websites on which they applied their algorithms. They generated revenue through syndication, not advertising. For example, Inktomi syndicated its search technology to the Yahoo search portal, and Teoma syndicated to Ask Jeeves (among others). Google is (and always has been) both a search engine provider (syndicating its services) *and* a search portal.

By 1998, the big money in the internet industry was thought to be in banner ads, and in creating content-rich pages that supported banner advertising, not in providing impartial navigation tools to other web portals. This logic prompted the search engine provider (and also portal) Alta Vista to fashion its search portal into a content page and try to entice users to stay for a while and view the banner ads. Thus, Google, Alta Vista, AlltheWeb, Teoma, and Inktomi were only moderately successful companies. The industry breakthrough happened with the startup Goto.com in 1998. Goto began combining its impartial (now widely referred to as "organic") algorithmic searches with a database of advertisers. Unbeknownst to

users, Goto's searches put the highest advertising bidder at the top of its search engine lists. This placement was significant: studies had indicated by then that users didn't click past page three. Consequently, with users clicking on sponsored links thinking they were legitimate search results, and with traffic steering effectively to the pages of the paying customers, business clients flocked to Goto's increasingly popular service.

Goto charged its customers per click-through in a scheme called "Pay-For-Performance." Because Goto shared a portion of its revenue with the search portals it partnered with, search portals were as happy as advertisers. In fact, the entire business world exalted at the success of what is now termed keyword advertising. With banner ads failing to make an impact, a commercially viable advertising system on the internet had arrived. Besides, the growing justification among internet industry marketers was that people wouldn't know, and if they did they wouldn't care, if search engines deliberately (even clandestinely) skewed content to serve their sponsors. Shopping, it was widely agreed, was the future of internet information seeking: no ethical lines about information access were being crossed.

Towards the end of 2000, GoTo renamed itself Overture (a name, the company felt, that better represented its mission), disbanded the GoTo search portal, and concentrated solely on its new role as a commercial advertising database. This for-profit database was then syndicated to Web portals, where the Overture database could be cross-listed with the search lists generated by one of the top five "impartial" search engine providers (Google, Inktomi, AlltheWeb, AltaVista, or Teoma). It was up to the portal to decide how to identify the new sponsorship-infusing result lists. Some portals clearly indicated in the sidebars or banner bar where the sponsored links fell ("sponsor results," "sponsored matches," "sponsored links"). Others surreptitiously hid the links within the organic list itself.

By 2002, Overture had signed up 80,000 advertisers (Overture, 2003) and was distributing its for-profit search results to tens of thousands of web portals across the internet, including MSN, Yahoo!, Lycos, AltaVista, HotBot, Netscape, AOL, Infospace, Fast, and ESPN.com. Regardless of visible above-the-line sponsorship or invisible within-the-list sponsorship, the advent of Overture's rise meant that a search within a search portal was more and more likely to yield commercial results, especially since money exchanged hands each time a portal facilitated a click-through.

Overture's success prompted Google to begin its own keyword advertising initiative in 2002, called AdWords. Amassing its own index of commercial sites, the Internet's most popular search engine provider and portal began cross-listing keyword searches with its sponsored links. Unlike Overture, however, Google took the higher road and refused to facilitate integrated sponsorship within organic search results. The company also settled on a different payment plan, essentially

combining Overture's auction system of selling key words and placement to the highest bidder with an algorithm that factored relevance, or the ad's click-through rate, into placement. With the almost immediate success of keyword advertising, marketing strategists advised their clients to cover all their bases and sign up with both the Overture and Google AdWords advertising plans (Schachter, 2003).

The Success of Paid Inclusion

Introduced by the search engine provider Inktomi in 2001, paid inclusion was the first time an impartial search engine provider intentionally infused its organic, "impartial" results with commercial listings. Recall that portals were already integrating Overture's sponsorship within their search result lists. Now with paid inclusion, the navigation tools themselves were finding enormous monetary benefits to inviting paying customers to infiltrate their result lists.

This is how paid inclusion works: customers/advertisers who pay a flat fee are *guaranteed to be included* every time an algorithmic search engine completes a search. Because search engines do not search the entire web, only parts, and because many sites can easily slip away until algorithms are refreshed or updated, paid inclusion guarantees constant inclusion in that search engine provider's index. Niche topics that have fewer websites to compete with, like "Plantar Fibromatosis," are especially beneficial for the advertiser. Inktomi's model, while causing controversy among consumer advocates, was soon copied by every major search engine provider except Google, meaning that, by 2002, Inktomi, AlltheWeb, Teoma, and AltaVista were all offering paid inclusion as part of their overall syndicated package.

As surely as paid inclusion is profitable to search engine providers, there is also a noteworthy fringe benefit to advertisers investing in paid inclusion: part of the flat fee involves advice on how to write advertisers' listings to further enhance their position. "Since [commercial search engines] alone understand how the algorithms inside their search engine 'black boxes' work," *Financial Times* reporter Richard Waters observed, "they generally know how to game the system, though it is a power they claim to use responsibly" (2003: 30). In other words, even if paid inclusion clients didn't pay for prominent placement directly, at least they got the tools to figure out how to get there.

In the big picture, the addition of paid inclusion meant that search engine providers had a highly profitable revenue stream beyond mere syndication. As such, a wave of mergers and acquisitions began by early 2003: Yahoo! acquired Inktomi ($235 million) and Overture scooped up both AlltheWeb ($70 million) and AltaVista ($140 million). "The paid-inclusion model is really icing on the cake," said Yahoo! Chief Financial Officer Sue Decker in 2003. "That alone really justifies the price of the transaction" (Reuters, 2003:2). A few months later, Yahoo! then acquired Overture. Thus, within half a year Yahoo! had gobbled up three of the

five leading search engine providers, as well as a massive advertising index with which to cross-list all the searches. Yahoo! could now compete head on with Google.

Google, Search Engine Marketers (SEMS), and Contextual Linking

Even though Google vowed to be more forthcoming by promoting a fire wall between its Adwords program and its organic listings, and by opting out of paid inclusion, Google's search engine lists are nevertheless deeply affected by commercial bias. The heavy presence of commercial pages within Google's result lists is due, in part, to the rise of search engine marketing, which solely exists to influence placement within the databases of search engine providers and maximize the overall visibility of their clients' web site. The search engine optimization (SEO) market, which offers "positioning" and "advisory & marketing" services to its clients, is flourishing.

Moreover, Google, which remains the most popular search engine provider, is a sort of Holy Grail for search engine marketers (SEMs), and a tough nut to crack because the company is especially secretive about its constantly-evolving search algorithm. However, because Google's PageRank algorithm strategy is partially based on the number of links pointing to a site (ostensibly making it more "popular," and therefore more worthwhile to most web searchers), SEMs have become especially savvy to the linking game, working with their clients to increase the number of links leading to their clients' web sites. We hear about this practice in popular culture: for example, pranksters and political activists turned official websites for 2004 Democratic nominee John Kerry and President George W. Bush into the top listings for search terms like "waffle" or "miserable failure" respectively. The ranking was achieved when large numbers of people added links to their pages (connecting a particular search term, like "waffle," to an inserted link, like Kerry's official homepage), and thus increasing the page rank's association with a key term. This strategy is called "Google bombing" in the mainstream media, and considered a harmless novelty. Meanwhile, enterprising SEMs (whom we don't tend to hear about) use the term "horizontal marketing," and do anything they can to increase linkage for paying clients. This includes specializing in particular areas such as health and insurance-related sites better to shape web rings of reciprocal links. Blogrolling is another common way SEMs have generated more links: by applying the popular software supplied by Blogrolling.com, a user can add links to a blogpage with one easy click, which in turn more easily leads to link-farming, the practice of creating web sites with nothing but links. As Jill Walker (2005) explains in her helpful analysis of the link economy:

> ***There is a black market for links. You can pay dollars or kroner or yen to buy links to your site from link farms, circles of sites with nothing***

> ***but links. There is also a common law perception of link prostitution or link slutting: shamelessly selling one's own integrity for links.***
>
> *(p. 3)*

Because Google's market success is dependent upon its perceived credibility (the company's motto is "Don't Do Evil"), Google has heavily discouraged link farming, and has punished link-farming companies and their clients with lower search results. Meanwhile, as SEMs continue to be punished and are, as a consequence, losing a successful marketing tool, Google has emerged a winner on two counts. First, the company can continue to boast its commitment to search engine integrity. Second, with the weakening of link farming, advertisers have become increasingly dependent on Google's very successful AdWords program (Goodman, 2003). They have also become increasingly dependent on reciprocal linking as a necessary marketing tool. In this regard, Google is doing a fine job to accelerate this trend, which in the business world is referred to as "contextual linking."

A contextual link is a link to another site that matches the context of the main Web site. Today, users can click on contextual links at the bottom of nearly every online article in a commercial publication. But, small as they are, they are effective far beyond the advertising spot on a given page; the act of linking is also an act of endorsement, and consequently increases the company's PageRank standing in Google search results. Moreover, the more prominent the website that contains the contextual link, the greater the endorsement, and the higher the page rank.

Link farms have never been nearly as effective at influencing Google's ranking system as individual links from a highly prominent web site. A link, for example, to the used pick-up truck company Bronco Graveyard (broncograveyard.com) that appears on the home page of the popular trucking magazine *Truckin'* (truckin.com) can do a world of good in terms of enhancing Bronco Graveyard's visibility; a link on the popular Tennis.com website to the less-known raquetdepot.com also helps increase the small web site's "popularity," and thus its ranking on regular search results. As Walker explains, "The economy of links is not product-oriented. It is service-oriented, and the service is the link. The link is an action rather than an item; an event rather than a metaphor" (2005: 2).

Contextual links are thus highly valued, with commercial online publications quickly jumping into the contextual link business by giving advertisers the option to buy links on their home page, as Tennis.com does through its "Tennis Magazine MarketPlace Program." Although SEMs have worked hard to establish reciprocal links between smaller sites, it turns out that it's the more well-connected and powerful search engine companies, Google and Yahoo!/Overture, that are the most busy brokering contextual link deals through their massive index of advertisers.

Google introduced its contextual linking program, called AdSense, in 2003, while Yahoo! introduced Content Match a month later (Acohido, 2003). Both programs

broker contextual links on content Web sites.[2] Yahoo!, for example, supplies sponsored links to CNN.com and Wall Street Journal.com. Google supplies sponsored links to *U.S. News & World Report*, the Weather Channel, and ABC.com (Mangalindan, 2003). Its purchase of Sprinks in 2003 (a pay-per-click advertising network owned by media conglomerate Primedia), and a resulting relationship with Primedia (which, among other media products, owns the largest number of niche magazines, all of which have an online presence), will allow Google to supply contextual links to all these publications. Google's drive to plant more and more contextual links among prominent pages across the web, a process that increases the prominence of all these commercial pages within the Google PageRank system, very much undermines the company's own line about search engine result integrity.

To repeat, with the above developments, both Google and Yahoo! are becoming more like general online ad agencies than search engines, and like ad agencies they increasingly measure "ad" performance and collect consumer data. Consequently, both measure the results of ads by tracking the clickstream data, cookies, pixel tags, and contact/personally identifying information of search engine users. While Yahoo!/Overture relies on large, web-based company Doubleclick for this purpose, Google uses its new subsidiary, Kaltix, a start-up company that has developed profile-tailoring software to better target individual users by tracking their choices on the web. Google purchased Kaltix in October 2003 (Mangalindan, 2003). As *Wall Street Journal* reporter Mylene Mangalindan observed, "By gathering more data on each Google user, the reasoning goes, the search engine would know that a search for 'apple' is one for fruit rather than computers" (2003: B1). Both Yahoo! and Google are also working towards providing successful geo-location functions to their marketing toolkits. These ad services identify users' specific locations, and thus enable local advertisers to use search engines as a marketing strategy.

Interestingly, Google defines these local business opportunities in terms of greater democracy. In 2003, Google's director of product management argued that her company enabled democracy because anyone, even small advertisers, could advertise via Google (Mangalindan, 2003). Accordingly, in the world of search (as spoken by representatives from the "search engine of integrity"), the notion of online democracy no longer has anything to do with regular users—the democracy of ideas—but applies only to the advertising world—the democracy of the marketplace. The word "relevancy" has also come to have new meaning in the world of search business-speak. Rather than attempting to deliver the most

---

[2] ***With Yahoo! increasingly mirroring Google's PageRank system, and with such prescribed contextual linking in place, the search results of both search engine provider/portals are now nearly indistinguishable, especially in terms of their promotion of the most prominent sites (with which they have advertising relationships). As such, the notion of search engine variance is discredited. As Hindman et al (2003) have observed, "All modern search engine algorithms—including those radically different from Google's PageRank—tend to return these most connected sites first" (p. 27).***

relevant information to users, the task is now for search engines to deliver the most relevant consumer information to consumers. As business reporter David Crowe stated in the *Australian Financial Review*, "keeping advertisers happy with their paid searches is now the most important objective for the big search companies" (2003:29).

The Oligopoly

Three main companies, all American-based, dominate the search industry: Google, Yahoo!, and Microsoft. Google has been involved in search technology from the company's inception (1998), syndicating its search engine provider services and becoming a favorite search portal among users. Yahoo!, which began as a directory and search portal (outsourcing to Inktomi, and then Google, for its search technology), became a dominant player with its purchase of Inktomi, and then Overture (which in turn had just purchased AltaVista and AlltheWeb), in 2003. Since these acquisitions, Yahoo! has been in direct competition with Google for search portal prominence. Yet Microsoft, with its powerful MSN portal, is on the horizon as potentially the most dominant player of all.

Microsoft's attempt to acquire Google in 2003 was not surprising given the company's nearly limitless resources and stated ambitions in content acquisition. As early as 1995, Gates was talking about going "well beyond simply providing a pipe for bits" (1995:241-242). Microsoft, in Dawson and Foster's words, is "interested in moving up 'the economic food chain' from the delivery and distribution of bits at the bottom to computer applications and services and content at the top. Such companies want to own the bits, not simply deliver them" (1998: 60). However, Google denied the partnership or takeover opportunity (at least for the time being), and Microsoft turned to Plan B.

For most of 2003, MSN was still relying on Inktomi to power its organic searches, Overture to power its commercial searches, and Looksmart as a fortifying directory. In other words, MSN was deeply dependent on subsidiaries owned by Yahoo! But by October, after the Google talks unraveled (Teather, 2003), MSN ditched Looksmart and started work on its own search engine platform, resolving to drop Inktomi and Overture sometime after 2006, when it would roll out its "Google-killer" search engine algorithm (Bazeley, 2004). Thus, Microsoft had begun to amass a proprietary index of sites from which to conduct searches and was hiring hundreds of engineers to work on web-searching algorithms to top Google. Moreover, the software company planned to integrate its search technology *directly into its Windows operating system* under a project code-named "Longhorn" (Mangalindan, 2003).

For anyone familiar with Microsoft's history of annihilating the competition, this strategy seems similar to Microsoft's triumph over Netscape in the web browser business. "Today we are number one in email, we are number one in messenger. Our ambition is to be number one in search," Sharon Babyle, the general manager

of MSN's consumer Internet service, said at the end of November 2003 (Conners, 2003). Given Google's subsequent release of its new desktop computer search software (Google Desktop Search), which allows users to search their desktops far more efficiently than Microsoft (Bazeley, 2004), it is clear that, if anything, Google will put up a good fight. But, now that Google's future requires the company to attend to the demands of shareholders, many analysts are forecasting damage to Google's search integrity. As the opening sentence of a story in *Wired* plainly said, "The world's biggest, best-loved search engine owes its success to supreme technology and a simple rule: Don't be evil. Now the geek icon is finding that moral compromise is just the cost of doing big business" (McHugh, 2003). The first step in moral compromise was the Google Desktop Search program, which increases the company's ability to target users with personalized advertising. Likewise, the company's Gmail program, introduced in March 2004, examines the content of individual emails and sends users' marketing information back to company headquarters. The chances for more moral compromise and commercial intrusions into Google's search listings increase as Microsoft and Yahoo! ramp up the competition.

These days, even Google, the "ethical" search engine with the company motto "Don't Be Evil," is now focusing most of its attention on ad placement, either through the sponsored links it brokers on its own search pages, through contextual links on other content pages, or through data mining. Reflecting on the company's motto after Google went public in 2004, a *New York Times* editorial stated "Such idealistic talk out of Silicon Valley, so seemingly empowering back in 1999, seems embarrassingly naïve now that the party's ended, at least for the rest of us" (Googling, 2004:10).

Conclusion: Digital Archiving, Subject Gateways and Open Source Software

Despite the considerable implications of search engine commercialization for knowledge access, the topic has not gained much attention in academic and library spheres. One reason for this lack of attention is good public relations: the search engine industry continues to highlight "integrity," "relevance" and "objectivity" as a mantra, meanwhile marginalizing noncommercial, educational, or even merely controversial (or unpopular) information (Gerhart, 2004). Another reason is the general silence in the U.S. mass media when it comes to any criticism of commercialism (since the mass media themselves are major participants in advertising-supported commercial media culture). Consequently, educators and citizens in general, for the most part, continue to have a largely optimistic outlook on search engines as helpful and trustworthy educational, research, and informational tools. Meanwhile, search engines continue to be the primary destination for student research (Griffiths and Brophy, 2005; Rainie and Hitlin, 2005).

Despite its heavy commercialism, the web's potential as a place for online scholarship and diversity is still evident. But, to realize the web's educational and noncommercial potential, educators and librarians need to move away from promoting individual skills (advanced searching techniques, web page evaluation skills) as a way to cope with excessive commercialism. The problem here is that, while it may feel empowering to teach or possess these skills, a wholesale critique of the commercial web structure and the future of the internet (commercial interest vs. public interest) remains sidelined (Fabos, 2004).

Indeed, educators, librarians and citizens need to address the complex and economically-charged structure of the web that affects all search results regardless of how well one crafts an individual search. We should address the increasing difficulties of locating content that is *not* commercial, and the misleading motives of the commercial, publicly-traded internet navigation tools, and the constant efforts among for-profit enterprise to bend the internet toward their ends. Evolving internet marketing strategies are in plain view in the pages of business journals and industry trade magazines. They are being taught, globally, in university business courses. They are being celebrated at search engine marketing conferences, of which there are many. And there are books, such as the recent release, *Search Engine Marketing, Inc: Driving Search Traffic to Your Company's Web Site*, which illustrate the growth and competition with regard to dominating search. "You are not the only person joining the search marketing game," Moran and Hunt write. "At the beginning of this book, we trotted out all the 'gee-whiz' numbers to show you how search marketing is growing. So the good news is that you are catching the wave, but the bad news is there are a lot of other surfboards out there to contend with" (Moran and Hunt, 2006: 490).

If we want to go beyond a mainstream, commercialized, sponsored online information repository we need to turn to a different structure that offers a more inclusive, democratic information environment. As it turns out, there is hope (although it comes with acronyms that are a lot harder to remember than catchy commercial search engine names like Yahoo! and Google). Numerous computer scientists and digital librarians have been developing open source technology, such as the Open Access Initiative for Metadata Harvesting Protocol (OAI-PMH), iVia, and Data Fountains, that offer (and enhance) a user's ability to search across multiple (that is, thousands of) subject gateways. These digital repository harvesting services imitate the functions and interface of a search engine, but they can be molded to search in specific academic areas. In other words, one can create completely noncommercial searching environments that offer the scope and feel of a search engine. These developments have profound implications for academic research.

The initial framework OAI-PMH was developed in 2001, and since then has been embraced by digital librarians worldwide as a means for sharing metadata across subject gateways. As Shreves, Habing, Hagadorn and Young explain, "The mission

of the Open Archives Initiative, the entity responsible for the protocol, is to 'develop and promote interoperability standards that aim to facilitate the efficient dissemination of content'" (2005: 577). Some good examples of subject gateways that involve OAI-PMH are the Resource Discovery Network (a UK-based subject gateway network), Renardus (an academic subject gateway collective that serves Europe), The National Science Digital Library (which is funded by the National Science Foundation and focuses purely on science-related sites), INFOMINE (based at the University of California at Riverside), and OAIster, based at the University of Michigan.

Beyond the OAI protocol, which basically links gateways together, other archiving initiatives aim to make it easier to build subject gateways, combine subject gateway networks, and help digital librarians standardize and automate data collection. For example, Edward Almasy, co-director of the Internet Scout Project, has with his colleagues developed two open source software packages called the Scout Portal Toolkit (SPT) and the Collection Workflow Integration System (CWIS). These technical resources are a user-friendly means for building high-quality subject gateways. "They allow a group or organization (or even an ambitious individual)," Almasy writes, "to share a specific knowledge base via a full-featured portal on the Web, with little or no investment in technical resources or infrastructure" (2005: 621). INFOMINE co-founder Steve Mitchell has also developed the iVia and Data Fountains software platforms, which each expedite data collection-building once a subject gateway is built (Mitchell, 2005). Other initiatives have aimed to help digital librarians standardize subject gateway metadata, so more and more gateways can be effectively linked together (Kelly, Closier, and Hiom, 2005).

One also can't underestimate the important expansion of the collaborative subject gateway movement. Contributor-run archives such as ibiblio.org, the Linux Documentation Project, the Degree Confluence Project, Merlot, and Etree.org are examples of technologically inventive portals supported by passionate volunteers who cooperate to build these open source services (Jones, 2005). Perhaps the fastest-growing area of subject gateway development is the wiki movement. Wiki (which means "quick" in Hawaiian) is a social software trend that enables any user to edit and build a given web page within a wiki site. Wikipedia, for example, is a collaboratively-built encyclopedia that involves more than 350,000 volunteers, who by 2005 had contributed more than 2 million entries, and made it the most popular online reference site (Auchard, 2005). Volunteer contributors change entries in an atmosphere of trust and public goodwill, and all former entries are archived so a user can see how a certain topic area has evolved. No librarians are involved in these collaborative gateways, just experts and public citizens dedicated to sharing their knowledge and/or creative efforts with others.

The ongoing work toward subject gateway development—all of it developed as free, open source software—provides a small but growing countervailing force to

the commercialization of "the universe of knowledge." Underlying all these efforts is the understanding that, for a democracy to function properly, one needs access to all kinds of information, not just information with a commercial purpose. Also underlying these efforts is the understanding that, in our commercial system, educators, librarians and citizens interested in nurturing a public sphere must work together to control the destiny of the internet—or somebody else will.

References

Acohido, B. (2003) 'Simplifying the Info Hunt.' *USA Today* 11E November: 17

Almasy, E. (2005) 'Tools for Creating Your Own Resource Portal: CWIS and the Scout Portal Toolkit.' *Library Trends* 53(4) 620-636

Auchard, E. (2005) 'News Junkies Discover Wikipedia.' *Seattle Times* E3 September: 12

Bazeley, M. (2004) 'Google Rolls Our Search Software.' *San Jose Mercury News* B2 October: 15

Conners, E. (2003) 'Microsoft in Search of Top Spot.' *Australian Financial Review* 32 November: 25

Crowe, D. (2003) 'In Search of a Profit Engine.' *Australian Financial Review* 29 October: 18

Dawson, M. and Foster, J. B. (1998) 'Virtual Capitalism.' In *Capitalism and the Information Age: The Political Economy of the Global Communication Revolution*. R. McChesney, E. M. Wood and J.B. Foster (eds.) New York: Monthly Review Press: 51-68

Fabos, B. (2004) *Wrong Turn on the Information Superhighway: Education and the Commercialization of the Internet*. New York: Teachers College Press

Gaither, Chris (2003) 'Searching For Dollars As Banner Market Flags, Net Giants Pin Hopes On Revenue From Targeted Ads Related To Specific Queries.' *The Boston Globe* F1 February: 23

Gates, Bill (1995) *The Road Ahead*. New York: Viking

Gerhart, S. L. (2004) 'Do Web Search Engines Suppress Controversy?' *First Monday*, 9(1) January [online] available from http://firstmonday.org/issues/issue9_1gerhart/index.html [23 January 2004]

Goodman, A. (2003) 'Cracking Google's New Algorithm.' *WebProNews* [online] available from <http://www.webpronews.com/ebusiness/smallbusiness/wpn-2-20031203Cracking GooglesNewAlgorithm.html> [24 September 2005]

Griffiths, J. T. and Brophy, P. (2005) 'Student Searching Behaviour and the Web: Use of Academic Resources and Google.' *Library Trends* 53(4) 539-554

Jones, P. (2005) 'Strategies and Technologies of Sharing in Contributor-Run Archives' *Library Trends* 53(4) 651-662

Kelly, B., Closier, A. and Hiom, D. (2005) 'Gateway Standardization: A Quality Assurance Framework for Metadata' *Library Trends* 53(4) 637-650

Mangalindan, M. (2003) 'Seeking Growth, Search Engine Google Acts Like Ad Agency.' *Wall Street Journal* B1 October: 16

McHugh, Josh (January 2003) 'Google vs. Evil.' *Wired Magazine* [online] 11 (1). Available from <http://www.wired.com/wired/archive/11.01/googgle_pr.html> [20 March 2004]

Mitchell, S. (2005) 'Collaboration Enabling Internet Resource Collection-Building Software and Technologies' *Library Trends* 53(4) 604-619

Overture (2003) 'Overture to Acquire Web Search Unit of Fast Search & Transfer—FAST' February: 25 [online] available from <http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=OVER&script=410&layout=0&item_id=385605> [18 April 2003]

Rainie, L., and Hitlin, P. (2005) 'The Internet at School.' Pew Internet and American Life Project August: 2 [online]. Available from http://www.pewinternet.org/PPF/r/163/report_display.asp [20 September 2005]

Reuters (2003) 'Yahoo! to Buy Software Maker.' *Los Angeles Times* Section 3, 2 December: 24

Schachter, K. (2003) 'The Pay-Per-Click Internet Market Grows into Multi-billion Dollar Business.' *Dolan Media Newswires* December: 5

Sherman, C. (2004) 'Microsoft Unveils Its New Search Engine—At Last.' *Search Engine Watch* November: 11 [online]. Available from <http://searchenginewatch.com/searchday/print.php/34711_3434261> [11 February 2005]

Shreves, S. L., Habing, T.G., Hagedorn, K., and Young, J.A. (2005) 'Current Developments and Future Trends for the OAI Protocol for Metadata Harvesting' *Library Trends* 53(4) 576-589

Teather, D. (2003) 'Microsoft Runs Search for a Way to Take Over Internet Giant Google.' *The Guardian* 1 November: 1

Walker, J. (2005) 'Links and Power: The Political Economy of Linking on the Web.' *Library Trends 53*(4) 524-529

Waters, R. (2003) 'Search Industry Scours Scrap Heap for Bargains.' *The Financial Times* 30 March: 12

Downloaded from **EastBound / Journal / 2006 / 1**
http://www.eastbound.info/journal/2006-1/