[Chapter in **Libr@ries: Changing Information Space and Practices**

Editors: Cushla Kapitzke (University of Queensland)

Chip Bruce (University of Illinois Champaign-Urbana)

Publisher: Erlbaum]

Search Engine Anatomy: The industry and its commercial structure

Bettina Fabos

> It is vitally important for business people to understand how search engines work, and how to use them … Search is how your business, whatever it is, will market itself. (The business section of *The Age,* Melbourne, Australia, 2003, November 4)

The topic of search engine enterprise—discussed relentlessly in the business world—has somehow slipped under the radar of librarians, educators, most academics, and the news media (save for the business section). Even as search engines become increasingly reconditioned to serve free enterprise, and now address people as *consumers* rather than *users*, the industry has been particularly good at sustaining four prevalent myths about their services: 1. Search engines are impartial information tools. 2. Search engines search the entire Web, gleaning the most relevant results. 3. Search engines vary greatly, thus offering choice and a competitive marketplace. 4. Search engines are the only place to go for relevant information on the web. Yet, if we do not acknowledge the commercial nature of search engines and believe that search engines are looking after users' best interests, it's easy to fall prey to these myths.

Indeed, these myths can be easily debunked (although the search industry certainly doesn't want that to happen). Search engines are not impartial or reliable. Most of the information they organize is, quite intentionally, commercial. Users wade through veritable strip malls of search engine results, often blaming themselves for not using better searching skills to get better results**.** The truth about search engines is that no knowledge of advanced searching practices (such as Boolean terminology and understanding the subtle differences between individual search engines) will change the main focus of commercial search engines, which is to connect consumers to their advertisers. Moreover, since most "search engines"—which I will later distinguish as "search portals"—are fed by an extremely small number of increasingly consolidating search providers, all of which mimic each others' algorithm strategies, the notion of choice between search engines is almost moot.

What is so unnerving about these four myths is the extent to which users believe them, using search engines as the main gateway to all web information (Griffiths & Brophy, 2005), and the extent to which the search engine industry benefits from this ignorance. Consequently, rather than spending time learning about search engine technology, advanced searching skills, Boolean terminology, the differences

between search engines, or webpage evaluation skills, it is much more important to understand the industry itself—and the economic and political forces that control it—if one wants to be a truly knowledgeable user of the web. This chapter, or "anatomy lesson," is thus an attempt to bring the discussion of the web's evolving commercial structure into the sphere of educators, academics, and others located on the periphery of the information technology business world. That there are alternatives to search engines—as this book clearly attests—is the answer to a more democratic and relevant web. The challenge for all of us, then, is to first understand the way our information tools are shaped by economic forces, and second, to learn that there are alternatives to excessively commercialized web services we have come to rely so heavily upon.

Anatomy lesson 1: Search industry structure

Search engines were once considered a failed business idea because they were only a conduit to other pages. In other words, they lacked stickiness; no one stayed long enough to see the advertising. In response to this crisis, search engine portals tried to develop new services to attract and retain users. For example, AltaVista spent millions to develop new portal content that it hoped would make it a comprehensive web portal for not only searches but other activities such as news, travel, and shopping. Google resisted such efforts, and insisted instead on focusing on being the best-syndicated search engine provider, with the most relevant search results. However, analysts mocked Google for its seeming lack of a means to make money from its singular mission of search excellence.

Then, search engine portals began experimenting with sponsored links—a list of two or three paying sites that appear above the actual search results. Because sponsored links are so highly targeted (they directly relate to the search terms that users type in), they became enormously profitable. A small company dealing with specialized golf equipment, for example, could sponsor a link that accompanied a user's search on golf, directly targeting the golfer. Oftentimes because users didn't know the difference between sponsored and actual searches, they were clicking sponsored links 12 to 17 percent of the time (Waters, 2003) far in excess of the less than one percent banner ad click-through rate today (Harvey, 2003b). And every time a user clicked on a sponsored link, the search engine earned money. Not surprisingly, search engine services barely distinguished between the sponsored and non-sponsored categories in order to generate as many click-throughs as possible.

Let's start with a little dissection. To understand the search engine industry and its gradual and quiet commercialization, it is necessary to distinguish between the four facets of the search engine industry: directories, search engine providers, search engine portals, and *commercial* search providers. The four areas of the search engine industry have their own completely separate functions. Certain companies, like Google, may perform all the functions within a single company; other companies contract out to smaller ones, which each perform a singular functions, and are merged to form the basis of what most people consider to be a "search engine."

Directories

Often mistaken for search engines, directories are nothing more than comparatively small databases that may or may not feed a search engine. Directories predate search engines. Librarians began cataloguing Web sites in the early 1990s, but it was Yahoo! who developed the first commercial directory in 1994 by hiring numerous editors to compile web pages and place them into logical categories. The Directory was steadily growing as individuals, organizations and companies submitted sites for Yahoo! employees to review and (hopefully) add to a category.

As the largest *commercial* directory on the Web, Yahoo! became instantly recognized as the place to look for information. Yahoo!'s initial income came from banner advertising, but also from individuals, organizations, and (mostly) companies who paid Yahoo! for "express submissions" to the Yahoo! Directory (a recurring annual fee of $299 by 2004). Other online directories followed Yahoo!'s model, like the Australian-based directory Looksmart, which launched in 1996. Another significant directory (especially to the search engine industry) is the Open Directory Project, which was launched in 1998. This enormous non-commercial endeavor, constructed entirely by a global army of volunteers (or as they are referred to, "net-citizens") is by far the largest directory on the Web and continues to grow in size every day.

Directories have become an important component to the search industry in that they provide massive lists of indexed pages to feed search engine providers (more on these below). The Open Directory Project is the most significant directory to this end for two reasons: it's the largest directory on the Web, and it has licensed its content for open content distribution. That means that any search engine provider can draw upon the Open Directory Project as it conducts its searches. Search engine providers all claim to have amassed their own proprietary webpage databases through which they conduct their searches, but the Open Directory Project makes up a large portion of each of these databases (Fine Brand, 2003). Yahoo!'s directory has also been an important database that feeds the company's larger search portal, and the MSN portal has long relied on Looksmart's directory to distinguish its search results from other search services (although that relationship ended in 2004).

Search engine providers

Search engine providers are the tools that actually do most (but not all) of the searching. Sometimes referred to as "algorithmic search engines" or "crawler based search engines," search engine providers license highly complicated crawling software to search the web. Search engine provider companies have all got their start in a similar way: Various very smart teams of computer engineers first put their heads together and came up with a unique mathematical formula, or algorithm, for determining webpage relevance. They then got venture capital to amass a database containing billions of web pages, and, through extensive automated text indexing tools, built a "proprietary web index"—the bigger the better. Search engine providers then began to syndicate their service to other search portals (what we commonly think of as "search engines"). Because the job of creating and updating these web search algorithms, and then amassing and managing (i.e., updating, cleansing) these web indexes is so huge, there are actually only a handful of search engine providers in existence. At the time of this writing, the most prominent search engine providers can be counted on one hand: AltaVista, Inktomi, Google, AlltheWeb, and Teoma. The finite number of search engine providers is one reason why search results from different "search engines" are so similar—they are powered by the same provider.

AltaVista and Inktomi are the oldest existing search engine providers, both emerging in 1995.[1] The AltaVista algorithm was developed by scientists at the Digital Equipment Corporation in Palo Alto California, and the Inktomi algorithm was created by a computer science professor and his graduate student at Berkeley. The reason that AltaVista is better known than Inktomi is that AltaVista developed its own branded search portal (while also syndicating its services elsewhere); Inktomi opted to just syndicate, with Yahoo! and MSN being its first major customers. Google arrived in 1998, bringing a new kind of searching standard to the industry. Created by Larry Page and Sergey Brin, two grad students at Stanford, Google's formula became instantly recognized as *better*. While AltaVista and Inktomi had based webpage relevance on keywords (the greater number of

keywords on a page, the more relevant), Google based relevance on links (the number of links that point to a given webpage, the more relevant). Google became the most sought-after search engine provider as a result of its superior algorithm (for example, Yahoo! dumped Inktomi for Google in 2000), and initially made half its revenue from selling its search technology and access to its massive web index, to various web sites (Harvey, 2003a). The search engine provider became an overnight success as a branded search portal as well.

Two more search engine providers emerged after Google. AlltheWeb was created by a group of scientists at the Norwegian University of Science and Technology in 1999, and became hugely popular throughout Europe. Finally, Teoma was born in a Rutgers University computer lab in 2000. There are other, smaller search engine providers, such as Wisenut, and numerous specialty search engines (focusing on medieval or ancient topics, or the Chinese language, for example), but the reality of the small world of search engine providers is that it's getting even smaller. By 2004, three search engine providers, AltaVista, Inktomi, and AlltheWeb had all merged under the Yahoo! umbrella (more details on this below), leaving just three main search engines providers, Google, the Yahoo! group, and Teoma, to power nearly *all* the searches on the web. Therefore, the myth that there are many search engines to choose from, each with varying algorithms, is simply untrue.

Search engine portals

Yahoo!, Excite, MSN, AOL, Google, Ask Jeeves, Lycos—all these are search portals. They are web sites that offer a search toolbar, which is powered by a syndicated search engine provider (most often one of the top five). For example, at the time of this writing, the portals Netscape, AOL, Go, Google, and thousands of smaller, specific-purpose portals (e.g., ESPN.com, Amazon.com, Walmart.com, Mamamedia.com) are all powered by Google. The search portal Lycos is powered by AlltheWeb; MSN, Amazon.com, and eBay are powered by Inktomi; AskJeeves is powered by Teoma, and Hotbot is powered by four search engine providers—Google, AlltheWeb, Inktomi and Teoma. Again, these different relationships explain why some search engine services (e.g., Yahoo! and MSN) have more similar results than others at a given point and time: they are powered by the same search engine provider (Inktomi).

To complicate the matter of partnerships, a few search portals, such as Yahoo! and MSN, integrate directory listings alongside an algorithmic database. In other words, the Yahoo! search portal draws listings from its own Yahoo! Directory in addition to the results provided by Inktomi; MSN, until recently, was drawing listings from the Looksmart Directory and the Inktomi search engine provider. Not surprisingly, whichever search engine provider or directory a search portal decides to partner with, it's a big deal for that company. For example, Yahoo! vaulted Inktomi's profile in 1998 from a relatively unknown search engine provider to a significant player when it chose to outsource its searching technology to the company; then Yahoo! devastated Inktomi when it switched to Google in 2000; In a strategic move, Yahoo! then bought the diminished Inktomi outright in 2003, and now is in the business of syndicating Inktomi's technology.

Which brings us to another complication: some branded portals, such as the new Yahoo!, Google, and AlltheWeb, also syndicate their own search engine provider. Yahoo! and Google are now in intense competition, not only for portal prominence, but for search engine provider prominence. AlltheWeb remains an especially popular search engine portal throughout Europe and has a growing following in the U.S. as a search engine provider (powering the searches, for example, for Lycos). AltaVista, which operated one of the first popular search engine portals (and search engine providers), is now owned by Yahoo!, but it still operates its branded search portal. Its technology has now merged with Yahoo!'s Inktomi, and is syndicated

through Yahoo!'s growing partnership network.

Commercial search providers

Commercial search providers build and manage indexes of web pages (like search engine providers), provide an algorithm for searching these indexes (like search engine providers), and then syndicate this service to search portals (also like search engine providers). The main difference is that the indexes of commercial search providers are all advertisers who pay to be there. To date, every major search portal combines both algorithmic and commercial searches, which are conducted side by side—key words are used to identify both relevant web pages and relevant advertisers. Typically, commercial search results appear in separate locations on the search result list (top, bottom, side) as "sponsored listings" or "featured sites," but many times with as little demarcation as possible between the sponsored and non-sponsored listings so as to inspire more click-throughs. For example, in 2004 AllthyeWeb used a nearly imperceptible thin grey line to separate the first three sponsored links from "the rest"; AltaVista used a large font size, so the first four sponsored listings take up the entire page—a user must scroll down to get to the non-sponsored results. And MSN listed sponsored sites ahead of the regular search results, but employed the same consecutive numbering system. In other words, if there are three sponsored matches, then the first non-paid match begins as number four.

The company GoTo.com (now called Overture) has been instrumental in activating the commercial search business, and now provides commercial results to the majority of search portals on the web. GoTo began as a search portal in its own right, relying on Inktomi for its algorithmic searches and banner advertising for its income. However, in 1998 the company began a new income-generating scheme: it brazenly auctioned off placement within the portal's "impartial" search engine result list itself. The higher an advertiser's bid, the higher the web site appeared in GoTo's search result list; GoTo's search results were effectively stacked with paying customers. Although the move caused considerable controversy among consumer advocates (and within the industry), advertisers were delighted.

High placement within a search result list is important for two reasons. First, users trust this list because they believe it prioritizes relevant web sites according to the key terms entered. Second, users typically don't tend to look beyond the first two or three pages in a search result list, believing that the first two pages are the most relevant (Lasica, 2001) Recalling the four myths I introduced at the beginning of this chapter, search engines are less impartial that people typically think. Another GoTo breakthrough was its "Pay-For-Performance™" strategy, whereby advertisers paid GoTo *only* if a user clicked on a sponsored link. By 2000, the company began to syndicate its commercial search services to other search portals (AOL being its first major customer). Because portals kept a portion of the "Pay-For-Performance" revenue every time a user clicked on one of GoTo's paid placements, search portals were as happy as advertisers. Banner advertising had become increasingly ineffective, and finally there was a way, through search itself, to make money. One Lycos executive justified the practice this way:

> We thought long and hard and decided it doesn't matter if we are paid for a link, so long as the results are what the user wants … the industry has trained users to avoid anything that looks commercial. By calling them paid listings, it hurts the user. (Lasica, 2001, p. 2)

Indeed, the growing justification among internet industry folk was that people generally use the web for commercial purposes anyway. They use the web to find flower delivery services, or to purchase a barbecue grill.

Towards the end of 2000, GoTo renamed itself Overture, disbanded the GoTo search portal, and concentrated solely on its new role as a commercial search provider. By 2002, Overture had signed up 80,000 advertisers (Overture, 2003a) 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. These web sites retained their algorithmic search provider (e.g., Google), but cross-listed this database with Overture's growing index of sponsored web links. Yahoo!, for example, mixed its Yahoo Directory results with Google's algorithmic results and Overture's Pay-For-Performance sponsorship results, although it's not exactly clear whether the sponsored pages were always restricted to the "sponsored" section, or slipped down into the regular web results. Because search portal companies are not obliged to disclose exactly where the commercial influence lies in a given search result list—in the sponsored listings above a search result list (Kopytoff, 2003), or within the result list itself, revealing their mode of user deception has not been common practice. MSN, which had mixed the Looksmart Directory results with Inktomi's algorithmic results and Overture's pay-per-click sponsorship results, has been purportedly following the GoTo.com model by stacking its result pages (Fine Brand, 2003). Before AltaVista became a Yahoo! subsidiary, AltaVista experimented with selling its search results to the highest bidder independently of Overture (Fine Brand, 2003; Hansell, 1999). Back to the four myths, search engines are not impartial.

Regardless of visible above-the-line sponsorship or invisible within-the-list sponsorship, the advent of commercial search providers has meant that a search within a search portal is more and more likely to be heavily commercialized. In a single quarter of 2002, Overture facilitated 563 million "paid introductions" (click throughs) and made $126 million, compared to Google's approximately $15 million in revenue for its main business of running non-paid searches (Overture, 2003b). But Google was not about to stand by and watch. Google responded to Overture's success by amassing its own index of commercial sites and creating the "AdWords" program. AdWords essentially combines Overture's auction system of selling key words and placement to the highest bidder with an algorithm that factors relevance, or the ad's click-through rate, into placement. In addition, Google also established a more expensive sponsored links option for links appearing above, rather than on the side, of the search result list. Significantly, Google took the higher road by promising to never place sponsored links within the company's objective search results. Such company integrity, however, may be irrelevant, or, at worst, a convenient marketing tool. Already, plenty of other channels have evolved to undermine Google's promise of integrity. For this discussion, we need another anatomy lesson.

Anatomy lesson 2: The crafty craft of search engine sponsorship

First, a slight review is in order. As we've seen, search results on any commercial search portal are already skewed towards commerce due to a number of basic payment practices.

- Commercial directories charge a recurring annual fee for an express submission (in Yahoo!'s case, $299), which ensures a continued listing once a submission is approved. The Looksmart Directory, as of 2002, accepts *only* paid submissions from commercial sites to its directory. These enterprising strategies indirectly benefit well-endowed and for-profit sites that can afford elite treatment and positioning within a directory.

- Commercial search providers have instituted pay-for-placement deals designed to directly benefit commercial sites. The higher an organization

bids on a key word, the higher they are listed in the search portal.

Yet these developments are just the beginning of an evolving and commercially innovative search industry.

Paid inclusion

In 2001, Inktomi introduced a new variable that would serve commerce—and the search engine industry—extraordinarily well: paid inclusion. Paid inclusion means that customers who pay a flat fee are *guaranteed to be included* in every search completed by the Inktomi search engine. Search engines do not search the entire web, only parts (here, another myth is deflated). With paid inclusion, paying sites would always be incorporated into the searchable index (unlike many other sites, which simply slip away as algorithms are updated). Although paid inclusion does not guarantee the web site's rank within the search results, it does guarantee inclusion, somewhere, each time a search is conducted. For niche topics especially, this bodes well for the advertiser. As reporter Chris Gaither (2003) explains, "Internet companies have realized that, if someone is hunting for information on a topic like mesothelioma, the person is ripe for specialized advertising" (p. F1).

Inktomi's model was soon copied by every major search engine provider save Google, meaning that by 2002, Inktomi, AlltheWeb, Teoma, and AltaVista—were all offering paid inclusion as part of their overall syndicated package. As the practice of paid inclusion exploded, advice about paid inclusion practices appeared as unproblematic common sense in countless business newsletters and magazines by 2003:

> The key to success is finding the words that will drive traffic to your site—and, more important, convert those potential customers into sales. [Martin Child, Overture's managing director for Northern Europe] says generic terms such as travel may generate a large number of leads but many of those will be wasted if your firm offers only a niche service. "'Travel' may not convert as well as 'Icelandic expedition,'" he says—even though there would be far fewer searches that use the latter, much less expensive term. Bunis recommends a mix of generic and specific descriptions of your business. (Durman, 2003, p. 16)

With such a profitable commercial system, search engine providers were finally seen as money-makers in their own right, and as such came to be regarded more as lucrative properties rather than services for mere syndication. By early 2003, Yahoo! had acquired Inktomi ($235 million) and Overture had scooped up 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). A few months later, Yahoo! then acquired Overture.

As surely as paid inclusion is lucrative 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 so as 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" (Waters, 2003, p. 30). In other words, even if paying sites didn't pay for prominent placement directly, at least they got the tools to figure out how to get there. As advisory material from an online marketing firm called The Web Search Workshop related in 2004, these new "opportunities" could get their clients' web sites "more (and faster) exposure in a crowded market":

> There is some dissatisfaction that these paid services are now opening up a gap between those websites able or willing to pay and therefore changing the balance of search results being offered. However, this trend for providing a paid alternative in return for privileges is likely to stay and probably increase in the future. (The Web, 2004, p. 1)

For good reason, there was heady jubilation within the search industry and in business circles over the Web's commercial viability via search engine listings. Results had become so commercially skewed, however, that consumer advocacy groups were increasingly alarmed. A campaign, initiated by the media watchdog group Commercial Alert, led to a Federal Trade Commission (FTC) investigation into the practice of undisclosed, paid search results within search portals. Completed in June 2002, the study reported (not surprisingly) that the web's largest search engines did not reveal the preferred treatment they accorded to sponsors. The FTC's response was typical in terms of the current political climate: a gentle rebuke and a call for self-regulation. The rebuke however, did *seem* to be effective: it led to more differentiation between sponsored sites and "nonpaid" sites. For example, Yahoo!, AskJeeves, Lycos, and numerous other portals began stressing the "objectivity" of their web results by using bold red headings to demarcate a sponsored link from a non-sponsored link. These distinctions, though, more or less veiled the incursions of paid inclusion, which had surreptitiously become the industry norm.

Search engine marketers (SEMs)

Only one search engine provider/portal has resisted both pay-for-placement and paid inclusion. Google has taken an admirable stance on search engine integrity since its inception, and, although it has rigorously pushed its AdWords program, which matches sponsored sites to key words within a search, the company has widely publicized its refusal to allow any direct commercial influence in its search result lists. This is not to say, however, that Google's result lists are free from market influence. One of the most significant developments just outside of (but directly affecting) the search engine industry has been the rise of search engine marketing. This mini-industry 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. These small SEO companies (or SEMs) try to guarantee prominent listings for their clients. Although most business organizations know the four necessary steps towards Web site visibility (a subscription to Google's AdWords, Overture, Yahoo!'s Directory, and a submission to the Open Directory Project), SEMs aggressively act on their behalf.

Most SEMs are especially focused on Google, a sort of Holy Grail for SEMs. Google is the most popular search portal/search engine provider [conducting 83 percent of all searches in 2003 (Hindman, Tsioustsiouliklis & Johsnon, 2003)], yet a tough nut to crack for SEMs, because it doesn't give anyone access to its algorithm. In fact, one of the most typical promotional statements appearing on these companies' web sites concerns the ability to decode the patterns behind Google's objective search results. "We understand the 'spidering' schedule that Google employs," says Morevisibility.com. "By submitting at the appropriate intervals, we are able to systematically deep-penetrate the Google database" (MoreVisibility, 2003). Meanwhile, as Fiona Harvey of the *Financial Times* has reported, "so many small companies have sprung up in this field that Google engineers spend much of their time tweaking its search criteria in order not to fall prey to them" (2003b, p. 32).

Since 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 have 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." This strategy is called "Google bombing" in the mainstream media, and considered a harmless novelty. Meanwhile, enterprising SEMs (who 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 to better 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. One such firm, the Oklahoma City-based SearchKing, which had practiced link farming and was "punished" by Google, filed a federal lawsuit against Google in 2002. However, Google won the case in 2003 on the grounds that it has First Amendment rights to present its search results in any order it sees fit (Kopytoff, 2003). While the SearchKing vs. Google case pitted one commercial company's interests against another, the case has some interesting implications for any future efforts to de-commercialize search engine result lists. Like the landmark *Midwest Video* case in 1979, which entitled cable companies to pick and choose which channels to carry and escape Federal Communication Commission (FCC) regulation, this case allows search engine providers to escape any rules that would force them to disclose why some content (namely commercial content) appears more heavily concentrated than others. Of course, these developments may bode well for internet consumers (as they are now universally referred to in the search industry), but not so well for internet users with noncommercial tasks.

Because of Google's aggressive actions towards link farms and its win against SearchKing, link-farms have become more risky than worthwhile. 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 about its commitment to search engine integrity. Second, with the demise 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.

Contextual links

As it stands, link farms have never been nearly as effective at influencing Google's ranking system as singular 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. Called "contextual links," they are links to other sites that match 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. As Walker (2005) 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" (p. 2).

Contextual links are, not surprisingly, 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 AdSense program in 2003, while Yahoo! introduced Content Match a month later (Acohido, 2003). Both programs broker contextual links on content Web sites. 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, actually undermines the company's line about search engine result integrity.

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). Again, the myth about 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). These authors have also observed that the Web, via search engines, now operates much like traditional media, with heavily concentrated oligopolies serving as gatekeepers for entry into the rest of the community.

With these developments, both Google and Yahoo! are also becoming more 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 (what the industry now euphemistically terms "customizes") by tracking the clickstream data, cookies, pixel tags, and contact/personally identifying information of search engine users. While Yahoo!/Overture relies on the large, web-based company, Doubleclick for this purpose, Google relies on 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* business 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" (p. B1). Both Yahoo! and Google are also working towards providing successful geo-location functions to their marketing toolkits (Oct. 18). 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 relevant information to users, the task is now for search engines to deliver the most relevant consumer information to consumers (Jan. 6). As business reporter David Crowe (2003) stated in the *Australian Financial Review*, "keeping advertisers happy with their paid searches is now the most important objective for the big search companies" (p. 29).

In five short years, an industry study revealed that the pay-per-click market had grown from under $100 million a year to a multi-billion industry in 2003, and included these startling statistics: 42 percent of those who bought from online retail sites arrived via search engines (Schacther, 2003). So successful was search engine sponsorship that rates increased by 50 percent in 2003 alone, revealing that "the adoption of the internet as an advertising vehicle by traditional advertisers is truly taking place" (Hansell, 2003b). Industry reports also pointed to search engine sponsorship as an increasing, rather than diminishing trend: the business analyst group Bancorp Piper Jaffray projected that the search industry would swell to $7 billion by 2007 (Oct. 11).

Anatomy lesson 3: Google, Yahoo!, and the predatory goals of Microsoft

Three main companies, all American-based, dominate the search industry: Google, Yahoo!, and Microsoft. Google has been involved in search technology from its 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 the most dominant player of all. As CEO Bill Gates has said, "we will catch them," setting the date for the summer of 2005 (Feb. 1, Markoff, 2003).

The Google/Yahoo! rivalry

In 2000, Yahoo! hired Google to conduct the searches on its search portal, giving the company "its first big break" (Hansell, 2003a). By 2003, however, Google and Yahoo! had become the fiercest competitors. To understand the extent of this rivalry, it's helpful to look at the major acquisitions each company made prior, during, and just after 2003:

Oct. 2001	Overture achieves profitability, with "outstanding financial results" Overture's success establishes search as a key business strategy.
May 2001	Inktomi introduces "paid inclusion"
Oct. 2002	Google launches AdWords program, begins head-on competition

	with Overture
Dec. 2002	Google launches the Froogle shopping tool
Mar. 2003	Yahoo! enters the search engine provider business, acquiring Inktomi (and dumping Google)
Apr. 2003	Overture acquires AlltheWeb and AltaVista
Apr. 2003	Google acquires Applied Semantics (an online advertising software co.)
May 2003	Google launches its AdSense program
June 2003	Overture introduces Content Match, directly competing with AdSense
Sept. 2003	Yahoo! expands Yahoo! Shopping (using Inktomi technology), to compete with Froogle
Sept. 2003	Google acquires Kaltix (online profiling and marketing company)
Oct. 2003	Yahoo acquires Overture (which includes AltaVista and AlltheWeb)
Oct. 2003	Yahoo! introduces SmartSort (which helps users quickly narrow a search for certain products)
Oct. 2003	Yahoo! acquires 3721 Network Software (Chinese letter search engine)
Mar 2004	Yahoo! acquires Kelkoo, Europe's leading online comparison shopping service
Mar 2004	Yahoo!/Overture introduces Site Match (a controversial combination of paid inclusion and pay-per-click)
Mar 2004	Google introduces Google Local, allowing advertisers to connect with local users

The end result of this run of acquisitions, which mostly occurred during 2003, meant that two rival search engine companies provide the searching functions (both commercial and algorithmic) to nearly every search portal on the Web. The one remaining search engine provider, Teoma (owned by Ask Jeeves) is considered a ripe prospect for acquisition (Savitz, 2003). Google itself has been preyed upon by the software giant, Microsoft. Indeed, apart from the Google/Yahoo! rivalry, Microsoft emerged as the third dominant player by the end of 2003.

Microsoft makes a move

Microsoft's moves on Google were 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, p. 241-242). Microsoft, in Dawson and Foster's (1998) 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" (p. 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 algorithmic 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 had ditched Looksmart and had 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). In other words, 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 hauntingly familiar 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). During the run-up to Google’s public offering, which finally occurred in August 2004, Microsoft was working hard to destabilize Google and snag Google employees (Markoff, 2004). 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). For starters, the Google Desktop Search program 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. Moral compromise and commercial intrusions into Google’s search listings will certainly continue as Microsoft and Yahoo! continue the onslaught.

Conclusion

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 stealthily undermining internet users’ faith in search technology. 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, the four myths about search engine services resound: they are impartial, they are all-inclusive, they vary greatly, and they are the most reliable place to go for relevant online information.

We educators, for the most part, buy into these myths; we continue to have a largely optimistic outlook on search engines as helpful and trustworthy educational and research tools. Despite its heavy commercialism, the web’s potential as a place for online scholarship and diversity is still evident—when the right search terms are used—and positive experiences can certainly outweigh negative experiences, for now, anyway. Moreover, educators and librarians have heavily promoted individual skills (advanced searching techniques, web page evaluation skills) as a way to cope with excessive commercialism. Although it may feel empowering to teach or possess these skills, a wholesale critique of the commercial web structure remains sidelined (Fabos, 2004).

Fortunately, researchers in economics, political science, communication studies, applied physics and computer science have begun to scrutinize the structural components of the web against the framework of democracy and knowledge access. Economists Lucas D. Introna and Helen Nissenbaum (2000), for example, have examined search engines’ current trajectory as market-driven information tools and have questioned the future of the web as a public good. They argue “if search engines systematically highlight Web sites with popular appeal and mainstream commercial purpose, as well as Web sites backed by entrenched economic powers, they amplify these presences on the Web at the expense of others” (p. 28).

Hindman et al (2003), also economists, are also concerned with the increasing concentration (and visibility) of heavily linked sites. They have analyzed the hyperlinks surrounding political Web sites (measuring link structure on a massive scale), and have concluded that the number of highly visible sites is small, and the visibility drop-off is rapid. They have called this organizational structure "googlearchy" (the rule of the most heavily linked), and suggest that the Web is more like the current corporate mass media, with top-down control of limited content, rather than a diverse platform of ideas.

Another computer scientist, Susan L. Gerhart (2004), has researched a residual effect of web site consolidation: the way search engines inherently suppress controversy. In her study, which tracked search engine results about five controversial topics, Gerhart found that three out of five were buried by commercial sites, and concluded that significant amounts of controversial material are unattainable via search engines. In her words, controversy is important because:

> … controversies often express the richness and depth of a topic. Controversies dramatize change. Controversies may make a critical difference in life-altering decisions. Scientists, journalists, and intelligence analysts are professionally required to address multiple perspectives, facts, authorities, and opinions on topics. Search engines may significantly decrease their productivity or conceal incompetence if controversies are overly difficult to investigate. (Gerhart, 2004, p. 4)

According to Gerhart, unless a controversy has already been recently highlighted in the mainstream media, or unless a searcher is already aware that a controversy about a particular topic already exists (and can type the appropriate keywords), controversial topics are largely invisible. Computer scientist Jill Walker (2005), another scholar researching the implications of a link-driven online economy, eloquently summarizes the mainstreaming of Web content this way:

> We are participants in this power structure whether we like it or not. We can criticize it, reflect upon it, approve of it or try to subvert it. We must not ignore it. This standardization of links and their value will shape what the future finds. It defines what can be found. It defines knowledge. (p. 3)

Indeed, all of this research is important, because it goes beyond the net result of search engine result lists (what librarians and most educators are typically concerned with) and addresses the complex and economically-charged structure of the web that affects all search results regardless of how well one crafts an individual search.

As people habitually turn to commercial search engines to navigate a commercial environment, they are unaware of the increasing difficulties to locate content that is *not* commercial. They are unaware of the misleading motives of the internet navigation tools they use, and of the constant efforts among for-profit enterprise to bend the internet toward their ends. Communications scholar Robert McChesney (1999), who has been on the forefront of discussions about media concentration, is another researcher who has touched upon the internet as the latest in a long line of commercial media that have been colonized by corporate power, turning a so-called democratic medium into one wholly controlled by big media oligopolies. "Despite its much ballyhooed 'openness,'" he writes, "to the extent that it becomes a viable mass medium, it will likely be dominated by the usual corporate suspects" (p. 183).

Today, the internet is an oligopoly, with Google, Yahoo! and Microsoft as the three

companies controlling most of the internet's information flow. These companies are capable of bringing good things to the web search environment, and will do their best to appear as if they put the interests of users first, rolling out new plans to make search experiences more worthwhile and exposing "the invisible web." For example, Google has begun to digitize all the books in the Stanford University library system that date before 1923 (those that are in the public domain); the company will include this wealth of information in its database. In March, 2004, Yahoo! initiated its "Content Acquisition Program," part of which involves content partnerships with organizations like the Library of Congress, National Public Radio, The New York Public Library (owner of Project Gutenberg), the National Science Digital Library, and Michigan's OAIster digital archive). These considerable archives will supply "premiere" (that is, educational) listings on the Yahoo! search result list. And as Microsoft's search engine becomes a significant player, the company will more than likely make well-publicized efforts to improve its database with rich, non-commercial archives. As with television, there has to be some decent content to drive the ads—that's the basis of how commercial mass media operates in the United States.

Regardless of these seemingly good intentions, however, the myths surrounding search engines must be debunked. Commercial search engines will always be educationally compromised; they will always give advertisers premium treatment in their services. The other part of Yahoo!'s "Content Acquisition Program," for example, involves Site Match, a more vigorous paid inclusion program refreshed every 48 hours, which allows web-heavy companies more control over which of their pages are frequently indexed. The creation of Yahoo!'s Site Match also marks a change in the paid inclusion payment model. Now, a finder's fee is involved beyond the original flat fee: any time a user clicks on a commercial link, whether it's on page 1 or page 20 on the search results page, advertisers pay the search engine a small listing fee—naturally giving search engines more incentive to get users into the "click-through" mentality.

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. We need to discredit the final myth: that search engines are the only place to go for relevant online information. Some places to start are the non-profit world of subject gateways (e.g., The National Science Digital Library, INFOMINE, OAIster, or the Resource Discovery Network), issue networking sites such as Govcom.org, or the collaborative environments of wiki pages (e.g., wikipedia.org; disinfopedia.org), where people, not corporations, make the significant choices about the information—controversial and diverse—that is accessible online.

NOTES

[1] Both Lycos and WebCrawler, which emerged in 1994, pre-date AltaVista and Inktomi, but stopped developing their algorithms in favor of outsourcing search services from AlltheWEb (Lycos) or multiple search engine providers (Webcrawler).

References

Acohido, B. (2003, November 17). Simplifying the info hunt. *USA Today,* 11E.

Bazeley, M. (2004, October 15). Google rolls our search software. *San Jose Mercury News*, p. B2.
Conners, E. (2003, November 25). Microsoft in search of top spot. *Australian Financial Review*, p. 32.
Crowe, D. (2003, October 18). In search of a profit engine. *Australian Financial Review*, 29.
Dawson, M. and Foster, J. B. (1998). Virtual capitalism. In R. McChesney, E. M. Wood & J.B. Foster (Eds.), *Capitalism and the information age: The political economy of the global communication revolution* (pp. 51-68). New York: Monthly Review Press.
Durman, P. (2003, Nov. 9). Customers clicking with online adverts. *Sunday Times* (London), 16.
Fabos, B. (2004). *Wrong turn on the information superhighway: Education and the commercialization of the internet*. New York: Teachers College Press.
Fine Brand (2003). A brief history of search engine marketing on the Web. [Online.] Availailable: http://www.Fine Brand.com/ideacenter/search-engine-optimization/ abriefhistory-2.cfm [2004, March 7].
Gaither, Chris (2003, February 23). Searching for dollars as banner market flags, Net giants pin hopes on revenue from targeted ads related to specific queries. *The Boston Globe*, F1.
Gates, Bill (1995). *The road ahead*. New York: Viking.
Gerhart, Susan L. (2004, January). Do Web search engines suppress controversy? *First Monday*, 9(1). [Online]. http://firstmonday.org/issues/issue9_1gerhart/index.html. (2004, January 23).
Goodman, A. (2003). Cracking Google's new algorithm. *WebProNews* [Online]. Available:http://www.webpronews.com/ebusiness/smallbusiness/wpn-2-20031203CrackingGooglesNewAlgorithm.html
Griffiths, J. T. and Brophy, P. (2005, Spring). Student Searching Behaviour and the Web: Use of Academic Resources and Google. *Library Trends*, 53(4), pp. XX-XX.
Hansell, S. (1999, April 15). Altavista invites advertisers to pay for top ranking. *New York Times*, C2.
Hansell, S. (2003a, April 7). Yahoo plans improvements in effort to regain lost ground. New York Times, 2.
Hansell, S. (2003b, October 9). Yahoo's resurgent profit is led by strong advertising. New York Times, C7.
Harvey, Fiona (2003a, January 15). Online ads await click-through to profitability: Questions have been raised about whether the medium works, but analysts are optimistic. *Financial Times*, FT Report, 1.
Harvey, Fiona (2003b, March 4). Putting a price on results: Recent acquisitions suggest the sponsored search sector is a bright spot in a gloomy market. *The Financial Times*, 32.
Hindman, M, Tsioutsiouliklis, K, and Johnson, J. (2003). "Googlearchy": How a few heavily-linked sites dominate politics on the Web. [Online]. Available: www.princeton.edu/~mhindman/googlearchy--hindman.pdf [2004, 10 April].
Introna, L. and Nissenbaum, H. (2000). Shaping the web: Why the politics of search engines matters. *Information Society, 16*(3): 189-186. [Online]. Available at: http://www-us.ebsco.com/online/Reader.asp (if your library is subscribed).

Kopytoff, V. (2003, November 9). Google's future looking good; Bulked-up search firm beats up rivals. *San Francisco Chronicle*, I1.
Lasica, J.D. (2001, July 23). Search engines and editorial integrity. *Online Journalism Review* [Online]. Available: http://www.ojr.org/ojr/technology/1017778969.php. [2001, December 12].
Mangalindan, M. (2003, October 16). Seeking growth, search engine Google acts like ad agency. *Wall Street Journal*, p. B1.

Markoff, J. (2003, October 31). Microsoft and Google: Partners or rivals? *The New York Times*, C1.
Markoff, J. (2004, February 1). The coming search wars. *The New York Times*, section 3, 1.
McChesney, R. (1999). *Rich media, poor democracy: Communication politics in dubious times*. Urbana: University of Illinois Press.
McHugh, Josh (2003, January). Google vs. Evil. *Wired Magazine, 11*(1) [Online]. Available: http://www.wired.com/wired/archive/11.01/googgle_pr.html. [2004, March 20].
MoreVisibility platform three: E-commerce initiative (2003). More visibility.com [Online]. Available: http://www.morevisibility.com/serices_three.html. [2003, April 18].
Overture (2003a, February 25). Overture to acquire web search unit of Fast Search & Transfer—FAST. Press release; Overture, Inc. [Online]. Available: http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=OVER&script=410&layout=0&item_id=385605 [2003, April 18].
Overture (2003b). Vision: Corporate overview. Overture, Inc. [Online]. Available: http://www.content.overture.com/d/Usm/about/company/vision.jhtml [2003, May 25].
Reuters (2003, December 24). Yahoo! to buy software maker. *Los Angeles Times*, Section 3, 2.
Savitz, E. (2003, October 30). Why are Jeeves shares so high? *The Toronto Star*, K07.
Schachter, K. (2003, December 5). The pay-per-click Internet market grows into multi-billion dollar business. *Dolan Media Newswires*.
Teather, D. (2003, November 1). Microsoft runs search for a way to take over internet giant Google. *The Guardian*, 1.
The Web Search WorkShop (2004). Materials: Paying for search engine and directory listings. [Online]. Available: htttp://www.websearchworkshop.co.uk/paying.htm. [2004, March 21].
Walker, Jill (2005, Spring). Links and power: The political economy of linking on the web. *Library Trends, 53*(4), pp. XX.
Waters, Richard (2003, March 12). Search industry scours scrap heap for bargains. *The Financial Times*, 30.