

INTO A GOOGLE WORLD: RETHINKING UBIQUITY

by Peter McDonald

At the recent annual conference of ALA (Orlando 2004), a number of cheery librarians proudly sported T-shirts emblazoned with the slogan: "Libraries – Google without the Garbage!" Well, they got it partly right. Because it is a sad fact that Google (and the other search portals by extension) even without their garbage fail mostly to reveal much of the richness, the diversity, indeed the culturally magnificent, socially interactive and physically welcoming quiddity that represents even our most humble rural libraries, even the one-room in Sully County, South Dakota, still less the majesty of our major research and big city library systems where the cultural history of our national heritage resides.

Google-like portals do however have their place. As the ubiquitous search engines of choice the world over, they have grafted themselves onto the information seeking habits of most people and are likely to remain central to our Internet habits. I use Google myself. It is easy, efficient, and more often than not, retrieves enough "useful" information on any given topic to make all of us, myself included, Google junkies of a sort. Google is purported to handle 200 million searches a day. Indeed 95% of all Web searches are handled by two portals, Google and Yahoo, with the likes of Ask Jeeves and AltaVista a distant third. For quick fact-finding these search engines are excellent. But when we consider the possibility of these commercial portals somehow replacing the average citizen's trip to the library, that surely is troubling possibility.

On June 21st, 2004, *The New York Times* carried a front page under-the-fold article titled: "Old Search Engine, the Library, Tries to Fit Into a Google World," by Katie Hafner. Hafner's main thesis is that libraries everywhere are under assault as researchers of every stripe, from high school sophomores, to Ivy faculty, are turning not to libraries for authoritative sources of information but to online resources, and chief among those online sources are Google-like search engines. Says Hafner: "While the accuracy of online information is notoriously uneven, the ubiquity of the Web means that a trip to the stacks is no longer the way most...research begins." And she elaborates: "The biggest problem is that search engines like Google skim only the thinnest layers of information that has been digitized."

Hafner's point is well taken. Neither Google nor Yahoo search on proper meta-data tags, still less on Dublin Core standards of indexing, the current library norm. Indeed these portals seem positively obdurate in their unwillingness to adopt Open Access standards of indexing that would reveal at least some of the "deep Web" to the average searcher. Sadly, the majority of the deep Web, which comprises tens of millions of pages of primary material digitized by universities, special collections, research centers, historical societies and scholars, is formidably invisible to the average Google neophyte typing away in her search-box. More to the point, commercial search engines have seemed equally reluctant to incorporate some sort of Z39.50 compliance standard that might allow a user to search through local online library catalogs. While this article was being written, OCLC announced that it would offer its entire collection of 53.3 million items connected to 928.6 million library holdings for "harvesting" by Google and Yahoo! And Google has launched a beta version of Google-Scholar intended to harvest just these sorts of resources. It's unclear at this time whether the initiative began with OCLC or the search engine community, but regardless, if implemented, this is good news for Web surfers everywhere. It's unclear at this time whether the initiative began with OCLC or the search engine community, but regardless, if implemented, this is good news for Web surfers everywhere.

There is also this to consider, that both Google and to a lesser extent Yahoo preside over what amounts to little more than an online popularity contest, arguably the biggest in history. Google's innovation and success as a search engine rests on the fact that it's underlying technology gathers data at lightning speed on the number of links between pages, rather than the actual pertinence of content requested. So a Google search becomes by default a sort of popularity vote, where the number of links far out-weigh content in retrieval algorithms. Matthew Hindman and Kenneth Cukier, joint fellows at Harvard's Kennedy School, wrote in an insightful op-ed in the *New York Times*:

Behind Google's complex ranking system is a simple idea: each link to a page should be considered a vote, and the pages with the most votes should be ranked first.... But what is good for Google is not necessarily good for the rest of the Web. The company's technology is so strong that its competitors have adopted a similar approach to organizing online information, which means they now return similar search results. Thus popular sites become ever more popular, while obscure sites recede ever further into the ether. (*NYT* 8/23/04)

And there's the rub, it's a winner-take-all system. (See also: "How Google Works" *The Economist*, v. 372, no. 8393, Sept. 18th 2004 pp.32-34) In a corporate dominated society such as ours, where the voice of the marginalized and the revolutionary and the anti-establishmentarian go largely unheard, Google has the potential to aid the "do it yourself" web progressive with a message but meager resources to broadcast it. Web-casting becomes

the optimal vehicle for dissemination. But Google won't get you there if the link path is not already highly traveled to begin with, which may be problematic on sites such as these. In this respect, Google is not much different than other major media outlets, residing smugly in the middle of a disquieting trend toward consolidation of information, managed by the few at the expense of the many, where diverse radical voices are often stifled. True, it is Google's technology that makes it so, rather than its editorial or political philosophy, but it nevertheless wields immense power in the world of information management. In that capacity, it should be viewed with rightful caution as an impartial information broker.

But there is room, I'd argue, for wider and closer collaboration between libraries and commercial search engines where our joint interests should intersect. There has been some movement in this direction. Google, for one, has experimented with the term "searchable online database" with quotation marks in its search string, which may or may not retrieve "deep Web" material particularly well, but it is a start. All the same, one is left to wonder why they shy away from indexing more thoroughly. Why not mine our rich library collections now available online? Obviously this does not apply to licensed databases, which are only accessible through login or IP authentication.

On the flip side, it is also worth noting that to date, Google and its clones, to their credit, have stood firm with libraries against filtering, despite the hysteria surrounding CIPA and other anti-pornography rulings. Type "gay sex pics" in Google and there they are, a feat even a six-year-old with some typing and spelling skills can do. But Yahoo has admitted filtering some bestiality sites, and just a few years ago bowed to Christian Right pressure to block several gay sites (see below). As is common with the corporate world, their bedrock commitment to "free speech" is uneven.

Somewhat more troubling, most of the major commercial search engines (Google claims otherwise) have been shown to accept corporate funds to place paying sites at the top of hot topic search retrievals. This is known in the business as "paid placement". How are we to know which ones are which? Type in "insomnia" and doubtless the top ten sites retrieved will try to sell you a sleeping pill to alleviate it. Well fact is, we won't know which sites paid for placement and which didn't. Is that top site retrieved really the most authoritative, or just the site with the deepest pocket? (See for example: "Search engines charged with false advertising" Shimm, Suzy. *The Local Planet*. Spokane, WA.: Aug 1, 2001 v.2/30, p.9) As for the brouhaha over paid placements, the most quoted refrain in reply is that being commercial entities themselves, these search engines have a right to turn a buck too on the free enterprise flywheel, who are we to gainsay them?

In short we never know what we are getting from a random search on a commercial portal. By way of example, type in "AIDS" in any search engine and you will likely retrieve over 20+ million hits – can this truly be useful? Is any of it authoritative? We have no way of knowing. But for

a person with AIDS searching at a public terminal at the branch library in Brownsville TX, say, these search results may well be worthless in practical terms. The number of hits is simply too large to sift through meaningfully especially given how many may indeed be "garbage." In other words, if information literacy isn't your long suit, how can you tell a good site from a bum one? (See for example: "Dotcom Doctors: Both useful medical advice and potentially dangerous misinformation are available on the Internet." Strickland, Eliza. *Gambit Weekly*. New Orleans, LA: Feb 03, 2004. v.25/5 p.39)

But the *NYT*'s Hafner, an internet booster of long standing, has it wrong to suggest the "biggest" problem of these search engines is their poor indexing. Far more troubling is the fact that until recently, actually until August 2004 to be exact, Google was a privately owned corporation, beholden to no one except its owners. True it has largely been free of scandal, and Google is generally conceded to be reasonably democratic, though voices differ regarding its Google.news service. (See for example: "News at the Speed of Google" *Columbus Alive*. Columbus, OH: Jan 16, 2003. v.20/2 p.3) Conversely, library activist Chuck Munson notes that Google.news banned San Francisco IndyMedia over an Isreal/Palestine controversy, when the indy site used the term "zionazi" in an online article. The term was offensive to many, and Google blocked the site in March of 2004 after numerous complaints were lodged. However, as Munson notes, Goggle does not block sites that use derogatory terms for Arabs such as "towel-head" "raghead" or "Islamofascist". (See: Infoshop.Org: <http://www.info-shop.org/inews/stories.php?story=03/09/04/0267960>)

Be this as it may, like so many other corporate firms, in late 2003, Google followed the industry norm, and opened its latest research and development facility in Bangalore. Bangalore of course is India's Silicon Valley, where decent programmers come at a fraction of the cost of counterparts in San Jose, California. To date, no lay-offs at the Google headquarters at Mountain View, California, have been announced but the trend is obvious. But now that it is beholden to stock holders and their demand for return on investment, the future at Google is predictable. As the joke says: What do you call a Silicon Valley programmer? Waiter!

Further, what are we to make of the recent Google IPO offering? As one stock site glowed: "Seldom has a financial event been so eagerly awaited by the investing world... [I]t will be the culmination of years of anticipation by stock market participants." But the dot.bomb explosions of 1999/2000 doubtless still reverberate despite the giddy projections for Google's future. In typical Google fashion, the company did not follow normal IPO protocols when it rolled out its public offering in August 2004, but instead used an internet auction to set its share price, bypassing the usual Wall St. investment houses altogether. But Google did warn that, once listed on NASDAQ, its initial share price might plunge once the hoopla subsided. But should we not ask, amazing as Google is at searching the Web, what in fact is it offering? Where's its marketable product? Advertising is at a

minimum. It sells nothing itself. It's search engine is free. One is left to wonder, why is it worth billions, \$27 billion in fact at last count?

Publicly-traded Yahoo and other similarly traded search sites have corporate shortcomings too of course. In 2002, Yahoo was accused of unfair labor practices when it supported the firing of janitors who tried to unionize Team Services Inc., a company that hires cheap migrant workers to scrub Silicon Valley's toilets, Yahoo's included. As noted, Yahoo has also disabled sites it was hosting on a number of occasions because they were deemed "objectionable" solely because of content. Gay sites have been particularly targeted, such as the Guerrilla Queer Bar in Los Angeles, which the Christian Right had called "amoral." Yahoo blocked the site, only to turn it back on when the outcry from the gay community grew louder than the yelps from the Christian Right. And in the spirit of Enronesque greed-head behavior, Yahoo Inc. Chief Executive Terry Semel exercised options to buy 2 million shares of his company's stock this past July (2004), then immediately sold all of it for a profit of more than \$42.8 million. Same as it ever was.

But there is another more troubling fact that should be of true concern to libraries, more alarming to my mind than the stock-cashing behavior of Yahoo's CEO. Far back as 2001, in the wake of the shameful passage of the USA PATRIOT Act by our supine Congress, the Electronic Frontier Foundation warned:

Be careful what you put in that Google search. The government may now spy on web surfing of innocent Americans, including terms entered into search engines, by merely telling a judge anywhere in the U.S. that the spying could lead to information that is "relevant" to an ongoing criminal investigation. The person spied on does not have to be the target of the investigation. This application must be granted and the government is not obligated to report to the court or tell the person spied upon what it has done. [EFF 10/31/01 post on: <http://www.eff.org/>]

The ACLU has seconded this word of caution, and has stated publicly that Google records can be summarily subpoenaed by federal authorities under provisions of the USA PATRIOT Act. Indeed, several reports in the alternative press revealed that Google had itself agreed in principle to abide by such rulings if and when they might be asked to turn over their (actually our) search records.

If this weren't bad enough, countless hundreds of library home pages, from public libraries to those at private colleges, use the Google search box as their search engine of choice. What a Trojan Horse have we created here? We tout our staunch resistance to the USA PATRIOT Act as champions of civil liberties on the one hand, then give away the keys to the citadel by letting Google manage our patrons' search habits. Ashcroft doesn't have to

send some bumbling flatfoot to the nearest library of scrutiny, his investigators can just drop by Google's headquarters and ask for the records of any library using the Google search box, and Google will doubtless comply. (See IndyMedia@UK: <http://www.indymedia.org.uk/en/2003/10/278746.html>) But in some respects Google and its clones are merely the logical tip of the digital iceberg. It is no news to anyone reading the *Progressive Librarian* over these past years, that the civic commons our public and academic libraries represent, is being eroded in any number of ways by the onslaught of our online addictions. Certainly the gate counts at academic libraries have been going down steadily since the early 1990s, when access to electronic resources became ubiquitous at a distance. Don't *walk* in, *modem* in! This is less true of public libraries, but in some respects the challenge to reclaim the physical library as the civic commons of our society remains just as pressing. One might well ask if the steady foot traffic heading to our public libraries is largely due to the fact that there are free internet service available. Is this really a public good? The point is surely debatable. But in an age when many studies show whitecollar workers stare at their monitors upwards of six hours a day, doubtless more than they engage the members of their own families, (it is certainly true in our family where face time is at meals), this rush to the keyboards is a trend worth examining with a critical eye.

Hard to say definitively where the Googlization of the world is leading us as citizens of an information saturated world. Doubtless it is a two-edged sword. Indeed this dichotomy should be the fodder for any number of master's theses at our library schools. But where are the articles in the peer-reviewed press that ask critical questions on this topic? Many of the few citations on this topic in EBSCO Online's Information Science Abstracts simply posit ways for libraries to mimic Google's success but only to do it better. What are we to make of this? The paucity of critical analysis in these peer-reviewed articles seems to suggest a singular unwillingness among the status quo to really question where Google mimicry will lead the library profession as a whole. I would argue that this is a race librarianship will never win. Nor, perhaps, want to win.

Instead, in defense of libraries, let us ask: Where else but in our nation's libraries do people gather freely with the sole purpose to educate themselves, entertain their minds, enrich their lives, and explore new horizons with willing guides at their elbows, namely exceptional reference professionals? Where else but in our physical libraries does serendipity, browsing, and discovery play such a key role in the betterment of people's lives? Where else but in our libraries, and it certainly is not the mall, is the true civic commons of our society? I would posit nowhere. No matter how gussied-up Google and its cohorts get, they will never attain to these lofty achievements. They are a pale simulacrum of what began in 1731 in Philadelphia as the first public library in America, whose motto was: "To pour forth benefits for the common good is divine." Libraries to this day hold that truth to be self-evident. The Googles of the world do not hold a candle to this rush light.