[Faculty Publications, UNL Libraries](#)   [Libraries at University of Nebraska-Lincoln](#)

7-1-2006

Want to See the Sites? Better Find a Better Guide: Do Popular Search Engines Return Librarian-Recommended Sites?

David C. Tyler
University of Nebraska - Lincoln, dtyler2@unl.edu

Scott Childers
University of Nebraska-Lincoln, scott.childers.sels@gmail.com

Beth McNeil
University of Nebraska-Lincoln, memcneil@purdue.edu

Alicia Dostal
University of Nebraska - Lincoln

Want to See the Sites?
Better Find a Better Guide:
Do Popular Search Engines Return
Librarian-Recommended Sites?

David C. Tyler
Scott Childers
Beth McNeil
Alicia Dostal

ABSTRACT. This paper presents the results of a study of the utility of several popular search engines and of two newer search engines with respect to librarian-selected lists of Web resources and Internet searching behaviors. This study addresses whether said resources are returned where Internet searchers could reasonably be expected to find them and whether the search engines employed serve as acceptable substitutes for the expert advice of librarians. Search engines included in the study were Google, MSN.com, Yahoo, Lycos, AskJeeves, Icerocket, and Acoona. Searches for the study were based on the topics/titles of the "Internet Resources" columns from *College & Research Libraries News* for 2004. Finally, the paper addresses methodological concerns and proposes possible directions for further research. doi:10.1300/J136v11n03_06 *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH.*

David C. Tyler is Reference and Liaison Librarian, Sociology and Anthropology (E-mail: dtyler1@unl.edu); Scott Childers is Assistant Systems Librarian (E-mail: schilders1@unl.edu); Beth McNeil is Associate Dean (E-mail: mmcneil1@unl.edu); and Alicia Dostal is Student Research Assistant, all at 318 Love Library, University of Nebraska-Lincoln, Lincoln, NE 68588-4100.

KEYWORDS. Web sites, search engines, user behavior, searching behavior

Throughout modern civilization, people have wanted to make the world of information available to them manageable and useful, and there have been any number of efforts to fit this world of information into an organization and delivery system that would put the best information readily to hand when it is needed.[1] In the current climate, librarians and information scientists make great efforts as reference providers, resource reviewers, and catalogers to provide direction for, organization of, and a conceptual map to what philosopher Karl Popper called the third world, "The world of the products of the human mind" (Popper 1972, 1978). However, we appear to be nowhere near achieving complete success. In the interim, information seekers on the Web have had to make do with search engines, delivery systems which seem often to be their first choice when searching for information. Librarians and libraries are sometimes a third or fourth option, when they are an option at all (Weiler 2005; Chelton and Cool 2004).

Since most information seekers seem to want to turn to the Web as a resource and to search engines as a convenient organization and delivery vehicle, what we wondered was whether, where the credibility and manageability of information resources are concerned, Internet searchers can use and depend upon search engines as an acceptable surrogate for knowledgeable librarians and their expertise. We wondered whether there might not be some sort of beneficially symbiotic if inchoate or unexplained relationship to be found between search engines and librarians and whether search engines might, despite the vast number of sites and pages that they index, provide ready access for their searchers to the sorts of sites that librarians would recommend to their patrons.

To this end, as an initial and exploratory foray into this topic, we decided to conduct a series of searches for librarian-recommended sites and pages using the five search engines–Google, Yahoo, AskJeeves, MSN, and Lycos–named most popular in a survey of 2000 web users conducted by Keynote Systems, Inc, and reported in *The Wall Street Journal* (Delaney 2005) plus two newer engines, Acoona and Ice-Rocket. These engines were included because at the time of the study

they were both receiving a certain amount of attention from the press as new and innovative products, so we included them because we were interested in seeing whether they would show a marked improvement over the older and more established five engines. Of course, these search engines are not the only popular search engines, or necessarily the best, but they do appear to be frequently used by the general population.

METHODOLOGY

In order to conduct this informal exploration, we gathered together eleven Internet resource lists from the 2004 issues of *College & Research Libraries News* and, based upon the content of each list, came up with keyword searches that we thought an apt, though not necessarily expert, web searcher might employ when searching for the sort of sites listed. In building these keyword searches we were influenced by our personal experiences with grading the "web searching" exercise offered by our library instruction course and patterned our searches after the more successful answers provided by students. It should be understood, from recent research, that the searches below are reflective of a slightly above-average grasp of how to use the Internet as a search tool. Apparently, the majority of searchers use just one or two words to compose their searches. For example, a study by the Nielsen Norman group in 2004 reported that six in ten web searchers typed in only one word in order to find what they were looking for; people using one or two words comprised four-fifths of all searchers; just 3% of web searchers surveyed placed quotations around words in order to conduct phrase searches; and a scant 1% used any advanced search techniques at all (Ward 2004). These behavior patterns and numbers are generally consistent with those of similar studies of both Web searching and the searching of library OPACs (Spink et al. 1999; Allison and Childers 2002; Novotny 2004; Shenton and Dixon 2004). The reader ought to keep in mind that the resources from the *C&RL News* lists are obviously not the only librarian-recommended resources available for these topics on the web. The lists were chosen for use in this study for two reasons: they were already extant at the time of the study, and they list sites and pages recommended by librarians. The reader also should note that this is not a user study of Web searchers behavior, but a study of how search engines perform when presented with the searches of a hypothetical and above-average searcher.

The topics for the eleven resource lists and our hypothetical keyword searches appear in Table 1. The July/August Web resource list had three separate search strings because the list had three distinct and discrete subsections.

In the late spring of 2005, searches for each of the topics, using the predetermined key words based on the article topic, were completed in each of the search engines. Because of the peculiarities of the several search engines not all of the searches were executed exactly as displayed in Table 1. In some instances multiple searches involving different arrangements of the search terms were employed, and the results from what seemed the most effective search were used. Searches for each topic area were completed in a particular search engine before moving on to the next keyword search. After having conducted the searches we attempted to determine how many of the recommended sites and pages were returned within the first one hundred results for the

TABLE 1. Resource List Title, Number of Sites, and Keyword Search

2004 *C&RL News* issue	Resource List Title	Number of sites	Keyword Search
January	Indigenous nations: Sites of interest	33	indigenous and (peoples or nations)
February	Knowledge management: A guide to resources on the Internet	28	knowledge management
March	Gray literature: Resources for locating unpublished research	35	(gray or grey) and (literature or research)
April	Book arts on the Web: An introduction to selected resources	33	(book or books) and (handmade or arts or artists)
May	Asssessing student learning: Available resources	33	assessing and student and learning
June	Success at every stage: Web sites for a career in academic librarianship	24	academic and (librarian or librarianship) and (career or careers)
July/August	The electoral college, political parties, and elections: Sites to help you through the voting process section: The electoral college section: Political parties section: United States 2004 Presidential Election	21	electoral college (6 sites) political and (party or parties) and united states (9 sites) 2004 and presidential election and united states (6 sites)
September	The Civil Rights Movement: Sites for students and researchers	23	civil and rights and movement
October	The Middle Eastern World: Starting points for research	30	middle and (east or eastern)
November	Alternative assessment in higher education: Web sites for a learner-centered approach	33	alternative and assessment and higher education
December	U.S. military and defense studies: Online resources	21	united states and (defense or military)
TOTAL (N)		314	

search engine. Our assumption in looking at one hundred results was that even the more determined web searcher would begin to be frustrated by that point. Again, we caution the reader to keep in mind that our efforts here were exceptional rather than the norm. In connection with the aforementioned Nielsen Norman Group survey, Dr. Jakob Nielsen remarked, "If it is beyond the first page, it is as if it did not exist" (Ward 2004); and a recent report issued by the search marketing firms Enquiro and Did-it, and by Eyetools, a software company which provides tools and services to measure eye movement as people look at web pages, suggests that there is a small "golden triangle" of search results wherein most searchers scan just the top three results of a search in an F-shaped pattern, with just a few glancing at the fourth or fifth result. According to this report, rather astoundingly, there is just a 50% chance that searchers will bother to look beyond the fifth returned item in a search (Sherman 2005). We then calculated how many sites were returned per hit within these top 100 results in order to determine how the ratio of noise to worthwhile content returned within the results that a determined searcher would be likely to browse. We also recorded how many total results were returned for each search as a sort of measure of the absolute noise each search engine generates. The results for our exploratory study are presented in the next section.

RESULTS

The results of this study should be considered a sort of snapshot of what the search tools returned at the time of the project, the results over a fixed interval for a particular population, not a sample from which one could make gross statistical inferences. What the search engines index is of course extremely volatile on a daily basis, with sites constantly being added in and other sites disappearing due to link rot (Tyler and McNeil 2003). The engines relevancy ranking algorithms may be periodically changed to prevent certain types of spamming and to prevent unscrupulous manipulation of their systems. The company providing the engine may decide to change its focus. For example, since our data collection period IceRocket has changed its focus in order to emphasize their blog crawled content and become a "tracking engine," not a general search tool (Cuban 2005). They are still, however, providing Web searching services and so we have elected to include their results.

Undoubtedly, the first question to be answered would be how many of the selected URLs did each of the search engines return within their

top one hundred results. For the searches mentioned in Table 1, the engines returned the following results as shown in Table 2.

As Table 2 indicates, the various search engines returned somewhere between 0 and 14 of the selected URLs within their top 100. This measure is not so much an indication of how well each of the search engines performed in terms of the coverage of the selected URLs as it is a quick measure of the patron aggravation that each of these search engines is likely to return. For our hypothetical internet searcher executing one of these searches and hoping to discover a reliable resource (of the sort listed in the *C&RL News*) each of these search engines returned between 86% and 100% potential "noise" in their results.

Which brings us to the second question: how well did each of the search engines perform relative to the URLs available? In other words, how well did each of them cover the list of selected resources within their top 100 results? As Table 3 shows, several of the search engines look a bit better when one considers their coverage of the lists as opposed to the return rate for the selected items within their top 100 results.[2]

Although most of the search engines and our keyword searches returned less than 20% of the selected URLs for the year as a whole, there were, however, several instances where the search engines returned anywhere from 21% to 62% of a particular month's selected URLs. Perhaps an interesting feature to note as one glances across Table 3 is that where one search engine did well with a particular topic, several of the others did well, also. See for example, the uncommonly good results across the board for February (Knowledge Management) and July/Au-

TABLE 2. Number of Hits per Keyword Search in the Top 100 Results

	Google	MSN	Acoona	Ice Rocket	Yahoo	Ask Jeeves	Lycos
January	3	1	0	3	3	2	6
February	11	10	6	9	12	13	14
March	2	2	4	3	5	5	3
April	4	0	6	5	7	4	1
May	1	2	2	4	5	2	3
June	0	1	2	3	3	4	0
July/August	8	10	5	13	11	10	11
September	3	1	2	5	2	5	6
October	0	0	3	0	0	3	0
November	0	1	5	3	4	5	3
December	0	0	1	1	4	3	4

TABLE 3. Percentage of Hits in the Top 100 Results for Librarian-Selected URLs (Rounded)

	Google	MSN	Acoona	Ice Rocket	Yahoo	Ask Jeeves	Lycos
January	9	3	0	9	9	6	18
February	39	36	21	32	43	46	50
March	6	6	11	9	14	14	09
April	12	0	18	15	21	12	03
May	3	6	6	12	15	06	09
June	0	4	8	13	13	17	0
July/August	38	48	24	62	52	48	52
September	13	4	9	22	9	22	26
October	0	0	10	0	0	10	0
November	0	3	15	9	12	15	09
December	0	0	5	5	19	14	19
GROUP (N = 314)	10	9	11	16	18	18	16

gust (Electoral College/Political Parties/United States 2004 Presidential Election).

Although they in several instances appeared to have done a very good job of covering the URLs from some of the librarian-vetted resource lists, our results to this point also show that the search engines have returned what seems an irritating amount of "noise" along with the preferred URLs. We initially were hoping that the search engines would, regardless of whether a sizable number of the preferred URLs appeared within their top one hundred results, return a reasonable number of hits for our keyword searches, a manageable list of hits that perhaps a particularly determined searcher could reasonably be expected to browse through. As Table 4 amply illustrates, our hopes were dashed.

The search results above ranged from unwieldy to unimaginable. It is highly unlikely that a patron would be able to wade through tens of thousands of results much less tens of millions, and one should keep in mind that the searches from Table 1 are not "average" searches. As mentioned above, most patron searches are comprised of only one or two terms and do not employ advanced search techniques. Our results here illustrate that as a guide to Karl Popper's "world of the products of the human mind," each of these engines has little to offer to the inquisitive information-seeking patron. This is a gross exaggeration, of course, but our hypothetical patron would seemingly be as well off wandering the streets and asking strangers for information in the hopes of stumbling upon a librarian or other expert from a relevant field.

TABLE 4. Total Results Returned for Each Keyword Search

	Google	MSN	Acoona	Ice Rocket	Yahoo	Ask Jeeves	Lycos
January	2,272,000	290,171	109,388	1,860,000	2,690,000	2,240,000	552,400
February	108,000,000	28,281,452	7,980,731	110,000,000	87,900,000	74,200,000	40,100,006
March	218,000	76,001	13,956	3,530,000	167,000	136,000	27,900
April	655,000	38,699	25,771	676,000	320,000	279,000	81,100
May	5,580,000	113,654	155,268	5,500,000	1,770,000	1,380,000	1,022,000
June	89,900	7,290	2,594	89,300	22,500	27,500	4,650
July/ August	2,900,000 23,000,000 18,500,000	116,232 568,502 205,260	337,095 3,556,450 1,665,925	2,720,000 24,000,000 18,200,000	4,100,000 17,100,000 15,600,000	3,670,000 14,000,000 12,500,000	1,166,000 2,249,000 1,316,000
September	13,100,000	3,247,219	867,073	13,200,000	9,730,000	3,619,000	4,374,000
October	1,220,000	3,552,888	1,936,581	117,000,000	29,600,000	4,955,000	6,067,000
November	16,700,000	41,560	409,758	16,900,000	3,870,000	1,302,000	1,409,000
December	37,900,000	1,230,507	1,463,959	25,800,000	38,700,000	4,648,000	5,350,000

Our final question then would be, after having searched these engines separately and found them wanting, might our hypothetical Internet searcher improve his or her results by employing them in concert, as a sort of team? As Table 5 shows, the answer would largely be, once again, "no."

For the February and the July/August Web resource lists, the seven search engines performed adequately, covering roughly 68% and 76% of the preferred URLs, respectively, within their top 100 results. For the resource lists for the rest of the year, however, the search engines as a team returned far less than 50%. One could argue that the poorness of these search results is a product of the vagueness of our keyword searches. For example, the results for the October search were particularly low and our search was particularly broad. One also might be inclined to argue that someone researching a topic on the Middle East would have employed a narrower search. However, current research shows that most searches employed by most patrons are more vague and less particular than the searches we offer here. Thus, our keyword searches are more narrowly focused than the average search, yet the results returned are still absolutely unmanageable in size.

CONCLUSION AND DIRECTIONS
FOR FURTHER RESEARCH

What we have provided here is quite obviously not a definitive study of the utility of search engines as they relate to Internet searchers and

TABLE 5. Total Hits for Combined Search Engines As a Percentage of Selected URLs

Months	%
January	24.24
February	67.86
March	17.14
April	33.33
May	21.21
June	20.83
July/August	76.19
September	34.78
October	13.33
November	27.27
December	42.86

librarians as reviewers, but more of what cognitive psychologist Gerd Gigerenzer would call "a fast and frugal" assessment (Gigerenzer 1999). For a fuller picture, more study would be necessary. In fact, there are several likely follow-up studies to this project. If one were inclined to compare the utility of the several search engines, one follow-up study would involve executing the same searches and asking a panel of relevant experts from the field of librarianship to identify how many of the top 100 returned results they would be likely to recommend to their patrons. A second study might involve re-performing the searches performed here several times in order to obtain a more "diachronic" assessment of the search engines. Our study as performed pertains just to the selected resource lists during a brief interval and to discrete search results. If one were interested in the efficacy of Internet searchers' behaviors relative to the search engines' performance, a third study, obviously, would involve an actual user study in which Internet searchers of various abilities were asked to provide their own searches and search vocabularies.

Again, our results should hardly be taken as the final say on this question, but it certainly seems as though the Web as a reference resource would benefit from the sort of pointed, value-added reviewing and indexing that librarians and services like Yahoo provide. Our results suggest that whatever relevancy/ranking algorithms search engines are using are actually remarkably good, at least in several instances, at pushing the librarian-selected URLs nearer to the top of their results lists, and the numbers of total hits that some of the engines provided were also impressive in the raw. However, the searching and evaluative behaviors of actual web searchers would seem to render the search engines' efforts moot, and the unwieldy and enormous result sets tended largely to drown out our selected librarian-recommended sites in a seemingly endless sea of Web noise.

NOTES

1. See, for example, the early efforts of Melville Dewey, John Cutter, Paul Otlet and Henri LaFontaine in Europe (Rayward 1997), the subsequent efforts of the American Documentation Institute in the United States (Williams 1997), or the speculative musings of H.G. Wells on the "World Brain" (Wells 1938) and Vannevar Bush on the Memex (Bush 1945).

2. If one were to assume that all 28 of the librarian-selected URLs from February's list were randomly distributed across Google's 108,000,000 hits (which the results rather obviously suggest was not the case) and that each hit returned was to a unique

URL (which actually was not always the case with all the search engines), then one would expect that it would be very unlikely indeed for even one of the 28 selected URLs to appear in the first 100 results (approximate probability = 0.000025925). That 11 of the selected URLs would be returned in the top 100 results, as was the case with Google, would be astronomically unlikely (approximate probability = 5.0375×10^{-59}).

REFERENCES

Allison, D.A. and S. Childers. 2002. Index Relativity and Patron Search Strategy. *portal: Libraries and the Academy* 2(1): 145-53.

Bush, V. 1945. As We May Think. *Atlantic Monthly* 176(1):101-08 [accessed at http://www.theatlantic.com/doc/194507/bush (Nov. 20, 2005)].

Chelton, Mary Kay and Colleen Cool. 2004. *Youth Information-Seeking Behavior: Theories, Models and Issues.* Lanham, MD: Scarecrow Press.

Cuban, Mark. 2005. "Google responds on Blogspot–and Icerocket.com continues to grow." *blog maverick: the mark Cuban weblog* (October 21, 2005) [accessed at http://www.blogmaverick.com/entry/1234000013064482/ (November 20, 2005)].

Delany, Kevin J. (2005). "Google's rivals narrow search gap." *Wall Street Journal.* Jan. 13: 4.

Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. 1999. *Simple Heuristics that Make Us Smart.* New York: Oxford University Press.

Novotny, E. 2004. I Don't Think I Click: A Protocol Analysis Study of Use of a Library Online Catalog in the Internet Age. *College & Research Libraries* 65(6): 525-37.

Popper, Karl. *Objective Knowledge: An Evolutionary Approach*, Oxford: Clarendon Press, 1972.

Popper, Karl. "Three Worlds." The Tanner Lecture on Human Values. University of Michigan, April 7, 1978. [accessed at: http://www.tannerlectures.utah.edu/lectures/popper80.pdf (Nov. 20, 2005)].

Rayward, W. B. 1997. The Origins of Information Science and the International Institute of Bibliography/InternationalFederation for Information and Documentation (FID). *Journal of theAmerican Society for Information Science* 48(4): 289-300.

Sherman, C. 2005. A New F-Word for Google Search Results, *Searchenginewatch* (March 8, 2005) [accessed at http://searchenginewatch.com/searchday/article.php/3488076 (November 20, 2005)].

Shenton, A. K. and P. Dixon. 2004. Issues arising from youngsters' information-seeking behavior. *Library and Information Science Research.* 26(2):177-00.

Spink, A., J. Bateman, and B. J. Jansen. 1999. Searching the Web: A survey of EXCITE users.*Internet Research-Electronic Networking Applications and Policy.* 9: 117-28.

Tyler, D. C. and B. McNeil. 2003. Librarians and Link Rot: A comparative analysis with some methodological considerations. *portal: Libraries and the Academy* 3(4):615-32.

Ward, M. 2004. "How well do you use the web?" *BBC NEWS UK Edition* (March 29, 2004) [accessed at http://news.bbc.co.uk/1/hi/technology/3578149.stm (November 20, 2005)].

Weiler, A. 2005. Information-seeking behavior in Generation Y students: Motivation, critical thinking, and learning theory. *Journal of Academic Librarianship* 31(1): 46-53.

Wells, H. G. *World brain*. (1938) Garden City, N.Y.: Doubleday.

Williams, R. V. 1997. The Documentation and Special Libraries Movements in the United States, 1910-1960. *Journal of the American Institute for Information Science*. 48(9):775-81.