

Search Magic: Discovering How Undergraduates Find Information

American Anthropological Association
Annual Meeting 2011
Montreal, Canada

Andrew D. Asher, PhD
Scholarly Communications Fellow
Bucknell University
andrew.d.asher@gmail.com

Introduction

Searching for information might seem like one of the most routine and commonplace activities of university life. However, as students work within an information environment that is increasingly open and dynamically changing, research assignments also represent a complex and potentially daunting task, and one that is fraught with embedded social and cultural processes and relationships.

The Ethnographic Research in Illinois Academic Libraries (ERIAL) Project was a two-year study of student research practices involving a collaborative effort of five Illinois universities: DePaul University, Illinois Wesleyan University (IWU), Northeastern Illinois University (NEIU), the University of Illinois at Chicago (UIC) and the University of Illinois at Springfield (UIS). Using a mixed-methods approach that integrated nine qualitative research techniques and included over 600 participants, the ERIAL project sought to gain a better understanding of undergraduates' research processes based on first-hand accounts of how they obtained, evaluated, and managed information for their assignments.

In this paper, I will focus only on the subset of this data related directly to the search itself. This data is principally drawn from two research methods: 156 semi-structured ethnographic interviews, in which students were asked to demonstrate searches they had

conducted for a recent research assignment, and 60 research process interviews, in which a project anthropologist accompanied and observed students as they located resources.

Algorithmic Culture

My goal in this paper is not only to describe some of the results from the ERIAL study, but also to illustrate how an ethnography of students' search practices might demonstrate and contribute to the concept of "algorithmic culture." Following Ted Striphas (2011a), I use the term "algorithmic culture" to describe how some aspects of the work of culture--"the sorting, classifying, hierarchizing, and curating of people, places, objects, and ideas"-- are becoming the purview of "machine-based information processing systems." Furthermore, "some of our most basic habits of thought, conduct, and expression. . .are coming to be affected by algorithms, too. It's not only that cultural work is becoming algorithmic; cultural *life* is as well" (Striphas 2011a).

Algorithms are also cultural artifacts themselves, and can be understood as embodying a set of socially and culturally embedded negotiations, decisions, judgments, biases, politics, and ideologies. For example, PageRank, the ranking and relevancy algorithm that comprises the core of Google search, is fundamentally premised on a concept of aggregated social judgment, that is, the assumption that a mathematical calculation based on the number of links to a website combined with an evaluation of the relative importance of the websites from which those links originate, can be used as a proxy for evaluating the quality or value of a site in a way that is analogous to how citations are used to evaluate academic papers (Brin & Page 1998; Page et. al. 1999; see also Battelle 2005:75-76). In addition to PageRank, the Google search algorithm uses a total of more than 200 "signals"¹ to rank its search results--including measures related to localization, personalization, timeliness, and quality (e.g. spam/content farms)--each of which represents a specific decision about the relative value of information. Because of these

¹ <http://www.google.com/about/corporate/company/tech.html>

embedded judgments, “algorithmic culture encourages us to see computational process not as a window onto the world but as an instrument of order and authoritative decision making. The point of algorithmic culture, both terminologically and methodologically, is to help us understand the politics of algorithms and to approach them and the work they do more circumspectly, even critically” (Striphas 2011b).

One central characteristic of the politics of algorithms is their secrecy. Secrecy is fundamental to the algorithmic culture expressed both by Google and by other search tools. While the general parameters of Google’s search algorithm are publicly available, its details are proprietary and closely guarded corporate secrets. One justification for Google’s secrecy is the argument that it helps ensure the efficacy and impartiality of search results by preventing websites from gaming the algorithm to artificially inflate their ranking (for example, see Battelle 2005:159-163). Indeed, an entire industry--search engine optimization--has materialized to reverse engineer Google’s algorithm. It is, of course, also in Google’s economic interest to carefully control the availability of information about its algorithms. In this way, the algorithmic culture established by Google, as well as other proprietary search algorithms, ultimately rests on a tension between proprietary knowledge on the part of the corporation and trust on the part of the user (Battelle 2005:183-185). Because search systems can not be properly interrogated by their users, (except, perhaps, by a few with very sophisticated technical skills), these users must simply put their faith and trust in the algorithm and the people who designed it.

By shaping the processes through which information is found, and by extension, becomes known, search algorithms perform an epistemological function. By structuring the discovery of information, search algorithms express a form of Foucaultian disciplinary power that provides the scaffolding for how students complete their academic work and profoundly structures the

way students acquire knowledge. The secrecy inherent to these search processes and tools should therefore be critically addressed by educators and students alike.

Google's Simple Search

Google has become the primary starting point for students for both everyday and academic research, as evidenced not only by the ERIAL project, but also by almost every recent study of student search habits (Connaway and Dickey 2010:28–29; Head and Eisenberg 2009:15; De Rosa 2006:1–7; Prabha, Connaway, and Dickey 2006:13–14,16–18; Griffiths and Brophy 2005:550, 545). For this reason, understanding the embedded cultural politics of Google's search algorithm is critical to understanding students' search practices

Not surprisingly, for the students interviewed in ERIAL project, Google was by far the most prevalent search tool used. 88% of the students interviewed discussed Google, and Google was mentioned over three times more frequently than the next most popular search tool (JSTOR). Google's dominance may also be affecting academic search in other subtle, yet critical, ways. Google's simplicity and single search box seems to have created the expectation among students of a specific search experience for academic search systems: in particular, a single search box that quickly accesses many resources, and an overreliance on simple keyword search (see Hampton-Reeves et. al. 2009:45; CIBER 2008:14).

During the 60 research process interviews conducted for the ERIAL study, the research team observed 161 unique searches.² 80% (128) of these searches were for unknown items (e.g., when a student was attempting to discover sources related to a research question, rather than a specific book title or journal article). The vast majority of these searches were simple searches.

² For the purposes of our analysis, we defined a search as any time a student opened a new resource to search for information. If the student changed his search terms within a resource, we did not count this as a new search. Therefore we observed 161 searches encompassing 238 separate sets of search terms.

Students generally treated all search boxes as the equivalent of a Google search box, and searched “Google-style,” using the “any word anywhere” keyword search as a default, even when it was not appropriate or effective to do so. In total, 202 of the 238 (85%) observed sets of search terms used this approach (see also CIBER 2008,14; Hampton-Reeves et. al. 2009, 45). A junior in nursing explained, “I basically throw whatever I want into the search box and hope it comes up. . . . But it’s like Google and I use it like Google. I don’t know how to use it any other way.”

Students’ overuse of the simple search leads directly to the problems of obtaining too many or too few search results. These twin problems of “too little” and “too much” information are really one and the same, as both issues stem from a lack of sufficient conceptual understanding of how information is organized and how to build an effective search query. Almost all of the students interviewed by the ERIAL Project exhibited difficulties evaluating and narrowing down (or expanding) search results. When faced with unsatisfactory results, students usually changed the search, either by entering new search terms or trying a different database altogether, rather than using more advanced search tools to expand or refine the search. Perhaps because of their experience with Google, students often appeared to believe that if they could only find the magic words or phrase, whatever piece of information they were looking for would be revealed to them and a manageable list of results would be returned. In this way, the search experience is thus iterative rather than determinative. This practice can lead to students using lower quality or less accurate search terms because they return fewer results (especially in a typical full text search).

This is another important expression of the algorithmic culture of search: the Google search experience has conditioned students that search should be simple. Academic libraries and

information providers have likewise embraced simplicity as a positive value for search, and have sought to reduce the “cognitive load” of using the various and fragmented catalogs, databases, and interfaces contained on a typical library website (CIBER 2008:30 see also Wong et. al. 2009:6). Much effort has recently gone into making the experience of searching library databases more like Google by the creation of “discovery tools” that can simultaneously search across a library’s many platforms, catalogs, and databases to provide aggregated results. Like Google, these search systems provide a single search box that is easily queried with natural language, and like Google they rely on complex and proprietary algorithms to produce relevancy rankings.

However, by making search easier for students these tools can counterintuitively decrease the quality of the search. By enabling students to get to information faster and easier, these systems can also reinforce unreflective research habits that contribute little to the overall synthesis of a research paper or academic argument.

Database Choices

This process can be illustrated by the usage of academic research databases observed in the ERIAL project. While students routinely began with Google, the ERIAL project and other studies (Gabridge, Gaskell, and Stout 2008:516–17; Head and Eisenberg 2009:3) have observed that students working on academic assignments do eventually consult library databases, particularly when seeking reliable or scholarly sources.

When choosing a database to conduct a search, students typically returned repeatedly to a resource that had worked in the past, even if it was not the best or most appropriate for the task (see also Head and Eisenberg 2009:3). Usage statistics gathered during to ERIAL project at the Ames Library at Illinois Wesleyan University reflected this. Of the 101 databases that had one

year of comparable search data, the top three databases (JSTOR , PsycINFO, and Academic Search Premier) accounted for 38.7 percent of searches. Usage then fell off quickly, with the next seven databases accounting for 20.4 percent of searches. The remaining 91 databases accounted for 40.9 percent of searches, with 76 of these databases holding a less than 1 percent share of total searches (see Table 1).³

Database	Number of Searches	Percentage of Total Searches
JSTOR	32,116	14.78%
PsycINFO	27,906	12.84%
Academic Search Premier	24,082	11.08%
Top 3	84,104	38.7%
CINAHL Plus with full text	9481	4.36%
Hoover's Online	6501	2.99%
MLA International Bibliography	6190	2.85%
Science Citation Index	5789	2.66%
LexisNexis	5515	2.54%
Social Sciences Citation Index	5478	2.52%
Arts & Humanities Citation Index	5400	2.49%
Top Ten	128,458	59.11%
Remaining 91 Databases	88,840	40.88%
Total	434,596	100%

Table: 1 Database use at IWU, AY 2008-2009.

The popularity of the JSTOR database is perhaps illustrative of the privileged place simplicity takes in the algorithmic culture of search. JSTOR is a multidisciplinary content

³ These statistics encompass usage by the entire university. Unfortunately, it is impossible to differentiate student searches from other users. However, given that IWU students vastly outnumber faculty, it is reasonably safe to assume that this usage is student-driven.

provider of full-text academic journals, and allows students (and scholars) to easily search across up to 1400 journals via a single search interface. JSTOR was extremely popular among the students who were interviewed by the ERIAL project. JSTOR was referenced more than five times as often and by twice as many participants than then next most popular database, Academic Search Premiere. Students appeared to rely on JSTOR disproportionately compared to other academic databases, to an extent that surprised the librarians working on the project, who often did not view JSTOR as the optimum resource for students' research assignments. However, for students, JSTOR was usually sufficiently robust to meet the minimum requirements of a particular assignment—typically around five sources. JSTOR simply works for a wide range of assignments across a wide range of disciplines, providing fast access to full-text and reliable resources.

Students generally did not realize—and had not investigated—the limitations of the database that might make it inappropriate for a given task. For example, students regularly used JSTOR to search for current information, not realizing that JSTOR often does not provide access to the most recently published articles (articles typically only appear in JSTOR after 3–5 years, depending on publisher).⁴ Nor did students think to investigate whether or not there was a database that would be more focused on their topic of choice. Students found JSTOR effective because it fit in well with their established work practices. Unfortunately, because it provides access to full-text materials, as well as its flexibility and wide coverage of topics, JSTOR also enabled students to succeed using subpar search strategies because it worked well enough.

⁴ This changed in 2011 when JSTOR will began offering a “Current Scholarship Program” containing up-to-date content from over 200 journal titles.

Trust Bias: “I Never Go Past the First Page”

In an information environment where the retrieval of information is increasingly trivial, students’ ability to effectively evaluate information becomes preeminently important. Unfortunately, in the searches observed by the ERIAL project, students’ evaluation of potential sources appeared cursory (see also CIBER 2008:10). Students typically made rapid appraisals of a source’s usefulness, often based only on its title or a superficial scan of its abstract. When evaluating search results, students seldom examined citations past the first page of results, an observation that is supported by Griffiths and Brophy’s recent study of search engine use (2005:551).

These practices are also an expression of the disciplining effects of algorithmic culture. Through the act of ordering and ranking, search systems’ relevancy algorithms impart (and reinforce) a sense of authority and credibility in the results. Users regularly assume that information that is objectively “best” will be ranked first. This “trust bias” is well documented in the literature on search engines (Vaidhyathan 2011:59; Hargittati et. al. 2010; Hargittai 2007; Pan et. al. 2007). Because it holds the power to create a list of results, the search engine self-validates the quality of its results. In this recursive loop, users depend solely on their trust in the search algorithms brand, be it Google, or JSTOR, or something else.

This belief that credible and quality resources should appear on the first page or two of search results caused many students observed by the ERIAL Project to assume that if they cannot quickly find information on a topic, then the information must not exist and they should give up on that topic. Only rarely did students conclude that a lack of search results might, in fact, reflect incorrect search terms or an ineffective search strategy.

For example, when discussing a recent research paper, a sophomore international studies major observed, “Originally I had a different topic. I was thinking about something that had to do [with] the discrimination of Jews in sixteenth-century London, and I realized that finding information on that would be almost impossible. ’Cause I’m interested in the really obscure topics that you would be like, “that’s really interesting.” But no one really has done anything on that, so it’s really hard to find. So, [crime in nineteenth century London] seemed like it would be easy to find information on, so I decided on that one. . .”

Students regularly overestimated how “obscure” a particular topic actually is, and demonstrated remarkable ease in changing topics to fit easily found information. Because of this, students often passed up unique or interesting topics in favor of topics with widespread coverage.

Search Magic

Search algorithms can thus reveal or conceal information depending on the skills of the user. Unfortunately, the students who participated in the ERIAL project did not appear to adequately understand conceptually how information is organized or how search works. Of all the students who were asked, none could correctly explain how a search in Google (or any other search engine) works or organizes results. Search results were returned “as if by magic.”

Arthur C. Clarke’s observation “That any technology, sufficiently advanced, is indistinguishable from magic” (1973[1962]:21) has been used in relation to Google perhaps to the point of cliché (see Battelle 2005:129; Vaidhyanathan 2011:53). However, we should still attend to the reason why search feels like a magical experience.

Siva Vaidhyanathan argues that Google seems magical because of its usefulness for helping its users find meaning by providing a managed and ordered set of actionable choices for

a query (2011:53). Coupled with a speed that seems near-instantaneous (see Vaidhyathan 2011:53-54), this experience makes it easy for users to forget that Google (and other search algorithms) are simply tools, especially since their workings are made intentionally opaque. This secrecy makes it difficult for students to fully understand the embedded politics of how information is organized and retrieved. This lack of “algorithmic literacy” potentially renders students vulnerable to the disciplinary power contained in search systems, as well as subjects, rather than agents, of algorithmic culture.

Indeed, and students described experiences of anxiety and confusion as they searched for resources. A senior in women’s studies described her difficulties conducting a search, “. . . finding ways to narrow down, there was just so much information. . . how do I weed out what my specific topic is from the general larger topic? . . . How do I find specifically my information when there’s not a book titled [on] this topic? So, I guess just being overwhelmed with the amount of literature out there [that] doesn’t really relate to my topic and how do I pull my stuff out of it? ’Cause I feel like I was very much kind of blindly branching out and a lot of times by chance finding things and then going on from there.

One challenge for educators and librarians is to balance facilitating ease of use with a conceptual understanding of how search works. Search shouldn’t be magic; it’s only when its processes and algorithmic culture are demystified that our students become empowered to use it effectively.

Works Cited

- Battelle, J.
2005. *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York: Portfolio.
- Brin, S., and L. Page.
1998. "The anatomy of a large-scale hypertextual Web search engine* 1." *Computer networks and ISDN systems* 30 (1-7): 107–117.
- CIBER (Centre for Information Behaviour and Evaluation of Research).
2007. *Information Behaviour of the Researcher of the Future: A CIBER Briefing Paper*. London: CIBER.
- Clarke, Arther C.
1973[1962] *Profiles of the Future: An Inquiry into the Limits of the Possible*. Revised edition. New York: Harper & Row.
- Connaway, Lynn Silipigni, and Dickey, Thomas.
2010. *The Digital Information Seeker: Report on Findings from Selected OCLC, RIN and JISC User Behaviour Projects*. OCLC Research.
- Griffiths, J. R, and P. Brophy.
2005. Student Searching Behavior and the Web: Use of Academic Resources and Google. *Trends* 53(4): 539.
- Hampton-Reeves, S., C. Mashiter, J. Westaway, P. Lumsden, H. Day, and H. Hewertson.
2009. *Students' Use of Research Content in Teaching and Learning*. A report for the Joint Information Systems Council (JISC). Centre for Research-informed Teaching, University of Central Lancashire.
- Hargittai, E.
2007. "The social, political, economic, and cultural dimensions of search engines: An introduction." *Journal of Computer-Mediated Communication* 12 (3): 769–777.
- Head, Alison J., and Michael Eisenberg.
2009. *How College Students Seek Information in the Digital Age*. Project Information Literacy Progress Report. University of Washington.
- Hargittai, E., L. Fullerton, E. Menchen-Trevino, and K.Y. Thomas.
2010. "Trust online: young adults' evaluation of Web content." *International Journal of Communication* 4: 468–494.

- Page, L., S. Brin, R. Motwani, and T. Winograd.
1999. "The PageRank citation ranking: Bringing order to the web."
- Pan, B., H. Hembrooke, T. Joachims, L. Lorigo, G. Gay, and L. Granka.
2007. "In Google we trust: Users' decisions on rank, position, and relevance." *Journal of Computer-Mediated Communication* 12 (3): 801–823.
- Prabha, C., L. S Connaway, and T. J Dickey.
2006 *The Whys and Hows of College and University User Satisficing of Information Needs. Phase IV Report: Semi-Structured Interview Study*. Report on National Leadership Grant LG-02-03-0062-03, to Institute of Museum and Library Services, Washington, D.C. Columbus, Ohio: School of Communication, The Ohio State University.
http://imlsproject.comm.ohio-state.edu/imls_reports/imls_PH_IV_report_list.html.
- Striphas, Ted.
2011a. "Who Speaks for Culture?" posted Sept. 26, 2011,
<http://www.thelateageofprint.org/2011/09/26/who-speaks-for-culture/>

2011b. "Culturomics," posted April 5, 2011,
<http://www.thelateageofprint.org/2011/04/05/culturomics/>
- Vaidhyanathan, Siva.
2011. *The Googlization of Everything (and Why We should Worry)*. Berkeley: University of California Press.
- Wong, William, Stelmaszewska, Hanna, Bhimani, Nazlin, Barn, Sukhbinder, and Barn, Balbir.
2009. *User Behaviour in Resource Discovery: Final Report*. JISC.