

The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction

Eszter Hargittai

Department of Communication Studies Northwestern University &
Center for Advanced Study in the Behavioral Sciences Stanford University

Search engines are some of the most popular destinations on the Web—understandably so, given the vast amounts of information available to users and the need for help in sifting through online content. While the results of significant technical achievements, search engines are also embedded in social processes and institutions that influence how they function and how they are used. This special theme section of the Journal of Computer-Mediated Communication explores these non-technical aspects of search engines and their uses.

doi:10.1111/j.1083-6101.2007.00349.x

Introduction

Search engines are some of the most commonly accessed websites online (see a listing of the most popular online properties at ranking.com, including top spots held by msn.com, google.com, and yahoo.com). Millions of people turn to them to find content on a daily basis (Fallows, 2005), submitting billions of queries each month (comScore, 2007). In fact, search engine use rivals email as the most common activity undertaken by Internet users (Rainie & Shermak, 2005). People turn to search engine services day after day to find information about current events, health concerns, products, government services, natural disasters, their new neighbors, prospective employees or dates, and a myriad of other topics ranging from the mundane to the utmost serious. Despite their central role in how people access information, however, little social science work has focused on the non-technical dimensions of search engine tools, the companies that run them, or the practices of the users who rely on them. The goal of this special theme section of the *Journal of Computer-Mediated Communication* is to consider the social, political, economic, and cultural dimensions of large-scale search engines.

An undertaking of this sort must be an interdisciplinary effort given the many dimensions of the relevant questions, ranging from who uses search engines and for what purposes to the distribution of skill in search engine uses, and from whether all

content has equivalent chances of being included in the result listing of these tools to whether they can be manipulated. Accordingly, contributors to this special section represent several fields including communication, sociology, psychology, library and information science, and media studies. Not surprisingly, the vast array of questions raised by search engines attracts a diverse group of scholars.

The articles included here can be grouped into three categories based on their focus. Two articles consider the cognitive dimensions of searching while taking into consideration social factors such as trust and user demographics. Another two articles focus on the social context of individuals' information seeking online, considering the role of their social networks and their experiences in their information-seeking behavior. The remaining three articles examine the tools themselves to consider what content search engines present to users and what are the behind-the-scenes processes that go into the decisions about what to include and how to display results.

Earlier Work

To be sure, the non-technical aspects of search engines and their uses is not uncharted territory. Depending on how widely one casts the net, one can find considerable relevant work in past literature across fields. For example, researchers in the areas of library and information science have been interested for a long time in how people find material using various interfaces and databases, and these projects are not unrelated to the questions addressed by the articles in this collection (for a review of some of this work, see Bar-Ilan, 2003; Hsieh-Yee, 2001). However, work in this domain often only focuses on very small and non-representative samples and rarely considers the social context of searching (for a more elaborate discussion of this point, see Hargittai & Hinnant, 2005). Thus, a collection of pieces focusing on the social, political, economic, and cultural dimensions of search engine use addresses a gap in the literature.

What do we already know? Thanks to existing work, we do know that search engine use is one of the most popular activities among Web users (Fallows, 2005; Rainie & Shermak, 2005). We also know that when asked about their search abilities, many users tend to be confident (Fallows, 2005), although research observing people's online information-seeking behavior tends to find discrepancies depending on user attributes (e.g., Hargittai, 2002), and logs of search queries suggest that the majority of users do not take a particularly sophisticated approach to searching but rather often rely on only one or two terms in their queries (e.g., Spink, Jansen, Wolfram, & Saracevic, 2002).

We also have evidence suggesting that search engine users are not particularly savvy about the behind-the-scenes of search engines. For example, when asked in one study whether they were aware of the distinction between paid and unpaid results, the majority (62%) indicated that they were not (Fallows, 2005). These findings were mirrored by another report asking similar questions in which 56% did not know the

difference between the two types of results (iCrossing, 2005). Moreover, the latter findings suggested that this know-how is not randomly distributed among users, but rather that men and younger adults claimed to be more informed about this aspect of search engines than women and older users.

Regarding the role of search engines in channeling user attention, although researchers started considering the possible gatekeeping implications of these services years ago (Hargittai, 2000; Introna & Nissenbaum, 2000), little empirical work has followed to examine the extent to which search engines may or may not discriminate against certain types of content while (perhaps unduly) favoring others. Some case studies have examined the censorship of certain types of material by some search engines in particular national contexts (e.g. Finkelstein, 2004; Zittrain & Edelman, 2002), but there is little systematic work considering less controversial materials and their relative chances of inclusion.

Recent trends suggest that the search engine market is shrinking, with fewer large players guiding users' online behavior than ever before. This suggests that decisions made by just a few players in this landscape can have considerable repercussions for what material is realistically within the reach of users. Accordingly, a critical look at what factors determine inclusion and exclusion criteria in search results and how users approach them is increasingly important in order to gain a better understanding of how users' access to content is being mediated by a handful of commercial services.

While previous work has examined questions related to those explored in this special theme section of *JCMC*, a unique contribution of this collection is that it presents together articles exploring the social, economic, political, and cultural dimensions of search engines instead of having them scattered across different outlets. By showcasing these articles that represent a diverse set of methods, approaches, and questions all relating to search engines in one compilation, this special theme section highlights the diversity of issues at hand and how the perhaps seemingly unrelated explorations are interconnected and can benefit from one another.

Methodological Challenges

All fields face methodological challenges, but new areas of inquiry must also tackle uncharted terrain, which adds to the complexity of the undertaking. Data necessary to examine important questions concerning the social aspects of search engines are scarce despite search engines themselves generating voluminous data sets based on the logs of their users' actions. The reasons for data scarcity with respect to academic research are manifold and include the limitations imposed by proprietary data sets as well as other factors discussed below.

Search engine companies have enormous amounts of data about the use of their services, but these data tend to be proprietary and are rarely released to researchers. Companies are very concerned about anonymizing such data, which is a non-trivial process and thus requires considerable effort. For example, in 2006, researchers from AOL released a seemingly anonymized data set containing over 20 million search

queries from over 650,000 users spanning three months (Pass, Chowdhury, & Torgeson, 2006) as a resource to the non-commercial community for research purposes. However, due to the level of detail in the data set (including ID numbers attached to each query), an analysis of the data led to the identification of some users (Barbaro & Zeller, 2006). Given the controversial response and repercussions from the case (Wray, 2006), it is even less likely that companies will make such information available to researchers in the future.

Even if log data were made more readily available, there are limits to how much one can learn about users simply based on logs (Hargittai, 2007). Such data are rarely accompanied by the types of covariates about user attributes that make certain types of fine-grained analysis possible. Moreover, given that users are not randomly distributed across search engines (see “Future directions” below), knowing information about the users of one site does not mean that one can necessarily generalize to Internet users overall.

Another challenge in researching this domain is that Internet use is very much a moving target with ever-changing terminology to describe the various associated phenomena. While a user may have experiences with a particular service such as Yahoo! or Ask, he or she may not know that the search services on these sites are called “search engines.” This lack of understanding may seem implausible to some, but data from the *General Social Survey* (2000, 2002) suggest that users are not always clear about the concept of search engines, sometimes confusing them with Web browsers. Additionally, data I have collected on young adults’ Internet uses show that some users who claim to use specific search engines concurrently report not knowing what a search engine is when asked whether they use any (Hargittai, 2007). One way to get around these issues of terminology is to ask about the use of specific services. However, often surveys—except for ones by marketing professionals—shy away from asking about proprietary services. While understandable for some reasons, this tendency to omit specific services from studies may occasionally hurt the research process.

At a different level of analysis—taking the search engine as the focus of study—researchers are faced with other challenges. Studying search engine coverage is difficult, for example, because random sampling of websites is impossible given that there is no comprehensive listing of all existing sites. Also, because the algorithm of search engines is proprietary information, it is impossible to know what gets covered and what does not by the various services. In fact, in some cases results change by user and user location, so a study conducted on one machine in one location by a particular user may not be possible to replicate on another machine under different circumstances, even soon after the initial inquiry. This poses significant challenges for the replication of research results, which is a basic tenet of scientific investigation.

For the above reasons, most of the authors in this collection could not draw on existing data sources and have had to rely on original data collection for their studies. The Wirth, Böcking, Karnowski and van Pape article and the Pan, Hembrooke, Joachims, Lorigo, Gay and Granka article are both based on studies conducted by the researchers in lab settings. Kayahara and Wellman base their study on interviews

and, in some cases, in-person observations of people's online behavior. Howard and Massanari analyze survey data from the *Pew Internet & American Life Project*. Van Couvering conducted in-depth interviews with search engine producers. Vaughan and Liwen used automated techniques to generate a random list of domain names and then checked their coverage by four search engines. Bar-Ilan identified pages popularized by aggressive targeted linking and looked at changes in coverage by considering what other sites link to them compared over time. Overall, the articles in this special section draw on a diverse set of interesting and unique data to address the questions to which they seek answers.

In addition to data source challenges, an emerging area of study also faces the evolution of new terminology. The articles in this section address many similar issues but in earlier drafts used varying terms to describe similar phenomena and ideas. Where possible and where conventions have started to emerge, the terms have been consolidated for the sake of consistency and comprehension. For example, in the marketing and search engine industry, the results listing of a search engine has come to be referred to as "SERP"—the search engine results page. Other acronyms are popular as well, such as SEO, which stands for search engine optimization. The authors have clarified such terms when they are introduced in order to facilitate access to the studies by non-specialists.

In this Special Section

The authors whose work is included in this collection represent numerous disciplines and are based in institutions across several countries. Research on search engines is very much an interdisciplinary undertaking, as evidenced by the studies featured here.

Approximately fifty abstracts were submitted in response to a call for proposals. The selection of just a portion of these for further review was difficult given the fascinating topics proposed by authors. In some cases, fit was more of an issue than quality. I invited seventeen full papers for submission. Of these, 11 were actually submitted, and the review process resulted in the seven articles included here. In this section, I briefly summarize each.

The first two articles examine people's search engine uses at the level of responding to results pages. The study by Wirth and colleagues looks at how people make decisions concerning the results with which they are presented after performing a search. The researchers find that users engage in both heuristic and systematic processing of results depending on task type and user attributes. In the second article, Pan and colleagues focus on a particular aspect of approaching search engine results: the extent to which users trust one particular service's ranking of results. Based on observations of college students' use of Google in an experimental setting, they conclude that students are influenced by the order of results even if those do not reflect actual rankings or relevance and are due to study manipulation instead.

The next cluster of articles considers the larger social context of searching. Kayahara and Wellman report on a study of how people find information concerning cultural activities. Rather than restricting their project to Internet searching only,

they consider this question in the context of people's social and media environments, both online and offline. Based on interviews, they find that people tend to turn to those in their networks first or traditional media such as the local paper or broadcast programs for recommendations before then proceeding to look up information about the suggested activities online. In the study that follows, Howard and Massanari use survey data about Americans' online activities and understanding of search engines to consider how a user's socioeconomic status and experience with the Internet relate to engaging in search activities on the Web. They argue that experience online can overcome some of the disadvantages posed by certain people's income and educational background as regards their Internet use. As time goes by, it will be interesting to see whether these findings hold up for later adopters as well once they accumulate more experience using the Web.

The last three articles focus on material covered by search engines, including a look at what decisions influence the content and presentation on these services, a comparison of search engines in different countries, and the possible manipulation of one such tool. Van Couvering presents results from in-depth interviews with program managers and engineers of major search engine companies. She finds that these people mainly think about the quality of their products in market and scientific/technological terms. The managers and engineers devote little discussion to issues of fairness in representation when discussing the quality of the tools that they develop. In the next article, Vaughan and Liwen compare the extent to which different search engines index a randomly generated group of websites based in various countries. Their results suggest that U.S.-based sites are more likely to be included in search engines, even when a search engine's focus is on coverage of material from another country. Their analyses consider different types of sites—commercial, educational, and government—from four countries covered on four search engines. Finally, Bar-Ilan explores a search engine manipulation method—Google bombing—and its long-term consequences for Google's results in response to certain queries. After considering the history of nine Google bombs and classifying them by type, the author finds that some manipulations are longer lasting than others. Longevity seems to be related to the original underlying goal of influencing rankings.

Future Directions

It is often the case that interesting research raises more questions than it answers. The various articles in this collection discuss directions for future inquiry resulting from their findings. Whether with regard to the processing of search results or the social context of search, ample room remains for future exploration.

Based on the many proposals received for this theme section, I would like to point out one issue concerning research in this domain that I think warrants consideration. Interestingly, but perhaps not surprisingly, a disproportionate number of submissions in the response to the call for papers concerned the study of a single search engine, most often Google. While there are certainly questions one can

ask about specific services (and some of the articles included in this collection adopt such an approach), it is important to remember that every day millions of people turn to a *variety* of online sources to satisfy their information-seeking needs. Thus depending on the questions asked, research should not limit itself to just one service unless there is solid evidence to suggest that its users are representative of a larger group of users and its services representative of other tools as well. As this area of scholarship matures, it will be important to move beyond studies of single search engines to more inclusive analyses.

A report called "How America Searches" by the marketing agency iCrossing tabulated information about users by search engine experience (iCrossing, 2005). The study found that users of different search engines exhibit varying levels of understanding regarding searching and tend to engage in different activities, suggesting that users of different services are not identical. For example, users of Google are considerably more likely to do searches on professional and business topics or research on products than are users of some other search engines. The study also found that Google users are considerably more likely to know the difference between sponsored and non-sponsored results (56% as compared to 42% of Yahoo users or less for users of other services such as MSN, AOL, and Ask). Whether this gap is due to user characteristics or features of the search engines, the fact is that differences exist by service, and thus results based on the users of one tool may not be generalizable to the entire Internet user population.

Aggregate data on search engine uses suggest that less than half of Internet users in the U.S. run their queries on Google sites (comScore, 2007). Of course, measuring search engine popularity is a complex undertaking and claims about relative user share depend on the specifics of the approach (Hargittai, 2004). For example, it may be that a considerably larger proportion of searchers than suggested by the above figure see results from Google's index of sites given that many sites other than google.com are powered by Google's search engine. However, as I have argued elsewhere (Hargittai, 2004), site-specific layout and presentation also matter; thus use of a particular site versus a particular engine should not be confounded.

In sum, depending on the research questions at hand, it is important to conduct studies in a more holistic manner than simply relying on the services that academics happen to think most people use or happen to assume are interchangeable with others. The focus of research needs to be an aggregation of tools to account for the online actions of various people, not just a select segment that may not be representative.

Conclusion

Overall, the collection of articles in this special section suggests that far from being solely technical phenomena, search engines and their uses are embedded in a myriad of social processes that are important for social scientists to consider in their research in order to understand the social implications of these important tools of our time. Given their popularity, search engines are important brokers of information, and

knowing more about how they represent content and how they are used is vital to understanding patterns of information access in a digital age. By looking at individual search actions, the social context of search, and the landscape of search engine results, the articles in this special section offer interesting food for thought and directions for future research concerning the social, political, economic, and cultural dimensions of search engines.

Acknowledgments

I would like to thank *JCMC* Editor Susan Herring for supporting this special section. I am grateful to the Center for Advanced Study in the Behavioral Sciences for affording me the time and supportive environment necessary for such a project. Moreover, the peer review system of journal publishing would not work if it were not for the many generous colleagues who give their time and effort to reading and commenting on others' writing. For this collection, I relied on researchers from many disciplines and institutions for feedback. I am grateful to the following people for having served as reviewers: Alessandro Acquisti, Lada Adamic, Eytan Adar, Elisabeth Anderson, Paul Baker, Dania Bilal, Michaela DeSoucey, Martin de Santos, Julian Dierkes, Corey Fields, Santo Fortunato, Jeremy Freese, Jason Gallo, Anne Holohan, David Huffaker, Divya Kumar, Sheree Josephson, Dan Li, Adrienne Massanari, Sara Nephew, John Quiggin, Soo Young Rieh, Andrea Tartaro, David Tewksbury, Jan van Dijk, Gina Walejko, James Webster, and Elaine Yuan. Their comments helped improve the manuscripts in this collection.

References

- Bar-Ilan, J. (2003). The use of Web search engines in information science research. *Annual Review of Information Science and Technology*, *38*, 231–288.
- Barbaro, M., & Zeller, T. (2006, August 9). A face is exposed for AOL searcher no. 4417749. *New York Times*. Retrieved March 18, 2007 from <http://www.nytimes.com/2006/08/09/technology/09aol.html?ex=1312776000&en=f6f61949c6da4d38&ei=5090>
- comScore. (2007). *comScore Releases January U.S. Search Engine Rankings*. Reston, VA: comScore Networks.
- Fallows, D. (2005). Search engine users. *Pew Internet & American Life Project*. Retrieved March 18, 2007 from http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf
- Finkelstein, S. (2004). *Jew watch, Google, and search engine optimization*. Retrieved March 12, 2007 from <http://sethf.com/anticensorware/google/jew-watch.php>
- Hargittai, E. (2000). Open portals or closed gates? Channeling content on the World Wide Web. *Poetics*, *27*(4), 233–253.
- Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. *First Monday*, *7*(4). Retrieved March 18, 2007 from http://firstmonday.org/issues/issue7_4/hargittai/index.html
- Hargittai, E. (2004). Do you "Google"? Understanding search engine use beyond the hype. *First Monday*, *9*(3). Retrieved March 18, 2007 from http://www.firstmonday.org/issues/issue9_3/hargittai/index.html

- Hargittai, E. (2007). Beyond gigs of log data: The social aspects of Internet use. On *Tech Talks*. Mountain View, CA: Google, Inc. Retrieved March 18, 2007 from <http://video.google.com/videoplay?docid=6884513936578531954&q=beyond+gigs>
- Hargittai, E., & Hinnant, A. (2005). Toward a social framework for information seeking. In A. Spink & C. Cole (Eds.), *New Directions in Human Information Behavior* (pp. 55–70). New York: Springer.
- Hsieh-Yee, I. (2001). Research on Web search behavior. *Library and Information Science Research*, 23(2), 167–185.
- iCrossing. (2005). *How America Searches*. New York: iCrossing.
- Introna, L., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The Information Society*, 16(3), 169–185.
- National Opinion Research Center. (2000). *General Social Survey*. Chicago: University of Chicago.
- National Opinion Research Center. (2002). *General Social Survey*. Chicago: University of Chicago.
- Pass, G., Chowdhury, C., & Torgeson, C. (2006, May). *A picture of search*. Paper presented at the First International Conference on Scalable Information Systems, Hong Kong.
- Rainie, L., & Shermak, J. (2005). Search engine use. *Pew Internet & American Life Project*. Retrieved March 18, 2007 from http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf
- Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *IEEE Computer*, 35(3), 107–109.
- Wray, R. (2006, August 2). AOL sacks three workers for releasing customers' data. *The Guardian*. Retrieved March 18, 2007 from <http://business.guardian.co.uk/story/0,,1855398,00.html?gusrc=rss&feed=24>
- Zittrain, J., & Edelman, B. (2002). *Localized Google search result exclusions. A statement of issues and call for data*. Cambridge, MA: Berkman Center for Internet and Society, Harvard University. Retrieved March 18, 2007 from <http://cyber.law.harvard.edu/filtering/google/>

About the Author

Eszter Hargittai is Assistant Professor of Communication Studies and Sociology and Faculty Associate of the Institute for Policy Research at Northwestern University where she heads the Web Use Project. In 2006–2007, she is a Fellow at the Center for Advanced Study in the Behavioral Sciences at Stanford. Her research focuses on the social and policy implications of information technologies with a particular interest in how digital media may contribute to or alleviate social inequalities. Her research projects have looked at differences in people's Web-use skills, the evolution of search engines and the organization and presentation of online content, political uses of information technologies, and how digital media are influencing the types of cultural products people consume.

Address: Department of Communication Studies, 2240 Campus Dr., Evanston, IL 60208, USA