

Tel.: +315 443 5617; fax: +315 443 6886

E-mail address: pzhang@syr.edu

URL: <http://melody.syr.edu/pzhang>

Available online 8 February 2006

doi:10.1016/j.ipm.2006.01.001

John Battelle, *The search: How Google and its rivals rewrote the rules of business and transformed our culture*, Penguin Group, 311 pp., ISBN: 1-59184-088-0, Cost: \$25.95

The following three quotations summarize the major points of the book.

“Search as a problem is about five percent solved.” [Udi Manabar, the CEO of Amazon’s A9.com search engine; p. 12].

“We’re one bad story away from being seen as Big Brother.” [Anonymous Google executive; p. 14].

“I would like to see search engines become like the computers in Star Trek. You talk to them and they understand what you’re asking.” [Craig Silverstein, Google employee number one; p. 15].

Therefore, I will organize the review around these three statements.

“Search as a problem is about five percent solved.”

When I first read this statement, I recalled Michael Lesk’s statement in *Digital Libraries: Books, Bytes, and Bucks* (Lesk, 1997) that textual search as a problem is nearly solved [p. 2]. It highlights, for me, (1) the disconnect that so often exists between academics and practitioners and (2) how difficult it is to predict where a field or technology is going. Lesk’s comment comes before the massive increase of content on Web, the reliance on Web search engines in everyday information seeking, the increase in Web spam content to deceive search engines, and the growth of paid search as a business model for online commerce.

Battelle does a good job of supporting Manabar’s position by outlining various issues including technical, social, linguistically, and commercial that Web search faces today. Surprisingly to me, some of the technical issues were the most interesting. For example, the author comments that Google has more computers (approximately, 175,000) servicing its search index than existed in the entire world in the 70s. Battelle touches on issues familiar to most researchers in the field, including the effect of stop words, term matching, and meta-data, along with some more uncommon ones including paid search and spam.

“We’re one bad story away from being seen as Big Brother.”

In the social area, Battelle uses a “Database of Intentions” metaphor to describe the possible uses of and knowledge from the massive amount of search data that Google, Yahoo!, and the other major search engines have concerning the desires, issues, and interests of millions of searchers. He addresses the implicit trust between searchers and search engines and the increasing strain that exists (or will exist) between this “implicit trust” and the pressures from government and others to use this information for a variety of purposes, especially national security. The author does an excellent job of highlighting the simultaneous and ongoing changes in technology, information and people that has brought us to the point where this is a concern. In fact, the databases are already being used for purposes other than improving search (c.f., Ettredge, Gerdes, & Karuga, 2005; Jansen, Jansen, & Spink, 2005).

The tension between protecting this “implicit trust” of searchers and utilizing the information contained within for other purposes raises many concerns and difficult situations. There may not be a right answer for all situations. For example, take national security. Certainly, it seems reasonable that one expects a degree of privacy when using a search engine. However, I then picture in my mind the image from 9/11, of the Towers

coming down. If there is information contained within the search engine databases that could save lives, shouldn't one use it to save those lives? You cannot get much more "Don't be evil" than that.

Perhaps technology will provide a "happy middle ground". As an analogy, most organizations have pornography detection software that is reasonably non-intrusive running on their networks due to sexual harassment, illicit use of equipment, and productivity concerns. Why not some "terrorist detection software" for these search engine databases using a similar approach? Whatever the answer, this is tough situation that society is going to have to address soon, as the author states.

"I would like to see search engines become like the computers in Star Trek. You talk to them and they understand what you're asking."

The discussion here is one that researchers in the field will be very familiar with, namely "How to get the retrieval engine to 'know' what you are looking for?" The book both sets the stage with Web search in this area and paints the step-by-step picture of technical and paradigm advances that are getting us to this goal.

Along the way, it may be difficult to tell which technological innovations will actually be beneficial. In a confirmation of what I had heard anecdotally, Brin and Page recount how their research paper on the link ranking feature now known as PageRank was rejected by the SIGIR conference as "disjointed" and "not IR" enough. (It would be interesting to find out who the reviewers were!) In hindsight, one can make a good case that the implementation of Web linking with anchor text was one of the major advances in Web search. Brin and Page's subsequent paper describing this work (Brin & Page, 1998) has become one of the most cited papers in information retrieval. For those interested, see CiteSeer, ISI Web of Science, and Google Scholar for the citation counts. So, for the legions of researchers who have had papers rejected from SIGIR, take heart!

These are the three major themes of the book, but there are many other interesting aspects also. One captivating idea presented in the book is "search" as an interface metaphor, which makes good sense given that the browser was designed to display Web results and is the de facto standard for application interfaces today. In addition, there are interesting correlations presented in the book on how search is influencing our communication and term usage patterns. I found it a very interesting and provoking line of reasoning.

The book is 311 pages, which includes an 11 page index and 11 pages of notes. The main part of the book is divided into 11 chapters (what is with the 11 thing?), plus an epilogue and acknowledgements sections (only 8 pages total, not 11). The first two chapters are an overview of the issues and influences of Web search. Chapter 3 is a review of pre-Google search on the Web—yes, it existed. Chapters 4–11 trace Web search as Google develops as an organization, a technology, and a Web brand.

Overall, the book is very well organized and informative. The author does a good job of not making the book "just about Google"; rather, the author places Google within the progression of Web search. There is ample discussion of many of the other "pre-now" search engines, such as AltaVista, Lycos, and Excite. The book discusses Yahoo! and acknowledges Yahoo! as a major Web force, but the author seems to put Yahoo! in this "distant place" for most of the book. I found this interesting.

During the seemingly endless string of "hot new search engines"—Lycos, Hotbot, Excite, NorthernLight, Infoseek, AlltheWeb.com, AltaVista, plus maybe a dozen other ones—that have come and gone, Yahoo! has watched the competition come and go, serves relevant results as good as any other engine in the field (c.f., Jansen & Molina, 2006), has always been a major player, and is still a major player today. Now, that is the interesting story of Web search! How has Yahoo! done it for so long when so many others have failed?

There are several inline references to both private Web research companies' studies and academic articles, although these are weighted to the articles from search engine researchers and not university academics. I found the 89 endnotes in the notes section especially interesting. These endnotes added a lot of value to the book. The ample combination of both references and endnotes provide an additional degree of legitimacy many times not seen in popular press books.

The book reads as the author's journey through this investigation. One feels the author's presence throughout the book. There are a handful of "*Hey, look at me! Ain I great?*" comments. However, the statements are never over the top. So, I found these occasional remarks more humorous than irritating, thus, they added to the pleasurable read rather than distracting from it.

I enjoyed the book. Web search is my research area, so I expected the book to be just a confirmation of what I already knew. However, I learned a lot! It is a must read for any researcher or practitioner of Web information searching or information retrieval.

References

- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1), 107–117.
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using Web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11), 87–92.
- Jansen, B. J., Jansen, K., & Spink, A. (2005). Online job seeking and recruiting. *Internet Research—Electronic Networking Applications and Policy*, 15(1), 49–66.
- Jansen, B. J., & Molina, P. (2006). The effectiveness of Web search engines for retrieving relevant ecommerce links. *Information Processing & Management*, 42(4), 1075–1098.
- Lesk, M. (1997). *Practical digital libraries: books, bytes, and bucks*. San Francisco, CA, USA: Morgan Kaufmann.

Bernard J. Jansen
*School of Information Sciences and Technology,
The Pennsylvania State University,
329F Information Sciences and Technology Building,
University Park, PA 16802,
USA*

Tel.: +1 814 865 6459; fax: +1 814 865 6426

E-mail address: jjansen@acm.org

URL: http://ist.psu.edu/faculty_pages/jjansen

Blog: <http://jimjansen.blogspot.com>

Available online 28 February 2006