

Them! Google's Ambivalence toward Library and Information Science

by Shawne D. Miksa

In the classic science fiction film *Them!* [1] the forces of good (sturdy law enforcement and geeky science) seek out and destroy the forces of evil (gigantic mutant ants) that have resulted from foolish tampering in 1945 by the forces of good with the forces of nature (atomic energy) in the desert. In a key scene, the lone survivor of an ant attack is shocked out of her comatose state via a whiff of formic acid (essence of ant) and shrieks “Them! Them! Them! Them!” Given the criticism of Google by some in our field it is quite easy to imagine librarians as the comatose victims shrieking in terror as they relive their encounter with a gigantic monster on the information frontier. In the movie, the forces of good come to realize that, while the ants may be mutant monsters, their natural instinct to band together, propagate, spread out and forage for food remains unchanged. They have no plan to replace humans – they simply do what they do. Despite this insight, it is clear the ants must be destroyed for sake of humankind’s survival.

It is a bit extreme to paint librarians and Google locked in the same battle, but the questionable characterization by some librarians of Google as a mutant technology lamentably exists alongside those who admire it for its many innovations. From all accounts, Google has no interest in replacing libraries. They express [2] a great appreciation of libraries. “Even before we started Google, we dreamed of making the incredible breadth of information that librarians so lovingly organize searchable online,” said Larry Page, Google co-founder and president of products. However, their awareness of the library and information science (LIS) beyond that of supplying people

to stock and staff book storehouses is vague at best. Their notice of the field manifests itself much like the ants, who only bother with the humans when they are presented as a possible food source.

In order to understand this ambivalence it may help to contrast Google’s search engine with traditional library classification and to understand the original intentions of Google’s creators, Brin and Page. There is a limit to what we can find out about Google’s search engine by delving into the resources about the company and its technology, but what is there provides interesting food for thought.

“Google,” John Battelle [3, p. 37] writes, “is currently our culture’s grandest declaration of the power of search – but it is by no means the first.” In his description of the development of search engines, Battelle relates the motivations and actions behind the rise, and sometimes downfall, of well-known search engines such as Google, Alta Vista, Excite and Yahoo!. In particular, he lovingly expounds on the “Database of Intentions,” which he declares to be “...the most lasting, ponderous and significant culture artifact in the history of humankind...the aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result.” (p. 6) The overall focus of Battelle’s book is search (that is, web traffic), and in particular, paid search, which some librarians may see as an uncomfortable contradiction to, say, the *ALA Library Bill of Rights* [4]. We don’t sell access to our users, although we may all be secretly envious of the perks enjoyed by those who do (for instance, Google reported revenues of \$3.87 billion for the quarter ending on June 30, 2007 [5]). According to the Google Book Search Help Center FAQ on the Google website [6], “There are currently no ads on books from the Library Project.” However, there are links to online books stores, but no evidence of linkage for donations to library funding.

Shawne D. Miksa is assistant professor in the School of Library and Information Sciences at University of North Texas. She can be reached by email at smiksa@unt.edu

This simple division of purposes is quite central to the argument of the pros and cons of Google. Sandler [7] calls Google a “disruptive technology” whose “catalyzing changes are already underway.” (p. 21) He asks, “How can we expect libraries to maintain relevance in a changed world?” (p. 21) Anderson [8] echoes this theme as he chides librarians for not taking more seriously the effect Google’s services have on the role of the traditional library. Brewster Kahle of the Open Content Alliance (OCA) contends that the “idea of making all books accessible online in new and different ways is all good news. But if you do this in a way that the materials that have been housed in libraries for centuries are made available only through one corporate interface, this is an Orwellian future.” [9] Jeanneney [10, p.71] cautions that “accessibility to everything without Ariadne’s thread to guide our curiosity may cause us to lose our way.”

Motivations Driving Google’s Work

There are plenty of resources recounting the story of Google’s origins, including many studies of the accuracy and efficiency of its growing stable of information tools (for example, comparing Google Scholar with Web of Science is currently quite popular). An interesting fact is that Brin and Page used the very simple premise of citation analysis (with the proper accolades to Eugene Garfield and citation indexing) as the basis for PageRank, the primary Google algorithm that employs the hyper-linking structure of the web to calculate the ranking of websites. (Depending on what corner in LIS one inhabits this origin may be common knowledge or a surprisingly unknown fact.) There still lingers, however, the disconnect between LIS and computer science. Langville and Meyer [11] preface their exploration of the science of search engines, including PageRank, with a short history of information retrieval and compile the following list:

Notable artifacts that belong in our information retrieval museum are a few lists of individual library holdings, sorted by title and also author, as well as examples of the Dewey decimal system (1872), the card catalog (early 1900s), microfilm (1930s), and the MARC (Machine Readable Cataloging) system (1960s). (p. 2)

An exploration of the About Google web pages, in addition to external literature written about Google, reveals that perhaps the search empire is still conflicted (or ignorant) as to whether its technology provides meaningful searches with context based on ranking or just ranking. Battelle reports “the Google service made no pretensions of actually reading a particular site, or of understanding its content. It simply laid bare the often ugly truth of how well connected a site happened to be.” (p. 80). Google’s Company Overview page [12], however, states that “Google’s mission is to organize the world’s information and make it universally accessible and useful.” We can add an exception – it is universally accessible except for where censored by certain countries that use Google (for instance, China), or when Google censors itself [13].

Ranking does provide a type of organization, but the question remains as to whether PageRank also gives a sense of context and relevance before users obtain the material or if just the fact that they obtain it is enough of a justification to consider a search successful. Brin and Page [14] write that, “our notion of ‘relevant’ is to include the very best documents since there may be tens of thousands of slightly relevant documents. This very high precision is important even at the expense of recall.” (p. 108-109) However, Milner [15] suggests

If we had a computable theory of meaning then it should be possible to get a document, such as an email, summarized by software. Currently, it cannot be done. Even Google has to do the next best thing and show you passages from the document, which may match your search. (p.193)

Library classification is one process of information organization that does not work based on the concept of popularity as employed by Google. Most classification systems are based on a concept of subject, in which like subjects are grouped together, while unlike subjects are not. Assignment of classification numbers is based on subject analysis, or content analysis, powered by human judgment as to what something is about and where in a finite collection of information objects it should be placed in relation to everything else in the collection. A classifier would thus assign a classification number to a web page based on analysis of its content.

By contrast PageRank's main thesis, as described by Langville and Meyer, is that "a page is important if it is pointed to by other important pages." (p.31) In addition, Google tells us [16] the following

[The engine] then conducts hypertext-matching analysis to determine which pages are relevant to the specific search being conducted. By combining overall importance and query-specific relevance, Google is able to put the most relevant and reliable results first.

Google's purpose and many projects have inspired both praise and criticism, some of which carries equal ambivalence. One of the most recent, mentioned above, is a work entitled *Google and the Myth of Universal Knowledge* [10] by Jean-Noël Jeanneney, president of the Bibliothèque nationale de France. Jeanneney writes that "behind its majesty, Google is hiding frailties, like any company founded on a (currently, at least) single, albeit profuse, activity, without any parent corporation or other major body to lean on." (p. 62) He is especially critical of the Google Book Search because of potential biases toward English-only literature, as well as biases in ranking (what he terms the "gondola end"). Jeanneney asserts the following:

Due to the criteria of frequency and density of links, the pages most often recognized by the engine will in turn be more easily called up by the other users clicking on the links, and we can be sure (thanks to the principle of lending only to the rich) that the pages that are already overwhelmingly "selected" will continue to be so. (p. 45)

As a countermeasure, Jeanneney suggests that a European algorithm be defined. Google Book Search is now available in nine languages, according

to the Google website. Despite that statement, there isn't a "language" search feature visible on the Book Search interface, although searching with non-English terms will net some non-English books.

In their 2005 "Letter from the Founders," [13] Brin and Page discussed the negative reaction to their Book Search. They write, "We believe one of the greatest services we can provide to users around the world is to increase people's access to human knowledge" and that Google respects copyright and does not have permission to break copyright on any work. Their reasoning for the criticism from authors and publishers is that the "...transition to the online world is a huge change, and one they understandably view with some trepidation." In the 2006 letter [17], the founders state that "much of the highest quality information in the world may be found in tens of millions of books tucked away in libraries and on publishers' shelves. These books can be tremendous assets – but only if people know that they exist." From a librarian's perspective there is an underlying tone of disregard for the purpose of libraries prior to Internet search engines and the work in bibliographic control that is currently practiced, hence the resentment and concerns over competition between libraries and Google. In actuality Google [18] states, "We consider our primary competitors to be Microsoft and Yahoo."

If libraries were to suddenly develop a business plan for making money then we might one day make it to the competitors list. Until then, it would be beneficial to continue analysis of Google's information tools and provide the type of guidance to information users that has been the mainstay of our field. ■

Resources Mentioned in the Text

- [1] Weisbert, D. (producer). (1954). *Them!* [motion picture] United States: Warner Bros. Pictures.
- [2] Google.com. (2006). Google checks out library books. Retrieved August, 15, 2007, from www.google.com/press/pressrel/print_library.html
- [3] Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York: Portfolio.
- [4] American Library Association. (1996). *The library bill of rights*. Retrieved August 15, 2007, from www.ala.org/ala/oif/statementspols/statementsif/librarybillrights.htm
- [5] Google.com. (2007). Google announces second quarter 2007 results. Retrieved August 15, 2007, from <http://investor.google.com/releases/2007Q2.html>

Resources, continued

- [6] Google Book Search Help Center FAQ. (2007). How are Library Project books displayed? Are there ads?" Retrieved August 15, 2007, from <http://books.google.com/support/bin/answer.py?answer=43742&topic=9082>
- [7] Sandler, M. (2005). Disruptive beneficence: The Google Print Program and the future of libraries. *Internet Reference Services Quarterly*, 10(3/4). Issue published simultaneously as *Libraries and Google*. Binghamton, NY: Haworth.
- [8] Anderson, R. (2005). The (uncertain) future of libraries in a Google world: Sounding an alarm. *Internet Reference Services Quarterly*, 10(3/4). Issue published simultaneously as *Libraries and Google*. Binghamton, NY: Haworth.
- [9] Albanese, A.R. (2007). Scan this book! In the race to digitize the public domain, is the future of the library at stake? An interview with the Open Content Alliance's Brewster Kahle. *Library Journal Online*. Retrieved August 15, 2007, from www.libraryjournal.com/article/CA6466634.html?industryid=47175
- [10] Jeanneney, N. (2007). *Google and the myth of universal knowledge*. Tr. Teresa Lavender Fagan. Chicago: University of Chicago Press.
- [11] Langville, A. & Meyer, C. (2006). *Google's PageRank and beyond: The science of search engine rankings*. Princeton, NJ: Princeton University Press.
- [12] Google.com. (2007). *Company overview*. Retrieved August 15, 2007 from www.google.com/corporate/index.html
- [13] Brin, S. & Page, L. (2005) Letters from the founders. Retrieved August 15, 2007 from http://investor.google.com/2005_founders_letter.html
- [14] Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30, 107-117.
- [15] Milner, B. (2004). *Google and the mission to map meaning and make money*. London: Electric Book Co.
- [16] Google.com. (n.d.). Technical overview. Retrieved August 15, 2007, from www.google.com/corporate/tech.html
- [17] Brin, S. & Page, L. (2006) Letters from the founders. Retrieved August 15, 2007, from http://investor.google.com/2006_founders_letter.html
- [18] Google.com. (2007). Investor FAQ. Retrieved August 15, 2007, from <http://investor.google.com/faq.html#competitors>