November 2013

People Profile: Bonnie Klein

Editor

Follow this and additional works at: http://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Recommended Citation
DOI: https://doi.org/10.7771/2380-176X.2746

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Like the earlier E-Government Act of 2002 (P.L. 107-347), the new bill assigns the responsibility for policy, guidance and oversight to the Office of Management and Budget (OMB), Office of Information and Regulatory Affairs (OIRA). In my opinion, the current policy in OMB Circular A-130 “Management of Federal Government Information” already covers the search capabilities provision by directing agencies to “use electronic media and formats, including public networks, as appropriate and within budgetary constraints, in order to make government information more easily accessible and useful to the public.”

At the December 11, 2007 Senate Committee hearing on “E-Government 2.0: Improving Innovation, Collaboration, and Access,” Karen Evans, Administrator of the Office of Electronic Government and Information Technology at OMB, reported on the progress the government has made in getting services and information online and available to citizens. One avenue is USAGov, the official U.S. Government Internet portal and centralized point of entry for locating government information, benefits, and services. In FY 2007, USAGov received approximately 97 million visits during the year or 1.87 million visits per week.

At the same hearing, John Lewis Needham, Google’s Manager for Public Sector Content Partnerships, testified that: “The government produces a lot of information and these databases cannot be navigated by Web crawlers.” Needham correctly stated that the most prevalent technical barriers to search engine access to “deep Web” government information are: (1) agency use of dynamic query-based databases, (2) Robots.txt. files that prevent crawling and (3) outdated links.

Needham also opined that “Agencies are concerned more about how information is presented than if users are finding it.” The fact is that agencies are concerned about both. To meet reporting requirements and scorecards, Government agencies want the searching public to readily discover, recognize, and choose the agency as their preferred trusted and authoritative information provider.

Rumors from page 26

suit? In fact, I was in some airport or other the other day and walked into a bookstore. I noticed right away that all the books were displayed “face out.” I don’t know if I like it or not, you see – “Borders tries about-face in shelf space display.” by Jeffrey A. Trachtenberg, The Wall Street Journal, 3/12/2008. yordispatch.inyork.com/yd/business/ci_8527433 online.wsj.com/article/SB120528554163329185.html?mod=yahoo hs continued on page 36

Seek and Ye Shall Find?
The premise of the proposed legislation is that if agencies make their data searchable, it will be indexed and discoverable. Hear ye, citizens, seek and ye shall find. Well, maybe. It depends on where you search, what you are searching for, and how you are searching.

Most search engine users expect and accept that they must sort through a large amount of material, much of it irrelevant to their purpose. To aid users in narrowing results, Google and other search engines offer options that limit a search to material types such as images, video, maps, news, and books or by specific interest groups such as Scholar and Finance. Google also offers a U.S. Government option that searches U.S. federal, state and local government domains and sites; but this option resides under “Special Searches” and is not readily apparent to most Google users.

If agencies do apply sitemap or another indexing standard, will search engines rank the federal government information higher in search results? The answer is “No.” Google states in its Public Sector Frequently Asked Questions (FAQ) that “implementing sitemaps does not affect the ranking of a Webpage in search results.”

The answer to the FAQ “What pages will Google index? Will they appear in Google.com or Google’s US Government Search?” is both a disclaimer and business policy. Google “can’t guarantee that we’ll include all pages that we crawl on your agency’s Website in our index. However, we’ll include all pages we believe are relevant to our users, so that they appear in search results of Google.com and Google’s US Government Search, as well as other Google services.”

Instead Google assesses relevancy based on its PageRank technology. Donna Bogatin in her January 26th, 2007 ZDNet post “Google search PageRank excludes relevant Websites” observes that “By requiring that Web pages have inbound links from third-party Web sites, the PageRank based algorithm may result in automatic exclusion of the most relevant pages for a given query simply because no other Websites have linked to them.” You’ll have to take it on faith, but there is a lot of esoteric and eclectic government information that only a few, if any, would seek or need to find.

We also need to keep in mind that Google and other search engines are commercial entities, not public utilities. Barbara Frist’s description of the search engine business model is: “Google gets content for free, gives it away for free, and makes its money by being an enormous distribution channel for everything from physics research to 19th century scanned books to the latest YouTube video.” Content is a means to an end. In 2007, Google had 57% of the market share and reported 4th quarter revenue of $4.83 billion, a 50% increase over 2006. AdSense revenue increased 30%, amounting to $1.45 billion of the total. Business operations and revenue-generating advertising partnerships, not altruism, factor into page ranking.

As I said earlier, when federal agencies have taken the initiative to open deep Web databases, commercial search engines do not always rank the government-source content above commercial or for-fee suppliers. The page rank depends on what, where and how one searches. I offer the experience of my agency as an example.

continued on page 34