

Libraries and the Future of Search

Judith O'Dell

Central Michigan University Libraries

Mount Pleasant, Michigan

Introduction

Libraries have begun experimenting with search engines that use the Google model for their catalog or other internally-produced systems. Some librarians see this as “giving in” to a younger generation, and believe that, because it produces results but requires very little understanding of the systems used for organizing information, the results are therefore of low quality. Libraries have traditionally created records classifying and describing sources of information which are then organized into a system used to retrieve or locate those sources. Google, in contrast, is focused on creating a “smart” search engine that in theory would alter or eliminate the need to organize records.

Eric Schmidt, CEO of Google, has stated that the original mission of Google was to “make all the world's information universally useful and accessible” (Schmidt 1). When asked where the future of search is going, Schmidt said that the user should be able to ask a question and get a single answer, which would also be the right answer in their language. He stated that to do so the system needs to know what you meant, which “ultimately means understanding how you think,” and both depend on knowing a lot about you (Schmidt 6). Schmidt admitted this would take longer than his lifetime but presented this as the challenge for Google to conquer.

What is being described is essentially a smart search engine; one that knows the wants, needs and thinking of the user. Assuming such a thing can be created, a question is what it will mean for the future of searching in libraries, and whether it will negate the need to organize and classify and replace those processes with a system that essentially catalogs and classifies the user. Users might no longer need a basic understanding of the organization of information because that capacity will be built into the search engine's data or text mining capabilities.

The Smart Search Engine

To make this goal a reality, there are technical, business, social, and legal issues to be addressed; some of them overlap. Technical issues include questions about the feasibility of creating such a system. Business issues include competition from others such as Microsoft or Yahoo, who might reach the goal first or move searching in a different direction. Social issues include the public's willingness to accept changes in societal values integral to use of the new technology. The most obvious legal issue is the question of privacy. Looking at the current direction of these concerns might provide an insight into the future. Unless otherwise stated, the word “search,” is used in this article to mean the process of developing a query, conducting a search using that query, processing the query, and receiving a response.

The primary technical question is whether the technology will be available to create this “smart” search engine. There are many individuals and organizations working on ways to optimize the search experience by enhancing the way a search engine captures, stores, and retrieves content. In essence,

they are trying to improve relevance, the ability to effectively match content with the user's needs. And, the determination of relevance is central to the concept of "getting the right answer." But determining what is relevant is altered by one's needs, interests, and context (Yu, et al. Introduction 1). The ability of a search engine to retrieve relevant information can be determined by looking at recall (the ability to find all the relevant information available), precision (the proportion of results relevant to the search), and ranking (the degree to which a result set is ordered as to list the most relevant information first) (Yu et al. Introduction 1). What are developers doing to improve relevance?

There are different ways to group or define the characteristics of search engines under development; some are listed below:

- level of automation: a continuum between a completely manual system and a completely automated system
- types of software
- open source: non-priority software where users have access to the source code (Breeding 27)
- proprietary: the source code is owned by an individual or organization (Breeding 27)
- single source: searches one source
- federated: searching more than one source with a single search query
- types of content management or document management systems
- enterprise information
- scholarly information
- web
- broad-based/horizontal search: software directed at indexing a generic list of websites
- vertical search: software directed at indexing websites that are focused on a specific topic
- types of data
- structured: having a consistent format
- unstructured: having varied formats
- visual
- textual
- numerical
- metadata: structured data describing content
- full text
- types of taxonomies
- expert: predefined sets of terms based on the knowledge of the creator/s.
- user based: creating identifiers or tags based on the number of times that term is used by searchers.
- clustering: developing taxonomies using automated systems.
- hierarchical clustering: use of ascending or descending methods of putting content into ranked groups through a series of steps.
- partitional clustering: all clusters are determined in one step.
- conceptual clustering: assigns descriptive information to groups.
- analysis: processes that identify useful information within content
- word search: matching keywords
- natural language processing: includes all attempts to use computers to process language used by humans (Coxhead 1)
- mining
- data mining: "...the process of analyzing large data sets using statistical, pattern recognition, and knowledge discovery techniques to determine meaningful... trends and information" (Inniss, et al. 8).
- text mining/text analytics/information extraction: "...extraction ...of patterns, useful information or knowledge from natural language text," (8 Innis et al. 8) and "...is designed to handle structured data from databases or XML files, working with unstructured or semi-structured data sets..." (Fan, et al. 78)

- topic tracking: predicting useful information by matching it with previously used sources listed in user profiles (Fan 78)

In practice these categories are not mutually exclusive. A given system will often feature two or more of the views described and some categories amount to different ways of looking at the same problem. An example is clustering, where the product is a taxonomy or index but the process is a type of text analysis (Fan 78). Examples of products being developed are listed below.

One effort includes the development of a community-based, personalized web search system that defines relevance as the preferences of search communities. Single searches are placed in a group or community along with other searches that represent like interests. It is a collaborative process where preferences are defined as queries plus user selections from results lists chosen by members of communities of interest, which when collected and grouped form the basis of community expertise. Earlier page selections are ranked based on their relevance to queries that are similar to the current query. Using this information, certain pages are promoted as likely selections that would interest the current searcher because they were of interest to other users with similar interests. This system works with user-generated content, grouping it to create social networks and communities of interest. The selection of previous individual community members, as a group, helps define the context of words in a new query, a limiting factor that refines the search process used by generic Internet search engines. When combined, these community preferences are viewed as a community's expertise (Smyth).

Current Initiatives

ChoiceStream has developed their "RealRelevance" platform which has personalized the process of selecting entertainment options, such as movies, music and sponsored links. ChoiceStream uses this system as a basis as a way of personalizing search (Patel 1). Their approach to personalization is the Attributed Bayesian Choice Modeling (ABCM). It classifies web pages using attributes related to each content domain, rather than by user preferences. An example of a content domain is television programs. The attributes of each content domain are identified and each item is classified based on those attributes. In addition, a user profile is created using multiple methods of input. The result is a profile of user preferences for the various attributes. ChoiceStream views their system as an advance because it can produce results with less data, since it is not necessary to retain large databases containing the history of user ratings or selections (Review of Personalization Technologies).

A current focus of many search engine builders is the creation of vertical search engines that focus on a niche market by including only one topic. These are growing in number but do not appear to have a fundamentally new way of searching. They simply eliminate a major problem of broad-based search engines, which bring together different types of data, by narrowing the topics included, which in turn narrows the types of data to be integrated. There are other search engine producers that are tackling the problem of diverse content by focusing on conceptual search.

Text Analytics is an approach being used to conquer the problems of conceptual search. This approach automates the categorization of content by using sample text or documents as a starting point. Using these samples, the content is analyzed for concepts that are used to create the categories, after which further sources are analyzed for concepts that relate to the original sample. Concepts are defined by their proximity within a defined conceptual space. In essence, words are translated into concepts based on the other words or other content near them (Challenges We Solve).

While no organization is close to creating the search engine described by Eric Schmidt, progress is being made in this direction. What is not clear is if the goal will be achieved and whether Google or any other group will achieve the goal. The fact that so many organizations are involved suggests that the end product will be the result of the efforts of many organizations, not just one. An advantage to having the search engine do the work is that all information can be placed in a single system; individual organizations will not have to create their own classification or records.

A lot of effort being put into personalized search. Personalizing search places emphasis on what the user is asking for from the users' perspective, including the question of what a user means by a search request. Answering this question is the goal that Eric Schmidt set for Google. The reason most often given for personalizing search is that it empowers the user. The user can interpret relationships as he or she sees them, without having to learn the logic of experts, subject specialists, or librarians.

To accomplish this goal, the search engine must be able to create personal ontologies, which in turn requires the ability to understand the way a user thinks. This implies the ability to interpret concepts. Conceptual searching, natural language processing systems, and artificial intelligence are much more difficult than originally thought and no radically new ideas are on the horizon (Yu, et al. Inside 1). The key may be the right algorithm or a future technology, or both. But, lacking this ability, current research seems focused on building a better keyword search by creating a better system for determining relevancy. If Eric Schmidt's goal is to become a reality, one needs to resolve the problem of conceptual searching (Yu, et al. Inside 1), or develop a system that works around the problem.

Library Search

Most libraries have an electronic version of the card catalog. When the catalog was originally digitized, it was thought that searching would be significantly easier because of the power of the search engine to scan the material quickly and scan the entire record. The card catalog is different than the online catalog in at least two ways.

In the past, the library was a primary collector of information and people used the library as a primary source for finding information. The card catalog was intended to reflect the materials owned by or located in a specific place or within a specific collection. While OPACs still reflect information from a specific collection, the Web contains limitless amounts of information from a limitless number of sources, in a variety of formats. Some Web sources are a part of many OPACs. The card catalog was limited in scope and existed in a contained environment. OPACs are used in an environment that does not operate by the same rules. Finding information effectively in the digitized world requires the user to understand multiple systems for retrieving information.

A second difference is that a user could see the card catalog while searching through it, and could therefore get at least a basic sense of how it was put together. As information structures have gone behind the scene, the user becomes less conscious of their existence. With the online catalog, the user is looking at a search box with little sense of how the information behind the box is structured. This is not a problem for those knowledgeable about the ways information is organized, but could be quite confusing to the uninitiated user. While it is true that anyone can do at least a basic keyword search, the more sophisticated ways of using the catalog require some knowledge of how it is constructed.

One survey showed that "generation Y," who are ages 18-29, are the heaviest users of libraries when they need problem-solving information, but most people turn to the Internet to begin their search (Estabrook and Rainie, summary iii). In his book, *Illlicit*, Moises Naim stresses the idea that the Internet allows users to bypass traditional structures in which connections are made and make connections of their choice. The library is no longer the primary locale for those seeking information. This is demonstrated by the fact that about 80 percent of students start their search for information to complete assignments somewhere other than a library website (Marcum 6-7).

If users eventually find their way to a library site, they are confronted with a tool that is constructed differently from other online tools. Studies have found that most users have an understanding of how to effectively use any online search tool (Lewandowski 141-142), but a Web search engine is more likely to retrieve an answer than the list of resources on a library website, if used ineffectively. This is due, in part, to the fact that when using a general search engine, one does not have to go through multiple divisions (i.e. the catalog, databases, etc.) or multiple layers of classification of sources to retrieve a list of answers from the entire collection of sources.

With most Web search engines, the user types words in a single search box. According to Todd Miller, “the paradox demonstrated so elegantly by Google is that the most powerful information access approach also happens to be the simplest and easiest. The most complex and least intuitive interfaces wind up securing information, not facilitating information access (1-2).” Dirk Lewandowski states that users generally only look at the first page of results from a given search. In essence, users only look at results found without scrolling (142). Given this fact, is it likely that users will be willing to start their search at a library website when it requires that they wade through divisions and layers to find information? Why should the users learn the more difficult system, which requires an understanding of thesaurus terms, and which can often only be used to retrieve information from that specific database or aggregator's databases.

Even librarians are becoming overwhelmed by the need to search multiple platforms. Federated search engines are presented as a solution to searching multiple databases, but federated searching has its own problems, such as authentication, de-duplication, and searching various platforms. (The Truth about Federated Searching)

Thesauri are more precise, but that may mean that search engines have not solved the problem of conceptualized searching or automated classification, and not that taxonomies and subject headings devised by experts are simply better. Automation of true natural language processing might eliminate the need for thesaurus terms used to manually classify or catalog. Given the growth of information, it may not even be realistic to continue cataloging and classification for all the information sources being produced.

An automated system for creating metadata has been proposed as a solution to cataloging and classifying all information sources. This assumes that detailed metadata is needed for all sources, even those that are online and full text. Deanna B. Marcum raised the question “do we need to provide detailed cataloging information for ... digitized materials? Or can we think of Google as the catalog (Marcum 7)?” Others see a combined system of creating records, using both automated and manual input (Dutra 3). Some have questioned the use of hierarchical structures to organize records when there are technologies such as mapping or other visual imaging tools. Some have even suggested that the use of text will be diminished as visual and voice technologies improve. EBSCO is an example of a database aggregator that is moving to the use of visual search. Such changes will help empower individuals who have learning styles incompatible with reading text.

Users do not want to learn multiple information systems or systems they find structurally difficult to use, and in many cases do not want to understand even one system. A study of information-seeking behavior conducted at the University of Idaho Library found that “time” was most frequently mentioned by undergraduates, graduates and faculty as a concern in finding information. That expresses a desire for a system that is easy to use and understands them (Weiler 50).

Empowering the user is an important purpose of technology. The University of Idaho Library study asked participants to describe a, “dream information machine.” Common answers were

- mind reader
- intuitive, and could determine their information needs without them having to verbalize them.
- one-stop source for information needs, using voice recognition and natural language to search to return a comprehensive collection of information sources
- portability
- ubiquitous access 24/7

The authors state that they find it difficult to imagine this type of machine being built (Weiler 49-50), but this is in large part the search engine described by Eric Schmidt.

Catalogs empower librarians, who understand how to use them. The hierarchical structure of the catalog is both its strength and its weakness. The subject terms bring a consistency that improves

precision. The use of subject headings relies on thesaurus terms, carefully constructed by librarians and subject specialists, which most closely reflect the state of knowledge or the structure of various disciplines.

While academic libraries are likely to have some users who are well-informed about particular disciplines, this may not apply to new users, first- and second-year undergraduates who have not studied in depth. Aside from whether users understand how to use library tools, there is the question of the extent to which the tools are used. In 1984, before most bibliographic tools were stored on computers, Stephen K. Stoan reported on a series of studies that indicated the majority of faculty do not use what he calls "access tools" as their starting point for doing research, but rather use footnotes and bibliographies from the primary literature (books and journals) (Stoan 100-101). Citations found in context of the subject being searched are more useful than descriptors filtered through another layer of human logic (Stoan 103). The catalog and serials lists then become locator tools (Stoan 104).

Some things are not easily classified and forcing them into a group alters the way they are viewed. For example, is the tomato a fruit or a vegetable? It depends on the perspective. A botanist will see it differently than a cook (Yu, et al. Structured Data 1). The Library of Congress rules for filing rules in card catalogs called for cards with subject headings to be arranged according to the hierarchical structure of concepts within each academic discipline. This method of filing, which required users to know that structure, was rejected by most libraries, particularly undergraduate libraries, in favor of the ALA filing rules, which used alphabetical arrangement.

If the expertly-organized catalog were replaced with a search engine capable of reproducing a user's worldview, index and search terms would change continuously according to the user's thought patterns. This might undermine the carefully-constructed view of disciplines, since conceptual relationships would also be subject to constant alteration. Rather than being taught the structure of a discipline as defined by experts, users would collectively redefine the content and parameters of the discipline, which would be reflected in the use of selected information, which is then combined with other data collected in communities of interest.

The strength of a search engine is its ability to search by topic. Today, much credible information is not being published in traditional sources that libraries are likely to collect. Institutes, foundations, and other organizations publish a lot of information, often on their websites (Naim 206). While libraries can create links to these sources in the catalogs, how much labor is involved in finding and creating records for these sources? Search engines are better designed to locate and bring together information from multiple sources because "ownership" is not a criterion for inclusion. The typical online catalog is not taking advantage of the technology.

North Carolina State University structures online searching of their catalog by topic, by allowing users to browse the catalog using the subject-based LC classification system through their Endeca software. More recently, others such as Aqua Browser and Primo create tag clouds that can be used to modify or limit keyword searches of the catalog. One similarity in all of these approaches is that search is done on a collection by collection basis. Endeca and Aqua Browser still use expert-designed taxonomies to structure the searches. while Primo and others have moved to user-designed taxonomies.

Expert taxonomies have consistency on their side but the world of search is moving to user-based taxonomies. User-based taxonomies may result in nonsensical results, whether intentional or not, e.g., [I Can Has Cheezburger](#), where viewers are invited to recaption pictures. To effectively search by topic, collections must be unified. Sources must be brought together so users search a single universe of information, e.g., the Google Library Project. One problem is deciding who should control this universe of information. Should it be a business enterprise motivated by business concerns? Many think not (Hafner 1).

Most search engines are still based on keyword searching, which limits their accuracy in interpreting human logic. To move close to the goal set for Google requires more. To understand a user, a system must understand how humans think.

What happens to librarians if an effective personalized system for searching is created? Teaching people to use tools for searching is the core of library instruction. Some have suggested that librarians move away from teaching people to use tools and become more involved in the development of those tools. This is what many early librarians did. The problems that occur when librarians are not involved in the development of search technology were demonstrated during the introduction of Univac at the 1962 Seattle World's Fair. An exhibition ostensibly showed the library of the future, but hardware vendors had little understanding of library processes. As a result, the exhibition demonstrated novelties rather than substantive ways that a computer might truly be useful (Downey).

Librarians could be involved in the development of the technology, the management of the content, or the creation of content. Librarians are currently involved in developing content management systems, institutional repositories, digitizing projects, and blogs and similar resources. Tracy and Hayashi surveyed reference librarians about ways in which Information and Communication Technology (ICT) was affecting their positions as librarians. The conclusion states:

The academic library's heightened reliance on information and communication technology challenges librarians' status and role as knowledge workers by altering workplace tasks and the overall conditions under which the librarians labor. ...Responses to our survey suggest how, unlike the frequently optimistic perspectives and predictions about technology that are found in the professional literature, librarians' experiences with and views toward ICT and how it relates to their profession are much more complex...

Scholars have contested hopeful forecasts about technology and work for some time, but librarians are equally concerned about how their profession and the public service mission of the library are being refashioned to match the prerogatives of ICT content and infrastructure providers (Tracey and Hayashi 65-66). The role of libraries and librarians as content providers is expressed in Google's goal, "to organize the world's information and make it universally accessible and useful." That role is not necessarily appropriate for a corporation, and is part of the mission of libraries (Vaidhyanathan 3)

Privacy

Determining how users think requires collecting information about user preferences, which brings us to the issue of privacy. What information will organizations be allowed to collect about their users? What will they be allowed to do with the information they collect? What liabilities will be attached to the misuse of that information? The answers to these questions will likely have an impact on the ability to build Google's intended search engine. As Eric Schmidt stated, the system needs to know what you meant and how you think and both depend on knowing a lot about you. How a system collects information about user preferences can determine to what extent privacy becomes a concern. Privacy is an accumulation of concerns and issues that include government intrusion into personal information, company policies directing what information is kept about individuals and who has access to the data, security of personal data and intrusion or misuse. What legislation, if any, will be needed to protect personal privacy? What technology is available to assure privacy? While attitudes about personal privacy are changing and people are more willing to divulge information about themselves, will this attitude of openness continue to grow and will it grow in all societies?

The legal system defines the rights and responsibilities of individuals and organizations. Neil M. Richards discusses the current state of privacy law in a review of *The Digital Person: Privacy and Technology in the Information Age*, by Daniel J. Solove. Richards uses Solove's book as a basis for discussing the Information Privacy Law Project, which "refers to a collective effort by a group of scholars¹ to identify a law of 'information privacy' and to establish information privacy law as a valid field of scholarly

inquiry (Richards 2).” Richards states that “despite the seemingly incessant attention paid to privacy, legal theorists and policymakers continue to have little idea just what our legal conception of ‘privacy’ is and to the extent there is a ‘law of privacy,’ it remains a piecemeal, poorly understood, and only partially successful body of jurisprudence.”² There is no clear definition of privacy law and by extension no clear definition of privacy.

According to Richards the fundamental problem is twofold. One, the scholars who are active in the “Information Privacy Law Project” insist on making a distinction between information privacy and decisional privacy. Information privacy is defined as the right to control information about one's self, drawing mainly upon tort law of privacy, privacy legislation and constitutional protections guaranteed by the first and fourth amendments. Decisional privacy is defined as the right to make decisions about one's personal and reproductive autonomy and is based on constitutional law (Richards 3). Richards believes that the insistence on dividing the two hampers the development of a unified law of privacy because both are about the autonomy of individuals (Richards 6,12). He states that “Such a broadening of the focus of privacy law more generally could also help information privacy scholars deal with some of the more intractable policy problems facing this area of the law, such as the copyright/privacy overlap, the problems of privacy in expressive personal records, and the tension between privacy rules and the First Amendment (Richards 6).”

The second problem Richard's identifies is the lack of identifying or unifying principles upon which the field of privacy law is to be based. “For new areas of scholarly inquiry like information privacy, it is essential to identify and establish the basic assumptions and claims of the field so that scholars participating in the movement have a shared intellectual space in which to work (Richards 10-11).”

Solove states that:

A distinctive domain of law relating to information privacy has been developing throughout the twentieth century. Although the law has made great strides in dealing with privacy problems, the law of information privacy has been severely hampered by the difficulties in formulating a compelling theory of privacy. The story of privacy law is a tale of changing technology and the law's struggle to respond in effective ways. Information privacy law consists of a mosaic of various types of law: tort law, constitutional law, federal and state statutory law, evidentiary privileges, property law, and contract law (Solove 56).

If information privacy and decisional privacy are combined into a single theory of privacy law, the interpretation of one may affect how the other is interpreted and determine the legal authority for future decisions. We do know if there is a developing theory of information privacy, and this area is likely to set precedents affecting the flow of information. Privacy as a unified concept has not yet taken shape and it is unknown how privacy will be regulated, but it is an issue of significant concern to many individuals and organizations, with laws in the process of being developed.

The Patriot Act of 2001 (Jaeger, Bertot, and McClure 295-296) brought libraries into the forefront of the debate over government intrusion into personal records. The Patriot Act amended many other acts relating to information policy, but the Foreign Intelligence Surveillance Act (FISA) was most changed (Jaeger, Bertot, and McClure 296, 298). FISA was passed in 1978 and clearly defined investigative conduct as it related to criminal investigation and foreign intelligence investigations (Jaeger, Bertot, and McClure 297). Jaeger identifies seven significant ways that the Patriot Act alters FISA as it relates to the collection and analysis of personal information; four of these affect libraries. The Patriot Act allowed court orders to be obtained for surveillance that would not have been allowed under FISA and increased the types of records that can be searched, including library records. The 2001 Act contained a secrecy clause that prohibited the disclosure of information about the investigation and allowed information obtained during an investigation to be shared with other law enforcement organizations (Jaeger, Bertot, and McClure 298-300). The renewal of the Patriot Act in 2006 contained changes that made it more difficult to obtain records, allowing subpoena recipients to contact an attorney and challenge gag orders (Congress, Bush Renew Patriot Act 1).

David Vise notes that every time a search is done using Google, Google saves the search indefinitely. This is also true of emails. He points out concerns such as the possibility of government investigators gaining access to the data or of having the information subpoenaed (Google's Growth 3). John Battelle states that Google is doing this to make products and services better, using what they know about their users to target information to an individual's needs. The problem is that Google is a corporation and corporations, over time, change management and policies (Google's Growth 3-4). How long can we trust Google to follow their current policies? We only have to look at the willingness of Google and other industry leaders to censor information in the Chinese market (Kirchgaessner 1) and Yahoo's alleged provision of names to the Chinese government (Kopytoff C1,1-2) to see that policies are easily forgotten and business interests take precedence when pressured by government.

Jeffrey Brown notes that many people are concerned about [Google Earth](#), which provides satellite imagery and lets users focus in on specific addresses (Google's Growth 1). This has spawned competition from Microsoft with its [Live Search Maps](#). Mapping is a very useful service for most people who are legitimately trying to find a specific location, but what dangers exist from the public having easy access to locating a person's home or business address? Should we also be concerned about sites such as [FlightAware](#) that allow live tracking of private and commercial air traffic in the United States? Improvements that help us also have the potential to hurt us.

Government regulation of the Internet is an interesting issue itself. On June 27, 2005 the U. S. Supreme Court decided the case *Metro-Goldwyn-Mayer Studios inc., et al., Petitioners v. Grokster, Ltd., et al (Metro-Goldwyn-Mayer Studios)*. It held that online file-sharing services can be held liable for copyright violations if they encourage use of the file sharing software (Fiorino 1). But what if the file-sharing service being used is not a US site? Would only individuals in the US and possibly other countries with copyright protection be prevented from obtaining the downloads? Since the MGM Grokster decision, Australia, South Korea and Taiwan have ruled against operators who encouraged users to download copyrighted material (Lash 2).

Societies vary in their attitudes toward privacy, particularly attitudes toward specific issues. Monie Lee, in a study of direct marketing, noted a number of other studies supporting this. Milne, Beckman, and Taubman in 1996 and found that while 67 percent of US respondents found target marketing practices to be acceptable, only 16 percent of the Argentinean market agreed. Target marketing was defined as gathering background and past purchasing information on people (Lee 1). Lee notes that, "direct marketers assert that their interest in consumer information is only to better identify and serve their potential customers (Lee 2)." Like Google, direct marketers see themselves as serving the customer by retaining information, but may not see the enormous responsibility they accept for maintaining private information about the customer and how the customer might be harmed by the misuse of the information. From the perspective of connecting buyers and sellers, or in the case of Google helping people to locate information, they may well be helping consumers.

The issue of personal data is much larger than the ability to communicate. Google's recent decision to make search logs anonymous after a number of months suggests that they are becoming aware of such issues (Helft 1). Unfortunately, Facebook did not heed the warning when they gathered information about their users and tracked their activities, publishing this data without the user's permission. The intent was to present targeted ads to users (Musthaler 1). Professor Barry Smyth noted that the use of a community-based personalization system, rather than user-based personalization, can help obviate this concern, because individual information becomes anonymous within the community profile 3.

The actions of Google or its competitors depend on factors beyond their control. The information Google or any organization will be able to keep and use in the future will depend in large part on how information privacy is ultimately defined. Privacy law is a new field, and it will probably be a long time before there is a clear legal definition of privacy. It will be a long time before search engine producers have a clear notion of their rights and responsibilities regarding the information that flows through their systems.

Conclusion

Today we are making decisions about the direction of libraries with inadequate technologies and insufficient data about future trends. Steven J. Bell writes about this issue in, "Submit or Resist: Librarianship in the Age of Google" (Bell 68-71). The "googlelizers" prefer to give users what they want, while "resistors," as defined by Judy Luther, oppose the "dumbing down" of information systems (Bell 70). Many librarians believe that making library resources more like Google is the best way to bring users back to the library (Bell 68). This assumes that the primary reason users prefer Google is the simple search interface. The segmented and layered structure of library materials, which forces users to repeat their query multiple times and mine through levels of classification, is also a factor.

Bell is a resistor who believes that libraries should teach users how to differentiate between high quality research and that which is just good enough (Bell 68). With fewer users coming into the library building, there is a need for effective ways of educating users remotely. Methods include text or chat reference, online instruction, and integrating instruction in social networks such as Second Life and Facebook. There is disagreement on how effective these efforts have been and to what extent they should be used. But they are moving reference and instruction into the virtual world.

Bell sees Google as a competitor to libraries. He sees "googlelization" as "copy-your-competitor thinking" (Bell 70). Another view is that Google is trying to create the ultimate search experience by improving the quality and capabilities of the search engine. Libraries are trying to improve the search experience by making quality information available and organizing to make retrieval effective. Traditional structures may not be adaptable to the automated arena. They can be digitized, but they may not be the most effective way of searching in that arena. Google and libraries are working toward the same goal, approaching it from a different perspective. One future for librarians is to help create that ultimate search engine.

There are many players who will have an impact on the future definition and direction of information privacy. Privacy is often discussed, but is just beginning to come together as a theoretical concept. As such, it is difficult to identify its future direction. Because it has an impact on so many people, with implications directly connected to technology and how technology is used, it is likely that this issue will be at the forefront for many years to come.

Notes:

1 Richards refers to the following as the source for the list of scholars: Paul M. Schwartz & William M. Treanor, "The New Privacy," MICH. L. REV 101 (2003): 2163, 2177 & n.33 (identifying some of the major participants in this project).

2 Richards cites one of his earlier works: Neil M. Richards, "Reconciling Data Privacy and the First Amendment," UCLA L. REV 52 (2005): 1149, 1154-55.

3 This was a comment made during the presentation.

Works Cited

Bell, Steven J. "Submit or Resist: Librarianship in the Age of Google." *American Libraries* October 2005: 68-71.

Breeding, Marshall. "An Update on Open Source ILS." *Computers in Libraries* (Mar 2007): 27-29.

"Challenges We Solve." Content Analyst Company Home Page 2007, 5 March 2008. [1-7].
<http://www.contentanalyst.com/html/challenges/challenges.html>

"Libraries and the Future of Search," Judith O'Dell. *Library Philosophy and Practice* 2009 (May)

Coxhead, Peter. "An Introduction to Natural Language Processing." 2001: [1-8]. 8 March 2007. (Dr. Coxhead is a Senior Lecturer, College of Engineering & Physical Sciences School of Computer Science, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK p.coxhead@cs.bham.ac.uk) <http://www.cs.bham.ac.uk/~pxc/nlpa/AI-HO-IntroNLP.pdf>

"Congress, Bush Renew Patriot Act, Despite Continuing Opposition." *HR on Campus* 1 April 2006: 1. Legal News. LexisNexis Academic. Central Michigan University, 28 July 2006. <http://www.lexisnexis.com/>

Downey, Greg. "The Librarian and the Univac: Automation and Labor at the 1962 Seattle World's Fair." In *Knowledge Workers in the Information Society*. Ed. Catherine McKercher and Vincent Mosco. New York: Lexington Books, 2007. 37-52.

Dutra, Jayne. "Enterprise Search: Rethinking it in a Web 2.0 World." *FreePint*: feature article 29 November 2007: [1-5]. <http://www.freepint.com/issues/291107.htm>.

Estabrook, Leigh and Lee Rainie. "Information Searches That Solve Problems; How People Use the Internet, Libraries, and Government Agencies When They Need Help," Pew Internet & American Life Project and the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, executive summary 30 Dec. 2007: i-x. http://www.pewinternet.org/report_display.asp?r=231.

Fan, Weiguo, et al. "Tapping the Power of Text Mining." *Communications of the ACM* 49.9 (September 2006) 77-82. *The ACM Digital Library*, 3 December 2008. <http://portal.acm.org/citation.cfm?id=1151030.1151032&coll=GUIDE&dl=ACM&CFID=13213934&CFTOKEN=37615964>

Fiorino, Larry. "Commentary: Supreme Court says File-sharing Firms are on Hook for Songs, Movies." *The Daily Record* 1 July 2005: [1-2]. Legal News. LexisNexis Academic. Central Michigan University, 5 July 2006 <http://www.lexisnexis.com/>.

"FlightAware," 2008, *FlightAware*, 20 November 2008 <http://flightaware.com/>.

"Google Earth," 2008, Google, 19 November 2008 <http://earth.google.com/>.

"Google's Growth." *NewsHour with Jim Lehrer* PBS 30 Nov. 2005 News Transcripts. LexisNexis Academic. Central Michigan University, 15 December 2005: [1-4]. <http://www.lexisnexis.com/>.

Hafner, Katie. "Libraries Shun Deals to Place books on Web." *New York Times* 22 October 2007: [1-4]. *New York Times Lexis-Nexis Academic*. Central Michigan University 9 December 2008 <http://www.lexisnexis.com/>.

Helft, Miguel "Google Adds a Safeguard on Privacy for Searchers." *New York Times* 15 March 2007: [1-2] Section C; Column 6; Business/Financial Desk. U.S. News LexisNexis Academic. Central Michigan University 7 May 2007 <http://www.lexisnexis.com/>.

"I can has cheezburger." 5 December 2008 <http://icanhascheezburger.com/>

Inniss, Tasha R., et al. "Towards Applying Text Mining and Natural Language Processing for Biomedical Ontology Acquisition" *Conference on Information and Knowledge Management, Proceedings of the 1st International Workshop on Text mining in Bioinformatics*, November 10, 2006 : 7-13. ACM Special Interest Group on Information Retrieval. Association for Computing Machinery Arlington, Virginia. 21 November 2008. <http://portal.acm.org/citation.cfm?id=1183535.1183539&coll=GUIDE&dl=GUIDE>

Jaeger, Paul T., John Carlo Bertot, and Charles R. McClure. "The Impact of the USA Patriot Act on Collection and Analysis of Personal Information under the Foreign Intelligence Surveillance Act." *Government Information Quarterly* 20.3 (2003): 295-314. WilsonSelectPlus. FirstSearch. Central Michigan University, 28 July 2006 <http://www.oclc.org/default.htm>.

Kirchgaessner, Stephanie. "Web Giants on Defensive over Chinese 'Holocaust' Remarks." *Financial Times* 16 February 2006, USA 2 nd edition: [1-2]. ABI/INFORM Global. ProQuest. Central Michigan University, 26 July 2006 <http://www.proquest.com>.

Kopytoff, Verne. "New charge Yahoo aided China Crackdown; 4th Dissident Jailed with Online Data, Rights Group Says." *San Francisco Chronicle* 29 April 2006: C1, [1-2] InfoTrac Custom Newspapers, Thomson Gale. Central Michigan University, 26 July 2006. <http://www.gale.cengage.com/>

Lash, Steve. "Copyright Official, Senator Diverge Over Clarity of Law." *Chicago Daily Law Bulletin* 28 September 2005: [1-2]. Legal News. LexisNexis Academic. Central Michigan University, 5 July 2006 <http://www.lexisnexus.com/>.

Lee, Monie. "Attitudes Toward Direct Marketing, Privacy, Environment, and Trust: Taiwan vs U.S." *International Journal of Commerce & Management* 14 (2004): 1-17. ABI/INFORM Global. ProQuest. Central Michigan University, 5 July 2006 <http://www.proquest.com>.

Lewandowski, Dirk. "Web Searching, Search Engines and Information Retrieval." *Information Services and Use* 25 (2005): 137-147.

"Live Search Maps." 4 December 2008 <http://maps.live.com/>.

Marcum, Deanna B. "The Future of Cataloging." *Library Resources & Technical Services* . 50.1 (2006): 5-9.

Metro-Goldwyn-Mayer Studios inc., et al. v. Grokster, ltd., et al., no. 04-480, Supreme Ct. of the US, 27 June 2005.

Miller, Todd. "Federated Searching: Put in Its Place." *Library Journal* 15 April 2004: [1-2]. 6 Jan. 2006 <http://www.libraryjournal.com/article/CA406012.html>

Musthaler, Linda. "Facebook Fiasco Highlights Privacy Concerns," *Network World* 3 January 2008: [1-2]. US Newspapers & Wires. LexisNexis Academic. Central Michigan University, 6 January 2008. <http://www.lexisnexus.com/>.

Naim, Moises. *Illicit*. New York: Anchor Books, 2005.

Patel, Jay. "Personalizing Search: Promise & Pitfalls." Search Engine Meeting, April 23-24 2007: 1-5. Infonortics. Boston, Massachusetts: (Notes prepared for remarks)

"Review of Personalization Technologies: Collaborative Filtering vs. ChoiceStream's Attributized Bayesian Choice Modeling" Technology Brief, *ChoiceStream*, July 2007: 1-13. <http://www.choicestream.com/solutions/technology/>.

Richards, Neil M. "The Information Privacy Law Project." *Georgetown Law Journal* 94 (July 2006): 1-51. Law Reviews, LexisNexis Academic. Central Michigan University, 12 July 2006 <http://www.lexisnexus.com/>.

Schmidt, Eric. "Interview with Charlie Rose." *Charlie Rose Show*. PBS. 3 June 2005. 1-12. Transcript.

Smyth, Barry. "Collaborative Web Search: Social & Personal." Search Engine Meeting, April 23-24 2007: [1-26]. Infonortics. Boston, Massachusetts.

Solove, Daniel J. *The Digital Person: Technology and Privacy in the Information Age*. New York: New York University Press, 2004.

Stoan, Stephen K. "Research and Library Skills: An Analysis and Interpretation." *College & Research Libraries* 45 (1984): 99-109.

Tracy, James F. and Maris L. Hayashi. "A Librariat? Labor, Technology, and Librarianship in the Information Age." In *Knowledge Workers in the Information Society*. Ed. Catherine McKercher and Vincent Mosco. New York: Lexington Books, 2007. 53-67.

"The Truth about Federated Searching." *Information Today*: feature article November/December 2003: [1-2]. 6 Jan 2006 <http://www.infotoday.com/it/oct03/hane1.shtml>.

Vaidhyanathan, Siva. "A Risky Gamble with Google." *Chronicle of Higher Education* 2 December 2005: [1-9].

Weiler, Angela. "Information-Seeking Behavior in Generation Y Students: Motivation, Critical Thinking, and Learning Theory." *Journal of Academic Librarianship* 31 (2005): 46-53.

Yu, Clara, et al. "Inside the Mind of a Search Engine." *Patterns in Unstructured Data: Discovery, Aggregation, and Visualization*. Presentation to the Andrew W. Mellon Foundation, May 2002: [1]. National Institute for Technology and Liberal Education, 26 July 2007 <http://www.knowledgesearch.org/lsi/>

Yu, Clara, et al. "Introduction – the Need for Smarter Search Engines." *Patterns in Unstructured Data: Discovery, Aggregation, and Visualization*. Presentation to the Andrew W. Mellon Foundation, May 2002: [1-2]. National Institute for Technology and Liberal Education, 26 July 2007 <http://www.knowledgesearch.org/lsi/>

Yu, Clara, et al. "Structured data – Everything in its Place." *Patterns in Unstructured Data: Discovery, Aggregation, and Visualization*. Presentation to the Andrew W. Mellon Foundation, May 2002: [1]. National Institute for Technology and Liberal Education, 26 July 2007 <http://www.knowledgesearch.org/lsi/>