Bibliographic Essay Unit 5 -- Searching
Erin Horanzy and Rosanna Longenbaker
June 22-27

The world of searching is vast and complex. The types of searches and search engines available to conduct one's research are not only multitudinous but somewhat difficult to differentiate at first glance. As Nancy Davis Kho observes, "As search technology has evolved and digital information has expanded and taken on new forms, "search" is less about finding and more about doing: integrating search and discovery into workflow to improve and speed decision making." (Kho 2010) In order for librarians to serve serve patrons sufficiently, it is necessary to be capable of doing searches at a high level of competency.

There are three types of search engines: crawler-based, human powered, and hybrid. A crawler-based search engine uses crawlers, automatic searches that follow hyperlinks, to find content. Google is an example of a crawler-based engine and is best used for general all-around information gathering. A human powered search engine is also called a directory, and uses editors who respond to submissions in order to create search results. Yahoo! directories is an example of this type of engine, and is considered good for second-line searches. Of course, the hybrid combines the properties of the first two. Bing bills itself as a "decision engine" and appears to do just that. However, Microsoft's search algorithms are somewhat opaque. (Werts 2011)

Now, on to the searches themselves. There are several types of searches: Boolean, keyword, and semantic (natural language). The keyword search is the simplest, but the least intuitive. If you know many synonyms of the word you're searching for and you are willing to conduct several searches, or you are looking for something extremely specific, uses this search. It is what it sounds like: simply input a word or phrase into the search engine or database and it uses the keywords to conduct a search.

A Boolean search is similar, in a way, to a math problem that affects the number of search results. There are three types of connectors, AND, NOT, and OR. Using the connector OR is like adding - it expands the number of results. Using AND and NOT is like subtracting-- AND requires that both keywords be present in the results, and NOT requires that the keyword following NOT not be in the results. Additionally, using a + requires that the exact keyword be included in the results, and quotation marks require that the phrase be included as is in the results. However, it is important to note that Google will return results for a non-quotation marked phrase with higher ranks for the complete phrase as written. If a search is done correctly, using the Boolean operators in a logical way, a search can be narrowed to the point where all or almost all of the results are pertinent to the initial search. (Spencer 2011)

A semantic search, also known as a natural language search, is a search conducted using the context of the words within the structure of the search phrase. Instead of inputting keywords or Boolean operators, the searcher simply inputs a phrase or question, from which the search engine extracts the salient keywords and provides the searcher with, hopefully, the appropriate answer. A semantic search, according to Tamas Doszkocs, "produces meaningful results, especially when the retrieved items do not contain any of the query terms or the search involves no query text at all." (Doszkocs 2010) Instead of finding the response to the input, it finds the

answer to the implied question. To a user inexperienced in Boolean, the semantic search is the most intuitive and most likely to provide the desired answers. However, there are a few hazards to this type of search-- for example, if the search concerns something homonymic such as "mouse." Should the results come up with the small rodent or the device used to manipulate a computer? In this case, a better search would be one that includes Boolean operators to specify the search term.

Therefore, the most important thing to keep in mind as one searches is that it is important to run multiple searches in the hopes that different results will arise from different keywords or semantics. As Marydee Ojala writes, in her article "Black Thumb Searching," "Search is not static. Black thumb [overly shallow] searches can turn green [deep, complex and fruitful]... I think this is another mark of a sophisticated searcher versus an amateur. We return and redo searches. It's not hit and run; it's taking a second and third look to see what has changed." (Ojala 2011) Searching is no longer simply asking for information and receiving it; instead it has become an active interaction where the information is contextually provided, improving search quality. (Brynko 2010)

Even more important than perseverance is using the appropriate search terms. If one term is bringing in too many results, it's necessary to think of a narrower synonym. For example, "dogs" gets 619 million results in Google, whereas "Chihuahua-Great Dane mixes" gets a mere 1.6 million hits. Brett Spencer, in his article entitled "Lesson Plans for Google Search Specificity", reminds searchers to consider the type of information desired. Everyone knows to look for cheesecake recipes but fail to consider searching for knitting tutorials rather than "how to knit". (Spencer, 2011) Spending a small amount of time in thinking carefully about the appropriate search terms can save a lot of time and effort overall.

The smart googler (noun: a person who searches Google) also considers the wide variety of alternate services that Google provides such as searches for images, videos, maps, news, email, and shopping. Knowing that the search is, for example, for a recipe can narrow the search down by searching only in recipes. Each subsearch also provides a number of narrowing choices, including but not limited to the size of an image, some ingredients in a recipe, the dates in a news article and the price of items for sale. David Mattison, in his article "Time, Space, and Google," discusses at length these services as well as several useful collaborative services such as Google groups, sites, documents, and Wave. (Mattison 2010)

A final smart method of searching is to use a metasearch engine. A metasearch engine is an engine that searches other search engines. They don't come up with their own search results, but pass on the results garnered by others. A few examples of metasearch engines are Dogpile, Metacrawler, Deeperweb and Kayak. Dogpile, for example, runs a search through Google, Yahoo! and Bing simultaneously and shows a list of results with a small credit line to the appropriate search engine below each one. Understanding the concept of metasearch engines aids in the understanding of metadata.

The simplest definition of metadata is simply a set of data that describes and gives information about other data. "Metadata is sometimes referred to as data about data" (Rubin, 2010, p. 157). Traditionally, metadata can be found in a library's card catalog (literally or figuratively within the card catalog online). In a library, the data is found within the books, articles, and other media stored for patron use. The card catalog therefore gives data about these

pieces of data: title, author, date, physical specifications, and medium among many other data points. “Metadata can be used not only for locating current materials but for archival and preservation purposes as well” (Rubin, 2010, p. 157).

Tamara Pianos examined the portals used in Germany to access academic information. She noted that academics would like portals to have search features that are easy to find on their homepages. Users want to find information easily and for little or no cost. Also, she noted that people did not want to sign up with multiple passwords. Academics would like to be able to access multiple sources with the same password. She wrote about Vascoda, which is “an interdisciplinary academic information portal” (Pianos, 2008, p.123). Originally, vascoda used “meta-search-techonlogy” it now uses “new search engine technology” (Pianos, 2008, p. 123).

In the article, “Metadata Models of the World Wide Web,” the concept of using metadata to search the Web is explained along with the work that is necessary to make that possible. Usually, libraries have used metadata while searching their catalogs. This metadata comes from databases with MARC records. However on the Web, information is searched for using key words. Search engines have helped to make that material accessible. It is hoped that metadata can be used to search the Internet using the “Semantic Web.” In order for this to succeed a standardized system for labeling websites with metadata had to be created (Metadata Models). “the Semantic Web developers sought to develop the basic structure on which all other metadata would be developed. This basic structure is called the Resource Description Framework, or RDF. RDF itself relies on the Uniform Resource Identifier, the standard identifier format for the Web, and extensible Markup Language (XML), a set of rules for encoding documents and data electronically” (Metadata Models, 2010, p. 13)

Jane Hutton (2008) explored the ability of college libraries to provide online students with access to free e-books. She discussed that while there are many projects to make records and books available online for free these documents are part of the hidden web. To check availability, Hutton searched for the same list of free e-books using the websites of several different college libraries. She also used Internet search engines to find the same material. Through this study it was found that libraries were not succeeding in finding the e-books, which are available online for free. Internet Archive, INFOMINE, and OAIster are some attempts to make it easier for people to find the full-text books that are available online.

At the base, the problems that she faced were the overly complex methods of finding the e-books. If the sites providing the e-books were using standardized metadata, the entire endeavour would be much simpler. In her paper, Hutton mentions the difficulties that arise in finding full-text e-books because there are no standards for records. “Unlike MARC records from a single library catalog, which can be merged into a searchable database of consortial library holdings, the data fields for one set of digital document records may not match a set of digital document records from another source” (Hutton, 2008, p. 497-498).

Some efforts are underway to standardize metadata on the Web. The Dublin Core Metadata Element Set , “defines a set of elements for resource description in the online environment” (Rubin, 2010, p. 158). It will enable people who do not have experience in cataloging to “describe their Web sites and assign terms that could significantly aid others in accessing them” (Rubin, 2010, p. 158).

"There are many types of search tools, but the most popular versions can be categorized as (1) individual search engines, (2) subject directories, (3) metasearch engines, and (4) specialized search engines." (Williams & Sawyer, 2011, pg 74) There is great importance in proper searching on a multitude of platforms, including search engines, metasearch engines, and using metadata. If there isn't proper understanding of the platform and the search terms, then adequately thorough research cannot be done.

WORKS CITED

Brynko, Barbara. (17 June 2011). SEM: The Cutting Edge of Search. Information Today 27.6: 29-30. Computers & Applied Sciences Complete. EBSCO. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=iih&AN=50995593&site=ehost-live

Doszkocs, Tamas. (17 June 2011). SEMANTIC Search Engines Mean Well. Online 34.4: 36-42. Computers & Applied Sciences Complete. EBSCO. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=iih&AN=52686610&site=ehost-live

Hutton, J. (2008). Academic libraries as digital gateways: Linking students to the burgeoning wealth of open online collections. Journal of Library Administration, (3/4), 495-507. Retrieved from EBSCOhost. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=lxh&AN=35653208&site=ehost-live

Kho, Nancy Davis. (17 June 2011). SEARCH: It's More Than Finding; IT'S DOING. EContent 33.5: 24-29. Computers & Applied Sciences Complete. EBSCO. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=iih&AN=51122504&site=ehost-live

Mattison, David. (17 June 2011). Time, Space, and Google. Searcher 18.4: 20-31. Computers & Applied Sciences Complete. EBSCO. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=iih&AN=50315140&site=ehost-live

Metadata models of the World Wide Web. (2010). Library Technology Reports, 46(2), 12-36. Retrieved from EBSCOhost. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=lxh&AN=48651558&site=ehost-live

Ojala, Marydee. (17 June 2011). Black Thumb Searching. Online 34.3: 48-50. Computers & Applied Sciences Complete. EBSCO. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=iih&AN=50140891&site=ehost-live

Pianos, T. (2008). A comparison of academic information portals. Information Services & Use, 28(2), 123-125. Retrieved from EBSCOhost. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=lxh&AN=34723205&site=ehost-live

Rubin, R. E. (2010). Foundations of Library and Information Science. New York: Neal-Schuman Publishers, Inc.

Spencer, Brett. (17 June 2011). Lesson Plans for Google Search Specificity. Online 35.3: 29-33. Computers & Applied Sciences Complete. EBSCO. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=iih&AN=60435952&site=ehost-live

Werts, Cybée Elaine. (17 June 2011). COMPARING SEARCH ENGINES for Quick and Dirty Answers. Online 34.1: 32-35. Computers & Applied Sciences Complete. EBSCO. http://0-search.ebscohost.com.www.consuls.org/login.aspx?direct=true&db=iih&AN=48218625&site=ehost-live

Williams, B.K., Sawyer, S.C. (2010). Using information technology: a practical introduction to computers and communications. New York: McGraw-Hill.