So You Think You're an Expert: Keyword Searching vs. Controlled Subject Headings

April Grey, University of Louisiana - Lafayette & Christine R. Hurko, Niagara University

"If our goal is to compile a set of results and sift through those results to get the best ones … keywords may not always be the best way to accomplish the task".
--William Badke[1]

Abstract

The authors examine the differences between keyword searching and controlled subject headings in conducting research and the relative issues of each. They highlight the habits of students who frequently use Google to search and how this colors the students' ability to search in controlled environments like traditional databases. The benefits of taxonomies (the structure of LCSH) are examined as a means of more effective searching. Folksonomies and social tagging are also discussed. The authors highlight that LCSH is more in line with effective information literacy instruction and conclude that controlled subject heading searching is truly a more effective means of conducting research.

So You Think You're an Expert: Keyword Searching vs. Controlled Subject Headings

Introduction

According to Standard 2, Performance Indicator 2 of the Association of College & Research Libraries (ACRL) Competency Standards for Higher Education, information literacy standards state that an information literate student "...selects controlled vocabulary specific to the discipline or information retrieval source."[2] Information literate students must be able to search the library catalog using subject headings to find relevant resources. For the Google generation, which is defined as people "born after 1993, growing up in a world dominated by the internet," keyword searching has become second nature.[3] Librarians should be able to explain *why* 5.4 million results are received in 0.13 seconds using keyword searching via a search engine. When first introduced to an online catalog, some undergraduates are unaware that there is a more effective way to search beyond keywords. Controlled subject headings are more efficient than keyword searching or social tagging and yield higher quality results. Librarians can show patrons how to increase relevant results by adhering to ACRL standards and by employing controlled vocabulary searching.

The Impact of Web Search Engines

How often have you noticed students typing an entire sentence, including punctuation, into a search box in hopes that the OPAC or database will return the results they need? In a study conducted in 2002, Yu and Young analyzed the effect Web search engines had on subject searching in a library's OPAC.[4] They noted that regardless of the visual and systematic differences between an OPAC and an Internet search engine,

patrons continued to search using uncontrolled keywords as if they were searching the Web. Similarly, Antel and Huang stated that, "…subject searching is difficult for patrons, unlikely to be very successful and becoming less frequent as patrons' behavior is shaped by keyword search engines."[5] Today's students do not realize that Web searching and OPAC searching differ and the students predominantly use keywords to query a search. Unlike an OPAC, Web searches are *keyword-dependent* and require the use of an exact string of characters to bring back results. Websites lack controlled vocabulary, or access points, to describe categories such as title, author or subject. Like the children's book character Amelia Bedelia, who doesn't understand homonyms, web search engines cannot determine the meaning of a word in context. They are designed to scour full-text resources for the presence of keywords regardless of context or meaning and bring back results based on the string of characters that was typed in. The authors have coined this "Amelia Bedelia Syndrome." For example, just Googling the word "dressing" brings up numerous results because it has more than one meaning. It can mean dressing as in getting dressed, salad or food dressing, dressing as in preparing a chicken, or bandage as in wound dressing. Unless the descriptive data is specifically written out on a Web page, chances are the results returned will be diverse especially when the search term is so ambiguous. The user is then forced to revise his or her search multiple times before finally receiving useful results.

Keywords, key words, "key words"

As the use of keyword searching in OPACs and databases has increased with the emergence of over 200 available search engines[6], librarians have become more concerned with how information is sought and evaluated by patrons. The transition from "searching

dominated by metadata-enabled searching (academic library card catalogs) to the present full-text or algorithmic searching (Web search engines) [has]… occurred without sufficient analysis of the weaknesses of full-text searching."[7] Many searches are deemed unsuccessful by the seeker because the use of specific keywords does not retrieve all relevant items, as many are overlooked by search engines. According to Thomas Mann, the weakness of full-text keyword searching is that it looks at the "granular level" instead of considering the complete picture.[8] Often, a variety of results will be returned for students' keyword searches that include both the anticipated meaning as well as potential unintended meanings. However, students do not want to evaluate scores of search results in order to find items that are applicable to their query.

Conversely, spelling and grammatical errors will decrease the accuracy of search results. Contemporary, popular and slang vocabularies are not included in subject headings and therefore do not show up when searching the library OPAC by subject. This may lead students to believe that the library doesn't offer any material on the information they seek, leading to a negative research experience. According to Badke, "the ambiguities of language, the possibility of describing the same thing in multiple ways and the fact that some concepts are almost impossible to put into words, mean that keywords are going to be treacherous friends."[9]

However, not all keyword searching provides poor results and there are some advantages. When pursuing a research topic, McCutcheon acknowledges that in reference to the ease of use, keywords are the simplest way to acquire basic information. For the Google generation, this is usually where searching stalls. Once a researcher is required to find scholarly sources, "...keyword searching alone is unable to provide results that

enable a seeker to gain a comprehensive overview or in-depth understanding of a given subject."[10]

To Control, or Not to Control: That is the Question

Are LCSH and controlled vocabularies necessary for smart information retrieval? In the library world, there have been two schools of thought – those who want to do away with controlled subject headings, and those who want to continue them.[11] Yu and Young believe controlled vocabularies are critical for information retrieval. Without controlled vocabulary terms being assigned in MARC records, the keywords searched by users would not produce as many relevant results and as researchers tack on more keywords in their searches, they run the risk of search failure, or retrieving "zero" hits.[12] In fact, it may take researchers less time and result in more useful results if subject headings are used to search a catalog. Unlike keywords, subject headings are consistent. They are concepts that are arranged in a taxonomy or a predetermined string of words, and they describe a work in hierarchical form. This is the essence of how subject headings work in any catalog. All items that have the same subject heading will be returned in a results list when a particular subject heading is searched. Controlled vocabulary can also have numerous subdivisions in order to provide more precise results. The additional information narrowed down by subdivisions includes dates, geography and form. This is another way that students can narrow down results when using LCSH to find topical information for research.

Because catalogers provide controlled subject access to materials, library users are able to do more than search by keyword, author, and title. Often, the keyword searched will be present in one or more of the words in the subject headings allowing for

the retrieval of relevant records. Arguably, LCSH are more significant to the search function of the OPAC than many librarians realize. In a 2005 study conducted by Gross and Taylor on the effect controlled vocabularies had on keyword searching, it was found that, "...if subject headings were to be removed from or no longer included in catalog records, users performing keyword searches would miss more than one third of the hits they currently retrieve..."[13] However, a keyword search in a library's catalog can prove fruitful, if it is used in conjunction with an understanding of subject headings. Once a record with a meaningful LCSH is found by the student, the hyperlinked LCSH will then lead to other related items.

Some librarians believe that subject analysis isn't necessary in the age of digital retrieval since students are using keyword searches as their main search method of choice. According to Antell and Huang, keyword searching is overwhelmingly commonplace. They report that in 1995, Arlene Taylor, at the Association for Library Collections and Technical Services, in a program titled, "Crisis in Subject Cataloging and Retrieval", noted that library administrations were questioning whether catalogers and controlled subject headings were even necessary.[14] Libraries encounter significant costs in assigning controlled subject headings to materials, in the form of hiring catalogers and maintaining resources. However, having controlled subject headings improve the searchable material in the OPAC and increases the students' access to the library materials. The benefit of accurate search results outweighs the cost of the professional resources required for thorough subject analysis.

According to Thomas Mann, taxonomies such as the Library of Congress subject headings are "linked to broader, related, and narrower terms".[15] Subject headings

establish relationships with other headings and strings of subject headings are vital to reducing search results to a usable number. Whittling down the number of items returned helps students find sources that are relevant to their topics. LCSH provides a system of increasingly narrower search parameters set in a strict taxonomy. The bottom line is that librarians need to keep students engaged in library research, which means that they need to retrieve relevant results nearly every time they search, and keyword searches or social tagging are at best occasionally effective. Since LCSH are housed in an authority file in the library's OPAC, a thoughtful and informed search of the terms in the catalog will always bring back at least one item.

Social Tagging and the Collaborative Creation of Content

A folksonomy is a way to classify user-generated social tags that are written in natural language. Lyons and Tappeiner argue that combining taxonomies and folksonomies is beneficial to improving subject access to resources. For example, new technologies are harder to describe using controlled subject headings because the vocabulary isn't updated instantly and catalogers must choose descriptions from a predetermined list of "outdated" headings. Folksonomies can be used in conjunction with traditional subject headings to improve access using free-form descriptions. Allowing library users to participate in describing library material by creating tags can provide a more appropriate description of the item "specific to their peer groups."[16] Scholarly research in continuously advancing subjects such as the sciences and engineering would greatly benefit from user-generated tagging because access of articles by those in the field with knowledge of specialized vocabulary will increase. Nevertheless, there are weaknesses in this system. Inconsistencies in user-generated tags mean abbreviations and misspellings will affect the

search results negatively, and this is yet another reason why controlled subject headings are vital to catalog records.[17]

In fact, a hybridized version of controlled and user-generated tags would offer the best of both worlds. Librarians should understand the value of a static taxonomy, yet as researchers we realize that the world is constantly changing and that in order to stay relevant and connected to the Google generation, we must update LCSH continuously. Search terminology regularly changes because vocabulary is not static. However, updating subject headings in LCSH is a very slow and laborious process. For decades, radical cataloger Sanford Berman has advocated for the need to update subject heading terminology from the archaic to the contemporary.[18] In theory, user-generated tags enhance traditional subject headings and assist users by providing a current subject thesaurus. Unfortunately, patron generated tags are too broad in scope in their current usage. In a study regarding the inclusion of social tags as indexing terms to predict applicable subject headings, Yi noted that, for any given search, "the most valuable finding in terms of resource may be the decrease of performance with more than the top-five popular social tags." [19]

Making Information Literacy Informative

Searching for and analyzing information can cause students to become frustrated. This feeling of dissatisfaction and displeasure has been referred to as "search fatigue." [20] Specifically, this condition of distress during library research is caused by the dependence on keyword searching. To combat this affliction students must be taught how to use subject headings in order to take advantage of the significant amount of available

research via academic resources. They will reap the benefits of comprehensive results that controlled vocabulary brings to scholarly research.

Everywhere, library instructors face the same problem - how to teach students numerous search tactics in a short amount of time. Besides the usual handouts and demonstrations on the computer via projector, instructors need to engage student "experts." PowerPoint presentations are all well and good, but they can be monotonous, causing students' eyes to glaze over midway through the first slide. It's hard enough to try to teach a subject most students think they already know, let alone try to keep them from browsing Facebook and Twitter while trying to stress the importance of controlled subject headings in an OPAC. Unfortunately, "users are not willing to devote much of their time to learning to use these systems. They just want to get their search results quickly and expect the catalog to be easy to use with little or no time invested in learning the system."[21]

One way that library instructors can show the benefits of LCSH is to explain how controlled vocabulary is linked in the OPAC. Once students understand how subject headings work, they can click on a controlled subject heading which will lead to them to other materials with the same subject heading. Librarians should explain the ease of using hypertext linked subjects in the OPAC. Using the pre-assigned subject terms will bring students to multiple records in one click. Although it may seem time consuming for students to search for a subject heading that is close to their topic of choice, it is more productive than stumbling through an ambiguous keyword search. They will then be able to find materials that are directly related to the information they seek. They will also learn

that scholarly research is not quick and easy. It takes time and practice to become a proficient researcher.

Conclusion

Over the last few decades there has been much debate over the hasty emergence and growing popularity of full-text keyword searching, and the decline in the use of traditional controlled subject headings. As catalogers and librarians struggle to come to grips with a generation so accustomed to the swiftness in which Web search engines spew back results, one must ask how we can continue to preserve the integrity of the library catalog while still accommodating today's users. We must remember that the backbone of every library collection is the quality controlled subject headings assigned by professional catalogers. Enhancements to the catalog such as user-generated tags may increase access but will never fully replace controlled subject headings.

---

1. Badke, William. "The Treachery of Keywords." *Online* 35, no. 3 (May/Jun 2011): 52-54. *Academic Search Complete*, EBSCO*host* (accessed October 11, 2011).

2. Association of College and Research Libraries. "Information Literacy Competency Standards for Higher Education." Association of College and Research Libraries. http://www.ala.org/ala/mgrps/divs/acrl/standards/informationliteracycompetency.cfm (accessed September 22, 2011).

3. Nicholas, David, Peter Williams, Paul Huntington, Maggie Fieldhouse, Barrie Gunter, Richard Withey, Hamid R. Jamali, Tom Dobrowoiski, and Carol Tenopir. "The Google Generation: The Information Behaviour of the Researcher of the Future." *Aslib Proceedings* 60, no. 4: 290-310. *Academic Search Complete*, EBSCO*host* (accessed October 11, 2011).

4. Yu, Holly, and Margo Young. "The Impact of Web Search Engines on Subject Searching in OPAC." *Information Technology & Libraries* 23, no. 4 (December 2004): 168-180. *Academic Search Complete*,

EBSCO*host* (accessed October 10, 2011).

5. Antell, Karen, and Jie Huang. "Subject Searching Success Transaction Logs, Patron Perceptions, and Implications for Library Instruction." *Reference & User Services Quarterly* 48, no. 1 (Fall 2008): 68-76. *Academic Search Complete*, EBSCO*host* (accessed October 10, 2011).

6. The Search Engine List. "Comprehensive List of Search Engines." Accessed October 11, 2011. http://www.thesearchenginelist.com.

7. Beall, Jeffrey. "The Weaknesses of Full-Text Searching." *Journal of Academic Librarianship* 34, no. 5 (September 2008): 438-444. *Academic Search Complete*, EBSCO*host* (accessed October 11, 2011).

8. Mann, Thomas. "The Peloponnesian War and the Future of Reference, Cataloging, and Scholarship in Research Libraries." *Journal of Library Metadata* 8, no. 1 (March 2008): 53-100. *Academic Search Complete*, EBSCO*host* (accessed October 11, 2011).

9. Badke, 52.

10. McCutcheon, Sevim. "Keyword vs. Controlled Vocabulary Searching: The One with the Most Tools Wins." *The Indexer* 27, no. 2 (June 2009): 62-65. *Library, Information Science & Technology Abstracts*, EBSCO*host* (accessed October 10, 2011).

11. Calhoun, Karen. "The Changing Nature of the Catalog and its Integration with Other Discovery Tools." prepared for the Library of Congress, 2006, available at: http://www.loc.gov/catdir/calhoun-report-final.pdf (accessed October 11, 2011).

12. Yu, 168.

13. Gross, Tina, and Arlene G. Taylor. "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results." *College & Research Libraries* 66, no. 3 (May 2005): 212-230. *Academic Search Premier*, EBSCO*host (*accessed October 10, 2011).

14. Antell, 69.

15. Mann, 62.

16. Lyons, Catherine, and Elisabeth Tappeiner. "Cataloging 2.0: Metadata Research and Initiatives at a Community College Library." *Journal of Library Metadata* 8, no. 2 (2008): 155-157. *Academic Search Complete*, EBSCO*host* (accessed October 10, 2011).

17. Badke, William B. *Research Strategies: Finding Your Way Through the Information Fog*. 3rd ed.

Bloomington, IN: Iuniverse, 2008.

18. Berman, Sanford. *The Joy of Cataloging: Essays, Letters, Reviews, and Other Explosions.* Portsmouth, NH: Greenwood-Oryx Press, 1981.

19. Yi, Kwan. "A Semantic Similarity Approach to Predicting Library of Congress Subject Headings for Social Tags." *Journal of the American Society for Information Science & Technology* 61, no. 8 (August 2010): 1658-1672. *Academic Search Complete*, EBSCO*host* (accessed October 11, 2011).

20. Beall, Jeffrey. "Search Fatigue: Finding a Cure for the Database Blues." *American Libraries* 38, no. 3 (March 2007): 46-50. *Academic Search Complete*, EBSCO*host* (accessed October 10, 2011).

21. Yu, 178.