

Keyword Searching and Subject Headings at the University of Manitoba Libraries: a Research Project

by

Pat Nicholls

Library catalogues have been under attack for a number of years for being costly to create and maintain and difficult to use. With the advent of web catalogues and Google, there has been even more pressure on libraries to change how catalogues are created but only a limited amount of research has been undertaken to determine how useful the data that is provided in our catalogues really is. This study attempts to fill that gap in a limited sense, for the University of Manitoba Libraries' catalogue.

A couple of reports have been published recently which recommend removing subject analysis from the cataloguing process. The recommendations of the Bibliographic Services Task Force at the University of California published late last year included the following recommendation:

III.2c: Consider using controlled vocabularies only for name, uniform title, date, and place, and abandoning the use of controlled vocabularies [LCSH, MESH, etc] for topical subjects in bibliographic records. Consider whether automated enriched metadata such as TOC, indexes can become surrogates for subject headings and classification for retrieval.¹

Karen Calhoun recently completed a study for the Library of Congress of the current status of library catalogues. She recommends changes that research libraries should consider from a business perspective. Among the recommendations that she makes, the following is included:

4.2.3 Abandon the attempt to do comprehensive subject analysis manually with LCSH in favor of subject keywords; urge LC to dismantle LCSH [14]²

Neither of these recommendations, however, were based on research studies that looked at the actual results of catalogue use studies.

¹University of California Libraries. Bibliographic Services Task Force. Rethinking How We Provide Bibliographic Services for the University of California: Final Report. 2005. Retrieved 10 July 2006 from: <http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>

²Karen Calhoun, *The Changing Nature of the Catalog and its Integration with Other Discovery Tools*, 2006. Retrieved 12 April 2006 from: <http://www.loc.gov/catdir/calhoun-report-final.pdf>

Literature Review

Catalogue use studies have a long and checkered history in library literature^{3,4} but the evolution of the online catalogue has opened a new variety of possibilities including transaction log analysis. Several of the articles that I reviewed that studied transaction logs reported interesting results. The Transaction Log Analysis Task Force at the University of Illinois at Chicago “analyzed transaction logs of the university’s OPAC alongside existing online user aids, evaluated screen displays, and recommended changes to enhance the system’s user friendliness.”⁵ An investigation of a sample of the transactions of 114 sessions on the online catalog of University of Missouri-Columbia was used to determine what types of searches were conducted and what search modes and fields were used.⁶

The most recent study and most similar to the one I conducted, was done by Tina Gross and Arlene G. Taylor which examined a sample of 227 searches from a transaction log of 2,270 keyword searches to determine what proportion of records retrieved by a keyword search has a keyword only in a subject heading field. Their results, with a smaller sample, more carefully screened, found that, “on average, 35.9 percent of hits would not be found.”⁷

Background at the University of Manitoba

Using files that are created for backup purposes, which record virtually all use of the catalogue, it is possible to examine the searches that have been done on the catalogue. The University of Manitoba Libraries catalogue, locally known as BISON (<http://bison.umanitoba.ca>), is a Web2 catalogue searching our SirsiDynix Unicorn

³Andrew Large and Jamshid Beheshti, “OPACs: A Research Review,” *Library & Information Science Research*, 19 (1997): 111-133.

⁴Scott Hamilton and Helen Thurlow, “Transaction log analysis @ State Library of Queensland”. *Evolution of Evidence: Global Perspectives on Linking Research with Practice*. Retrieved 13 April 2006 from: <http://conferences.alia.org.au/ebi2005/Hamilton.pdf>

⁵Deborah D. Blecic, et al, “Using Transaction Log Analysis to Improve OPAC Retrieval Results,” *College & Research Libraries*, 59 (1998), p. 41.

⁶Lynn Silipigni Connaway, John M. Budd and Thomas R. Kochtanek, “An Investigation of the Use of an Online Catalog: Users Characteristics and Transaction Log Analysis”, *Library Resources and Technical Services*, 39 (1995), p. 142-151.

⁷Tina Gross and Arlene G. Taylor, “What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results”, *College and Research Libraries*, 66 (2005) 212-230.

database. Our customization of the Web2 software provides keyword searching and browsing of the catalogue by all of the standard indexes, including author, title and subject, but the default search is a keyword search of virtually the complete MARC record. Our database contains over 1.3 million bibliographic records representing 99.9% of our collections, with over 2 million item records. Our policy is to catalogue all material acquired (or for which access has been purchased), regardless of format, including books, journals, digital resources, government publications, maps, manuscripts, microforms, and other materials. Examining these log files can provide much information about our users and how the cataloguing information that we provide is used to access our materials.

The web log files have been saved and stored on a monthly basis for statistical purposes. Here is a summary of the keyword searches of our catalogue for 2005:

Table 1

| BISON Web Search Statistics | | | | | | | | |
|------------------------------------|----------------|-------------------------------------|---------------|---------------|----------------|-----------------|--------------|----------------------------|
| Keyword Searches | | | | | | | | |
| Month | Total Searches | Total Keyword Searches ⁸ | % of searches | Title Keyword | Author Keyword | Subject Keyword | ISSN Keyword | Other Keyword ⁹ |
| 2005 January | 205,327 | 153,399 | 75% | 20,225 | 14,872 | 6,943 | 13,699 | 8,446 |
| 2005 February | 293,782 | 219,703 | 75% | 27,112 | 17,933 | 9,959 | 21,601 | 12,001 |
| 2005 March | 228,263 | 173,994 | 76% | 20,376 | 13,318 | 7,376 | 19,253 | 8,379 |
| 2005 April | 155,055 | 109,260 | 70% | 15,067 | 9,609 | 3,719 | 9,411 | 7,179 |
| 2005 May | 126,744 | 88,731 | 70% | 12,017 | 7,526 | 3,410 | 7,326 | 7,388 |
| 2005 June | 118,420 | 83,229 | 70% | 12,755 | 7,620 | 3,143 | 7,544 | 6,713 |
| 2005 July | 110,583 | 75,997 | 69% | 10,358 | 6,976 | 2,439 | 7,594 | 6,335 |
| 2005 August | 142,609 | 98,070 | 69% | 14,938 | 10,259 | 2,691 | 8,200 | 9,660 |
| 2005 September | 180,111 | 139,802 | 78% | 16,751 | 11,551 | 6,025 | 15,913 | 7,494 |
| 2005 October | 290,917 | 224,218 | 77% | 24,097 | 16,460 | 9,443 | 28,634 | 13,306 |
| 2005 November | 279,267 | 215,820 | 77% | 22,612 | 15,625 | 8,574 | 23,252 | 11,477 |
| 2005 December | 129,096 | 96,089 | 74% | 12,079 | 8,647 | 2,817 | 7,866 | 6,919 |
| 2005 Total | 2,260,174 | 1,678,312 | 74% | 208,387 | 140,396 | 66,539 | 170,293 | 105,297 |
| Monthly average | 188,348 | 139,859 | 74% | 17,366 | 11,700 | 5,545 | 14,191 | 8,775 |

Browse searches are tabulated separately.

⁸ Total keyword Searches includes unqualified keyword searches as well as the following index specific keyword searches.

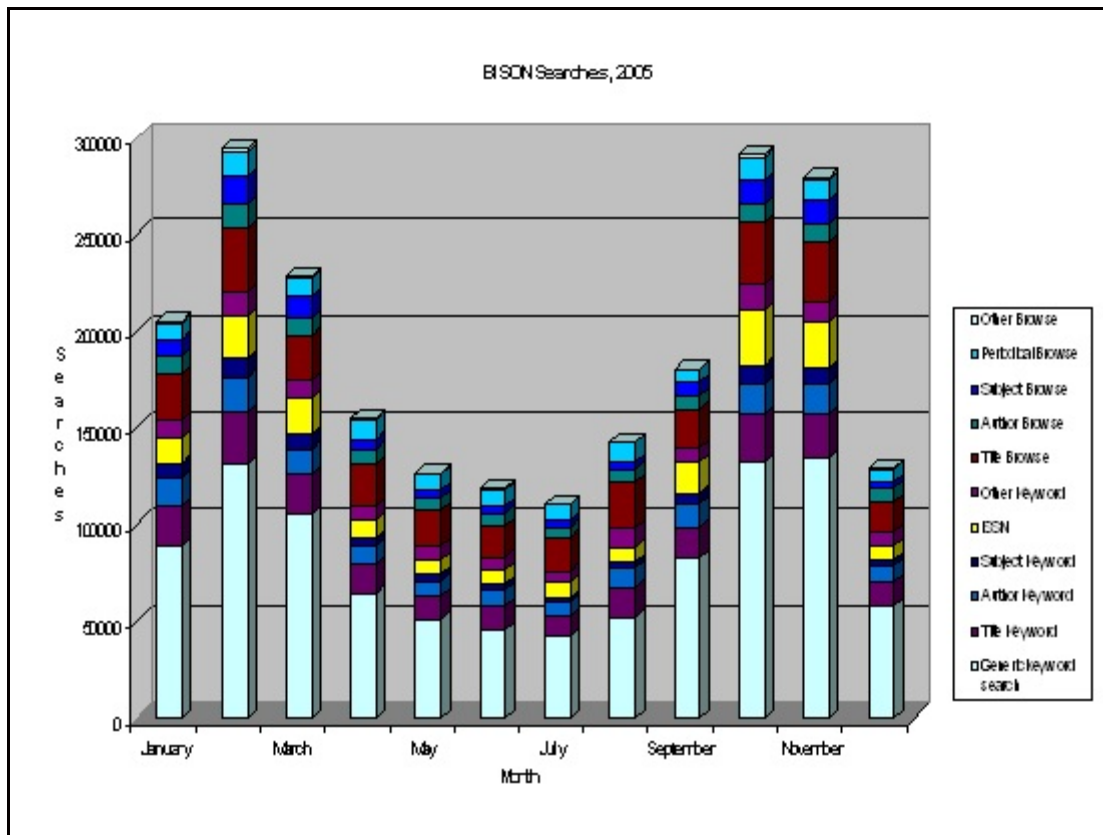
⁹ Other keyword searches include periodical title, series title and ISBN searches.

Table 2

| BISON Web Search Statistics | | | | | | | | |
|-----------------------------|----------------|-----------------|-----|--------------|---------------|----------------|-------------------|--------------|
| Browse Searches | | | | | | | | |
| Month | Total Searches | Browse searches | % | Title Browse | Author Browse | Subject Browse | Periodical Browse | Other Browse |
| 2005 January | 205,327 | 51,928 | 25% | 24,557 | 8,734 | 9,033 | 8,335 | 1,269 |
| 2005 February | 293,782 | 74,709 | 25% | 33,437 | 12,320 | 14,722 | 12,416 | 1,814 |
| 2005 March | 228,263 | 54,269 | 24% | 24,118 | 8,611 | 11,215 | 9,317 | 1,008 |
| 2005 April | 155,055 | 45,795 | 16% | 21,406 | 8,073 | 5,153 | 10,278 | 885 |
| 2005 May | 126,744 | 38,013 | 30% | 18,335 | 6,269 | 4,403 | 8,043 | 963 |
| 2005 June | 118,420 | 35,191 | 30% | 16,062 | 6,560 | 3,822 | 7,994 | 753 |
| 2005 July | 110,583 | 34,586 | 31% | 17,534 | 5,327 | 3,216 | 8,184 | 325 |
| 2005 August | 142,609 | 44,539 | 31% | 24,029 | 6,658 | 3,491 | 10,100 | 261 |
| 2005 September | 180,111 | 40,309 | 22% | 20,019 | 6,507 | 6,992 | 6,541 | 250 |
| 2005 October | 290,917 | 66,699 | 23% | 31,545 | 9,763 | 12,424 | 11,952 | 1,015 |
| 2005 November | 279,267 | 63,447 | 23% | 29,948 | 9,583 | 12,187 | 10,589 | 1,140 |
| 2005 December | 129,096 | 33,007 | 26% | 15,641 | 6,652 | 3,480 | 6,707 | 527 |
| 2005 Total | 2,260,174 | 582,492 | 26% | 276,631 | 95,057 | 90,138 | 110,456 | 10,210 |
| Monthly average | 188,348 | 48,541 | 26% | 23,053 | 7,921 | 7,512 | 9,205 | 851 |

In graphic form, the preponderance of unqualified keyword searching is even more apparent:

Figure 1



The above statistics demonstrate that, although keyword searching is the predominant search method of our users, there is a group of users who continue to use subject searches, either keyword subject searches or browse searches:

Table 3

| BISON Web Search Statistics | | | | | |
|------------------------------------|------------------|------------------|---------------|-----------------|-----------------------|
| Keyword Searches | | | | | |
| Month | Total Searches | Keyword Searches | % of searches | Subject Keyword | % of keyword searches |
| 2005 January | 205,327 | 153,399 | 75% | 6,943 | 5% |
| 2005 February | 293,782 | 219,703 | 75% | 9,959 | 5% |
| 2005 March | 228,263 | 173,994 | 76% | 7,376 | 4% |
| 2005 April | 155,055 | 109,260 | 70% | 3,719 | 3% |
| 2005 May | 126,744 | 88,731 | 70% | 3,410 | 4% |
| 2005 June | 118,420 | 83,229 | 70% | 3,143 | 4% |
| 2005 July | 110,583 | 75,997 | 69% | 2,439 | 3% |
| 2005 August | 142,609 | 98,070 | 69% | 2,691 | 3% |
| 2005 September | 180,111 | 139,802 | 78% | 6,025 | 4% |
| 2005 October | 290,917 | 224,218 | 77% | 9,443 | 4% |
| 2005 November | 279,267 | 215,820 | 77% | 8,574 | 4% |
| 2005 December | 129,096 | 96,089 | 74% | 2,817 | 3% |
| 2005 Total | 2,260,174 | 1,678,312 | 74% | 66,539 | 4% |
| Monthly average | 188,348 | 139,859 | 74% | 5,545 | 4% |

Although 4% is a very small percentage, it is my hypothesis that there is an additional use for subject headings in the overall searching of the catalogue. That is that the terms from subject headings improve the chances that users find material that they consider useful, even if they are not informed about the use of a fixed vocabulary

Methodology

For the purposes of this study, I selected a sample of three months from the past year, processed the files to extract only the searches that were done, and then extracted only the keyword searches that were unqualified to a particular index. Those searches that had produced a hit were then re-searched to determine what in the record had resulted in it being found. The hypothesis was that the subject and/or contents notes that we add to the records improve the search success. A total of 1,534 searches which had produced one or two hits were examined. The study was limited to these searches to ensure that the examination of the resulting records would be manageable.

The results of the search done during the study did not always match the original results recorded in the transaction logs. In some cases this was a result of additions

to the catalogue since the original searches were done. These records were eliminated from the study. In other cases it was necessary to return to the original log files to determine that the searches had been limited by one of the Optional Search Limits that are provided for the University of Manitoba Catalogue. These are: Limit by Format, Language, Library and Year. A total of 92 searches were limited by one or more of these properties.

Findings

Of the 1,543 searches that were examined, 154, or approximately 10%, were successful only because of a subject heading that had been added to the record. In addition, another 374 searches retrieved records because of a contents note or a summary or annotation note.

Table 4

Analysis of Searches

| | Searches examined | Success Due to Subject Headings: | | | | Success Due to Contents Notes | | Success Due to Summaries | | Total | |
|----------------|-------------------|----------------------------------|---------|--------|---------|-------------------------------|---------|--------------------------|---------|--------|---------|
| | | No | | Yes | | Yes | | Yes | | | |
| | | Number | Percent | Number | Percent | Number | Percent | Number | Percent | Number | Percent |
| May, 2005 | 1,070 | 962 | 89.9% | 108 | 10.1% | 195 | 18.2% | 18 | 1.7% | 213 | 19.9% |
| February, 2005 | 257 | 229 | 89.1% | 28 | 10.9% | 100 | 38.9% | 7 | 2.7% | 107 | 41.6% |
| October, 2005 | 216 | 198 | 95.7% | 18 | 8.7% | 49 | 23.7% | 5 | 2.4% | 54 | 26.1% |
| Total | 1,543 | 1,389 | 90.5% | 154 | 10.0% | 344 | 22.4% | 30 | 2.0% | 374 | 24.4% |

Figure 2

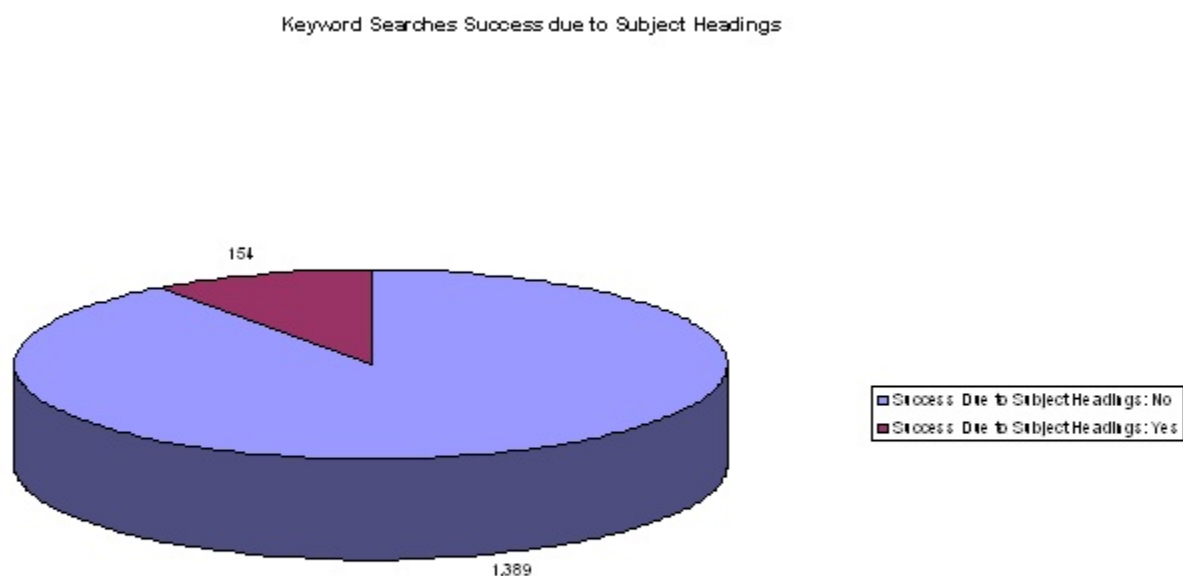
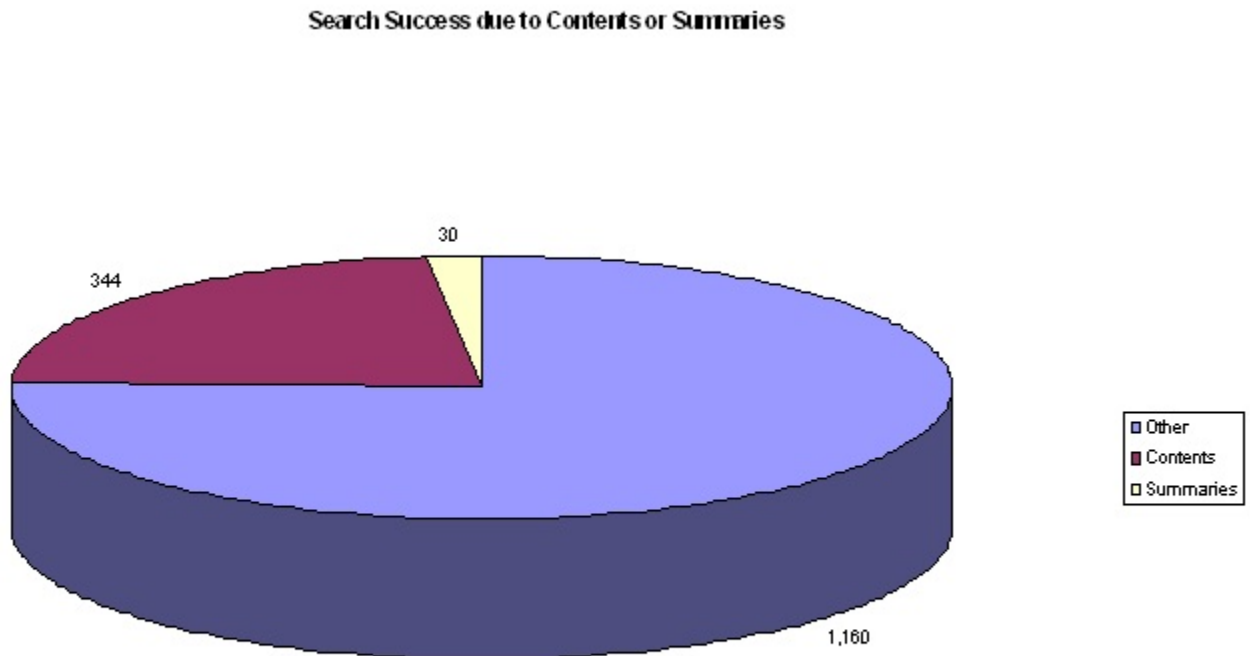


Figure 3



When combined with the searches where users intentionally search the subject indexes, subject headings become a significant resource for finding materials in the catalogue. As well, it can be seen from the results of this study that the table of contents notes and the summary/annotation notes that are added, also increase, significantly, the hits that are retrieved when using keyword searching.

Other Findings and Searches of Interest

Examining the searches performed by our users provided other insights beyond the parameters of the research study. Some of these are worth noting. For example, in each month that I examined, there were a number of searches done that included a url. (February, 35, May, 20 and October, 43). In each month there are searches such as pin change (and variations, February, 2; May, 2; October 8 and claim ID (February, 5; May, 4 and October, 17. There are also many searches for NETDOC, our local name for our networked databases, which are successful, for BISON, which might be

searches for the animal, or the University of Manitoba sports teams or our catalogue and 84 searches for yahoo mail in May and October. The search with the most hits was a search for t.p which returned 1,261,568 records in 11,593 milliseconds (i.e. 11.593 seconds).

Conclusion

Subject headings and contents and summary notes do improve retrieval of material from our libraries' catalogue. Although keyword searching has become the mainstay of our catalogues as our students and faculty apply their "Google" skills to our catalogues, the subject headings that we add to the catalogue records do help to ensure that searches are more successful. The contents notes and annotations that we have been adding to our records for approximately 10 years, also enhance the retrieval capabilities of our catalogues.

In addition, examining the searches that our users do in our catalogue provides useful insight into the issues that they have while using our libraries.

Bibliography

- Blecic, Deborah D. et al, "Using Transaction Log Analysis to Improve OPAC Retrieval Results," *College & Research Libraries*, 59 (1998), p. 41-50.
- Calhoun, Karen. *The Changing Nature of the Catalog and its Integration with Other Discovery Tools*, 2006. Retrieved 12 April 2006 from:
<http://www.loc.gov/catdir/calhoun-report-final.pdf>
- Connaway, Lynn Silipigni, John M. Budd and Thomas R. Kochtanek, "An Investigation of the Use of an Online Catalog: Users Characteristics and Transaction Log Analysis", *Library Resources and Technical Services*, 39 (1995), p. 142-151.
- Gross, Tina and Arlene G. Taylor, "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results", *College and Research Libraries*, 66 (2005) 212-230.
- Hamilton, Scott and Helen Thurlow, "Transaction log analysis @ State Library of Queensland". *Evolution of Evidence: Global Perspectives on Linking Research with Practice*. Retrieved 13 April 2006 from:
<http://conferences.alia.org.au/ebf2005/Hamilton.pdf>
- Hunter, Rhonda N. "Successes and Failures of Patrons Searching the Online Catalog at a Large Academic Library: A Transaction Log Analysis." *RQ*. 31 (1991), pp. 395-402
- Large, Andrew and Jamshid Beheshti, "OPACs: A Research Review," *Library & Information Science Research*, 19 (1997): p. 111-133.
- Morgan, Eric Lease. A "Next Generation" library catalog. Retrieved 12 July 2006 from:
<http://dewey.library.nd.edu/morgan/ngc/>
- Novotny, Eric. "I Don't Think, I Click: a Protocol Analysis Study of Use of a Library Online Catalog in the Internet Age." *College and Research Libraries*. 65 (2004), pp. 525-537.
- University of California Libraries. Bibliographic Services Task Force. Rethinking How We Provide Bibliographic Services for the University of California: Final Report. 2005. Retrieved 10 July 2006 from:
<http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>
- Wallace, Patricia M. "How do Patrons Search the Online Catalog When No One's Looking? Transaction Log Analysis and implications for Bibliographic Instruction and System Design." *RQ*. 32 (1993), pp. 239-252.