Is Keyword Searching the Answer?

Joy Tillotson

This study looks at three aspects of keyword searching to see if defaulting to keyword searches might serve as a solution to the problems people have with subject searches in online public access catalogs (OPACs). It investigates whether keyword searching produces useful results, if people who use keyword searches to find information on a subject say they are satisfied with the results, and how keyword searching and controlled vocabulary searching are offered and explained in currently available OPAC interfaces. The conclusion is that both keyword and controlled vocabulary searching ought to be easily available in an OPAC, and that improvements need to be made in explanations and help offered to subject searchers.

ince subject searching is often unsuccessful in library catalogs (studies report success rates from 12 percent to 75 percent), it is tempting to keep looking for ways to help users get more satisfactory results.[1] Because I had observed that reference librarians use keyword searching more than library users (and this observation was confirmed in a study from Adelphi University), I decided to investigate what might happen if all searches were automatically done as keyword searches, and what effect use of keyword searches has on the level of satisfaction of library users.[2] Concluding that some changes in searching styles might be useful, I surveyed available OPAC interfaces to see whether keyword searching was easily available and whether the instructions and help offered to users of controlled vocabulary searching mentioned keyword searching.

PART 1: EVALUATION OF KEYWORD SEARCH RESULTS

Joan Cherry found that searches done as keyword rather than as exact searches (ones where the OPAC expects controlled vocabulary to be used) are more likely to produce some results.[3] She evaluated the results of the searches to the extent that she reported the searches as unsuccessful if they retrieved more than 500 citations. Jennifer Rowley reported two concerns with keyword searching. One concern is that the number of irrelevant citations in a keyword search might overwhelm the relevant ones, especially when all words in the record are searched. The other concern is that keyword searches do not retrieve all the relevant material.[4] It is possible that these concerns make interface designers reluctant to choose keyword searching as a default method. I designed the first part of this study to test these beliefs by examining the results of keyword searches for relevance.

Methodology

In this part of the study I looked at 400 subject searches in two sizes of catalog. The OPAC at Memorial University of Newfoundland has about 700,000 records and the one at the University of Toronto has about 7 million records. This

Joy Tillotson is Head of Information Services, Queen Elizabeth II Library, Memorial University of Newfoundland.

allowed me to check whether the size of the catalog affected the number of relevant records retrieved by keyword searching. The 400 searches for the study were subject searches from transaction logs on the two catalogs—200 from each catalog. Half the searches chosen had produced no hits when originally searched; i.e., they had been entered as if they were Library of Congress Subject Headings, but they did not match correct subject headings closely enough to be found by browsing nor did they match cross-references provided by the OPACs. I did each of the 400 searches as a keyword search on the catalog it came from and recorded the following:

- the total number of items retrieved;
- the percentage of relevant items on the first two screens of short records displayed;
- the number of relevant items among the first 30 short records; and, where possible,
- the overlap between the keyword searching set and a set produced by searching using appropriate Library of Congress Subject Headings.

There is some evidence from a study by Joseph Janes that experienced librarians are able to judge reasonably well which citations a user would find relevant.[5] There is also an advantage in having a single person judge the relevance in each case, namely, that the results are more consistent than is the case when using the judgments of individual users. As a check, I randomly selected a subset of 100 searches from the 400 and had a senior library school student repeat these searches on both catalogs. I chose searches to produce a sample that consisted of equal numbers of searches from each site and equal numbers of searches that had produced some or no hits.

According to Janes, experienced library school students are not quite as good as librarians at judging relevance.[6] In this case, the student was generally more conservative in her judgments of the number of relevant citations. Because the student did the same searches in both catalogs, while I did the searches only in the catalog whose transaction logs they came from, the results cannot be compared exactly. Therefore I have used only the student's results to compare differences that may be because of the database's size.

The user whose behavior both the student and I attempted to approximate is the undergraduate looking for a couple of books on a topic. To do this, we checked the short bibliographic records displayed on the first two screens of results (7 to 12 records) for relevant items. Since not all OPACs display 7 to 12 records on the first two screens, the student and I looked also at the first 30 titles for each search. This should make it easier for people to compare our results with a similar project using their own OPAC.

I established guidelines about what would count as a relevant search and we both followed them. For example, the student and I counted a record as relevant if the search words (or synonyms or grammatical variants, e.g., sex for sexual) appeared in the title or elsewhere in the brief record. To get an idea of recall, we chose an appropriate subject heading or headings by consulting Library of Congress Subject Headings or by inspecting the subject headings in good titles retrieved by keyword searches (or a combination of the two strategies). We did a search using the controlled vocabulary term(s) and combined the resulting set of citations with the set from the keyword search using a Boolean AND to find the amount of overlap between the two sets.

Results

One argument against keyword searching is that the user will get too many useless references. The keyword searches did produce large results in some cases—an average of 1,063 citations per search at the University of Toronto (the larger database) and an average of 431 citations per search at Memorial. The results were not ten times larger in the larger database, however, nor were the resulting sets that much larger than the sets retrieved by controlled vocabulary searching (an average of 894 in the large database and 294 in the

smaller one). Despite the large sets, useful citations were often found right away. Because we were trying to model the behavior of a person who is looking for a few good books, the first thing we looked at was whether there were any good citations on the first two screens of results. The first two screens of citations (i.e., the first 7 to 12 citations) included at least one relevant item most of the time (see table 1).

In each search set, 50 percent of the searches had produced no hits at all when done as controlled vocabulary searches. Therefore, these results show a significant improvement resulting from using keyword searching. Recall that the library school student was doing the same searches on both databases and found useful results slightly more often in the smaller database, which is what might be expected.

A more standard method of measuring success in searching is to look at precision and recall. Precision is the percentage of citations that are relevant in the set of citations retrieved. I did not look at the whole set of citations because many sets were large and there is little evidence that users look at every citation in large sets. I report here only the precision of the first 30 citations. Table 2 shows that precision is somewhat better in the smaller catalog, again as might be expected. The average precision that was found in 14 studies of online bibliographic databases varied from 17 percent to 81 percent. It is hard, therefore, to draw any conclusion about how precise keyword searching is, compared to other methods, except that it is on the lower end of the range.[7]

The other argument against using keyword searching is that the searcher will not get all the relevant material. Since it is very difficult to establish the set of all relevant material in such large databases, we created sets of relevant material. The student and I achieved this by doing subject searches using the search terms if they matched LC Subject Headings or by finding the closest matches in Library of Congress Subject Headings. Sometimes no good match

was found, particularly with very narrow topics. We assumed that the set of citations retrieved by the controlled vocabulary search was "all the relevant material" and checked to see how much of it was retrieved by the keyword search of the same topic. Table 3 shows what percentage of the relevant material was retrieved by the keyword searches.

The figures in table 3 look very high, given reported average recall rates of 41 to 61 percent.[8] Of course, half the searches were exact or near matches of LC subject headings; as a result, about half the time the keyword search retrieved 100 percent of the relevant material. A keyword search retrieved all of the relevant materials 33 percent of the

TABLE 1
PERCENTAGE OF SEARCHES WITH
RELEVANT CITATIONS ON FIRST
TWO SCREENS USING KEYWORDS

	% Larger Database	% Smaller Database
Librarian	75	71
Student	69	72

TABLE 2
PRECISION OF
KEYWORD SEARCHING

	% Larger Database	% Smaller Database
As judged by librarian		
On first two screens	40	40
In first thirty titles	31	43
As judged by student		
On first two screens	27	31
On first thirty titles	20	22

TABLE 3
AVERAGE RECALL OF
RELEVANT MATERIAL

	% Larger Database	% Smaller Database
Librarian	68	73
Student	65	63

time when a controlled vocabulary using the same words retrieved nothing. In 10 percent of the cases, a keyword search retrieved none of the relevant citations. In general then, it seems searchers could expect to retrieve many of the relevant citations and that the sets, although large, would not be so full of irrelevant citations as to be useless. Also, as previously discovered, keyword searches were often successful when an exact search on the same terms produced no results.

From these results it seemed likely that people who used keyword searching were more likely to be satisfied with the results. The next step was to try to verify this by asking users.

PART 2: ONLINE SURVEY AND TRANSACTION LOGS

Since it seemed likely from part 1 of the study that people who used keyword searching would be more satisfied with the results, part 2 looked at the satisfaction reported by users who used keyword searching compared to that of searchers who used controlled vocabulary searching. An online survey of user satisfaction with the OPAC was conducted at the University of Toronto.[9] The OPAC uses Data

Research Associate's Information Gateway interface that offers users a choice of exact and keyword searching methods on a high level menu (see figure 1).

One survey question was "What did you find in your search session today?" The possible responses were: nothing you were looking for, some of what you were looking for, enough of what you were looking for, exactly what you were looking for, and more than you were looking for. For 189 of the survey respondents, transaction logs were available so that I could check whether they had used exact or keyword searches and relate the search technique to the response to the survey question. Of these 189 respondents, 29 people said in the survey that they were planning to search by subject only. Others were doing more than one type of search, e.g., searching by subject and looking for a specific item. Eleven of those searching by subject used controlled vocabulary searches only, and 18 used keyword searching of some kind. Figure 2 shows that people using keyword searches were less satisfied with their results.

I do not like to draw major conclusions from so few searches, but the results do

```
 ┌──────────────────<UTLink -- Searching: UTCat>──────────────────┐
 │ Find...              UTLink Menu...          Welcome          Help... │
 │ ┌──────────────┐                                                      │
 │ │Author        │  For:                       For:                     │
 │ │AuthOr keyword│                                                      │
 │ │Title         │     * other libraries          * starting over       │
 │ │TitlE keyword │  s  * Internet resources                             │
 │ │Subject       │     * electronic publications                        │
 │ │SubJect keyword│    * University information                         │
 │ │Keywords      │     * exiting/logging off                            │
 │ │Numeric...    │                                                      │
 │ ├──────────────┤                                                      │
 │ │Choose Database│ Type the highlighted letter to choose a menu        │
 │ └──────────────┘                                                      │
 │                              OR                                       │
 │                                                                      │
 │           Press the NEXT key for an introduction to using UTLink      │
 │                                                                      │
 │                  Press the HELP key at any time.                     │
 │──────────────────────────────────────────────────────────────────── │
 │ Press the NEXT key for more.                                         │
 │             Press RETURN to begin an Author search.                  │
 └──────────────────────────────────────────────────────────────────────┘
```

FIGURE 1
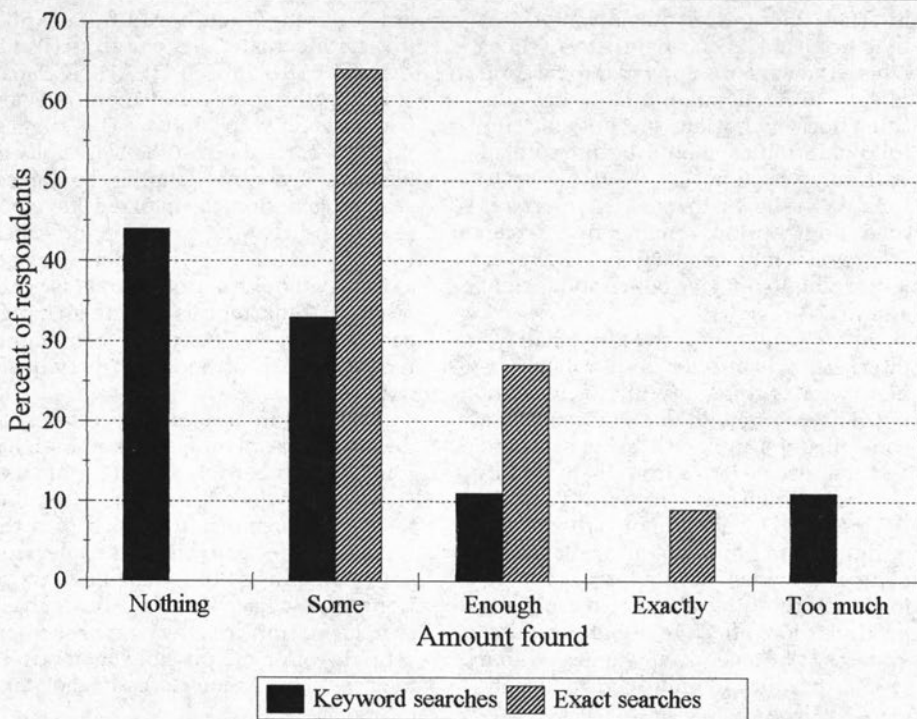Choices Offered to the Searcher on the University of Toronto OPAC

FIGURE 2
What Twenty-nine Subject Searchers Said They Found:
Comparison of Keyword and Controlled Vocabulary (Exact) Searches

not point to keyword searching as a panacea for the problems of subject searching. Part of what happened is that people resorted to keyword searches when an exact search failed and then found nothing they liked with the keyword search either. Some keyword searches produced citations that appeared to be about the topic, but were still declared unsuccessful by the searcher.

PART 3: SURVEY OF OPAC INTERFACES

I concluded from parts 1 and 2 that both keyword and controlled vocabulary searching ought to be available. Since other writers have come to the same conclusion, I decided to complete the study by seeing how well this idea was implemented in OPAC interfaces currently in use.[10] I decided to look at

whether keyword searching was an obvious option in most OPAC interfaces and how well the differences between keyword and controlled vocabulary searching were explained. My curiosity on these points was fueled by two recent research results. One was Terry Ballard's finding that a change in the amount of keyword searching followed a change in the position of the keyword search option on the menu.[11] The other was the fact that about half (48.3 percent) of the respondents in the survey discussed above were not aware that they should be using Library of Congress Subject Headings when they chose to search by subject.

I looked at 17 OPAC interfaces available on the Internet, using the list provided by Hytelnet (version 6.3). I chose the first North American (or first English-language) site listed for each

interface type and left out only interfaces that occurred at a single site or ones where the versions appeared to be completely different at different sites. Because local variations are possible, the following tables should be interpreted with some caution.

Table 4 shows that many interfaces offer both options on the first screen where searching can be done. However, 35 percent do not give both options right away.

Table 5 shows that fewer than half the interfaces which offer both options explain what is going on with exact searching. That is, they do not mention that a controlled vocabulary is being searched and that use of terms from that vocabulary might give better results.

This lack of explanation is disturbing in light of literature about mental modeling and evidence from technical writing literature that better instructions produce more effective use of computer systems.[12-15] Steven Zink suggested in a 1991 article that an improvement in subject searching might result from explicitly stating that "Use of LIBRARY OF CONGRESS SUBJECT HEADINGS may result in the location of more relevant materials on your topic."[16] This suggestion was carried out at the University of Nevada, Reno, by changing the Subject searching option to invoke a keyword search and creating a new option called LC Subject Heading which does an exact search. Although a formal study is yet to be done, Zink reports that this change has resulted in "far fewer complaints regarding lack of books on previously noted subjects."[17]

Although interfaces may offer both methods of searching and explain what is meant by them, it is almost certain that some users will not pay close attention to the instructions. Table 6 shows the type of help available to people who enter something other than a Library of Congress Subject Heading at a subject search prompt. It is very rare for interfaces to offer all possible methods of help, and many interfaces do not take

TABLE 4
SEARCH OPTIONS BY INTERFACE TYPE

Interface and Location	Both Search Methods on First Screen	Keyword Search Only on First Screen	Exact Search Only on First Screen
BLCMP (Birmingham)		X	
BuCAT (Athabasca)		X	
CATS (Cambridge)	X		
DOBIS (Emory)			X
DRA Atlas (Abilene)	X		
Dynix (Alma)	X		
GEAC (Atlantic School of Theology)	X		
INNOPAC (Ada)	X		
Libertas (Bristol)		X	
LS/2000 (Appalachian)	X		
Mutilis (Baylor)			X
NOTIS (Auburn)	X		
PALS (Creighton)	X		
UNICORN (Kennesaw)	X		
URICA (Australian National University)	X		
VTLS (William and Mary)	X		
Z39.58 (Dartmouth)		X	

TABLE 5
EXPLANATIONS OF CONTROLLED VOCABULARY SEARCHING

Interface and Location	Explains Exact Searching	Does Not Explain	Does Not Have Exact Searching
BLCMP (Birmingham)			X
BuCAT (Athabasca)			X
CATS (Cambridge)	X		
DOBIS (Emory)		X	
DRA Atlas (Abilene)		X	
Dynix (Alma)		X	
GEAC (Atlantic School of Theology)	X		
INNOPAC (Ada)	X		
Libertas (Bristol)			X
LS/2000 (Appalachian)		X	
Mutilis (Baylor)		X	
NOTIS (Auburn)		X	
PALS (Creighton)	X		
UNICORN (Kennesaw)		X	
URICA (Australian National University)	X		
VTLS (William and Mary)	X		
Z39.58 (Dartmouth)		X	

TABLE 6
HELP OFFERED TO SEARCHERS WHO GET
NO HITS WITH A SUBJECT SEARCH

Interface and Location	Uses LC Cross-references	Displays Alphabetical Near Matches	Uses or Suggests Keyword Searching
BLCMP (Birmingham)			X
BuCAT (Athabasca)			X
CATS (Cambridge)	X	X	
DOBIS (Emory)		X	
DRA Atlas (Abilene)	X		
Dynix (Alma)	X	X	
GEAC (Atlantic School of Theology)		X	
GEAC Advance (Boise State)	X		
INNOPAC (Ada)	X	X	X
Libertas (Bristol)			X
LS/2000 (Appalachian)		X	
Mutilis (Baylor)			X
NOTIS (Auburn)	X		X
PALS (Creighton)			X
UNICORN (Kennesaw)		X	X
URICA (Australian National University)			X
VTLS (William and Mary)	X	X	
Z39.58 (Dartmouth)			X

advantage of the possibility of keyword searching.

CONCLUSIONS

From part 1 of this study, I concluded that if users had entered the same terms as keywords instead of subjects, they would have had more satisfactory results. However, part 2 did not confirm that users were happier with the results when they used keyword searching. It did indicate that, given an easy choice, people will choose to use both keyword and controlled vocabulary searching; as a result, I concluded that both should be available. From part 3, it was clear that OPAC interfaces are doing reasonably well at offering both options but less well at explaining the use of controlled vocabulary and offering help with unsuccessful searches. In light of the fact that increasing numbers of users are using our catalogs from remote locations with no access to lists of acceptable subject headings, it would seem that improvements in these last two areas should be a priority for improvements to OPAC interfaces.

REFERENCES AND NOTES

1. Thomas A. Peters, "The History and Development of Transaction Log Analysis," *Library Hi Tech* 11, no. 2 (1993): 41–66.
2. Terry Ballard, "Comparative Searching Styles of Patrons and Staff at a University Library," *Library Resources and Technical Services* 38, no.3 (July 1994): 293–305.
3. Joan M. Cherry, "Improving Subject Access in OPACs: An Exploratory Study of Conversion of Users' Queries," *Journal of Academic Librarianship* 18, no. 2 (May 1992): 95–99.
4. Jennifer Rowley, "The Controlled Versus Natural Indexing Languages Debate Revisited: A Perspective on Information Retrieval Practice and Research," *Journal of Information Science* 20, no. 2 (1994): 108–19.
5. Joseph W. Janes, "Other People's Judgements: A Comparison of Users' and Others' Judgements of Document Relevance, Topicality, and Utility," *Journal of the American Society for Information Science* 45, no. 3 (Apr. 1994): 160–71.
6. Ibid.
7. C. H. Fenichel, "The Process of Searching Online Bibliographic Databases," *Library Research* 2 (Summer 1980–81): 107–27.
8. Ibid.
9. Joan M. Cherry, Joy Tillotson, and Marshall Clinton, "Online Survey of OPAC Users in a Large Academic Library: A Profile of Users and Their Satisfaction" (unpublished report, University of Toronto, 1994).
10. Martha M. Yee, "System Design and Cataloging Meet the User: User Interfaces to Online Public Access Catalogs," *Journal of the American Society for Information Science* 42, no. 2 (Mar. 1991): 78–98.
11. Ballard, "Comparative Searching Styles."
12. William K. Horton, *Designing & Writing Online Documentation* (New York: Wiley, 1990), 37.
13. Nathaniel S. Borenstein, "Help Texts vs. Help Mechanisms: A New Mandate for Documentation Writers," in *ACM SIGDOC '85: The Fourth International Conference on Systems Documentation: Proceedings* (New York: Assn. for Computing Machinery, 1985), 78–83.
14. Henry Ledgard et al., "The Natural Language of Interactive Systems," *Communications of the ACM* 23, no. 10 (Oct. 1980): 556–63.
15. M. A. Sullivan and A. Chapanis, "Human Factoring a Text Editing Manual," *Behaviour and Information Technology* 2, no. 2 (Apr./June 1983): 113–25.
16. Steven D. Zink, "Monitoring User Search Success through Transaction Log Analysis: The WolfPAC Example," *Reference Services Review* 19, no. 1 (1991): 49–56.
17. Steven D. Zink, electronic mail message to author, May 16, 1994.