

Subject No-hits Searches in an Academic Library Online Catalog: An Exploration of Two Potential Ameliorations

Rumi Y. Graham

This paper describes a study that explored ways in which users' subject-searching problems in a local online catalog might be reduced. On a weekly basis, the author reviewed catalog transaction logs to identify topics of subject searches retrieving no records for which appropriate information resources may actually be represented in the catalog. For topics thus identified, the author explored two potential ameliorations of the no-hits search results through the use of authority record cross-references and "pathfinder" records providing brief instructions on search refinement. This paper describes the study findings, discusses possible concerns regarding the amelioration methods used, outlines additional steps needed to determine whether the potential ameliorations make a difference to users' searching experiences, and suggests related areas for further research.



The landmark series of online catalog studies sponsored by the Council of Library Resources (CLR) in the early 1980s established unequivocally that subject searching was the most popular type of online catalog search and that catalog users experienced the most difficulty with subject searches.¹⁻² Spurred by the findings of the CLR studies, discussions and research efforts to improve subject searching in online catalogs flourished throughout the 1980s and early 1990s.

Researchers attempted or suggested ways to improve subject searching by, for example, enriching the subject content of the catalog database; enhancing browse displays and other aspects of the user-system interface; increasing the sophisti-

cation and power of the catalog search engine; and exploring user-centered rather than system-centered philosophies of, and approaches to, system design and improvement.³⁻¹⁹ Curiously, despite the momentum built up during this period of concentrated research, the online catalog sector of the subject access research front became comparatively quiet over the subsequent decade.

Web Solutions?

Reasons for the loss of momentum in research addressing the improvement of online catalog subject searching over the 1990s remain unclear. However, we may speculate that the slowdown was perhaps indicative of the extremely complex nature of the remaining subject -searching

Rumi Y. Graham is Manager, Collections and Database Services, at the University of Lethbridge Library in Lethbridge, Alberta, Canada, and a doctoral student in information studies at the University of Toronto; e-mail: graham@uleth.ca.

problems. Another possibility is that research efforts were diverted elsewhere toward, for instance, the development of Web interfaces to online catalogs and other information retrieval systems or the creation of new Web-based information retrieval tools and resources.

The mid-1990s emergence of Web interfaces to online catalogs heralded the commencement of significant expansions to the accessibility and content of library catalogs. For all their added reach and content breadth, however, it is disappointing that the subject-searching capabilities of Web catalogs appear to be much the same as those of pre-Web, second-generation systems.²⁰

In a discussion of new approaches to online subject access at the end of the 1990s, Sandy Roe listed five common subject-searching problems encountered "first in the OPAC, often still in Web-based library catalogs, and certainly in general Web searching today:" The first was "no items retrieved," and the second was "too many items retrieved."²¹ A year later, Chris Evin Long undertook a critical evaluation of subject searching in sixty Web catalogs and concluded that of the "Web-based OPACs currently in operation ... many deficiencies present in earlier generations of online catalogs have been passed down to the next generation."²²

Subject-searching Problems in a Local Catalog

The proposition that the problem of "no items retrieved" (hereafter referred to as no-hits searches) persists in today's Web catalogs is supported by use statistics for the University of Lethbridge library online catalog, whose existence has spanned the pre-Web and Web interface eras. This catalog had a character-based interface from its installation in 1990 up to 1997, when the library began to offer an additional Web-based interface. From 1999, the Web interface became the primary catalog interface, although the original character-based interface remained available.

The year 1999 also marked the implementation of an agreement between the University of Lethbridge and Medicine Hat College, located 200 kilometres apart in southern Alberta, to share the university's Innopac automated library system, which included the online catalog.²³ From 1990 to 2002, the highest percentage of all searches in this catalog most often occurred in the subject index, and the subject index consistently accounted for the highest or second-highest percentage of all no-hits searches.²⁴ Thus, conversion to a Web interface alone did not result in marked changes in the frequency of subject no-hits search results in the University of Lethbridge catalog, which remained significantly high.

Suppositions and Goals of the Study

The impetus to investigate subject-searching problems in the University of Lethbridge catalog arose from three suppositions:

- Some subject no-hits searches represent topics covered in cataloged library materials that may not have been discovered by the user.
- Some subject no-hits searches may recur periodically (e.g., from semester to semester or year to year).
- Some subject no-hits search results may be preventable if different catalog responses can be engineered, leading to the retrieval of some records that future users may judge to be relevant or worth pursuing.

These suppositions gradually coalesced out of many years of gleaning indications of subject-searching problems from experiences of assisting students with their research assignments at the reference desk and from the examination of two types of transaction log data: descriptive statistics on catalog use, and the text of users' subject no-hits search terms (which the author scanned periodically).²⁵ On the one hand, we know that transaction log data alone provide no information about users' intentions or needs. That is, some searches resulting in no retrievals may be judged by the user to be successful because the no-hits results were

TABLE 1
Studies of Subject No-hits Searches in Online Catalogs

Authors	Data Collection Period	# Subject No-hits Searches Analyzed/Logged	Subject Searches Resulting in No-hits
Lester (1989)	120 days**, 1989	1,524	60.1%
Tolle (1983)	17 days, 1982 (SULIRS)	9,235	52.8%
Kern-Simirenko (1983)	1 day, 1982 (System A)	204	46.0%
Wilkes & Nelson (1995)	14 days*, 1991	351	40.6%
Kern-Simirenko (1983)	1 day, 1982 (System C)	69	39.0%
Kern-Simirenko (1983)	1 day, 1982 (System B)	231	35.0%
Barrett & Maticka (1990)	5 days, 1989?	3,215	32.0%

Data collection period durations originally reported in weeks* or months** were converted to days for ease of comparison.

expected and, conversely, some searches retrieving at least one record may not be successful because the user judges all records to be irrelevant or an overwhelming number of records are retrieved.

On the other hand, we tend to think of no-hits search results as representing unsatisfactory results most of the time. Because users' exact no-hits search terms can be captured in transaction logs, analysis of these data may assist us in taking initial steps toward achieving a better understanding of the characteristics of subject no-hits searches and toward developing methods to ameliorate them effectively. An interest in discovering what could be done locally to improve subject searching led the author to embark upon an exploration of the third supposition using transaction log data.²⁶ The goals of the present study were to examine users' subject no-hits search terms systematically in order to determine what types of searches are logged as subject no-hits searches in the University of Lethbridge catalog and to explore possible methods of enhancing the database that could potentially reduce the occurrence of subject no-hits searches.

Previous Research

Defined as "the study of electronically recorded interactions between online information retrieval systems and the persons who search for the information

found in those systems," transaction log analysis is a popular means of studying system performance and use of online catalogs and other information retrieval systems.^{27,28} A number of studies have used transaction log analysis to investigate subject no-hits search results in online catalogs, a sampling of which is presented in table 1.²⁹⁻³³ As well, a longitudinal transaction log study used analyses of no-hits search results to improve online catalog searching through enhancements to the catalog search interface, and one result of this study was a significant decrease in subject no-hits search results.³⁴

Another possible means of reducing subject no-hits search results is the inclusion of actual no-hits search terms as cross-references in authority records, thereby automatically leading the user from no-hits search terms to potentially relevant controlled vocabulary terms such as the Library of Congress Subject Headings (LCSHs). The potential of new authority record cross-references for reducing the frequency of no-hits searches in online catalogs has been explored in several studies, but with a focus on author searches or author and title searches, representing known-item searches rather than unknown-item (subject) searches.³⁵⁻³⁹

In 1982, a project led by Pauline A. Cochrane established a procedure for staff

at other libraries to forward suggestions to the Library of Congress (LC) for consideration as new LCSH cross-references.⁴⁰ In an appendix to the project's final report, Cochrane explained that the intent behind the project was to improve LCSH for use in online catalogs. This successful project laid the groundwork for a cooperative approach to maintaining the LCSH authority file, now known as the Subject Authority Cooperative Project (SACO) coordinated by LC. No other published studies were located that attempted to improve subject searching in online catalogs through new LCSH cross-references, although, in her assessment of the state of reference structures in online catalogs, Jane Greenberg noted that "there are many examples of librarians manipulating authority control modules and authority data in order to improve reference structure access."⁴¹

Methods

For the duration of the study, on the first working day of each week, complete lists of subject no-hits searches from the previous week were downloaded from transaction logs of the University of Lethbridge catalog. The author analyzed each no-hits search in the lists using an exploratory and interpretive process of seeking answers to a series of questions:

- Is the search string intelligible (i.e., can plausible meaning be inferred)?
- If intelligible, what topic or topics might the search string represent?
- Does the catalog contain information that might be relevant to the search?
- If the catalog potentially contains information, how might the database be enhanced to prevent the recurrence of the no-hits search result?

To answer the last two questions, topics potentially represented by the subject no-hits searches were searched again in the Innopac cataloging subsystem where the author was able to view, update, and create MARC authority and bibliographic records, as needed. After some initial experimentation, the following two processes were developed as potential en-

hancements to selected subject no-hits search results when the catalog was found to contain records that might be relevant to an inferred topic of a subject no-hits search:

1. creating or upgrading an authority record to provide a link between terms comprising a no-hits search and a valid LCSH represented in at least one bibliographic record in the catalog; or
2. creating a "pathfinder" record using the bibliographic record format to provide a brief set of instructions on how to retrieve some potentially relevant records using the catalog's "LIMIT" feature.⁴²

Enhancement 1 involved adding cross-references comprising the verbatim text of selected subject no-hits searches to MARC authority records whose authorized headings were interpreted to be possibly the same as, or related to, the new cross-references.⁴³ Verbatim no-hits search terms were used as cross-references because these terms were users' natural language vocabulary and thus represented terms that other catalog users might use in subject searches. The author assumed that converting subject no-hits search terms to be in compliance with LC policy governing the creation LCSH cross-references may be less helpful, a supposition partially corroborated by a study of catalog users' understanding of LCSH that found that users assigned correct meanings to less than half of the LCSHs they examined.⁴⁴

Most subject no-hits searches potentially representing topics covered by materials in the library's cataloged collections were found to be amenable to enhancement 1. Some, however, were discovered to represent combinations of topics (e.g., weight and body image, and ethics in sport psychology) for which no single LCSH could be found. Although it is possible to add a given cross-reference to more than one authority record, the author was doubtful that users would find two or more "see" references to be helpful, if individually, none matched their exact search topic.

Thus, enhancement 2 was developed as a more direct means of leading users to potentially relevant records when appropriate single LCSHs could not be found. It involved the adaptation of the bibliographic record format to serve as a "pathfinder" record, defined here as a brief record whose purpose is to provide instructions on how to conduct two-step searches on particular topics. The authority record format is unable to perform the function of a pathfinder record because it does not support the direct presentation of search instructions to catalog users when a field containing an authorized heading is not present.

For each no-hits search interpreted to be intelligible, the author determined whether application of enhancement 1 or 2 could lead to the retrieval of at least one potentially relevant record, but not too many. However, what constitutes "too many" is a situation-specific and subjective judgment that can be made only by the user and varying suggestions are found in the research literature on the number of retrievals that users consider to be too many.⁴⁵ Lacking

empirical data on what University of Lethbridge catalog users considered to be too many records retrieved, the author attempted to minimize information overload for users by adding database enhancements resulting in fewer than 100 records retrieved, whenever possible.⁴⁶

The author recognized that, on the one hand, regarding the two enhancements as true improvements to the database is easily challenged because both involved significant deviation from standard subject cataloging practice. On the other hand, the purpose of exploring nonstandard uses of authority and bibliographic records was to support the main goal of subject cataloging, which is, in the author's view, the effective communication of the main subject matter of documents to catalog users. When we suspect that a significant proportion of users' needs may not be adequately met, it is legitimate to consider alternatives to current cataloging practices and policies in order to serve catalog users better. Limitations represented by the nonstandard nature of the two database enhancements

FIGURE 1
Enhancement 1 Example

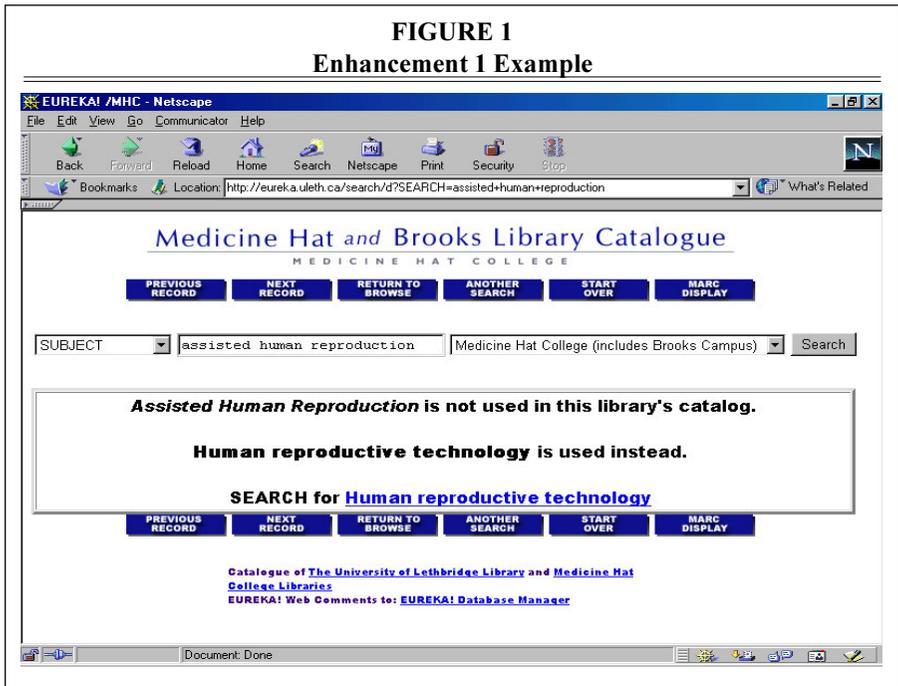
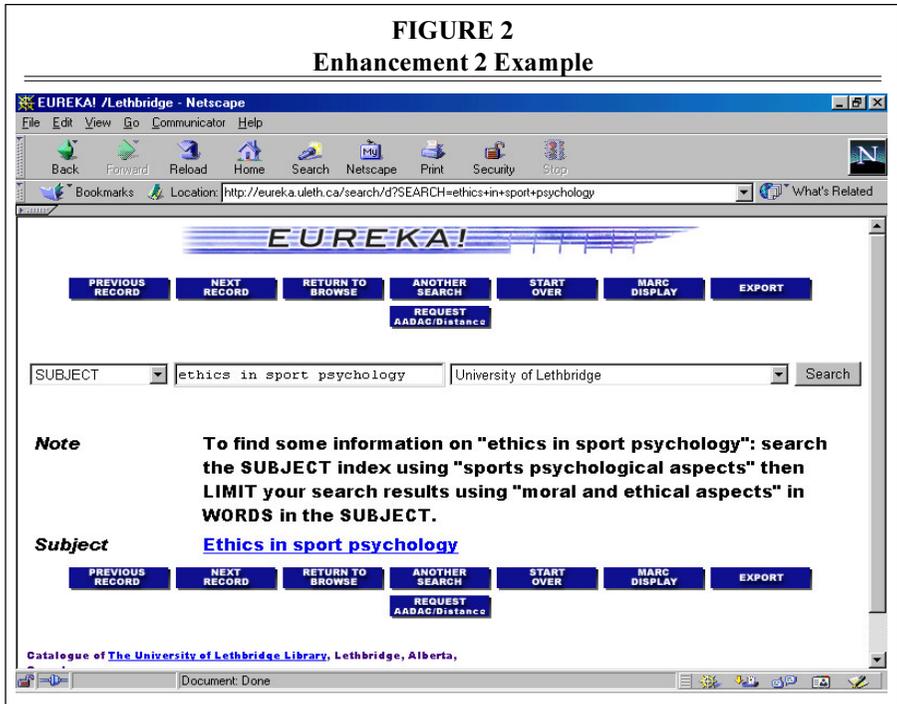


FIGURE 2
Enhancement 2 Example



explored in this study are further considered in the discussion section.

Database Enhancement Examples

“Assisted human reproduction” is an example of a subject no-hits search result that was modified using enhancement 1. This search was logged in the Medicine Hat College subset of the shared University of Lethbridge catalog on March 20, 2002. Before modifications were made, the catalog response to this search was a “nearest neighbors” browse display, with the highlighted message “Your entry *assisted human reproduction* would be here – *Search as Words.*”

If the user selected the option to “Search as Words,” the catalog would have automatically reexecuted the search for “assisted human reproduction” in the keyword index, where some records may have been retrieved and perhaps judged by the user to be useful. Hence, it is possible that the user judged the outcome of this no-hits search to be successful, although transaction log data show that the “Search as Words” option is rarely used.

Changing the database using the enhancement 1 methods involved adding “assisted human reproduction” as a cross-reference to the subject authority record for “human reproductive technology.” “Human reproductive technology” is an LCSH that was interpreted by the author to be similar in meaning to assisted human reproduction and was present in at least one catalog record in the database. After this database enhancement was made, searching the subject index for assisted human reproduction invoked a new system response illustrated in figure 1. As long as at least one bibliographic record containing a subject heading beginning with “human reproductive technology” continues to exist in the catalog, if a user encountering the display in figure 1 clicks on “Human reproductive technology,” the catalog automatically performs another subject search that retrieves all records containing this LCSH.

An example of enhancement 2 is found in the “before” and “after” catalog responses to the subject search for “ethics in sport psychology,” which was logged

as a no-hits search on March 19, 2002. Before enhancement, the catalog presented the standard “nearest neighbors” browse display containing the user’s original search terms with the option to re-search the terms in the keyword index. Once again, if the user selected the “Search as Words” option from this display, it is possible that the search for ethics in sport psychology actually resulted in useful retrievals via an automatically reexecuted search in the keyword index.

Because no LCSH incorporating all of the topical elements of “ethics in sport psychology” was identified, the author was unable to create an authority record cross-reference for this no-hits search. The author thus determined that enhancement of the catalog response to this particular no-hits search was better handled in a pathfinder record containing the following note: “To find some information on ‘ethics in sport psychology’: search the SUBJECT index using ‘sports psychological aspects,’ then LIMIT your search results using ‘moral and ethical aspects’ in WORDS in the SUBJECT.” The enhanced catalog response to the subject search for ethics in sport psychology is illustrated in figure 2.

Following the instructions in the new pathfinder record for ethics in sport psychology on October 30, 2002, led to the retrieval of one record: *The Making of High Performance Athletes: Discipline, Diversity, and Ethics*. The two LCSHs in the retrieved record were “Sports—Psychological aspects” and “Sports—Moral and ethical aspects.” Arguably, these two headings together cover the topic ethics in sport psychology, whereas, singly, neither one addresses all three component concepts of sports, psychology, and ethics.

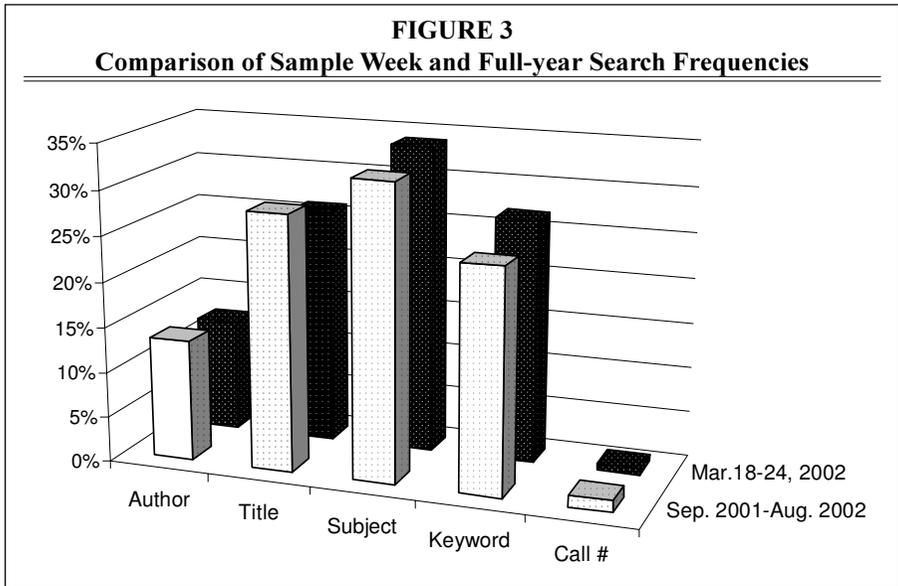
Findings

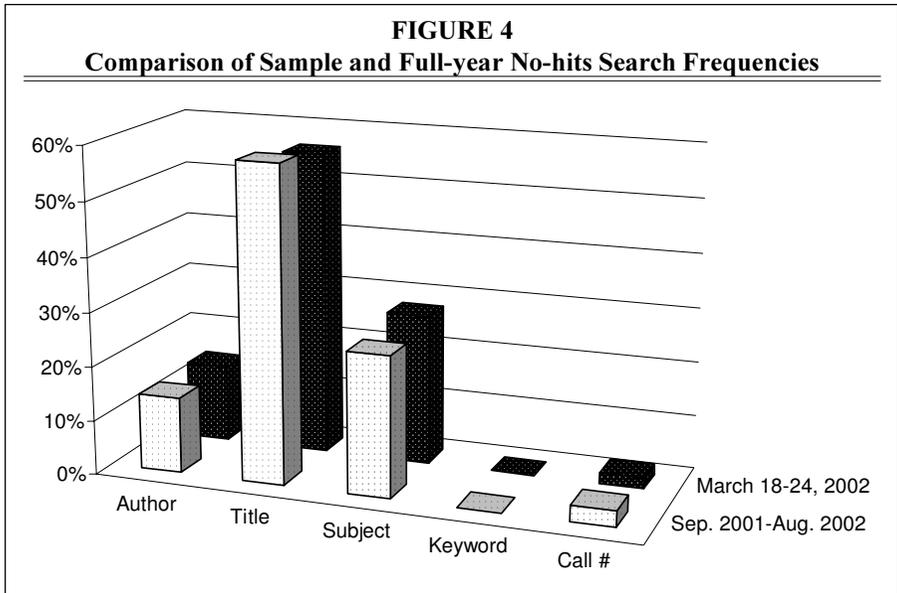
The findings of this exploratory study are reported in two sections. A detailed analysis of a one-week sample of subject no-hits searches is presented first, followed by a statistical summary of longitudinal study findings.

One-week Sample Analysis

The sample analysis is a retrospective look at how the subject no-hits searches for the week of March 18–24, 2002, were reviewed and enhanced during the study. This week was selected because it occurred in the most recent regular academic term (spring 2002) and was the

FIGURE 3
Comparison of Sample Week and Full-year Search Frequencies





week in which the highest number of subject no-hits searches (1,940) was recorded.

As rough measures of whether searches in this sample week were representative of the full year's searches, data on *total* searches and *no-hits* searches for the week of March 18–24, 2002, were compared to those for the complete 2001–2002 school year. Both comparisons reveal that frequencies of all searches (figure 3) as well as no-hits searches (figure 4) across the five types of search indexes were quite similar, thereby suggesting that the sample week was not atypical of searching done over the full year.

Of the 1,940 subject no-hits searches, 419 were discarded because they were duplicates, leaving 1,521 unique searches logged during the week of March 18–24, 2002. The author analyzed each unique search by seeking answers to four questions described in the methods section of this article and found that 515 (approximately 34%) were amenable to enhancement. In 90 percent of cases, the enhancement was achieved through the addition of new authority record cross-references and 10 percent through the addition of a new pathfinder record.

To determine what types of searches were represented in the list of 1,521 no-

hits searches, a categorization scheme was created for the retrospective analysis. Proceeding heuristically, nine categories emerged from the list, which are coded and defined in table 2.⁴⁷

After categorizing the 1,521 sample no-hits searches according to the scheme in table 2, percentages distributions were calculated for:

- the number of sample searches (1,521) in each of the nine no-hits categories;
- the number of enhanced searches (515) in each of the nine no-hits categories;
- the number of sample searches (1,521) that were enhanced in each of the nine no-hits categories.

These three sets of percentages are presented in table 3.

By far the greatest proportion (71%) of the sample subject no-hits searches occurred in category 1 (correctly spelled topical search, but not in LCSH format), as did the greatest proportion (86%) of enhanced no-hits searches. As well, 41 percent of category 1 no-hits searches were found to be amenable to enhancement, and this percentage is second only to that of category 7 (correctly spelled geo-

TABLE 2
No-hits Categories Used in One-week Sample Analysis

Category Code	No-hits Category Definition	Examples
1	Topic; correctly spelled but not in LCSH format	• <i>Health risks with older mothers</i>
2	Personal/corporate name; misspelled or in a nonstandard format	• <i>Terry Fox</i>
3	Topic; incorrectly spelled and not in LCSH format	• <i>Heirarchy in the church</i>
4	Topic; incorrectly spelled, but in LCSH format	• <i>Lierature and society</i>
5	Title; should have been searched in title or a course reserves index	• <i>Canadian Journal of Political Science</i> • <i>History 2290</i> • <i>APA citation guide</i>
6	Undetermined meaning	• <i>Eduus</i> • <i>I love you</i>
7	Geographic place name; correctly spelled, but not in LCSH format	• <i>The Caribbean</i>
8	Topic; correctly spelled and formatted LCSH	• <i>Dams—Australia</i>
9	Personal/corporate name; correctly spelled and formatted	• <i>Aga Khan</i>

graphic name, but not in LCSH format) by a very narrow margin.

Category 2 (personal or corporate name containing spelling errors or in a nonstandard format) represented the second largest percentages of no-hits searches and of enhanced searches, but the percentages values were small (9% and 7%, respectively). It is worth noting

that name searches (categories 2 and 9) may have represented searches for information *about* a person or organization or information *by* a person or organization; in the latter case, these searches might have been more appropriately conducted in the author index.

Table 4 provides examples of database enhancements made in response to subject

TABLE 3
Distribution of No-hits and Enhanced Searches in One-week Sample

Category Code	Abbreviated Meaning	% of 1,521 No-hits Searches	% of 515 Enhanced Searches	% of No-hits Searches That Were Enhanced
1	Non-LCSH topic	71%	86%	41%
2	Nonstandard name	9%	7%	25%
3	Misspelled non-LCSH topic	8%	1%	4%
4	Misspelled LCSH	3%	2%	24%
5	Wrong index	3%	2%	24%
6	Undetermined	2%	0%	0%
7	Non-LCSH geographic name	1.5%	2%	42%
8	Valid LCSH	1.5%	0%	0%
9	Valid name	1%	0%	0%

TABLE 4
Examples of Enhanced No-hits Searches

Category	Original No-hits Search String	New Cross-reference	New Pathfinder Record*
1	canadian scholarships	Scholarships—Canada —Directories	
1	aboriginal parents		1. Parents 2. Indians of North America
1	industrial globalization		1. Globalization 2. Industries
2	charlotte whitton	Whitton, Charlotte, 1896–1975	
3	shakespear sonnets	Shakespeare, William, 1564–1616. Sonnets	
4	capital punishment ethic aspects	Capital punishment— Moral and ethical aspects	
5	the crc handbook of chemistry and physics	Chemistry—Tables	
5	APA citation guide	Publication manual of the American Psychological Association	
7	the caribbean	Caribbean Area	

* Entries in the New “Pathfinder” Record column indicate 1. the LCSH to be searched, and 2. the LCSH or subject words to be used to limit the results of searching 1, where these two steps are described in the statement: “To find some information on [no-hits search terms]: search the SUBJECT index using “[1.]” then LIMIT your search results using “[2.]” in WORDS in the SUBJECT.”

no-hits searches in each no-hits category that received at least one enhancement. In most cases, cross-references were added to subject authority records, but on occasion exceptions were made. For example, the two no-hits examples listed for category 5 in table 4 were interpreted to represent searches for titles, but only the first one was enhanced by the addition of a new cross-reference in a subject authority record. The second example, “APA citation guide,” was interpreted to be a search for the American Psychological Association’s publication manual, a popular reference source used heavily by students. Because no single LCSH could be identified that would retrieve catalog records describing only, or

primarily, the APA’s publication manual, an authority cross-reference was created to lead users from the subject no-hits search “APA citation guide” to a title search for “Publication manual of the American Psychological Association.”

Longitudinal Study Findings

Table 5 presents data for subject no-hits search analyses that took place for each calendar year of the study. Subject searches represented, on average, 27 percent of all no-hits searches annually. Over the course of the study, approximately 23 percent of all subject no-hits searches received database enhancements, representing a smaller percentage of searches ame-

nable to enhancement than was found in the one-week sample (34%), but totaled almost 38,000 enhanced-subject no-hits searches.

Discussion

This section considers two questions: How do the findings of this study compare with those of previous studies of subject no-hits searches in online catalogs? What are the study’s limitations?

Comparison with Previous Studies

Although no other studies were located that attempted to improve subject searching by systematically enhancing the catalog database in direct response to no-hits searches, several online catalog studies were identified from the 1980s and early 1990s that analyzed the frequency of subject no-hits searches, a sampling of which was presented in table 1. One can see that the present study’s data collection period (44 months, or approximately 1,320 days) and number of no-hits searches analyzed (162,663) far exceed those pertaining to the studies in table 1. Moreover, the average percentage of subject no-hits searches obtained over the course of the present study (27%) is lower than that of the other studies, which ranged from 32 to 60.1 percent.

It is possible that the present study’s lower percentage of subject no-hits

searches is attributable, at least in part, to its considerably larger sample size and longer data collection period. Other factors that may account for some of the observed differences in the results of these studies and perhaps may reduce their comparability are the variety of online catalog systems studied, the use of variant definitions of subject searches or no-hits searches, and variations in data collection, sampling, or analysis methods.⁴⁸

At the same time, the database enhancements explored in the present study may have played a role in the slight reduction in the frequency of subject no-hits search results. The 1999 expansion of the University of Lethbridge’s catalog user community to include Medicine Hat College users led to an unsurprising 10 percent increase in the frequency of subject index searches, which remained 6 to 10 percent higher than pre-1999 levels to the end of the 2001–2002 school year. However, a corresponding proportional increase in subject no-hits search frequencies was not found: from 1997–1998 to 1999–2000, the percentage of no-hits searches occurring in the subject index remained constant at about 29 percent and then fell gradually to 26 percent by 2001–2002. Hence, a possible explanation for the slight decrease in subject no-hits search percentages when we otherwise would expect them to have

TABLE 5
Longitudinal Study Findings

	1999	2000	2001	2002*	Total
Subject no-hits searches	42,134	47,044	48,396	25,089	162,663
Subject no-hits searches as % of all no-hits searches	31%	25%	25%	25%	27%
No. of enhanced subject no-hits searches	10,487	10,413	10,493	6,594	37,987
Percent of subject no-hits searches enhanced	25%	22%	22%	26%	23%
New authority records created	3,454	2,772	2,735	1,559	10,520
New pathfinder records created	298	536	815	663	2,312
Ratio of new authority records to new pathfinder records	11:1	5:1	3:1	2:1	4:1

* Data collection for 2002 covered eight months only (January to August).

remained constant may be the enhancements made to the database in the course of conducting the present study, which spanned January 1999 to August 2002.

Study Limitations

Although the present study deliberately set out to explore new techniques for improving subject searching, there is much to question regarding its enhancement methods. A general concern pertains to their subjective nature. Because transaction log data provide no information about users' intentions, the process of analyzing the subject no-hits searches was necessarily a fallible, interpretive process, and the author may not have correctly discerned the intended meanings of these searches. Thus, it is possible that from the user's viewpoint, the post-enhancement catalog response represented an irrelevant or less-helpful response than did the original no-hits response.

With regard to enhancement 1, which involved the creation or updating of authority record cross-references, the non-standard coding of authority records is a limitation. In many cases, the new cross-references comprising verbatim no-hits search strings did not represent headings formatted according to LC subject cataloging policy and practice.⁴⁹ Some of the no-hits search strings contained spelling and grammatical errors, and therefore the authority record cross-references contained these same errors (e.g., "Malayasia," and "rythm and blues"). It may be argued that these types of cross-references are not enhancements because they deliberately introduce misspelled cross-references into the database. In addition, they do not bring errors to the attention of the user, thus perhaps missing an opportunity for learning to take place. If automatic spelling error detection becomes a standard feature of online catalogs, however, the potential usefulness of this type of cross-reference may diminish.

As well, new authority records were keyed in USMARC or MARC 21 format, but no attempt was made to code each one fully. The goal of creating these brief

records was to generate the desired new cross-references quickly in order to complete the analysis of the current week's subject no-hits searches before the subsequent week's list was generated. Thus, the utility of these new authority records may be limited to the local catalog, as their relevance for other catalog databases is unknown.

There is equal, or perhaps even more, cause for concern when we consider the limitations of enhancement 2, which involved the creation of pathfinder records. First of all, using bibliographic records solely to present information about search refinement is unorthodox and could be viewed as unwise tampering with a record format that was not designed for this purpose.⁵⁰

Second, an undesirable outcome of introducing pathfinder records into the database is the interfiling of the verbatim text of subject no-hits searches in the controlled vocabulary subject index. This is a requirement because, otherwise, pathfinder records have no access point that can be used to retrieve them in the catalog. These non-LCSH headings, therefore, appear in the online catalog subject index where users may encounter them while browsing the subject index and mistake them for valid LCSHs. In subject index browse displays in the University of Lethbridge catalog, pathfinder record access points (subject no-hits search terms) are visually indistinguishable from other valid headings.

Another undesirable feature of pathfinder records is the appearance of verbatim no-hits search strings masquerading as valid subject headings in pathfinder records themselves. In a regular bibliographic record display, the user may click on any of the subject headings included in the record, labelled "Subject," because this action automatically executes a subject index search for the selected heading without requiring the user to type in a new search. In a pathfinder record, however, the subject no-hits search term is labelled and is selectable as if it is a valid LCSH in a bibliographic record. (See the example in figure 2.) If

the user clicks on the term displayed in the "Subject" field of a pathfinder record, a potentially confusing loop is formed because the catalog responds with a browse display containing the original no-hits search string, and selecting this "heading" from the browse display put the user back into the pathfinder record.

An additional potential limitation of enhancement 2 is the users' ability to understand and use pathfinder records effectively. Users may choose not to read or heed the suggestions for search refinement contained in pathfinder records. If the suggestions are not used, the pathfinder records perhaps pose more of a hazard than a help to catalog users. Even the willing searcher who wishes to act on the suggested steps for reexecuting a search differently may not understand how to carry out the instructions. Although it can be a powerful tool for search refinement, transaction logs show that the catalog LIMIT option is used infrequently. Use of the LIMIT option was present in each pathfinder record's instructions for search refinement.

Table 5 indicates that the ratio of new authority records to new pathfinder records decreased steadily over the course of the study. This may have been due to changes over time in the author's perceptions of how best to handle enhancements. Other possibilities are that users' search topics became more interdisciplinary, thus rendering their topics less likely to be representable by a single LCSH, or perhaps users began to construct online catalog subject searches in a manner similar to that habitually used with Internet search engines. In any case, the creation of pathfinder records likely contravenes more aspects of standard cataloging and authority control practice than does adding non-standard authority record cross-references. If widespread use of pathfinder records is not advisable, alternative means for enhancing catalog responses to interdisciplinary or multitopic no-hits searches are needed.

Last but not least, the enhancements explored in this study are not easily au-

tomated or scalable to larger databases or user communities. At the busiest times of the academic year, completing the analysis of the current week's list of subject no-hits searches required the equivalent of one to two full workdays, while other regular work responsibilities continued to receive first priority. However, it would be difficult to automate the time-consuming intellectual analysis portion of the study because of the complex array of sources that informed the author's interpretations and subsequent selective enhancements of subject no-hits searches. These sources included knowledge of the university's curriculum, areas of research specialization on campus, current student research assignments, popular research topics, areas of library collection strengths and weaknesses, idiosyncrasies of the Innopac automated system, LC subject cataloging policy, and MARC 21 bibliographic and authority record coding conventions.

Further Research

Two distinct areas of further research are considered in this section. The first involves next steps required to determine whether, in fact, the potential subject no-hits search ameliorations explored in this study make a difference to users' subject-searching experiences. The second area of further research comprises additional research questions related to this study's explorations.

Next Steps

The present study developed and explored two potential ameliorations of subject no-hits search results and found some indication that enhancements 1 and 2 may have reduced the frequency of no-hit subject searches. However, to determine whether, in fact, enhancements 1 and 2 enhance subject searching in the University of Lethbridge catalog, we also need to discover whether they make a positive difference in users' subject-searching experiences.⁵¹ How might an evaluation of the effectiveness of enhancements 1 and 2 be undertaken?

Unobtrusive monitoring of the use of only new or enhanced authority records and pathfinder records resulting from this study is currently not possible in the Innopac system. If obtrusive data collection methods are used, perhaps quantitative and qualitative data on a cross-sectional sample of individual users' searching behaviors, intentions, and judgments could be collected in an experimental setting involving study participants searching for topics shown to be frequent subject no-hits search topics in previous transaction logs.

It may be possible to gather transaction log data if the computers or participants involved in such an experiment could be uniquely identified, thus allowing study participants' searches to be tracked under a distinct identification code in the transaction logs. Alternatively, perhaps screen-capture software could be used to record the content of Web catalog screens encountered by each participant in the experiment. However, the information we might wish to elicit from participants of such a study begs the question of what aspects of online catalog users' subject-searching behavior we have the greatest need to understand better.

In this regard, it is helpful to consider Christine L. Borgman's discussion of two sources of problems in searching online systems: differences among individual users, and differences in individual system features. She suggested that "on any given system, people will search in different ways, with different levels of success and satisfaction. Until we can identify the factors responsible for low success rates, it will be difficult to narrow the performance range, moving users toward the upper end of the success scale and thus removing barriers to access."⁵² Borgman further indicated that research has revealed several possible individual user factors that may influence success rates, including frequency of use of a given database, amount of training received, and academic major. Controlling for some or all of these user factors may therefore be advisable in a formal evalu-

ation of the enhancements explored in this study.

Another evaluation approach, based on a suggestion by Ray R. Larson for evaluating subject searching in online catalogs, is to measure precision (the extent to which results of a search contain relevant items) and recall (the extent to which all relevant items are retrieved) of enhanced subject no-hits searches.⁵³ Because the database enhancements resulting from the present study can be suppressed and unsuppressed from the public catalog, it may be possible to compare search results on selected no-hits topics obtained in two different states of the catalog, one containing the database enhancements and the comparison state lacking the database enhancements, to see whether differences in precision and recall are found.

Others have suggested that precision and recall measures of relevance in automated information retrieval are indicative of an objective, system-centered view concerned only with the presence or absence of a match between the topic of a subject search and the topic of a document to the exclusion of the user.⁵⁴ Proponents of an alternative user-centered approach view the judgment of relevance (e.g., the success of subject searches) as a multidimensional cognitive process that may be informed not only by the stated or inferred topics of documents, but also by other changeable, subjective, and time- and context-dependent factors.⁵⁵

An approach to the evaluation of the present study's database enhancements from the subjective relevance point of view could involve real-world information-seeking needs and contexts of catalog users, and elicitation of their cognitive processes as they interact over time with the catalog, its database enhancements, and the actual documents retrieved and used. Such a longitudinal study could involve periodically interviewing a small group of users as they progress through a complex library research task to elicit information about their experiences and cognitive processes. Concurrently, users' online catalog

subject searches could be monitored via transaction logs, and no-hits searches could be enhanced, when possible, using the methods described in the present article. Thus, interviews could include discussion of particular subject no-hits searches as well as users' views on whether subsequent catalog database enhancements are useful in the context of their particular information-seeking task.

Related Research Questions

In addition to further evaluative research, other issues have emerged from this exploratory study that may be worthy of investigation. For example, it has been suggested that in online catalogs, information overload (retrieving too much) and search failure (retrieving nothing) can be equally problematic.⁵⁶ Although captured in past transaction log reports, data on differences in retrieval set sizes across the University of Lethbridge catalog's various search indexes have not been analyzed longitudinally. However, an examination of the retrieval set sizes for the one-week sample from March 18–24, 2002, showed that although at least 80 percent of searches in all indexes retrieved fewer than a hundred records, the highest frequency of searches retrieving a hundred or more records occurred in the subject index (20%). Determining the extent of information overload could be achieved through a retrospective analysis of retrieval set sizes by index over the years. It also may be useful to determine whether information overload and search failure have been correlated in this catalog by comparing retrieval set size and no-hits search frequency patterns across the different search indexes.

Two other questions about subject searching in online catalogs are worthy of further investigation:

- How does an individual user's subject-searching behavior and cognitive processes change over time as catalog searching experience is acquired?
- What types of subject-searching enhancements might be suggested by answers to the preceding question?

Larson noted that "a longitudinal study of a set of users from their first introduction to the online catalog through one or two years of use, using transaction monitoring to record search behavior, and questionnaires and interviews to trace changing attitudes and needs" is one way to address questions that are not directly answerable using transaction log data alone.⁵⁷

Conclusions

Because the University of Lethbridge and its partner, Medicine Hat College, are primarily undergraduate institutions, on a continuing basis, many users of their shared library catalog are novices. We know that novice catalog users most often search the subject index, that library users in general have little understanding of the LCSHs that comprise the controlled vocabulary used in the catalog's subject index, and that users experience the most difficulty searching the subject index, which is frequently manifested in high proportions of no-hits searches.^{58–60} Thus, finding ways to improve the subject-searching experiences of users of this shared catalog by, for example, reducing the frequency of subject no-hits search results, is of particular concern.

Database enhancements 1 and 2 explored in the present study may have contributed to the slight reduction in the proportion of subject no-hits search results observed from 1999 to 2002, which occurred at a time when the proportion of subject searches logged as no-hits searches would otherwise have been expected to remain constant after expansion of the catalog's primary user community in 1999. However, the more significant question of whether these database enhancements actually make a difference to users' subject-searching experiences requires further research.

If the present study's methods of enhancing subject searching are shown to be beneficial to users' subject-searching experiences, a reexamination of aspects of standard cataloging practice may be worthwhile. Current policy governing the

updating of LCSH cross-references states that additions, changes, and deletions to 4XX cross-references require citation of "sources that support the form of the proposal or references."⁶¹ This requirement, based on literary warrant, presents problems when applied to the evaluation of LCSH cross-references because it is author centered. That is, terms used by authors (which are selected into the LCSH controlled vocabulary by subject catalogers) are not necessarily the same as those known to users.

Use warrant, explored in the present study, is an alternative principle for guiding the creation of new LCSH cross-references. In 1990, Elaine Svenonius noted that relying on use warrant, or common usage, for the selection of controlled vocabulary terms posed difficulties in the past, but she also suggested that "the relatively new possibility of analyzing transaction logs of users' searches in online databases offers some promise of progress in the understanding of terms commonly used in information seeking."⁶² The

present study demonstrated that it is possible to respond in a timely manner to users' natural language (no-hits) search terms by selectively expanding the entry vocabulary using two database enhancement techniques. The results of this exploratory study can be viewed as support for the suggestion that use warrant bears reconsideration as a criterion for evaluating proposed new LCSH cross-references in addition to literary warrant.

Because many of the well-documented problems of subject searching in online catalogs have remained the focus of various research efforts for more than two decades, progress in improving subject searching may continue to evolve gradually on different research fronts and from a variety of research approaches. Thus, small-scale studies such as the one described here may be worthwhile for those preferring to explore alleviations to perceived subject-searching problems of local online catalog users that might be achievable within a more immediate time frame.

Notes

1. Joseph R. Matthews, Gary S. Lawrence, and Douglas K. Ferguson, eds., *Using Online Catalogs: A Nationwide Survey: A Report of a Study Sponsored by the Council on Library Resources* (New York: Neal-Schuman, 1983).
2. Karen Markey, "Users and the Online Catalog: Subject Access Problems," in *The Impact of Online Catalogs*, ed. Joseph R. Matthews (New York: Neal-Schuman, 1986).
3. Carol A. Mandel, "Enriching the Library Catalog Record for Subject Access," *Library Resources & Technical Services* 29 (Jan./Mar. 1985): 5-15.
4. Markey, "Subject-Searching Experiences and Needs of Online Catalog Users: Implications for Library Classification," *Library Resources & Technical Services* 29 (Jan./Mar. 1985): 34-51.
5. Alex Byrne and Mary Micco, "Improving Opac Subject Access: The ADFA Experiment," *College & Research Libraries* 49 (Sept. 1988): 432-41.
6. Shirley Anne Cousins, "Enhancing Subject Access to Opacs: Controlled Vocabulary vs Natural Language," *Journal of Documentation* 48 (Sept. 1992): 291-309.
7. Mia Massicotte, "Improved Browsable Displays for Online Subject Access," *Information Technology and Libraries* 7 (Dec. 1988): 373-80.
8. Jeffrey C. Huestis, "Clustering LC Classification Numbers in an Online Catalog for Improved Browsable," *Information Technology and Libraries* 7 (Dec. 1988): 381-93.
9. Dorothy McGarry and Elaine Svenonius, "More on Improved Browsable Displays for Online Subject Access," *Information Technology and Libraries* 10 (Sept. 1991): 185-91.
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 (Mar. 1991): 78-98.
11. Mark T. Kinnucan, "Fisheye Views as an Aid to Subject Access in Online Catalogues," *Canadian Journal of Information Science* 17 (Jul. 1992): 25-40.
12. Tamas E. Doszkocs, "Cite NLM: Natural Language Searching in an Online Catalog," *Information Technology and Libraries* 2 (Dec. 1983): 364-80.
13. Stephen Walker, "Improving Subject Access Painlessly: Recent Work on the Okapi Online Catalogue Projects," *Program* 22 (Jan. 1988): 21-31.
14. Ray R. Larson, "Classification Clustering, Probabilistic Information Retrieval, and the

Online Catalog," *Library Quarterly* 61 (Apr. 1991): 133-73.

15. Mary Micco, "The Next Generation of Online Public Access Catalogs: A New Look at Subject Access Using Hypermedia," *Cataloging & Classification Quarterly* 13, no. 3/4 (1991): 103-20.

16. Micheline Hancock-Beaulieu and Stephen Walker, "An Evaluation of Automatic Query Expansion in an Online Library Catalogue," *Journal of Documentation* 48 (Dec. 1992): 406-21.

17. Marcia J. Bates, "Subject Access in Online Catalogs: A Design Model," *Journal of the American Society for Information Science* 37 (Nov. 1986): 357-76.

18. Micheline Hancock, "Subject Searching Behaviour at the Library Catalogue and at the Shelves: Implications for Online Interactive Catalogues," *Journal of Documentation* 43 (Dec. 1987): 303-21.

19. Micheline Hancock-Beaulieu, "Evaluating the Impact of an Online Library Catalogue on Subject Searching Behaviour at the Catalogue and at the Shelves," *Journal of Documentation* 46 (Dec. 1990): 318-38.

20. The notion of "generations" of online catalogs was introduced by Charles R. Hildreth, "Pursuing the Ideal: Generations of Online Catalogs," in *Online Catalogs, Online Reference: Converging Trends*, ed. Brian Aveney and Brett Butler (Chicago: ALA, 1984).

21. Sandy Roe, "Online Subject Access," *Journal of Internet Cataloging* 2, no. 1 (1999): 69-78.

22. Chris Evin Long, "Improving Subject Searching in Web-based Opacs: Evaluation of the Problem and Guidelines for Design," *Journal of Internet Cataloging* 2, no. 3/4 (2000): 159-86.

23. In this paper, references to the University of Lethbridge catalog covering 1999 or later should be understood to be referring to the shared catalog of the University of Lethbridge and Medicine Hat College.

24. From 1990 to 1995, no-hits search results occurred most frequently in the subject index and second most frequently in the title index, but from 1996 this pattern has been reversed. The pattern reversal is likely due to increased searching in the various online periodical indexes, which became available to University of Lethbridge users in greater numbers from about 1996. To locate desired articles identified in online periodical index search results, users must search the online catalog title index, but in many cases these searches retrieve no records because the university owns a very small percentage of the journals indexed in the online periodical index databases available to its users.

25. Patricia Flaherty, "Transaction Logging Systems: A Descriptive Summary," *Library Hi Tech* 11, no. 2 (1993): 67-78.

26. Exploration of the third supposition was possible without the need to gather any additional data because transaction log data were monitored as a part of normal work responsibilities. It was not possible to pursue the additional resources and time to explore the other two suppositions at the time the study was conceived.

27. Thomas A. Peters, Martin Kurth, Patricia Flaherty, Beth Sandore, and Neal K. Kaske, "Transaction Log Analysis," *Library Hi Tech* 11, no. 2 (1993): 37.

28. Thomas A. Peters, "The History and Development of Transaction Log Analysis," *Library Hi Tech* 11, no. 2 (1993): 41-66.

29. Marilyn Ann Lester, "Coincidence of User Vocabulary and Library of Congress Subject Headings: Experiments to Improve Subject Access in Academic Library Online Catalogs," (Ph.D. diss., Univ. of Illinois at Urbana-Champaign, 1989).

30. John E. Tolle, *Current Utilization of Online Catalogs: Transaction Log Analysis, Final Report to the Council on Library Resources: Volume 1 of Three Volumes* (Dublin, Ohio: OCLC Online Computer Library Center, 1983).

31. Cheryl Kern-Simirenko, "OPAC User Logs: Implications for Bibliographic Instruction," *Library Hi Tech* 1 (winter 1983): 27-35.

32. Adeline Wilkes and Antoinette Nelson, "Subject Searching in Two Online Catalogs: Authority Control vs. Non-Authority Control," *Cataloging & Classification Quarterly* 20, no. 4 (1995): 57-79.

33. Beverley Barrett and Margaret Maticka, "An Analysis of User Failure in Subject Searching an Online Catalogue," in *Garbage In, Garbage Out: The Need for Quality in the Age of Automation*, ed. Alan Bundy and Judith Bundy (Adelaide: Auslib Pr., 1990), 38-49.

34. Deborah D. Blecic, Josephine L. Dorsch, Melissa H. Koenig, and Nirmala S. Bangalore, "A Longitudinal Study of the Effects of OPAC Screen Changes on Searching Behavior and Searcher Success," *College & Research Libraries* 60 (Nov. 1999): 515-30.

35. Arlene G. Taylor, "Authority Files in Online Catalogs: An Investigation of Their Value," *Cataloging & Classification Quarterly* 4, no. 3 (1984): 1-17.

36. Elisabeth Sinnott, "Fewer Errors Resulting from the Users' Misconception of the Opac in 1992 Than a Decade Ago: A Comparative Study of No Direct Hits and Zero Hits in Author Searches," *Cataloging & Classification Quarterly* 18, no. 1 (1993): 75-101.

37. Nirmala S. Bangalore, "Authority Files in Online Catalogs Revisited," *Cataloging & Classification Quarterly* 20, no. 3 (1995): 75-94.

38. ———, "Mystery Names: Plausible Sounding Names Resulting in Zero Hits in Opacs,"

Technical Services Quarterly 14, no. 1 (1996): 15–31.

39. Jean Dickson, "An Analysis of User Errors in Searching an Online Catalog," *Cataloging & Classification Quarterly* 4, no. 3 (1984): 19–38.

40. Pauline A. Cochrane, *LCSH Entry Vocabulary Project: Final Report to the Council on Library Resources and to the Library of Congress* (Washington, D.C.: Council on Library Resources, 1983).

41. Jane Greenberg, "Reference Structures: Stagnation, Progress, and Future Challenges," *Information Technology and Libraries* 16 (Sep. 1997): 108–19.

42. The Innopac LIMIT feature becomes available only after a user has first conducted a search retrieving more than one record. Clicking on the LIMIT THIS SEARCH button presents a menu of criteria by which the original search may be limited or sorted. The third item in this menu allows limiting by "Words in the Author, Title, Subject, or Publisher." The default is "Words in the Author," but users may switch to one of the other choices through a pull-down menu.

43. Coding details for specific types of fields in MARC authority records are found in *Marc 21 Format for Authority Data: Including Guidelines for Content Designation* (Washington, D.C.: Library of Congress Cataloging Distribution Service, 1999).

44. Karen M. Drabenstott, Schelle Simcox, and Eileen G. Fenton, "End-user Understanding of Subject Headings in Library Catalogs," *Library Resources & Technical Services* 43 (Jul. 1999): 140–60.

45. Stephen E. Wiberley, Robert Allen Daugherty, and James A. Danowski, "User Persistence in Displaying Online Catalog Postings: LUIS," *Library Resources & Technical Services* 39 (July 1995): 247–64; Stephen E. Wiberley and Robert Allen Daugherty, "Users' Persistence in Scanning Lists of References," *College & Research Libraries* 49 (Mar. 1988): 149–56; Marcia J. Bates, "The Fallacy of the Perfect Thirty-Item Online Search," *RQ* 24 (fall 1984): 43–50.

46. In some cases, avoidance of large retrieval sets was not possible. For instance, the subject no-hits search for "America politics and government" was interpreted to be equivalent to the LCSH, "United States—Politics and government," and was therefore added as a cross-reference to the LCSH authority record. The number of records retrieved by the LCSH, "United States—Politics and government," however, exceeded a hundred.

47. Other researchers such as Barrett and Maticka, "An Analysis of User Failure" and Peters, "When Smart People Fail" created no-hits categories as part of their analyses of "failed" controlled vocabulary catalogue searches. However, the researcher determined that none of these schemes adequately covered the subject no-hits searches analyzed in the present study. Therefore, a customized categorization scheme was developed.

48. Allyson Carlyle discussed similar problems of LCSH "matching" studies in "Matching LCSH and User Vocabulary in the Library Catalog," *Cataloging & Classification Quarterly* 10, no. 1/2 (1989): 37–63.

49. See sections H195, H373, and H200 9b in Library of Congress, Cataloging Policy and Support Office, *Subject Cataloging Manual: Subject Headings*, 5th ed. 4 vols (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1996).

50. Pathfinder records were keyed as non-MARC bibliographic records. MARC and non-MARC bibliographic records can coexist in this catalog database (an Innopac system) with few differences in how they are indexed or displayed.

51. Although a formal evaluation of the effectiveness of the database enhancements was not within the scope of the present article, the library reference staff at the University of Lethbridge as a whole endorsed the idea of experimenting with enhancements 1 and 2 as possible ways to ameliorate catalog users' subject-searching problems. Over the course of the study, the author received unsolicited positive feedback on the database enhancements from several reference staff members and one problem with pathfinder records was reported, which was described in the study limitations section of this article.

52. Christine L. Borgman, "Why Are Online Catalogs Hard to Use? Lessons Learned from Information-Retrieval Studies," *Journal of the American Society for Information Science* 37 (Nov. 1986): 387–400.

53. Larson, "Between Scylla and Charybdis: Subject Searching in the Online Catalog," *Advances in Librarianship* 15 (1991): 175–236.

54. For example, see Carol L. Barry, "User-defined Relevance Criteria: An Exploratory Study," *Journal of the American Society for Information Science* 45 (Apr. 1994): 149–59; and Stephen P. Harter, "Psychological Relevance and Information Science," *Journal of the American Society for Information Science* 43 (Oct. 1992): 602–15.

55. A useful overview of research on the meanings and uses of relevance in the context of information seeking and use is found in Linda Schamber, "Relevance and Information Behavior," *Annual Review of Information Science and Technology* 29 (1994): 3–48.

56. Larson, "The Decline of Subject Searching: Long-term Trends and Patterns of Index Use in an Online Catalog," *Journal of the American Society for Information Science* 42 (Apr. 1991): 197–215.

57. *Ibid.*

58. Joseph R. Matthews and Gary S. Lawrence, "Further Analysis of the CLR Online Catalog Project," *Information Technology and Libraries* 3 (Dec. 1984): 354-76.
59. Drabenstott, Simcox, and Fenton, "End-user Understanding of Subject Headings in Library Catalogs."
60. Markey, "Users and the Online Catalog."
61. Library of Congress, "Subject Cataloging Manual," H200 9b.
62. Elaine Svenonius, "Design of Controlled Vocabularies," in *Encyclopedia of Library and Information Science*, v. 45, suppl. 10, ed. Allen Kent (New York: Marcel Dekker, 1990), 82-109.