

# The Impact of Web Search Engines on Subject Searching in OPAC

Holly Yu and  
Margo Young

*This paper analyzes the results of transaction logs at California State University, Los Angeles (CSULA) and studies the effects of implementing a Web-based OPAC along with interface changes. The authors find that user success in subject searching remains problematic. A major increase in the frequency of searches that would have been more successful in resources other than the library catalog is noted over the time period 2000–2002. The authors attribute this increase to the prevalence of Web search engines and suggest that metasearching, relevance-ranked results, and relevance feedback (“more like this”) are now expected in user searching and should be integrated into online catalogs as search options.*

In spite of many studies and articles on Online Public Access Catalogs (OPAC) over the last twenty-five years, many of the original ideas about improving user success in searching library catalog have yet to be implemented. Ironically, many of these techniques are now found in Web search engines. The popularity of the Web appears to have influenced users’ mental models and thus their expectations and behavior when using a Web-based OPAC interface. This study examines current search behavior using transaction-log analysis (TLA) of subject searches when zero-hits are retrieved. It considers some of the features of Web search engines and online bookstores and suggests future enhancements for OPACs.

## Literature Review

Many studies have been published since the 1980s centering on the OPAC. Seymour and Large and Beheshti provide in-depth overviews on OPAC research from the mid-1980s through the mid-1990s.<sup>1</sup> Much of this research has addressed system design and user behavior including:

- user demographics,
- search behavior,
- knowledge of system,
- knowledge of subject matter,

- library settings,
- search strategies, and
- OPAC systems<sup>2</sup>

OPAC research has employed a number of data-collection methodologies: experiment, interviews, questionnaires, observation, think aloud, and transaction logs.<sup>3</sup> Transaction logs have been used extensively to study the use of OPACs, and library literature reflects this. While the exact details of TLA vary greatly, Peters et al. define it simply as “the study of electronically recorded interactions between online information retrieval systems and the persons who search for the information found in those systems.”<sup>4</sup> This section reviews the TLA literature relevant to the study.

## Number of Hits

TLA cannot portray user intention or actual satisfaction since relevance, success, or failure are subjectively determined and require the user to decide. Peters recommends combining TLA with another technique such as observation, questionnaire or survey, interview, or focus group.<sup>5</sup> In spite of the limitations of TLA, many studies (including this one) rely on it alone. Typically, these studies define failure as zero hits in response to a search. Generalizing from several studies, approximately 30 percent of all searches result in zero hits.<sup>6</sup> The failure rate is even higher for subject searches: Peters reported that about 40 percent of subject searches failed by retrieving zero hits.<sup>7</sup>

Some researchers also define an upper number of results for a successful search. Buckland found that the average retrieval set was 98.<sup>8</sup> Bleic reported that Cochrane and Markey found that OPAC users retrieve too much (15 percent of the time).<sup>9</sup> Wiberly, Daugherty, and Danowski (as reported in Peters) found that the median number of postings considered to be too many was fifteen, although when fifteen to thirty postings were retrieved, more users displayed them all than abandoned the search.<sup>10</sup>

## Subject Searching

Some studies have specifically looked at subject searching. Hildreth differentiated among various types of searches and defined one hundred items as the upper limit for keyword searches and ninety as the upper limit for subject searches.<sup>11</sup> Larson defined reasonable subject retrieval as between one and twenty items and found that only 12 percent of subject searches retrieved the appropriate number.<sup>12</sup>

---

**Holly Yu** (hyu3@calstatela.edu) is Library Web Administrator and Reference Librarian at the University Library, California State University, Los Angeles. **Margo Young** (Margo.e.young@jpl.nasa.gov) is Manager of the Library, Archives and Records Section at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena.

Larson is not the only researcher to have reported poor results in subject searching. For more than twenty years, research has demonstrated that subject or topical searches are both popular and problematic. Tolle and Han found that subject searching is most frequently used and the least successful.<sup>13</sup> Moore reported that 30 percent of searches were for subject, and Matthews et al. found that 59 percent of all searches were for subject information.<sup>14</sup> Hunter found that 52 percent of all searches were subject searches and that 63 percent of these had zero hits.<sup>15</sup> Van Pulis and Ludy referred to Alzofon and Van Pulis's earlier work in 1984 where they reported that 42 percent of all searches were subject searches.<sup>16</sup> Hildreth found that 62.1 percent of subject searches and 35.4 percent of keyword searches failed.<sup>17</sup> Larson categorized the major problems with online catalogs as follows:

- users' lack of knowledge of Library of Congress subject headings (LCSH),
- users' problems with mechanical and conceptual aspects of query formulation,
- searches that retrieve nothing,
- searches that retrieve too much, and
- searches that retrieve records that do not match what the user had in mind.<sup>18</sup>

During an eleven-year longitudinal study, Larson found that subject searching was being replaced by keyword searching.<sup>19</sup>

No consistent pattern in the number of search terms has emerged in the literature. Van Pulis and Ludy reported that user searches were typically single words.<sup>20</sup> Markey contended that users' search terms frequently matched standardized vocabulary in large catalogs.<sup>21</sup> None of Markey's researchers consulted LCSH, and only 11 percent of Van Pulis and Ludy's did so, notably in spite of their library's user-education programs. Peters reported that Lester found that the average search was less than two words and fewer than thirteen characters.<sup>22</sup> Hildreth found that more than two-thirds of keyword searches included two or more words and 42 percent of these multiple-word searches resulted in zero hits.<sup>23</sup> The proportion of zero-hit keyword searches rose with the increasing number of words in the search.

Subject headings have been a matter of considerable study. Gerhan examined catalog records and surmised their accessibility in an online catalog. He contended that when a keyword from the title only is accessed, only 50 percent of all relevant books would be found and that title keywords would lead a user to subject-relevant records in 55 percent of cases while LCSH would lead a user successfully in 85 percent of the cases.<sup>24</sup> In contrast, Cherry found that 42 percent of zero-hit subject searches would have been more fruitful as keyword or title searches than by following cross references retrieved from the subject field.<sup>25</sup> She recommended converting zero-hit subject queries to

other types of subject searches (keyword). Thorne and Whitlatch recommended that subject searchers should select keyword rather than subject headings as their first access strategy.<sup>26</sup>

## Types of Problems in Subject Searches

Numerous studies have categorized reasons for search failure (typically in zero-hit situations), but Peters reports that a standard categorization has not yet been established.<sup>27</sup> In cases where more than one error is made in a search (and Hunter reported this to be frequent), there is no consistency in how that is assigned. Nonetheless, some major categories of problems stand out:

- **misspelling and typographical errors**—Peters found that these errors accounted for 20.8 percent of all unsuccessful keyword searches, while Henty (reported by Peters) concluded that 33 percent of such searches could be attributed to this.<sup>28</sup> Hunter found that 9.3 percent of subject searches had typographical and spelling errors.<sup>29</sup>
- **keyword search**—Hunter found 52.6 percent of zero-hit searches used uncontrolled vocabulary terms.<sup>30</sup>
- **wrong source or field**—Hunter concluded that 4.5 percent of searches should have been done in a source other than the catalog, while 1.3 percent of searches were of the wrong type (an author search in the subject-search option).<sup>31</sup>
- **items not in the database**—Peters found that searches for items not held in the database accounted for 39.1 percent of unsuccessful searches, while Hunter found that problem in only 2.5 percent of the problem cases.<sup>32</sup>

In addition to these problems, Hunter also found that index display and rules relating to the systems accounted for 27 percent of errors.<sup>33</sup>

## Resulting Recommendations for Change

While Hildreth stated, "There has been little research on most components of the OPAC interface" in 1997, he proposed two options to improve user success: increased user training or improved design based on information-seeking behavior.<sup>34</sup> Wallace pointed out that there is a very short window of opportunity when searchers are amenable to instruction and that successful screen designs should therefore focus on presenting the quick-searching options employed by the majority of users first.<sup>35</sup> Large and Beheshti observed "that too many options simply caused confusion, at least for less experienced OPAC users," and they summarized that OPAC-

interface research focuses on menu sequence, browsing, and querying.<sup>36</sup>

## Menu Sequence

In terms of menu sequence, Hancock-Beaulieu indicated that "the menu sequence in which search options are offered will influence user selection."<sup>37</sup> Ballard found that the amount of keyword searching was affected by its position on the menu.<sup>38</sup> Scott reported that both keyword- and subject-search success improved when the keyword was placed at the top of the menus.<sup>39</sup> Thorne and Whitlach used a combination of methods in their study and concluded that several interface changes should be implemented:

- strongly encourage novice users to start with keyword (list keyword above subject heading),
- relabel "keyword" to "subject or title words," and
- relabel "subject heading" to "Library of Congress Subject Heading."<sup>40</sup>

Blecic et al. studied transaction logs over six months to track the impact of "simplifying and clarifying" OPAC introductory screens. After moving the keyword option to the top, keyword searching increased from 13.30 percent to 15.83 percent of all search statements. Blecic et al. found her original tally of 35.05 percent of correct searches having zero hits decreased to 31.35 percent after screen changes.<sup>41</sup>

## Querying

OPAC-interface design has been based on an assumption that users come to the catalog knowing what they need to know. In either text-based OPAC or Web-based OPAC, query-based searches are still mainstream. Searchers are required to have knowledge of title, author, or subject. Ortiz-Repiso and Moscoso observed that Web-based catalogs, like all library catalogs, basically fulfill two functions: locating works based on known details and identifying which documents in the database cover a given subject.<sup>42</sup> Natural-language input has long been considered a desirable way to overcome this shortcoming.

## Browsing

Relevance-ranked output and hypertext were considered by Hildreth to be promising in 1997.<sup>43</sup> OPACs have not been conceived within a true hypertext environment, but rather they maintain the structure of their original formats, principally machine-readable cataloging (MARC), and therefore impede the generation of a structure of nodes and links.<sup>44</sup> In addition to continuing to employ MARC format as its underlying structure, the concept of main entry and added entry, field label, and display logic all reflect cataloging rules. Amazon.com and Barnes and

Noble have completely moved away from this century-old structure to provide easy access to book information. In the Web environment, the concept of main entry loses its meaning to multiple-access points and linking capabilities of author, subject, and call number.

Another prominent drawback of Web-based OPACs is that they have not taken advantage of thesaurus structure and utilized the thesaurus for searching feedback. The hierarchical relationship in LCSH is underutilized in terms of the relationship between terms and associations through related terms. Web-based OPACs have failed to make use of this important access.

The persistence of these drawbacks in OPAC-interface design is rooted deeply in cataloging rules that were derived from the manual environment more than a century ago. It reflects the gap between "concepts typically held by nonprofessional users and those used in library practices."<sup>45</sup>

In her article "Why Are Online Catalogs Still Hard to Use?" Borgman concludes:

Despite numerous improvements to the user interface of online catalogs in recent years, searchers still find them hard to use. Most of the improvements are in surface features rather than in the core functionality. We see little evidence that our research on searching behavior studies has influenced online catalog design.<sup>46</sup>

## Catalog Content

Users misunderstand the scope of the catalog. In questionnaire responses, 80 percent of Van Pulis and Ludy's participants indicated they had considered looking elsewhere than the library catalog, as in periodical indexes.<sup>47</sup> Blazek and Bilal reported a request for inclusion of journal-article titles in one response to their questionnaire.<sup>48</sup> Libraries responded to these requests by acquiring databases on CD-ROM, loading them locally (sometimes using the catalog system to mount a separate database), and, most recently, providing access to databases over the Internet. However, seldom have libraries responded to these requests by integrating search access through a single front end as the default search.

## Impact of Web Search Engines

Blecic et al. found that keyword searching increased from 13.3 percent to 28.3 percent over her four-year series of logs. At the same time, zero hits in keyword increased from 8.71 percent to 20.78 percent while subject zero hits dropped from 23 percent to 13.69 percent. She surmised that the influence of Web interfaces might have affected the regression-fluctuation in search syntax, initial articles, and author order.<sup>49</sup>

The Web is so prevalent today that many who use it to search OPAC while in the library, especially those in academic libraries, are likely to be users of the Web outside of the library and will bring their mental models of Web searching to OPACs. New users of Web-based OPACs are likely to have used the Web and thus they, too, bring that familiarity with them.

Jansen and Pooch conducted the most thorough comparison between Web search engines and the OPAC interface to date. In analyzing the collected studies on user searching behavior, they compared a number of categories of user searching behavior using Web search engines and OPAC (see figure 1).

According to Jansen and Pooch, the majority of searchers on both OPACs and Web search engines use approximately two terms in a query, have an average of two queries per session, do not use complex query syntax, typically view no more than ten documents from the result list, and rarely use Boolean operators.<sup>50</sup>

With all of these similarities, why do searches seem to be more successful in Web search engines than in OPAC? The authors believe the answer lies in the interfaces of OPAC, the search models, the sequence of menu options, and document collection. While document collection is beyond the scope of this paper, the impact of Web search-engine interfaces and search models reflected in literature is discussed.

## Transaction-Log Analysis at CSULA

TLA was performed on CSULA's library catalog over three consecutive quarters (summer 2000, fall 2000, and winter 2001), and then repeated in the winter quarter for two more years (winter quarters 2002 and 2003). For consistency, all sampling was done in the second half of the quarter. The focus of the studies in 2000–2001 was to examine the problems encountered by students—the majority of whom speak English as a second language (ESL)—in searching the library catalog, and whether the types of problems were different in an ESL environment. That study found a high rate of spelling and typographical errors as well as problems with abbreviations. Changes were made in the search interface and engine, and the study was continued to determine the effect of the changes. As in any unobtrusive study, user intent or actual user satisfaction remained unknown. Actual time logs were not studied, and it was not determined whether the user eventually achieved success.

The library uses Innovative Interfaces, which provides an automatic-logging facility for zero-hit searches as well as the reporting capability of the number of searches performed and the number of hits retrieved. Zero-hit searches were reviewed by one of the authors. Each zero-hit search

Category	Web Search Engine	OPAC
Session length (number of queries per user per session)	1–2	2–5
Query length (number of terms per query)	2	1–2
Number of relevant documents viewed per session	10 or less	less than 50
Use of advanced features (percentage)	9	8
Use of Boolean (percentage)	8	1
Failure rate (percentage)	10	7–19

**Figure 1.** Comparison of User Searching Behavior Using Web Search Engine and OPAC

was assigned to as many categories as it appeared to belong to (typographical error, number, and subject syntax). This is in contrast to both Peters and Hunter, who assigned errors to the least problematic and most difficult problem respectively. Hunter's finding, that many transactions had more than one problem, was borne out in this analysis.

## Categories of Problems

Categories of problems were developed dynamically as searches were reviewed. Categories created in the first year of the study were carried over to subsequent years. While the categories were developed independently from earlier studies, the categories used in this study are similar to those used by Peters and Hunter. The analysis resulted in the categories listed in table 1. When one matches a Peters or Hunter category, the Peters or Hunter category name is listed in parentheses:

- wrong type of search or wrong field (search),
- incorrect syntax,
- typographical error (typo),
- spelling error (spell),
- abbreviation or number problem,
- inadequate definition of topic, and
- inappropriate for a library catalog search (source).

For instance, when a search would have likely had a more successful result in a field other than subject, it was assigned "wrong type of search or wrong field." Titles entered as subjects were assigned to this category (e.g., "ASCE journal" or "the story of"). In most cases, the search was reentered into the database to see if the results were improved in the new field.

A search was assigned to the category "incorrect syntax" when the search was topical but the structure was

incorrect for LCSH (e.g., “women roles in fairy tales,” “women under yuan dynasty 1260 1368,” “latin american children books”).

When the user typed something that appeared to be a spelling error (e.g., “feenix” for “Phoenix”), the search was assigned into the spelling category. In particular, phonetic spellings were assigned as spelling errors. However, spelling errors could have been typographical errors. When a number was entered (e.g., “world war 2”) or an abbreviation entered (e.g., “ADS”) which might have had better results spelled out, these searches were assigned to the number or abbreviation categories, respectively.

Searches that appeared ill-defined or were known to the authors to be word-for-word out of course assignments were assigned the category “inadequate definition of topic” (e.g., “what role arts play in our society” or “key indicators” or “different ethnic sports”).

Very specific searches for facts or current information were categorized as “database,” implying that the user would have achieved success more quickly in a database or in the index of a particular book (e.g., “evil characters in fairy tales”). This latter category of searches often was also considered a concept problem for subject searching.

### Keyword Search Problems

In the summer and fall of 2000, logs of zero hits on keyword searches were analyzed to gain an understanding of the problems users were encountering in searching the catalog. At that time, approximately 8 percent of searches were categorized as more likely to succeed in another search field (Title or Author), 26 percent were spelling errors, 22 percent were typographical, and 32 percent had abbreviations that might have caused problems. Only 5 percent appeared to be more appropriate for a database or reference book. Seven percent had significant problems with the search concept.

### Text-based to Web-based Interface

In 1999, the library began the transition from a text-based interface to a Web-based interface using Innovative Interfaces’s WebOPAC module. By 2000, the library was using both text-based and Web-based interfaces. In the text-based interface, keyword was listed almost halfway

**Table 1.** Zero Hits in Subject Searches

Category of Problem	Peters/Hunter equivalent (if any)	% in 2001 (N = 302)	% in 2002 (N = 779)	% in 2003 (N = 2,244)
Title	Wrong type of search	2.8	6.3	3.8
Keyword	Uncontrolled vocabulary termused, wrong type of search	47.6	65.5	70.1
Author	Wrong type of search	1.9	2.8	2.4
Subject syntax		42.1	20.8	11.9
Subject concept		6.8	6.7	8.0
Spelling	Spelling error	10.5	4.7	4.3
Number		3.8	5.3	2.7
Plural		3.0	3.3	2.5
Typographical error	Typographical error	3.6	8.7	5.6
Stopword	Initial article entered	4.5	3.3	0.5
Abbreviation		2.1	2.6	2.0
Database	Recommend source other than catalog	9.1	32.3	28.9
Discount	Senseless string of characters, garbage entered, malicious entry	2.6	0.9	0.9

down in the menu (figure 2), and received only 11 percent of the searches. In the Web-based interface where keyword was the default search (figure 3), it represented 25 percent of the searches. This increased use of options higher in the menu supports and confirms the findings of both Bleic and Ballard.

### Search and Interface Changes

In January 2001, the library implemented Innovative Interfaces’s Advanced Keyword Search (AKS) option and changed the interface (figure 4).

The AKS option ignores leading articles in user input and, in the case of zero-hits, broadens the search iteratively. The search-broadening techniques are applied in the following order until results are found. Innovative Interfaces allows libraries to set how many of these options are implemented before returning a zero-hits response to the user. The library implemented options one through three.

1. implied adjacency is modified to AND,
2. implied adjacency is modified to OR,
3. field limits (SU or AU) are ignored, terms are combined with OR,

4. first word only is searched, and
5. words are truncated.

Once the AKS option was implemented, there were no zero-hits in keyword search beginning in winter (January) 2001. However, there were still zero-hits in subject searches.

The Web interface was changed in the fall of 2001—keyword searching was placed higher on the menus and made the default while subject searching was moved farther down on the search options and relabeled “LCSH Subjects.” These changes closely model the recommendations made by Thorne and Witlatch. The desired effect was to discourage use of the subject search when a user really had keywords and not subject headings. Keyword searching rose dramatically while subject searching fell almost as much and title searching also dropped. Figure 5 shows how the relative use of each major type of search changed over the four years of the study.

The subject search required LCSH to be used for a successful search. Approximately 50 percent of these searches retrieved no results, while almost 20 percent retrieved more than one hundred hits. Subject-heading problems included keyword searches, spelling, abbreviations, numbers, and plurals.

### User Success

Following implementation of the revised interface (figure 4), user success (defined merely by retrieving between one and one hundred hits) from 2001 to 2002 remained the same in keyword searches (70 percent). Success in subject searches increased from 34 percent to 41 percent. However, there was a large drop in the percentage of subject searches: from 29 percent down to 5 percent. This may indicate that an increased number of users who chose to search subject understood the syntax of LCSH.

Searches that retrieve too many results, defined as one hundred or more, may be as unsatisfactory as searches that retrieve zero hits. Because standard practice on the Web is to display results ranked by relevance, the issue of many results is less problematic. INNOPAC’s Advanced Search allows the library to choose to display results by relevance, by date, or in alphabetical order. The CSULA library chose to display by date since the library’s users often requested recent material. However, if users view only one or two pages (as indicated by Jansen and Pooch), a combination of relevance and date ranking should be considered since Advanced Search does increase the number of hits.<sup>51</sup>

Zero-hit searches that were caused by improperly formatted subject searches (incorrect subject-heading syntax), as shown in figure 6, declined from more than 40 percent to approximately 12 percent, reinforcing the assumption that those few users who chose to do a subject search had some knowledge of the structure of LCSH.

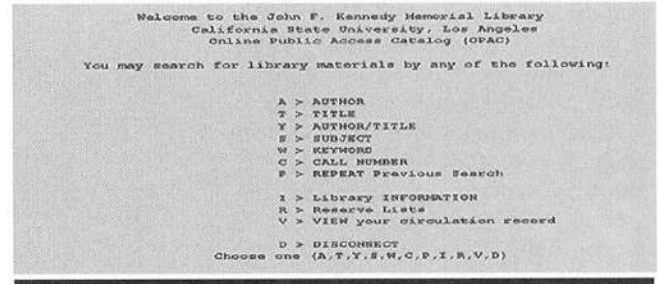


Figure 2. Text-based Search Screen (1999–2000)

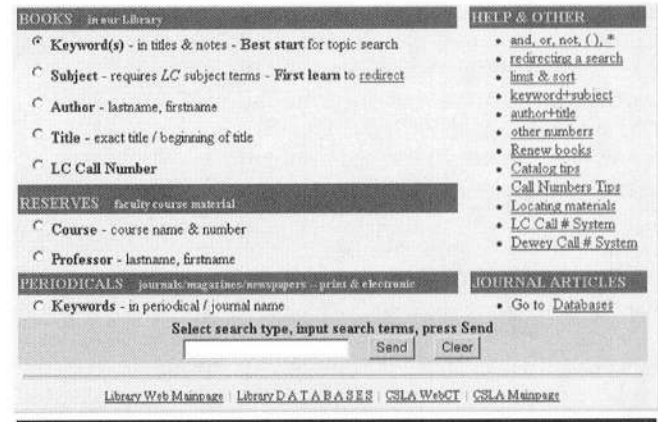


Figure 3. Web-based OPAC Search Screen (2000)

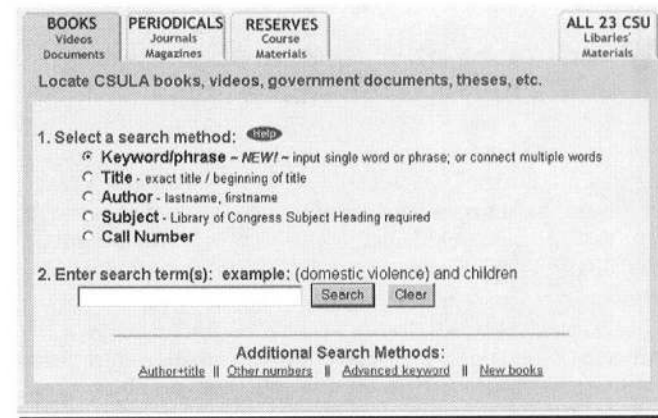


Figure 4. Search Interface after Implementing Advanced Search (January 2001)

Searches categorized as more appropriate for databases or book indexes increased almost 20 percent over these three years—from 9 percent to 28 percent. This type of problem thus became the second most common situation

after the consistent number one problem throughout—"search more appropriate for a keyword index." Users appear to be using the catalog as a single hammer rather than taking advantage of the array of tools a library presents to the user, and this trend appears to be accelerating. The authors hypothesize that the prominence of Web search engines into which users can type anything (Google's single-search box) has led users to assume that a library catalog searches everything (journal articles, full-text resources). Blazek and Bilal and Moore found that users wanted journal-article titles included in the library catalog even two decades ago.<sup>52</sup> The dramatic increase in catalog searches that would be more successful in other sources indicates to the authors that an all-in-one or metasearch approach is desired by users and is, in fact, already assumed to exist. Users may wish to limit results or select from a "product list" (otherwise known as formats), but the authors propose that this would best be done to refine the search rather than to begin it. As Bates pointed out, browsing from a list of options is easier than making the initial search input.<sup>53</sup>

## Influence of Web Search Engines and Online Bookstores

Jansen and Pooch report a CommerceNet/Nielsen Media 1997 statistic that 71 percent of Web users use search engines.<sup>54</sup> In her recent article, Luther asserts that "Google has radically changed users' expectations and redefined that experience of those seeking information." She concludes that for many searchers only "the quality of the results matters less than the process—they just expect the process to be quick and easy." Luther further comments that Google provides "good enough" answers by relying on algorithms that include relevancy ranking of popular culture. She predicts that metasearch technology could be a way to "meet the expectations and needs of the Google generation."<sup>55</sup> The use of Web search engines is sufficiently ubiquitous to have an impact on users' expectations of a Web-based OPAC. In spite of the prevalence of Web searching, Muramatsu and Pratt reported in 2001 that users have multiple misconceptions about how search engines process their queries and that this leads to poor decisions and dissatisfaction with some search

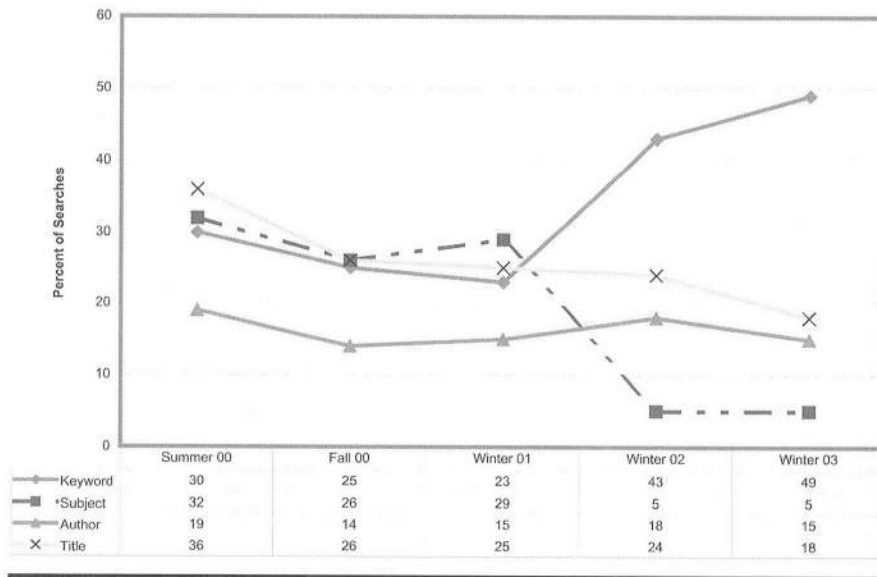


Figure 5. Types of Searches over Time

engines.<sup>56</sup> The authors believe that the same lack of understanding applies to OPACs and leads to dissatisfaction with the library itself.

Web search engines and online bookstores have a number of features that are not typically incorporated into OPACs. These functions include: natural-language entry, automated mapping to controlled vocabulary, spell-checking, similar pages, relevance-ranked output, popularity tracking, and browsing. "Search inside the book," recently implemented by Amazon, includes full-text searching of books as part of the regular keyword search for books. This feature will raise users' expectations of library catalogs even further.

## Natural-Language Searching—Free Text versus Controlled Vocabulary

The Web contains hundreds of millions of pages of information, but they are not cataloged under a controlled vocabulary or subject headings. Search engines typically rely on full-text searching. Research by Jansen and Pooch has shown that search-engine users have a difficult time developing search queries.<sup>57</sup> Search engines do not in most cases require users to enter complicated command strings that use search statements, subject headings, or Boolean operators. Search inputs tend to be simple, averaging two or three words, and Boolean operators are used infrequently and incorrectly. Luther points out that "searching with Google is as easy as entering keywords in a single search box."<sup>58</sup> Web search interfaces often allow users to enter titles with initial article and author names in any order. Bleic et al. observe that the influence of Web

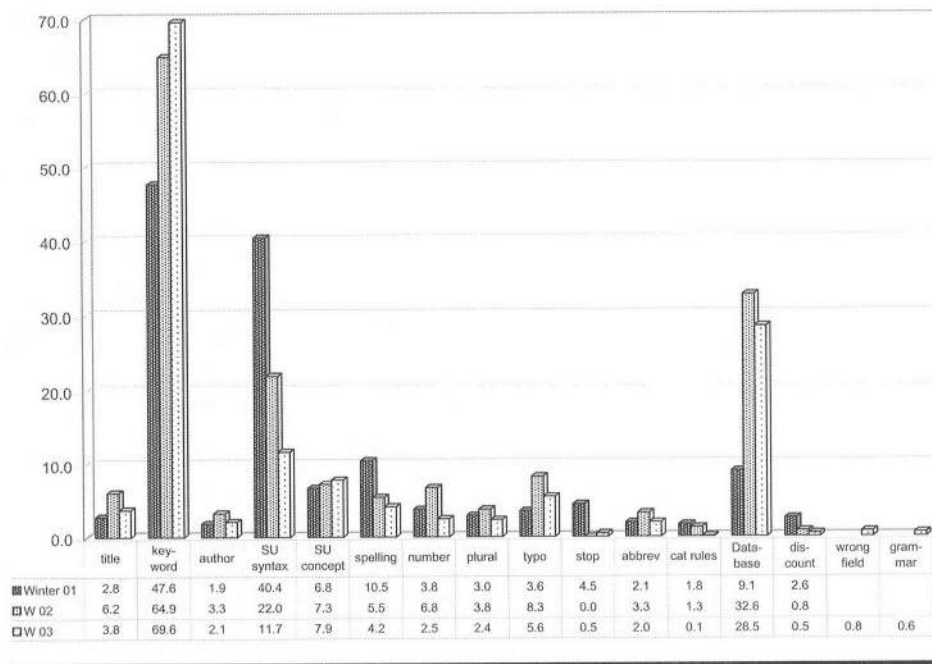


Figure 6. Reasons for Zero Hits in Subject Searches

search engines may have had an impact on the declining percentage of both correct syntax and zero-posting searches (such as title searches that included the initial article, and author searches entered in the incorrect order).<sup>59</sup> Users appear to be “unaware that search engines operate differently from other information-retrieval systems they may use, such as a library online catalog, and this appears to contribute to inappropriate search queries.”<sup>60</sup> OPACs need to keep pace and allow natural-language searching capability.

### Automated Mapping to Controlled Vocabulary

OPAC researchers have found that users build queries using only terms they have in mind and do not consult a thesaurus before a search is conducted. They perform only minor query modifications and refinements, and overwhelmingly apply screen browsing and trial-and-error strategies, the two least effective strategies.<sup>61</sup> The authors’ results continue to confirm this observation. Thus, the automatic expansion of a free-text keyword search helps get users started and reduces the number of zero-hit searches.

Hypertext linking of subject terms and authors is a first step. A user can click on the subject headings or controlled vocabularies in one book record and utilize a subject-heading search without knowing how to construct a subject heading. This feature helps users find works related to one another and complements a free-text search.

Further enhancements could be automatic mapping and taxonomy development. NLM’s MedlinePlus takes natural-language queries and converts these queries into controlled vocabulary. Taxonomies, constructed based on the content at hand and serving to divide the content into manageable chunks, typically using a faceted classification, are regularly being constructed in the business environment to aid in retrieval of internal Web content. In a sense, this is customized grouping of subject headings for each local collection and is in direct contrast to the library approach of shared-copy cataloging.

### Spell Checking

Spelling and typographical errors account for over 15 percent of the zero-hits keyword searches in

this study. Google’s “Do you mean this?” feature has proven to be an effective way to provide correct spellings for users. Google’s spell-checking software looks at the query entered by the user and checks to see if the spelling uses the most common spelling of a word. If it calculates that an alternative spelling will generate more relevant search results, Google asks, “Do you mean: [a more common spelling]?” Google’s spell check is based on occurrences of words used on the Internet, and “it is able to suggest common spellings for proper nouns (names or places) that might not appear in a standard spell-check program or dictionary.”<sup>62</sup> Misspellings recorded by catalog transaction logs could be reused as a basis for a function like Google’s spell check at a library-specific level.

### Similar Pages

Salton defined relevance feedback as a classic information-retrieval (IR) technique that reformulates a query based on documents identified by the user as relevant.<sup>63</sup> Relevance feedback has been widely used in IR systems and has been applied in Web search engines. There are different forms of relevance feedback—text analysis, used by most search engines; page status, which Google utilizes; and “communities of interest and expertise” created by Teoma.

GoogleScout and Teoma’s communities of interest and expertise offer similar page options to the user for Web sites with similar results. GoogleScout technology



automatically scouts the Web for pages that are related to its results so it can find a large number of resources very quickly without requiring the user to select the right keywords. Teoma structures the appropriate communities of interest on-the-fly and ranks the results on a range of factors including authorities and hubs (good resources pointing to related resources). Google offers an option of "similar pages." While the subject-redirect function in a Web-based OPAC emulates this, it succeeds only if the user's initial search term yielded the right result. OPAC users have the option of clicking on hyperlinked headings (author, title, subject headings) but cannot ask the system to perform a more sophisticated search on their behalf.

### User-Popularity Tracking

Amazon and Barnes and Noble Web sites present enhanced information about items by user-popularity tracking. Circulation statistics or user comments could serve as a form of "recommender system" to help novices narrow their selections. Messages such as "other students who checked this book out also read these books" could be dynamically inserted in bibliographic records. Users could also be allowed to provide comments on materials in the catalog, thus providing an interactive experience for OPAC users.

### Summary of Web Features

There are positive and negative impacts of Web search engines and online bookstores on Web-based OPAC users. Users who find Web pages to be comfortable, easy, and familiar may make greater use of Web-based OPACs. While they bring with them their knowledge of search engines, they also bring their misperceptions. The possibility of using similar tools to those found on Web search engines can greatly "reinforce the usefulness of the catalog as well as the positive perception that the end user has of it."<sup>64</sup> Given the diversity of the errors that users experience, a combination of approaches is necessary to improve their search success. Automatic mapping of free-text-to-thesaurus terms, translation of common spelling mistakes, and links to related pages are tools already in use in the Web search engines. "See similar pages," extensive use of relevance feedback, and popularity tracking along with natural language are less common.

## Recommendations for Web-based OPACs

The authors' TLA revealed a continuing problem with subject-heading searches and showed a trend toward

searching topics that are not typically answered in a book catalog. The former problem has a well-documented history, while the authors believe the latter problem stems from the influence of the Web and Web search engines. Several changes to typical OPACs are recommended to address the trends observed in the course of this study.

### Metasearching

The recent trend of incorporating databases and OPACs into a single search reflects the necessity of expanding information resources and simplifying access to resources. This study's empirical results clearly indicate a need to expand this integration into one search. While some argue that this metasearching will further augment the syntax digression and prevent users from becoming information literate, others believe that metasearching, along with the option of searching each individual database, is an ultimate goal for online search. Like it or not, the metasearch technology, also known as federated or broadcast search, "creates a portal that could allow the library to become the one-stop shop their users and potential users find so attractive."<sup>65</sup> One-search-for-all cannot solve all problems; however, guiding users to where they are most likely to find results quickly (the quick search) should satisfy the needs of the majority of users.

### Menu Sequence

Effective screen design has a positive effect on user success. The menu sequence for search options plays a significant role in user selection. This research and others it have demonstrated that users choose an option higher rather than lower in a list. Too many options "simply cause confusion, at least for less experienced OPAC users."<sup>66</sup>

### Browsing Feature

Browsing is a natural and effective approach to many information-seeking problems and requires less effort and knowledge on the part of the user. The literature suggests that a great deal of the use of the Web relies on known Web sites, recommended sites, or return visits to sites recently visited—thus relying on browsing rather than on searching. Jenkins, Corritore, and Widenbeck found that domain novices seldom clicked very deep—out and back—while Web experts explored more deeply.<sup>67</sup> Holscher and Strube note that Hurtineene and Wandtke claim that only minimal training is necessary for browsing an individual Web site, while Pollok and Hockley claim that considerably more experience is required for querying and navigating among sites.<sup>68</sup>

Hancock-Beaulieu found that between 30 percent and 45 percent of all online searches, regardless of the type of search, are concluded with browsing the library shelves.<sup>69</sup>

A browsing option can assist users to effectively find the desired documents by clustering related documents based on terms in a thesaurus.

## Results Display

Effective display helps users find what they are looking for. Page layout or position of labels, text, and instructional information "eases eye movement and improves the clarity of the overall screen."<sup>70</sup>

Web searchers expect graphics, so as Chan indicates, it is possible to make use of icons both to display information and to provide direction to the catalog user.<sup>71</sup> New display options are increasingly taking advantage of the graphical capability of the Web. A recent example by Union College's Schaffer Library in New York State (<http://libraryopac.union.edu/>) presents a promising new format for both citation display and full-record display (figures 7 and 8, respectively). Incorporating graphics makes the screen easy to read and field labels easy to understand.

## Relevance Ranking

Study after study has demonstrated that users will not scroll through more than one or (at most) two pages of references. Web search engines display results based on their calculation of relevance. While full-text searching makes relevance ranking more helpful due to the size of each result and the number of occurrences of each word, users now *expect* relevance ranking. Library systems could design a relevance-ranking algorithm based on the criteria of users in determining relevance. This would likely include currency (date of publication) and subject headings and terms. User popularity and term frequency in tables of contents might also be among the factors used.

## Helpful Hints

The literature has long discussed helpful hints in OPACs, including help by means of query expansion and search tips.

Context-sensitive help may be used to assist users in question negotiation. The CSULA catalog utilizes context-sensitive search examples on its basic search screen. As shown in figure 9, when the user selects a search method, a context-sensitive example appears above the search-input box that tells the user how to input an author search, a title search, or a keyword search when that option has been selected. This approach minimizes clutter on the screens.

Brajnik et al. explain that help should be provided autonomously by the system without the user's request, offering tools and concepts that will enable users to generate better strategies.<sup>72</sup> Another approach is

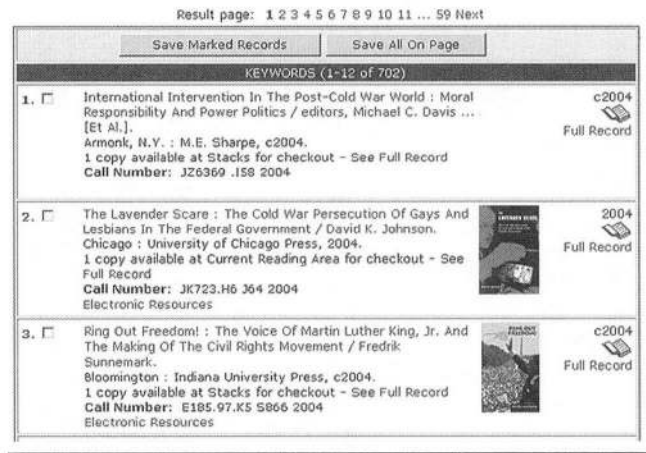


Figure 7. Citation Display



Figure 8. Full-Record Display

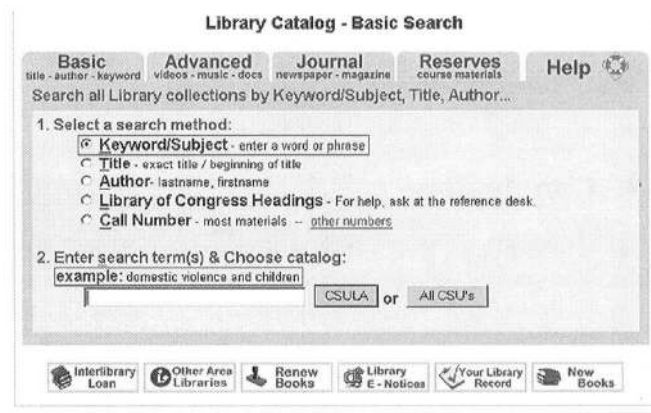


Figure 9. CSULA Context-Sensitive Example

to implement user help through tips or tactics selected and accumulated from a collection of common user-search mistakes. In such a case, the system would play a more active role by generating relevant search tips on the fly and using zero-hits search results as a basis for generating a spell check or suggesting alternate wording.

An ideal scenario is that OPAC allows the user to pursue multiple avenues of an inquiry by entering fragments of the question, exploring vocabulary choices, and reformulating the search with the assistance of various specialized intelligent assistants. Borgman suggests that an OPAC should be judged by whether the catalog answers questions rather than merely matches queries. She suggests the need to design systems that are based on behavioral models of how people ask questions, arguing that users still need to translate their question into what a system will accept.<sup>73</sup>

## User Instruction

On-site training and online documentation can help make it easier to use OPAC. With the advent of information literacy, the shift in library instruction from procedure-based query formulation to question-being-answered has taken place. At CSULA, instruction for entry-level classes focuses on formulating a research statement and then identifying keywords and alternate terms. The instruction sessions that follow the initial-concept formulation are short and focus on how to enter keyword or subject, author, and title, and the use of Boolean operators. This approach may improve success until the systems provide the tools to improve search strategies or accept an untrained user's input.

As an increasing number of users access online library catalogs remotely, assistance needs to be embedded into intuitive systems. "Time invested in elaborate help systems often is better spent in redesigning the user interface so that help is no longer needed."<sup>74</sup> Users are not willing to devote much of their time to learning to use these systems. They just want to get their search results quickly and expect the catalog to be easy to use with little or no time invested in learning the system.

## Conclusion

The empirical study reported in this paper indicates that progress has been made in terms of increasing search success by improving the OPAC search interface. The goal is to design Web-based OPAC systems for today's users who are likely to bring a mental model of Web search engines to the library catalog. Web-based OPACs and Web search engines differ in terms of their systems and interface design. However, in most cases, these differences do not

result in different search characteristics by users. Research findings on the impact of Web search engines and user searching expectations and behavior should be adequately utilized to guide the interface design.

Web users typically do not know how a search engine works. Therefore, fundamental features in the design of the next generation of the OPAC interface should include changing the search to allow natural-language searching with keyword search first, and focus on meeting the quick-search need. Such a concept-based search will allow users to enter natural language of their chosen topic in the search box while the system maps the query to the structure and content of the database. Relevance feedback to allow the system to bring back related pages, spelling correction, and relevance-ranked output remain key goals for future OPACs.

## References and Notes

1. Sharon Seymour, "Online Public-Access Catalog User Studies: A Review of Research Methodologies, March 1986–November 1989," *Library and Information Science Research* 13 (1991): 89–102; Andrew Large and Jamshid Beheshti, "OPACs: A Research Review," *Library and Information Science Research* 19 (1997): 2, 111–33.
2. *Ibid.*, 113–16.
3. *Ibid.*, 116–20.
4. Thomas A. Peters et al., "An Introduction to the Special Section on Transaction-Log Analysis," *Library Hi Tech* 11(1993): 2, 37.
5. Thomas A. Peters, "The History and Development of Transaction-Log Analysis," *Library Hi Tech* 11 (1993): 2, 56.
6. Pauline A. Cochrane and Karen Markey, "Catalog Use Studies since the Introduction of Online Interactive Catalogs: Impact on Design for Subject Access," in *Redesign of Catalogs and Indexes for Improved Subject Access: Selected Papers of Pauline A. Cochrane* (Phoenix: Oryx, 1985), 159–84; Steven A. Zink, "Monitoring User Success through Transaction-Log Analysis: The WolfPAC Example," *Reference Services Review* 19 (Spring 1991): 449–56; Michael K. Buckland et al., "OASIS: A Front End for Prototyping Catalog Enhancements," *Library Hi Tech* 10 (1992): 7–22.
7. Thomas A. Peters, "When Smart People Fail: An Analysis of the Transaction Log of an Online Public-Access Catalog," *Journal of Academic Librarianship* 15 (1989): 5, 267.
8. Michael K. Buckland et al., "OASIS," 7–22.
9. Deborah D. Blecic et al., "Using Transaction-Log Analysis to Improve OPAC Retrieval Results," *College and Research Libraries* (Jan. 1998): 48.
10. Peters, "History and Development of Transaction-Log Analysis," 2, 52.
11. Charles R. Hildreth, "The Use and Understanding of Keyword Searching in a University Online Catalog," *Information Technology and Libraries* 16 (1997): 6.
12. Ray R. 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 and Technology* 42 (1991): 3, 210.

13. John E. Tolle and Sehchang Hah, "Online Search Patterns: NLM CATLINE Database," *Journal of the American Society for Information Science* 36 (Mar. 1985): 82-93.

14. Carol Weiss Moore, "User Reaction to Online Catalogs: An Exploratory Study," *College and Research Libraries* 42 (1981): 295-302; Joseph R. Matthews et al., *Using Online Catalogs: A Nationwide Survey—A Report of a Study Sponsored by the Council on Library Resources* (New York: Neal-Schuman, 1983), 144.

15. Rhonda N. Hunter, "Success and Failures of Patrons Searching the Online Catalog at a Large Academic Library: A Transaction-Log Analysis," *RQ* 30 (Spring 1991): 399.

16. Noelle Van Pulis and Lorne E. Ludy, "Subject Searching in an Online Catalog with Authority Control," *College and Research Libraries* 49 (1988): 526.

17. Hildreth, "The Use and Understanding of Keyword Searching," 6.

18. Ray R. Larson, "The Decline of Subject Searching," 3, 60.

19. Ibid.

20. Van Pulis and Ludy, "Subject Searching in an Online Catalog," 527.

21. Karen Markey, *Research Report on the Process of Subject Searching in the Library Catalog: Final Report of the Subject Access Research Project* (report no. OCLC/OPR/RR-83-1) (Dublin, Ohio: OCLC Online Computer Library Center, 1983), 529.

22. Peters, "The History and Development of Transaction-Log Analysis," 2, 43.

23. Hildreth, "The Use and Understanding of Keyword Searching," 8-9.

24. David R. Gerhan, "LCSH in vivo: Subject Searching Performance and Strategy in the OPAC Era," *Journal of Academic Librarianship* 15 (1989): 86-87.

25. Joan M. Cherry, "Improving Subject Access in OPACS: An Exploratory Study of Conversion of Users' Queries," *Journal of Academic Librarianship* 18 (1992): 2, 98.

26. Rosemary Thorne and Jo Bell Whitlatch, "Patron Online Catalog Success," *College and Research Libraries* 55 (1994): 496.

27. Peters, "The History and Development of Transaction-Log Analysis," 2, 48.

28. Ibid.

29. Hunter, "Success and Failures," 400.

30. Ibid., 399.

31. Ibid., 400.

32. Peters, "The History and Development of Transaction-Log Analysis," 2, 56.

33. Hunter, "Success and Failures," 400.

34. Hildreth, "The Use and Understanding of Keyword Searching," 6.

35. Patricia M. Wallace, "How Do Patrons Search the Online Catalog When No One's Looking? Transaction-Log Analysis and Implications for Bibliographic Instruction and System Design," *RQ* 33 (winter 1993): 3, 249.

36. Large and Beheshti, "OPACS: A Research Review," 125.

37. M. M. Hancock-Beaulieu, "Online Catalogue: A Case for the User," in *The Online Catalogue: Developments and Directions*, C. Hildreth, ed. (London: Library Association, 1989), 25-46.

38. Terry Ballard, "Comparative Searching Styles of Patrons and Staff," *Library Resources and Technical Services* 38 (1994): 293-305.

39. Jane Scott et al., "@\*#&@ This Computer and the Horse It Rode in On: Patron Frustration and Failure at the OPAC" (in "Continuity and Transformation: The Promise of Confluence":

# USABILITY, ACCESSIBILITY, YOUR LIBRARY.

## User Interface Consulting

FEDERATED SEARCH ENGINES

LIBRARY PORTALS & DATA CENTERS

OPACS & CHILDREN'S CATALOGS

DIGITAL LIBRARIES

**Ezra Schwartz**

(773) 256-1418

ezra@artandtech.com

<http://www.artandtech.com>

Proceedings of the ACRL 7th National Conference, Chicago: ACRL 1995), 247-56.

40. Thorne and Whitlatch, "Patron Online Catalog Success," 496.

41. Bleic et al., "Using Transaction-Log Analysis," 46.

42. Virginia Ortiz-Repiso and Purificacion Moscoso, "Web-Based OPACS: Between Tradition and Innovation," *Information Technology and Libraries* 18, no. 2 (June 1999): 68-69.

43. Hildreth, "The Use and Understanding of Keyword Searching," 6.

44. Ortiz-Repiso and Moscoso, "Web-Based OPACS," 71.

45. Ibid., 75.

46. Christine Borgman, "Why Are Online Catalogs Still Hard to Use?" *Journal of the American Society for Information Science* 47 (1996): 7, 501.

47. Van Pulis and Ludy, "Subject Searching in an Online Catalog," 53.

48. Blazek and Bilal, "Problems with OPAC: A Case Study of an Academic Research Library," *RQ* 28 (winter 1988): 175.

49. Deborah D. Bleic et al., "A Longitudinal Study of the Effects of OPAC Screen Changes on Searching Behavior and User Success," *College and Research Library* 60, no. 6 (Nov. 1999): 524, 527.

50. Bernard J. Jansen and Udo Pooch, "A Review of Web Searching Studies and a Framework for Future Research," *Journal of the American Society for Information Science and Technology* 52 (2001): 3, 249-50.

51. Ibid., 250.

52. Blazek and Bilal, "Problems with OPAC: A Case Study," 175; Moore, "User Reaction to Online Catalogs," 295-302.

53. M. J. Bates, "The Design of Browsing and Berry-Picking Techniques for the Online Search Interface," *Online Review* 13 (1989): 5, 407-24.
54. Jansen and Pooch, "A Review of Web Searching Studies," 238.
55. Judy Luther, "Trumping Google? Metasearching's Promise," *Library Journal* 128 (2003): 16, 36.
56. Jack Muramatsu and Wanda Pratt, "Transparent Queries: Investigating Users' Mental Models of Search Engines," *Research and Development in Information Retrieval* Sept. 2001. Accessed Mar. 10, 2003, <http://citeseer.nj.nec.com/muramatsu01transparent.html>.
57. Jansen and Pooch, "A Review of Web Searching Studies," 235.
58. Luther, "Trumping Google," 36.
59. Blecic et al., "A Longitudinal Study of the Effects of OPAC Screen Changes," 527.
60. Susan M. Colaric, "Instruction for Web Searching: An Empirical Study," *College and Research Libraries News*, 64 (2003): 2.
61. A. G. Sutcliff, M. Ennis, and S. J. Watkinson, "Empirical Studies of End-User Information Searching," *Journal of the American Society for Information Science and Technology* 51 (2000): 13, 1213.
62. "All About Google," Google. Accessed Dec. 10, 2003, [www.google.com](http://www.google.com).
63. G. Salton, *Introduction to Modern Information Retrieval* (New York: McGraw-Hill, 1983), 18.
64. Ortiz-Repiso and Moscoso, "Web-Based OPACs," 71.
65. Luther, "Trumping Google," 37.
66. Maaiké D. Kiestra et al., "End-Users Searching the Online Catalogue: The Influence of Domain and System Knowledge on Search Patterns. Experiment at Tilburg University," *The Electronic Library* 12 (Dec. 1994): 335-43.
67. C. Jenkins et al., "Patterns of Information Seeking on the Web: A Qualitative Study of Domain Expertise and Web Expertise," *IT and Society* 1 (Winter 2003): 3, 74,77. Accessed May 10, 2003, [www.ItandSociety.org/](http://www.ItandSociety.org/).
68. C. Holscher and G. Strube, "Web Search Behavior of Internet Experts and Newbies," *9th International World Wide Web Conference*, (Amsterdam. 2000). Accessed Mar. 28, 2003, [www9.org/w9cdrom/81/81.html](http://www9.org/w9cdrom/81/81.html); A. Pollock and A. Hockley, "What's Wrong with Internet Searching," *D-lib Magazine* (Mar. 1997). Accessed May 10, 2003, [www.dlib.org/dlib/march97/bt/03pollock.html](http://www.dlib.org/dlib/march97/bt/03pollock.html).
69. M. M. Hancock-Beaulieu, "Online Catalogue: A Case for the User," 25-46.
70. Wilbert O. Galitz, *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques* (Chichester, England: Wiley, 1996).
71. Juliana Chan, "An Evaluation of Displays of Bibliographic Records in OPACs in Canadian Academic and Public Libraries," MIS Report, Univ. of Toronto, 1995. [025.3132 C454E]
72. Giorgio Brajnik et al., "Strategic Help in User Interfaces for Information Retrieval," *Journal of the American Society for Information Science and Technology (JASIST)* 53 (2002): 5, 344.
73. Borgman, "Why Are Online Catalogs Still Hard to Use?" 500.
74. Ibid.