2013-12-04

Manually Classifying User Search Queries on an Academic Library Web Site

Chapman, Suzanne; Desai, Shevon; Hagedorn, Kat; Varnum, Kenneth J.; Mishra, Sonali; Piacentine, Julie

https://hdl.handle.net/2027.42/101884

http://creativecommons.org/licenses/by-nc-nd/3.0/us/

Manually Classifying User Search Queries on an Academic Library Web Site

Suzanne Chapman, Shevon Desai, Kat Hagedorn, Ken Varnum, Sonali Mishra, and Julie Piacentine

Abstract

The University of Michigan Library wanted to learn more about the kinds of searches its users were conducting through the "one search" search box on the Library Web site. Library staff conducted two investigations. A preliminary investigation in 2011 involved the manual review of the 100 most frequently occurring queries con- ducted through the site search box over the course of a month. Those 100 search terms accounted for 16 percent of total queries and were largely one-word searches for databases. In the follow-up investigation, the Library embarked on a more ambitious exploration of the 454,443 searches conducted during the winter 2011 semester, devising a method for selecting, categorizing, and summarizing user search queries. A sample of 1,201 searches from the search query logs was examined; after eliminating duplicate searches, there were 992 unique terms available for categorization. Using a non-overlapping sample of queries, a rubric was developed for categorizing user searches. Each of seven library staff members reviewed all unique terms in the sample to categorize them into the best fitting category from the rubric. After establishing a threshold for reliability among the individuals categorizing the queries, 862 unique search terms were analyzed. Based on this analysis, the most frequent kinds of searches conducted in the winter semester in 2011 on the University of Michigan Library's Web site were specific databases (28 percent), topical/exploratory types of queries (28 percent), and books (including searches by title, ISBN, call number, or a combination thereof) (16 percent).Within the sample, known-item searches comprised nearly half (44 percent) of searches in the sample. An- other fifth (20 percent) of total searches were categorized as "exploratory," supporting the need to provide broader, subject-based paths to information through the site. Somewhat surprisingly, there were a small number of article searches (article titles, or mixed searches of journal names and authors and/or title words) in the search box—an indication that users understand the University of Michigan Library primary search box is not for articles.

Introduction

The University of Michigan Library (MLibrary) launched a Web site[1] in 2009 with a focus on the search box as the primary means for users to interact with and discover resources. Because an analysis of popular searches was used during the Web site's design process to influence the organization of search results, there was some confidence that the Web site performed well for the most commonly entered search keywords, but it was unclear whether it performed equally well for searches in the "long tail"—that is, the vast number of searches conducted only a handful of times. Therefore, this study investigated the types of queries users conducted when searching the library Web site.

Background

As is common with libraries at other large research institutions, the MLibrary acts as a gateway to a vast collection of resources, including: full text of online books, journals, and newspaper articles, catalog records for the 8 million books on the shelves, tens of thousands of online journals, hundreds of databases, locally digitized collections, and more. Providing access to these resources is one of the main goals of the search box on the library's Web site, where ten to fifteen thousand keyword searches are conducted each day.

The library's primary search box appears near the top left corner of every page in the library Web site (Figures 1 and 2). The default search encompasses content found in six content areas:

1. Databases (from Search Tools, the library's database finder)

2. Mirlyn, the library's catalog

---

[1] http://www.lib.umich.edu/

3. Online journals (from our A-Z list of online journals)

4. Research guides

5. Library web pages

6. Collections

Figure 1. MLibrary gateway page search box collapsed

Figure 2. MLibrary gateway page search box expanded on mouse-over

A search automatically includes all areas, unless the user deselects one or more of the search content areas in the interface. Only a small minority of searches, less than 10%, is limited to particular content areas. Search results have been presented in a "bento box" interface[2] since 2009. In this style of interface, results are presented "pre-faceted" in groupings based on some predetermined method. In MLibrary's case, the above six content areas were included. Also included was "Deep Blue" (MLibrary's institutional repository), which was a subset of "Library Webpages" but was presented separately to highlight the materials owned by the library (See Figure 3 for the search results page as it existed during this investigation).

---

[2] The term "bento box" to describe search results is credited to Tito Sierra of North Carolina State University.

Figure 3. MLibrary multi-search results page for the search "social work" (early 2011)

Manually Classifying User Search Queries on an Academic Library Web Site

In order for library staff to better understand search interactions with the library Web site, a search logging tool was built. For all user interactions with the site, including search queries, the web server created an entry in its log file. Every night, a custom script read this log file and extracted certain data elements for storage into a MySQL database for later analysis. While the full log files were routinely purged after two weeks, the anonymized search data were kept indefinitely. For each query, the database noted:

- the search query

- the user's geographic location (if on campus, whether the user was in a library building, on a wireless network, or elsewhere; if off campus, the country)

- the date and time of the query

- the URL of the page the user was on when the query was executed

- any search limits specified.

Initial Usage Analysis

In an internal study undertaken in early 2011, library staff reviewed the search log data for all search queries using the default search box from the month of January 2011. At that time, the new search interface had been in place for three full academic semesters and there was a desire to better understand if the interface was meeting user expectations.

For this initial analysis, staff examined the top 100 most frequent queries entered via the search box on the MLibrary Web site in January 2011. These frequent queries represented a total of 22,177 individual searches (roughly 16% of all searches conducted that month). Each query was categorized using five distinct categories created by the staff reviewing the 100 queries.

These categories were: "specific database," "library service," "publication," "keyword/subject,"
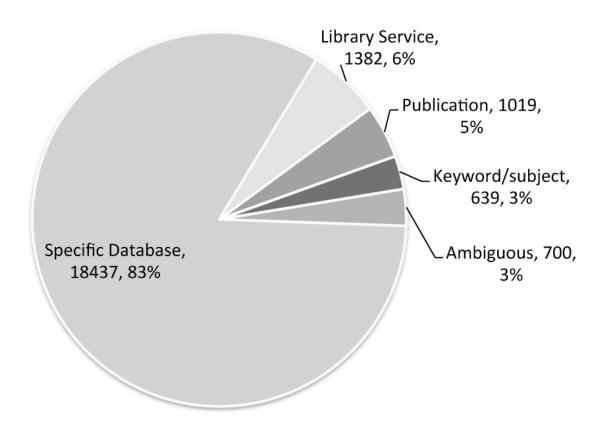
or "ambiguous" (See Figure 4).

Figure 4. Top 100 most frequent queries by category (numbers represent total number of

searches)

Of the most frequently searched terms, a significant majority (83%) was assigned to the

"specific database" category. Assignments to other categories were: "library service" (6%),

"publication" (5%), "ambiguous" (3%), and "keyword/subject" (3%). Given this preliminary

review, it was concluded that databases are the most sought-after items from the library search

box. However, there are over 400,000 searches conducted each semester; investigating only the

top 100 most frequently searched queries does not take into account the "long tail" of searches

conducted. In other words, the most frequently searched queries only represent a small percentage of overall searches. By not investigating all the searches from a particular time period, the majority of searches were not examined.
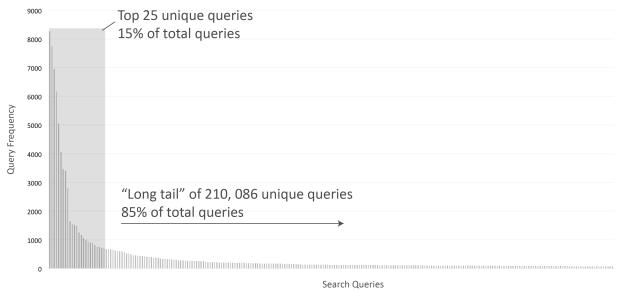

Figure 5. "Long tail" of searches on MLibrary site (January 2011)

Although this initial study provided insight into what was most frequently searched through the library's default search box, it did not shed much light onto the breadth of activity on the site. We wanted to better understand a fundamental question: what were our users searching for and did the structure of search results pages reflect those needs? We wanted to understand what was going on with the other 85% of searches out in the long tail.

Literature Review

There has been much discussion about academic library Web sites, including how sites and homepages have evolved over the years. Geoffrey Little discusses these changes in his 2012 article "Where Are You Going, Where Have You Been? The Evolution of the Academic Library Web site." He points out that many times academic libraries do not consciously think about "differences between how we design, package and market our web sites and services and our

OPACs [...] Librarians think about web sites and OPACs as two very different beasts. From the user's perspective, however, this division is not obvious and makes no real sense" (p. 124). The MLibrary "one search" on the library Web site was created partly in response to these concerns. While the "one search" does not search across databases at the article level, it does search the library catalog, collections, and policy pages on the Web site, as well as our research guides.

Arguably, the MLibrary search is one of the most important pieces of the library homepage. While it is unclear as to whether a single search box on a library homepage is the most effective design for patrons (Swanson and Green, 2011), this approach had already been implemented. Therefore it was worth examining what kinds of search queries were being entered into the MLibrary search. While usability studies had been conducted on various aspects of the University of Michigan library Web site, there had never been an in-depth data analysis of the search itself, specifically examining what terms users searched.

There are many examples of published research performed on the topic of search or log queries and their analysis, in particular within the computer science literature. Some of these articles looked at query data for commercial web search engines (Broder 2002; Gabrilovich 2009) while others addressed queries conducted on a more limited Web site or content collection, such as the Utah State government Web site (Chau et al. 2005) or a specific digital image portal (Hollink et al. 2011).

The majority of these studies used some form of automated categorization or semantic analysis. Surprisingly, very little precedent was found for a manual semantic analysis, rather than an automated process. The following review focuses on those studies that manually applied a semantic categorization to the search query information being analyzed. Although few examples of semantic classification of search queries were uncovered, the approach of this study used a

qualitative method of data analysis. Categorization or classification is recognized as an important part of any method of analytic induction (Preissle 2008).

In several instances, search queries have been grouped or categorized, but are predominantly grouped by "format" or "type" of query. Andrei Broder (2002) looked at data gathered from the AltaVista search engine, and created a "taxonomy of web search," organizing queries into three large categories: navigational (also called "known item"), informational (trying to fulfill a subject information need, through links to additional resources or directly to the content), and transactional (where it appeared that the user's intent was to perform some sort of web-mediated activity). It is interesting to note that Broder made a point of acknowledging that there is no sure way of knowing user intent, "inferring the user intent from the query is at best an inexact science, but usually a wild guess" (Broder 2002, 5). However, few details related to the method of the manual log analysis, such as how many people looked at the search queries, how decisions were made regarding the choice of taxonomical classification, etc, were provided.

Steve Jones et al. (1999) conducted a log analysis of approximately 30,000 queries of a subject-specific digital library. The study used both automated and manual analysis of logs, and examined such details as the most common query terms, query length, query refinement, and use of Boolean operators. While Steve Jones, et al. indicated that there was some human analysis of the queries (such as identifying queries as "problematic" from a "semantic point of view"), there are no details about how this human analysis was conducted, or even if all 30,000 queries (or just a sample) were inspected by human eyes (Jones 1999, 165). There was no attempt to semantically group search queries, possibly because of the limited, subject-specific nature of the collection.

There have been other examples of query and log analyses within the library environment, including examinations of search transactions on library Web sites (Lown et al., 2011; Madle et al. 2006). In the study by Cory Lown and colleagues (2011), 1.4 million transaction logs were examined in order to examine recorded searches and resulting clicks. However, the content of the search queries themselves was not analyzed (2011). Gemma Madle and her colleagues looked at the web logs of a subject-focused digital collection, the National Electronic Library of Infection (2006). Although search queries were not categorized semantically, they were grouped by "format" into single word searches (over half the queries) and those of Boolean construction (only 3.5%) (Madle et al. 2006, 144).

One study was found that did employ some level of manual analysis (Song, et al. 2009). These researchers used human annotators to classify "ambiguous" queries on a Web site search engine, in order to eventually "observe that query ambiguity is to some extent predictable" (Song 2009, 216). Five reviewers examined each of 60 ambiguous terms; ambiguity is defined as those terms for which the query has more than one meaning (differentiated from a "broad" query, which has many sub-topics). Similar to the method of human analysis in this article, each reviewer performed a series of steps in order to determine the "type" of search—ambiguous, broad, or clear. Tools that were used included dictionaries, thesauri, and web search results. Reviewers then "voted" on the "type" for each search query. This corresponds with the steps used in the current study for human reviewers to determine categorization of search queries and to reach a consensus on each.

Methods

Developing a method

Following the initial usage analysis, researchers sought a valid method for analyzing MLibrary's search queries that would include all the search queries (not just the top 100) from an appropriately chosen period of time. With the complete set of 454,443 queries searched via the MLibrary search box from January 1 through April 30, 2011, a random sample of 1,201 was selected for deeper examination. This sample size was chosen in consultation with our on-campus Center for Statistical Consultation and Research (CSCAR).[3] Microsoft Excel extracted the random sample set by choosing every 378th term in the data set. Only unique queries were examined within the sample set. To analyze repeated, identical queries would have added unnecessary time to the analysis. By consolidating identical queries, the final random sample was reduced to 992 queries. In addition, variants in capitalization (e.g., "JStor," "jstor," and "JSTOR") were considered to be identical. However, any queries that appeared to be different names for the same resource (e.g., "isi," "web of science," and "isi knowledge") were not consolidated. Fifteen queries that could have been either spam or hacking attempts,[4] were also eliminated.

Developing the classification scheme

The original, collaboratively developed categories used to analyze the top 100 queries (see Figure 4) were not sufficiently comprehensive or descriptive for this larger sample and were therefore redefined. A pilot set of 50 queries (these were the first 50 queries from the full set of

---

[3] More information about CSCAR can be found at http://www.cscar.research.umich.edu/. We consulted CSCAR because we wanted to make sure our sample size would be sufficiently large to indicate statistical confidence in our results. As classification was a manual process involving multiple librarians reviewing the entire sample, the smallest sufficient sample size was desired. CSCAR provided confirmation that our proposed sample size – roughly 1000 searches – would be enough.

[4] All the eliminated queries followed the same pattern of a complex Boolean search that included a single gibberish word in parentheses for each term, as in this example: "(bQZZEsPMTxhzln) AND Author:(TovDuaJcqeBq) AND Title:(fMshrtYbRWLJBPg) AND SubjectTerms:(ulEtCzUXiAerNgdehg)."

queries) was examined. Without discussing categories beforehand, each member of the analysis group looked at the pilot set and individually came up with a classification scheme with categories that he or she thought were appropriate to use in organizing the queries. Each individual was told that each query must fit inside one category, so all queries were assigned a category. Queries could not belong to more than one category.

The group worked to develop categories that reflected the actual terms searched but that did not necessarily mimic the categories currently used on the library Web site (e.g., Mirlyn, Databases, Online Journals, etc.). After each individual had categorized the 50 queries, individual classification schemes were compared as a group. A shared scheme emerged based on our discussion. Next a second pilot test was run on the next group of 50 queries—also outside the random sample—to test the shared classification scheme. Once complete, the classification scheme was finalized by the group.

Recognized issues and limitations of the classification scheme

The group had to address the following issues and limitations to the classification scheme:

- **To what extent could each search query be investigated for clues about its meaning before being categorized?** Because the individuals in the group may not understand every query, the group had to decide on guidelines for the amount of research an individual needed to do on a query. Research could have entailed individuals performing searches on the MLibrary Web site, correcting for apparent spelling errors and typographical errors, conducting a search in Google or WorldCat, etc. It was determined to be beneficial for those categorizing to conduct a small amount of research, using discretion, in order to make a more informed decision about the category for the query.

- **Could perceived intent be factored in?** The researchers struggled with this question. There was no way the researchers could know conclusively what the patron was intending or expecting to retrieve with their search. For example, a search for "proquest" could be a search for the Proquest database but it could also be a search for information about the company Proquest. The group ultimately decided that some small assumptions about intent were allowable—in the above example, that the intent was far more likely to be a search for the database than about the company—in order to help with categorization.

- **How important was the individual confidence in the categorization of a query and should that confidence factor into the final analysis?** Initially, a discussion ensued about including a rating of individual confidence level for each search query. Instead, a category was included in the classification scheme called "not enough information," defined to mean that the individual categorizers were not sufficiently confident in the assessment of the query to assign it a category.

- **How narrow of a classification scheme could the researchers design?** It became clear that the researchers needed to address not just the nature of the item or the subject being searched (the "what" or "item type" of the query) but also the means of that search (the "how" or "query type"). For example, it was not enough to notice that a patron was looking for a book in a particular query, but that the ISBN, title, author, or a combination of these was being used in the search for that book. This factored heavily into how the classification scheme was designed.

- **How could reliability be addressed across human reviewers and how could the random sample of 992 queries be divided among the reviewers?** No matter how clearly defined the classification scheme was, ultimately the categorization process was

going to be subjective. This would obviously lead to problems related to how one individual categorized a query versus how a different individual categorized the same query. Initially, it was thought that it would be sufficient to have each search query categorized by at least two individuals in the group, with each person analyzing between 200-400 terms each. However, due to concerns over the difficulty of categorizing consistently across the group, it was decided that limiting the sample size handled by each reviewer would not provide consistent results and that each individual would need to categorize the entire set of 992 queries.

To address reliability across human reviewers and determine the final count of the frequency of queries in each category, an official category designation was needed for each query. In consultation with CSCAR, the researchers determined that if at least five out of the seven analysis group members categorized a query with the same category, that category would become the "official" one for that query. Queries that did not pass this threshold were marked as having "no agreement" and were excluded from our data analysis. For example, the search "Ancient Sleeping" received four votes for "exploratory" and three votes for "not enough information," so it was marked as "no agreement," while "American culture" received five votes for "exploratory" and so was officially categorized as "exploratory."

It should be noted that this decision resulted in approximately 25% of queries being categorized as "no agreement." Although this may seem like a large percentage, it would be a mistake to artificially force these queries into a category just for the sake of keeping them within consideration. Weakening the criteria of agreement between five of seven reviewers would weaken confidence in our data, making the study itself less meaningful.

The classification scheme

As discussed above, the entire set of queries was categorized using two dimensions:

a. The nature of the item or the subject being searched. The researchers called this the "what" or the "*item type*" (second column, Table 1 below). Item types included books, databases, multimedia, articles, periodicals, databases, and library Web site content (e.g., hours, collections, staff, etc.).

b. The words the patron used to get at the meaning of the subject, i.e., how did the patron decide to describe the subject of the search query. Researchers called this the "how" or the "*query type"* (third column, Table 1). Query types included title, DOI, ISBN, etc.

As the classification scheme emerged, we needed to determine whether the query was intended to reveal a specific, known-item, or if it was broader and intended to reveal a range of information. To answer this, the "item types" were divided into two different groupings: known-item searches and non-known item searches (first column, Table 1). Queries in which the patron entered the title of something that was recognized as a unique, identifiable object were categorized as known-item searches. Examples include books or articles ("Imputing Missing Data: A Comparison of Methods for Social"), standard numbers such as a DOI ("10.4271/2004-01-1076") or ISSN ("1468-0173"), and databases ("academic search premier"). While there was some debate over whether a search for a specific item was a search *for* that item or for information *about* that item, the group decided to treat them as "for" searches and to categorize them as known-item searches. The other grouping, non-known item searches, was a catch-all for other searches—topics, authors, and anything too vague to categorize as a known item search.

Table 1. Final classification scheme

Type of Search	Item Types	Query Types	Examples
Known item searches	Article	Title	Imputing Missing Data: A Comparison of Methods for Social
		Multi	- sports medicine - Hewett acl men and women
		DOI	10.4271/2004-01-1076
	Book	Title	introduction to hospital and health system pharmac practice
		Multi	unbearable lightness portia de rossi
		ISBN	1577664612
		Call number	HV5825.K89
	Database	Title	academic search premier
		Multi	cengage onefile[5]
	Multimedia	Title	back to the future
		Multi	Black, Brown, and Beige by Duke Ellington
	Multiple formats	Title	Harry Potter and the Half Blood Prince
	Periodical	Title	international journal of clinical practice
		Multi	finite elements wiley
		ISSN	1468-0173
	Specific item	N/A	astm c33
Non-known item searches	Exploratory	N/A	critical access hospital planning
	Creator	N/A	jay galbraith
	Library website content	N/A	online journals
	Not enough information	N/A	- ag .com - alphabetical list

---

[5] The "Database—Multi," "Periodical—Multi," "Periodical—ISSN" and "Specific Item" examples were not found in our final query set of 992 queries, but are included here as good examples that were actually found in one of our pilot query sets.

For many item types there was also a "multi" query type, used to describe queries that used more than one kind of metadata. For example, a search done for "giver lowry" would be categorized as "Book—Multi," since the item type was deemed to be a book and the query contained part of the book's title and part of the author's name.

Some of the "Item Types" in Table 1 require further definition:

- *Multiple formats*. This category refers to queries that could belong to more than one item type. For example, "pride and prejudice" could refer to either the book or a film adaptation.

- *Specific item*. This category refers to queries that are known items but didn't fall neatly into one of the other categories. For example, "astm c33" is a specific standard, but creating a specific category for these few query instances was decided to be overkill.

- *Exploratory*. This category refers to queries that were not known-item searches and that did not have a clear item type. Generally, these are what were considered to be topic and keyword searches. When investigated queries remained ambiguous, researchers decided to bias the results towards *exploratory* on the grounds that categorizing an ambiguous query as a particular item type was a bigger assumption than categorizing it as exploratory. Because *exploratory* was a more vague category that did not involve naming specific types, it implied less about the search on the dimensions we were interested in. Under this system, queries like "probability models" were categorized as exploratory even though there is a book by this title. The group also used clues present in the queries to help make distinctions. For example, capitalization within the search query was treated as a clue that the query was in fact a known item search. For example, if the query had been "Probability Models," it would have been more likely to be categorized as a book

instead of exploratory. Inclusion of qualifying words or prepositions/conjunctions also was a factor in categorization. For example, if the query had been "introduction to probability models," even though the words were not capitalized, it was more certain that the query was intended for a specific book and more likely to be categorized as "Book—Title."

- *Creator*. Queries involving an author were originally categorized with the appropriate item type plus a query type of "author." However, since authors may have written multiple item types, e.g., both articles and books, it was determined that this was too unreliable and recategorized these queries using a separate item type of "creator." This change was made after all the queries had been categorized by individuals, and it was a global change—changing any instance of "author" as a query type to "creator" as an item type. This change also simplified the question of intent for these kinds of queries since these queries may have been intended as searches for information about a person or for a list of works they authored. For example, "Henry Kissinger" could have been a search for information about Henry Kissinger or for materials written by Henry Kissinger.

- *Library Web site content*. Library Web site content was considered to be content that is created by library staff (as opposed to the varied content we provide access to via one of our search services or tools). This could include staff profiles, library hours, research help guides, or information about each library and our services, etc.

- *Not enough information*. As described above, searches were categorized as "Not enough information" if the researchers weren't able to make a well-informed decision on categorization.

Results & Analysis

Once all the categorization results from all seven individuals were combined, the researchers tabulated voter consensus (See Figure 6).

Known-item searches

Out of the 992 queries analyzed, 530 (44%) were grouped as known item searches. The high proportion of known-item searches suggests that most users of the MLibrary Web site have a basic sense of what they are looking for—whether it be the title of a book or the database with which they intend to start their research (e.g., knowing that they want to start with JSTOR). This might also indicate that users understand that MLibrary search is useful for these kinds of searches.

Non-known-item searches

Out of the 992 queries analyzed, 332 (28%) were categorized as non-known-item searches. Unlike the known-item search type, this set of item types (exploratory, creator, and library Web site) tends to be less concrete. These searches were conducted on a variety of subjects, ranging from specific topics (e.g., "impacts of illegal file sharing on motion picture industry") to general keyword searches (e.g., "diabetes"). Although we did not specifically categorize for subject area, a review of these queries shows that a wide range of disciplines were represented.

The MLibrary Web site is designed to act as both a finding aid for users seeking known items and a discovery tool to introduce users to new resources. That being the case, it was a surprise that there were a relatively low number of non-known-item searches compared to

known-item searches in the data set. However, it is not clear that the preponderance of known-item searches means that the MLibrary search performs better as a known-item finder than as a discovery tool. One possible interpretation of the tilt towards known-item searches is that the MLibrary Web site is only one of many possible mechanisms for identifying resources, while in many cases it may be the primary or only vehicle for acquiring those resources. For example, a student may learn about a source through searches outside the MLibrary site, through word-of-mouth from faculty or fellow students, or through a citation in an article they have already found. However, their options for accessing the source may be limited to purchasing the item or accessing it for free through MLibrary.

It is also worth noting that the performance of the entire MLibrary Web site as a discovery tool cannot be judged by the search, which is only one part of the site. Exploratory searches do not represent the full range of exploratory activities within the site, which might also include browsing and serendipitous discovery through Web site features like spotlights.

Not enough information

Out of the 1,201 queries analyzed, 32 (3%) received an official categorization of "not enough information" because the queries were deemed to be too vague to even attempt a categorization. For example, the queries "mcl500.7" and "movie" were deemed to not have enough information to classify.

No agreement (not categorized)

Of all the queries, 75% received at least five out of seven common votes, leaving 25% (307 queries) as "no agreement."
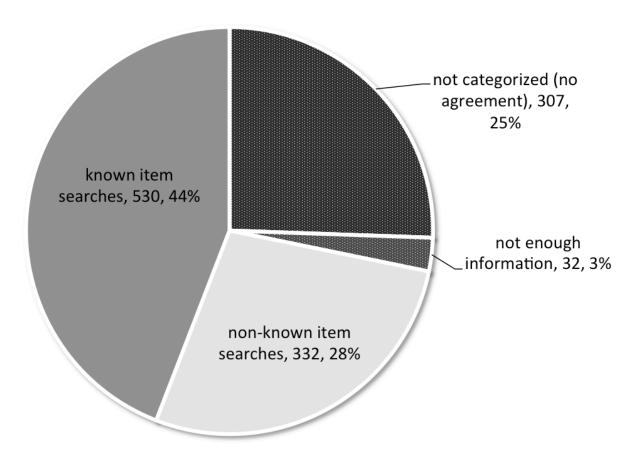
Figure 6. Categorization of the queries, plus those that were not categorized because of "no agreement"

To further analyze the data in Figure 6, queries were organized by item type in Figure 7. This figure shows that databases were the most common type of query, followed by exploratory searches, book, periodical, creator, library Web site content, article, multimedia, and items that could appear in multiple formats.

- *Database (20% of queries analyzed, 246 queries)*: Compared to the initial usage analysis finding that 83% of the top 100 searches were for databases, it was clear that,

as anticipated, the long tail was comprised of diverse search types, with searches for

databases comprising a much smaller proportion of the whole.

- *Exploratory (20% of queries analyzed, 239 queries):* Perhaps unsurprisingly, users

  did as much exploratory searching as they did database searching, which highlights

  the importance of facilitating both types of searching. It was important to stress that

  due to the categorization schema, some of these exploratory searches may have been

  searches for known items. If an ambiguous query was particularly short or lacked

  clues to indicate that the searcher had indeed intended a known item search, the query

  was categorized as exploratory. Because of this, the proportion of queries categorized

  as exploratory searches may be slightly inflated.

- *Book (12% of queries analyzed, 140 queries) and Periodical (8% of queries analyzed,*

  96 queries): Since search is the primary way to retrieve specific books and

  periodicals, it was not surprising that these rank second and third among known-item

  searches.

- *Creator (5% of queries analyzed, 59 queries)*: Our efforts to categorize searches for

  named persons highlighted the ambiguity of this search type. This points to a need to

  better accommodate searches for works *by* a named person as well as searches for

  works *about* a named person.

- *Library Web site content (3% of queries analyzed, 34 queries)*: Although searches

  weren't specifically categorized by topic, a quick review of these queries showed be a

  mix of searches for library staff members (e.g., "paul courant") and queries about

  library services (e.g., "guest borrower card"). Based on usage statistics from this

same period,[6] 17% of clicks on the search results pages are in "library Web site" related sections. While still relatively low compared to use within other sections, this is not as low as the 3% of queries classified in this category. Still, based on both numbers, search activity related to library Web site content is in the minority. This highlights the need to better accommodate other types of searches both in terms of search functionality and search results screen real estate. Currently, over one-third of the screen real estate on the search results page is dedicated to library Web site-related results.

- *Article (2% of queries analyzed, 22 queries)*: Since the primary article search tool (Articles*Plus*) is not included as a content area in the MLibrary search box, the low number of queries for articles could be an indication that users understand that articles are not a part of this search option. Of course, it is worth noting again that many of the "exploratory" and "creator" queries may have been intended as article queries.

- *Multimedia (1% of queries analyzed, 19 queries)*: The small number of multimedia searches may relate to the fact that a larger proportion of the holdings are print materials. In fact, given that multimedia items make up approximately 2% of the collection,[7] the 1% representation of multimedia queries within the sample suggest

---

[6] This figure was calculated using Google Analytics through a custom JavaScript. When a user clicked on a particular item in the search results, Google Analytics recorded a coded URL that included the search query, the section in which the result was located, and the sequential number of the result in the results set. The 17% figure is undoubtedly higher than the number of clicks to the "correct" resource because Google Analytics treated all search results clicks as equal. Google Analytics would treat a user who clicked on three results, only getting to the intended resource on the third click, exactly the same as if three different users conducting the same search had clicked three different results. Additionally, the "library Web site" term in Google Analytics subsumed several sections of the search results page.

[7] This figure is a rough estimate based on the following method. A search for an empty string on Mirlyn (UM's library catalog) in early September 2011 yielded 7,388,238 results. Out of these, 72,480 items are tagged with 'audio' (which appears to include tapes, LPs, CDs and spoken word items); 69,539 items are tagged with 'visual material' (which

that searches for multimedia are approximately proportionate to the occurrence of
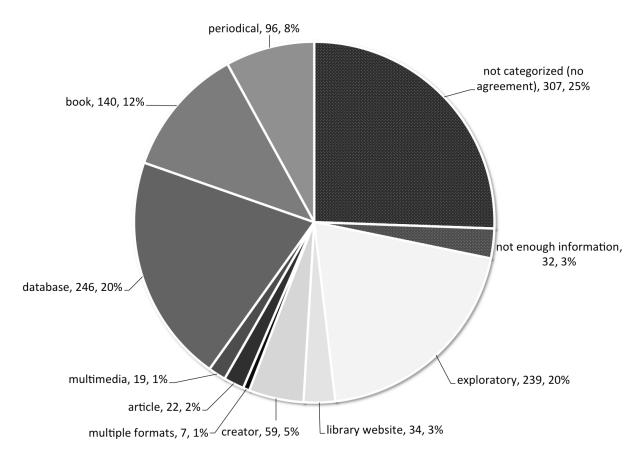
multimedia in the collection.

Figure 7. Organization of categorized search queries by item type, plus those not categorized

through "no agreement"

- *Multiple formats (1% of queries analyzed, 7 queries)*: Given that the classification

  scheme was designed to differentiate between multiple types and formats of queries,

  there were not many queries that ended up being categorized as possibly belonging to

  multiple formats (e.g., could be for a book or a film).

---

appears to include DVD, VHS, slide and motion picture items); 3,676 items are tagged with 'CDROM'; and 2,302 items are tagged with 'video games.' Adding these together totals multimedia making up 2.003% of our holdings.

- *Specific item (0 queries)*: Though this category emerged from the initial subset of 100 queries used to establish the classification scheme, in the final analysis of 1201 queries, none were categorized as specific items.

Future Research

Several areas arose during the course of this study that merit further investigation. While log data indicates that users understand the distinction between the MLibrary and Articles*Plus* tabs, it is not entirely evident why this distinction is clear while the distinction between the library catalog (Mirlyn) and Articles*Plus* tabs is less obvious.

The absence of article searches in the MLibrary search box, in particular, is quite interesting. From looking at search logs from the library catalog, it is known that users do seek articles in the library catalog. Additionally, search log analysis from the "Articles*Plus*" tab, which provides access to full-text online articles, shows that many users conduct article searches using that tool.[8] It seems that users are clear on the distinction between the first (MLibrary) and second (Articles*Plus*) tabs.

The question of causality is also of interest. Have users adapted to the way the Web site works, or is the Web site truly reflective of user research needs? These questions are not answerable from log analysis alone, but require more detailed individual interviews and focused research. Additionally, the sample method (every 378 queries) precluded any in-depth study of user behavior at the individual level. This study does not explore how particular users might have

---

[8] For the first fourth months of 2011 (January-April), during which time Articles*Plus* was still a relatively new service (launched in September 2010), 22.5% of searches (129,235 queries) on the library Web site were conducted using the Articles*Plus* tab. In comparison, for the same period of time in 2012, 41% of searches (282,780 queries) were conducted using the Articles*Plus* tab.

refined their queries in one user session, nor does it track what resources were clicked on after a search query was entered.

Likewise, the study involved assumptions about user intentions on the library Web site, but did not touch on the ultimate research goal of the individual user. Was the user's query the goal in itself, or was it a means to find the best place to conduct the actual research? Further research into the user's actions from a more holistic approach is needed to better understand users' goals, and how those relate to search.

Additionally, this process should be replicated at other research libraries to better understand the relationship between the findings presented in this study and the bento-box style search results employed on many libraries' Web sites. Do users of library Web sites that "pre-facet" their searches—asking users to identify the kind of resource they wish to search (databases, journals, books) before conducting the search—have similar categorizations of search kinds?

Conclusions

The University of Michigan Library set out to validate the search-centered interface for its Web site design through an analysis of the searches conducted using the search interface. There was concern that the search interface was not meeting the needs of the Web site's users. In general, the search-centric interface to the Web site was shown to match the kinds of searches that library users were conducting.

This study was one part of a series of investigations into various functions and interfaces provided through the Web site. This study, combined with other user studies, resulted in design changes to the search results screens to make them better match observed behavior. Specifically, the headings for the various search results sections were made more prominent to help users find

the most relevant section of results. Because of the relatively lower frequency of queries that seemed related to content created by (rather than resources provided by) the library, staff directory entries and institutional repository items were moved out of the main results. Fewer results were displayed in each section—with a more prominent "see all" link—than in the original design, largely because the most relevant results for the searches conducted in the course of the study appear in the first few results under each section. Not only did the study show that relevant results were appearing, it helped convince the researchers that the most relevant results were appearing higher in the list. This somewhat serendipitous discovery was not part of the original research plan.

The results of this study show that queries categorized as known-item searches comprised nearly a majority (44%, 530 searches) of the total. A fifth (20%, 239 searches) were categorized as "exploratory." The Web site's search interface is presented first in the standard navigation, followed by browse. This hierarchy broadly reflects the way Web site visitors conduct their searches, with a preponderance of interactions focused on known items rather than more abstract topics. At the same time, the number of exploratory searches is not trivial and reinforces the need to maintain a breadth of resources and multiple paths to them. This will benefit a significant portion of our Web site visitors.

Through an exploration of the "long tail" of search queries on the library website, we have gained a much better understanding of the kinds of searches our users are conducting. Known item searches are the plurality of search queries conducted on the site, meaning that relevance-based search technologies may be better suited to most searches than those focused on retrieval. At the same time, a significant minority of searches are exploratory, meaning that alternative approaches need to be offered. The broad-based, pre-faceted search results on the site,

together with links to appropriate exploratory tools, are effective tools for getting users to the resources they seek.

Acknowledgements

The authors wish to acknowledge the contributions of several individuals who aided in the study and analysis of the data. Giselle Kolenic at the University of Michigan's Center for Statistical Consultation and Research (CSCAR) helped us with devising our random sample methodology and in analyzing the results. Saurabh Koparkar helped with the categorization of the search queries. Justin Joque provided assistance with creating random samples using Excel and sorting results. Tom Burton-West helped with identifying resources to consider in the course of our literature review. Our thanks also go to Bill Dueber for his in-depth comments on an earlier draft.

References

Broder, Andrei. 2002. "A Taxonomy of Web Search." *SIGIR Forum* 36 (2): 3-10.
http://www.sigir.org/forum/F2002/broder.pdf.

Chau, M., Xiao Fang, and O. R. Liu Sheng. 2005. "Analysis of the Query Logs of a Web Site
Search Engine." *Journal of the American Society for Information Science and Technology*
56 (13): 1363-76. doi:10.1002/asi.20210.

Gabrilovich, Evgeniy, Andrei Broder, Marcus Fontoura, Amruta Joshi, Vanja Josifovski, Lance
Riedel, and Tong Zhang. 2009. "Classifying Search Queries using the Web as a Source of
Knowledge." *ACM Transactions on the Web* 3 (2): 5:1-5:28. doi:10.1145/1513876.1513877.

Hollink, Vera, Theodora Tsikrika, and Arjen P. de Vries. 2011. "Semantic Search Log Analysis:
A Method and a Study on Professional Image Search." *Journal of the American Society for
Information Science and Technology* 62 (4): 691-713. doi:10.1002/asi.21484.

Jones, Steve, Sally Jo Cunningham, Rodger McNab, and Stefan Boddie. 2000. "A Transaction
Log Analysis of a Digital Library." *International Journal on Digital Libraries* 3 (2): 152-69.
doi:10.1007/s007999900022.

Little, Geoffrey. 2012. "Where are You Going, Where have You Been? The Evolution of the
Academic Library Web Site." *The Journal of Academic Librarianship* 38 (2): 123-125.

Lown, Cory, Tito Sierra, and Josh Boyer. "How Users Search the Library from a Single Search
Box." *College & Research Libraries*. Published electronically December 13, 2011.
http://crl.acrl.org/content/early/2012/01/09/crl-321.

Madle, Gemma, Patty Kostkova, Jane Mani-Saada, and Anjana Roy. 2006. "Lessons Learned
from Evaluation of the Use of the National Electronic Library of Infection." *Health
Informatics Journal* 12 (2): 137-151. doi:10.1177/1460458206063809.

Preissle, Judith. "Analytic Induction." In *The SAGE Encyclopedia of Qualitative Research Methods*. SAGE Publications, Inc. 2008.

Song, Ruihua, Luo Zhenxiao, Nie Jian-Yun, Yu Yong, and Hon Hsiao-Wuen. 2009. "Identification of Ambiguous Queries in Web Search." *Information Processing & Management* 45 (2): 216-229. doi:10.1016/j.ipm.2008.09.005.

Swanson, Troy A., and Jeremy Green. 2011. "Why we are Not Google: Lessons from a Library Web Site Usability Study." *The Journal of Academic Librarianship* 37 (3): 222-229.