

# Measuring Query Complexity in Web-Scale Discovery

## *A Comparison between Two Academic Libraries*

**Rachael A. Cohen and Angie Thorpe Pusnik**

**Rachael A. Cohen** (*rachcohe@indiana.edu*) is Discovery User Experience Librarian, Indiana University Bloomington.

**Angie Thorpe Pusnik** (*atthorpe@iuk.edu*) is Digital User Experience Librarian, Indiana University Kokomo.

*This study reports on the examination of search transaction logs from web-scale discovery tools at two Indiana University campuses. The authors discuss how they gathered search queries from transaction logs, categorized queries according to the Library of Congress Classification schedule, and then examined queries using text analysis tools in order to identify which subjects were being searched and whether users were using advanced search options. The results of this investigation demonstrate how transaction logs may be used to communicate user interactions within discovery services. The findings offer detailed insight into the subjects and skills that teaching faculty and librarians should communicate to improve information literacy instruction. The search queries also uncover information needs that provide direction for collection managers.*

**T**o improve user experiences, libraries continuously seek methods to better understand how users interact with their services. Reviewing actual user interactions—such as chat or e-mail transcripts, online resource usage reports, and search transaction logs—provides the opportunity to identify recurrent themes among resources used, topics of inquiry, and potential research obstacles. Search transaction logs are a particularly attractive dataset for analysis

due to their comprehensive nature, as well as their ability to reveal both users' information needs and trends in search behaviors.

This study reports on the examination of search transaction logs from two web-scale discovery tools at two academic libraries. Libraries have increasingly adopted web-scale discovery tools over the past several years, and many libraries have implemented these systems as the first line of approach on their websites.<sup>1</sup> This prime placement invites usage from all types of users and results in a rich dataset that spans user disciplines and demographics. Analysis of discovery tool transaction logs is a choice assessment strategy because it is anonymous, non-intrusive, and comprehensive. In this paper, the authors discuss how they gathered and classified search queries from transaction logs and then used text analysis tools to identify which subjects were being searched, as well as the complexity of users' searches. The search transaction logs allowed the authors to develop a more captivating message for teaching faculty and librarians regarding the direct ties between discovery tool usage and assigned coursework. This messaging will help deepen campus partnerships to improve users' information literacy skills.

## LITERATURE REVIEW

### Web-Scale Discovery Assessment

Within the last decade, several of the major library database and system providers began offering index-based discovery services that offer users a variety of options for searching and retrieving materials from library collections.<sup>2</sup> Assessment of these services began as soon as libraries started to consider which discovery service to purchase. Among other product features, librarians investigated whether users took advantage of post-search filter options, whether discovery products offered exact and advanced search options, and where users succeeded and fell short during usability tasks.<sup>3</sup> In a nutshell, these studies “kicked the tires” of the various discovery service options on the market, revealing their strengths, weaknesses, and areas of opportunity, which have since driven product development.

In addition to providing centralized indexes for library collections, discovery services often come with significant price tags. In order to continue to justify these continuing expenses over time, libraries have sought ways to show the value and utility of these tools. Such evaluations have extended beyond usability tests to the ways in which discovery services impact library services such as instruction and reference. Cmor and Li discussed how their library’s adoption of a discovery service propelled them to realign their instruction course plans with their institution’s updated learning outcomes policy. Doing so allowed the library to concentrate more on teaching students how to understand and evaluate information rather than simply navigate interfaces.<sup>4</sup> Similarly, after reviewing student feedback on information literacy sessions, Debonis, O’Donnell, and Thomes realized that their discovery service needed to be incorporated into library instruction exercises.<sup>5</sup> The authors revised both their instruction and reference protocols in order to inspire students to carefully consider their sources throughout the search process. Integrating the discovery service into additional library services increases the likelihood that users will know when and how to use this tool. Feedback surveys such as the one Debonis, O’Donnell, and Thomes implemented are certainly useful for identifying ways to improve library services. However, they are not comprehensive because they only capture responses from participants, which is most likely not the library’s entire user population. To fully understand search behaviors in a discovery service, other data sources are necessary.

### Transaction Log Analyses

One such far-reaching data source is a transaction log. A variety of industries have adopted transaction log analyses in order to better understand user behavior on websites.<sup>6</sup> A transaction log analysis is an examination of electronic records—that is, transactions—of interaction that have occurred during searching episodes between a web search

engine and users searching for information on that web search engine.<sup>7</sup> Transaction log analyses are usually used to evaluate system performance, system architecture, or searcher actions. Spink and Jansen documented numerous analyses from the e-commerce, medical, and adult entertainment sectors until 2004 in their book.<sup>8</sup>

Although library users are also likely web searchers, their behaviors across channels may not be identical. A primary conclusion from Jansen et al.’s study of *Excite* search queries was that, in fact, “web search users seem to differ significantly from users of traditional IR [Information Retrieval] systems.”<sup>9</sup> Included among these “traditional” systems are library OPACs. Recognizing the wealth of data stored within their systems, and the fact that users may interact differently with their systems than with commercial options, libraries have also adopted transaction log analyses. The majority of these analyses have focused on OPACs, and the overarching goals have been to improve library systems.<sup>10</sup> Common points of observation among these studies include query length, type of search option (e.g., basic vs. advanced), typographical errors, and use of Boolean operators.<sup>11</sup> A study by Villén-Rueda, Senso, and de Moya-Anegón diverged from others by investigating the distribution of subject queries across major areas of knowledge, such as experimental sciences, health sciences, or engineering.<sup>12</sup> As library technologies have evolved over time, transaction log analyses have moved to other online library resources, including digital libraries, federated search tools, and websites.<sup>13</sup>

Very few transaction log analyses have been performed for web-scale discovery systems, though. Meadow and Meadow evaluated nine hundred search queries from Summon and categorized each query into one of seven types, such as URL, natural language, or known item, prior to further analysis.<sup>14</sup> This study broke new ground by applying transaction log analysis to a discovery system, which Brett, German, and Young carried into their assessment of the tabbed-search interface on their library website.<sup>15</sup> These studies categorized queries into broad groups, such as “Database/Journal,” “Subject,” and “Known Item.” They did not, however, delve into the specific subjects users searched. General suggestions may be drawn from these results, but curricular themes and distribution of searches across schools or departments remain unknown.

### Collection Relevance

Discovery assessment at the transaction log level brings to light the information needs of many library users. This type of analysis is particularly important due to staffing and time limitations that may prevent librarians from having complete knowledge of the entire curriculum taught at their institutions. Transaction logs also reveal the materials for which users are actually looking, which indicates this is a prime data source for collection development purposes, too. This is an understudied area within the literature. Libraries have measured the impact of discovery service implementation

on library collection use, with mixed results. Kemp found that, following the implementation of Summon, print circulations and link resolver activity increased while database, e-journal lookup, and OPAC searches decreased.<sup>16</sup> Looking at a different discovery service, Calvert alternatively observed a decrease in print circulations after her library implemented EBSCO Discovery Service (EDS).<sup>17</sup> She did, however, notice an increase in abstracts viewed and full text articles retrieved in several of her library's subscription EBSCOhost databases. These studies focused on evaluating possible effects of discovery service adoption on the library's existing collection, but they did not use discovery service usage reports to appraise or improve the collection.

In 2016, Siegel noted that "there still appears to be very little currently written on the topic of utilizing a discovery service's search query data in order to discover holes within a particular library's collection."<sup>18</sup> He began to fill this gap by using the top fifty unique queries from a year's worth of discovery transactions at a Virginia academic library in order to prove search queries may be used for collection development purposes. After categorizing the queries according to specific disciplines, he repeated user search queries in order to identify low search result terms, which signified potential gaps in the library's collection. This study thus illustrated that search queries may be used as a collection development tool.

The current study continues to fill the gap identified by Siegel. It also advances discovery service assessment by adding to the limited transaction log analyses that have occurred to date. For libraries, transaction logs may help answer questions regarding how users search for materials, whether they take advantage of advanced search options (e.g., Advanced Search, Boolean operators, and field codes), and which subject areas are more or less frequently searched. The answers to these questions impact pedagogy—particularly with regard to learning activity design, sequencing, and evaluation—and collection development.

---

## METHOD

### Data Collection

During the fall 2015 semester, two librarians at two Indiana University (IU) campuses initiated a semester-long research project to examine user search terms from EDS, the discovery tool used at both campuses. The authors sought answers to two research questions:

1. What queries and/or themes recur at each institution?
2. What are the similarities and differences between recurrent queries across the two campuses?

To answer these questions, the authors analyzed text data recorded within each campus's EDS search transaction logs from the fall 2015 semester (August 24–December 18). These anonymous transaction logs were harvested from Google

Analytics. The authors were interested in the second research question because, although they are both IU librarians, their campuses greatly differ in size and areas of study. The first school in this study, Indiana University Bloomington (IUB), is the flagship campus of the IU system, offers degrees in more than two hundred majors, has a Carnegie classification of "Doctoral University: Highest Research Activity," and had an FTE of 41,165 during the 2015–2016 academic year.<sup>19</sup> The second school in the study, Indiana University Kokomo (IUK), is the smallest regional campus of the IU system. It offers degrees in more than thirty majors, has a Carnegie classification of "Baccalaureate College: Diverse Fields," and had an FTE of 2,676 during the 2015–2016 academic year.<sup>20</sup> IUB implemented EDS in August 2011, and IUK launched EDS in September 2011. Both IUB and IUK upgraded to the EBSCO Google Analytics—Advanced tracking code within their respective EDS platforms in summer 2015 to gather search data for fall 2015. The EBSCO Google Analytics—Advanced tracking code cleanly harvests search terms to Excel, with little to no cleanup needed. The search query logs did not contain any personally identifiable information for human subjects, so Institutional Review Board (IRB) research approval was not required for this project.

In spring 2016, the authors exported the first 18,000 EDS search queries for the fall 2015 semester from each of their Google Analytics accounts. To reduce the datasets to a manageable quantity of queries, the authors calculated a random sample by setting an error rate of 3 percent and a confidence rate of 99 percent for the query population of 18,000. These parameters produced a sample size of 1,677 queries per campus. The authors used the first 18,000 queries because it was the closest major interval to IUK's total EDS searches (18,555) for the fall 2015 semester. IUB recorded 122,607 total queries during this time period.

### Search Queries

Next, the authors reviewed each of the 1,677 search queries and assigned up to two classes and subclasses using the Library of Congress Classification schedule (<http://www.loc.gov/catdir/cpsolcco/>). For example, the search query "(folklore) AND (death)" was assigned the class of "G - Geography, Anthropology, Recreation" and subclass of "GR: Folklore." To address interrater reliability, the authors used IUCAT, IU's shared OPAC, to search for queries and review call numbers. The authors also each reviewed ten identical, initial queries, assigned classes and subclasses, and then discussed their results. This discussion produced changes in the review methodology, which the authors then applied to an additional ten queries. Following the review of the second ten queries, the authors finalized their procedures, which are captured in images 1–4.

The procedures also included:

1. Do NOT categorize:
  - a. Database names (e.g., Academic Search Premier)

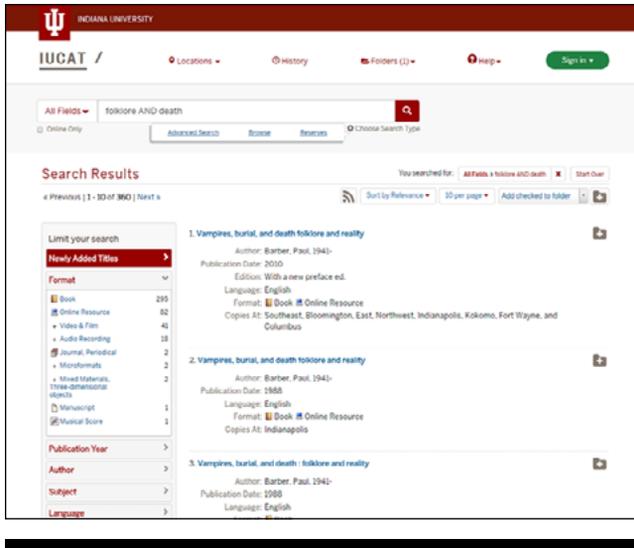


Image 1. IUCAT search results for “folklore AND death”

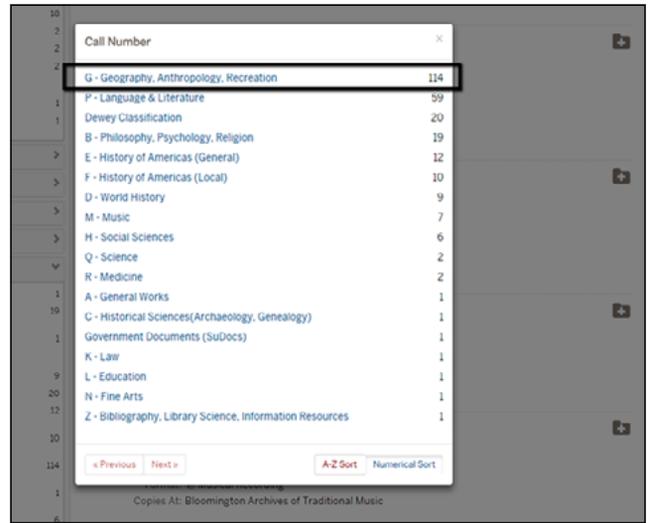


Image 3. Call number class selection in IUCAT

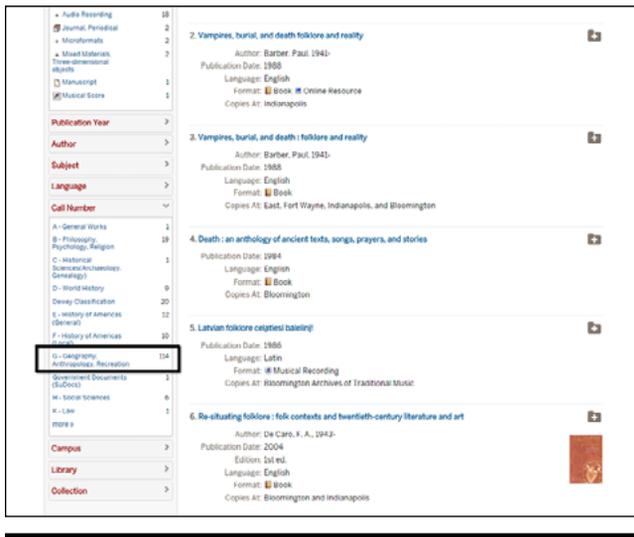


Image 2. Call number facet in IUCAT

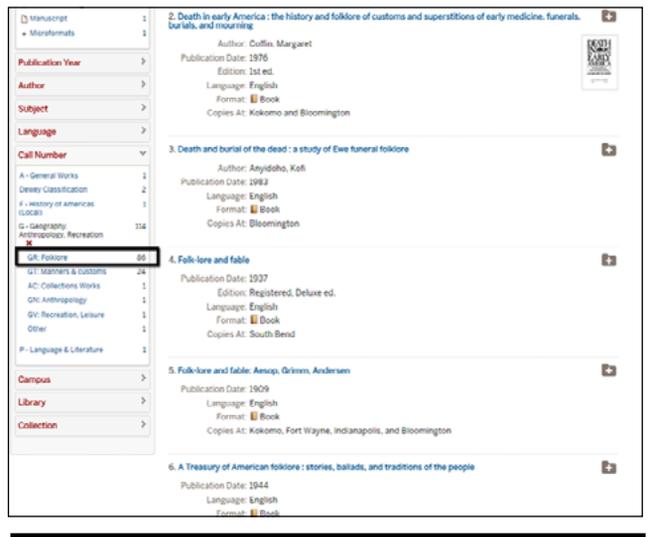


Image 4. Subclass identification in IUCAT

- b. Journal name/title
- c. Source types lacking additional context (e.g., “Literature review”)
- 2. DO categorize:
  - a. Article titles (i.e., the main subject of the article)
  - b. The primary subject area, but OK to use two if necessary
  - c. If a search is ambiguous, use the top two classifications.

Databases, journal titles, and source types were not assigned classes because one or even two specific classes could not necessarily be defined from the name alone. For example, Academic Search Premier is a database that contains content on hundreds of subjects, and thus it was impossible to assign only one or two classes to it. A journal title search query was *Journal of Purdue Undergraduate Research*. This journal publishes undergraduate research

across a variety of disciplines, so this, too, was an unclassifiable query.

However, article titles, as well as chapters and essays, known titles (i.e., books), and keywords were assigned classes. The authors aimed to identify the central subject of each keyword and article search query. When the authors were unsure of the subject or there were zero results in IUCAT, they would use a general web search engine to determine the main topic. For example, a search for “Richard M Kavuma,” a Ugandan journalist and the editor of several newspapers, returned zero results in IUCAT, but a general web search provided insights into Kavuma’s identity, which allowed the authors to categorize this search query. Two classes were applied when a search query equally fit two subject areas. For example, queries relating to espionage were assigned to both political science and military science classes because both classes could yield relevant results on this topic.

## FEATURE

**Table 1.** Search query notation in EDS transaction logs

Query	How EDS records query in transaction logs
american disabilities act	(american+AND+disabilities+AND+act)
“poverty”	poverty
[Advanced Search] deforestation AND zoos	(deforestation)+AND+(zoos)

In addition to categorizing each search transaction, every query was tagged as either a Basic Search or an Advanced Search and was marked if a field code was used in conjunction with the advanced search. EDS distinguishes Advanced Search queries from Basic Search queries in transaction logs by using the all capital letters “AND” between two sets of parentheses. Table 1 presents examples of how different types of search queries are recorded within EDS transaction logs.

EDS offers eight field codes as Advanced Search options to improve the precision of user searches. These field codes are:

- TX – All Text
- AU – Author
- TI – Title
- SU – Subject Terms
- SO – Source
- AB – Abstract
- IS – ISSN
- IB – ISBN<sup>21</sup>

The authors also denoted frequent queries to identify potential search query patterns. Topics with more than ten search queries were grouped together as “popular queries.” These query groupings reflected not only repeated searches on the same or similar queries but also frequent searches on books written by certain authors or thematic queries, such as myths.

## RESULTS

### Query Subjects

Social Sciences was the most common class for both IUB and IUK search queries. Figure 1 shows the complete distribution of Library of Congress Classification totals for the first class each query was assigned at both campuses. At IUB, 473 searches (30.3%) were classified as Social Sciences, and IUK recorded 422 searches (26.9%) within this class. Social Sciences was the only class in which at least one search query was recorded for every subclass. Figure 2 is a visualization of the combined distribution of Social Science queries across the subclasses for both campuses. The thickness of the line corresponds with the number of queries identified for each subclass: The thicker the line, the greater the number of pertinent queries in the transaction logs. Across the two campuses, HV, HQ, and HD were the top subclasses, whereas

HS, HA, and HX each recorded only minimal search queries.

### Query Complexity

By default, both IUB and IUK direct users to the Basic Search option in EDS. Thus, usage data for Advanced Search options reflects a measured

choice from users: they took a specific action to perform an Advanced Search, rather than rely on the Basic Search. Figure 3 shows the breakdown of Basic, Advanced, and Advanced + field code searches at both campuses. IUB users performed nearly 3 times as many Advanced Searches as did IUK users and approximately 2.5 times as many Advanced Searches with field codes. Although queries that use field codes are automatically considered Advanced Searches, since users must be on the Advanced Search screen to view the field code drop-down menu, the majority of Advanced Searches at both campuses did not include any field codes. Figure 4 presents the distribution of field code use at the campuses. IUB searchers used the author (AU) and journal name (JN) field codes more often than did IUK searchers. IUK users, on the other hand, selected the title (TI) field code much more often than did IUB users. Journal and book identifiers (ISBN and ISSN), however, were not used at all at either campus. It is also important to note that two field codes—JN and DE—are not found in the dropdown “Select a field” menu on the EDS Advanced Search page. These are valid and functional field codes, but their use indicates these searches were conducted by expert users with special knowledge. The authors theorize these searches indicate librarian usage.

At both campuses, users searched for known titles, articles, and authors without using field codes. Article titles were searched 77 times (4.6%) at IUB and 136 times (8.1%) at IUK. Known titles were searched less, with 22 such searches (1.3%) recorded at IUB and 65 (3.9%) at IUK. Author searches without the AU field code occurred least often: 13 (0.8%) users searched for authors at IUB, and only 3 (0.2%) searched at IUK. Although these query counts are not strikingly high, they do reveal an opportunity for more user instruction on how to use field codes to search more precisely.

### Query Missteps

Other transaction log analyses have reported on search failures, which have been defined as searches that result in zero hits.<sup>22</sup> After reviewing each search query within the two random samples, the authors decided to focus on two specific aspects of query failure analysis: typographical errors and questions. The authors chose these elements over others—such as query length and type of search (e.g., author, title, or subject)—because they hypothesized these would be the most common user miscues. Both issues may also be

relatively easily addressed through system features, such as spellcheck, and library instruction sessions. Typographical errors were divided into four categories: addition (e.g., serveral rather than several); deletion (e.g., elderly – elderlry); substitution (e.g., mighnt – midnight); and inversion (e.g., presenec – presence).<sup>23</sup> The authors also identified queries that contained more than one category of typographical error and those that were gibberish (e.g., sdf). Figure 5 shows that IUK users searched for more queries containing typographical errors than did IUB users. Deletions were the most common error at IUK, and deletions and substitutions tied for frequency at IUB. Overall, though, less than 8 percent of IUK user search queries contained typographical errors, and less than 2 percent of IUB queries contained these errors.

Past studies have explored whether users enter search queries in question format into commercial search engines.<sup>24</sup> Libraries are in the business of answering questions, so reviewing the types of questions users enter into a discovery service helps librarians better understand their users’ needs. The authors searched transaction logs for twelve question starters, shown in figure 6. User behavior differed between the two campuses. On the whole, IUB users searched for fewer questions, and the most frequent questions included “is” or “do.” IUK users conducted more searches that contained questions; “how” and “what” questions recorded the highest totals. Still, at

IUB, less than 2 percent of all search queries included questions, and this percentage only rose to 3.6 percent at IUK.

### Popular Queries

Finally, the authors examined transaction logs for repeated queries on topics. A search query with more than ten separate searches was labeled a “popular query.” Additionally, the text analysis software R was used to identify the ten most popular individual keywords across both query datasets. These terms, shown in figure 7, allow for the identification

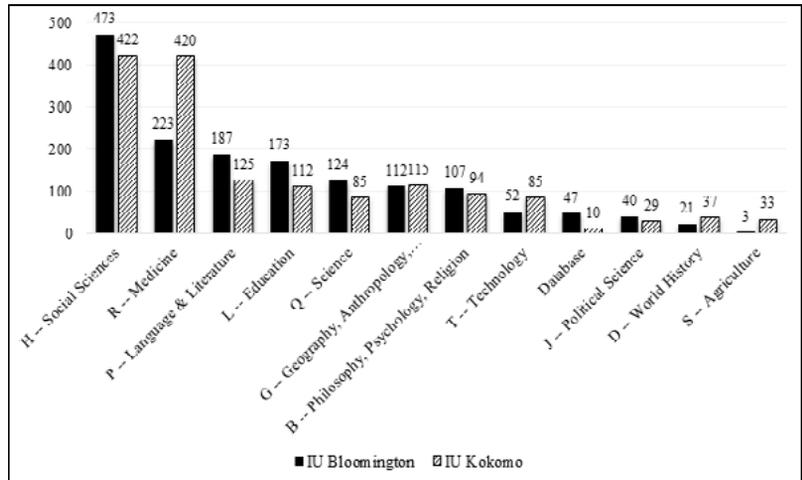


Figure 1. Complete distribution of LC Classification totals for both campuses

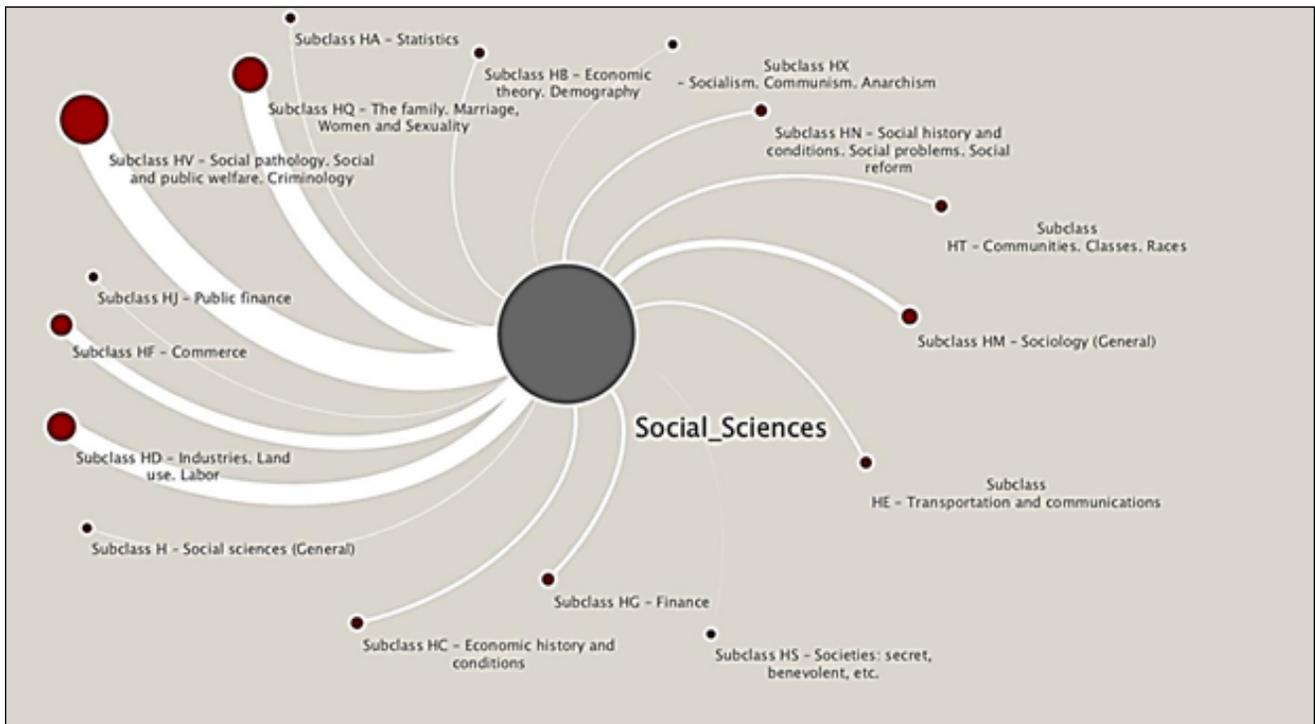


Figure 2. Combined distribution of Social Science queries across the subclasses

## FEATURE

of broad topical patterns, such as the use of “Education” at IUB and “health” at IUK. The popularity of these terms matches some expectations while disrupting others, based on enrollment figures for specific majors.

The value of this dataset lies in being able to not only dig in to common terms but also common themes. Themes may not show up in search query frequency reports because they require additional knowledge of how search terms may be connected within an institution’s curriculum. Tables 2 and 3 present emergent themes from both IUB and IUK. The authors identified two tendencies for popular themes: (1) frequent searches may take the form of phrases, such as known items, that could reflect course-adopted text(s) for particular classes and/or instructors; and (2) frequent searches may reveal the exploration of certain subjects, such as “folklore + (fill-in-the-blank)” or combinations of searches regarding medical ethics. IUB identified eighteen recognizable popular themes while IUK observed eleven themes. Besides the recurrence of these queries, their variations suggest more disparate interactions than a single user simply modifying his or her search over and over again.

## DISCUSSION

Transaction log analyses provide granular evidence of how, and for what, users search library resources. This study found that the majority of EDS searches related to social science and medical topics. This was of particular interest for IUB because some of the social science liaisons deliberately steer students away from EDS in favor of subject databases. The results of this study suggest that information literacy instructors at both IUB and IUK should add or further expand upon EDS search strategies during their instruction sessions.

The need for further EDS instruction does not stem only from the quantity of searches performed. The quality of the searches also illustrates this necessity. Natural language queries towered above other query types, and they were subject-driven. For example, less than half of IUB users conducted an Advanced Search, and fewer than 20 percent of IUK users took advantage of this search option. Since the search logs at both campuses indicated users performed searches for known items, such as books and articles, a simple guide to EDS field codes may help users obtain more precise results

while simultaneously introducing them to more advanced search options. The lack of queries in the form of questions suggests the majority of EDS users are comfortable searching just by keyword. This is encouraging for instruction

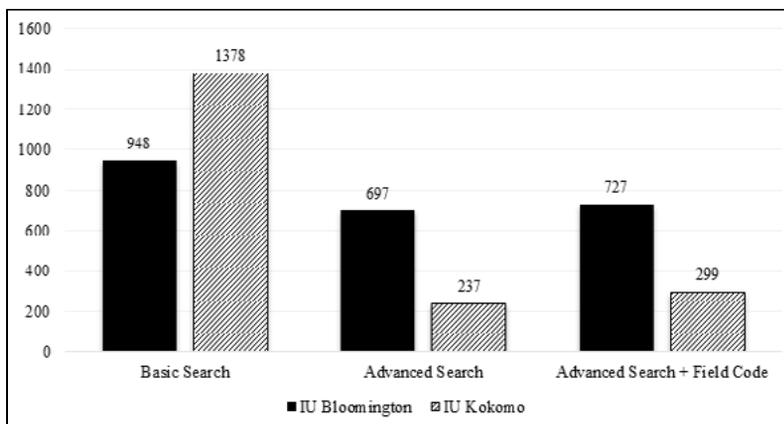


Figure 3. Basic vs. Advanced vs. Advanced + field code searches breakdown

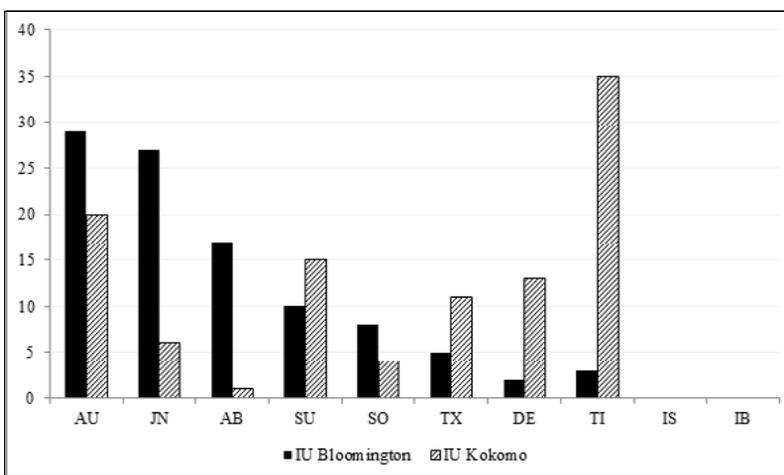


Figure 4. Distribution of field codes

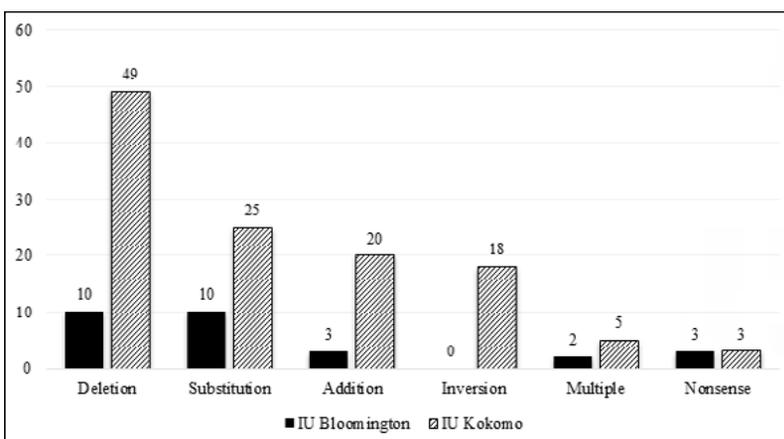


Figure 5. Distribution of typographical errors

because it implies users liken EDS to commercial search engines, not question-and-answer services.<sup>25</sup> Instruction can thus begin by building on existing consumer search skills and adapting them to library resources, rather than starting from scratch.

In addition to generating ideas for specific strategies to teach in information literacy sessions, the results of this study also divulge aspects of institutional curricula. The popular queries and themes that emerged in each campus's transaction logs reveal opportunities for instruction in specific courses, perhaps where such needs were previously unknown. Transaction log analysis may, then, provide a new rationale for teaching discovery where this previously did not occur. Libraries have experienced more or less success transitioning to discovery service instruction. Buck and Mellinger distributed an online survey to institutions that had implemented the web-scale discovery service Summon. Fifty-eight percent of respondents reported they felt the implementation of Summon had changed their instruction, with the most frequent change being how much class time was spent emphasizing which database to choose. Respondents indicated they were able to spend more time on topics such as refining search terms, research as an iterative process, or higher-level search skills.<sup>26</sup> This is a critical shift in libraries. Rather than spending precious time teaching which resource to search, librarians may now concentrate on teaching how to search resources. Sharing user search queries enables this type of higher level learning: by understanding authentic user search queries, librarians and teaching faculty may develop targeted strategies to hone students' existing search skills, rather than beginning with sifting through a list of dozens of databases. New partnerships may form as colleagues—including teaching faculty, instruction librarians, and other public service librarians—discuss the search skills students already exhibit and how those can be developed to more advanced levels.

Finally, the implications of these results do not apply only to instruction and public services. The popular queries and themes also have obvious implications for collection development: If a library does not own titles that are frequently searched, acquisition of the material should be considered. If the title is already owned, perhaps a librarian should coordinate with the pertinent faculty member to place the item on print or electronic reserve, if possible. Frequent themes may also reveal changing disciplinary focal points. These may be areas for the library to research—in consultation with both faculty and library service providers—for additional collection development.

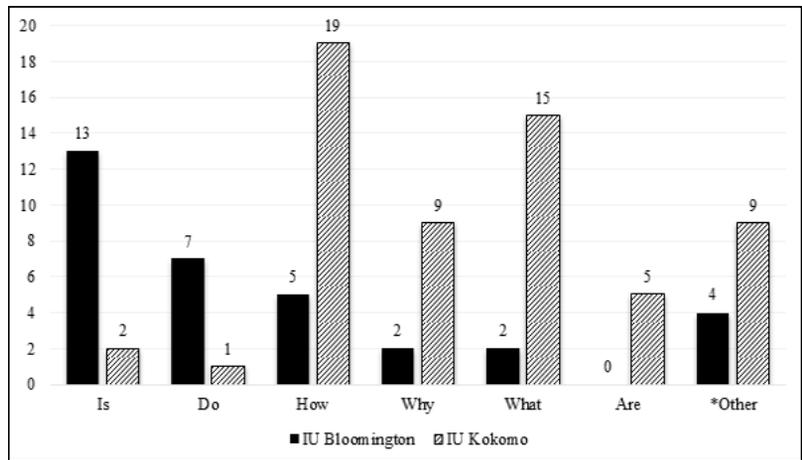


Figure 6. Distribution of question starters

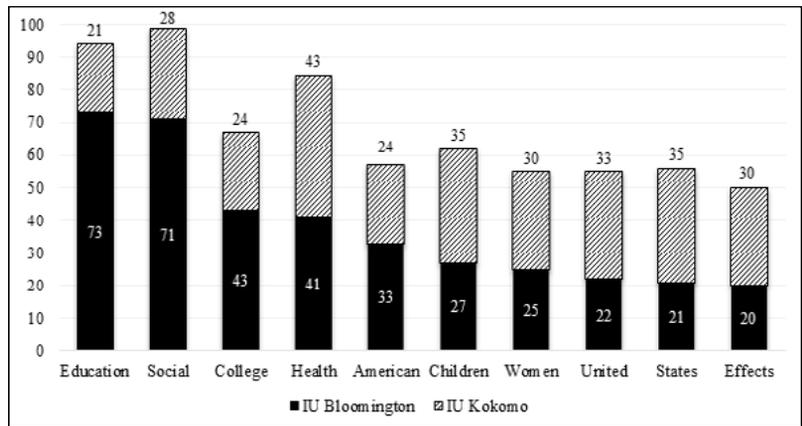


Figure 7. Top 10 search terms

### Limitations

This study acknowledges a few limitations, the first being that it was extremely time intensive. Each search query was individually evaluated and categorized by a human being; no such automatic or systematic mechanism for this process yet exists. Additionally, without the use of a call number facet in the IUCAT OPAC, categorizing each search query would have been virtually impossible, as the authors are not catalogers. Third, the analyzed search queries represented a random sample from one semester of web-scale discovery activity at each campus. The distribution of queries across LC classes may differ between fall and spring semesters or even academic years. These results, then, are not exhaustive, but they do reflect a snapshot from a particular time frame, and they may definitely still be used to open new discussions with instructors. Finally, LC classes do not necessarily neatly map to specific courses. Courses may address a variety of topics, and the authors' method of assigning up to two classes to each search query indicates the potential interdisciplinary nature of user searches. However, the popularity of

## FEATURE

**Table 2.** Bloomington popular query examples

Query	Example Search	Searches
Fashion	(fashion) AND (style) (fashion) AND (fashion trend)	38
Folklore	(folklore) AND (comics)	24
Their Eyes Were Watching God	("their eyes were watching god" AND mcgowan")	19
Psychology myths	(10% AND "of" AND our AND brains) (men AND are AND better AND "at" AND math AND than AND women)	21

**Table 3.** Kokomo popular query examples

Query	Example Search	Searches
Espionage	(war AND Spies) (espionage AND cases)	24
Feminism	(feminism AND fairy AND tale AND social AND norm)	19
James Bond	(007 AND british AND empire)	20
Types of intelligence	multiple intelligence	17

certain classes and subclasses over other subclasses indicates where additional library outreach may be most impactful.

### Next Steps

The results of discovery service transaction log analyses have persuasive implications for information literacy instruction. A cohesive approach to discovery instruction is necessary because disparate pedagogies may result in dissimilar student research skills.<sup>27</sup> Fawley and Krysak further state, "When integrated into lesson plans with learning outcomes that emphasize critical thinking skills, discovery tools offer a chance to teach evaluative techniques and higher-level refining skills that are transferrable across subject specific databases."<sup>28</sup> Transaction log evaluations are one way to achieve this goal: understanding who is and who is not using the discovery service, based on subjects and resources searched, informs librarians where lesson-plan integration of the discovery service should occur. Furthermore, search queries reveal where librarians' instruction efforts are succeeding and where additional effort should be invested.

Popular search queries should be discussed with teaching faculty and instruction librarians. At a minimum, librarians may reach out to the faculty whose students are using EDS, based on transaction logs. Transaction logs may also serve as internal discussion points among library employees. The logs answer questions about how students are searching, which can inform outreach strategies to departments with both high and low discovery adoption. These conversations are necessary in order to determine the extent to which the library is integrated into different academic departments and schools. If the library is not integrated into the curriculum, additional discussions should be held in order to determine why not, whether non-library resources are instead being

used, and how the library might better serve non-user subject areas. Transaction logs provide evidence of user behavior, but improving the success of user searches requires collaboration with instructors in- and outside of the library.

Transaction logs also suggest messaging strategies. At IU Kokomo, librarians have juxtaposed EDS with Google and Amazon to help frame the discovery service with students' existing mental models of familiar search engines. However, search queries such as database names and questions suggest that students may be applying this metaphor too literally. Instruction should clarify that discovery is designed for subject and keyword searching, and improvements are being made for known-item searching. Source types (e.g., "articles on" or "books about") or interrogative phrases (e.g., "what is" or "how to") are unnecessary, and the discovery service offers its own tools—such as the Source Type facet—to refine search results, which will likely produce more precise results than keywords. In addition to facets, librarians should teach and encourage the correct use of Advanced Search and field codes. Emphasizing these features will help students receive more relevant and useful results, as well as get more value out of the discovery service.

Public service librarians who do not teach information literacy sessions, such as reference and access services librarians, also stand to benefit from the results of this study. Identifying search themes enables reference librarians to prepare for probable reference questions. This may consist of simply familiarizing oneself with useful resources to answer likely questions, or it may extend to creating reference guides, print or online, for easy distribution to interested students. Similarly, identifying frequent known item searches allows access services librarians to recommend titles for course reserves. In consultation with colleagues, access services librarians may be able to compile a list of teaching faculty

to approach regarding course reserves prior to the start of a new semester.

It is important to note that the current study only examined transaction log data from a single fall semester at each campus. A future area for research, then, is to repeat the study, either for a spring semester, subsequent fall semester, or both. Repeating the study would allow for the identification of persistent or recurrent popular queries and themes. The researchers would also be able to evaluate changes in user behavior patterns, such as the use of advanced search options.

An additional direction for future research is to compare discovery transaction logs with those from other databases. If librarians and instructors teach disparate database pedagogies, it would be worthwhile to evaluate search queries from different resources to determine whether more sophisticated user behaviors are more prevalent in certain resources.

## CONCLUSION

The purpose of this study was to report the results of two transaction log analyses from two web-scale discovery tools at two academic libraries. The findings show that the subjects most frequently searched across both institutions stemmed from social science and medical fields. The majority of users conducted basic searches, and more users opted to conduct an Advanced Search without a field code than with one. These results underscore the need for additional instruction on higher-level search techniques.

This study encourages additional communication between librarians and teaching faculty. Search query data helps everyone involved with information literacy instruction to better understand how users are actually utilizing the discovery tool. From this shared understanding, librarians can collaborate with instructors to improve students' skills. The demonstration of refinement tools, Boolean operators, and overall scaffolding of information literacy at different stages in students' academic careers would likely improve students' keyword selection and subsequent search results.<sup>29</sup> This is win-win for the library and teaching faculty: a happy student searcher who finds what is needed for a project is more likely to succeed in coursework, as well as return to the library's resources for the next assignment.

Finally, the authors also experienced personal unexpected benefits that are worth noting. The process of categorizing search queries according to the Library of Congress Classification scheme improved the authors' knowledge of this system. This led to improvements in curricular awareness during reference shifts in successive semesters. The ability to recognize course themes from individual reference questions spurred deeper interactions with students regarding the nature of their assignments and who their instructors were so that the library might reach out to those teaching faculty for additional engagement opportunities. This study showed the extent to which discovery is used, and the results

encourage opening new discussions regarding how discovery may be taught for the most benefit.

## References and Notes

1. Marshall Breeding, "Looking Forward to the Next Generation of Discovery Services," *Computers in Libraries* 32, no. 2 (2012): 28–31, <https://librarytechnology.org/document/16731>.
2. Marshall Breeding, "Relationship with Discovery," *Library Technology Reports* 51, no. 4 (2015): 22–25, <https://journals.ala.org/index.php/ltr/article/view/5688/7067>.
3. Noah Brubaker, Susan Leach-Murray, and Sherri Parker, "Shapes in the Cloud: Finding the Right Discovery Layer," *Online* 35, no. 2 (2011): 20–26; Maria M. Pinkas et al., "Selecting and Implementing a Discovery Tool: The University of Maryland Health Sciences and Human Services Library Experience," *Journal of Electronic Resources in Medical Libraries* 11, no. 1 (2014): 1–12, <https://doi.org/10.1080/15424065.2013.876574>; Anita K. Foster and Jean B. MacDonald, "A Tale of Two Discoveries: Comparing the Usability of Summon and EBSCO Discovery Service," *Journal of Web Librarianship* 7, no. 1 (2013): 1–19, <https://doi.org/10.1080/019322909.2013.757936>.
4. Dianne Cmor and Xin Li, "Beyond Boolean, Towards Thinking: Discovery Systems and Information Literacy," *Library Management* 33, no. 8/9 (2012): 450–57, <https://doi.org/10.1108/01435121211279812>.
5. Rocco Debonis, Edward O'Donnell, and Cynthia Thomes, "(Self-)Discovery Service: Helping Students Help Themselves," *Journal of Library & Information Services in Distance Learning* 6, no. 3–4 (2012): 235–50, <https://doi.org/10.1080/1533290X.2012.705648>.
6. Thomas A. Peters, "The History and Development of Transaction Log Analysis," *Library Hi Tech* 11, no. 2 (1993): 41–66, <https://doi.org/10.1108/eb047884>.
7. Bernard J. Jansen, "Search Log Analysis: What It Is, What's Been Done, How to Do It," *Library and Information Science Research* 28, no. 3 (2006): 407–32, <https://doi.org/10.1016/j.lisr.2006.06.005>.
8. Amanda Spink and Bernard J. Jansen, *Web Search: Public Searching of the Web* (Dordrecht, Netherlands: Kluwer Academic, 2004).
9. Bernard J. Jansen, Amanda Spink, and Tefko Saracevic, "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," *Information Processing & Management* 36, no. 2 (2000): 207–27, [https://doi.org/10.1016/S0306-4573\(99\)00056-4](https://doi.org/10.1016/S0306-4573(99)00056-4).
10. Eng Pwey Lau and Dion Hoe-Lian Goh, "In Search of Query Patterns: A Case Study of a University OPAC," *Information Processing & Management* 42, no. 5 (2006): 1316–29, <https://doi.org/10.1016/j.ipm.2006.02.003>.
11. Thomas A. Peters, "When Smart People Fail: An Analysis of the Transaction Log of an Online Public Access Catalog," *Journal of Academic Librarianship* 15, no. 5 (1989): 267–73; Lau and Goh, "In Search of Query Patterns"; Helen Georgas, "Google vs. the Library (Part II): Student Search Patterns and Behaviors When Using Google and a Federated Search Tool," *portal: Libraries and the Academy* 14, no. 4 (2014): 503–32; Megan Dempsey and Alyssa M. Valenti, "Student Use of Keywords and Limiters in Web-Scale Discovery Searching," *Journal of Academic Librarianship* 42, no. 3 (2016): 200–206, <https://doi.org/10.1016/j.acalib.2016.03.002>; Hao-Ren Ke et al., "Exploring Behavior of E-journal Users in Science and Technology: Transaction Log Analysis of Elsevier's ScienceDirect OnSite in Taiwan," *Library & Information Science Research* 24, no. 3 (2002): 265–91, [https://doi.org/10.1016/S0740-8188\(02\)00126-3](https://doi.org/10.1016/S0740-8188(02)00126-3).
12. Luis Villén-Rueda, Jose A. Senso, and Félix de Moya-Anegón, "The Use of OPAC in a Large Academic Library: A Transactional Log Analysis Study of Subject Searching," *Journal of Academic Librarianship* 33, no. 3 (2007): 327–37, <https://doi.org/10.1016/j>

- .acalib.2007.01.018.
13. Steve Jones et al., "A Transaction Log Analysis of a Digital Library," *International Journal on Digital Libraries* 3, no. 2 (2000): 152–69, <https://doi.org/10.1007/s007999900022>; Susan Avery and Daniel G. Tracy, "Using Transaction Log Analysis to Assess Student Search Behavior in the Library Instruction Classroom," *Reference Services Review* 42, no. 2 (2014): 320–35, <https://doi.org/10.1108/RSR-08-2013-0044>; Stephen Asunka et al., "Understanding Academic Information Seeking Habits through Analysis of Web Server Log Files: The Case of the Teachers College Library Website," *Journal of Academic Librarianship* 35, no. 1 (2009): 33–45, <https://doi.org/10.1016/j.acalib.2008.10.019>.
  14. Kelly Meadow and James Meadow, "Search Query Quality and Web-Scale Discovery: A Qualitative and Quantitative Analysis," *College & Undergraduate Libraries* 19, no. 2–4 (2012): 163–75, <https://doi.org/10.1080/10691316.2012.693434>.
  15. Kelsey Brett, Elizabeth German, and Frederick Young, "Tabs and Tabulations: Results of a Transaction Log Analysis of a Tabbed-Search Interface," *Journal of Web Librarianship* 9, no. 1 (2015): 22–41, <https://doi.org/10.1080/19322909.2015.1004502>.
  16. Jan Kemp, "Does Web-Scale Discovery Make a Difference? Changes in Collection Use after Implementing Summon," in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, edited by Mary Pagliero Popp and Diane Dallis, (Hershey, PA: Information Science Reference, 2012), 456–68, <https://doi.org/10.4018/978-1-4666-1821-3.ch026>.
  17. Kristin Calvert, "Maximizing Academic Library Collections: Measuring Changes in Use Patterns Owing to EBSCO Discovery Service," *College & Research Libraries* 76, no. 1 (2015): 81–99, <https://doi.org/10.5860/crl.76.1.81>.
  18. Timothy Siegel, "Utilizing Discovery Service Queries for Collection Development Purposes," *Current Studies in Librarianship* 32, no. 2 (2016): 91–118.
  19. Indiana University, "Historical Enrollment, Hour and FTE: Bloomington: Fall 2006 through Fall 2015," Indiana University Fact Book, University Institutional Research and Reporting, Indiana University, 2015, [https://wwwiu.edu/~uirr/reports/standard/factbook/2015-16/Bloomington/Student\\_Data/Enrollment/Historical\\_Enrollment](https://wwwiu.edu/~uirr/reports/standard/factbook/2015-16/Bloomington/Student_Data/Enrollment/Historical_Enrollment); Indiana University, "Indiana University-Bloomington," The Carnegie Classification of Institutions of Higher Education, 2014, [http://carnegieclassifications.iu.edu/lookup/view\\_institution.php?unit\\_id=151351&start\\_page=lookup.php&clq=%7B%22first\\_letter%22%3A%22I%22%7D](http://carnegieclassifications.iu.edu/lookup/view_institution.php?unit_id=151351&start_page=lookup.php&clq=%7B%22first_letter%22%3A%22I%22%7D); Rachael A. Cohen and Angie Thorpe, "Discovering User Behavior: Applying Usage Statistics to Shape Frontline Services," *The Serials Librarian* 69, no. 1 (2015): 29–46, <https://doi.org/10.1080/0361526x.2015.1040194>.
  20. Indiana University, "Indiana University-Kokomo," The Carnegie Classification of Institutions of Higher Education, 2014, [http://carnegieclassifications.iu.edu/lookup/view\\_institution.php?unit\\_id=151333&start\\_page=lookup.php&clq=%7B%22first\\_letter%22%3A%22I%22%7D](http://carnegieclassifications.iu.edu/lookup/view_institution.php?unit_id=151333&start_page=lookup.php&clq=%7B%22first_letter%22%3A%22I%22%7D); Indiana University, "Historical Enrollment, Hour and FTE: Kokomo: Fall 2006 through Fall 2015," Indiana University Fact Book, University Institutional Research and Reporting, Indiana University, 2015, [https://wwwiu.edu/~uirr/reports/standard/factbook/2015-16/Kokomo/Student\\_Data/Enrollment/Historical\\_Enrollment](https://wwwiu.edu/~uirr/reports/standard/factbook/2015-16/Kokomo/Student_Data/Enrollment/Historical_Enrollment).
  21. EBSCO Information Services, "What Field Codes Are Available When Searching EBSCO Discovery Service (EDS)?" Support—EBSCO Help, accessed January 17, 2017, [http://support.epnet.com/knowledge\\_base/detail.php?id=3198](http://support.epnet.com/knowledge_base/detail.php?id=3198).
  22. Rhonda N. Hunter, "Successes and Failures of Patrons Searching the Online Catalog at a Large Academic Library: A Transaction Log Analysis," *Research Quarterly* 30, no. 3 (1991): 395–402, <http://www.jstor.org/stable/25828813>; Jones et al., "A Transaction Log Analysis of a Digital Library."
  23. Adam Brown, "A Singaporean Corpus of Misspellings: Analysis and Implications," *Journal of the Simplified Spelling Society* 3 (1988); Graeme Hirst and Alexander Budanitsky, "Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion," *Journal of Natural Language Engineering* 11, no. 1 (2005): 87–111, <https://doi.org/10.1017/S1351324904003560>; Andre-Roch Lecours, "Serial Order in Writing—A Study of Misspelled Words in 'Developmental Dysgraphia,'" *Neuropsychologia* 4, no. 3 (1966): 221–41, [https://doi.org/10.1016/0028-3932\(66\)90029-7](https://doi.org/10.1016/0028-3932(66)90029-7).
  24. Seda Ozmutlu, Huseyin C. Ozmutlu, and Amanda Spink, "Are People Asking Questions of General Web Search Engines?" *Online Information Review* 27, no. 6 (2003): 396–406, <https://doi.org/10.1108/14684520310510037>.
  25. Johannes Leveling, "A Comparative Analysis: QA Evaluation Questions versus Real-World Queries," paper presented at 2010 Workshop on Web Logs and Question Answering (WLQA 2010), May 22, 2010, Valletta, Malta, <http://doras.dcu.ie/16035/>.
  26. Stefanie Buck and Margaret Mellinger, "The Impact of Serials Solutions' Summon on Information Literacy Instruction: Librarian Perceptions," *Internet Reference Services Quarterly* 16, no. 4 (2011): 159–81, <https://doi.org/10.1080/10875301.2011.621864>.
  27. Nancy Fawley and Nikki Krysak, "Learning to Love Your Discovery Tool: Strategies for Integrating a Discovery Tool in Face-to-Face, Synchronous, and Asynchronous Instruction," *Public Services Quarterly* 10, no. 4 (2014): 283–301, <https://doi.org/10.1080/15228959.2014.961110>.
  28. Fawley and Krysak, "Learning to Love Your Discovery Tool."
  29. Yin-Leng Theng et al., "Scaffolding in Information Search: Effects on Less Experienced Searchers," *Journal of Librarianship and Information Science* 48, no. 2 (2016): 177–90, <https://doi.org/10.1177/0961000615595455>; Dempsey and Valenti, "Student Use of Keywords and Limiters," 205; Meadow and Meadow, "Search Query Quality and Web-Scale Discovery," 172.