Adam P Laubheimer. Just What I'm Looking For: The Prevalence of Automatic Query Suggestion Services in Academic Library Online Catalogs. A Master's Paper for the M.S. in L.S degree. March, 2012. 32 pages. Advisor: Ronald E. Bergquist

As a larger proportion of academic library users are familiar with web-based search engine functionality, there has been a great deal of interest among the library community in implementing advanced search features on online public access catalogs. This content analysis seeks to determine the prevalence of several next-generation catalog discovery features, with an emphasis on a relatively new technology, automatic query suggestion.

Headings:

Search algorithms

Web Search Engines

Federated Searching

Querying

JUST WHAT I'M LOOKING FOR: THE PREVALENCE OF AUTOMATIC QUERY SUGGESTION SERVICES IN ACADEMIC LIBRARY ONLINE CATALOGS

by
Adam Page Laubheimer

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

March 2012

Approved by

_____

Ronald E. Bergquist

Table of Contents

Introduction

"Because it is catalog data that has made collections accessible over time, to fail to define a strategic future for library catalogs places in jeopardy the legacy of the world's library collections themselves" (Calhoun, 2006, p.7)

It is a frequently stated perception (Calhoun, 2006) that traditional library online public access catalogs (OPACs) are challenging for novice users. Furthermore, the argument continues that for users trained to look for information on a web search engine paradigm, traditional library catalogs are archaic, clumsy and require too much library-specific expertise to employ effectively (Ballard, 2010). As a result, in the recent past, many research institutions have made significant investments (either financially or in systems deployment resources) in so-called next-generation catalogs. These next-generation catalogs prominently display many technologies and features that were once the exclusive domain of web search purveyors; technologies such as auto-suggestion of queries in realtime, faceted search refinements, federated database or article search integrated into the main search functionality, spelling correction, and relevance ranking have been slowly making an appearance in some of the more ambitious OPAC services.

Typically, these features are referred to as the *discovery layer*, indicating the separation from the metadata and full-text content subsystems. While traditional OPACs excel at known-item search, they are relatively poor systems for browsing or discovering new materials and resources; the traditional OPAC provides a relative paucity of

discovery tools for those not versed in the language of Library of Congress Subject Headings.

Jeff Wisniewski (2009) provided a list of commercial products and open-source solutions that offer some to all of these next-generation features. Some are integrated directly into part of the core integrated library system (ILS), some are third-party discovery layer tools. Additionally, auto-suggest functionality can be implemented via the JQuery Autocomplete Javascript library in conjunction with Apache Solr (Pennell and Sexton, 2010). This approach, though technical and not an off-the shelf implementation, has shown to be a useful addition to the UNC Library catalog (Pennell and Sexton, 2010).

The domain of this content analysis is focused on auto-suggestion (or query suggestion, or auto-complete) search, a feature commonly found in web search; the hallmark of auto-suggest search is that as the query is being entered, a list of other queries related to the one being entered are displayed below the search bar. While many of the other major next-generation catalog features listed above are more widespread currently, query autosuggestion is a relatively new feature offering. As auto-suggest is now a pervasive feature of web search, appearing in ubiquitous web services such as Netflix, Amazon, and Google, users are accustomed to the presence of such a feature on online searching systems. This content analysis seeks to determine the prevalence of auto-suggest search among academic library online catalogs, and to determine if there is a relationship between the presence of auto-suggest and any other next-generation catalog features.

Literature Review

As a greater preponderance of university library users have had years of informal information search training on popular web search services, user expectations have developed about what features should be present in robust, modern information systems. Today's younger users tend to prefer web search-style platforms for information discovery than the traditional library OPAC. In a study of younger academic library users, it was noted that

> Students usually prefer the global searching of Google to more sophisticated but more time-consuming searching provided by the library, where students must make separate searches of the online catalog and every database of potential interest, after first identifying which databases might be relevant. In addition, not all searches of library catalogues or databases yield full-text materials, and NetGen students want not just speedy answers, but full gratification of their information requests on the spot" (Lippincott, 2005, p.57)

OCLC's seminal 2005 report bears out this statement with data: when beginning their research, 89% of college students begin with web-based search engines, and only 2% start with the library catalog. Furthermore, 68% of these students remarked that Google was the search engine most commonly used for this purpose. In a stunning result from this study, 96% of college students agreed that Google provides worthwhile information, and only 84% agreed that library catalogs provide useful information (Calhoun, 2005).

Holman (2010) concurred in a later study, noting that the "Millennial" generation is overwhelmingly reliant on Google and similar search engines for fulfilling information needs; in a study of first-year student searching behaviors at the University of Baltimore conducted in 2008, 76% of students began an information seeking task with a Web search engine (rather than an online database or encyclopedia), and of this vast majority, 72% of them began with Google (Holman, 2010).

Wisniewski (2009) discussed why users prefer web-based tools for discovery, articulating the issue as a problem of convenience and perception:

> We (libraries) have quality. The search engines have convenience and ease of use. In short, we have competition and the competition is winning. Users need information, and they need it now, so if your decidedly rich information is locked inside a decidedly unfriendly discovery layer, users will continue to bypass the library and its tools to use something easier. (Wisniewski, 2009, p.54)

Data from Fast and Campbell's (2004) qualitative study corroborate this statement, pointing out that Western Ontario University student users consistently preferred Google to their university library's OPAC, and the most common comment indicated that this was based on the perception of lower investment of time and effort with web searching. Furthermore, the speed or time constraint of which the students spoke was not related to system speed or page load times, but to the number of steps from composing the query to accessing the resource (Fast and Campbell, 2004, p.140).

As today's college students so clearly prefer Google and other search engines to library OPACs, it is important to understand what sorts of information-seeking activities that these search engines support better than OPACs. Classically, OPACs have excelled at known-item search, providing a number of avenues to locating a specific, previously-

identified resource. Young and Yu (2004) undertook a study of Subject searches in a university OPAC using transaction log analysis (TLA), noting that "OPAC-interface design has been based on an assumption that users come to the catalog knowing what they need to know... Searchers are required to have knowledge of title, author, or subject" (p. 170).

Yet, in the same paper, Young and Yu (summarizing the findings of Jansen and Pooch, 2001) noted the crucial fact that

> [T]he majority of searchers on both OPACs and Web search engines use approximately two terms in a query, have an average of two queries per session, do not use complex query syntax, typically view no more than ten documents from the result list, and rarely use Boolean operators (Young and Yu, 2004, p. 171)

This research clearly indicates that search systems are increasingly used as platforms for information discovery based on simple keyword searching (Ballard, 2010), rather than simple known-item quests; attendant to this, users expect search systems to function as "decision engines" (Kutub, 2010), making subtle suggestions as to what the user may mean by an ambiguous query. Users expect, more and more, that the search services they encounter provide a more direct role in suggesting materials related to a broad query.

User desire to see broad keyword queries interpreted by subject is a major challenge for OPAC developers and libraries. Libraries and OPACs are heavily invested in controlled vocabularies for subject access, but users seek to search more semantically. As Calhoun et al (2009) discuss in their recent study of the variation between user

preferences and librarian's expectation of user preferences in OPAC design, users state that they want "more subject information" about resources.

> It is unlikely, given the relatively few unique subject-rich words contributed to a catalog description by controlled subject headings, that they mean more controlled subject headings. Given end-user survey respondents' top choices for catalog enhancement and what end-user focus group participants reported, "more subject information" is more likely to be interpreted as subject-rich data elements not generally included in a standard catalog description (Calhoun et al, 2009, p. 52).

Simplicity of use (or the perception of simplicity) is also a major factor in how patrons use an OPAC. Alison (2010) noted that, "one of the least popular features of the online public access catalog (OPAC) has been the ability to limit post-search. It appears most individuals are not willing to sort through complicated input forms and prefer a simple keyword search" (p. 376).

This user frustration with faceted limiting, though a popular and commonly-integrated next-generation feature, was demonstrated by Emanuel (2011), in a usability study of the VuFind open-source discovery layer at University of Illinois at Urbana-Champaign. VuFind mainly focuses on faceted searching, and the users in that study expressed near universal confusion with both the meaning and function of the facets presented.

Novotny (2004) found in a student OPAC search study that users "typed in broad keyword searches and expected that the 'computer' would interpret their search and process the results" (p. 531). Students having extensive familiarity with the web searching paradigm had difficulty understanding the conventions of traditional OPAC searching, and had particular problems with evaluating the relevance or usefulness of

results on relatively unstructured results listings. Particularly for more novice OPAC users, impatience ruled the day. "I don't think, I click" (Novotny, 2004, p. 530) is the infamous user quip from this study, indicating a desire to have the system present relevant, rational results quickly to a broad query. The desire to not have to evaluate the results, but simply to immediately select a highly-ranked resource, was displayed by most young users.

In addition, it was noted that users tended not to look at extensive results, but would limit their resource examination to the highest-ranked items (Novotny, 2004, p. 530). Novotny's study is a classic example of the need for relevancy ranking in OPAC results.

This was corroborated by another study, stating that

> [L]ittle time is spent in evaluating information, either for relevance, accuracy or authority and [those born after 1993] have been observed printing-off and using internet pages with no more than a perfunctory glance at them. Researchers have similarly found young people give a consistent lack of attention to the issue of authority. (Rowlands, 2008, p. 13)

Emanuel (2009) elaborates upon this idea further, expressing a fundamental difference in search habits and strategies between librarians and web-savvy patrons:

> "librarians develop a strategy before actually searching. This can take many forms, including limiting a search to a particular format... which is done by using limiters on the main search screen of a catalog... However, users do not search like librarians. They are accustomed to entering a keyword in a single search box, seeing what comes up, and then limiting on the basis of the results" (Emanuel, 2009, p.119).

This search-and-refine method relies upon the system to make a reasonable judgment, and anticipate the relevance of items to a user's ambiguous initial query. From this point, the user is able to make decisions about how to pursue further refinements.

This unarticulated desire by users for direction in searching is contrary to the user experience of classic catalogs. While classic catalog products excel at representing the physical collection and providing known-item search, the high learning curve of the interface frustrates many users. Ballard (2011) notes a study in which the New York Law School added a next-generation catalog (Innovative Encore) in parallel to its classic catalog, making both available. Both products were enrolled in Google Analytics, a free service that analyses user behavior. The differences in how users responded to the two different catalogs were dramatic, where

> "[next-generation catalog] users spent nearly four minutes in the system, but classic catalog users only stayed in a session for about 90 seconds. Bounce rate, or the number of users who left a session without searching was 33 per cent for [next-generation] users and 85 per cent for classic catalog users. Encore users looked at more than four pages per session and classic catalog users looked at 1.3... a searcher was 15 times more likely to refine a search if they were in [the next-generation catalog] than the classic online catalog" (Ballard, 2011, 267).

The clear implications of this study are that users interacted with the next-generation catalog in a more dynamic, discovery-oriented fashion than the classic catalog, and a larger number of users walk away from the classic catalog frustrated.

As these studies have demonstrated, younger, more web-savvy patrons show a distinct preference for search engine functionality and ease of use, many libraries have made a major push toward implementing these next-generation features in their OPAC, or as a separate discovery layer presented in parallel. Yang and Wagner (2010) provided a

solid framework for evaluating next-generation feature inclusion in academic libraries' OPACs. They created a checklist of twelve useful discovery tools that have been widely discussed in the library literature as being characteristic of a next-generation catalog:

1. *Single point of entry for all library information* (faceted searching)
2. *State of the art web interface*
3. *Enriched content* (reviews, summaries, cover art, etc.)
4. *Faceted navigation*
5. *Simple keyword search box on every page*
6. *Relevancy by circulation statistics*
7. *Did you mean…?* (spell correction)
8. *Recommendations/Related Materials*
9. *User Contribution* (folksonomy/tagging, reviews, etc.)
10. *RSS Feeds*
11. *Integration with social networking sites*
12. *Persistent links*

(Yang and Wagner, 2010)

Yang and Wagner's checklist instrument was used in Yang and Hofmann's (2011) study of the prevalence of these next-generation features in 260 academic libraries in the US and Canada. They found that 16% of library catalogs displayed no next-generation features whatsoever, 20% had only one feature, and 18% had two features. 73% lacked federated searching, 41% had no spell correction, and none of the systems displayed relevancy rankings by circulation statistics or recommendations based on patron transactions. Clearly, the next-generation catalog has not yet fully arrived.

Though the checklist instrument (Yang and Wagner, 2010) above presents a broad swath approach to assessing the next-generation catalog, it is heavily invested in looking for social features, particularly in regards to user-generated metadata (folksonomy keywords/tags, reviews, integration with prominent social networks, etc). While a useful instrument for investigating these sorts of features in a next-generation catalog, this study

does not seek to evaluate user-generated metadata in a research library context, but focus on resource discovery aids in the next-generation catalog. One major resource discovery feature present in some next-generation catalogs has been largely unexamined in the literature (and in Yang and Wagner's instrument) thus far: auto-suggest.

As several studies discussed above have shown, speed of results and actionable feedback from which a user can make decisions are two very important elements of the next-generation catalog. Novotny (2004), and Holman (2010) further indicated that younger users have a much more difficult time articulating their information needs into useful keywords, and boolean searching is even more problematic. Auto-suggest technology provides immediate feedback in the form of sample queries that have direct results from the catalog; this could very well have a demonstrable effect on user behavior once implemented. This view is extremely well articulated by Ji et al (2009), "users often feel 'left in the dark' when they have limited knowledge about the underlying data, and have to use a try-and-see approach for finding information. A recent trend of supporting autocomplete in these systems is a first step towards solving this problem" (p. 371).

The University of North Carolina began in January 2010 (Pennell and Sexton, 2010) to offer auto-suggest in their OPAC search service. A study by Pennell and Sexton (2010) indicated from query logs of the UNC OPAC that the auto-suggest service has proven itself useful; the average number of queries that used a suggestion from the query-suggestion service has been 18.06% (Pennell and Sexton, 2010). However, as useful as metrics such as dwell time, skips and clicks are to get an estimation of the usability of the discovery layer, these measures are at best implicit metrics.

As there is a paucity of other current literature looking at the implementation of auto-suggest technologies in library catalogs, the logical next step is to formally examine the prevalence of auto-suggest among academic libraries, and to see how its implementation is tied to other next-generation features.

Methodology

OPACs that feature next-generation features are large, complex systems. The costs of these systems vary widely; some are relatively expensive commercial products, whereas some are free (or low-cost) open-source projects. Either type of system will require significant resources, whether they are used to purchase a ready-made commercial product, or deploy and maintain an open-source system. As such, there is likely to be a significant difference between feature implementations at the most highly-funded research institutions and smaller academic institutions. As a result, this study looked at two separate domains of academic libraries: large research universities, and accredited four year and above colleges and universities.

Large research universities were defined as Research Universities of High or Very High Research Activity in the Carnegie Foundation for the Advancement of Teaching's Carnegie Classification. Smaller academic institutions were a larger grouping of Carnegie Classification types, including Doctoral Research Universities (a distinct group from Research Universities), Master's Colleges and Universities, Baccalaureate Colleges, and some of the Special Focus Institutions (Medical schools, Schools of engineering, Other technology-related schools, Schools of business and management, Schools of art, music, and design, and Schools of law). All of these schools were limited to 4-year or above, Public, or Private not-for-profit. Private for-profit schools were eliminated from

this study, as these institutions are usually either trade schools with little to no library-style resources, or institutions with no physical library location. Care had to be taken in selecting which special institutions to exclude, as many types of special institution have either very small libraries, or no libraries at all: schools that only offer Associate's degrees, Theological Seminaries and Bible Colleges, Other Health Professions Schools (mainly schools of acupuncture and alternative health modalities), Tribal Colleges, and Special Other (mainly schools of culinary arts and mortuary science).

The Carnegie Classification provides free data sets with listings of all institutions in the US, sorted and filtered by a variety of classification rubrics[1]. This study uses Basic Classifications. The first data set included RU/VH or RU/H, which returned 207 instutions. The second data set used the following filter:

> Basic = "DRU or Master's L or Master's M or Master's S or Bac/A&S or Bac/Diverse or Bac/Assoc or Spec/Med or Spec/Engg or Spec/Tech or Spec/Bus or Spec/Arts or Spec/Law" and Level = "4-year or above" and Control = "Public" or "Private not-for-profit"

which returned 1579 institutions. Both data sets were exported as separate CSV files for use with Microsoft Excel and JMP 9.

For each of the two main groups, 40 institutions were randomly selected as the sample, and each set was examined separately. The True Random Number Service[2] was chosen to create two separate lists of 40 random integers each; this tool was selected because it uses atmospheric noise to generate true random numbers, rather than the pseudorandom sequences found in statistics packages. The random sequence feature (which does random sampling without replacement) was chosen for convenience; after defining the appropriate upper and lower limits (lower limit was always 2 because of use

of an Excel row for labels) the first 40 numbers generated were used. Each institution's library catalog web page was discovered through Google searches, and the URI recorded for later evaluation of the catalog.

The protocol for each assessment session was to navigate to the specific library URI, and look for a search bar or a link to catalog search. When there was more than one option for catalog search, the option marked 'new', 'beta', or 'next-generation' was chosen. The assumption behind this behavior was that new or beta search tools would be more likely to contain next-generation features than older systems. On each system, four search strings were used, and, if there was an option, keyword search was chosen. The search terms used to evaluate the OPAC search features were (in order of use):

1. As we may think
2. Huckleberry Finn
3. Edward Tufte [intentionally not reversed]
4. Polysaccherides [sic]

The rubric for evaluation of each institution's catalog page was the following table:

Institution Name	ARL?	Autocomplete?	Federated?	Relevancy Ranked?	Faceted?	Spelling Correction?

Each of these metrics were evaluated on the following criteria:

- ARL: whether or not the institution is a member of the Association of Research Libraries, a selective organization representing libraries with the largest collections in North America. This was determined by looking at the ARL membership rolls[3] on February 2, 2012.
- Autocomplete: the presence or absence of any UI element near the search box that populates with a selectable list of query suggestions while typing in the main search box.

- Federated: did the results page return results to specific journal/serial/periodical articles or databases that are not usually cataloged monographically in the main catalog?
- Relevancy Ranked: does the results list have an option to sort by relevancy?
- Faceted: does the results page have a list of filters to select by subject, format, location, language, etc?
- Spelling Correction: did the OPAC at any point offer a spelling suggestion for 'polysaccharides' when 'polysaccherides' was entered?

After these data were collected, analysis was conducted to identify percentages of institutions whose library OPACs contain each of the main features listed above. Additionally, Fisher's Exact tests were done between each pair of variables to determine if there is any statistically valid relationship between each pair of variables. A value of $p<.0.05$ was considered enough to reject the null hypothesis that the variables are independent. The statistical analysis was carried out using JMP 9.

---

[1] http://classifications.carnegiefoundation.org/lookup_listings/standard.php

[2] http://www.random.org/

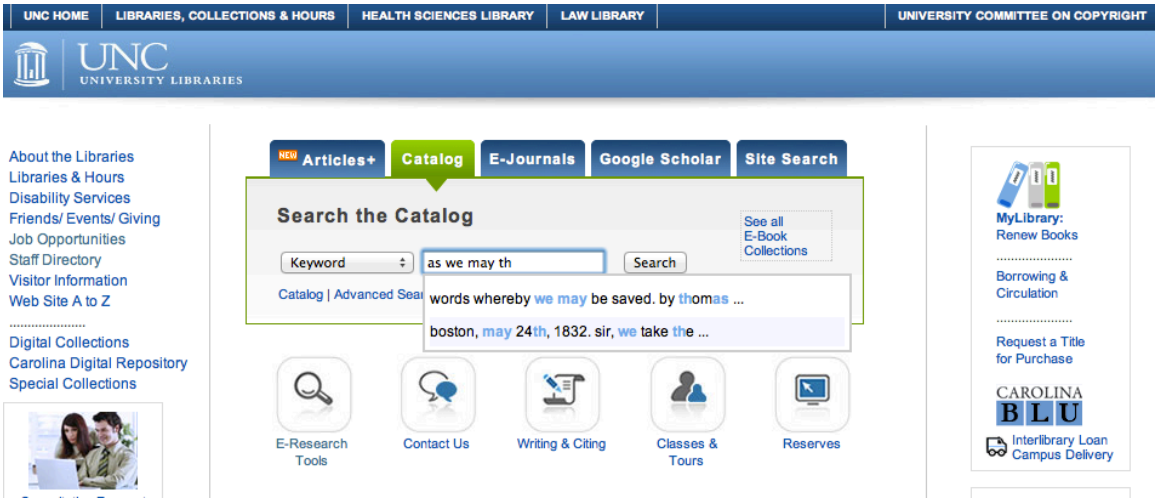
[3] http://www.arl.org/arl/membership/members.shtml

Findings and Discussion

The vast majority of institutions surveyed had web access to their library catalogs with no authentication credentials required. Among the research institutions data set, only one (2.5% of the sample) institution (University of Arkansas) was not available due to password-protection. Among the smaller academic libraries data set, four institutions (10% of the sample) were not able to be collected, due to there not being a web-accessible catalog available (Southwest Christian College, East-West University, Arkansas Baptist College, and Universidad Politecnica de Puerto Rico-Orlando Campus). It was sometimes possible to identify the system powering the next-generation features due to some sort of branding or disclosure on the web interface, but not always. Additionally, the user interfaces for the systems may have had custom CSS or other custom appearance features, making conclusive identification of the product further obscured. As a result, there will be no speculation on the identity of the specific products used by each institution.

Some form of auto-suggest for queries was present on 25% of Research Institution's OPAC services (10 institutions), versus 15% for the smaller academic library set (6 institutions). Interestingly, some (17.5% of the Research data set, 7.5% of the Smaller data set) featured an auto-suggest technology that was not available on the search

bar embedded in the library's home page, but appeared when using the search bar on the results page.

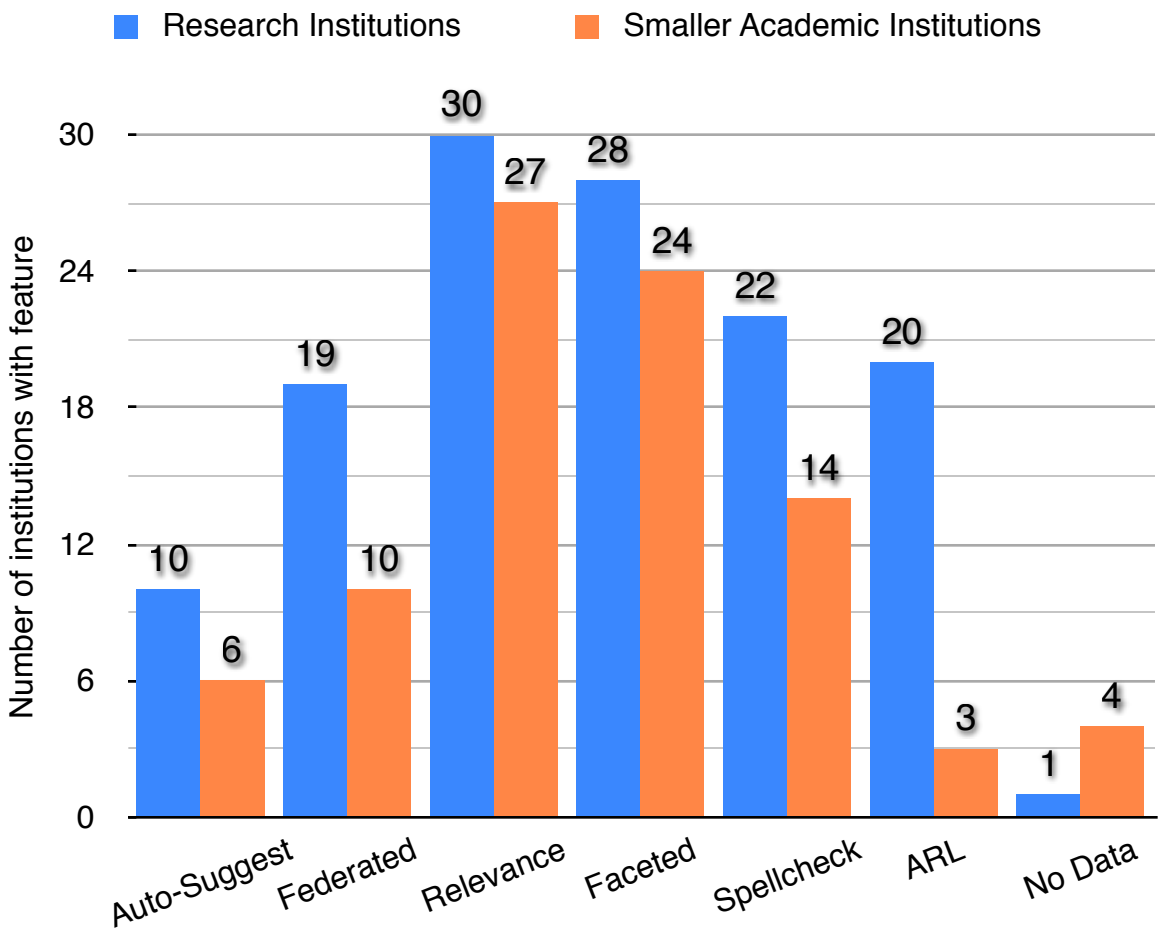


An example of Auto-suggest, from the UNC Library Catalog.

A form of relevance ranking was the most common next-generation feature available, appearing in 30 of the large institutions (75%), and 27 of the smaller institutions (67.5%). Faceted search limiting was also very frequent, appearing in 28 (70%) of the large institutions and 24 (60%) of the smaller academic libraries. These two features have long been bandied about by the library community as useful to catalog search; it is unsurprising that they appear so prevalently, with a small adoption gap between the large and small institutions.

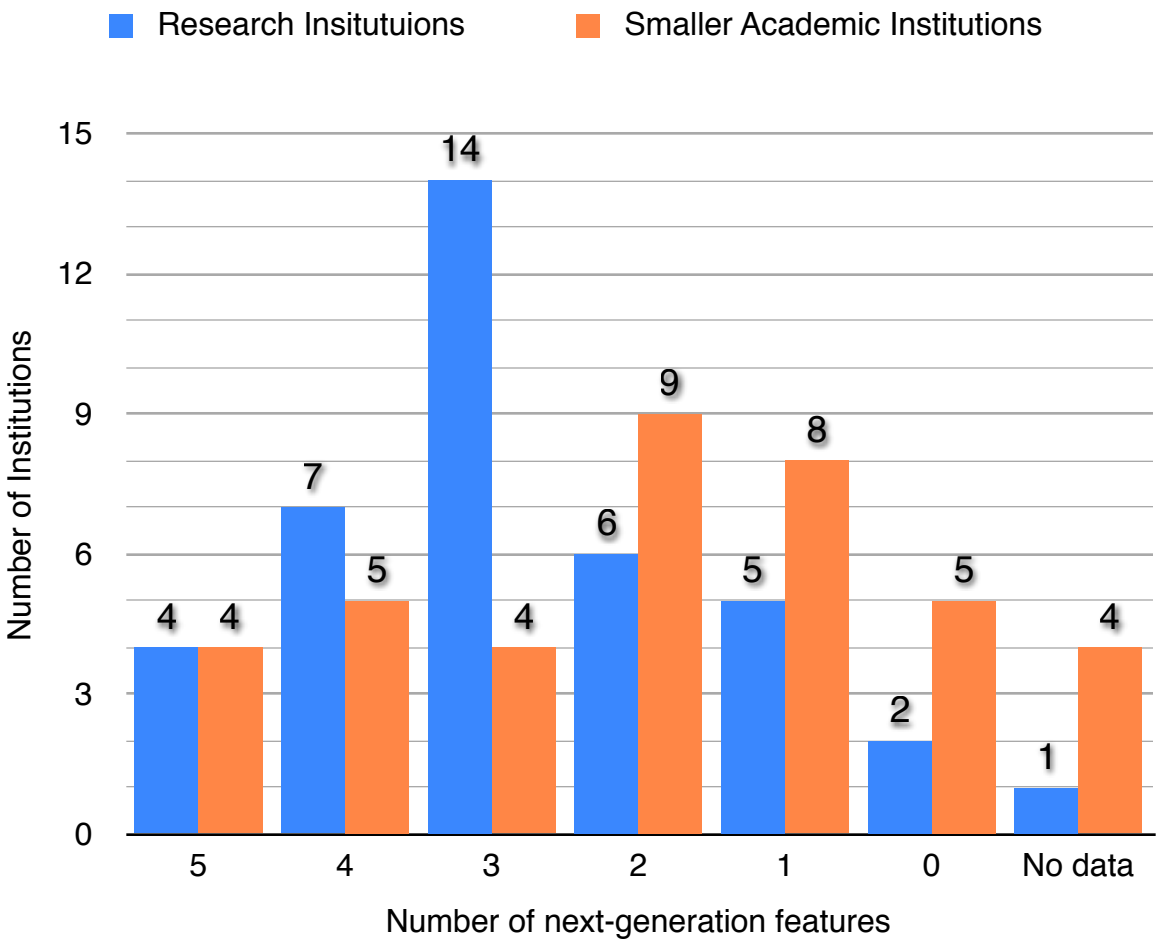
A larger gulf between institutional categories was found with the remaining next-generation features. Nearly half of the large research institutions offered federated article searching (47.5%), as opposed to a quarter of the smaller academic libraries (25%). Some sort of implicit spell-correction or explicit "Did you mean...?" feature was present on more than half of larger institutions (55%), compared to the 35% of smaller libraries.

7.5% of academic institutions belonged to the Association of Research Libraries, due to being considered part of the larger state university system library. These three schools were University of Minnesota-Morris (part of the University of Minnesota Library system), the Cleveland Institute of Art (part of Case Western Reserve University Libraries), and Oklahoma State University Center for Health Sciences (part of Oklahoma State University). All three of these institutions shared a catalog with their larger parent institution. Among research institutions, 50% belong to ARL.



Among large research institutions, four had all five next-generation features, seven had four features, fourteen had three features, six had two features, and five had

only one feature. Among smaller academic institutions, four had all five next-generation features, five had four features, four had three features, nine had two features, and eight had just one feature. Among the research-classified institutions, only two library catalogs (5%) contained no next-generation features whatsoever, George Mason University, and University of Oklahoma - Norman Campus. Among the smaller institutions, five had no next-generation features (12.5%).



After performing Fisher's Exact two-tailed tests between groups of variables to test for independence, some interesting associations appeared. Interestingly enough, in the research institution group, there was no indication of any association between ARL

membership and the presence of any other feature. As ARL membership serves as a reasonable proxy for overall library funding, it is all the more interesting that there is no association between ARL membership and next-generation features. However, in the research institution group, auto-suggest was associated with federated article search (p=.0033), relevance ranking (p=.0321), and faceted limiting (p=.0373). In this same group, federated search was associated with faceted limiting very strongly (p=.0001), and relevance ranking was strongly associated with spelling correction (p=.005).

Among the smaller academic institutions, auto-suggest was associated with only federated article search (p=.0001). This makes a great degree of sense, as they are the two least frequent next-generation features found among both groups. Federated article search was also associated with relevance ranking (p=.0394), faceted article search (p=.0146), and spelling correction (p=.0262). Finally, relevance ranking and faceted article search were also found to be associated (p=.0362).

These data clearly reinforce the idea presented in Yang and Hoffman (2011) that the next-generation catalog has still not arrived. Most libraries, whether larger research institutions or smaller academic libraries, have implemented only 1-3 next-generation features on their catalogs. Additionally, the features most commonly implemented were relevance ranking and faceted limiting, both of which tend to be more incremental changes from classic catalogs, rather than the major paradigm shift into a discovery-oriented platform. Though outside of the scope of this research, the quality of the various relevance ranking algorithms encountered in this study was highly variable. Some OPACs, though meeting the criteria in this study for being considered relevance

ranked, had a very poor implementation of this feature. The implications of this, and directions for further research will be discussed in the conclusions and implications section of this study.

Federated search, spell correction and auto-suggest are the most search engine-like next-generation features, and have the least presence in library OPACs. These discovery-oriented features are challenging to implement, and require significant investment in either vendor-specific products, or open-source implementation resources. Auto-suggest, despite being the least represented feature in this study, could have huge benefits to users seeking a more search engine like experience with an OPAC.

Conclusions and Implications

The importance of integrating modern information technologies to facilitate searching is clear, and a wide range of academic libraries have expended the resources to implement some of these next-generation features. Of the variety of next-generation features, those that direct discovery are tools that readily fit the needs of web-savvy users. As noted earlier, users are looking for a "decision engine" type of service (Kutub, 2010), that makes suggestions.

Auto-suggest is an excellent example of a simple, easily used discovery tool, and an immediate signifier of a robust information system; the immediate feedback it presents is a clear signal to users that they are interacting with a powerful, modern search tool. Unfortunately, auto-suggest is not currently a widespread feature in academic library catalogs, likely due to it being both a relatively new technology, and challenging to integrate into an OPAC. As auto-suggest is a pervasive feature in web search, more study is needed to determine what effect, if any, its presence or absence creates on the searching behaviors (and perceptions) of younger users.

Future research is also needed to determine the importance of a larger range of next-generation features to users. While it is certainly important to note self-identified responses of users to next-generation features, the likelihood of gathering useable, actionable data with this method is low; this technique is more useful as a bellwether indicating whether such improvements are in the right direction. A more in-depth,

longitudinal study examining user behavior before and after next-generation feature implementation is needed to see whether or not these features make a behavioral difference to the user. Transaction log analysis, after controlling for various population changes is a sufficient method to determine any substantive changes that arise from feature implementation.

Many users of academic library catalogs are steeped in the search and discovery paradigm of the web, and find traditional OPAC functionality to be cumbersome. College students have repeatedly been shown to prefer services like Google to library catalogs due to speed, convenience, and ease of use. Traditional library catalogs, though powerful, are increasingly perceived by younger users as outdated, unwieldy, and slow. However, as this study shows, academic libraries have shown a strong push toward closing the feature gap with popular web search services. Though auto-suggest is not yet a mainstream feature in academic library OPACs, academic libraries are to be congratulated on their progress in an era of tight budgets and expanded expectations.

Appendix A: Sample Set

Research Institutions:

1. Boston College
2. Brown University
3. Claremont Graduate University
4. Clark University
5. Clemson University
6. George Mason University
7. Louisiana State University and Agricultural & Mechanical College
8. Loyola University Chicago
9. Michigan State University
10. Michigan Technological University
11. Montana State University
12. New Jersey Institute of Technology
13. North Carolina State University at Raleigh
14. North Dakota State University-Main Campus
15. Northwestern University
16. Ohio University-Main Campus
17. Old Dominion University
18. Portland State University
19. Rensselaer Polytechnic Institute
20. Rockefeller University
21. SUNY at Albany
22. The University of Alabama
23. The University of Tennessee
24. The University of Texas at El Paso
25. University of Arizona
26. University of Arkansas
27. University of California-Berkeley
28. University of California-Santa Cruz
29. University of Kansas
30. University of Kentucky
31. University of Memphis
32. University of Missouri-Kansas City
33. University of North Carolina at Chapel Hill
34. University of Oklahoma Norman Campus

35. University of Rochester
36. University of Southern California
37. University of Wisconsin-Madison
38. Vanderbilt University
39. West Virginia University
40. University of Southern Mississippi

Smaller Academic Institutions:

1. Southwestern Christian College
2. Saint Michael's College
3. University of Minnesota-Morris
4. Corcoran College of Art and Design
5. The University of Tennessee at Chattanooga
6. Goshen College
7. East-West University
8. Huntingdon College
9. Arkansas Baptist College
10. McPherson College
11. Concordia University
12. Charleston Southern University
13. DePaul University
14. Wayne State College
15. University of the Southwest
16. Fitchburg State University
17. California Lutheran University
18. Shorter College-Professional Studies
19. Fort Valley State University
20. American Conservatory Theater
21. Peace College
22. Universidad Politecnica de Puerto Rico-Orlando Campus
23. Judson University
24. Le Moyne College
25. Hollins University
26. Texas A & M University-Kingsville
27. The Boston Conservatory
28. Cleveland Institute of Art
29. Lake Forest College
30. University of Central Missouri
31. Middlebury College
32. Albany Medical College
33. Clayton State University

34. Oklahoma State University Center for Health Sciences
35. Carroll College
36. Cedar Crest College
37. CUNY York College
38. Bennett College for Women
39. Marian University
40. Holy Names University

References

Allison, D. (2010). Information portals: The next generation catalog. *Journal of Web Librarianship, 4*(4), 375.

Ballard, T. (2011). User search-limiting behavior in online catalogs. *New Library World, 112*(5-6), 261.

Calhoun, K., Cantrell, J., Gallagher, P., Hawk, J., Gauder, B., Limes, R., . . . Cellantani, D. (2009). *Online catalogs: What users and librarians want: An OCLC report*. Dublin, Ohio: OCLC Online Computer Library Center, Inc.

Calhoun, K. (2006). *The changing nature of the catalog and its integration with other discovery tools*. Library of Congress.

De Rosa, C., Cantrell, C., Hawk, J., Wilson, A., Gauder, B., Havens, A., . . . Smith, S. (2006). *College students' perceptions of libraries and information resources: A report to the OCLC membership.* Dublin, Ohio: OCLC Online Computer Library Center, Inc.

Emanuel, J. (2009). Next generation catalogs: What do they do and why should we care? *Reference and User Services Quarterly, 49*(2), 117.

Emanuel, J. (2011). Usability of the VuFind next-generation online catalog. *Information Technology and Libraries, 30*(1), 44.

Fast, K. V., & Campbell, D. G. (2004). I still like google: University student perceptions of searching OPACs and the web. *Proceedings of the American Society for Information Science and Technology, 41*(1), 138-146.

Holman, L. (2011). Millennial students' mental models of search: Implications for academic librarians and database developers. *The Journal of Academic Librarianship, 37*(1), 19-27.

Jansen, B. B. J., & Pooch, U. (2001). A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology, 52*(3), 235-246.

Ji, S., Li, G., Li, C., & Feng, J. (2009). Efficient interactive fuzzy keyword search. Paper presented at the *Proceedings of the 18th International Conference on World Wide Web,* Madrid, Spain. 371-380.

Kutub, M., Prachetaa, R., & Bedekar, M. (2010). User web search behaviour. Paper presented at the *Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on,* 549-554.

Lippincott, J. K. (2005). Net generation students and libraries. *EDUCAUSE Review, 40*(2), 56.

Novotny, E. (2004). I don’t think I click: A protocol analysis study of use of a library online catalog in the internet age. *College & Research Libraries, 65*(6), 525.

Pennell, B. (2010). Implementing a real-time suggestion service in a library discovery layer. *The code4lib Journal,* (10)

Rowlands, I. (2008). The google generation: The information behaviour of the researcher of the future. *Aslib Proceedings, 60*(4), 290-310.

Wisniewski, J. (2009). Next-gen OPACs: No time like the present. *Online (Weston, Conn.), 33*(5), 54.

Yang, S. (2010). Evaluating and comparing discovery tools: How close are we towards next generation catalog? *Library Hi Tech, 28*(4), 690.

Yang, S. (2011). Next generation or current generation?: A study of the OPACs of 260 academic libraries in the USA and canada. *Library Hi Tech, 29*(2), 266.

Young, M., & Yu, H. (2004). The impact of web search engines on subject searching in OPAC. *Information Technology and Libraries, 23*(4), 168.