

Resource Discovery System Usage Report, 02-01-2012 to 08-25-2014

Beginning in 2011 the User Interfaces department began making sweeping changes to the UNT Libraries' resource discovery interfaces following producing a comprehensive report in which we examined the state of library resource discovery in general and the state of discovery specifically at UNT. In fall of 2012, our changes culminated in the release of a wholly redesigned website. Now we are ramping up for more changes in fall of 2014 into 2015. To help guide us, we need to look at usage statistics during the relatively stable period following the last website redesign.

This report looks at usage statistics for the two portions of our discovery environment that are most comparable and see the largest amount of use: the catalog and Summon. Since implementing Summon, the UNT Libraries organization has had questions about how people use Summon, what effect Summon use might have on the catalog, what the relationship between Summon and the catalog should be, and—ultimately—how we should present Summon to our users. In a few cases speculation about the answers to these questions has been contentious.

The easiest way to address some of these questions—the lowest hanging fruit, so to speak—revolves around examining how people actually search our systems. What kinds of searches are they conducting most frequently? Which systems are they searching more, when? Does their searching reflect that they make the same distinction that UI makes between the two systems—articles versus books? This report is an initial attempt at gleaning what information we can, focusing on *search* data because search data is easy to get and easy to compare between the two systems. We plan to delve more deeply into our usage data and more widely across our additional search platforms during the coming year.

Methodology

To create this initial report, we've used the Google Reporting API to extract data from our Google Analytics accounts for Summon and the catalog, and we've used Microsoft Excel to graph the data. We've primarily used pageviews as a metric to help us approximate the number of searches conducted, and we've filtered pages based on URL patterns to help determine where and how people are conducting searches. For instance, when users search the catalog from the *Books & Media* search tab in the *Find* box, it uses a particular URL pattern when it sends the search to the catalog. When we extract usage stats, we can filter page URLs using regular expressions so that we only get the URLs that conform to the relevant pattern, showing us only the searches that used the *Find* box.

Pageviews are only an approximation of searches, as activity such as refining a search (e.g., by clicking facets in Summon) and sorting results can generate additional pageviews. Paging through results does *not*, however, as our URL patterns filter out paged results. Furthermore, not all pageviews represent a search from an actual search form—it's not uncommon to bookmark a link to a result set in some way, either for personal use or for others to use. Although we could filter out most of this traffic, it seems to make sense to account for it in our data. Following a link to a canned search is qualitatively different than entering the search on your own, but it still represents a user entering the system to conduct a search.

Finally, note that, when we look at catalog searches, we're intentionally filtering out course reserve searches and Database searches. These types of searches aren't part of the *Books & Media* tab on the *Find* box, so, even though they are technically conducted in the catalog, we don't think they're as comparable to searches conducted in Summon. This is an effort to keep our comparison of catalog usage to Summon usage as fair as possible. (We will treat searches for databases separately in the future when we look at usage statistics for the other *Find* box tabs.)

Aggregate Patterns

To get an overall picture of how people are searching these two systems, we'll first examine aggregate patterns, using a broad URL filter to discover all comparable searches. Which system do people seem to be using more often? Figure 1, below, compares searches in the catalog to searches in Summon.

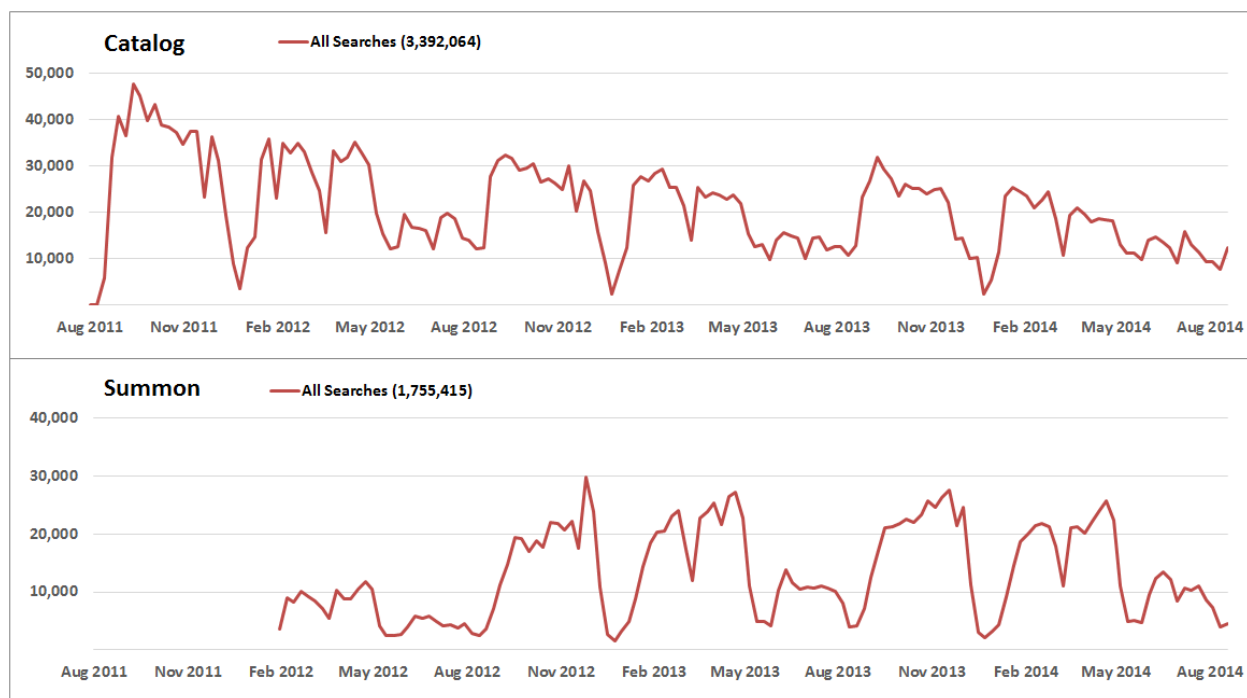


Figure 1: Comparison of Aggregate Catalog and Summon Searches, August 2011 to August 2014

Summon Beta versus Production

Perhaps the first feature in Figure 1 that stands out is Summon's surge in use after putting it into full production: in September 2012, usage suddenly more than doubles. This isn't surprising. During the beta period, the search was relegated to a secondary page on the website, but, in production, it became the default search. Its usage since has remained steady.

Effect of Summon on Catalog Use

Our decision to acquire Summon was predicated on the idea that we had a largely unmet need in our resource discovery ecosystem: we lacked a way for people to search for full-text journal articles. If our judgment was correct, based on the fact that we implemented Summon as an online article search, then

we'd expect Summon not to have had too much of an effect on catalog use, since the two systems serve two separate purposes.

Here, the data does show a pattern of declining catalog use. However, the decline doesn't appear to represent a wholesale switch from the catalog to Summon. The amount of use that Summon gets is disproportional to the decline in catalog use, which indicates that people are using both the catalog *and* Summon. Summon probably has absorbed some amount of catalog use, but—by and large, people are also still using the catalog.

The Academic Calendar

Library system use has always tended to follow the academic calendar, with usage remaining highest during the long semesters and lower (but steady) during the summer. We generally see troughs between semesters and during holidays when fewer people are using the systems.

This pattern is clear both in Summon and in the catalog. However, there are two interesting sub-patterns that differ between them.

First, during each long semester, Summon use increases during the course of the semester, peaking at the end, while catalog use starts high and decreases. One possible explanation for this pattern is that students have research projects and papers due at the end of the semester, and they tend toward needing to get full-text materials more quickly as the due date approaches, which makes them gravitate more toward Summon during that time.

Second, the troughs that happen between semesters are deeper and wider in Summon than they are in the catalog. When a semester ends, Summon use drops off immediately and remains low until the next semester begins, while catalog use is a bit steadier during semester breaks. Drawing any firm conclusions about this pattern is difficult, but intuitively it makes sense that the library catalog would continue to see use between semesters from people who either remain on campus or who otherwise come to campus when they would not normally do so. Whatever the case, it does seem to imply that Summon use revolves around the academic calendar even more so than does catalog use.

Distance Use

UNT has long had a large population of commuter students, and so, for the UNT Libraries, serving distance users has always been a priority. In Figure 2, we've plotted data to show the total number of searches, the number of searches from users located in Denton, and the number of searches from users on the unt.edu domain for each system. Although some users in Denton are local and some are not, and, likewise, some users on the unt.edu domain are local and some are not, this gives a general idea about how much use each system gets from local users and distance users.

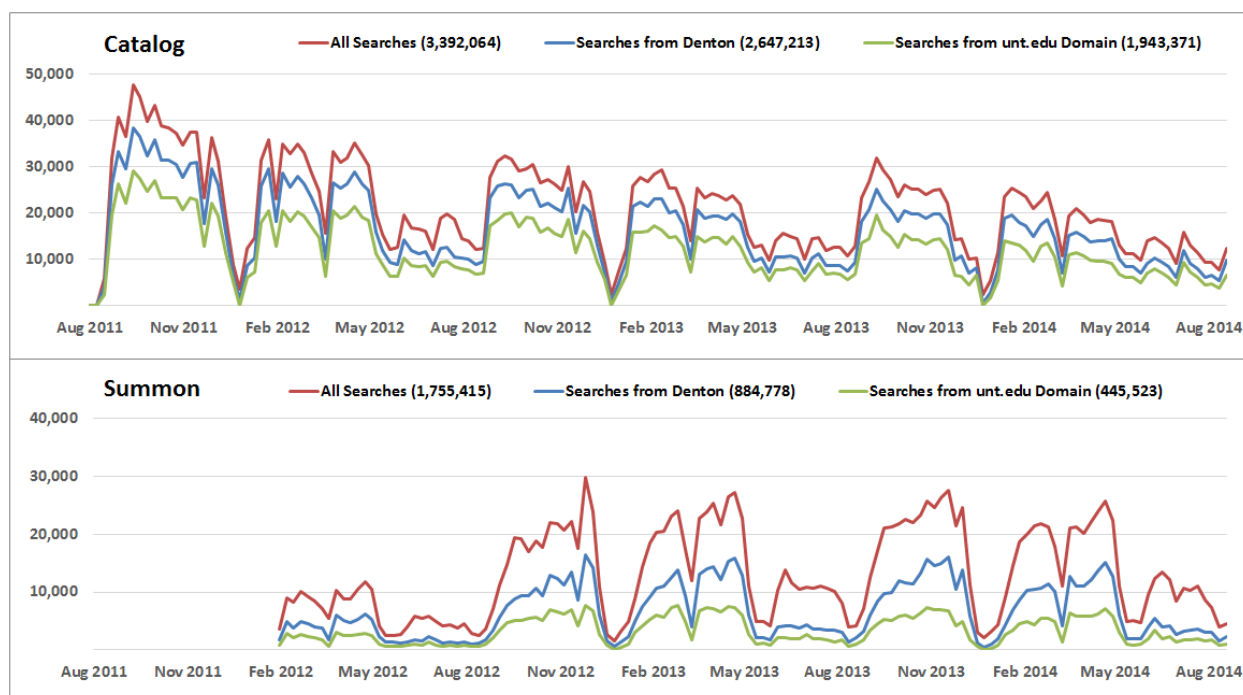


Figure 2: Aggregate Use Compared to Local Use, August 2011 to August 2014

In each graph, the red line represents total searches, the blue line represents searches coming from Denton users, and the green line represents all searches coming from users on the unt.edu domain.

We can see that the graph representing Summon use has a much bigger gap between total searches and local searches than the one representing the catalog, indicating that distance users—those who are not in Denton and not on the unt.edu domain—make up a significantly higher proportion of Summon users than they do catalog users. The numbers indeed reflect this: in the catalog, Denton users account for 77.31% of search traffic, and unt.edu users account for 56.54%. However, in Summon, Denton users account for only 50.40%, and unt.edu users account for 25.38%.

These statistics make sense and fall within our expectations; Summon’s role as an articles-search system means it focuses on full-text, immediate-access resources. Although the catalog does contain records for online resources, it still serves as the main guide to our physical collections, and so it sees higher levels of use from people who are on campus trying to access items in those collections.

***Find* Box – Tab Usage**

Since we launched the redesigned library website in September 2012, we’ve intended that the *Find* box that appears prominently on the home page should serve as a major entrance point to many of our search systems. But how often do people actually use that box?

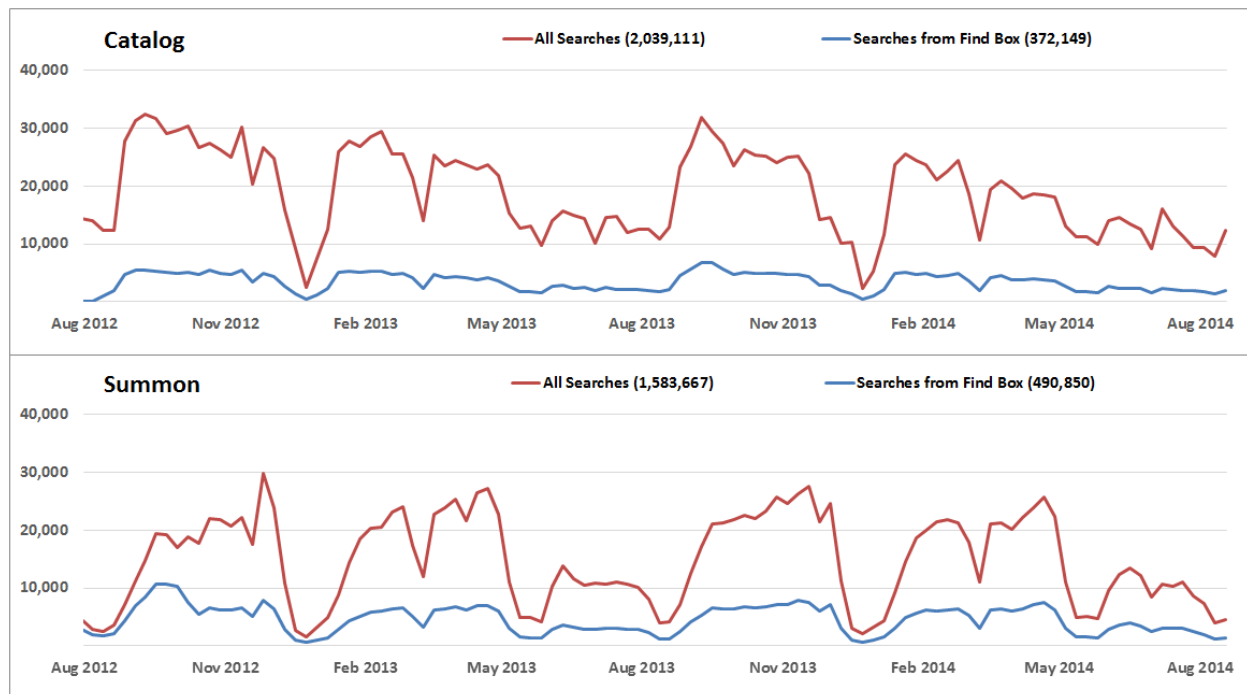


Figure 3: Searches Conducted via the *Books & Media* Tab and the *Online Articles* Tab, August 2012 to August 2014

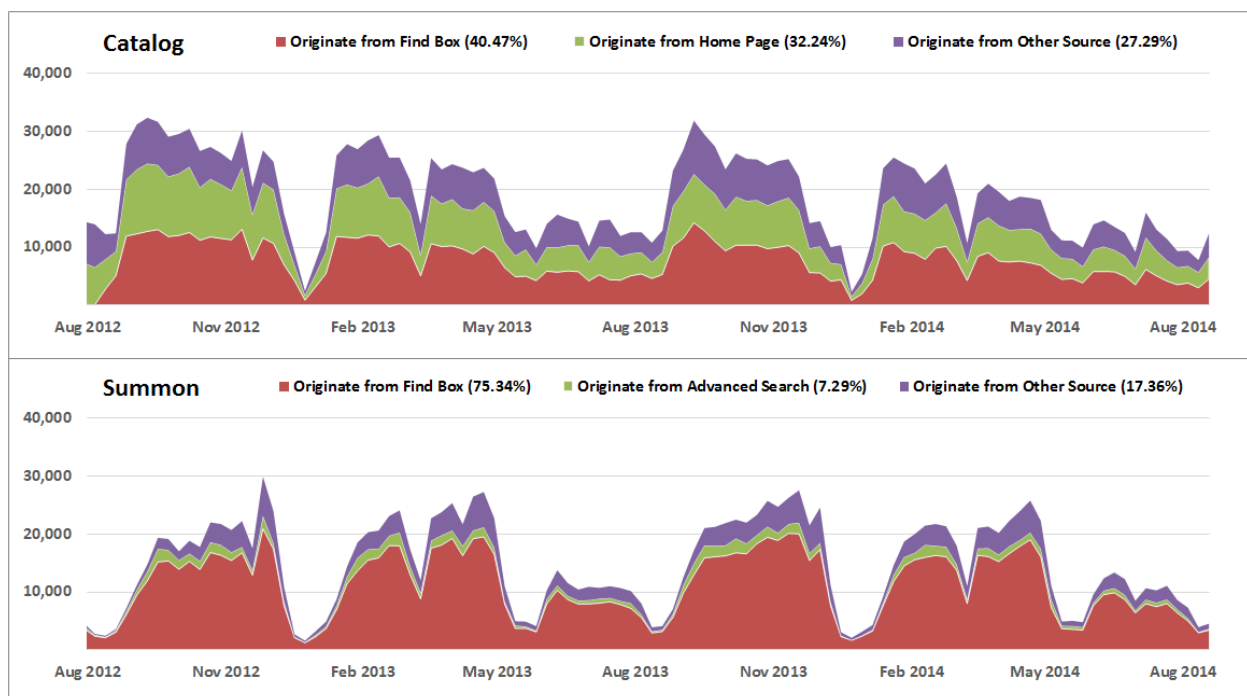


Figure 4: Searches by Session Source, Catalog versus Summon, August 2012 to August 2014

The graphs in Figure 3 show the number of searches conducted in each system that come directly from the *Find* box (the blue line) compared to the total number of searches (the red line). Note that, after a

user enters a system, subsequent searches do not utilize the *Find* box and are therefore not included in this statistic.

As a point of comparison, the graphs in Figure 4 show searches in each system divided into groups based on how users entered the site to begin their search session. In other words, if a user enters the system by searching the *Find* box and then continues searching, all of those searches are captured as having originated from the *Find* box. In both graphs, the red area shows the proportion of searches where users entered via the *Find* box; the green area in the Catalog graph shows where users started from the catalog homepage; the green area in the Summon graph shows where users started from the Summon advanced search screen; and, in both graphs, the purple area shows the proportion of searches originating from all other points.

Finally, note that we only have meaningful stats for the *Find* box beginning the semester that it was launched, Fall 2012—so these graphs start at August 2012 rather than August 2011 or February 2012.

The Online Articles Tab versus the Books & Media Tab

The default tab selection for the *Find* box has been a somewhat contentious issue. UI had decided unilaterally that the Online Articles tab would be the default. Our reasoning was that the *Find* box was primarily aimed at undergraduates and new users, and our user data suggested that these were the user groups that would most benefit from a find-articles search. Plus, we assumed that more advanced library users, such as librarians and faculty, would tend to go straight to the catalog homepage and bypass the *Find* box altogether. In the following months, several of these users were vocal in letting us know that they would indeed prefer Books & Media to be the default *Find* box tab, seemingly contrary to what we believed our earlier data had suggested.

Is this a case of a vocal minority skewing our understanding of the situation, or are their views more representative than we thought? Can the statistics shown in figures 3 and 4 be used to address this debate?

Figure 3 shows that the Online Articles tab does get more use than the Books & Media tab—about 31.9% more. The question is, is this because people actually find this type of search more useful, or is it simply because it's the default? Although there is not enough information to tell for sure, the fact that usage patterns remain mostly static from semester to semester suggests that people are continuing to use the system in similar ways. If they were using it incorrectly and not finding useful results, we might expect usage to spike and then drop off.

One interesting feature shown in Figure 3 is that the Online Articles tab use *does* spike right away in September 2012 before settling into a more normal pattern. This spike is interesting for a few reasons. A spike at that time does not occur in the Books & Media tab use statistics. The spike is not reflected in the overall searches for Summon at that time (indicating it was *only* a spike in Online Articles tab usage and didn't actually generate more Summon searches overall). And a spike like that never recurs. It's tempting to imagine the spike resulted from user trial and error immediately following the release of the new website—that some users did find the default tab to be confusing for a short time but then figured it out.

Ultimately, the debate is nearly moot. In a few months we will release a Bento Box search that will become the new default and will incorporate all of the searches on the tabbed *Find* box into one search interface.

Proportion of Find Box to Non-Find Box Searching

Figure 4 shows that most Summon search sessions (75.34%) come from the *Find* box while relatively fewer catalog search sessions (40.47%) do. These figures basically conform to our expectations.

First, the catalog is a more established system with its own preexisting user base that has its own use patterns. One of those use patterns is, as we suspected, to go straight to the catalog home page in order to conduct a search. Other common entry points for searching include the Media and Music scope pages along with direct traffic (e.g., links and bookmarks).

Second, Summon is a new system, and, as there is no real advantage to searching Summon via its home page, we have made no effort to point people there. The *Find* box has been the primary method available for searching Summon, and the statistics tell us that is in fact what people do. The portion of search sessions that start from the advanced search screen is negligible—only 7.29%—which is somewhat telling, since we’ve made no effort to hide the advanced search screen; that option is available directly from the Online Articles tab. Other common entry points to Summon include direct traffic (again, such as links and bookmarks) and LibGuides.

Find Box – Search Options

Both the Books & Media and Online Articles search tabs offer options for controlling one’s search. From Books & More, you can change which index you want to search (Keyword, Title, Author, etc.), and you can change which collection to search (UNT Dallas, Eagle Commons, Music, Media, etc.). From Online Articles, you can toggle a “peer-reviewed only” and a “show only articles that have the full-text available online” option. But how often do people actually change these default settings when they search using the *Find* box?

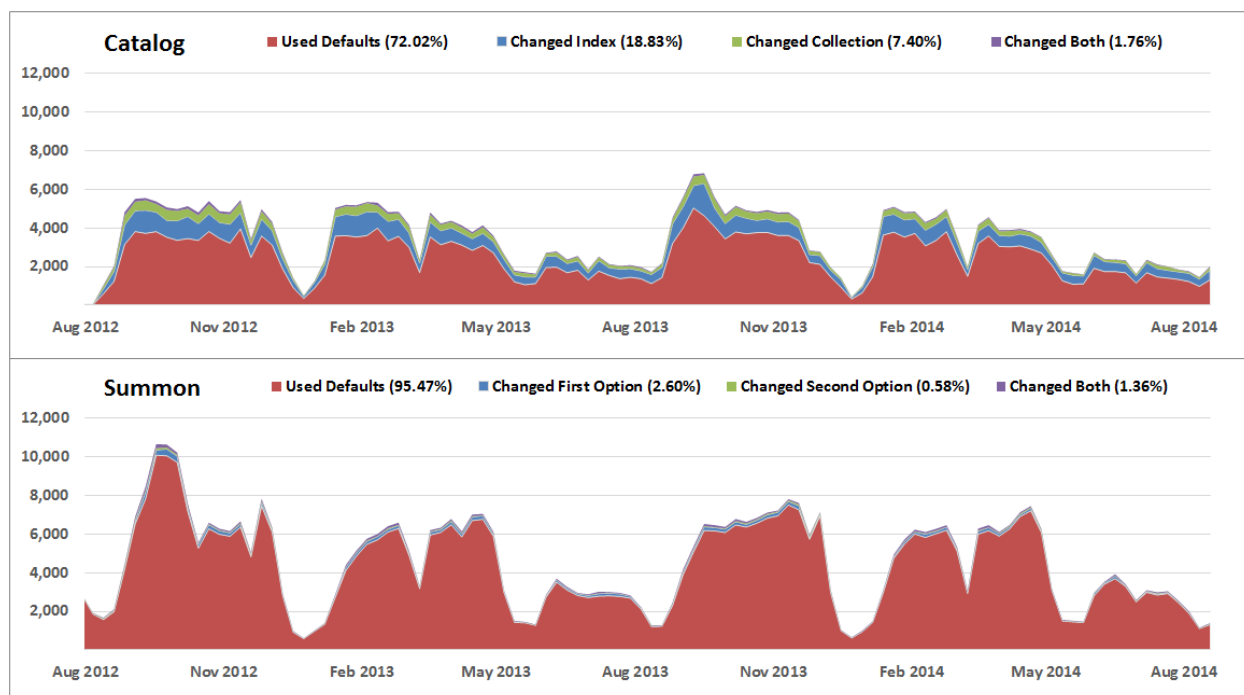


Figure 5: Find Box Searches, Which Options Did Users Choose? August 2012 to August 2014

Figure 5 shows the total number of *Find* box searches for the catalog and for Summon grouped by the available options that users changed. The red area shows the proportion of searches where users did not change the default options; the blue and green areas show where users changed one option but not the other; and the purple area shows where users changed both options.

That users rarely change default settings is something of a general rule in UI design, so the findings here aren't very surprising. Summon users especially do not change the default settings very often; whether this is because we have chosen the default settings well, because they don't understand the options, or because they simply don't care is open to debate. Perhaps the most we could say is that the default settings we have chosen for Summon—at least the two available on the *Find* box—are generally good enough.

What is perhaps more interesting is that catalog users actually *do* change the Books & Media options as often as they do. Our users especially seem to value the ability to conduct fielded searches in the catalog. We do not provide a similar option for the Online Articles search, so we don't know if they value this feature in general or if they value this feature specifically in the catalog. Furthermore, we wonder if they might seem to value this feature in the catalog because the catalog's keyword search is relatively poor or simply because this is what they expect to be able to do in a library catalog.

Catalog Use – Find-Box Searches versus Non-Find-Box Searches

Next we delve more deeply into the specific options people use when searching the catalog to find out if the way people use the *Find* box is representative of how people use the catalog as a whole.

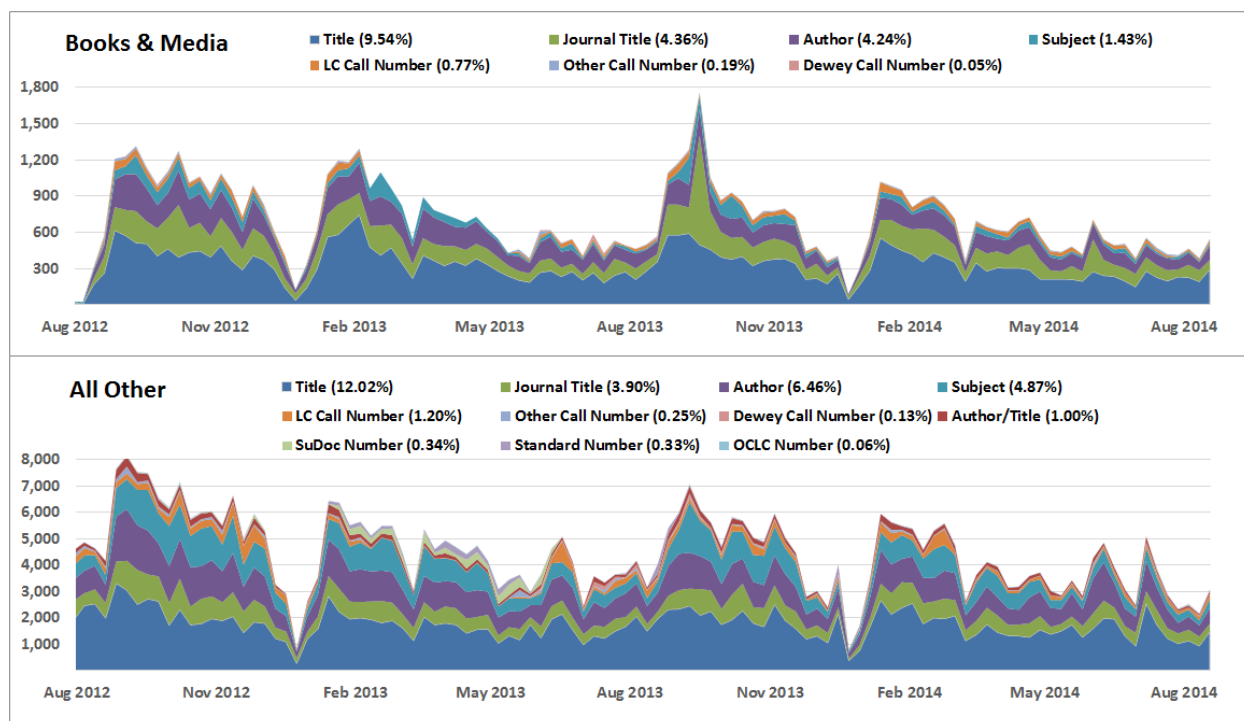


Figure 6: Catalog Index Searches except Keyword, Find Box (Books & Media tab) versus All Other, August 2012 to August 2014

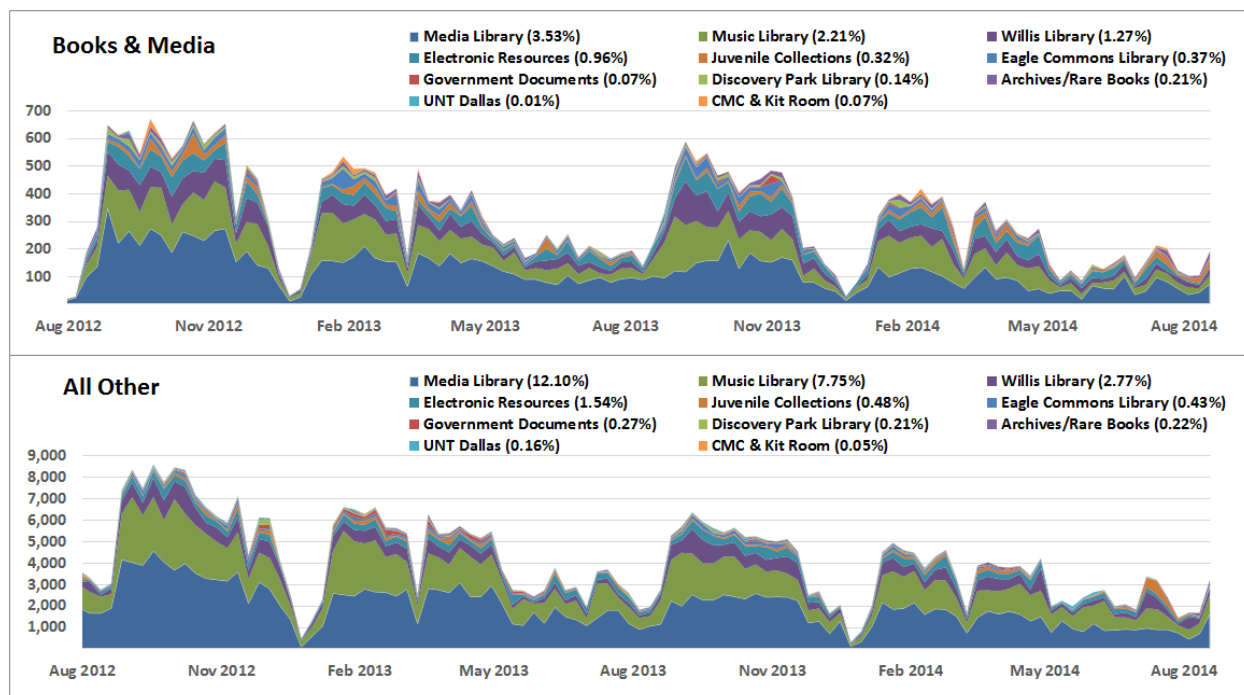


Figure 7: Catalog Collection (Scope) Searches except "All Collections," Find Box (Books & Media tab) versus All Other, August 2012 to August 2014

Figures 6 and 7 show the searches conducted via the Books & Media tab compared to the same index searches conducted from all other starting points, broken down into groups by each of the options available on the Books & Media tab both for index (Figure 6) and collection (Figure 7). Keyword and All Collections searches have been left off of the graphs because they are so large they make it difficult to see the breakdown of the other options. But the percentages reflect the breakdown of the full total, including keyword and All Collections searches.

Note that, for the sake of completeness, the index searches available in the catalog but not the *Find* box are included in Figure 6: Standard Number, OCLC Number, SuDoc Number, and Author/Title.

Dominance of the Default

Even though the default options aren't actually represented on the graphs, it's clear that they dominate all types of use: keyword searches represent 79.41% of all Books & Media searches and 69.94% of all other catalog searches; All Collections searches represent 90.85% of all Books & Media searches and 74.03% of all other catalog searches. Again, this is unsurprising: not only do users rarely change the default as a general rule, but these particular defaults also represent the most general search options.

However, we should point out that users stick with the defaults when searching via the *Find* box much more frequently than they use the general keyword and All Collections options overall. This seems to match up with our intention that the *Find* box be used as a more general search tool, where users conduct more specialized searches and search refinement *in* the catalog.

Index Searches

From Figure 6 we see that indexes used from the *Find* box are not too different than indexes used elsewhere. Title searches are the most common, followed by Journal Title, Author, and Subject searches. From the *Find* box, users tend to search for journal titles more often than authors or subjects, while author and subject searches are otherwise proportionally more common than journal title searches. It's possible this can be explained by the fact that clicking a name or subject heading from a bibliographic record fires off an author or subject index search, respectively, boosting those statistics. Call number searches are extremely rare. In fact, SuDoc and standard number searches, which are *not* present on the *Find* box, are searched more frequently than Dewey and local ("other") call numbers.

Collection Searches

Looking at Figure 7, we see two main differences in the use of collection searches when comparing *Find* box to non-*Find* box usage. The largest difference is in the use of the Media and Music collections. Although Media and Music *are* the most frequently searched collections in searches conducted from the *Find* box, those collections are searched much more often elsewhere. It's likely that this search activity mostly comes from the Media and Music scope search pages—e.g., <http://iii.library.unt.edu/search~S6/X> and <http://iii.library.unt.edu/search~S7/X>, which are linked from the catalog homepage and available via the catalog navigation as the *Media Library Catalog* and *Music Library Catalog*.

The other difference is that, almost across the board, other collections are searched slightly more often from sources other than the *Find* box—but only slightly.

Declining Searching by Collections

Figure 7 also shows a clear decline in this type of search (aside from the odd spike in the use of the Juvenile Collections near the end of July 2014). This trend does not seem to be explained by the overall relatively smaller decline in catalog use that we see in Figure 1 that we discussed earlier, as the decline in collection use seems more pronounced than the overall trend. In addition, each collection appears to be declining—there is no single collection that is wholly responsible for the decrease. Note that this decline does not mean people are using fewer Music resources, fewer Media resources, fewer Discovery Park resources, etc.—it simply means people are less frequently using the collection (or scope) functionality in the catalog to narrow their searches.

Summon Use – Search Refinement

Our default Summon search is somewhat narrow, and, as we’ve seen, the few options we do present to users on the *Find* box’s Online Articles tab hardly get used. But this isn’t the whole story. After users enter Summon, they have the opportunity to further refine their search. How do they tend to do this? How often do they conduct advanced searches? How often do they start over with the broadest possible Summon search? How often do they use facets to filter their results, and which facets do they use?

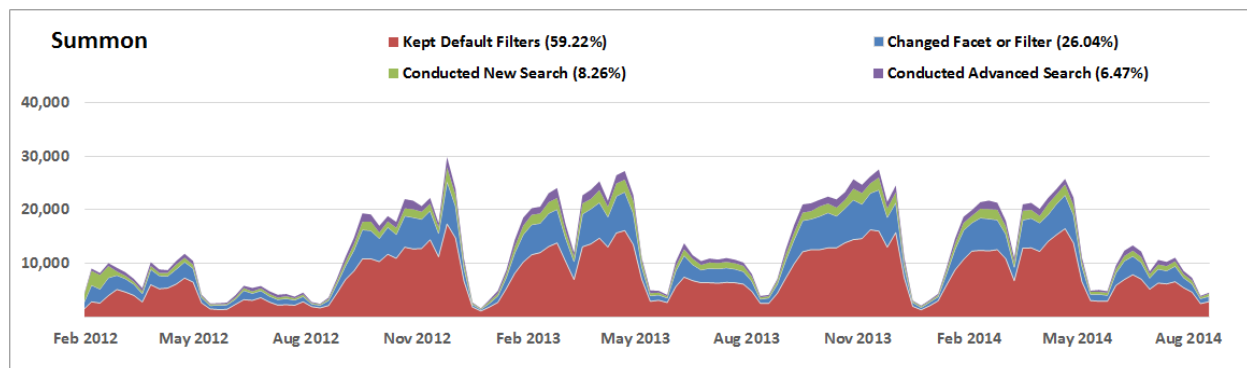


Figure 8: Summon Searches by Refinement Type, February 2012 to August 2014

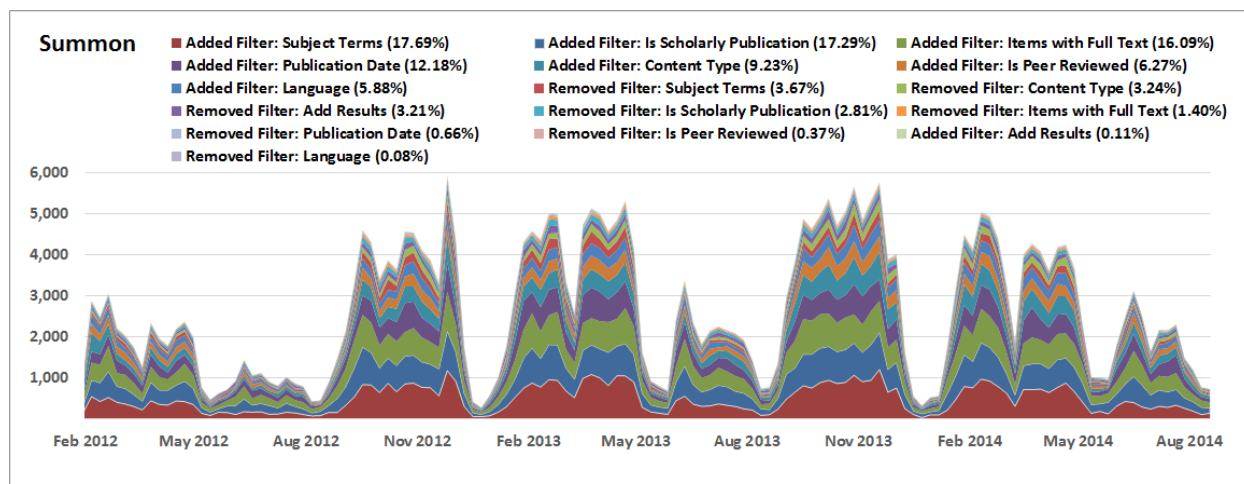


Figure 9: Distribution of Summon Facet and Filter Use, February 2012 to August 2014

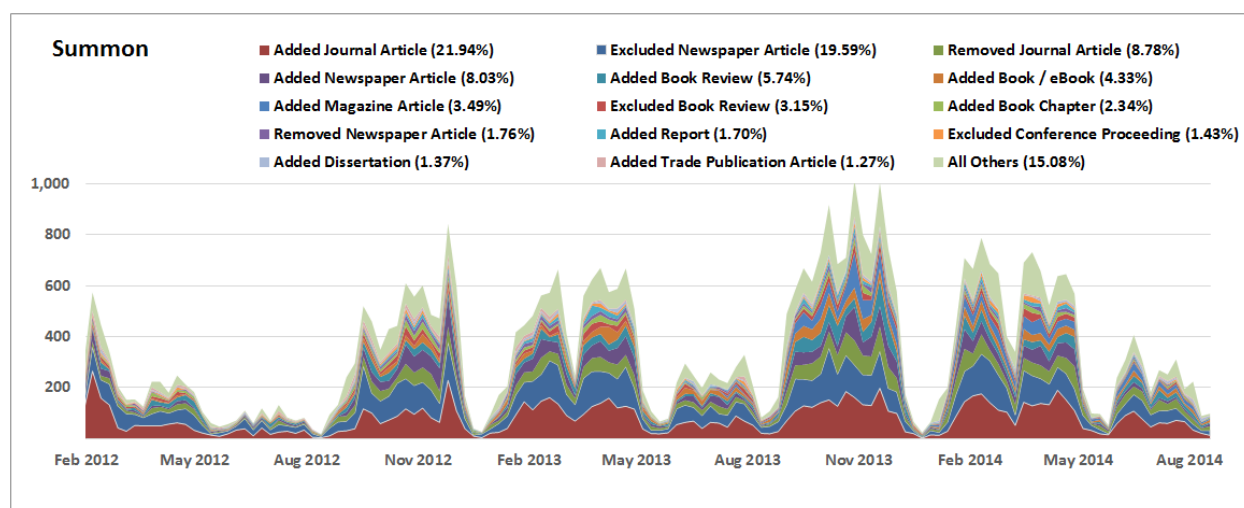


Figure 10: Summon Content-Type Facet Use

In Figure 8, we've divided Summon searches into four groups. First, represented by the red area, are searches that used *only* the default filters. Results may have had their sort order changed or users may have refined their search terms—but no facets or filters other than the defaults were ever selected. The second group, represented by the green area, includes advanced searches that were never further filtered. The third group, represented by the purple area, includes searches that used no filters at all—e.g., basic searches conducted from the Summon homepage or those resulting from electing to conduct a new search instead of keeping search refinements. The fourth and final group, represented by the blue area, includes searches where users refined their search by changing one or more facets in the interface after they conducted a basic or advanced search. Note that all of these groups are mutually exclusive. When a user first conducts an advanced search, that search would be included in the *advanced search*

group. If that user then adds or removes a facet filter, the new search would be included in the *changed facet or filter* group.

Figure 9 shows the distribution of exactly which facets and filters were used whenever a filter or facet was changed. Note that these only include searches where a facet or filter was changed *after* an initial search was conducted—they do not show, for example, advanced searches where particular filters were chosen as part of the advanced search.

Finally, Figure 10 shows which content types were added, excluded, and/or removed most often. The full data set contains 129 different categories, but we show the top 14 because they represent about 85% of all content-type facet use.

How Good are our Summon Defaults?

One of the things we're most interested in seeing from Figures 8, 9, and 10 is how useful the default Online Articles search seems to be. We try to gauge this by measuring how often people change the defaults and by looking at what options they choose when they do change the defaults.

As we see in Figure 8, the default options we've selected for Summon are kept in approximately 59.22% of searches. This means that, well over half the time, users search Summon without ever bothering to use facets or otherwise refine their search beyond entering new search terms. This is similar to the trend we see in the catalog where users tend not to deviate from the default options.

Furthermore, looking at Figures 9 and 10, we see that the specific default filters we chose for our Summon online articles search are all at or near the top of the filters that are most often chosen when people *do* use facets to refine their search. So, even when our default options are not selected (either because the users have removed them or because they've started a new or an advanced search), people often add them again while they're refining their search—in most cases, significantly more often than they choose other options.

Finally, Figure 10 shows that users do sometimes remove the journal article content type—it accounts for 8.78% of content type activity. When you consider that most Summon searches have the journal article type selected by default, removing it is comparatively quite rare.

Facets are Underutilized

Search facets are a “next generation catalog” feature meant to try to help make it easier for users to get more precise search results. Facets allow users to refine their searches on the fly instead of requiring them to formulate a complex search query all at once (such as when using an advanced search form). However, we see in our Summon stats that only 26.92% of all Summon searches make use of a facet or filter that was not already there when users first conducted their search, making them relatively underutilized.

Search Queries

In addition to looking at usage data that tells us how people are conducting searches, we've also extracted the actual queries people are using when searching our systems. This report only scratches the surface of search query analysis because, frankly, it's difficult to get completely meaningful results. Although looking at your top 100 queries is easy, it may not be as useful as you might hope because your top 100 queries may only represent 1 or 2% of all searches, which means you're ignoring the enormously long tail of less frequent queries. That isn't to say that the top queries aren't important, but they only tell you what things people are using the same search query to search for—they don't necessarily tell you what things or concepts people are searching for using different queries.

A more thorough search query analysis might require something like clustering—putting all queries into a limited number of useful categories, such as by purpose, by topic, etc. But this requires use of advanced techniques such as natural language processing and machine learning. For now, continuing with our “low-hanging fruit” approach, we are conducting the easier task of looking at our top 100 queries. Despite that this only shows us a small percentage of queries representing a small percentage of searches, we make the assumption that our top 100 queries are in some ways representative of the whole corpus and can tell us basically the kinds of searches that people are doing. We have also attempted to normalize our search data in ways that we hope will give us further insights.

Search Query Analysis Methodology

We followed the same general methodology outlined in the beginning of this report and used the Google Analytics Reporting API to extract search queries for Summon and the catalog from URLs, using pageviews as an analog for the number of searches conducted. We extracted four separate groups of queries: all Summon search queries, just Summon queries launched from the *Find* box, all catalog queries (except those using course reserves and database indexes), and just catalog queries launched from the *Find* box. We parsed the data for each of these groups in different ways and generated the following six statistical reports.

Raw – Reports total pageviews for each raw search query as it was entered by the user.

Normalized – Minimally normalizes raw search queries and sums the pageviews for the resulting normalized queries. Normalization includes: case normalization, Unicode character normalization, punctuation removal, space normalization, removing advanced keyword search syntax from catalog queries, and removing stop words.

Words – Sums pageviews for each word that appears in each normalized query and reports all words. For instance, if the normalized query “library science” has 20 pageviews and the normalized query “library” has 5 pageviews, then the word “library” is included with 25 pageviews and the word science is included with 20 pageviews.

Bigrams – Like words, but with sets of two words. If the normalized query “new york times” has 20 pageviews, then the bigrams “new york” and “york times” are both included with 20 pageviews.

Trigrams – Like bigrams, but with sets of three words.

Query Length – An analysis of raw queries based on the number of words present in each. This report simply sums the total pageviews for each one-word query, each two-word query, each three-word query, etc.

Because this data mostly consists of columns of words, instead of putting tables in this report, we have put the data into a separate document to make it easier to view (see the *queries.xlsx* Excel file that accompanies this report). This data file includes only the top 100 queries for each group (Summon, all queries; Summon, *Find* box queries; catalog, all queries; and catalog, *Find* box queries) for each type of report plus some comparative statistics. The full data set is available on request.

Differences Between Summon and Catalog Queries

Looking through the sets of top 100 queries, words, bigrams, and trigrams, we see both similarities and differences between the queries people use when searching the catalog compared to Summon. We see that the top queries in the catalog—both from the *Find* box and otherwise—include a substantial number of searches related to video games (*video games, xbox, wii, ps3, call of duty, bioshock*, etc.) and films and television (*comedy films, star wars, erotic films, horror, people like us, fight club, breaking bad*, etc.). We also see searches in the catalog for particular resources (*quarterly journal speech, encyclopedia library information sciences, new york times, oxford english dictionary, dallas morning news, chicago manual style, wall street journal*, etc.). As we look at the normalized queries, words, bigrams, and trigrams, we also see a substantial number of music-related terms and concepts being searched in the catalog.

On the other hand, the top queries in Summon are more based on topics or current issues and events: *autism, social media, psychological science, play therapy, gun control, teen pregnancy*, etc.

These results are encouraging—again, keeping in mind that we’re assuming that the top 100 queries are more or less representative of the rest, this seems to show that users generally do understand the main differences between the catalog (Books & Media) and Summon (Online Articles).

Similarities Between Summon and Catalog Queries

Looking just at raw queries, we see that searches for names of databases or other aggregate resources are prevalent: *jstor, ieee, naxos, ebSCO, worldcat, artstor, rilm, lexis*, etc. Clearly these types of resources are important to our users.

If we look at the word, bigram, and trigram data and treat it as indicative of general topics that are often searched in each system, we see some additional similarities. There are 51 words that appear in both the list of top 100 words for Summon and for the catalog (all searches). Bigrams and trigrams show less overlap, but we still see some common topics and themes: library and information science-related terms; terms for particular periodicals such as *new york times, wall street journal*, and *quarterly journal speech*; and other terms such as *higher education, play therapy, behavior analysis, african american, civil war, criminal justice, early childhood, social work, social media, special education, civil rights movement, applied behavior analysis, early childhood education, english language learners, human resource management, second language acquisition, supply chain management, university north texas*, and *world war ii*.

Query Length

Looking at the number of searches in each system by the length of search query demonstrates additional differences between how people search the catalog and how people search Summon. Figure 11, below, shows a graphical representation of the distribution of query lengths for the four groups of searches.

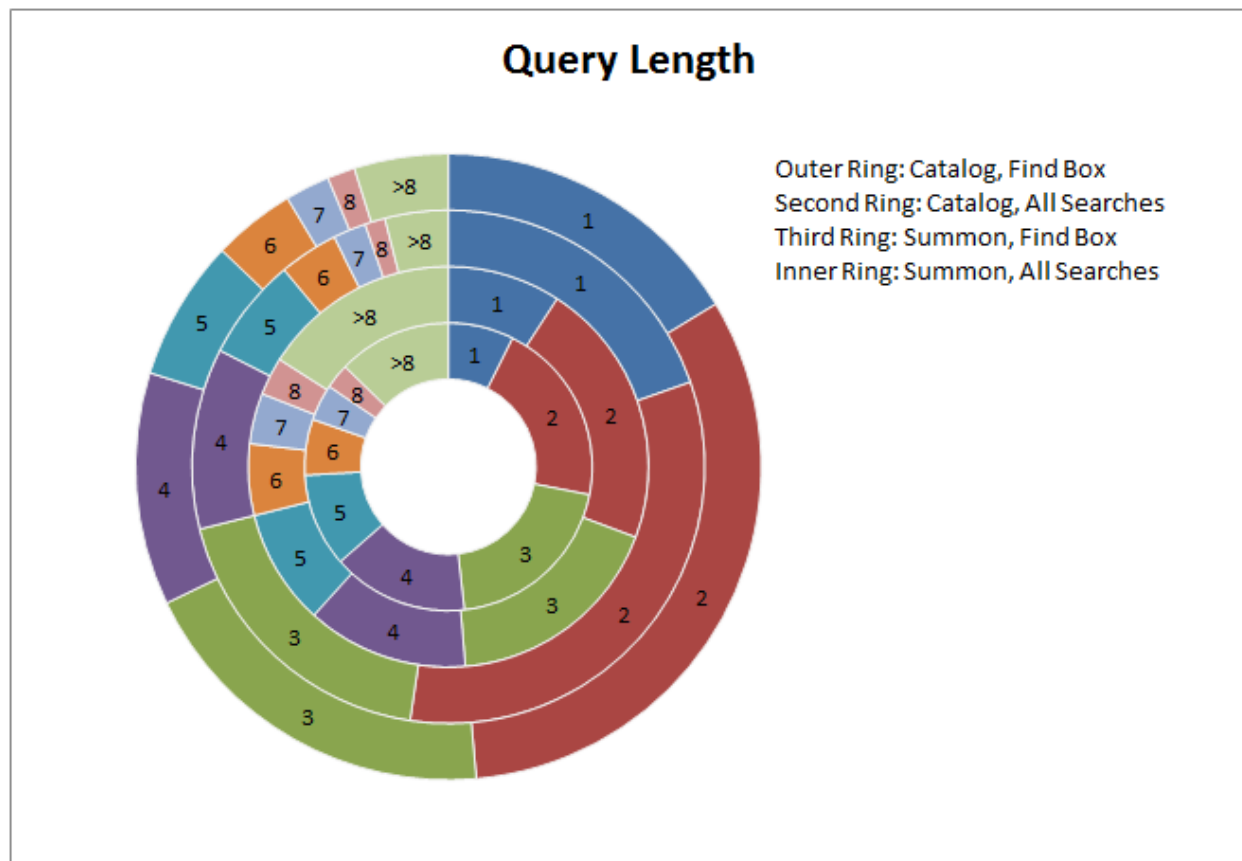


Figure 11: Query Length, Catalog versus Summon

In Figure 11, each numbered block corresponds to the proportion of searches using that number of search terms; the outermost ring represents catalog *Find* box searches, the next ring represents all catalog searches, the next ring represents Summon *Find* box searches, and the innermost ring represents all Summon searches.

We see from this graphic that both of the Summon groups tend to have longer search queries than the catalog groups. The catalog groups' distributions are more heavily weighted toward the front (one- and two-word queries), while the Summon groups' are more heavily weighted toward the end. The differences between the two systems seem to be greater than the differences between *Find* box searches and all searches.

Anecdotally, we understand that using Summon to find known items—articles, specifically—is common, and we think that the distribution of query lengths shows this to be the case, because longer queries

tend to be titles, citations, or full extracts from the text of a document. That isn't to say that the catalog isn't used for known-item searching; however, entering full citations into the catalog seems to be less common. This could also be another indication that people generally do use the Online Article search to find articles and the catalog search to find books, media, and periodicals, since article citations tend to be longer than citations for other types of resources.

Other Quantitative Statistics

In the Excel spreadsheet, we've captured a few additional statistics to help contextualize our query analysis. For each group and each type of data, we've included the total number of searches, the total number of searches represented by the top 100, and the percentage of searches represented by the top 100. We've also included statistics about the total number of unique queries.

The below graphs show, for each type of data, a comparison of the cumulative percentage of total searches that each query rank position represents between the four groups—all Summon searches, Summon *Find* box searches, all catalog searches, and catalog *Find* box searches.

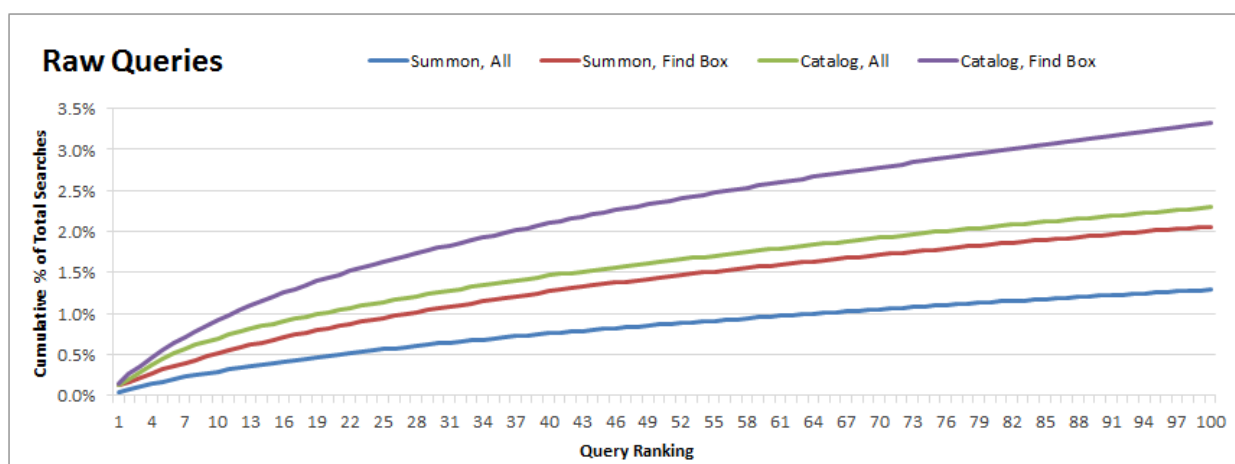


Figure 12: Raw Query Search Frequency Distribution, Top 100 Raw Queries

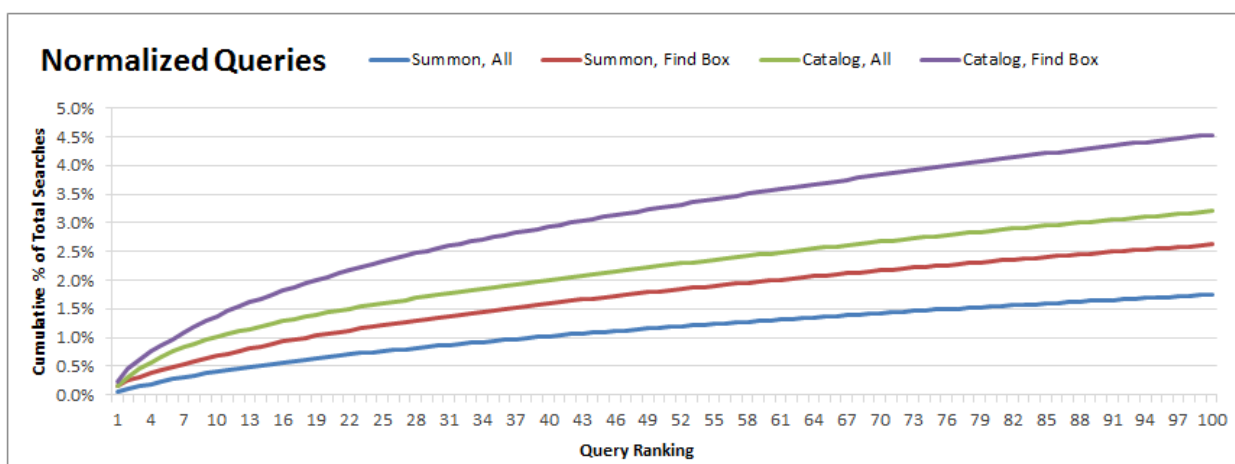


Figure 13: Normalized Query Search Frequency Distribution, Top 100 Normalized Queries

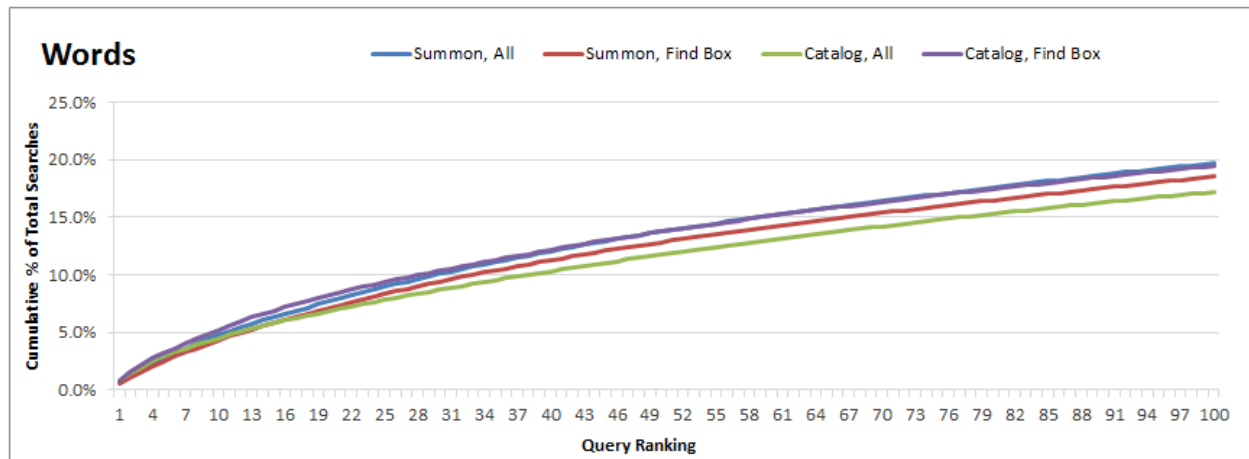


Figure 14: Word Search Frequency Distribution, Top 100 Words

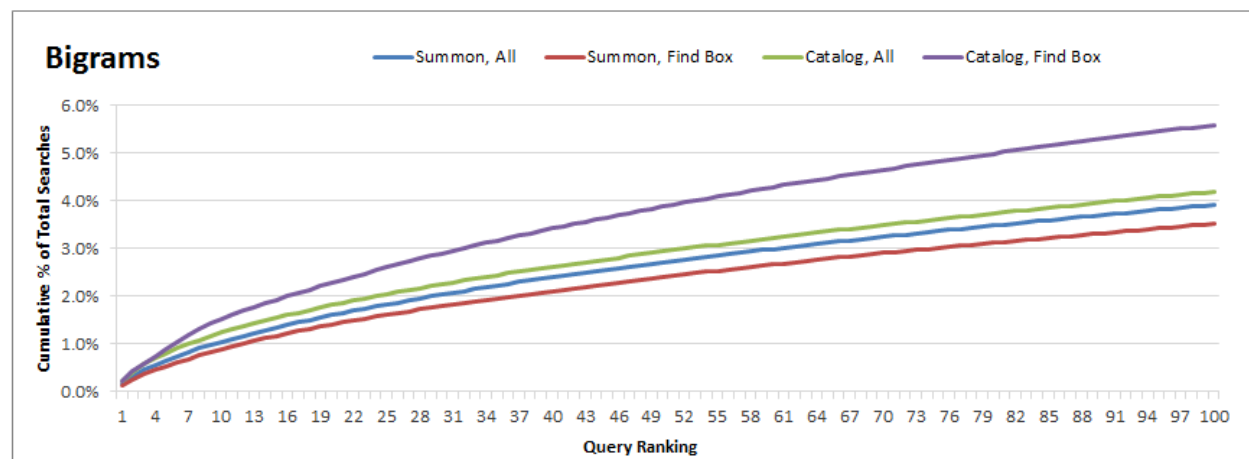


Figure 15: Bigram Search Frequency Distribution, Top 100 Bigrams

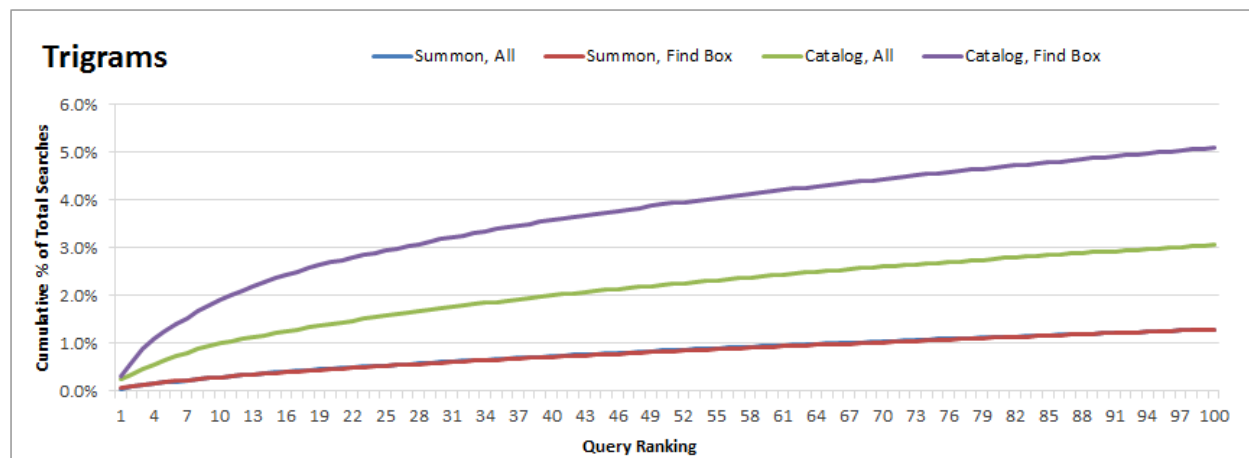


Figure 16: Trigram Search Frequency Distribution, Top 100 Trigrams

What do the graphs in figures 12 through 16 show? First, they show that catalog searches are more heavily weighted toward the top queries, especially catalog *Find* box searches, while Summon searches are more evenly distributed. This means that the top catalog queries are searched more often relative to other catalog queries than the top Summon queries.

Second, the graphs show that our normalization techniques did group queries better than simply looking at raw queries: we see that the top 100 normalized queries, words, bigrams, and trigrams do represent higher percentages of total searches compared to raw queries. But normalized queries are only marginally better, with a 1-2% increase, while words are much better, representing 17-19% of total searches for each group. This means the lists of top words are definitely more representative of their respective whole than the lists of raw and even normalized queries.

Summary and Conclusions

This report has considered a wide variety of usage data, and none of the results are too surprising. In fact, the majority seems to confirm that people use our discovery systems—at least, the two we’ve examined here—mostly as we expected and intended. Here are some general statements we can make about our systems that we think are supported by the available data.

By and large, users seem to understand searching one system for online articles and searching another for books, films, music, games, and other materials that they can get from the library.

First, Summon has not caused a wholesale switch away from the catalog. Although we *have* seen a decline in catalog searches since implementing Summon, this decline has not matched the surge in Summon use, and both systems have been used steadily since August 2012.

Second, the patterns in usage that we see in each system throughout the long semesters show that people are using the catalog more heavily at the beginning of the semester and Summon more heavily at the end. This supports that users see the catalog as the place to search for things you can get in the library (which require more time and effort to get) and Summon as the place to get online resources (which is what they need during the end-of-the-semester crunch time).

Third, Summon is used much more heavily by distance users; the catalog is used more heavily by local users. This lends further credence to the previous point.

Fourth, even though Summon does index some non-article content, our data shows that users overwhelmingly use it to search for article content. They tend not to change filters from our default Online Article search, and, when they do—or when they conduct a search that gets rid of the filters—they tend to switch them back.

Finally, our search query analysis, while it does not have the depth I would like, I think does clearly show that users tend to search the appropriate system for article content and for books/media content. Not only do we see many more catalog queries for movies, television, music, games, and books than

Summon queries, but we also see that queries in Summon are longer than queries in the catalog, which is what we would expect from an articles-search interface.

Summon usage seems to revolve around the academic calendar and academic use more than the catalog.

Summon use is high during each semester, drops off as soon as a semester ends, and doesn't pick up again until the next semester begins. Catalog use of course does drop off between semesters, but not as quickly. Aside from Christmas, there seems to be a higher baseline of catalog use.

From our search query analysis, we see that a lot of the top catalog searches are for video games, movies, television shows, and other things from popular culture. Although some of this use may be academic, it's perhaps safe to assume that much of this is for entertainment purposes. Aside from "video games," Summon has no game, film, or TV related terms in the top 100.

Media Library, which is where game, film, and TV-related resources are housed, is the most-used collection (or scope) in the catalog.

Users mostly search Summon from the *Find* box, while they mostly search the catalog from places other than the *Find* box.

From a resource discovery perspective, the catalog is more complex than Summon. While we've (necessarily) simplified the concept of what's in Summon (online articles) and the concept of what's in the catalog (books and media), *books and media* is perhaps more of a simplification of the catalog than *online articles* is a simplification of Summon.

It isn't surprising, then, that the catalog is mostly searched from elsewhere than the *Find* box. Many users use the catalog homepage as their starting point rather than the library homepage. Anecdotally, we know that there are smaller (but still significant) groups of users who use the Media Library and Music Library scope pages as their starting point. Even though we excluded databases and course reserves from our statistics, these also see significant use.

Moving forward, as we design our bento box search utility and we begin thinking about implementing interfaces to replace the catalog, we should be remain aware of the many types of use the catalog sees and ensure that we plan to accommodate that use.

Most users don't deviate from default searches and options.

Our data supports the known UI principle that users generally tend to stick with the defaults. When looking at *Find* box usage, we see that users rarely change the default options. When looking at Summon usage, we see that users use facets and filters less often than we would like, and they rarely use advanced searching. When looking at catalog usage, we see that the majority of usage is still keyword searches not limited to a particular collection.

Throughout this report we've assumed that patterns of continued use demonstrate that people are finding what they need and thus continuing to use our systems in similar ways. However, the strong tendency to stick with the default may overshadow the need to find the most relevant information; people don't deviate from established behaviors even when they might stand to benefit from change. So which is it? Does high use of our defaults show that our defaults are good, or are our defaults used simply because they are the default? I suspect that our defaults may be "good enough," but future research should try to untangle that web to ensure that we are implementing good and useful defaults.

The minority of users still represents a lot of searches, and we shouldn't ignore them.

When looking at usage data we're looking at what *most* users tend to do. For instance, when looking at usage of Summon, we see that 59.22% of searches use the default *Find* box search and never modify facets, start a new search, or use the advanced search. But the flip side of that percentage is that 40.88% of searches *do* modify facets, start a new search, or use the advanced search. That's 717,614 searches. Especially when the minority is still large, we need to make sure we delve into what the minority of users are doing so that we know how to accommodate them.

So, while it may be true that most users seem to understand and accept that Summon is used for online articles and the catalog is used for our physical collection, we see indications that this doesn't make sense to all users all the time. Even though we designed our UI to make it pretty clear that Summon is not the place to search for databases and journals, we still see that the top Online Articles searches are for *jstor*, *new york times*, *naxos*, *ebsco*, and *ebscohost*. In these cases, we need to try to ensure that people still get meaningful results. Fortunately, this is easy to do in Summon with their *best bets* feature; as we build the bento box search interface, we need to keep this in mind.