

Toward a Twenty-First Century Library Catalog

Kristin Antelman, Emily Lynema,
and Andrew K. Pace

Library catalogs have represented stagnant technology for close to twenty years. Moving toward a next-generation catalog, North Carolina State University (NCSU) Libraries purchased Endeca's Information Access Platform to give its users relevance-ranked keyword search results and to leverage the rich metadata trapped in the MARC record to enhance collection browsing. This paper discusses the new functionality that has been enabled, the implementation process and system architecture, assessment of the new catalog's performance, and future directions.

Editor's Note: This article was submitted in honor of the fortieth anniversaries of LITA and *ITAL*.

The promise of online catalogs has never been realized. For more than a decade, the profession either turned a blind eye to problems with the catalog or accepted that it is powerless to fix them. Online catalogs were, once upon a time, "the most widely-available retrieval system and the first that many people encounter."¹ Needless to say, that is no longer the case. Libraries cannot force users into those "closed," "rigid," and "intricate" online catalogs.² As a result, the catalog has become for many students a call-number lookup system, with resource discovery happening elsewhere. Yet, while the catalog is only one of many discovery tools, covering a proportionately narrower spectrum of information resources than a decade ago, it is still a core library service and the only tool for accessing and using library book collections.

In recognition of the severity of the catalog problem, particularly in the area of keyword searching, and seeing that Integrated Library System (ILS) vendors were not addressing it, the North Carolina State University (NCSU) Libraries elected to replace its keyword search engine with software developed for major commercial Web sites. The software, Endeca's Information Access Platform (IAP), offers state-of-the-art retrieval technologies.

Early online catalogs

Larson and Large and Beheshti summarize an extensive

Kristin Antelman (kristen_antelman@ncsu.edu), **Emily Lynema** (emily_lynema@ncsu.edu), and **Andrew K. Pace** (andrew_pace@ncsu.edu) are respectively Associate Director for the Digital Library, Systems Librarian for Digital Projects, and Head, Information Technology, at the North Carolina State University Libraries, Raleigh.

body of literature on online public access catalogs (OPACs) and related information-retrieval topics through 1997.³ The literature has tapered off since then; however, as promising innovations failed to be realized in commercial systems, mainstream OPAC technology stabilized, and the library community's collective attention was turned to the Web.

First generation online catalogs (1960s and 1970s) provided the same access points as the card catalog, dropping the user into a pre-coordinate index.⁴ The first online catalogs, byproducts of automating circulation functions, were "intended to bring a generation of library users familiar with card catalogs into the online world."⁵ The expectation was that most users were interested in known-item searching.⁶

With the second generation of online catalogs came keyword or post-coordinate (Boolean) searching. While systems based on Boolean algebra represented an advance over those that preceded them, Boolean is still a retrieval technique designed for trained and experienced searchers. (Twenty years ago, Salton wrote, "[T]he conventional Boolean retrieval methodology is not well adapted to the information retrieval task."⁷) Boolean systems were, however, simple to implement and economical in their storage and processing requirements, important at that time.⁸

Soon after the euphoria of combining free-text terms across records wore off, the library community recognized that the major problem with first- and second-generation catalogs was the difficulty of searching by subject.⁹

The "next-generation" catalog

By the early 1980s, thinking turned to next-generation catalog features.¹⁰ Out of this surge of interest in improving online catalogs emerged a number of experimental catalogs that incorporated advanced search and matching techniques developed by researchers in information retrieval. They typically did not rely on exact match (Boolean) but used partial-match techniques (probabilistic and vector-based). Since probabilistic and vector-based models were first worked out on document collections, not collections of MARC records, adaptations were made to the models.¹¹ These prototype systems included Okapi, which implemented search trees, and Cheshire II, which refined probabilistic retrieval algorithms for online catalogs.¹² It is particularly sobering to revisit one system that was developed between 1979 and 1983. The CITE catalog, developed at the National Library of Medicine, incorporated many of the features of the Endeca-powered catalog, including suggesting (MeSH) subject headings, correcting spelling errors, stemming, as well as even more advanced features, such as term weighting, keyword suggestion, and "find similar."¹³

Where are we now?

As Belkin and Croft noted in 1987, “there is a disquieting disparity between the results of research on IR techniques . . . and the status of operational IR systems.”¹⁴ Two decades later, libraries are no better off: all major ILS vendors are still marketing catalogs that represent second-generation functionality. Despite between-record linking made possible by migrating catalogs to Web interfaces, the underlying indexes and exact-match Boolean search remain unchanged. It can no longer be said that more sophisticated approaches to searching are too expensive computationally; they may, however, be too expensive to introduce into legacy systems from a business perspective.

The Endeca-powered catalog

Coupled with the relative paucity of current literature on next-generation online catalogs is a scarcity of library industry interfaces from which to draw inspiration, RLG’s Red Light Green and OCLC’s FictionFinder being notable exceptions.

In June 2004, library automation vendor TLC announced a partnership with Endeca Technologies for joint sales, marketing, technology, and product development of the company’s IAP software. This search software underlies the Web sites of companies such as Wal-Mart, Barnes and Noble, IBM, and Home Depot.

NCSU Libraries acquired Endeca’s IAP software in May 2005, started implementation in August, and deployed the new catalog in January 2006.

Several organizational and cultural factors contributed to making this project possible. Of significance was an ongoing administrative commitment to fund digital-library innovation, including projects that involve some risk. Library staff share this feeling that calculated risks are opportunities to improve the library as well as to open up new challenges in their own jobs. Critically, they also believe that not all issues, particularly “edge cases,” (i.e., rarely occurring scenarios) must be resolved before releasing a new service. Finally, it was important that the managers who controlled access to programming and other resources were also the project leaders and drivers of the collective urgency to solve the underlying problem. All these factors also contributed to making possible a five-month implementation timeline.

Functionality

The principle functionality gained by implementing an advanced search-and-navigation technology such as the

Endeca IAP falls in three main areas: relevance-ranked results, new browse capabilities, and improved subject access. Most ILSs, including NCSU’s former catalog, presented keyword results to users in one order: last-in, first-out (i.e., system sort), while browsing within keyword result sets was limited to the links within individual records.

Searching and relevance ranking of results

Inhabiting the catalog search landscape now, somewhere between a second- and third-generation catalog, is Endeca’s MDEX Engine, which is capable of both Boolean and limited partial-match retrieval. Queries submitted to Endeca can use one of several matching techniques (e.g., matchall, matchany, matchboolean, matchallpartial). The current NCSU implementation primarily uses the “match-all” technique for keyword searching, an implied AND technique that requires that all search terms (or their spell-corrected, truncated form) entered by the user occur in the result. The user is not required to enter Boolean operators for this type of search; in fact, these terms are discarded as stopwords. The “matchboolean” technique continues to support true Boolean queries with standard operators; access to this functionality is provided through advanced search options.

Although classic information retrieval research tends to associate relevance ranking with probabilistic or vector-based retrieval techniques, Endeca includes a suite of relevance ranking options that can be applied to Boolean-type searches (i.e., implied AND/OR). These individual modules are combined and prioritized according to customer specifications to form an overall relevance ranking strategy, or algorithm.

Each search index created in the Endeca software can be assigned a different relevance ranking strategy. This capability becomes significant when considering the differences in the data being indexed for ISBN/ISSN as compared to a general keyword search. Since the Keyword Anywhere index contains the majority of the fields in a MARC record and is the default search operator, its relevance ranking strategy received the most attention. This strategy currently consists of seven modules. The first five modules rank results in a dynamic fashion, while the final two modules provide static ordering based on publication date and total circulation.

The NCSU Libraries, algorithm prioritizes results with the query terms exactly as entered (no spell-correction, truncation, or thesaurus matching) as most relevant. For multiterm searches, results containing the exact phrase are considered more relevant than those that do not. In addition, NCSU has created a field priority ranking, which

provides the capability to define matches that occur in the title as more relevant than matches that occur in the notes fields. The relevance algorithm also considers factors such as the number of times the query appears in each result and the term frequency/inverse document frequency (tf/idf) of query terms.

The unprecedented nature of using this particular set of tools to define relevance algorithms in library catalogs meant that the initial configuration required a best guess approach. The ability to quickly change the settings and re-index provided the opportunity both to learn by doing and test assumptions. Much work remains, however, including systematic testing of the “matchallpartial” retrieval technique. While not a true probabilistic or vector-based matching approach, the “matchallpartial” retrieval technique will broaden a search by dropping individual query terms if no results are returned. However, this type of retrieval technique creates the challenge of developing an intuitive interface that helps users understand partial matching (although many users must be aware that this is how Google works).

Spell correction, “Did you mean . . .,” and sort

Several other features are included in the basic Endeca IAP application. These include auto-correction of misspelled words, which uses an index-based approach based on frequency of terms in the local database rather than a dictionary. Due to the presence of unique terminology in the database (particularly author names), the relevance ranking has been configured to display any matches on the user’s original term before spell-corrected matches. A “Did you mean...” feature also checks queries against terms indexed within the local database to determine if another possible term has more hits than the original term in order to provide the user the option to resubmit the search with a different spelling. Various sort options are supported, including date, title, author, and “most popular.”

■ Browse

Whatever the shortcomings of the card catalog, a library user could approach it with no query in mind; any drawer could be browsed. With the advent of online catalogs, this is no longer possible: an initial search is required to enter the system.

Marchionini characterizes “browsing strategies” as “informal and opportunistic.”¹⁵ A good catalog browse should simulate the experience of browsing the stacks, even potentially improving upon it since the virtual browser can jump around. Many patrons cite the serendipity of browsing the stacks and “recognizing” relevant resources as a key part of their discovery process. With

more books moving to online formats and off-site storage (and therefore, unable to be browsed), enhancing virtual browsing in the catalog becomes increasingly important.

As Borgman points out, “Few systems allow searchers . . . to pursue non-linear links in the database.”¹⁶ Key browsing features provided by the Endeca software are faceted navigation and the ability to browse the entire collection without entering a search term.

Although most modern search engines support both fast response times and relevance ranking, the opportunity to apply Endeca’s Guided Navigation feature to the highly structured MARC record data was particularly intriguing. Guided, or faceted, navigation exposes the relationships between records in the result set. For example, a broad topical search might return thousands of results. Classification codes, subject headings, and item-level details can be used to define logical clusters for browsing—post-coordinate refinement—within the result set. Since these refinements are based on the actual metadata of the records in the result set, users can never refine to less than one record, (i.e., there are no “dead ends”). These clusters, or facets, are known as dimensions. Users are able to select and remove values from all available dimensions in any order to assist them as they browse through the result set.

Endeca’s dimensions, while able to be browsed, are not available only as post-coordinate search refinements, however. Using the Endeca application, library catalogs can once again give users the ability to browse the entire set of records without first entering a search term. Any of the dimensions can be used to browse the collection in this fashion, and the ability to assign item-level information (e.g., format, availability, new book), as well as bibliographic-record elements, to the dimensions further enhances the browsing functionality.

■ Improving subject access

Given the unsuitability of Library of Congress Subject Headings (LCSH) as an entry vocabulary, improving topical (subject) access in catalogs centers around keyword searching. While keyword searches query the subject headings as they do the rest of the record, most systems do not take advantage of the fact that subject headings are controlled and structured access points or use the subject information embedded in the classification number.

The Endeca-powered catalog, in addition to addressing classic keyword-search problems by introducing relevance ranking, implied phrase, spell correction, and stemming, also leverages the “ignored” controlled vocabulary present in the bibliographic records—subject headings and classification numbers—to aid in improving topical searching. This is a system design concept that has been discussed in the literature on improving subject

access but has not until now been manifest in a major catalog implementation.

As Chan noted, “subject headings and classification systems have more or less operated in isolation from each other.”¹⁷ The Endeca-powered catalog interface is an experiment in presenting users with these two different, but complementary, approaches to categorizing library materials by subject.

Classification

Several catalog experiments created retrieval clusters based on Dewey- and DDC-classification schemes and captions in order to improve subject access by expanding the entry vocabulary and as a way to improve precision and recall.¹⁸ Using the LC Classification is more challenging, however, as it is not hierarchical. Still, the potential of its use has been noted by Bates and Coyle; and Larson experimented with creating clusters (“classification clusters”) based on subject headings associated with a given LC class.¹⁹ In Larson’s system, the interface suggested possible subject headings of interest, an approach similar to that of displaying the subject facets alongside the result set in the Endeca catalog.

There is some evidence from early usability studies that exposing the classification, much as it was physically exposed in the card catalog, is useful and desired by catalog users. Markey summarizes findings of a 1981 Council on Library Resources study in which many institutions conducted usability testing. Positive aspects of card-catalog use that people wanted to see in the OPAC included, a “visual overview of what is available in the library,” and “serendipity.”²⁰ But there is a difference between using the classification scheme to identify subject headings and displaying the classification itself in the user interface. The latter can be problematic from a usability perspective, as Larson pointed out, because the classification scheme and terminology are not transparent.²¹ Imagine the would-be browser of a library’s computer-science collection having to know to select first Q Science, then QA1–QA939 Mathematics, and then QA71–QA90 Instruments and Machines before possibly recognizing that QA75–QA76.95 Calculating Machines included computer science?

Despite these potential problems, because the Endeca software supported display of the LC Classification as a dimension, NCSU decided to experiment with its utility by making it available on the results screen.

Entry vocabularies

Entry vocabularies or mappings apply to all types of retrieval models. They address the general problem of reconciling a user’s query vocabulary with the index vocabulary represented in the catalog or documents.²²

Studies show that users’ query vocabulary is large

(people rarely pick the same term to describe the same concept) and inflexible (people are unable to repair searches with synonyms.)²³ Because of this, Bates refers to the objective of the entry vocabulary as the “Side-of-a-Barn Principle.”²⁴

Several approaches have been taken to develop this functionality. Building on Larson’s “classification clustering” methodology, Buckland created an Entry Vocabulary Module by associating dictionaries created by analyzing database records.²⁵ The result was natural language indexes to existing thesauri and classification systems.

While the Endeca-powered catalog does not yet incorporate an entry vocabulary, its exposure of the index vocabulary to the user in subject dimensions could be said to be a limited side-of-a-barn approach. The limitation is that only controlled vocabulary from the retrieved records is exposed as dimensions on the results screen; relevant records not retrieved because of a lack of match between query vocabulary and terms in the record will not have their facets displayed. Were an entry vocabulary for LCSH available, Endeca’s synonym-table feature could be used to map between query terms and LCSH.

Implementation

The library’s Information Technology Advisory Committee appointed a seven-member representative team to oversee the implementation. Preparatory steps included sending key development staff to training and a two-day meeting with Endeca project managers to establish functional and technical requirements.

Architecture

Knowing that the Endeca application would not completely replace NCSU’s integrated library system, determining how best to integrate the two products was part of the implementation process. The Endeca IAP coexists with the SirsiDynix Unicorn ILS and the SirsiDynix (Web2) online catalog, indexing MARC records that are exported from Unicorn’s underlying Oracle database. Figure 1 depicts the integration of the Endeca software with existing systems.

Although the Endeca software is capable of communicating directly with the database that supports the Unicorn ILS, NCSU chose the easier path of exporting MARC records into text files for ingest by Endeca. The MARC4J API is used to reformat the exported MARC records (which include item-level information in 999 fields) into flat text files with UTF-8 encoding that are parsed by Endeca’s Data Foundry process. Nightly shell scripts export updated and new records from ILS, merge those with the base Endeca files, and start the re-indexing process. The indexing of seventy-three MARC

record fields and ten dimensions results in an index size of approximately 2.5 GB. The entire index resides in system memory. The Endeca Data Foundry can easily parse and re-index the approximately 1.7 million titles in NCSU's holdings nightly (in stark contrast to the more than 3 days of downtime required to re-index keywords in Unicorn). The relative speed of this process and the fact that it does not interfere with the front-end application prompted the decision not to implement "partial indexing" at the outset.

Though there was little doubt among staff as to the increased capabilities of keyword searching through Endeca, the implementation team decided that authority searching (author, title, subject, call number) would be preserved in the new catalog interface. This allowed NCSU to retain the value of authority headings, in addition to providing a familiar interface and approach to known-item searching. Since the detailed record in Web2 included the capability to save records, place requests, and send system-suggested searches ("more like this"), the implementation team also decided to link from titles in the Endeca-powered results page to the Web2 detailed record. Only slight modifications were required to stylize this display in a manner consistent with the new interface.

The front-end interface for keyword searching in Endeca is a Java-based Web application built in-house. This application is responsible for sending queries to the Endeca MDEX Engine—the back-end HTTP service that processes user queries—and displaying the results that are returned.

User-interface design

Because it is created by the customer, NCSU Libraries has complete control over the look, feel, and layout of the Endeca search-results page.

Indexes, properties, and dimensions

The implementation team began the process of making indexing decisions by looking at the fields indexed in the Unicorn keyword-index file. This list included 161 MARC fields and subfields, including more than thirty fields that are never displayed to the public. This kitchen-sink approach was replaced with a more carefully selected list less than half that number.

The implementation team defined eleven dimensions for use with Endeca's faceted navigation feature. Once users enter a search query, they can explore the result set by selecting values from these dimensions: Availability; LC Classification; Subject: Topic; Subject: Genre; Library; Subject: Region; Subject: Era; Language; and Author (see figure 2). The eleventh dimension is not displayed on the results page, but is used to enable patrons to browse new titles. Each dimension value also lists the number of results associated with it; most dimensions are listed in frequency order.

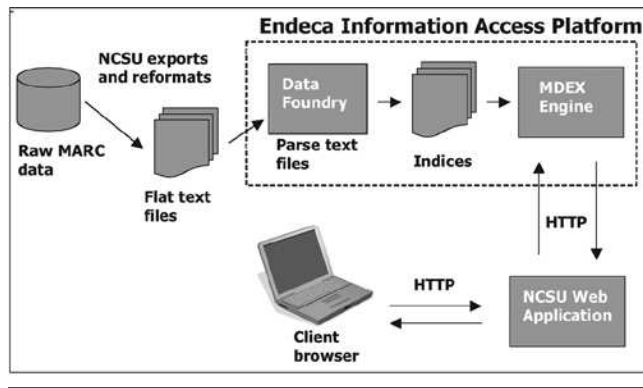


Figure 1. NCSU Endeca architecture

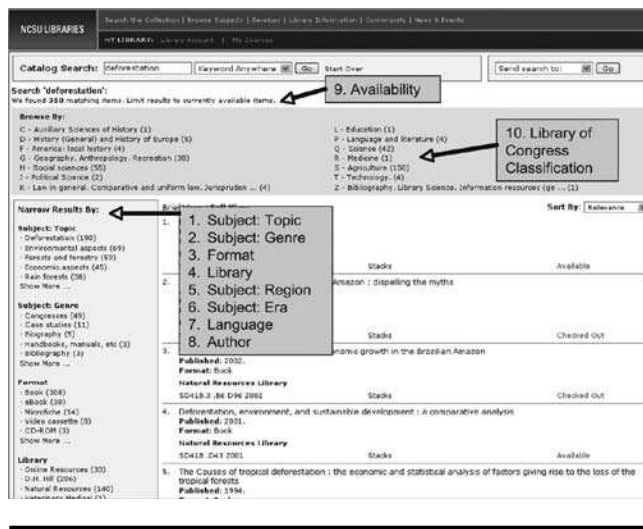


Figure 2. Dimensions

Search interface

Once the implementation team made some preliminary decisions regarding dimensions and search indexes, wireframes were created to assist in the iterative design process for the front-end application. While the positioning of the dimensions on the results page and the display of holdings information was well debated, the design of the catalog search page was an even hotter topic. Integration of both Endeca keyword searching and Web2 authority searching required an interface that could help users differentiate between the two tools.

A survey of the keyword-versus-authority searching distinction in a variety of library catalogs led to the development of four mock-ups. The implementation team chose a Search tab that includes separate search boxes for keyword and authority searching, as well as search

examples dynamically displayed based on the index selected. Authority searching was relabeled “Begins with” searching to let users know that this particular search box featured known-item searching (although it is also where LCSH searching is found) (see figure 3).

An Advanced Search tab re-creates the pre-coordinated search options from the Web2 search interface using Endeca search functionality. One unique new feature allows users to include or exclude reference materials and government documents from their results. A true Boolean search box is made available here, primarily for staff.

Browse

While users can submit a blank search and browse the entire collection by any of the dimensions, the Browse tab specifically supports browsing by LC Classification scheme (see figure 4).

This tab also includes a “New Titles” browse that can easily be refined with faceted navigation. At the time of this writing, there are plans to pull out other dimensions, such as format, language, or library, for browsing. This will be a great stride forward since there has traditionally been no way to perform a MARC codes-only search (in order to browse all Chinese fiction in the main library, for example).

Assessment

The Endeca-powered catalog seems self-evidently a better tool to help users find relevant resources quickly and intuitively. But since so much of the implementation involved uncharted territory, plans for assessment began before the launch of the interface, and the actual assessment activities began shortly thereafter. The library identified five assessment measures prior to implementation. One of these, however, requires longer time-series data (changes in circulation patterns), and another, the application of new and potentially complex log-analysis techniques (path

analysis). Other measures relate to use of the refinements, “sideways searching,” and objective and subjective measurements of quality search results, some of which can be preliminarily reported on here.

Log analysis

To learn more about how patrons are using the catalog, data from two months of search logs were analyzed. While authority searching using the library’s old Web2 catalog is still available in the new interface, search logs show that authority searching has decreased 45 percent and keyword searches have increased 230 percent. It is noted, however, that a significant—and indefinable—component of this increase in keyword searching is due to the fact that the default catalog search was changed from title to keyword.

Users are taking advantage of the new navigational features. Fifty-five percent of the Endeca-based search requests are simple keyword searches, 30 percent represent searches where users are selecting post-search refinements from the dimensions on the results page, and the remaining 15 percent are true browses with no search term entered (this figure includes use of Browse New Titles).

Dimensions

The horizontal space just above the results is used to display the full range of results within the LC Classification scheme (see figure 2). The first dimensions in the left column focus on the subject dimensions (topic and genre) that should be pertinent to the broadest range of searches. The following format and library dimensions recognize that patrons are often limited by time and space. When designing the user interface, it was not known which dimensions would be most valuable. As it turned out, dimension use does not exactly parallel dimension placement. LC Classification is the most heavily used, followed closely by Subject: Topic, and then Library, Format, Author, and Subject: Genre. Since no basis for the placement of dimen-



Figure 3. New catalog search interface



Figure 4. Browse by LC Classification and new titles

sions existed at the time of implementation, the Endeca product team plans to use these data, after some time, to determine if changes in dimension order are warranted.

Spell correction and “Did you mean . . .”

Approximately 6 percent of Endeca keyword searches responded to the user’s query with some type of spelling correction or suggestion: 3.6 percent performed an automatic spell correction, and 2.8 percent offered a “Did you mean...” suggestion. While NCSU has not analyzed how many of the spell corrections are accurate or how many of the “Did you mean...” suggestions are being selected by users, future work in this area is planned.

Recommender features

Two features in Endeca that have seen a surprising amount of use are the “most popular” sort option and the “more titles like this” feature available on the detailed-record page for a specific title. Both relate broadly to the area of recommending related materials to patrons.

The “most popular” sort option is currently powered by aggregated circulation data for all items associated with a title. While this technique is ineffective for serials, reference materials, and other noncirculating items, it provides users a previously unavailable opportunity to define relevance. To date, the “most popular” sort is the second most frequently selected sort option (after publication date, at 41 percent), garnering 19 percent of all sorting activity. Most-popular sorting was trailed by title, author, and call-number sorting.

When viewing a detailed record, users are given the option to find “more titles like this” or “more by these authors.” The first option initiates a new subject keyword search combining the phrases from the \$a subdivision of all the subject (6xx) fields assigned to the record. The latter option initiates an author keyword search for any of the authors assigned to the current record. While there are not good statistics on use of this feature, these subject strings appear regularly in the list of most popular queries in search logs.

Assessing top results

If relevance ranking was effective, one would expect to see good results on the first page. But what are “good” or “relevant” results? Greisdorf finds that topicality is the first condition of relevance, and Xu and Chen’s more recent study finds topicality and novelty to be equally important components of relevance.²⁶ While someone other than the searcher might be able to assess topical relevance, it is impossible to assess novelty, since it cannot be known what the searcher already knows.

Although researchers agree that relevance is subjective—that is, only a searcher can determine whether results are relevant—Janes showed that trained external

searchers do a reasonably good job of approximating the topical relevancy judgments of users.²⁷ The analysis reported here focuses on topicality (using a liberal interpretation of what might be topically relevant). NCSU Libraries sought to measure how many of the top search results are *likely* to be relevant to the user’s query in the old and new catalogs.

Methodology

One of the authors searched 100 topical queries (taken from 2005 search logs) in both Web2 and Endeca catalogs using “keyword anywhere.” Topical queries whose meaning was unclear (e.g., “hand wrought”) were excluded. The topical relevance of the top hits (up to five) was coded for each target.

Because not all search-result sets contained five records, success for each was measured as a ratio (e.g., $2/5 = .4$). Those searches that resulted in 0 records in both targets were discarded, while those that resulted in 0 records in target *a* but “found relevant results” in target *b* were counted as 0 in target *a*. The ratios were then averaged for each target and compared to determine the difference in relevance-ranking performance.

Finally, a random subset of forty-four of the queries was selected, and the placement in the Web2 results of the first result in Endeca was noted.

Results

On average, 40 percent of the top results in Web2 were judged to be relevant, while 68 percent of the top results in Endeca were judged to be relevant. That represents a 70 percent better performance for the Endeca catalog. If one makes the assumption that the first Endeca record is relevant (admittedly an assumption), based on these data, then one can look at the average position of that record in the old catalog. It was found that the first hit in Endeca fell between #1 and #4126 in Web2, with more than a third falling after the second screen of results, the maximum number of screens users are typically willing to examine.²⁸

While this level of increased performance is impressive, it masks some dramatic differences in the respective result sets. Looking at a broad search, “marsupial,” all of the top five hits in Endeca have “marsupial” in the title and “marsupials” or “marsupialia” as a subject heading. The result set includes seventy-eight records, thanks to this intelligent stemming. In the Web2 result set, just twenty-nine records, not a single one of the top five has “marsupial” in the title or subject headings (and the top two results, *Tributes to Malcolm C McKenna* and *Poisonous plants and related toxins*, are highly unlikely to be relevant). It is not until record #10 that you see the first item that contains “marsupial” in the title or subject. This single example demonstrates the benefit of both relevance ranking and stemming.

Usability testing

As a result of a long history of catalog-usability studies, there are things that are known about library catalog users. One is that people both expect systems to be easy to use and find that they are not.²⁹

Usability testing was conducted to compare student success in using the new catalog interface with that of students using the old catalog interface when completing the same set of ten tasks. Ten undergraduate students were recruited for the test. Five were randomly selected to use the old Web2 catalog, while the other five used the new catalog interface, which allows users to choose between a keyword search box powered by Endeca and an authority search box (begins with . . .) that is still powered by Web2. The test contained four known-item tasks and six topical-searching tasks (appendix A).

Task success, duration, and difficulty were recorded. User satisfaction was not measured since catalog usability studies have found that satisfaction does not correlate with success.³⁰

Task duration

Figure 5 shows the average task duration for the topical tasks (5–10) for Web2 and Endeca. Except for task 9, there is clearly a trend of significantly decreased average task duration for Endeca catalog users. The Endeca catalog shows a 48 percent improvement in the average time required to complete a task (01:34 in Web2 compared to 00:49 in Endeca). It is also noted that, although results from known-item searching tasks (1–4) are not reported in detail here, test subjects were just as successful in completing them using keyword searching in the Endeca catalog as they were using authority searching in Web2.

Task success and difficulty

In addition to task duration, the test moderator assigned a difficulty rating to each task attempted by the participants: easy, medium, hard, or failed. Figure 6 illustrates the overall task-attempt difficulty for topical tasks (5–10) in the Web2 and Endeca catalogs. The largest improvement is in the increased percentage of tasks that are completed easily in Endeca and the nearly equivalent decrease in the percentage of tasks that were rated as hard to complete. While a significant number of tasks were still failed using the Endeca catalog, many of these failures can be attributed to participants' propensity to select Keyword in Subject rather than Keyword Anywhere searches. In fact, the only instances where Keyword Anywhere search in the new catalog failed to lead to successful task completion were for a single participant who was unwilling to examine retrieved results closely enough to determine if they were actually relevant to the task question, assuming too quickly that the task had been completed successfully.

Terminology

Participants using both the Web2 and Endeca catalog interfaces expressed confusion over some of the terminology employed. One of the most problematic terms was "subject." A number of participants selected Keyword in Subject for topical searches because of the attraction of the word "subject." None of the participants recognized that this term referred to controlled vocabulary assigned to records. Coupled with a slight unfamiliarity with the term "keyword," not typically used in Web searching, this misunderstanding led participants to misuse (or overuse) Keyword in Subject searches when they could have found results more effectively using general keyword searching. This terminology problem appears to be an artifact of the usability testing, however. Looking at the search logs, more than 50 percent of the keyword searches were Keyword Anywhere searches, while only 4 percent represented Keyword in Subject searches.

Relevance

Relevance ranking of search results is clearly the most important improvement in the new catalog. Students in this usability test all looked immediately at the first few results on the first page to determine if their search had produced good results. If they didn't like what they saw, they were likely to retry the search with fewer or more keywords in order to improve their first few results. One participant

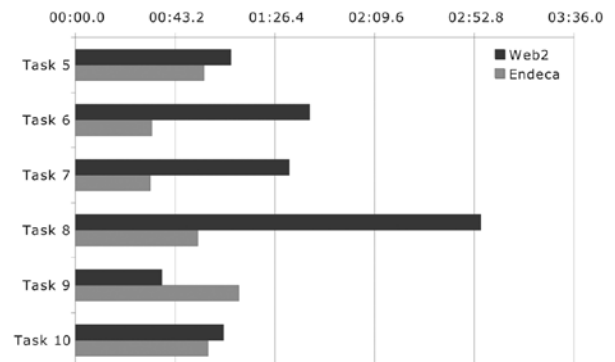


Figure 5. Average task duration: Web2 versus Endeca

* While task 9 may appear to be an aberration, it actually reveals effective use of new functionality. This task required users to locate an audio recording of poetry in Spanish. In Web2, three of five participants completed the task successfully, all using the material type and language limits available in the advanced search tab. The two participants who didn't locate this tool failed to complete the task. In Endeca, two participants used the same advanced search limits to complete the task successfully and two additional participants were able to locate and use Endeca dimensions to complete the task successfully. This suggests that the new interface is providing users with more options to help them arrive at the results they seek.

using the Web2 catalog expressed the need for relevance ranking, "Once I scroll through a page, I get pretty discouraged about the results."

The number of paging requests recorded in system logs confirms that users are focusing on the first result screen (with ten results per page); only 13 percent of searchers go to the second page.

Use of dimensions

When questioned after the test, all five participants who used the Endeca catalog intuitively understood that dimensions could be used to narrow results. However, only three used the dimensions during the test.

Throughout the tests, the student participants frequently attempted to limit their search at the outset, rather than beginning with a broad search and then refining. It is unclear whether this behavior is a function of the very specific nature of the test questions or experience with the old catalog. Log data show that users are indeed entering broad keyword searches with only one or two terms, which implies that dimensions may be more useful than this usability test indicates. It is also interesting to note that while none of the students understood that the LC Classification dimension represented call-number ranges, they did understand that the values could be used to learn about a topic from different aspects—science, medicine, education.

Future directions

Weeks before the initial application went live in January 2006, the list of desired features had grown long. Some of these were small "to do" items that the team did not have time to implement. Others required deeper investigation, discussion, and testing before the feature could be put into production. Still others may or may not be possible. A few of NCSU's significant planned development directions are summarized below.

Functional Requirements for Bibliographic Records

There is much interest in the utility of applying the

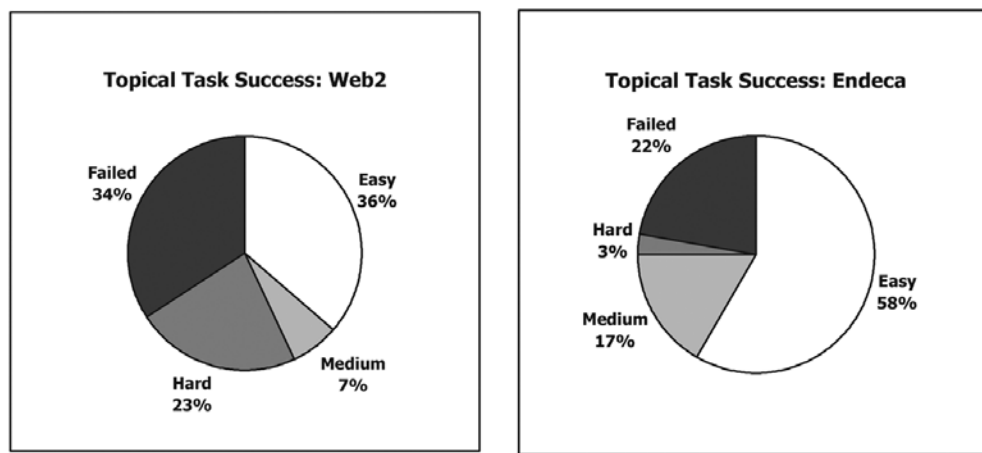


Figure 6. Topical task success and difficulty: Web2 versus Endeca

Functional Requirements for Bibliographic Records model to online catalogs.³¹ Endeca includes a feature called "record rollup" that allows retailers to group items together for example, different sizes and colors of a shirt. All that is required for this feature is a rollup key. NCSU, working with OCLC, has elected to try the OCLC work identifier to take advantage of this functionality and create work-level record displays in the Endeca catalog hit list.

Subject access

The collective investment libraries have made in subject and name authorities is leveraged with the faceted navigation features of Endeca. But only authorized headings in records are seen by Endeca, cross-references in the subject-authority record are not used. During implementation, the team looked at ways to improve the entry vocabulary to authorized-subject terms by loading the 1xx and 4xx fields from the subject-authority file into Endeca synonym tables so that users could be guided to proper subject terms. The team still views this as a promising direction, but simply did not have time to fully explore it prior to implementation.

Additional discussions with OCLC centered on their Faceted Access to Subject Terms (FAST) project. FAST terms are more amenable than LCSH headings to being broken up into topical, geographic, and time-period facets without losing context and meaning. The normalization of geographic and time-period subdivisions promises to be particularly useful. FAST has, to date, lacked a ready interface for the application of its data. While the FAST structure is more conducive to non-cataloger metadata creation and post-coordinate refinement, it still does not meet the need

for a user-entry vocabulary.³² Were such a vocabulary for LCSH to become available, it could be mapped to synonym tables to lead users to authorized headings.

Abandon authority searching?

The future of authority *searching*, however, is less clear. Although the usability testing described in this paper showed that the Endeca keyword search tools performed on a par with the old catalog for known-item searching, it is recognized that authority searching serves more functions. Clearly, collocation of all books on a topic is absent when a user does a topical search using keyword rather than a controlled subject heading. But there are more subtle losses as well. As Chan points out, one purpose of subject access is to help users focus searches, develop alternative strategies, and enable recall and precision.³³ This is not possible with a simple keyword search, unless the searcher discovers that he can search on a subject heading from a record of interest. The display of subject facets in the Endeca-powered catalog works to counter this weakness of simple keyword searching.

Another navigation aid in the traditional authority display that is lost in a simple keyword-search result is visible “seams.” As Mann points out, “Seams serve as perceptible boundaries that provide points of reference; without such boundaries, readers get ‘lost at sea’ and don’t know where they are in relation to anything else: they can’t perceive either the extent of what they have, or what they don’t have.”³⁴ Until users have confidence that a known item will appear at the top of a results list if the library holds that item, with a large keyword result set, one cannot confirm a “negative result” without browsing through the entire set. The Endeca-powered catalog interface does not help to address either the “seams” or the negative-result problem, which are two reasons why NCSU maintained authority searching.

An integration platform

Despite the vast improvements found in the Endeca catalog, the fact remains that it is still mainly books—as Calhoun says, “only a small portion of the expanding universe of scholarly information.”³⁵

There are two approaches to take with the Endeca platform: one is to take advantage of having control over the data and the interface to facilitate incorporation of outside data sources to enhance bibliographic records. The second is to put other, non-catalog data sources under the Endeca search-and-navigation umbrella. The middleware nature of the Endeca platform makes either approach more promising than the “square peg and round hole” problem of trying to work with library management systems ill-equipped to handle a diversity of digital assets. Whether as a feed of catalog data to a metasearch application or Web-site search tool, or as a platform for faceted access to

electronic theses, institutional repositories, or electronic books, Endeca has clear potential as a future platform for library resource discovery.

Conclusion

While it cannot be claimed that this Endeca-powered catalog is a third-generation online catalog, it does implement a majority of the third-generation catalog features identified by Hildreth. Most notably, through navigation of subject and item-level facets, the Endeca catalog supports two of his objectives, “related record search and browse” and “integration of keyword, controlled vocabulary, and classification-based approaches.” Spell correction, intelligent stemming, and synonym tables support “automatic term conversion/matching aids.” The flexible relevance-ranking tools support “closest, best-match retrieval” as well as “ranked output.” Much work remains, however. Three important features identified by Hildreth cannot be said to be implemented in this catalog at this time: “natural language query expression,” that is, an entry vocabulary, “expanded coverage and scope,” and “relevance feedback methods.”³⁶ Requirements for these features are either being reviewed or are already under development by both Endeca and NCSU Libraries.

NCSU views the Endeca catalog implementation in the context of a broader, critical evaluation and overhaul of library discovery tools. Like the library Web site, the catalog still requires users to come to it. When they do, it still sets a high threshold for patience and the ability to interpret clues. Still, at the end of the day it rewards the NCSU student searching “Declaration of Independence” with the book, *American Scripture: Making the Declaration of Independence* instead of the recent Congressional resolution, *Recognizing the Mexican holiday of Cinco de Mayo*.

References

1. Christine L. Borgman, “Why Are Online Catalogs Still Hard to Use?” *Journal of the American Society for Information Science* 47, no. 7 (1996).
2. Karl V. Fast and D. Grant Campbell, “I Still Like Google: University Student Perceptions of Searching OPACs and the Web.” In *Proceedings of the 67th ASIS&T Annual Meeting* (Providence, R.I.: American Society for Information Science and Technology, 2004).
3. Ray R. Larson, “Between Scylla and Charybdis: Subject Searching in the Online Catalog,” *Advances in Librarianship* 15 (1991); Andrew Large and Jamshid Beheshti, “OPACs: A Research Review,” *Library & Information Science Research* 19, no. 2 (1997).
4. Nathalie Nadia Mitev, Gillian M. Venner, and Stephen Walker, *Designing an Online Public Access Catalogue: Okapi, a Catalogue on a Local Area Network* (London: British Library, 1985).

5. Borgman, "Why Are Online Catalogs Still Hard to Use?" 495.
6. R. Hafter, "The Performance of Card Catalogs: A Review of Research," *Library Research* 1, no. 3 (1979).
7. Gerard Salton, "The Use of Extended Boolean Logic in Information Retrieval," in *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data* (New York: ACM Pr., 1984), 277.
8. Ray R. Larson, "Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog," *Library Quarterly* 61, no. 2 (1991).
9. Ibid.
10. Charles R. Hildreth, *Online Public Access Catalogs: The User Interface* (Dublin, Ohio: OCLC, 1982).
11. Larson, "Classification Clustering."
12. Mitev, Venner, and Walker, *Designing an Online Public Access Catalogue*; Ray R. Larson et al., "Cheshire II: Designing a Next-Generation Online Catalog," *Journal of the American Society for Information Science* 47, no. 7 (1996).
13. Tamas E. Doszkocs, "CITE NLM: Natural-Language Searching in an Online Catalog," *Information Technology and Libraries* 2, no. 4 (1983).
14. Nicholas J. Belkin and W. Bruce Croft, "Retrieval Techniques," in *Annual Review of Information Science and Technology*, ed. Martha E. Williams (New York: Elsevier, 1987), 129.
15. Gary Marchionini, *Information Seeking in Electronic Environments* (New York: Cambridge Univ. Pr., 1995), 100–18.
16. Borgman, "Why Are Online Catalogs Still Hard to Use?" 494.
17. Lois Mai Chan, *Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources: Issues and Challenges* (Washington, D.C.: Library of Congress, 2001), www.loc.gov/catdir/bibcontrol/chan_paper.html (accessed July 10, 2006).
18. Lois Mai Chan, "Library of Congress Classification as an Online Retrieval Tool: Potentials and Limitations," *Information Technology and Libraries* 5, no. 3 (1986); Mary Micco and Rich Popp, "Improving Library Subject Access (ILSA): A Theory of Clustering Based in Classification," *Library Hi Tech* 12, no. 1 (1994).
19. Marcia J. Bates, "Subject Access in Online Catalogs: A Design Model," *Journal of the American Society for Information Science* 37, no. 6 (1986); Karen Coyle, "Catalogs, Card—and Other Anachronisms," *The Journal of Academic Librarianship* 31, no. 1 (2005); Larson, "Classification Clustering."
20. Karen Markey, "Thus Spake the OPAC User," *Information Technology and Libraries* 2, no. 4 (1983): 383.
21. Larson, "Classification Clustering."
22. Marcia J. Bates, *Library of Congress Bicentennial Conference on Bibliographic Control for the new Millennium, Task Force Recommendation 2.3 Research and Design Review: Improving User Access to Library Catalog and Portal Information*, Final Report, 2003; Charles R. Hildreth, *Intelligent Interfaces and Retrieval Methods for Subject Searching in Bibliographic Retrieval Systems* (Washington, D.C.: Library of Congress, 1989); Bates, "Subject Access in Online Catalogs"; Belkin and Croft, "Retrieval Techniques."
23. Bates, "Subject Access in Online Catalogs"; Bates, *Library of Congress Bicentennial Conference on Bibliographic Control for the new Millennium*; Eric Novotny, "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, no. 6 (2004).
24. Bates, "Subject Access in Online Catalogs," 367.
25. Larson, "Classification Clustering"; Buckland et al., "Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies," *D-Lib Magazine* 5, no. 1 (1999).
26. H. Greisdorf, "Relevance Thresholds: A Multi-Stage Predictive Model of how Users Evaluate Information," *Information Processing & Management* 39, no. 3 (2003): 403–23; Yunjie (Calvin) Xu and Zhiwei Chen, "Relevance Judgment: What do Information Users Consider beyond Topicality?" *Journal of the American Society for Information Science and Technology* 57, no. 7 (2006).
27. Joseph W. Janes, "Other People's Judgments: A Comparison of Users' and Others' Judgments of Document Relevance, Topicality, and Utility," *Journal of the American Society for Information Science* 45, no. 3 (1994).
28. Bernard J. Jansen and Udo Pooch, "A Review of Web Searching Studies and a Framework for Future Research," *Journal of the American Society for Information Science and Technology* 52, no. 3 (2001); Novotny, "I Don't Think, I Click."
29. Borgman, "Why Are Online Catalogs Still Hard to Use?"
30. Brian Nielsen and Betsy Baker, "Educating the Online Catalog User: A Model Evaluation Study," *Library Trends* 35, no. 4 (1987).
31. IFLA Cataloging Section, "FRBR Bibliography," www.ifla.org/VII/s13/wgfrbr/bibliography.htm (accessed May 1, 2006).
32. Lois Mai Chan et al., "A Faceted Approach to Subject Data in the Dublin Core Metadata Record," *Journal of Internet Cataloging* 4, no. 1/2 (2001).
33. Chan, *Exploiting LCSH, LCC, and DDC*.
34. Thomas Mann, "Is Precoordination Unnecessary in LCSH? Are Web Sites More Important to Catalog than Books?" *A Reference Librarian's Thoughts on the Future of Bibliographic Control* (Washington, D.C.: Library of Congress, 2001), www.loc.gov/catdir/bibcontrol/mann_paper.pdf (accessed July 10, 2006).
35. Karen Calhoun, "The Changing Nature of the Catalog and its Integration with Other Discovery Tools," prepared for the Library of Congress, 2006, 24. Unpublished, www.loc.gov/catdir/calhoun-report-final.pdf (accessed July 7, 2006).
36. Charles R. Hildreth, *Online Catalog Design Models: Are We Moving in the Right Direction?* (Washington, D.C.: Council on Library Resources, 1995).

Appendix A: NCSU Libraries Catalog Usability Test Tasks

Known-Item Questions

1. "Your history professor has requested you to start your research project by looking up background information in a book titled *Civilizations of the Ancient Near East*."
 - a. "Please find this title in the library catalog."
 - b. "Where would you go to find this book physically?"
2. "For your literature class, you need to read the book titled *Gulliver's Travels* written by Jonathan Swift. Find the call number for one copy of this book."
3. "You've been hearing a lot about the physicist Richard Feynman, and you'd like to find out whether the library has any of the books that he has written."
 - a. "What is the title of one of his books?"
 - b. "Is there a copy of this book you could check out from D. H. Hill Library?"
4. "You have the citation for a journal article about photosynthesis, light, and plant growth. You can read the actual citation for the journal article on this sheet of paper." Alley, H., M. Rieger, and J.M. Affolter. "Effects of Developmental Light Level on Photosynthesis and Biomass Production in *Echinacea Laevigata*, a Federally Listed Endangered Species." *Natural Areas Journal* 25.2 (2005): 117-22.
 - a. "Using the library catalog, can you determine if the library owns this journal?"
 - b. "Do library users have access to the volume that actually contains this article (either electronically or in print)?"

Topical Questions

5. "Please find the titles of two books that have been written about Bill Gates (not books written by Bill Gates)."
6. "Your cat is acting like he doesn't feel well, and you are worried about him. Please find two books that provide information specifically on cat health or caring for cats."
7. "You have family who are considering a solar house. Does the library have any materials about building passive solar homes?"
8. "Can you show me how would you find the most recently published book about nuclear energy policy in the United States?"
9. "Imagine you teach introductory Spanish and you want to broaden your students' horizons by exposing them to poetry in Spanish. Find at least one audio recording of a poet reading his or her work aloud in Spanish."
10. "You would like to browse the recent journal literature in the field of landscape architecture. Does the Design Library have any journals about landscape architecture?"

Bailey continued from 127

Copyright © 2006 by Charles W. Bailey Jr.
This work is licensed under the Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard St., 5th Floor, San Francisco, CA, 94105, USA.