

Endeca

Another approach toward creating a next-generation library interface involves making use of technology from Endeca Technologies. The Endeca technology isn't a fully developed interface designed for libraries, but a powerful and feature-rich search engine that can be implemented for any organization's information environment. Endeca has earned a strong reputation for "Guided Navigation," an implementation of the faceted navigation that has been a common theme among the next-generation library interfaces covered in this report.

About Endeca

Endeca was founded in 1999 and is based in Cambridge, Massachusetts. In May 2006, Endeca was awarded a patent (Number 7,035,864) covering its "hierarchical data-driven navigation system and method for information retrieval," or Guided Navigation.¹

Endeca's products have been widely implemented by a number of large companies including Wal-Mart, Home Depot, Barnes & Noble, Costco, Bank of America, Hyatt, and Boeing to help power their Web sites.

Endeca provides state-of-the-art search and navigation technology. The Endeca search engine includes a wide variety of sophisticated features for determining matches to a query and for calculating relevancy ranking, determining appropriate facets, and giving spelling suggestions. With these capabilities, the Endeca technologies are well suited to libraries seeking to develop next-generation interfaces.

The Library Corporation formed a nonexclusive agreement with Endeca to distribute its Guided Navigation product beginning in June 2004. Libraries can also purchase the product directly from Endeca.

Creating an Endeca-Based Catalog

North Carolina State University Libraries was the first out of the chute to launch a library catalog based on the Information Access Platform from Endeca. NCSU crafted an interface to operate with its SirsiDynix Unicorn implementation, replacing its previous catalog based on the SirsiDynix Web2 product.

The Endeca product operates primarily behind the scenes, providing a search engine with a number of powerful features. As such, it gives the library a great deal of flexibility in how it implements a new interface based on this technology. This approach requires the library to take more responsibility for making decisions regarding search engine configuration options, relevancy-ranking algorithms, and interface design.

At the heart of the Endeca Information Access Platform lies the MDEX Search Engine, which performs most of the work during the search process. The initial task involves determining what items match the query. The search engine can be configured to process results according to a variety of options, which can be set differently for each index or dimension. The determination of results candidates from a query can be performed according to several different techniques, ranging from "match-all," where all the terms in the query must appear, to "matchany," where even a single query term triggers a match. The search engine also performs relevancy ranking based on the options and weightings. The Endeca search engine includes the ability to automatically detect and correct spelling errors in query terms, offering "did you mean" suggestions. These are but a few of the many features and options in the Endeca software.

Libraries interested in creating a catalog based on the Endeca search platform should be prepared to devote

considerable staff resources to the endeavor. Although NCSU was able to deploy its new catalog in only about six months, the project was an intensive effort involving many staff members. Libraries acquiring the technology through the Library Corporation can benefit from that company's services for implementation and integration.

The implementation of an Endeca-based catalog involves operating a search interface that is largely decoupled from the library's existing ILS. Records from the ILS will need to be exported and loaded into the Endeca search engine through a set of tools called the Data Foundry. This process builds the indexes according to the library's specifications, creating the various dimensions that will be available as facet categories.

An Endeca catalog is not completely disconnected from the ILS, however. Several aspects of the catalog rely on functionality provided by the existing online catalog or dynamically obtained in other ways from the ILS. The Endeca catalog might, for example, rely on the older catalog for the real-time status and availability of items and for patron service requests such as renewals, recalls, and holds.

The Endeca-based catalog as implemented by NCSU should be taken as only an example of how the technology can be implemented. Each of the implementations of library catalogs based on the Endeca platform looks and functions quite differently from the others.

A look at the results set from the NCSU catalog gives an overview of the capabilities of the Endeca software (see figure 3). Once a user enters a query, the initial list of results appears in a wide column. The results list appears in relevancy order by default, with the option to change to title or date sort through a drop-down selection tool.

A narrow column on the left side of the page displays groups of facets represented in the search. The groups

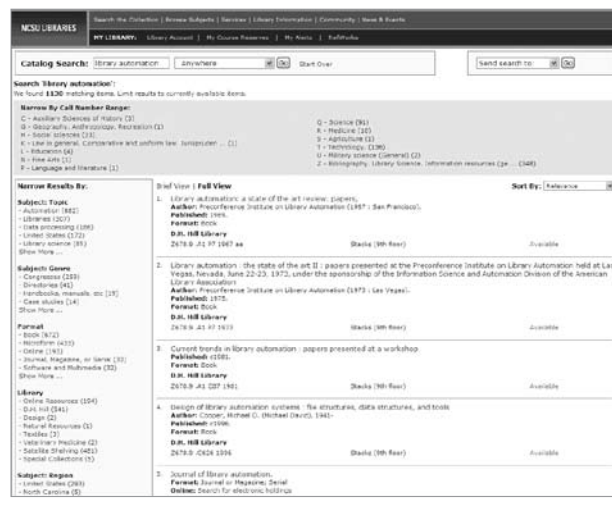


Figure 3: Initial results display from NCSU's Endeca-based catalog.

available include topical subjects, genres, format, library location, and geographic region. Following the standard conventions, under each category label appear the facets represented in the search results, with the number of items that will be returned enclosed in parentheses.

Above, spanning both columns, appears another set of links that allow the user to limit the results according to the Library of Congress call number ranges. This feature allows the user to home in on results based on the general organization of the library's collection and leverages the work done by catalogers. In addition to narrowing search results by LC classification, NCSU has also implemented the ability to browse through the catalog starting with these classification codes and drilling down through each sublevel. This feature emulates the way that users might explore the shelves of the physical library, browsing through the titles to see what items of interest they come across. The ability to browse and limit by LC classification is made possible by NCSU's decision to use these data from the MARC records as one of the dimensions configured in the Endeca indexing scheme (see figure 4).

NCSU also uses availability as one of the facet dimensions, making it easy to limit searches to only those items currently available.

The results listings display a brief view of each item, with some bibliographic information and basic holdings information, including library, call number, location, and availability. Clicking on an item launches a full-display window, tapping into the original ILS catalog, that includes more detailed information about the item as well as up-to-date holdings and status displays (see figure 5).

NCSU's new interface based on faceted navigation and relevancy ranking offers important improvements.

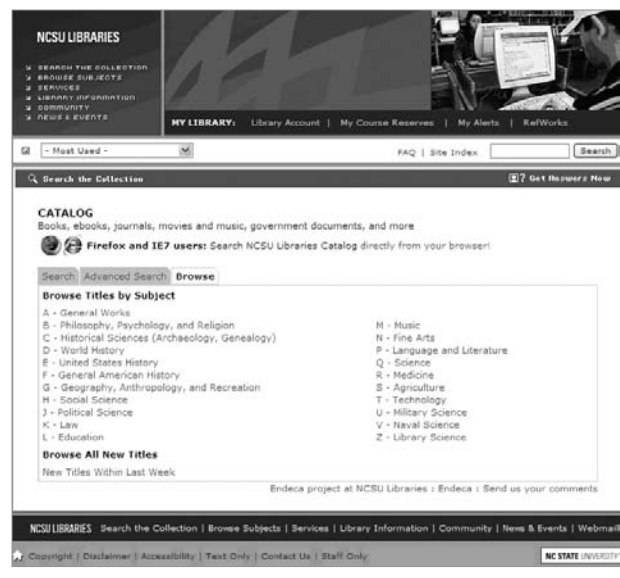


Figure 4: The NCSU catalog browse feature.



Figure 5:
Detailed item view from NCSU catalog.

In comparison to other next-generation library interfaces, however, it does not address some of the broader themes.

The new NCSU Endecca catalog does not aim to expand the overall scope of search. It provides access to the same body of material as the one it replaced. It does not bring the content of additional collections into its indexes or integrate with a federated-search environment, as we have seen with other next-generation library interfaces.

Another feature conspicuously missing involves enhancing the displays with cover art images or other extended content. The text-only displays of the NCSU catalog appear plain compared to those that blend in more graphic content.

The more recent implementation of Endecca at the Phoenix Public Library, while based on the same technology, takes a much more visual approach. Book jacket images and more color on the page result in a much flashier interface. This version, not surprisingly for a public library, does not offer the facets related to LC classification codes. While the facet groupings on the left side of the page are organized and labeled much like those for NCSU, they are missing the number if items associated with each of the facets (see figure 6).

McMaster University in Canada also recently implemented a new library catalog based on Endecca. The McMaster version looks significantly different from either the NCSU catalog or the one for Phoenix Public Library. Some of the characteristics of the McMaster catalog that vary from the other implementations include:

- **Collapsed facet groupings.** In this version, only the facet category headings display by default. Clicking on the category label expands the section, revealing the facets with item counts. An Expand All link opens all the categories at once. This approach allows the user to easily see all the options without having to scroll down on the page, but at the cost of an additional click to get to the actual facets.
- **Browse titles by subject.** Users can browse the entire collection by subject category. A portion of the search page lists all the subject categories along with the number of items classified in that area.

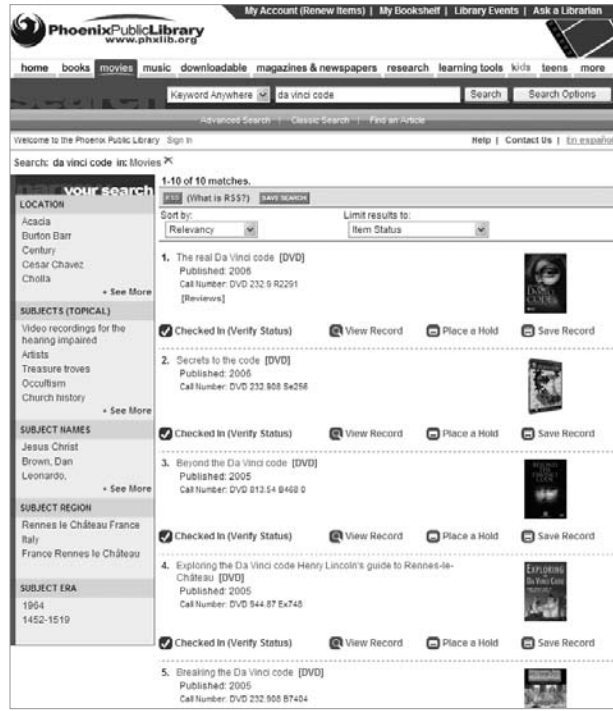


Figure 6:
Initial display of the Endecca-based catalog of the Phoenix Public Library.

- **Visual appeal.** McMaster's catalog includes cover art images.
- **Detailed view.** Instead of launching a new page when the user clicks on an item, it opens a cell that displays the details for the item. The details box closes automatically when the user clicks another item on the page or the close box.
- **Availability.** For physical items in the library's collection, a Check Availability link launches a new browser window that links to the item in McMaster's Horizon catalog.

A look at the three implementations of library catalogs using the Endecca technology reveals that while the software provides a common set of search features, focusing on faceted browsing with relevancy ranking, the interface options are extremely flexible.

North Carolina State University Libraries
www.lib.ncsu.edu/catalog

Phoenix Public Library
www.phoenixpubliclibrary.org

McMaster University Library
http://libcat.mcmaster.ca/index.jsp

Note

1. Adam J. Ferrari, David Gourley, Keith Johnson, Frederick C. Knabe, Daniel Tunkelang, and John S. Walter, "Hierarchical Data-Driven Navigation System and Method for Information Retrieval," U.S. Patent 7,035,864, filed May 18, 2000, and issued Apr. 25, 2006, available online at <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=7,035,864.PN.&OS=PN/7,035,864&RS=PN/7,035,864> (accessed June 1, 2007).