

Introduction

This issue of *Library Technology Reports* focuses on “next-generation library catalogs.” In this current phase of library automation, all eyes are focused on developing and deploying Web-based interfaces better suited to meet the expectations of the current generation of Web-savvy users. Over the course of the last year, a number of libraries have made bold moves to introduce new catalogs cast in a mold apart from their previous offerings. Library automation vendors have launched development efforts to create new catalogs and interfaces more in tune with today’s expectations.

Interest in a new generation of library catalogs springs from widespread dissatisfaction with many of the online catalogs provided as integrated modules of library automation systems. Although these catalogs have been designed to provide a rich set of functionality related to the library’s collections and services, they have not necessarily followed the broader conventions established in the context of the Web.

Terminology

While we can observe a lot of activity toward the development of new library interfaces, there is no single definition of what constitutes a next-generation library catalog. It’s not just that they were recently developed. Some of the products that we consider within this genre, such as AquaBrowser, have been around for a few years now. In broad terms, what constitutes a next-generation library catalog is its ability to transcend some aspect of the traditional mold of library catalogs.

The term *next-generation catalogs* presents problems from the outset. One might think of the term *next-generation* as describing something new that might be

developed in the future. Libraries seek next-generation catalogs here and now; the need isn’t future or abstract. Libraries do not necessarily have to wait. Most of the products on which this report centers, while in the early stage of their product life, are available now. So while it might be more accurate to use the term *current-generation catalogs*, that could lead to confusion with the online catalogs widely deployed today that were developed under a previous set of expectations. In this report, we will refer to these earlier products as *legacy catalogs*.

The term *catalog* also fails to completely capture the current vision of the tool for finding things in a library. It doesn’t do justice to the new, expanded vision of the library’s search environment. *Catalog* denotes a listing or directory of items in a collection. Part of the vision of the next generation of library interfaces involves searching within the items in the library’s collection.

The term *catalog* can be especially confusing in the context of a college or university Web site. Many academic institutions place a prominent link on their sites for the “catalog,” meaning a directory of the courses offered. Once the user selects the library, however, the catalog takes on the meaning of the search utility for finding books and other library materials. While we accept the term *catalog* in this report, we seek to expand the concept to something broader than it has meant in the past. The new generation of library catalogs differs in many important ways from its previous conceptions.

While the label *next-generation catalog* may not exactly capture this new generation of library interfaces, we will use it in this report, with all the caveats implied here. In this issue of *LTR*, we will explore some of the concepts, technologies, and user expectations that underlie this next generation of catalogs and will review some of the current and emerging products in this arena.

Problems with the Status Quo

In order to capture—or recapture—the attention of our users, libraries find themselves in catch-up mode. The Web permeates our culture. Almost all library users come with existing expectations set by their experience of the Web. To be taken seriously by users, the catalogs and other interfaces offered by libraries need to operate with the same levels of style and sophistication as other popular Web destinations.

Some of the shortcomings of legacy catalogs might include that they

- have complex search interfaces that might not be sufficiently intuitive
- are not consistent with well-established user interface conventions
- are unable to rank results according to relevancy or interest
- are too limited in scope
- are tied to print materials and are less able to address electronic content
- are unable to deliver online content to the user
- lack social network features to engage library users

While the legacy catalogs have their shortcomings in light of the heightened expectations set by the current state of the Web, they also have their strengths. When dealing with the library's physical collection, they offer a great deal of sophisticated functionality.

For example, the legacy catalogs carry a great deal of solid functionality related to patron services. Most allow a library user to sign in to an individual account for personalized services. A typical sign-in feature allows the user to view the items currently charged, see when they are due, perform renewals, pay fines, and initiate various types of requests.

Many aspects of the legacy catalogs must be incorporated into the next generation of library interfaces. We see many examples where new products rely on features provided by an existing library catalog, especially in areas involving personalized patron services and the display of the current status of library materials.

The Scope of the Catalog

One of the key problems with the traditional library catalog is its limitation in scope. A catalog that addresses only a subset of the library's collection falls short of the ideal. Asking library users to use one interface to find books and another to perform article-level searching adds a level of difficulty.

The shift in library collections toward increased proportions of electronic content has been underway for

a number of years and has reached the point in many libraries where expenditure for electronic content outpaces that for print. Libraries subscribe to many different products that provide access to article-level content. In the conventional library environment, no single product serves as a comprehensive gateway to all the components of the library's collection. One interface, the legacy catalog, specializes in print; libraries usually offer separate access methods for electronic content.

A typical library will develop lists of all the article-level databases to which it subscribes and will organize them by discipline or content focus. Some libraries have developed automatic finding aids that guide users to the best resource based on selections made in a scripted interface. In most cases, users will need to search multiple article databases to complete their research.

A lot of variables arise when we think about the scope of the library catalog, in terms of both what is common today and a more ideal approach. Is it realistic for all the content that the library provides to be in the catalog? Most library catalogs today provide access to some, but not all, components of the library's collections. Categories of content in a legacy catalog include

- **Books.** Separate records exist for each edition. Individual copies are represented in item-level records. Print and electronic book titles are included.
- **Multimedia materials.** These include CDs and DVDs.
- **Newspapers, magazines, and professional and scholarly journals.** These materials are represented only at the title level. The traditional library catalog does not describe individual articles.
- **Other categories, depending on the specialization of the library.** These might include government documents, musical scores, collection-level records for manuscript collections, and the like.

Legacy library catalogs rely solely on databases of MARC records. The universe of what can be searched is limited by the information that can be placed within this record structure. In most cases, MARC records provide metadata describing an item, not the full text of the item itself. Many complex standards underlie what we loosely call MARC, including AACR2 (Anglo-American Cataloging Rules, version 2), ISBD (International Standard Bibliographic Description), and MARC21 (Machine Readable Cataloging). While all these standards contribute to a very rich approach for describing bibliographic materials, the reliance on MARC alone can be limiting. A more expansive vision of library catalogs includes content and collections not well suited to MARC and might also embrace other metadata formats.

Over the last two decades, libraries have gradually shifted the focus of their collections toward electronic content. Yet finding information within these electronic

resources has not been integrated into the library catalog. Libraries subscribe to electronic indexing and abstracting products that provide their own interfaces for searching the contents of journals at the article level. Initially, these A&I databases based their search on metadata describing each article, such as titles, authors, and a few subject keywords or possibly an abstract. Increasingly, A&I interfaces now search the full text of articles in addition to the structured metadata fields.

While library collections have shifted toward increased proportions of electronic content, the traditional approach to library catalogs did not focus on providing access to this material. The scope of the legacy catalog usually does not include

- article-level searching
- online display of article content
- search and display of content from local digital library collections: photographs, manuscripts, local newspapers, genealogical materials, and the like
- contents of an institutional repository

Libraries increasingly subscribe to serials in their full-text electronic form, sometimes in addition to, but often instead of, the print version. Offering this content in electronic form vastly increases the opportunities for access and discovery. Once content is in electronic form, interfaces can be devised that search all the words, not just the metadata. As a library's e-book collections continue to grow, this full-text searching applies to books as well as articles in e-journals.

Now that content increasingly comes in digital forms, libraries have the opportunity to provide to their users extraordinarily powerful tools for discovery and access. Libraries subscribe to a myriad of products, each focusing on a particular content niche. By virtue of the electronic-content products to which a library subscribes, users have potential access to vast amounts of information available for online viewing.

Full-text electronic content is generally beyond the scope of the legacy catalog. Most catalogs provide a record for each journal or periodical. They lack the ability to allow users to search for articles within e-journals. While it is helpful to know to which e-journals the library subscribes, it would be far better to include the ability to search for the articles themselves. This limitation of scope underlies a frequent point of confusion for library users. They often don't understand that they must use some other interface to search for article-level content.

One of the characteristics of the vision for the next generation of library catalogs involves an expanded scope. In an ideal library interface, print and electronic content would stand on equal footing. Content to which

the library subscribes residing on servers external to the library would be just as easily accessed as the materials held physically in the library.

In this broader view, the catalog provides access to content delivered on all types of media. In addition to text-based materials, the next generation of library would provide tools for the discovery and display of other types of content, such as images, audio, video, and animations.

Federated Search

One of the major characteristics of the next generation of library catalogs involves a broader, more comprehensive approach to the body of material addressed in a search. In order to gain a better understanding of what is involved, it will help to consider one of the models popular today, that of using a federated search.

Apart from the online catalog, many libraries offer a separate search environment to address their ever-expanding collections of electronic content. As the collection of electronic resources grows ever larger, library users often have a difficult time knowing what products might contain information on their research topic. Searching many different information resources separately can be a time-consuming and tedious process. Federated-search products attempt to simplify the research process by offering an interface that allows a researcher to work simultaneously with multiple information resources.

A federated-search interface essentially serves as an intermediary between the user and a selected group of information resources. It relies on computer-to-computer conversations to automate the process of searching multiple resources.

A federated-search interface invites the researcher to enter a search and then casts that search to a selected group of resources, or targets. The federated-search utility will then receive the results from each of the targets and present them to the user, using its own presentation format. Depending on the capabilities and configuration of the federated-search product, the results sets from each of the targets will be further processed. Some of the processing might include cosmetic changes, such as presenting in a standard format and structure; it could also include de-duplication, where identical entries received from multiple targets are consolidated. One option with federated search involves combining the results from each of the targets into a unified set and ordering those results. The unified results might be ordered alphabetically, chronologically, or by relevancy.

The distributed query model of federated search has limitations in the speed of performance. The federated-search interface must establish a session with each of the targets, initiate the query, and retrieve the results. The

speed of the slowest responding target limits the speed of the overall process. Time-outs can be established to continue with the global process if one of the targets does not respond. Keeping the number of targets to a minimum and limiting the number of items requested from each target can help boost perceived performance. Many federated-search utilities request a small initial set of records from each target and retrieve additional results as the user delves deeper. Given constraints on bandwidth and the need to begin displaying some results to users fairly quickly, a federated-search interface will begin its initial display with a very limited number of results from each of the targets.

Ideally, a federated-search interface would transfer the complete results from each of the targets before it began processing the comprehensive results set. Only with the complete results in hand does it become possible to perform a credible sorting or ranking. Given the large number of items likely to match a given query, this approach would take far longer than an acceptable response time.

Ordering the results of a federated search by relevancy is a popular choice, given the dominance of this approach in the major Internet search engines. Federated-search engines, however, tend to be ill equipped to perform sophisticated relevancy ranking.

The distributed query approach faces difficulties presenting results according to relevancy ranking. The key problem relates to the shallow initial results sets from each of the targets, which we noted above as a strategy to boost perceived performance. While a typical query might return thousands or tens of thousands of items from each of the targets, the number of initial results may be as small as 30. If, for example, the query addresses 12 targets, and each returns 30 results, the initial results set will be based on 360 items. In many cases, the federated search does not have control over the order in which the results are returned by the targets. One cannot assume, for example, that an initial set of 30 results represents the most relevant or the most recent items that match the query.

While federated search is a pragmatic solution, it has some limitations. It sits apart from the library's catalog, so users must search separately for article-level content. It also makes a number of compromises in the way that it provides search services in order to gain acceptable performance. The ideal next-generation library catalog would subsume some of the ground currently held by federated search, taking on the responsibility for including article-level searching of the appropriate resources. The initial phase of this expanded scope might involve a tighter integration between the library catalog and a federated-search product. A later phase might include finding ways to include article-level information within the catalog's own indexes.

Delivering Content to the User

In addition to providing tools to help users find information on a topic, it is also important to provide the means for using that information. Providing a way to view the material online is ideal. This ideal is possible only when the content is available electronically, either as a freely available resource, as part of the subscriptions purchased by the library, or through pay-per-view commercial products.

Link resolvers based on the OpenURL specification serve as one of the key tools for delivering electronic content to library users. A link resolver, powered by a knowledge base of the electronic content to which a library subscribes, can present to the user a menu of options on how to view that content. If a record for an article turns up in a results set, the results set can present a button that invokes the link resolver, which can dynamically build a link to the full text of that article if it is available within the library's profile of subscriptions.

Many libraries choose to integrate a link resolver into their federated-search environment. The combination of a federated search to assist users in finding articles and a link resolver to help locate articles within the library's subscribed content results in a powerful environment for providing access to electronic content. Unfortunately, this environment can also be fairly complex and may not be entirely intuitive to library users.

For items available only in print form, the user would have to obtain the physical item. For books, key information might include whether it is owned by the user's primary library and whether it is currently available for checkout. If it is currently charged out by another user, delivery tools would include the ability to place a request to have the book next. If the book is not owned by the library, service options might include an interlibrary loan request or a link to an online bookstore where the book can be bought.

In the expansive view of the next-generation library catalog, discovering items of interest forms only the first half of the process. The second half involves putting the actual content in front of the user through the online viewing of electronic content or services related to providing physical materials to the user.

Integrated or Dis-integrated?

One of the key characteristics of library automation over the last few decades involves the concept of the integrated library system. This model of automation involves a modular system that provides a separate subsystem for each major category of a library's operation, supported by a common database infrastructure. The classic model of the ILS includes a cataloging module based on a central

bibliographic database that represents the library's collection and the tools for creating and maintaining that database. A serials control module deals with the specialized needs of periodicals and serials. Acquisitions deals with the business processes involving procuring materials for the library, payment transactions, budget control, and other related features. The circulation module provides inventory control, focusing on a broad set of features that automate the operations of the library's circulation desk. Through the circulation module, library staff can check out materials to library users, discharge returned materials, perform recalls and holds, charge fines for late materials, and the like. The online catalog module provides a direct interface to library users for searching the library's collections. Through the integrated online catalog module, library users can search the bibliographic database and perform a variety of tasks related to their use of the library, including placing requests for materials, viewing the items currently checked out, renewing materials, paying fines online, and the like.

Though the concept of the ILS involves discrete modules, in practical terms libraries acquire them as a suite of components from the same company. When it comes to the basic modules of the ILS—cataloging, circulation, serials, acquisitions, and the online catalog—libraries have not generally had the option of following a “best-of-breed” approach, choosing modules individually from different vendors. These core modules were not designed to operate independently of each other and have deep dependencies on the underlying database and other infrastructure components.

In recent years, this integrated approach has not held as fast. While the core modules continue to be intractably dependent on same-vendor suites, a new set of modules, mostly related to the management of electronic content, has emerged with a more flexible approach to integration. Utilities related to federated search, OpenURL linking, and electronic-resource management have been delivered to operate independently of the core ILS modules. The purveyors of these utilities intend for them to be used by any library, regardless of the ILS in place.

The ability of these utilities to operate independently of the ILS arises primarily from the competitive business environment. As libraries seek automation tools to address their increasing reliance on electronic resources, companies see an opportunity to extend their reach into their competitor's libraries.

While the core ILS modules remain tied to a single vendor, it is now theoretically possible to surround that ILS with components from other companies. It is not at all unusual for a library using the ILS from one vendor to implement federated search, linking, and electronic-resource management from other vendors.

The emergence of tools to manage electronic resources outside the core ILS development path has paved the way

for a dis-integrated approach to library automation as the library moves beyond traditional print materials.

Libraries pay a steep price for this dis-integration in terms of the staff resources involved in implementing multiple applications. Given that all these components must ultimately work together, the library takes on the role of integrator. As a library implements these additional modules, it must address a number of technical and procedural tasks. When a library implements an ILS, though the installation process can be quite complex, it comes as a well-integrated set of modules that do not need to be configured to work with one another. This isn't quite the case once the library begins implementing additional components, such as federated search, link resolvers, and electronic-resource management systems. These utilities tend to be designed to operate much more independently than the core modules, and the library will need to devote some attention to integrating the data, interface, and infrastructure of these components into its overall automation environment.

Until recently, the online catalog has been firmly within the realm of the core modules not eligible for dis-integration. The online catalog has deep dependencies on both the bibliographic database and the circulation subsystem. An integrated online catalog operates with complete synchronization with the other modules of the system. The search environment of the online catalog ties directly to the bibliographic database of the ILS. It immediately reflects new and changed items in the bibliographic database and dynamically shows circulation status and availability. Expected functionality of an online catalog includes the location, call number, and circulation status of each item it displays.

An integrated online catalog also leverages the database of library users maintained for the circulation system. At the circulation desk, the patron database connects individual users to the items charged to them. In the online catalog, the patron database provides the basis for a wider set of features. A typical online catalog includes capabilities for users to sign in to their own account, set preferences, view charged items, and perform other services. These features tie in through the ILS patron database.

So given the close relationship between the online catalog and the rest of the ILS, why would a library be interested in a less integrated model? The reason has to do with the limitations of online catalog modules delivered by ILS vendors. A library may have an interest in providing access to larger bodies of content, beyond what is managed within the ILS, and in gaining access to more advanced search technologies. As an integrated module, the library catalog lives at the mercy of the search technologies and other functions provided within the ILS. Just as important, the online catalog is tied to the development cycle of the ILS vendor. If the ILS vendor fails to

develop it aggressively and provide imaginative features, the library can fall behind in the eyes of its users.

Given that the ILS is designed to treat the online catalog as a deeply integrated core module, disentangling this module to function independently involves considerable work. The interactions between the online catalog and the rest of the ILS are deeply woven. In its search processes, the catalog interacts with the bibliographic database and its indexes. It displays information from the circulation module to show current location and availability status, and it interacts with the patron database for user sign-in and personalization features. No standard set of protocols exists to connect the online catalog with the other aspects of the ILS. The process of replacing the online catalog supplied with the ILS with another product involves finding ways to replicate both the search and the service components.

Given the advancements in search engine technology, it's possible to gain significant improvements over the often-rudimentary search capabilities provided within the core ILS.

One approach to replicating the search component of the online catalog through a separate product involves exporting the entire bibliographic database of the ILS into a new search engine. The library might, for example, run a utility that exports its entire database of MARC records and reformats them according to the requirements of a third-party search environment. Those records would then be loaded into the search engine, indexed, and made available through a new interface. Given the constant updates to the library's bibliographic database, it's necessary to keep an external search engine synchronized through frequent incremental updates and re-indexing.

The transfer of data out of the ILS into an external search platform can be a complex undertaking. The process might involve dealing with multiple databases within the ILS. In addition to the basic MARC record, it is necessary to layer in holdings and item-level information, which may be stored in the ILS in a proprietary format.

This process of exporting, reformatting, loading, and indexing presents a great deal of overhead in order to offer a new search environment. It's a process that can be largely automated and that can execute very quickly on the powerful computing platforms available today.

The task of providing accurate information regarding the status and availability of each of the items presents an even greater challenge. If this information is transferred in bulk between the ILS and the external search engine, the transfer must be performed very frequently in order for the system to be able to display the circulation status of items accurately.

Another approach to presenting status information to users in an external search environment involves linking back to the original online catalog displays for selected parts of the search and display process. This hybrid

approach makes it possible to take advantage of the more sophisticated features of an external search engine while maintaining the advantage of an integrated online catalog for status and service features. When we look at specific products later in this report, we will see a variety of approaches in how products tie back into the ILS to obtain current status information. Some, like WorldCat Local, use a behind-the-scenes protocol to interrogate the ILS and then present the status information within its own interface. Some, like the Endeca implementation at NCSU, rely on the new search environment to display lists of results, but when the user selects an item to view in more detail, it hands off that display to the local ILS, replete with its detailed status information and user service features.

The Next Generation: Putting All the Pieces Together

So, given the inadequacies of the legacy catalogs, what is the vision for the next-generation library catalog? There isn't one single answer. We will see a number of approaches, each attacking the problem somewhat differently. Some common threads include an expanded scope of search, more modern interface techniques, and search engines better at ranking results.

The status quo library interface takes a piecemeal approach. As we consider the next generation of library interfaces, we expect to gather a broader set of information, resources, and services into a single interface that is more comprehensive in scope and more modern in presentation.

One of the key goals for the next-generation library interface involves a single point of entry to all of the library's information. In an ideal world, the content of all the library's collections would be available through a single search interface. Wouldn't it be great if the library could offer a single search box that included all the traditional ILS content and the full text of all the electronic resources to which the library subscribes? An important part of the vision of the next generation of library catalogs involves exploring ways to expand the search universe. Instead of a catalog dealing mostly with print and a federated-search utility geared to work mostly with electronic content, a more expansive view involves a consolidated search environment that combines the features of both. This environment would combine the detailed features of the legacy catalog when it comes to finding library materials, including status display and request features, with discovery and online delivery of electronic content.

Today, a very large proportion of the content of periodicals, journals, and newspapers is available in electronic format. While problems remain with integrating all the different products in which this content resides, library

users currently reap great benefits from this full-text electronic content.

The availability of large amounts of the electronic full text of books seems to be on the horizon. A number of mass digitization projects are underway that promise the availability of millions of book titles in the next decade. The Google Library Print project, the Open Content Alliance, and the efforts of many individual libraries may soon bring to books the same benefits we currently see for serial publications.

Given the digital publication process, current books pose smaller technical problems. The primary obstacles to taking advantage of the full text of book content lie in the realm of copyright and contracts regarding intellectual property.

While the current tools focus on providing access to the full text of journal articles, the next phase in the progression of library interfaces may include providing online access to the full text of books.

A State-of-the-Art Web Interface

It's important that library interfaces compare well with other Web destinations in appearance and in navigation. Library users are increasingly well experienced with the Web and have become accustomed to the user interface conventions followed on other Web sites. Given the broader experience of the typical user, expectations are set high. When users interact with intuitive interfaces and visually appealing sites elsewhere on the Web, libraries feel challenged to offer interfaces that work just as well and look just as good.

Enriched Content

Legacy catalogs tend to offer text-only displays, drawing only on the MARC record. A next-generation catalog might bring in content from different sources to strengthen the visual appeal and increase the amount of information presented to the user.

Some categories of the content that can be blended with the basic bibliographic information from the ILS include

- **Cover art images, such as book jackets, movie cases, or any other visual representation that invokes the work.** Even a thumbnail-size image can spruce up the visual appeal of the record and convey information regarding the item. Especially for visually oriented users, a graphic can grab attention and draw in the user to read the text.
- **Tables of contents.** Especially for book-length works, a table of contents provides additional information that might not otherwise be captured in the bibliographic record. Individual chapter titles provide a higher level of detail.
- **Summaries.** Narrative summaries also provide additional information that might not be represented in

a basic bibliographic record. These summaries come from a variety of sources, ranging from promotional publisher blurbs to more objective abstracts.

- **Reviews.** Reviews assist the user in evaluating an item, especially when provided by authoritative experts in the field.

Record enrichment isn't that new a concept. Library catalogs have been gaining this capability gradually over the last few years. An increasingly large percentage of books and other library materials will have enriched content available.

Syndetic Solutions, now a subsidiary of Bowker, stands as the dominant provider of content to libraries to enrich the display of items in their catalogs. The company offers subscriptions that provide access to various levels of enrichment content. Syndetic Solutions has few competitors offering a prepackaged subscription service of enriched content for library catalogs. Therefore, the pricing can seem fairly aggressive to libraries with a modest budget.

Syndetic Solutions
www.syndetics.com

Much of the same types of enrichment can be obtained from Amazon.com. Part of Amazon's business model involves allowing others to make use of the e-commerce infrastructure that it has developed in creating their own online storefronts. Amazon benefits from increased exposure and commissions it receives on sales made by its partners. The Amazon Web Service, or AWS application programming interface, provides a mechanism that anyone can use to draw on the content and technologies within Amazon's environment. Of particular interest to libraries is the ability to display book images, reviews, summaries, and other content in their catalogs. While Amazon does not impose fees for the use of this content, it does require that any organization that makes use of any part of AWS provide a link on its site back to Amazon's.

AWS represents a more do-it-yourself approach to content enrichment that might be more technically challenging, but also more affordable. The requirement to provide a link to Amazon's Web site might be an obstacle for many libraries.

The standard approach to record enrichment involves layering the additional content only as a display feature. As the interface begins to display the record, it transmits the ISBN or other unique identifier to the enrichment supplier. If it finds a match, the service provides the images or other applicable data, which the local interface then presents to the user. Following this model, the enrichment takes place only upon display and does not affect

the search process. Words in the table of contents, for example, would not be indexed in the local search engine to increase the findability of the item.

An obvious extension to enrichment would involve retrieving the enriched content in advance as part of the indexing process for the local search engine. This approach adds more power to the overall interface, but requires a different workflow in the way that enrichment is added to the library's online catalog. There may also be contract and licensing implications for preloading enrichment data rather than using the on-the-fly presentation model.

Features Expected in the Next Generation

The vision of the next-generation library catalog expands far beyond the library catalogs of the past. The feature set includes much of what has evolved in earlier library OPACs, but blends characteristics found in many commercial Web destinations and social networking sites.

Faceted Navigation

A user-interface technique that has proven itself to be extraordinarily useful in the search process involves the use of facets that can be selected to narrow the results. Facets appear as links corresponding to words or phrases found within the results. A prevailing convention involves showing, usually in parentheses, the number of items that will be retrieved by selecting each facet. Most interfaces present the results-so-far in the middle of the display with the facets on either the right or left side. Depending on the complexity of the information being searched, facets may be grouped into categories.

Faceted navigation embodies a drill-down approach to searching an information resource. In this mode, a researcher begins with a general concept and incrementally homes in on a narrow results set by navigating through the specific terms revealed through facets. In a well-designed interface with faceted navigation, each time the user selects a term presented as a facet, a list of the items returned displays, along with new facets to further narrow the results. The process of incremental narrowing continues until the researcher is satisfied with the results and has reduced the items returned to a manageable quantity.

The process of faceted navigation allows the user to interact with an information resource by discovering the information held within rather than having to guess in advance. This approach stands in fundamental opposition to the traditional approach, in which the interface provides an advanced search page where the user can construct a complex set of qualifications at the beginning of the search process. The advanced search process gets users to express the exact concepts that match their

research when they initially approach an information resource. These concepts may or may not actually match the content of the resources. In many cases, a researcher will employ a trial-and-error approach by changing the selections in an advanced search page until the desired results are achieved.

Faceted navigation has been employed as a standard technique on popular Web sites for many years and has grown to be a well-accepted approach. For experienced Web users, faceted navigation isn't something that needs to be explained.

One of the challenges in constructing an interface with faceted navigation involves the mechanism used for presenting the terms that represent the facets. In many cases, the information resource already includes metadata that can be used to calculate facets.

The MARC record of a library's bibliographic database provides a wealth of fodder for creating facets. One can easily extract facets according to a number of different categories. Personal names, geographic areas, genres, topical subjects, date ranges, media types, and language are examples of the categories of facets that can easily be derived from MARC records.

Outside the realm of library bibliographic records, the raw data available for creating facets become much sparser. For information resources based on unstructured data, constructing facets can be much more difficult. In the absence of predetermined metadata, facets can be created through an analysis of the concepts reflected in the results sets.

The Vivisimo Clustering Engine offers an example of providing an interface much like faceted navigation without the need to preclassify the data. The Vivisimo technology analyzes the results of a search, groups items conceptually, and assigns each group a label. This label operates much like a facet, allowing the user to narrow the results set. Clustering technology works in the absence of classification terms, thesauri, or tags.

Vivisimo Clustering Engine

<http://vivisimo.com/html/clustering-about>

We noted that MARC records provide a wealth of data to populate a faceted interface. One of the problems that libraries have encountered in creating faceted interfaces based on their MARC records involves the length and complexity of facets derived from Library of Congress Subject Headings. While LCSH works well as a comprehensive tool for organizing library materials, it requires an expert to apply, and it generates long and complex headings that do not lend themselves to use in faceted interfaces.

FAST, or the Faceted Application of Subject Terminology, is an alternative approach for implementing sub-

ject headings developed by the OCLC Office of Research. This approach applies a much simpler set of rules in the assignment of headings to records and results in much shorter headings that work better with an interface based on faceted navigation.¹

Keyword Searching

Searching by keywords stands as a ubiquitous technique for finding information on the Web. Google and the other search engines have acclimated users to beginning their search process by entering a few words into a simple search box.

Legacy catalogs offer a variety of search options that can seem very complex compared to what users are accustomed to seeing elsewhere on the Web. Yet the advanced search options of legacy catalogs can be of great importance for some researchers. The single keyword query box may not be enough for most libraries. A more complete environment might include a link to an advanced search page that offers the user more precise search options.

Relevancy

Another key expectation well established by other Web interfaces involves the way that results return from a search. Almost all Web search engines return results according to some kind of relevancy ranking. The most important or interesting items appear first, followed by those of diminishing relevancy.

When relevancy ranking works well, it appears almost magical to the searcher. Type in a few words, thousands of items qualify as results, yet the best one appears at the top of the list.

Implementing a system that performs good relevancy ranking, however, can be incredibly difficult. Especially in response to a broad query, the number of potential results can be very large. Determining the ones most relevant poses a difficult technical problem. On one level, the search engine must determine the potential pool of result candidates. These are the items that in some fashion include the words. Next, one can perform some initial ranking based on how the keywords appear. If the query consists of multiple terms, items that contain all the terms can be ranked more highly than those that contain only some of them. Multiple occurrences of query terms increases relevancy. Where the terms appear can also increase the relevancy score. Query terms appearing in the document title or in section titles carry more weight than those that fall within text. These and many other nuances of textual analysis can be performed to determine a technical ranking of a results set.

Technical ranking, however, may not necessarily reveal the overall interest or importance of the items within the results set. Other factors, often having to do with more social measures, may be just as important. The number and quality of other documents that link to a

result candidate, and the number and quality of citations of the document increase relevancy.

To achieve successful relevancy ranking, one must explore a wide variety of clues that might reveal an item's level of interest and importance.

In a library context, a number of other factors might be considered. When ranking results from the library's book collection, the number of times that an item has been checked out could be considered an indicator of popularity. So might the number of copies owned by the library, or the number of libraries that, according to OCLC, own the book.

"Did You Mean . . . ?"

Another feature that has grown to be universally expected involves the ability for the search engine to detect common spelling errors in a query. Misspell a word in a Google search, for example, and we all expect it to respond, "Did you mean . . . ?" with a suggestion for a term that will work. A number of technical algorithms are available to help with this task. One can create an index of phonetically similar words that can be used to test the plausibility of a query. If a phonetically similar term returns a much larger results set than the one provided by the user, then the system might present that term as a suggestion.

A well-functioning "did you mean" feature goes beyond simply performing a spell-check on the query term. It's important to figure out if the term suggested will actually return more results than the original.

In the context of a library catalog, a "did you mean" feature can help prevent a very large number of failed searches. Giving the user a suggestion seems to be a much better option than the common message "No results found."

Recommendations

A common feature, especially in the e-commerce arena, involves proactively providing information about related materials. Amazon.com, for example, sports a prominent recommendation feature: "Users that bought X also bought Y." While the merchandising motivations of online bookstores may not apply to libraries, there may be a similar interest in promoting other materials in the collection. The challenge in the library environment might involve finding the user-behavior data on which to base these associations for recommendation.

Web 2.0: Enabling User Contributions

One of the key concepts of the Web 2.0 trend that has gained widespread interest involves blending aspects of social computing into a resource. In the spirit of Web 2.0, a resource isn't just a one-way presentation of information, but rather invites user participation and involvement. There are several ways in which this approach can be incorporated into next-generation library catalogs.

In addition to enriched content that might be obtained from external resources, a catalog can rely on its users to contribute supplemental content. Users can be invited to rate the items or to write reviews that express their opinions regarding works represented in the catalog. Other users might comment on these reviews or write their own.

While it is yet to be seen whether library users are interested in this level of involvement in the catalogs offered by libraries, enabling social features is gaining widespread use in many other arenas.

Tagging, another practice associated with the Web 2.0 movement, provides a convenient way for users to assign their own informal terms to items of interest. Completely unconstrained by formal rules, users make up their own tags, which they can use to find items later on. While some tags might relate only to a person's individual interests, some may be more community oriented. A "folksonomy" is a set of tags developed among a community to help classify a body of information.

Some very large collections, such as the popular photo-sharing site Flickr, rely solely on user-assigned tags for organization. Libraries, given their orientation toward more precise methods of organizing their collections, might balk at this approach. User-assigned tags can serve as an interesting supplement to the subject vocabularies traditionally found in library catalogs.

RSS

Distributing content through RSS (really simple syndication or rich site summary), in addition to conventional Web pages, provides the user with opportunities to use that content in more convenient ways. RSS delivers a set of related items through a simple XML protocol and today finds widespread use.

Library catalogs present a number of opportunities for the use of RSS:

- **Lists of new items in the collection.** This feature would allow a library user to subscribe to an RSS feed of new items as a notification service. Creating

feeds for each type of material or according to discipline makes the service even more powerful. A faculty member in the chemistry department who uses RSS to follow developments in the field could also subscribe to a library new-chemistry-books RSS feed, thus receiving this information from the library in a way that's highly integrated with the researcher's information-gathering style.

- **Lists of relevant items in other environments.** Offering search results as an RSS feed provides the ability to list relevant items within other portal environments, such as in a class page within the university's courseware application.

Each library or developer of library automation software has its own view of what constitutes a next-generation library catalog. We've explored some of the specific features that are found among the various products and projects. Each of the next-generation catalogs or interfaces follows a unique approach. The common thread among these products involves a desire to go far beyond the capabilities of the legacy library catalogs and give library users more powerful and appealing tools.

In the next chapters, we will examine some of the products that are currently available or soon to be available. The reports on each of the products aim to give a functional overview, focusing on the overall design and the features more than on the behind-the-scenes technology. These reports do not offer pricing information. As with most library software, the cost varies considerably depending on the size and complexity of the library and on the library's choices regarding optional components and services.

Note

1. See Edward T. O'Neill and Lois Mai Chan, "FAST (Faceted Application of Subject Terminology): A Simplified LSCH-Based Vocabulary," World Library and Information Congress: 69th IFLA General Conference, Aug. 1-9, 2003, Berlin, available online at www.ifla.org/IV/ifla69/papers/010e-O'Neill_Mai-Chan.pdf (accessed May 29, 2007).