How Do Users Search and Discover?

Findings from Ex Libris User Research

Christine Stohn

Senior Product Manager, Discovery and Delivery, Ex Libris

May 2015

Contents

Introduction

Users come to the library for a variety of reasons. For example, they may be seeking a specific item, be it a book, an article, an audio recording, or another type of material; they may need information for an undergraduate course assignment or a research project; perhaps they would like to obtain a quick overview of a specific topic; or they may want to find the latest publications in a specific field. Such research needs are often interconnected. A user might look up a specific article and then go on to seek an overview of the topic covered in the article.

In a discussion of users' research needs, additional factors come into play, such as the person's academic level and discipline. A student taking a course in Greek mythology who searches for *Andromeda* and a student of astronomy using the same query would be looking for different information altogether.

Discovery indexes cover a wealth of information. The Ex Libris Primo Central Index, for example, contains a billion records spanning materials of many types and varying depth. These materials are in addition to the vast local library collections, again a source of many valuable items available to users.

The challenging goal for library discovery systems, then, is to enable libraries to bring together users, their research intent and needs, and the wealth of available information. Intelligence about users' behavior, intent, and expectations in the context of the tasks that the users strive to accomplish is of key importance in endeavoring to achieve this goal. Such intelligence can help us, as developers of discovery systems, in a variety of areas, including:

- User interface design
- Search and ranking—understanding what the user's intent in a search is likely to be and what the system should push to the top of a result list to fulfill that intent
- Content decisions—determining what should and should not be included in a central index
- The normalization of data in a central index
- New features to support users in their work

Three types of research can provide particularly helpful intelligence about users' tasks, behavior, intent, and expectations:

- Log analyses, which shed light on users' behavior and preferences
- User studies, which help us understand what users need and what they need it for
- Usability studies, which demonstrate how well a system works for users

This paper describes the findings and conclusions from a recent Ex Libris user study and discusses how the conclusions apply to library discovery systems. The study included analyses of users' search logs and qualitative surveys conducted via workshops with librarians and interviews of users at different academic levels and from various countries and disciplines.

Research

Log Analyses

In an ongoing program at Ex Libris, usage logs are analyzed to shed light on various aspects of library users' search behavior and selection patterns. The logs that we study come from the Primo Central Index.

We look at several types of data in the Primo Central logs, including search queries; search refinements, such as the narrowing down of results through the use of facets; and selection activities, such as requests for the full text of electronic items. From these analyses, we can gain insight into what users search for and how they search for it. As a result, we know, for example, what types of searches need to be supported, and their variations.

Furthermore, by extracting such data from the logs, we can calculate key performance indicators (KPIs) and use them to assess the impact of modifications to the system.

Key Performance Indicators (KPIs)

To monitor trends and changes in search behavior, we calculate KPIs from the data in Primo Central log files. For example, a recent enhancement to Primo discovery and delivery solution provided additional support for searches for an entire citation, including the article title, author, journal title, volume, issue, and page numbers. A trend analysis of the logs showed that the number of such citation searches in which users subsequently selected the full text increased noticeably. This finding indicates that the enhancement was well received.

Examples of key performance indicators that we use are as follows:

- The amount of time users take to select an item from a result list
- The number of abandoned searches (searches in which users do not make any selections)
- The ratio of the number of selections to the number of searches and search sessions
- The ranking of the items that users select (in other words, do users click the first, second, or third item on the result list?)

Support for Various Types of Searches

Searches generally fall into two main categories: known-item searches and topic searches.

Known-item searches are those in which a user seeks a specific item, be it a book, an article, or another type of item. Primo Central logs reveal that *more than 50% of searches are for known items*.

Topic searches, also called exploratory searches, are those in which users are looking for information related to a certain subject but they do not have a specific item in mind. We distinguish between broad-topic searches, in which users are looking for general information on a topic, and narrow-topic searches, in which users typically seek more specialized information.

Through analyses of log files, one can identify characteristics of each kind of search. Queries for known items usually, but not always, consist of more than five words, can contain mixed case (which could be the result of copying and pasting from various websites, for example), may exhibit exactly the same phrasing as in the discovery system's title index, and occasionally consist of an entire citation (title, author, volume, issue, and page numbers) copied from a reference list or reading list and pasted into the search box. In addition, such queries may or may not include punctuation, and they may come with various combinations of metadata. Topic-search queries are usually shorter, containing up to five words.

Note that the classification of queries is not always straightforward. Some are ambiguous, such as a query that indicates a broad-topic search but exactly matches the title of an item, or a query that specifies authors and could therefore relate either to a particular work or to any work by the specified authors. For some examples, see Table 1.

Table 1. Types of queries identified from log files

Query from Log File	Type of Query
Motivating the academically unmotivated: A critical issue for the 21st century	Known-item
Introduction to Special Relativity,rindler	Known-item
economic dependency, gender, and the division of labor at home. american journal of sociology, 100, 652-688.	Known-item
cognitive systems research	Broad-topic
international marketing	Broad-topic
expert systems with applications	Narrow-topic
blaxter, hughes and tight	Author-related

To understand what topics users are searching for and how queries are most commonly formulated, we can extract the most popular searches from the log files. However, the parameter of the most popular searches cannot be used to determine which types of searches are most common. Users are not very likely to search for exactly the same known item with the same query as other users. But users are much more likely to create the same query to search for a particular broad topic. Therefore, among the most popular searches, one can find a large percentage of topic searches, although looking at the overall picture, we see that topic searches are actually a smaller percentage of all searches than known-item searches.

If the type of search is identified, the discovery system can adjust the result list accordingly. For example, when a query suggests that the user is conducting a broad-topic search, the system promotes overview articles and reference works in the result list. With such a query, the assumption is that the user does not have a specific work in mind but, rather, is trying to find out about the topic and start an exploratory search based on the first results.

All of these examples show how a search engine must support many variations in the phrasing of queries. As a result of the ongoing study of logs, along with feedback from librarians at Ex Libris customer sites, Primo is constantly enhanced to accommodate such variations. For example, support for citations that are pasted into the Primo search box has been expanded to include APA, Chicago, and MLA citation formats and variations thereof. Primo also handles common errors such as spelling mistakes, title words that are written in the wrong order, and leading articles that are missing.

In addition, if a search of metadata finds only a few results, Primo expands the search to include items in which the query terms are found in the full text. This expansion mechanism presents users with further material that may be relevant to their research needs.

User Studies

The purpose of user studies is to shed light on why users search and what they need the desired material for. Equipped with information from these studies, developers of discovery systems can add and improve features to help users achieve their goals.

Note that user studies are not the same as usability studies. The purpose of the latter is to assess "the ease with which a computer interface can be efficiently and effectively used, especially by a novice" (JM Reitz, *Online Dictionary for Library and Information Science* [ODLIS], http://www.abc-clio.com/ODLIS/odlis_about.aspx).

In 2014, Ex Libris conducted a user study whose results are to be applied for enhancing the discovery experience and strengthening its effectiveness. Specifically, the goals of the study were to gain a better understanding of the similarities and differences between the various types of users, detect recurring themes related to their searches, and identify significant characteristics of discovery and the supporting features that can help users excel in their academic goals. The subjects of the studies included active library users as well as those who seek to satisfy their information needs elsewhere, such as in Google.

Methodology

As explained earlier, the purpose of the user study was to obtain practical information that will help improve and enhance the discovery experience. The study was not intended to be a scientific investigation. For example, some users underwent interviews, whereas others filled out a questionnaire. And the questionnaire was later adjusted slightly as a result of the first set of responses.

As qualitative rather than quantitative research, the study focused on a relatively small sample of users; it was not large enough for statistically significant results, but it was sufficient for obtaining insights into what is needed for facilitating discovery. When we asked users about their mobile devices, for instance, we were not interested in finding out how many users employ a mobile phone or tablet; we wanted to learn about their attitude toward mobile devices and the ways in which they use such devices.

The first step of the study was to elicit user scenarios from librarians. We conducted workshops at two universities, involving librarians from 17 institutions, and consulted informally with librarians from other schools.

To obtain various types of information about users' needs and preferences, including user scenarios, we interviewed and sent questionnaires to students and researchers (all of whom we shall refer to as users). Our sample of users totaled 24, of whom 14 were female and 10, male. For details about the sample, see Table 2, Table 3, and Table 4.

Table 2. Geographic distribution of users in the sample

Country	No. of Users
USA	14
Germany	5
Canada	2
UK	2
Israel	1

Table 3. Research areas of users in the sample[1]

Area of Research	No. of Users
Business or economics	4
Humanities	4
Social sciences	4
Sciences	3
Medicine	2
Architecture	1
Biotechnology	1
Design	1
Engineering	1
Law	1
Psychology	1
Urban planning	1

Table 4. Academic levels of users in the sample

Academic Level	No. of Users
Undergraduate	7
Master's student	9
PhD student	5
Faculty member/researcher	3

In addition to a section about the person's area of study and academic level, the questionnaire included a variety of questions related to the person's research needs and habits, use of mobile devices, and likes and dislikes in library searching (for examples of questions, see Figure 1). The last question, in which users are asked to describe a task, was intended to elicit a user scenario.

[1] The respondents described their area of research in their own words. Therefore, the names of the research areas in this table are not necessarily consistent.

What literature do you need the most for your work?

- ☐ In-depth research articles
- ☐ Articles that provide an overview of a topic
- ☐ Definitions (to understand the topic better)
- ☐ Textbooks
- ☐ Other (please specify)

Where do you usually get literature from?

- ☐ I search for it myself.
- ☐ I just use what is provided by the course reading list.
- ☐ I go to the library and ask a librarian.
- ☐ Other (please specify):

Which of the following describes your situation best? (Choose 3)

- ☐ I need information quickly.
- ☐ I search and expect a few good results.
- ☐ I search and expect to find all the literature that is relevant for my topic.
- ☐ I like to browse and find interesting literature that I didn't know about.
- ☐ I like to expand my knowledge about the topic first. I start with Wikipedia, and after I have found enough terms, I look for more specific literature.
- ☐ Other:

Where do you usually start your search for literature?

- ☐ Google
- ☐ Google Scholar
- ☐ Library catalog
- ☐ Other (please specify):

Please describe a specific recent task, e.g., a course assignment or research project (or a repetitive task that you do often, such as keeping up to date in a specific area) that required you to search for literature. Please be as specific as possible; for example, list the title of the assignment or project and the type of information you needed for it.

Figure 1. Examples of questions and multiple-choice answers from the questionnaire

We encountered several methodological challenges with the questionnaire. Finding terminology that is clear to everyone is one such challenge. Variations in the users' background and origin meant that people might interpret terms differently from the intended meaning or just fail to understand a term altogether. In interviews, terminology issues can be dealt with, but questionnaires do not offer that flexibility. Similarly, people don't necessarily use the same terms in responding to questions, so comparing answers can be difficult.

Leading questions are a problem in questionnaires in general. If a questionnaire provides answer options, they can influence the respondent; on the other hand, the absence of options can lead to misinterpretation. In this study, we decided to use a multiple-choice format, but we were able to view the answers in context by looking at the user scenarios at the end of the questionnaire.

User Scenarios

One of the most important parts of the questionnaire was the last question, a request for a specific scenario involving a task that the user had recently carried out.

Here are three examples of user scenarios:

> **Undergraduate student in medicine**: One of this student's most recent tasks was to gather more information about a specific therapy (photopheresis) used for graft-versus-host disease. Most important to him was gaining a general overview of the topics and the current status of research. He started by searching for photopheresis in Google but didn't find anything useful. However, once he had found additional material, after asking his peers for help, he enjoyed delving into the subject.
>
> **Researcher in history**: In researching a very narrow topic about migration to the United States about 200 years ago, it was necessary to consult older publications dating from 1820 to 1940 to find more first-hand material. No single book or journal article was comprehensive, so this user had to review many titles to find the information.
>
> **Graduate student in physics**: This student needed to find research papers to prepare for a talk about semiconductor optical amplifiers. The goal was to present both recent developments in the field and original publications about the topic to explain the broader concepts.

Results and Discussion

The questionnaire and user scenarios revealed a number of recurring themes, some of which confirmed what we already knew and some of which were new to us.

First, we found that searching is accompanied by learning. For example, master's students often need to begin their research by learning the relevant terminology; they may start with creating a list of keywords that they then use for searching for material about their research area. Only after finding the general information do they drill down to specific topics. Undergraduates also often need to understand the general topic of their assignment before they move on to more specialized material.

Undergraduates tend to use resources that are on course reading lists. However, to do research for term papers and projects or when striving to achieve good grades, students often start with the reading lists and then navigate beyond them—an example that demonstrates the importance of integration between reading lists and the library's discovery system.

More experienced scholars are usually familiar with their field's core publications. Such researchers are often, but not always, interested in the latest publications. Starting, for example, from an article that they already know about, they like to find related material by looking at the list of references and following a citation trail or by looking at other publications by the same author. These researchers tend to prefer extensive result lists, whereas undergraduates tend to be satisfied with a few relevant items.

Graduate students often work on projects where they need both a general overview and specialized aspects of a topic, so they look for a mix of materials rather than just in-depth information.

The research needs of students also differ by discipline. Undergraduates in the humanities often write papers and essays and therefore need all kinds of information, including biographical material, secondary-source material, and primary-source material. In the sciences, learning is more exam oriented. Students tend to use textbooks and e-learning software to achieve their goals. However, the students also need to search for material when assigned a presentation on a specific topic.

Another theme that became clear is that research needs in a particular discipline can vary from one country to another because of the differences in the way the discipline is taught. For example, medical students in the United States start their practical work very early on and therefore have practical questions and needs, whereas in Germany, the first few terms of medical school are theoretical, so the students' needs differ from those of their American counterparts. Medical students in Germany often study from e-learning software and textbooks.

One of the general conclusions that we can draw from the user studies is that many students, undergraduate and graduate alike, include learning as a desired part of their information research. Also, researchers, graduate students, and undergraduates find it helpful to draw on the knowledge of their peers and like to seek relevant material by following navigation trails, in addition to conducting searches.

Regarding mobile devices, only a few of the respondents use tablets. Several students said that tablets are too expensive for them; these students consider laptops and smartphones more useful. One respondent works on a desktop computer. Those who have a tablet use it for quick tasks and for doing work on the go, such as reading articles.

The question now is how a discovery system can incorporate these findings to help researchers at all academic levels attain their goals.

Implications for Discovery

Discovery consists of several core concepts:

- Search and find
- Learning
- Exploration
- Personalization

The area of search and find was discussed earlier in reference to log analyses and remains a major driver in the development of discovery systems. Regarding learning, it became clear from our user studies that learning is part of the discovery process, closely related to exploration, and an outcome that users benefit from. In this section, the focus is on the concepts of exploration, learning, and personalization, and their implications for discovery systems.

Exploration and Learning

Users explore the sea of information with various tools, all of which support learning and can lead to serendipitous discovery. Many of these tools rely on relationships between different items. For example, recommendations are based on usage data; if one item is selected frequently along with another item, the recommender system considers them related. If one work cites or is cited by another (as seen in reference lists), they are considered related, as are items in topic clusters and items by the same author. Another exploration tool is a browsing functionality that enables a user to look at material that is similar to or related to a known title, author, or subject.

A typical exploration begins with the user's selection of an item—for example, a subject, an author, or a title. The discovery system presents another group of items, from which the user makes a selection. Then the user continues to navigate onward to additional sets of related items.

Take the browsing feature of Primo, for example. Starting with an author, subject, or title, the user can browse through a list and discover items by virtue of their alphabetical proximity to that author, subject, or title (Figure 2).

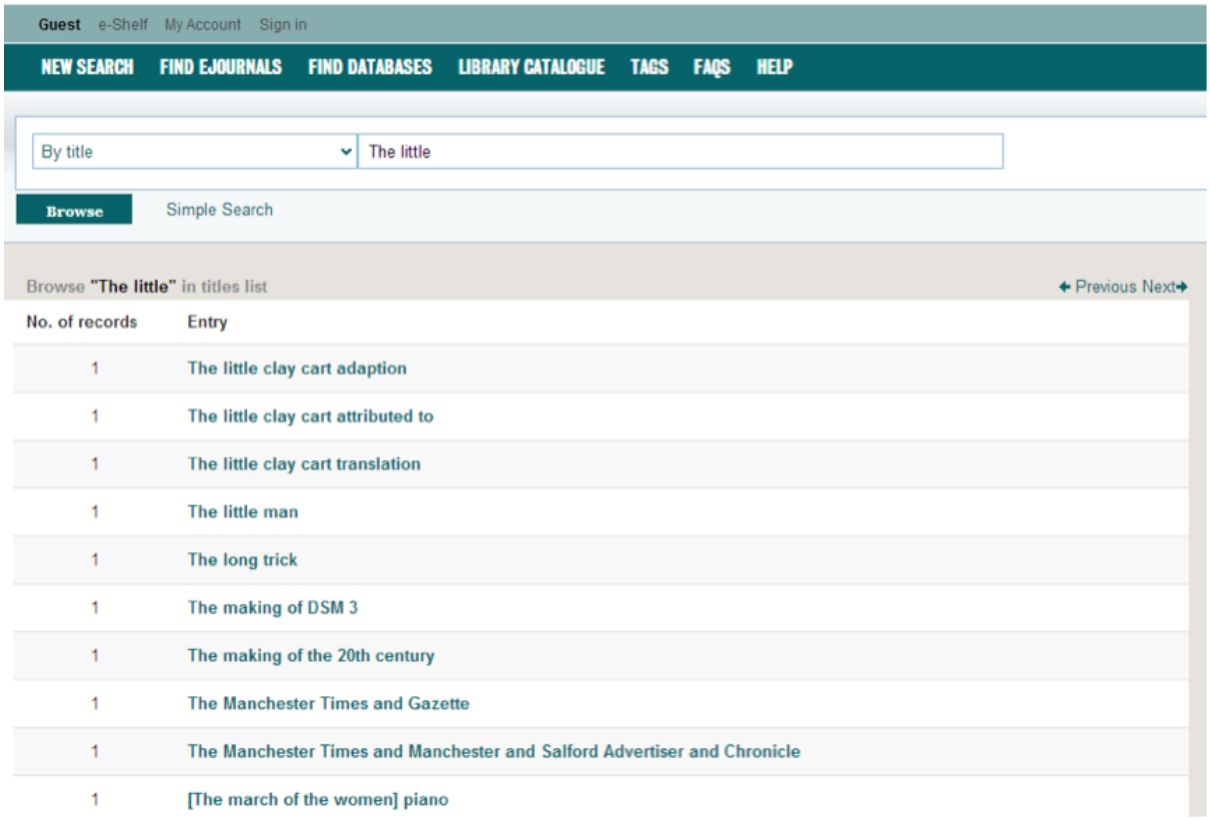

Figure 2. Exploring titles through browsing

Another useful feature is virtual browsing, the online equivalent of walking over to a bookshelf, finding a book of interest, and then discovering other books of interest near it on the shelf (Figure 3). This type of exploration is also very much about serendipitous discovery within the context of the first item that the user finds.

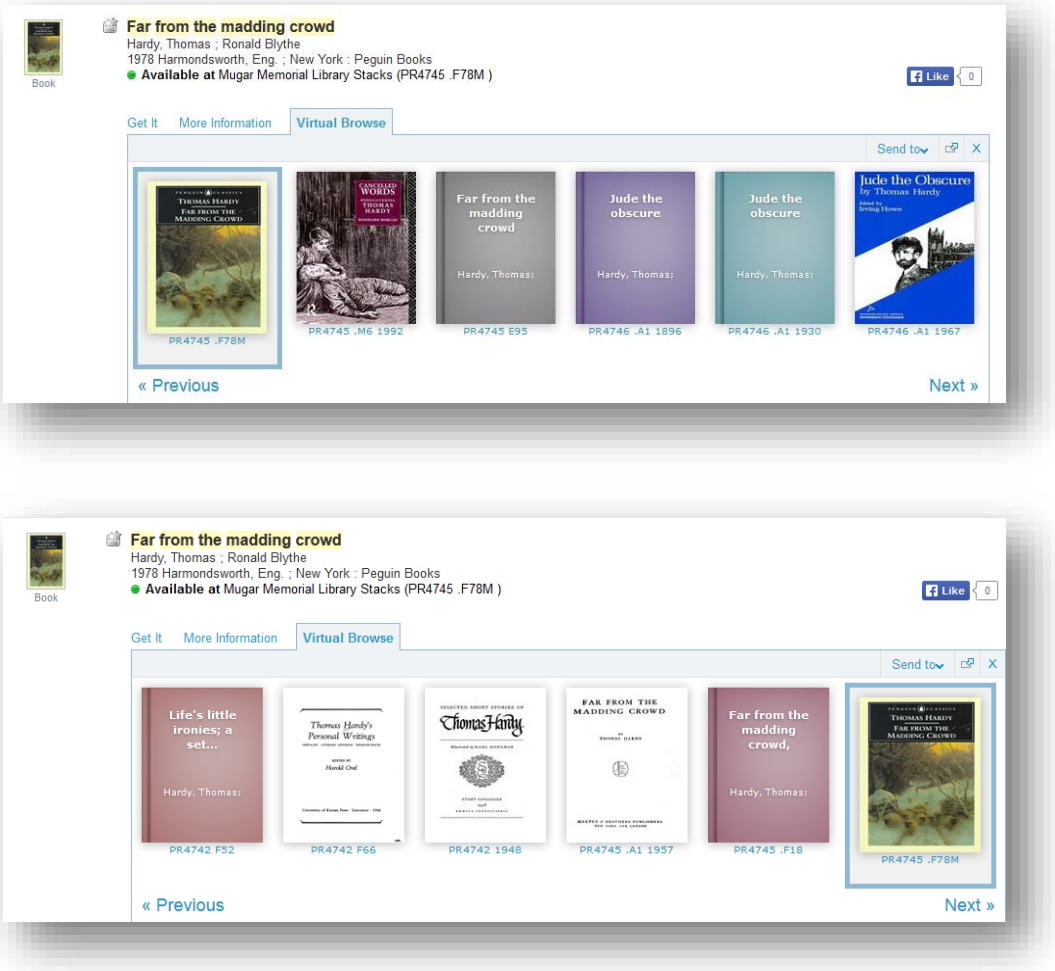

Figure 3. Exploring through virtual browsing

Recommendations provide additional sources of information. Just like the Amazon feature "Customers Who Bought This Item Also Bought…," the Ex Libris bX Recommender service can take users along a trail of discovery. For example, a student working on a paper about the sugar revolution would like to find information that goes beyond the course material. A search for *sugar revolution* yields a mixed result list, because of the query's ambiguity, and provides overview material on the topic. After selecting an article that appears at the top of the list and is related to economic history, the student looks at the recommendations and finds items on topics that are connected to the sugar revolution, such as the article "The Slave Trade, Sugar, and British Economic Growth, 1748-1776" (Figure 4). Clicking this article yields another list of recommended articles with additional useful information, and so the exploration continues. The process saves time, helps the student discover relevant material efficiently, and also constitutes a learning experience by helping the student understand the relationship between the sugar revolution, slavery, and British economic history in the 18th century. The exploration, therefore, not only introduces historical connections but also presents new facets of the topic that the student may not have known about.

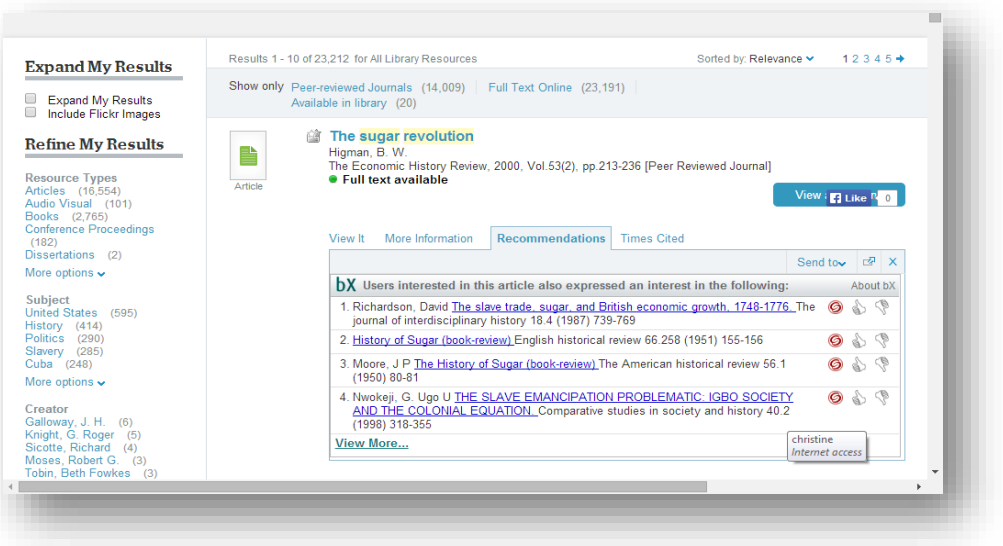

Figure 4. Exploring through recommendations

Another way to explore is via reference lists. When authors cite a work, it is clearly related to the topic at hand and may provide valuable information. To facilitate exploration via reference lists, discovery systems need to offer support for citation trails: from a selected article, users will be able to go to items that it cites and to items that cite it. As with recommendations, users will embark on a trail of discovery.

The scope and accuracy of exploration can be improved through the use of advanced technologies such as text mining and semantic enrichment. With such technologies, core topics and their meanings can be automatically extracted from an item and used for clustering—for example, to link an item to others with the same focus—and for enhancing searching. In today's world of big data, the opportunities for text mining and semantic enrichment are huge. They provide a scalable way to deal with hundreds of millions of items and can extract and normalize entities and concepts across the index by analyzing the items' metadata and full text. A number of scholarly publishers already use text mining to enhance discovery and exploration features for their content. Currently in the research stage at Ex Libris, text mining has great potential as a search and exploration aid in the discovery process.

Personalization

The user studies that we conducted and informal discussions that we held with librarians show that the same thing does not work for all users. One scholar who submits the query *game theory* might be seeking material in economics, and another scholar, material in mathematics.

Primo already incorporates personalization in the ranking of search results. If you identify yourself as a physics student and search for *graphene layer*, you will obtain many results. At the top will be results from physics journals. However, if you sign in as an engineering student and run the same search, you will receive the same number of results, but they will be ranked differently; Personalization means that the discovery system will adjust the ranking of results to a user's profile and thus will give more weight to engineering journals for the engineering student described in the previous example. Personalization options need to be available to all users, even those who are not logged on, to help them tailor search results to their needs and preferences.

In light of the masses of material available today and users' specific interests, personalization can play a central role in discovery. Users ought to be free to select the personalization options that they would like to take advantage of. These choices can influence the result list by ranking certain journals higher than others and boosting materials on the basis of their date of publication.

In the context of personalized result lists, it is worth digressing for a moment and looking at anticipatory discovery in the Google world. Anticipatory discovery adjusts the result list according to the user's search history; in other words, the system anticipates the user's intent on the basis of the user's past searches. The concept is appealing. However, in the library world, a number of concerns arise. First, the storage and analysis of an individual's discovery behavior raise privacy concerns. Second, such functionality requires a critical mass: the user has to have engaged the system enough times to leave a trail of analyzable patterns. Our user studies suggest that this critical mass is not necessarily reached. Undergraduates in the humanities, for instance, do not regularly seek new material at the library; they find most of their material in their reading lists. For a thought-provoking look at discovery and personalization (including privacy issues), see Robert S. Schonfeld's "Data for Discovery" (http://scholarlykitchen.sspnet.org/2015/02/05/data-for-discovery/).

The advantage of anticipating users' interests is that the result list is more targeted. But does this benefit also limit a user? Does the user get caught up in a bubble of information instead of learning to look beyond the known to find the unknown in a huge amount of data? Great opportunities lie in personalization and anticipatory discovery, but one must be sensitive to the privacy concerns of institutions and their users.

Discovery with and without Boundaries

Today subscriptions, purchases, and open access are not the only ways to obtain a document. Libraries can extend their reach through mechanisms such as patron-driven acquisition and article rental. However, many libraries choose to restrict users' discovery to only those materials that the libraries have access to physically or electronically. For undergraduate users, this limitation is often satisfactory, because they need only a few of the items that are available to them.

In a more versatile model, the library lets its users discover everything; if they want an item that the library doesn't have, the library obtains the item for them. This model offers users discovery without boundaries.

Navigation via citation trails and tools such as the bX Recommender service already present users with "everything" rather than just a restricted set of items. In fact, exploration is very much about discovery without boundaries. The flexible delivery methods that are available today can make this type of discovery a viable solution, especially for libraries that serve a diverse community consisting of undergraduates, graduate students, and researchers.

Conclusions

The core functionality of a discovery system is search and find. Users search in a variety of ways for specific items and for information on general topics. With the intelligence that log files provide on how users search, we are able to constantly enhance discovery capabilities to facilitate searching and help users satisfy their information needs.

However, search and find is not the only discovery practice that is important to the user of today. Our user studies clearly show that undergraduates and more advanced students rely on the search

process both to find material and to learn about the topic. They may start with a general topic and then become acquainted with various aspects of it while searching and looking at various materials. Exploration via connections between items, such as those provided by recommendations, citation trails, or other metadata connections, offers an efficient way to obtain relevant material and, by exposing users to items that they may not have known about, also fosters learning. Similarly, browsing and virtual browsing support serendipitous discovery. With today's big data technologies, data can be analyzed, enriched, and clustered across very large indexes. Ex Libris is investing development efforts in this area to enhance the discovery capability for the hundreds of millions items indexed in Primo Central.

The user studies that we conducted made clear that different users have different needs. We believe that a holistic approach based on the building blocks described here is essential for strengthening user engagement and delivering the fast, accurate, and intuitive discovery experience that patrons seek. At the same time, it provides libraries with the ability to increase the impact of the service that they provide.

About the Author

Christine Stohn is a Senior Product Manager for Ex Libris' Discovery and Delivery solutions. She focusses on areas such as the Primo discovery search engine, the Primo Central Index, and capabilities that enhance the user experience such as the bX Recommender. Christine holds over 20 years of experience in the library industry, both in the content and the technology aspects. Christine holds a degree in library sciences from the Free University in Berlin and an information systems degree from the Open University in the UK.