

## From Search to Discovery

**Tamar Sadeh**

Department or Division Name, Ex Libris Group, Jerusalem, Israel.

[tamar.sadeh@exlibrisgroup.com](mailto:tamar.sadeh@exlibrisgroup.com)



Copyright © 2013 by **Tamar Sadeh**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

---

### Abstract:

*In recent years, library users have shifted from searching in library catalogs and scholarly databases to searching in library discovery systems. This shift has introduced a fundamental change in the information-seeking process.*

*Discovery systems provide access to a large, diverse information landscape of scholarly materials—regardless of where the materials are located, what format they are in, and whether the library owns them or subscribes to them. At the same time, these systems typically offer simple, Google-like searching as the default option, to accommodate the expectations of today's users. With this type of searching, users do not spend much time formulating queries, and their queries often yield large result sets; therefore, discovery systems focus on relevance ranking and on tools that help users easily navigate and refine result sets.*

*Librarians have welcomed the advances in discovery services for their users. However, this new reality poses challenges to the practices that librarians have developed over the years and, in some cases, is at odds with the systematic, controlled approach to searching endorsed by librarians.*

*An examination of information-seeking and information-searching models together with a review of new technological capabilities of library discovery systems shows why such systems help today's searchers and facilitate their research more than the traditional systems.*

**Keywords:** Discovery systems, user experience, information seeking, relevance ranking

---

## Introduction

The immediacy of information, availability of communication channels, abundance and diversity of tasks that people routinely accomplish online, and the effects of social networks shape the expectations of users when they are looking for scholarly materials. Surveys and reports published since 2005 (reviewed in Sadeh 2011) clearly demonstrate that traditional library systems are lagging in popularity behind non-library information systems. These surveys and reports describe a reality in which library users have shifted from searching and accessing content via library services to attempting to satisfy their information needs through non-library services, such as web search engines, online bookstores, blogs, online news, and e-mail. Statistical data support these trends. For example, the Association of Research Libraries (ARL), whose membership spans North America, reports that reference transactions at ARL institutions decreased by 65% and circulation transactions decreased by 29% between 1991 and 2011, despite an increase of about 33% in the number of students enrolled in those institutions during that period (see <http://www.arl.org/storage/documents/service-trends.pdf>).

Markey points out that libraries' "failure to respond accordingly may permanently exile scholarly and scientific information to a netherworld where no one searches while less reliable, accurate, and objective sources of information thrive in a paradise where people prefer to search for information" (Markey 2007); this concern has driven librarians to think of ways to retain their users and maintain their leadership as academic information providers. However, to address users' expectations regarding the search interface, the breadth and relevance of services, and the comprehensiveness of the body of information available through a library system, libraries have had to undergo a major conceptual shift.<sup>1</sup>

The rise of "next-generation catalogs," which were introduced in January 2007 with the Endeca-based system at North Carolina State University marked a significant change in the way in which libraries engage with their users. Initially aiming to provide a modern, friendly user experience for library patrons, these systems, termed *discovery systems* or *discovery-and-delivery systems*, transformed the way in which users search in library catalogs and other local library collections such as institutional repositories, course materials, and even library web pages. However, only in 2009, when discovery systems shifted gear with the introduction of a central index,<sup>2</sup> could these systems serve as a single point of access to the entire library collection.

Discovery services that include a central index—such as Ex Libris Primo,<sup>3</sup> Serial Solutions Summon,<sup>4</sup> EBSCO Discovery Service,<sup>5</sup> and OCLC WorldCat Local<sup>6</sup>—have been adopted at a rapid pace, and thousands of institutions are making their local and global collections available to their users through such systems (Breeding 2007). Nevertheless, many librarians do not endorse these systems.

---

<sup>1</sup> For a discussion of user preferences, see Sadeh 2010, 2011.

<sup>2</sup> A central index that serves a library discovery system covers scholarly materials such as journal articles, e-books, newspaper articles, reviews, and legal documents (referred to in this paper as global materials). In addition to these materials, which all libraries use, a discovery system indexes local materials, such as the library catalog and digital repositories, for each participating library.

<sup>3</sup> <http://www.exlibrisgroup.com/category/PrimoOverview>

<sup>4</sup> <http://www.serialssolutions.com/en/services/summon/>

<sup>5</sup> <http://www.ebscohost.com/discovery>

<sup>6</sup> <http://www.oclc.org/worldcat-local.en.html>

## What do we mean by “discovery,” as opposed to “search”?

In this paper, traditional library information systems, such as library catalogs and databases, are referred to as *search systems*. Search systems offer structured search interfaces that are tailored to the specific data that they hold; the records are homogenous—they are cataloged in the same way, have the same data structure, and often relate to one topic (as in a subject-related database). Search systems typically expect users to possess medium-to-high searching literacy and enable users to accurately define their information need. Although in recent years such systems have also simplified their search interface and incorporated post-search refinement options, librarians encourage users to develop searching strategies and take advantage of the rich options of the systems’ search interface.

Library discovery systems, despite some differences between them, share several major characteristics, described in the following sections.

### User experience

The need to improve the user experience was the trigger for the development and deployment of discovery systems and has become the cornerstone of these systems. The introduction of the term *user experience* in the library realm—replacing the traditional term *user interface*—bears great significance. Although the term has been defined in various ways,<sup>7</sup> the following definition by Nielsen Norman Group is a good fit in the context of software systems:

"User experience" encompasses all aspects of the end-user's interaction with the company, its services, and its products. The first requirement for an exemplary user experience is to meet the exact needs of the customer, without fuss or bother. Next comes simplicity and elegance that produce products that are a joy to own, a joy to use. True user experience goes far beyond giving customers what they say they want, or providing checklist features. In order to achieve high-quality user experience in a company's offerings there must be a seamless merging of the services of multiple disciplines, including engineering, marketing, graphical and industrial design, and interface design.<sup>8</sup>

The designers of traditional library information systems, such as library catalogs and databases, were very focused on meeting the needs of librarians and expected that users would invest time and effort in learning how to use the system. The designers of discovery systems, driven by the needs of end users, strive to streamline the end-to-end process of finding information and make it as simple and friendly as possible. Rather than offering multiple options to enable users to describe their information need, discovery systems offer users simple search interfaces but complement these with multiple post-search options for assessing findings, refining results, and navigating to other results of possible interest. The look and feel of the interface is similar to that of other information systems that are familiar to users, such as web search engines and online bookstores. Furthermore, recognizing that today’s users spend hardly any time reading instructions, developers have made discovery systems very intuitive. The integration of relevant services such as a personal “e-shelf” for filing relevant results, OPAC functions (e.g., for requesting books and managing one’s loans), and citation management tools enhance the user experience and help make a discovery system more usable and comprehensive.

---

<sup>7</sup> See <http://www.allaboutux.org/ux-definitions>.

<sup>8</sup> Nielsen Norman Group (<http://www.nngroup.com/about-user-experience-definition/>)

## **Content**

One of the major concerns of users in the pre-discovery era was the fragmentation of the search scope. To satisfy their information needs, users had to search in various scholarly systems, such as library catalogs, digital repositories, and remotely hosted third-party databases. Metasearch systems, introduced at the beginning of the millennium, were the first step in addressing the need for a unified search capability that spanned multiple resources. Despite their impressive adoption,<sup>9</sup> metasearch systems have several drawbacks that are inherent to “just-in-time” processing. The main drawbacks are slowness (the speed of the search depends on the response time of the databases in which the user wishes to search) and the small number of results displayed (the searched databases initially return only a few results, which are arranged in an order dictated by the databases). In addition, metasearch systems do not cover the whole spectrum of library resources.

One of the major goals of discovery systems, particularly since the introduction of central indexes of global scholarly materials, has been to offer the same experience of comprehensiveness as web search engines, but in a manner that supports scholarly values and practices. The scholarly information universe, where content is described and controlled by professionals, is available to users of discovery systems for searching through one library-defined interface. Such a search environment is compelling for the majority of users: it supports simple, straightforward queries and frees users from having to select the most appropriate information resources for searching, especially when their information need is interdisciplinary. Furthermore, it enables them to find new materials that they were not aware of.

The content available to discovery systems—both local and global content—can be further enriched through the incorporation of data elements such as book cover images, tables of contents, abstracts, and information about authors. By integrating such data elements, discovery systems improve the search experience of the users and help them identify and assess the relevance of items. Today’s technology enables content from disparate systems to appear as if it came from one system, rendering the search experience easier, faster, and richer.

## **Search process**

Discovery systems were designed at a time when Google had already set the standard for searches, with its very simple search interface, excellent relevance ranking, rapid performance, and extensive content coverage. Rather than expecting the user to develop search strategies, Google accommodates a trial-and-error approach: the interaction with the search engine is so simple and quick and the relevance ranking is perceived by users as so good that if one of the first results does not satisfy the information need, the user typically rephrases the query and tries again.

Developers of library discovery systems aim to provide a similar experience. To that end, such systems leverage the specific nature of scholarly publishing practices and of the indexed data. Several factors contribute to the systems’ success at meeting the needs of library users.

First, the content indexed by discovery systems is formatted much better than that indexed by Google. Such content consists of bibliographic records (although not all of them conform to the same standard) and textual information (abstracts and full text, when available).

---

<sup>9</sup> Ex Libris alone reports that over 1,700 institutions have purchased the MetaLib metasearch system.

Structured data allows for a more efficient discovery: for example, the search engine can easily differentiate between information that is more significant—such as an article’s title, subject, or author—and information that has a smaller impact on the search process. Discovery systems leverage the data structure by offering options for searching in specific fields<sup>10</sup> and by assessing the relevance of items based on the fields in which the query terms were found.

Like web search engines, discovery systems do not assume great searching expertise on the part of users, who may well submit queries that yield large result sets. In addition to relevance ranking, which aims to display the item that the user was looking for at the top of the list, discovery systems offer a means of refining result sets through the selection of various attributes, or *facets*, of the desired items. Such facets may relate to administrative or structural information such as material availability (e.g., items that are available online or print items that are currently in the library), type of material, language, year of publication, the journal in which the item was published, or the collection in which it was originally made available. Facets may also relate to the content’s description; the user may opt to refine the search results by selecting a specific topic or author. Some discovery systems enable the user to include and exclude characteristics of the desired material—for example, the user may decide to refine the list of results for the query *global warming* to show items that relate to climate change and greenhouse gases but not to air pollution.

Searching is not the only way in which users find items of interest. Users often follow leads from one item to another. Therefore, an important feature of a discovery system is the interlinking within the broader information landscape. Once a user finds an item of relevance, this item becomes an anchor; the system then enables the user to navigate to related items, which the system singles out on the basis of similarities of data (the same author or subject, for example) or usage analyses (for example, to generate recommendations of the form “users interested in this item also expressed an interest in...”). The fact that a discovery system’s information landscape is so large and diverse enables searchers to find relevant materials that they could not specify.

### **Delivery and integration**

The discovery of relevant materials is only the first part of the interaction between the user and the system. The second part, which is equally important for the user, is obtaining an item, either physically or electronically. A discovery system integrates various services that facilitate delivery, such as context-sensitive link resolution, document delivery, citation manager tools, and hold or photocopy requests. Optimally, the system should serve users throughout their interaction with the library, offering personal services that are not necessarily related to a specific search, such as enabling users to change their password, edit their profile, review current loans (and renew them, if necessary), and view fines.

The seamless integration of a discovery system with other systems is crucial to its success as the library platform of choice. Such integration may relate to the institutional infrastructure—e.g., an authentication system—or other institutional platforms such as course management systems.

---

<sup>10</sup> Typically these options are offered as part of an “advanced” search functionality.

## **Why is the transition so difficult?**

Despite the rapid adoption of discovery systems, some librarians remain skeptical. At conferences where discovery systems are discussed, the implementation teams of such systems typically refer to the reluctance of librarians, primarily reference librarians, to embrace the new system.

First, librarians challenge the approach that does not require users to clearly formulate their information need. According to this approach, users deploy trial-and-error methods and trust that they will be able to identify the desired items. However, librarians attribute great importance to the mental query process, believing that the systematic articulation of an information need is a scholarly asset that must be learned and exercised. Whether or not an item is found is less important than the building of a search strategy.

Second, the information landscape indexed by a discovery system is heterogeneous, covering huge amounts of data of various types and from various sources. Consequently, the data is not uniform and groups of records may be subject to different cataloging rules, may adhere to different topics or subject-heading norms, and may vary in the amount of available metadata and data. Furthermore, at the time of this writing, there is not one discovery system that indeed covers the full spectrum of scholarly resources, even though all such systems cover the great majority of resources.

This reality makes librarians doubt users' ability to find the items that are likely to best satisfy their information needs. Librarians claim that users are happy to settle for "good enough" results and may not notice, or care, if they miss better ones. Such concerns arise, at least partly, from the fact that by breaking the traditional silos of information, discovery systems require that librarians invest more time and effort in learning the exact coverage available to their users.

Change has been occurring since the beginning of human civilization, helping our society move forward. Often, such change discards common beliefs, traditions, and practices. In many cases, people embrace change easily; however, at times they find it difficult to adjust and feel that the gain is not worth the loss. There is no doubt that the switch from search to discovery is a significant change for libraries, offering them new opportunities to better serve their users. Is the reluctance to adopt discovery systems as a replacement for structured, controlled, topic-related and material-related search systems justified? Or should librarians embrace the change, understand its benefits, and help develop it further?

To better address this question, let us look at models of human information seeking and see how well discovery systems support human behavior.

## **Modeling human information-seeking behavior**

In this paper, information seeking is defined as information behavior that consists of an active pursuit of information by scholars through the use of information systems. This definition acknowledges the reliance of searchers on finding information that they do not explicitly specify but that is nevertheless vital to their work tasks. Information-*searching* behavior is the aspect of information-seeking behavior that deals especially with active, directed searching in information systems for data that can be specified to some degree (Wilson 1999) (Figure 1).



Figure 1: Wilson's model: from information behavior to information searching. Modified from Wilson (1999), figure 11.

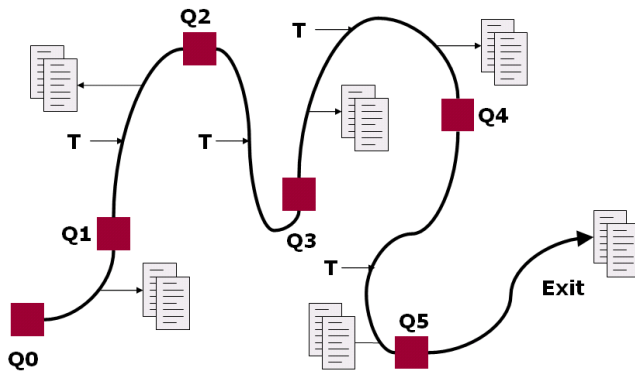
According to Wilson, information-searching behavior “is the ‘micro-level’ of behavior employed by the searcher in interacting with information systems of all kinds. It consists of all the interactions with the system, whether at the level of human computer interaction (for example, use of the mouse and clicks on links) or at the intellectual level (for example, adopting a Boolean search strategy or determining the criteria for deciding which of two books selected from adjacent places on a library shelf is most useful), which will also involve mental acts, such as judging the relevance of data or information retrieved” (Wilson 2000, 49).

Several models that shed light on the way people look for information have influenced much of the research in the area of information seeking and searching. In 1989, long before information systems were readily available to the general public, Bates suggested the berrypicking model (Figure 2). This model, which applies to a range of search techniques, offers a new perspective on the nature of the query and the search process and embeds the search process in an information territory.

Bates explains:

Users may begin with just one feature of a broader topic, or just one relevant reference, and move through a variety of sources. Each new piece of information they encounter gives them new ideas and directions to follow and, consequently, a new conception of the query. At each stage they are not just modifying the search terms used in order to get a better match for a single query. Rather the query itself (as well as the search terms used) is continually shifting, in part or whole.” (Bates 1989)

Bates also emphasizes that the search is taking place in the context of a “universe of interest,” which is within the larger context of the “universe of knowledge” (Bates 1989). The behavior of the searcher is the focus of attention in this model, and the “continuity represented by the line of the arrow is the continuity of a single human being moving through many actions toward a general goal of a satisfactory completion of research related to an information need. The changes in direction of the arrow illustrate the changes of an evolving search as the individual follows up various leads and shifts in thinking” (Bates 1989).



Q = query variation T = thought  documents, information

Figure 2: The Bates berrypicking model, with an evolving search. Modified from Bates (1989), Figure 2.

Although not necessarily applicable to computerized systems, this model fits well with scholars' use of discovery systems. With the exception of a known-item search—a search for a specific item (which is likely to appear at the top of the result list)—the initial search often leads to an interaction with the system whereby the user either rephrases or refines the search or follows leads, navigating from one item to another. Search logs indeed demonstrate such behavior (

<b>Searcher A</b>			
03:49:24	5432 document(s)	(296 ms) query '(forensic expert evidence in criminal trials)	[1,10]
03:57:02	1 document(s)	(375 ms) query '(Forensic botony)	[1,10]
03:57:16	1638 document(s)	(112 ms) query '(Forensic botany)	[1,10]
03:59:24	1638 document(s)	(135 ms) query '(Forensic botany)	[11,10]
07:23:35	42939 document(s)	(342 ms) query '(forensic evidence in court)	[1,10]
<b>Searcher B</b>			
05:36:01	191 document(s)	(195 ms) query '(forensic ink mass spectroscopy)	[1,10]
05:44:57	9619 document(s)	(291 ms) query '(forensic analysis glass)	[1,10]
05:45:18	1112 document(s)	(337 ms) query '(forensic glass mass spectroscopy)	[1,10]
05:45:28	2743 document(s)	(247 ms) query '(forensic glass mass spectrometry)	[1,10]
07:15:44	33949 document(s)	(403 ms) query '(forensicanalysis of paper)	[1,10]
07:47:25	15 document(s)	(916 ms) query '(Forensic Document Examiners? Skill in Distinguishing Between Natural and Disguised Handwriting Behaviors)	[1,10]

Figure 3).



<b>Searcher A</b>			
03:49:24	5432 document(s)	(296 ms) query '(forensic expert evidence in criminal trials)	[1,10]
03:57:02	1 document(s)	(375 ms) query '(Forensic botony)	[1,10]
03:57:16	1638 document(s)	(112 ms) query '(Forensic botany)	[1,10]
03:59:24	1638 document(s)	(135 ms) query '(Forensic botany)	[11,10]
07:23:35	42939 document(s)	(342 ms) query '(forensic evidence in court)	[1,10]
<b>Searcher B</b>			
05:36:01	191 document(s)	(195 ms) query '(forensic ink mass spectroscopy)	[1,10]
05:44:57	9619 document(s)	(291 ms) query '(forensic analysis glass)	[1,10]
05:45:18	1112 document(s)	(337 ms) query '(forensic glass mass spectroscopy)	[1,10]
05:45:28	2743 document(s)	(247 ms) query '(forensic glass mass spectrometry)	[1,10]
07:15:44	33949 document(s)	(403 ms) query '(forensicanalysis of paper)	[1,10]
07:47:25	15 document(s)	(916 ms) query '(Forensic Document Examiners? Skill in Distinguishing Between Natural and Disguised Handwriting Behaviors)	[1,10]

Figure 3: Two snippets of a Primo search log from May 30, 2013, showing the interaction of two searchers with the discovery system. The numbers in square brackets indicate which result is at the top of the screen (the first result in most of the examples) and the number of results on the page (10 in these examples). Note that in one example, the eleventh result is at the top of the page—the second page of results).

The interactive nature of the information-seeking process has also inspired Belkin et al. (1995), who describe this process as multiple interactions with an information system. They argue that “people’s conceptions of their information problems change through their interactions with the IR [information retrieval] system” (Belkin et al. 1995) and that each type of information need requires a different kind of interaction with the information system.

Belkin et al. propose an information-seeking behavior model that is based on “a multidimensional space of information-seeking strategies (ISSs)” (Belkin et al. 1995). Each such ISS is derived from a certain context, which consists of the person’s information-seeking goals and the knowledge that the person possesses before starting the process (e.g., the specification of the information needed). The ISSs establish interactions between the person and an information system, and several such interactions form an “episode” (Belkin et al. 1995). Through the information-seeking episode, the searcher’s knowledge and goals develop, changing the specific values of the ISSs. Four modes, or dimensions, of interaction are proposed: method of interaction, goal of interaction, mode of retrieval, and resource considered (Figure 4).

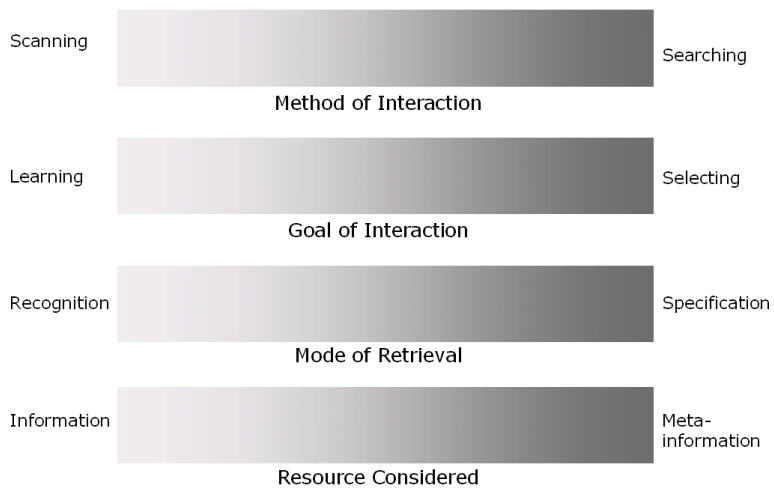


Figure 4: Belkin et al.'s modes of interaction. Modified from Belkin et al. (1995).

“According to our conceptualization,” note Belkin et al., “information-seeking behavior is characterized by movement from one strategy to another within the course of a single information-seeking episode, as a person’s problematic situation changes’ (Belkin et al. 1995).

A discovery system clearly supports most of the user interactions described through this model, and every type of interaction can be mapped along the four dimensions. Figure 5 and Figure 6 illustrate two of the most common interactions with a discovery system.

Figure 5 shows a search for a known item: the mode of interaction is searching, because the user is interested in a specific item and is not likely to scan long lists of results; hence, the mode of retrieval is specification rather than post-search recognition. The user is interested in obtaining an item, not just knowing that it exists, so the goal of interaction is selecting rather than learning, and the resource considered is the item itself, be it physical or electronic, rather than the metadata describing it.

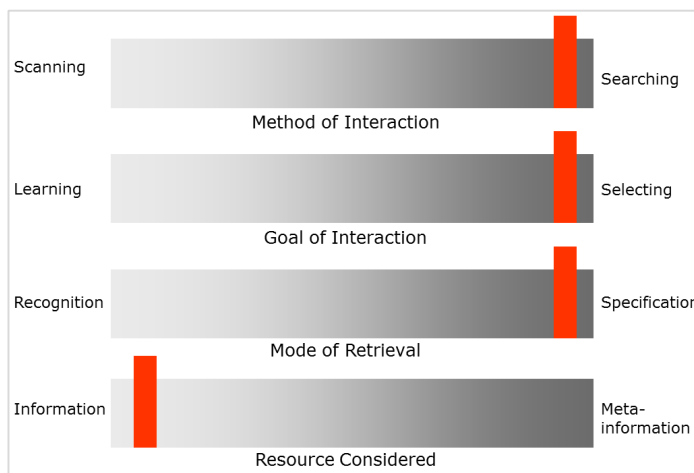


Figure 5: Search for a known item, mapped to Belkin et al.'s modes of interaction (1995)

Another example is a search for literature about a specific topic (Figure 6). In this case, the user is likely to start with a loose description of an information need and is therefore likely to receive a large result set. In the range between scanning and searching, the mode of interaction is around the middle. The goal of interaction is primarily learning, although the user may decide to select items. The mode of retrieval is mostly recognition, because the user is not referring to a specific item; and the resource considered is likely to be only the metadata (assuming it includes the abstract).

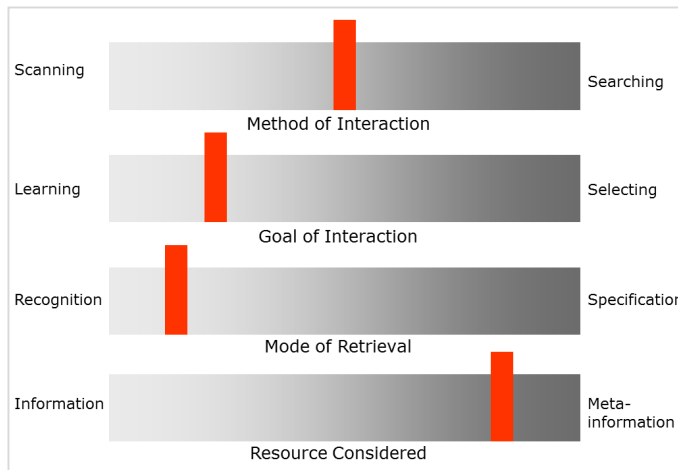


Figure 6: A literature search, mapped to Belkin et al.'s modes of interaction (1995)

In 2002, Bates introduced another model of information seeking, presented as a matrix that embeds information searching in the broader context of information seeking (Figure 7).

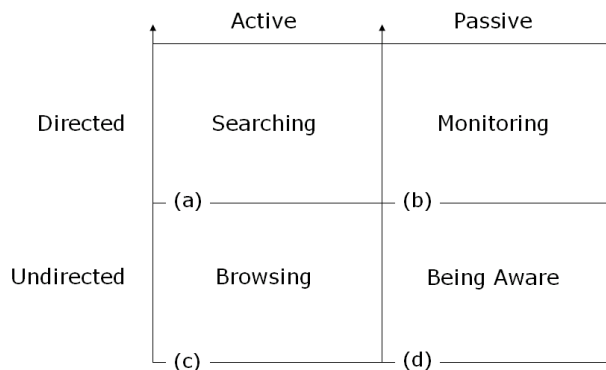


Figure 7: Modes of information seeking proposed by Bates. Modified from Bates (2002), Figure 3.

In this model, the first row (“directed” modes) refers to modes in which a person can specify, to some extent, her information needs, while the second row (“undirected” modes) refers to modes in which the person encounters information in a random manner. In one column, a person is actively looking for information, whereas in the other column, a person is just absorbing information in a passive way. According to this model, an academic searcher demonstrates all four combinations of modes: searching in an information system would be considered directed active behavior; browsing through enlisted newsletters, announcements, or personalized feeds of new submissions would be undirected active behavior; attending a conference or a workshop in one’s area of research would be directed passive behavior; and, finally, learning about new publications through haphazard conversations would be considered undirected passive information-seeking behavior.

A search system accommodates only the directed active mode. A discovery system opens up new options; for example, by suggesting articles related to one that a searcher has selected, the system brings to the attention of the searcher articles that do not necessarily share textual cues with the selected article. In this way, the system supports the undirected active mode of information seeking. Alerts defined by the user also fall into this realm. Passive information seeking, on the other hand, remains out of the scope of a discovery system.

The last model presented for this discussion represents the information-searching behavior of academic users (Figure 8) (Sadeh 2010).

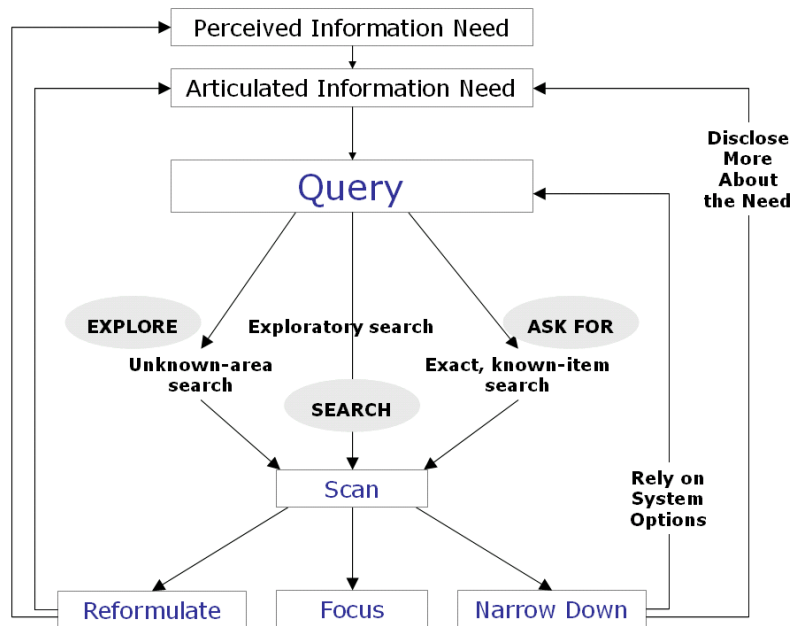


Figure 8: A model of information-seeking behavior of an academic user (Sadeh 2010)

Information seeking as an active, directed interaction starts with a *perceived information need*. The searcher has an idea of the material required and converts the perceived information need to an *articulated information need*, or query, in order to start the interaction with the information system. A query typically represents the searcher’s hypothesis about the specific keywords that are found in the metadata or the full text of the required documents. The distinction between a perceived information need and an articulated information need is crucial to the information-searching process. A failure of the system to return the results that a searcher is looking for can be attributed to the nature of the need itself (if, for example, the information to satisfy the need does not exist), the user’s failure to describe the need in a suitable way, or other factors.

The model addresses three kinds of query actions: Explore, Search, and Ask For. The mode depends on the information need and the way it is expressed. Queries vary from being very precise—when searchers describe a specific item—to very vague—typically when a searcher is not an expert in the field and thus cannot define the information need clearly. For example, when a searcher is looking for a specific article, she is likely to enter information such as the article’s title or the names of the authors. In this case, the user’s mode of searching is Ask For, as opposed to Search or Explore, even though the technical process is similar for the three modes. In the Ask For mode of searching, if the document is well defined by the user,

the result of the search is likely to be short, and a good relevance-ranking algorithm would position the article as the first item on the list.

The Explore mode can be exemplified by a novice user's search for information on a specific topic or a more experienced user's search for information in a field outside her area of expertise.<sup>11</sup> In these cases, the searcher is likely to be unfamiliar with the most appropriate search terms or with the kind of materials that the query results might offer. Therefore, the initial result list may not satisfy the user's information need.

Most queries, however, are somewhere between these two extremes: searchers are familiar enough with their field of interest to clearly define their information need yet do not know of a specific item that would provide the desired information. The variations of the possible queries in such cases relate to the amount of information that the searchers know about the topic or decide to provide when they articulate their perceived information need. The more information they know and provide, the closer the query comes to asking for a specific item. This mode of searching, which corresponds to what we intuitively perceive as "searching," can be described as Search mode. The balance between formulating a query with too little of the hypothesized information—which can lead to too many results—and including too much of the hypothesized information—which can eliminate relevant results—is of great concern to most searchers and triggers a trial-and-error mode of articulating the information need.

Once a query is submitted, the system displays a result list to the searcher. Even before scanning the results, the searcher obtains valuable information that the system has provided: the number of items in the list and the system's suggestions that relate to the query (such as Did You Mean...?) or relate to the result list. The latter might include, for example, post-search groupings and suggestions for new searches. All this information enables the searcher to understand right away whether the result list is worth exploring in more depth. For example, a lack of results may signify one of several possibilities: that the information does not exist, that the query contains wrong terms or even an error, or that the searcher has to rethink the articulated information need. Too many results may signify that the query was too broad. A system's Did You Mean...? suggestion may draw the searcher's attention to a misspelled name or a variation in a term. Topic lists and other information (date ranges, authors, , journal titles, and more) that serve as post-search groupings provide a brief summary of the result list; by looking at the terms displayed in these groups, the searcher can see the major characteristics that the items on the list have in common.

Searchers typically scan the first items in a result list before taking an action. Even a brief look at these items usually provides enough clues as to whether the searcher is on the right track, especially when the list is sorted by relevance. Often the first result, or one of the first results, is the requested item—especially when the query mode is Ask For. If none of the first items seems relevant, searchers typically reevaluate their query.

After analyzing the first screen, a searcher chooses one of the following options:

- Focus: If the result list is satisfactory and there are results that seem relevant, at least at first glance, the searcher may focus on a specific item.

---

<sup>11</sup> According to a survey by University of Minnesota Libraries, over 70% of undergraduate searches are exploratory; as searchers become more knowledgeable in their area, they tend to search more for known items. Discoverability: Phase 2 Final Report (2010), <http://conservancy.umn.edu/bitstream/99734/3/DiscoverabilityPhase2ReportFull.pdf>.

- Narrow down: If there are too many results, the searcher may choose to narrow down the list so that it shows only the items that are more relevant.
- Reformulate: If there are no results or the results do not seem relevant, the searcher may decide to reformulate the query.

Narrowing down can be carried out in two ways: either the searcher takes advantage of the system's options—perhaps by clicking a facet (such as a specific date range, topic, or journal name) or by choosing to see only materials available online—or the searcher decides to modify the articulated information need by providing more information.

However, if the searcher decides that the query was not adequately phrased, she needs to reformulate it. Reformulation can be minor or can be suggested by the system—when a name or term has been misspelled, for instance. However, in some cases, the searcher might need to consider the information need again, modifying the perceived information need and restarting the process.

The flow described by this model is iterative and may well span a long period of time. Sometimes users, particularly undergraduates, have a distinct, focused information need, and once the system satisfies the need, the process is terminated. However, in many other cases, especially with researchers, the interaction with the discovery system is ongoing.

This model of information searching covers only part of a user's information-seeking behavior. Other elements, such as undirected information seeking, are beyond the scope of this paper. However, navigation from one item to another, another aspect of information seeking, deserves special attention. Navigation is a very common practice of information seeking, regardless of the type of information that a user is looking for. In the scholarly domain, researchers carry out such navigation by following a citation trail: by moving backward and forward through references, a scholar can explore the origins of an idea and trace its impact on the academic discourse. As part of an information-seeking process, references may expand the knowledge of a searcher into adjacent areas that are only briefly discussed in the document that the scholar has found. Often, it is enough for a scholar to search only once and continue the information-seeking process through such navigation.

This theoretical model is well supported by discovery systems. Although conceptually the various modes of searching—Explore, Search, and Ask For—differ from one another, a working system supports them all through the same user-interface functions. The type of search determines the kind of options that the system makes available and that are useful in a specific context. For example, the display of facets is relevant only when the searcher is looking for materials on a certain topic and there are too many results for the searcher to scan.

## **What else can discovery offer?**

Like every new technology, library discovery systems started by increasing the efficiency of existing practices. Users were able to satisfy their information need before a discovery system was available; they had to use multiple information resources and were often challenged with complex workflows when trying to find relevant materials, but they would usually succeed in obtaining the required items. From the outset, discovery systems enabled searchers to use only one information resource through a user-friendly interface and streamlined the searchers' workflows from the initial search to the obtaining of the item. However, once this basic framework was set in place, the developers of discovery systems started introducing new capabilities that were not previously feasible.

### **Usage-based capabilities**

Following common practices in other systems, such as web search engines and retail websites, library discovery systems accumulate usage data. Information such as user queries, workflows, and selections is logged; however, unlike systems in other domains, library discovery systems do not associate usage information with individuals. Such information, which is now available in the library realm for the first time at such a scale and across such a large range of information providers and institutions, enables discovery systems to provide a variety of new capabilities.

### **System improvements**

There is no doubt that usage data may help developers improve their system by shedding light on the ways in which people use the system. Information such as the number of searches over time, with an indication of peak hours; the frequency at which various system functions are used; the success rate of search interactions; the time it takes a searcher to satisfy an information need; the entry point to the search functionality; and the physical device used by the searcher are all important clues. The success of relevance-ranking technology, to name a specific example, can be measured and the algorithm further tuned through the monitoring of key performance indicators that relate to ranking (such as the average position of the selected record in the result list) (Sadeh 2013).

### **Recommender systems**

The first usage-based service introduced in a discovery system was the Ex Libris bX article recommender. Derived from research by Johan Bollen and Herbert van de Sompel (2006) at Los Alamos National Laboratories, this service presents the searcher with a list of articles that are related to a given article. This relationship is determined by an analysis of selection patterns of the worldwide research community; if enough scholars select the given article together with another article during a search session, the two articles are considered related. Such recommendations enable a searcher to become aware of articles that are likely to be very relevant to the searcher's information need yet would not appear in the result list because they do not contain the search terms stated in the query. The presentation of related articles provides an element of serendipity that might greatly improve the searcher's chance of reaching materials that she had not thought of.

### **Trend analyses and popularity reports**

For end users, usage data may introduce another aspect of serendipity: rather than seeing just the items that they were looking for, searchers can find out about items that are currently the object of other scholars' attention<sup>12</sup> and can follow trends, such as the rate of adoption of open-access materials. Analyses of usage data can also help librarians gain insights about the ways in which library collections are used.

### **Support of collection development**

Usage data gathered by a discovery system and fed into a library management system greatly improves the collection development process. Furthermore, in an integrated environment where usage is analyzed and presented together with a cost analysis, librarians can more effectively make decisions regarding the materials that are required, the sources for obtaining these materials, and the most appropriate acquisition models. New trends such as patron-

---

<sup>12</sup> See the Hot Articles service provided by Ex Libris:  
<http://www.exlibrisgroup.com/category/bXHotArticles>.

driven acquisition are also supported by discovery systems that offer searchers an information landscape that is greater than that available to the library, while the actual acquisition may take place only upon demand.

### **Alternative (or complementary) assessment methods**

Traditionally, the number of citations of an academic work served as the basis for an assessment of the work, either directly or through more complex metrics, such as the journal's impact factor (which takes into account the overall citation patterns of all articles published in a specific journal). In today's environment, where the consumption of new publications is immediate and not all publications appear in journals that are assigned an impact factor score, a complementary, usage-based method of assessing the significance of scholarly publications is needed. Usage information may originate from the discovery system itself or may be accumulated from a variety of sources, as is exemplified in the altmetrics manifesto<sup>13</sup>. Because discovery systems are typically open and flexible, they can easily display information that originates from other sources—for example, altmetric.com<sup>14</sup>—thus seamlessly providing the user with better tools to determine whether and to what extent they can rely on a specific publication.

### **Personalization**

Another feature that has begun appearing in library discovery systems is their ability to tailor the search results to a specific searcher's needs. Although common in non-academic information systems, academic systems have kept away from making any judgment based on the searcher's attributes. However, with the ever-increasing number of publications that are available to users and the dismantling of subject-specific silos, a discovery system that takes into account a user's discipline and academic level is able to address the user's information need more accurately, especially when the user conducts an exploratory search.

### **Linking**

The availability of a very large corpus of data in a single system is one of the greatest advantages of a discovery system. The implementation of smart linking between items leverages this data and helps the system provide the searcher with usable information rather than lists of items. For example, the clustering of various aspects of a scientific work, including the research data, project reports, and published articles, is likely to provide searchers with a rich user experience and help them grasp the information in a more complete way. Furthermore, because institutional repositories are also harvested by discovery systems, the linking of various versions of a work, such as the preprint and post-print versions, becomes feasible. Another example relates to other entities, such as authors and institutions, which can be rendered into hubs of information, enabling searchers to discover other individuals and teams conducting research in their area and other departments that are dealing with similar projects.

## **Conclusions**

The rapid pace of adoption of discovery systems demonstrates the appeal of such systems to institutions and searchers. Discovery systems have transformed the academic search process, making it more intuitive and interactive and integrating it with other user activities. To support the broad, heterogeneous information landscape that is available today, the current

---

<sup>13</sup> <http://altmetrics.org/manifesto/>

<sup>14</sup> <http://www.altmetric.com>



discovery process puts more emphasis on the post-search activities and better supports serendipitous finding, leading users to a richer and more diverse learning experience.

The promise of discovery continues to evolve with the introduction of new services that could not have been implemented in the pre-discovery era. In the future, we are likely to see other innovations, such as the expansion of semantic search capabilities and the generation of tools that address additional aspects of the scholarly publishing and dissemination process.

## **Bibliography**

Bates, M.J. (1989) The design of browsing and berrypicking techniques for the online search interface, *Online Review*, 13(5), pp. 407-424. Version used for this study: <http://gseis.ucla.edu/faculty/bates/berrypicking.html> [last accessed: 3 December 2010].

Bates, M.J. (2002) Toward an integrated model of information seeking and searching. Paper presented at the 4<sup>th</sup> International Conference on Information Needs, Seeking and Use in Different Contexts, Lisbon, Portugal, September 11, 2002. *New Review of Information Behaviour Research*, 3, pp. 1-15.

Belkin, N.J., Cool, C., Stein, A., and Thiel, U. (1995) Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems, *Expert Systems with Applications*, 9(3), pp. 379-395.

Bollen, J., and Van de Sompel, H. An Architecture for the Aggregation and Analysis of Scholarly Usage Data. Paper presented at JCDL '06, Chapel Hill, North Carolina, June 11–15, 2006. ACM 1-59593-354-9/06/0006.

Breeding, M. 2007. Next-Generation Library Catalogs. *Library Technology Reports*, 43(4).

Markey, K. 2007. The Online Library Catalog. Paradise Lost and Paradise Regained? *D-Lib Magazine*, 13(1/2).

Sadeh, T. 2010. A Model of Scientists' Information Seeking and a User-Interface Design. PhD thesis, School of Informatics, City University London.

Sadeh, T. 2011. Discovery and Management of Scholarly Materials: New-generation Library Systems. *ProInflow* [online].

Sadeh, T. 2013. Optimizing Relevance Ranking to Enhance the User's Discovery Experience. Paper given at the IRCDL Conference, Sapienza Università di Roma, February 1, 2013.

Wilson, T.D. (1999) Models in information behaviour research, *Journal of Documentation*, 55(3), pp. 249-270. <http://informationr.net/tdw/publ/papers/1999JDoc.html>.

Wilson, T.D. (2000) Human information behaviour, *Informing Science*, 3(2), pp. 49-55. Special Issue on Information Science Research.