New Generation Discovery: A Systematic View for Its Development, Issues and Future

Yi Yu

King Abdullah University of Science and Technology, Saudi Arabia

Haider Ali

Yanbu Industrial College, Saudi Arabia

Author Note

Yi Yu, Library of King Abdullah University of Science and Technology, Saudi Arabia

Yi Yu is a Systems Librarian of the library of King Abdullah University of Science and Technology in Thuwal, Saudi Arabia. This paper is presented for the international conference: “Change and Challenge: Redefine the Future of Academic Libraries” on November 4-5th, 2012 in Beijing.

Contact: yi.yu@kaust.edu.sa

Haider Ali, Yanbu Industrial College, Saudi Arabia

Haider Ali, is a Lecturer of Yanbu Industrial College in Yanbu, Saudi Arabia. This paper is presented for the international conference: “Change and Challenge: Redefine the Future of Academic Libraries” on November 4-5th, 2012 in Beijing.

Contact: hali@yic.edu.sa

Abstract

Collecting, storing, discovering, and locating are integral parts of the composition of the library. To fully utilize the library and achieve its ultimate value, the construction and production of discovery has always been a central part of the library's practice and identity. That is the reason why the new generation (also called the next-generation discovery) discovery gets such striking effect since it came into library automation arena. However, when we talk about the new generation of discovery in the library domain, we should see it in the entirety of the library as one of its organic parts and consider its progress along with the evolution of the whole library world. We should have a deeper understanding about its relationship and interaction with the internet, the rapidly changing digital environment, and the elements and the chain of library services. To address above issues, this paper overviews the different versions of the definition for the new generation discovery by combining our own understanding. The paper also gives our own description for its properties and characteristics. The paper points out what challenges, which extends the technology domain to commercial interests and business strategy, are faced by the discovery applications, and how library and library professionals deal with those challenges. Finally, the paper elaborates on the promise brought by the new discovery development and what the next exploration might be for its future.

Keywords: discovery, new generation discovery, the next-generation discovery, discovery tools, discovery interface, discovery services, discovery environment, resource discovery, federated searching, web-scale discovery, OpenURL.

Introduction

Collecting, storing, discovering, and locating are integral parts of the composition of the library since it was conceived, no matter how much change and evolution has occurred to the library's style, its content, its format, its forms and its usage. To fully utilize the library's function and achieve its ultimate value, the construction and production of discovery - connecting users with the required and relevant information in a convenient manner - has always been a central part of the library's practice and identity. That is why it has such 'wow' factor and caused such excitement when so called new generation discovery tools come into the spotlight of library automation arena. The new discovery tools/services make big changes of the information retrieving and locating, and also causes great impact for information users and services suppliers. It gets praises but also brings disputes. This article will give a systematic review about what the new generation discovery means for all the related parts, what its current status and issues, how library professionals response the challenges in the new discovery environment, and what the information discovery future might be.

What Is The "New generation Discovery"?

In reality, there is no clear and uniformed definition regarding "new generation discovery". Luther and Kelly (2011) described it as "a Google-style approach for building and searching a unified index of available resources, instead of searching each database individually". According to Yang and Wagner (2010), it is a tool to "provide search and discovery functionality and may include features such as relevance ranking, spell checking, tagging, enhanced content, search facets". Yet there is a more direct and understandable term called "web-scale discovery" for it, even though the annotation of "web-scale discovery" may not cover all of the aspects of the "new generation discovery" (Matei, 2012). As it is really hard and might be misleading to attempt to give a definition in a single sentence without adequate and explicit description, a brief summery is given here from technology point of view: a next-generation of discovery can be delineated as web-scale discovery which is based on central indexing approaching integrated with federated search which based on important search and retrieval protocols and web 2.0 tools which applied to traditional online catalog for giving more intuitive interfaces and search and retrieval functionalities.

Dynamic and Unlimited Exploration Process

Then, what does “new generation discovery” really mean? What properties does it have? Is it a “next-generation catalog”, as most called it? Many articles seem to regard “next-generation discovery” and “next-generation catalog” as the same thing. However, next-generation discovery is much more than a next-generation catalog. The aim of the discovery is not only confined to the collection managed by the traditional library catalog, but also involves and expands to include non-print materials from heterogeneous resources in all formats. The discovery scope has been widened from including only the resources available in the library to also including the resources available on web. In this situation, the meaning of “catalog” - used to describe conventional flat environments - is not enough to present the depth and breadth of a multidimensional network environment.

Let us have a literature review of “catalog” and “discovery”. The definition of “catalog” in the Merriam-Webster.com dictionary is: 1) list, register; 2) a complete enumeration of items arranged systematically with descriptive details. Reitz (2012) defines catalog “A comprehensive list of the books, periodicals, maps, and other materials in a given collection, arranged in systematic order to facilitate retrieval (usually alphabetically by author, title, and/or subject).” The definition of “discovery” according to Merriam-Webster.com dictionary is: the act or process of discovering; display; exploration; something discovered. A “resource discovery” system implies the discovery of resources that might be unknown or new to the user (Nagy, 2009). A catalog is static, limited, fixed, and complete. It indicates a clear expectation. It is result oriented. It is focus on the ending. However, “discovery” is the act or process of discovering. It is dynamic, unlimited, transformed and expended. It is process oriented. It is focus on the whole effect.

View Discovery as a Process of the Whole System

Breeding (2010) explains that most often, this new type of discovery is being preferably called “discovery interface”. Yes, the term “interface” does indicate that the discovery, as an independent activity, has been liberated from the confines of library automation systems and separated from resource management. However, if we look back at the path of the development of discovery in library and academic environment, we can see that the development of discovery has not only stayed in the interface level. During the recent two decades, many technologies, protocols and standards have been continuously developed, promoted and adapted as part of

discovery applications, such as federated search solutions, link resolvers and the OpenURL, etc. those developments have been offering a variety of new discovery options.

More recently, “discovery services” became a prevailed term to represent the next-generation discovery. This might be the most appropriate term to represent it. As software services are a dominant software delivery model in today’s computing market, all of the features such as relevance ranking, spell checking, tagging, enhanced content, search facets, as well as those discovery process like multi-query, pre-indexing can be said to be software services. From the above discussion, we can view discovery as a process of the whole system, and not as an isolated entity. The major characteristics of this new type of discovery in library automation and computerization history can be epitomized as follows:

- **Discover further.** By the capacity of federated searches (based on a distributed query model), it breaks through the fetter of space in terms of the location, either physical or digital. The discovery territory is expanded from isolated local area to networked resources.
- **Discover deeper.** By the power of deep-indexing (based on pre-harvested content), it melts the segments of the bibliographic levels. The discovery result is unlimited within the traditional OPAC, it is beyond MARC record and expands to many metadata types for access the universe of content in the web.
- **Discover more.** By combining versatile new technology and an innovative approach, the discovery scope broadens in all directions and in all formats including library catalog records, e-journal articles, databases, newspaper articles, e-books, dissertations, institutional repositories, conference proceedings, grey literature, cited references, reports, etc.
- **Discover easier.** By realizing “one-step” Google-like searching plus incorporation with web 2.0 features, the intuitive interface and simplified navigation with rich functionalities – such as relevance ranking, faceted browsing, spell checking, tag clouds, search recommendations, and social networking – help make the discovery experience a more pleasant journey instead of a daunting process.

Current Issues

Accompanying the emergence and growth of this new realm of discovery, there have been much applause and cheer, but also not without doubts and disputes. The various concerns

are related to the major parties involved who play significant roles during the discovery process. They are resource publishers, service providers and users. Resource publishers can be a big information agent, an individual database, or e-journal vendors of a special subject or discipline. Most of them are commercial publishers and aggregators, but there are also non-commercial publishers or associated publishing organizations. Service providers can be software vendors who supply the discovery tools and features, libraries who provide access and retrieval of resources to the community of users, or infrastructure and network suppliers who support a resource discovery environment. Users can be specialized researchers of their research topic, general college students in their academic study, librarians with special needs for their reference services, or any casual user with their diverse information needs. The various voices of dissent among them can be generalized as the following three aspects:

Central Indexing Approach

The latest and most exciting achievement among the new generation discovery applications is web scale central indexing. The idea for this development is to overcome the drawbacks of federated search - slow speed and low relevance - by indexing the full corpus of information globally. However, two factors hinder to fully realize this approach.

The first factor is that there is no web scale discovery product which can exhaust all of the metadata from all of the publishers in the world. Without a fully comprehensive coverage, it cannot be called a “web scale” discovery. Needless to say, some discovery service producers are big resource aggregators for themselves. The method of central indexed coverage is big enough for some libraries. However, for libraries with large numbers of electronic subscriptions or for specialized libraries with specialized subject subscriptions, there is a need to combine the central indexing approach with federated search engines in order to cover those contents not in the scope of the central index. In this situation “many are familiar with the limitations of federated search technologies: slow speed, poor relevancy, ranking of results, and the need to configure and maintain sources and targets such problems remain with federated search products integrated with Web scale discovery services” (Jason, 2011). Thus, what is the advantage for the web scale center-indexed search engine?

The other factor is the publishers and database aggregators. They may not be willing to release their authority over the content to third party discovery service providers, especially if these service providers are also content providers themselves. A typical example of this situation

is when EBSCO removed its content from Primo's main database (Jastram, 2011). This may be because "some publishers and aggregators prefer to see their content made available through interfaces under their own control rather than through a third party discovery service where they have no particular way to influence the ranking or positioning of their content in search results." In particular, if this content provider also has the intent to develop their own new generation discovery tool, it will definitely not allow its competitors to have control of its resources. There is a prevailing concern regarding the competition within information marketing. The worry is that discovery service vendors might take advantage of the situation to further their own interests by promoting their own databases, and the publishers "are concerned about a loss of control in the way that library users experience their resources and that their resources may suffer variations in usage levels, which in turn will have an impact on whether or not libraries will maintain their subscriptions." (Breeding, 2012)

Thus we can say, the success of the central indexing approach depends on if it is able to get metadata from all of the content providers – if this goal is too farfetched, then at least from most of the them.

Relevancy Ranking

In the new discovery environment, the discovered scope has been widened from resources available within the library to the resources available throughout the internet. A "resource discovery" system implies that the users may not have a clear target, and that the resources discovered might be new or unknown to the user (Pradhan, Trivedi, & Arora, 2011). Compared to the world before the internet, we now have an abundance of information resources. However, users now have much less attention span for each resource and may not be willing to spend a long time on resource discovery, and are intimidated by the huge discovery opportunities.

Assessing the trade-off between the values of information gained versus the cost of performing the finding activity (Pirulli & Card, 1995) is a big consideration for the users. Thus, returning search results with the most relevant items on top of the list becomes a key feature of the new discovery tool. Although this function has been viewed with approval in many discovery applications, some users still feel that the relevancy ranking seemed a bit of odd/weak. Since the commercial factor still plays a significant role, (e.g. some search engines have "sponsored" links, some discovery services promotes their own databases above any others to an absurd level), how to win the sophisticated searchers' confidence "that they will be presented with the best

representation of the resources available from the library for their research” (Breeding, 2012) is still being tested.

Google-Like Search

One of the main features of the new generation discovery tool is to supply “Google-like” search, since the information search behavior and expectation of the millennium generation (Taylor, 2012) has been influenced so strongly by this unscientific search method. According to university investigation, this approach is popular amongst their students. However, “the tendency [of the faculty] is to go to databases directly” (Vaughan, 2012). The worry of “becoming accustomed to the convenience of a simple, Google-like search experience, patrons might neglect to search specialized databases that would give them better results” is a typical concern from academic librarians (Patrick Carr, 2011). The following comment represents the general opinion from a subject librarians’ point of view: “The idea of providing one-search for users in medicine — mostly physicians and medical students — is very difficult for me to justify (and teach). Further, it’s not appropriate in most search instances” (Giustuni, 2011). There is also a major concern regarding using one search across millions of records because it might be a potential weakness for users, since they may become overloaded by information. While it is fair to say that Google-like search is best used as a method of general browsing for general users, the comment that “many librarians and specialized users may even see [Google-like search] as a step backward” (Breeding, 2012) is not too harsh for this modern discovery tool.

What Role the Library/Librarian Should Play

When we review the whole history of discovery development, we can find that the dominant force is coming from the enterprising industry. When issues emerge, we often hear the lamentation from librarians: what is our position? They express strong doubt about “competition in the free market is the force looking out for library interests” (Jastram, 2011). However, there is still a lot of room for librarians to be part of the development, to contribute professional knowledge for this challenging and aspiring area.

Be A High Level Player of the Game

In the commercial world, it is inescapable for enterprise organizations to view the pursuit of profit as their main purpose. Some problems like the comprehensive central index “isn’t technical; it’s a matter of business decisions and strategies” (Breeding, 2012). However, it is the libraries’ responsibility to take the initiative to be involved in the development of the products

which should be based on the library's purpose and usage. We should use our influence to make the decision makers of the enterprise organizations to realize that sometimes sharing may be the win-win solution. Like the issues in central indexing, if the publishers see that their usage statistics is increasing due to using the central indexing discovery service, they might be willing to open their metadata to discovery service vendors. If the discovery service vendors treat all of the resources equally and do not give special privileges to their own databases, more publishers will be willing to join them, which then will result in their ability to supply better service and win more users.

Libraries are also able to give specific and direct help like metadata cooperation. Some thoughtful librarian once pointed out that "we probably don't need to create a cooperative metadata creation initiative for article-level metadata, because that metadata ... is already out there in the digital world. It's already been created; pretty much every publisher these days has electronic metadata for their articles published. We just need to collect it" (Rochkind, 2011). Certainly, how to collect this metadata is not a trivial job. Similar to many other technological developments, the products developed in the initial stage may only be for specialized purposes and resolving specific problems. As it grows, it may then experience the bottleneck effect as it expands to accommodate more users for more purposes. It will need protocols and standards as it develops into a high level and becomes more intricate, and will need broader analyses in order to have a better solution. Libraries have played a crucial role for removing the barriers between publishers, service vendors and users. For example, in order to make a comprehensive central index possible, there needs to be an organization that has no personal interests and no biased opinions to present the interests of the vast numbers of users. Thus, the Open Discovery Initiative, a workgroup recently launched by the National Information Standards Organization (NISO) came to be. It "aims at defining standards and/or best practices for the new generation of library discovery services that are based on indexed search ("Open Discovery Initiative," 2012).

Practice Librarianship in the New Environment

As the scope of new generation discovery expands to the web and Millennium users discards the traditional search method, there has been the question of whether the library is dead, and doubts about whether the library profession is still necessary. These are comprehensive questions that may not get a chance to be further discussed in this paper. However, we can get a basic impression about those questions when we have a better understanding of new generation

discovery. For example, usage of subject headings may influence the categories of facets in the new generation discovery tools. Consequently, the label of facets will affect the user's ability to locate the resource even when it is already discovered. As metadata creation and transformation is not a big issue in the digital world, how to word it exerts vocabulary control over the metadata that would directly affect the search results. The quality of the metadata would also affect the relevance ranking (Jastram, 2011). In central indexing discovery tools, the meshed index data comes from different regimes. For example, personal names in the biographic records of print copy usually have authority control practices, but it may not have any authority control in e-resources. In scholarly articles, names tend to be formed with the surname and initials, but in the Library of Congress authority file names are represented in more complete forms. In this situation, what does it mean when different types of data are mixed together? Thus, the new generation discovery has to have the capability to ensure the consistency of metadata as authority control.

Librarians not only can directly contribute their professional knowledge for the development of new discovery, they also can give suggestions and advice for the implementation and configuration of new discovery tools. In the example of relevance ranking, librarians can use their judgment and local needs to decide how to arrange the algorithms such as currency, full-text or subject headings. As no one system can be perfect for everyone, librarians should use their experience to their advantage and mediate the gap between the discovery tools and the users. For example, Google-like search is a controversial topic where librarians can collect different users' feedback and deliver them to the service vendor to aid in the improvement of their service (e.g. by adding advanced search options). Librarians can also use their judgment to make practical decisions based on the information and feedback they receive. If it is clearly indicated that "the web-scale discovery service is not the 'beginning and ending' for discovery" (Vaughan, 2012), it would not be wise to remove the classic online catalogue as a compliment to the web-scale discovery. If the discovery tool developer has indicated that "they were targeting their discovery service at undergraduate research needs" (Vaughan, 2012), it would not be prudent to insist on the faculty in your campus to use Google-like search for their research work. The success of the Google-style search engine is a significant motivation for other vendors to develop similar discovery tools in order to compete. However, Google's success is due to the fact that Google has an implicit understanding of its target audience. It allows for gathering as much

information as possible in as little time as possible with a simple word, which is perfectly suited for generating more users for more ad income in a “long tail” business strategy. While this strategy has helped Google become successful at generating revenue, it may be not suitable for non-profit institutions with specialized service audience (e.g. graduate students and faculty) and limited resources.

Be a Visionary of Discovery Future

All new things come from something already in existence. The visionary of the future is based on the good understanding of the present.

Customization & Personalization Currently, there are only a few discovery service vendors with similar functionalities in their discovery products on the market. However, there are thousands of clients as well as potential clients with different end users and different needs. How to satisfy the many various users' need is a big challenge for these existing discovery services. It is important for the services to have the capability of customization and personalization. Generally speaking, the services should allow their clients to realize their own influence on the discovery services as much as possible for the sake of their end users' needs.

There are already some personalized approaches in new discovery applications such as building personal e-shelves (the concept is the same as OPAC my book shelf). However, the new generation discovery should be able to do much more in terms of personalization. Using the users' personal information to supply personalized service is commonly used in marketing; it is being crucially applied in the academic area currently. If it is used well, discovery services will be able to go beyond what it had accomplished. For example, the application can simplify the user experience and tailor recommendations according to the users' information.

Discovery Workbench In the previous discussion, catalog, interface and service were summarized as the properties of new generation discovery. However, we think those properties are not enough to describe the extent of the capabilities of new generation discovery. “Discovery environment” should be introduced as a broader and deeper understanding about discovery. This means that the discovery is not just a registered record, a tool, or simply a resultant list from a search. It should be a part of learning, research, and the work of the users. It should be built around the user workflow. Carl Grant (2012) explained that it should be used as a workbench, termed “discovery workbench”, which be seamlessly integrated with all available tools in order to

benefit the user based on the user's needs. Just like air and water in the nature is essential for our lives, the services of discovery tools should be ubiquitous for our intellect.

Intelligent Discovery Besides worked as a workbench, discovery tools should be smart and understanding. Discovery services should be able to define the problem, to search, filter, evaluate, and interpret their findings. The display of discovery should be able to graphically represent the patterns and relationships that are visually detectable. The results of discovery should be corrected and replicated. The discovery in the digital world should preserve and strengthen the virtue of serendipity which human beings have been familiar with during the centuries of reading print materials. Ideally, the discovery application should be smart enough to read user's minds and know what he or she is searching for. If the comment that the "vast majority of researchers turn to web search engines to meet their information needs" (Law, 2010) is true currently, we believe that this situation will no longer be the case once discovery services realize the above discussed expectations. Certainly, this will not happen suddenly or in one day. It will be a gradual process and will need time. However, we hope it will not take too long.

Conclusion

One of the prime responsibilities of the library professionals is to assist the researchers in finding, guiding, utilizing library resources without demanding that they acquire specialized knowledge. It is presumed that the new generation discovery development has the capacity to connect library users with the vast information repositories and add value to library collections. Although the discovery development are in their extreme infancy, it is fascinating and aspiring to watch its maturity as it has tremendous potential to serve users' needs. We think, in order to assist our users to grasp this modern discovery tools and supply efficient information services in the new discovery environment, it is necessary to understand this new generation discovery comprehensively and profoundly. We conceive the best way to realize the power of the new generation discovery is to follow the ender users' needs and requirements and proceeded from the real world. We believe librarian still can play a crucial role in the new journey of the information services. We held high promise for the future of discovery service as it has plenty room for development and progress.

Reference

Breeding, M. (2010). The State of the Art in Library Discovery 2010. *Computers in Libraries*, 30(February), 31-35.

Breeding, M. (2012). Looking Forward to the Next Generation of Discovery Services. *Computers in Libraries*, 32(2), 28-31.

Catalog. 2012. *In Merriam-Webster.com*.

Retrieved July 15, 2012, from <http://www.merriam-webster.com/dictionary/catalog>

Discovery. 2012. *In Merriam-Webster.com*.

Retrieved July 15, 2012, from <http://www.merriam-webster.com/dictionary/discovery>

Environment. 2012. *In Dictionay.com*.

Retrieved July 15, 2012, from <http://www.dicitionary.com>

Giustini, D. (2011). Retrieved from Is Summon the search solution we want it to be?

<http://blogs.ubc.ca/dean/2011/03/is-summon-the-search-solution-we-want-it-to-be/>

Jason, V. (2011). Investigations into library web scale discovery services. [Preprint]. *Information Technology and Libraries* 61.

Jastram, I. (2011). Heads they win, tails we lose: Discovery tools will never deliver on their promise. Retrieved from <http://pegasuslibrarian.com/2011/01/heads-they-win-tales-we-lose-discovery-tools-will-never-deliver-on-their-promise.html>

Rochkind, J. (2011). More on aggregating article metadata. Retrieved from <http://bibwild.wordpress.com/2011/01/22/more-on-aggregating-article-metadata/>

Law, J. (2010). The Long Tail in Research Collections: Moment of Need Meets Discovery Speed
Retrieved August, 2012

Luther, J., & Kelly, M. C. (2011). The next generation of discovery: The stage is set for a simpler search for users, but choosing a product is much more complex. *Library Journal*, 136(5), 66-72.

Matei, D. (Mar 13, 2012). Next-Generation Library Catalogs: a resource guide. Retrieved July 12, 2012, from <http://infoguides.fuller.edu/nextgencatalogs>

Nagy, A. (2009). *A Conversation With Karen Schneider: VuFind and Summon*. Paper presented at the WiLSWorld 2009 Conference, Madison Pyle Center.

Open Discovery Initiative. (2012). Retrieved July 12, 2012, from <http://www.niso.org/workrooms/odi/>

Pirolli, P., & Card, S. (1995). *Information Foraging in Information Access Environments*. Paper presented at the CHI '95. http://www.sigchi.org/chi95/proceedings/papers/ppp_bdy.htm

Pradhan, D. R., Trivedi, K., & Arora, J. (2011). *Searching Online Resources in New Discovery Environment: A State-of-the-Art Review*. Paper presented at the CALIBER, Goa.

Reitz, J. M. (2012). Online Dictionary for Library and Information Science, from http://www.abc-clio.com/ODLIS/odlis_c.aspx

Taylor, A. (2012). A study of the information search behaviour of the millennial generation. *Information Research*, 17(1). Retrieved from <http://informationr.net/ir/17-1/paper508.html>

Vaughan, J. (2012). Investigations into Library Web-Scale Discovery Services. *Information Technology and Libraries*, 31(1), 32-82. Retrieved from <http://ejournals.bc.edu/ojs/index.php/ital/article/download/1916/1744>

Yang, S. Q., & Wagner, K. (2010). Evaluating and comparing discovery tools: How close are we towards next generation catalog? *Library Hi Tech*, 28(4), 690-709.