

Insights into One-Stop Searching in Discovery Tools

Sharon Q Yang
Rider University, USA
yangs@rider.edu

Abstract. *One stop searching is the most essential requirement for a discovery tool. However, this vision may never be fulfilled due to political and economic complications. The article discusses the difficulties arising from deploying a unified index, the solution to one stop searching in current discovery services. The integrated search, which combines a unified index with federated search technology, may come closer to what librarians want.*

Keywords: Library Catalog, Discovery Tools, Federated Search, Unified Index

1. One-Stop Searching-Librarians' Vision for the Next Generation Catalog

One-stop searching is defined as a single point of entry to search all library resources, which include the library's catalog, subscribed databases of journal articles, ebooks, local digital repositories and more. It is one of the important visions for the next generation catalog that has been under discussion in the library community for the past ten years. Until several years ago library patrons were forced to search for resources in multiple places, for instance, books and videos in a library's online catalog, journal articles in electronic databases, and images in local digital repositories. This fragmented approach brought difficulty and confusion to patrons who are used to simple approaches in Google and other Internet search engines. In a world where efficiency and speed are of the utmost importance, libraries are losing a competitive edge to less credible information sources. There has been a tendency of students and faculty turning away from the libraries towards the Internet for needed information. Past research has revealed that "*most students prefer the internet as a means to access information, which in turn posed a threat to libraries. With this, it suggests the need for libraries to adopt technological advances to bring back its role as a good source of information*" (Rose 2009).

Librarians have been searching for a better solution to serve their patrons for quite some time. The next generation catalog (NGC) embodied the visions of librarians from long discussions on the subject. One-stop searching at a single point of entry is at the center of attention and also the essential piece of the NGC. The first attempt at one-stop searching came as federated search in the mid- 2000s.

2. Federated Search

Federated search is also called metasearching, and distributed or broadcast search (Taylor, 2012). It is a single query to search multiple online resources and displays a single list of search results. A federated search engine does not have the elaborate features of the NGC that a discovery tool has. All it does is search, process the retrieved information, and display the results. Many federated search engines were developed more than ten years ago such as WebFeat, MetaLib, and 360 Search. Federated search was popular, and according to a 2007 study, an estimated 3/4 of big research university libraries deployed a federated search engine on their websites (Primary Research Group 2008).

A complicated and difficult task, a federated search engine pulls information from disparate sources through connectors or translators and displays the results to users. Connectors or translators are computer programs and each online data source needs its unique connector or translator that will map the search terms users entered from the federated search engine to search fields in the proprietary database. In other words, the connector will translate a user's query syntax and map it into the corresponding search fields in the target data source. The connector will retrieve the search result page and parse the results in the federated search engine for a meaningful display for users. From a technical point of view, it is not a formidable challenge to develop a connector program. A federated search system comes with a set of pre-

developed connectors. If a library subscribes to a database that is not on the list of the connectors or translators in the federated search engine, a library can pay a fee for the vendor to develop a connector for the resource. Real-time metasearching makes federated search attractive.

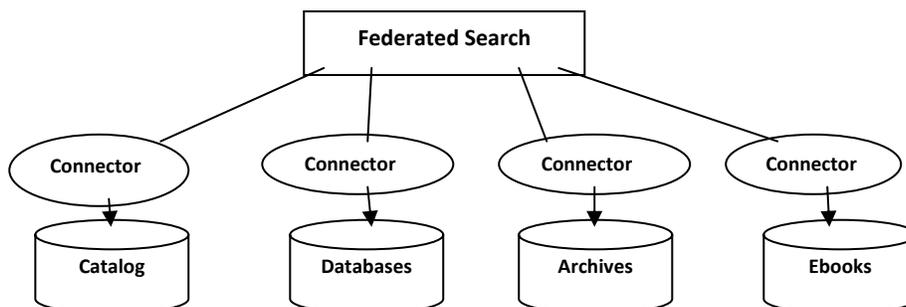


Figure 1. Federated Search Does Real-time Search through Connectors

While federated search is a progress that brought libraries closer towards achieving one-stop searching, it fell far short of expectations. For one thing, it had difficulty in the final stage of displaying search results. Processing a list of search results by reading and dissecting the HTML that a database provides is called “*screen scraping*” is a difficult way to obtain search results because the result data is not structured in a way that makes it easy to identify the fields in the result records. In addition, a federated search engine has to face the daunting task of what is called deduping, a process that involves moving the duplicate items from different databases and grouping them together for display. For another, databases change and improve interfaces and the connectors have to be modified accordingly to work with the new design. As Dalziel put it, “*While the advantages of metasearching are in acceleration and simplification of information searching and gathering, its downside is in maintenance problems because the system relies on connections to so many databases, and those connections are fragile and easily broken, making it a burden to maintain the system*” (Dalziel 2008), federated search is not totally satisfactory.

In the late 2000s, one-stop searching began to have a new solution with better displays and fast speed in information retrieval. The new inventions are discovery tools.

3. The Unified Index in Discovery Tools

With the advent of discovery tools, federated search became less popular and almost faded from the market. While an online public access catalog usually is a module of an integrated library system (ILS), a discovery tool is third-party software that is independent of a library system. It serves as a library user interface that can replace or live side by side with a library online catalog. Discovery tools provide many advanced features of the next generation catalog, while federated search does not. Gradually, discovery tool moved from local installation to the cloud. Today most discovery tools are offered as “*Software as a Service*” (SaaS) which is a subscription-based solution. Thus, many are called discovery services.

Not all the discovery tools are capable of one-stop searching, even though it was clear from the beginning that it was the most sought-after feature in a discovery layer. The past three years witnessed the releases of Web-scale discovery services such as WorldCat Local, Summon, Primo, and EBSCO Discovery Service. The Web-scale discovery services stand out from the rest because “*Web scale discovery services are a tool with major potential to transform the nature of library systems. These services are capable of searching quickly and seamlessly across a vast range of local and remote content and providing relevancy-ranked results in the type of intuitive interface that today’s information seekers expect*” (Cooppner 2013). At the center of Web-scale discovery services is a central or unified index and vertical search. The unified index is essential for one-stop searching. It provides a single point of entry where one search will retrieve results from multiple resources. Unlike federated search that goes out and does real-time searching in multiple resources, searching in a single unified index is fast and allows better deduplication and relevancy ranking. Pre-formed facets and many other advanced features make its user interface much more meaningful than federated search. A Web-

scale discovery tool allows tag clouds, book images, RSS, perm links, auto-completion of search terms, spell checking, and even FRBR relationships, all of which the federated search lacks.

The central or unified index is a database of pre-indexed metadata from disparate sources including remote full-text content and data in silos. It uses the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to harvest metadata from repositories provided by remote resources. To be more specific, OAI-PMH is based on the client-server architecture where the unified index harvests metadata from updating repositories based on a commonly understood data structure. *“The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) offers a simple technical option for catalogue and repository services to make their metadata available to other services, based on the HTTP and XML standards, resulting in facilitated discovery of distributed resources”* (Fegen 2007). Even though pre-indexing is not a new technology, its application in discovery tools is revolutionary. A library may have a vast array of licensed and local content containing millions upon millions of items. Pre-indexing licensed content involves negotiation and permission from publishers and content aggregators. Building a unified index is a huge project that needs strong financial and technical resources. That explains why only four discovery tools provide a unified index. Those Web-scale discovery services include EBSCO Discovery Service, Primo, Summon, and WorldCat Local.

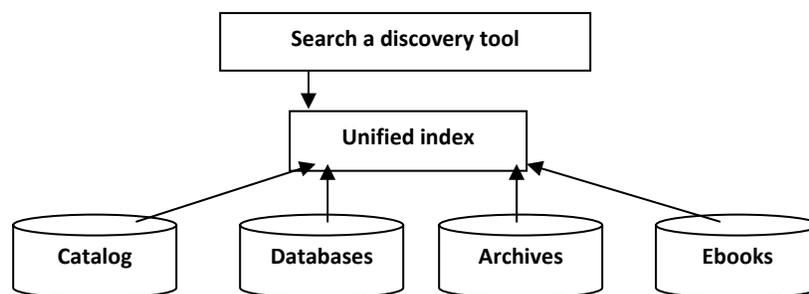


Figure 2. Unified Index Harvest Metadata and Catalog Records (Taylor 2011)

Research has indicated that adoption of a good discovery tool may increase the use of library resources by students, sometimes from 50% to 150% (Fitzpatrick 2010; Way 2010). More studies are needed to verify the findings. If this is true, the unified index plays the vital role in the success of proving a library’s relevance as an information keeper and provider.

Even with the inception of Web-scale discovery tools, one-stop searching is still more a dream than reality. The unified index is nowhere near perfection. Because each Web-scale discovery service creates its own unified index behind closed doors, no one knows how much those indexes differ or duplicate each other in content coverage. Furthermore, discovery services are limited in how much they can obtain from publishers and information aggregators, and these are important players in building a comprehensive index. One thing that can be ascertained is that none of the discovery tools can claim to cover a library’s resources fully. While it is a huge investment to subscribe to a discovery tool, libraries hardly know what they are paying for. Not all the discovery services are willing to release their content covered by the unified index. Thus an analysis or comparison of what each unified index covers remains impossible. It suffices to say that the unified index only provides pointers to most of or partial library resources, meaning one-stop searching does not truly exist.

Beside political and economic issues, the unified index has other setbacks. For instance, no matter how often a unified index is updated, it always lags behind in content. Known-item search has raised complaints, as the desired items may be hidden beneath thousands of hits in the search display. Sometimes the unified index does not cover what it claims to cover. The same search in the unified index may produce fewer or totally different results than that conducted in the native search interface of the resources. A unified index also lacks precision in information retrieval. Thus, serious researchers are still encouraged to search the disparate native resources. The value of the unified index lies largely in

the service it provides to freshmen or sophomores that are inexperienced in research, or for those who desire an interdisciplinary search experience.

4. Other Attempts at One-stop Search

Most discovery tools do not have a unified index, especially the open source ones. Nevertheless, there are various clever attempts to add one-stop searching for a discovery tool that do not come with a unified index. VuFind is the most popular discovery tool among open source and free solutions. Even though it does not have its own unified index, it can connect to the content search of a Web-scale discovery service through an application programming interface (API). A library must subscribe to a Web-scale discovery service before VuFind can search its unified index through the VuFind user interface. When installing VuFind, a user is provided the option to add the search capability of Summon or WorldCat Local in its configuration settings.

Serials Solutions' AquaBrowser is another discovery tool with many advanced features of the next generation catalog. Adding its native federated search engine 360 Search, it can query electronic databases and conduct article-level searches in addition to local holdings. Innovative Interfaces' Encore connects to resources by connectors, and does real-time searching. Significantly Innovative has formed a partnership with EBSCO Discovery Service by which the EBSCO unified index is searched by Encore (EBSCO Publishing 2012). In the next few years more discovery tools without federated search or unified index will look for ways to add one-stop searching. It has become a major setback for a discovery tool to miss this capability.

5. Future Outlooks

From a federated search to unified indexes, one-stop searching is still evolving. It slowly dawns on some that it might be a better solution to combine federated search with unified index. While each has its own pros and cons, these two approaches can be positioned to compliment, not replace, each other. This idea is not widely spread, but echoed somehow in the recent library literature. As early as 2011, Mike Taylor, who works in the field of data and indexing, explained the superiority of such a solution. The combined approach is called “*integrated search*” (Taylor 2011). There are resources whose metadata a unified index cannot harvest for various reasons, but federated search may include them in its search path.

The search results returned from a unified index are different from those of the federated search. They may come with pre-formed facets and relevancy ranking and may have completed the deduping and merging process. Integrated search must provide a way to combine those two totally different kinds of data into one single display list. Obviously integrated search has its own share of problems to deal with as well. In spite of the technical difficulties that accompany this hybrid solution, integrated search will expand the search scope and content coverage, carrying the dream of one-stop search even further towards the desired goal. It may not be true one-stop searching, but it certainly will come closer to it than federated search or unified index working separately.

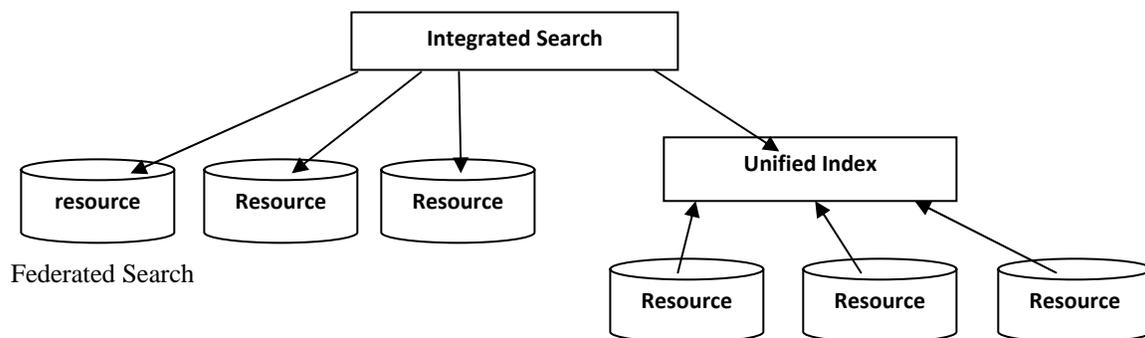


Figure 3. Integrated Search

A recent open-source discovery layer developed by a Japanese researcher, Prof. Takashi Harada, at the National Diet Library of Japan, is an example of integrated search. The new discovery tool, NDL Search (National Dietary Library is the name for the national library of Japan) uses the hybrid approach of a unified index and federated search. With this approach, it can search other national libraries in Asia as well as its local holdings in the pre-indexed metadata in one search.

One-stop searching is a vision that may be fulfilled one day, but not in the near future. Integrated search is the right path to take if one-stop searching is what libraries aspire to possess.

References

- DALZIEL, K. Metasearching: Not as good as we'd like it, *Nebraska Library Association Quarterly*. 39 (1) 14-20. 2008.
- EBSCO PUBLISHING. EBSCO Publishing and Innovative Interfaces Expand Partnership with EBSCO Discovery Service™ and Encore™, available at <http://www.ebscohost.com/newsroom/stories/ebsco-publishing-and-innovative-interfaces-expand-partnership> (accessed 24 May 2013). 2012.
- FEGEN, N. What Is the OAI Protocol for Metadata Harvesting, available at http://misc.jisc.ac.uk/refmodels/PLE_wiki/wiki.cetis.ac.uk/What_is_the_OAI_Protocol_for_Metadata_Harvesting.html (accessed 22 May 2013). 2007.
- FITZPATICK, S. Summon Summons 100th Customer, *American Libraries*. 41 (10), 24-25, Library Literature & Information Science Full Text (H.W. Wilson), EBSCOhost, (accessed 23 May 2013). 2010.
- HOOPNER, A. Ins and Outs of Evaluating Web Scale Discovery Services, available at <http://www.infotoday.com/cilmag/apr12/Hoepner-Web-Scale-Discovery-Services.shtml> (accessed 21 May 2013). 2013.
- PRIMARY RESEARCH GROUP. *Academic Library Website Benchmarks*, available at http://books.google.com/books/about/Academic_Library_Websites_Benchmarks.html?id=t4_gqG6wZaAC (accessed 19 May 2013). 2008.
- ROSE, S. Student views on library services: key lessons for developing libraries of the future, *SCONUL Focus*. Winter (47), 4-6. 2009.
- TAYLOR, M. When Worlds Collide: Metasearching meets Central Indexes, available at <http://www.nesc.ac.uk/talks/1114/Taylor-OpenEdge2011-when-worlds-collide%5B1%5D.pdf> (accessed 23 May 2013). 2011.
- WAY, D. The Impact of Web-scale Discovery on the Use of a Library Collection, *Serials Review*. 36 (4) 214-220, Academic Search Premier, EBSCOhost (accessed 23 May 2013). 2010.