

Implementation and Empirical Analysis of Web Discovery Tools with Special Reference to India

Vasanthkumar R.*, Gopal K.**

Abstract

This paper throws lights on how discovery solution works and provides the potential features of Discovery Solutions in context to Library environment. It also evaluates the available web discovery services offered by eminent organizations and also provides a practical analysis of available library discovery tools in context of the present-day explosion of available open search engines on the Internet. The focus of analysis include how discovery tools are expected to manage library collections, provide access to scholarly information content, as well as other factors. It discusses some of the possibilities regarding how these technologies, methodologies, and products might be able to adapt to changes in the evolving information landscape in scholarly communications and to take advantage of new technologies, metadata models. It also throws light on Indian scenario of web discovery service.

Keywords: E- Resources; Single Search Window; Relevance Ranking; Search Engines; Web Discovery Tools.

Introduction

In recent days Libraries have shown interest in the tools and technologies that offer more convenient way of accessing to the resources for the communities that they serve. These tools evolved steadily in recent decades, making great progress in the scope and depth of materials addressed.

The development seen in the successive generations of technology beginning from online catalogs, to Meta search tools, to the current generation of index-based discovery services represents an incredible improvement. The users of academic libraries including students, faculty, researchers and other users have left to augment Google results by searching library databases individually. Some of the libraries are using meta search engines also referred to as federal search engines, which are

simply searching across databases, but they have fallen short of libraries expectations in addition to this federal solution are slower of response, problems with relevance ranking and in adequate handling of duplicates are some of the major problems [1].

Web scale system really creating an information system that integrates nearly all library content to a single platform information professionals are coordinators of all these process and it is essential that they should have good understanding about the concepts of web discovery solutions, its technology, major players and evaluation parameters of selecting a discovery service, Our end users certainly are familiar with the idea of a single search across a comprehensive index that produces a large, relevancy-ranked results list.

Even though most patrons would not recognize the term web-scale discovery (WSD), it is what they have come to expect. More and more libraries are stepping up to meet their users' expectations by implementing WSD services.

Librarians around the world are trying to learn what these services are and how they work, evaluating the services on the market, selecting and implementing a service, and then teaching colleagues and patrons all about it.

Author's Affiliation: *Edutel Technologies Pvt. Ltd., Yeshwanthpur, Bengaluru-560022, Karnataka, India. **District Central Library, Chitradurga, Karnataka, India.

Reprint's Request: Vasanthkumar R., Edutel Technologies Pvt. Ltd., Yeshwanthpur, Bengaluru-560022, Karnataka.
E-mail: vasanth.kumar@edutelmail.in

Received on 04.01.2017, Accepted on 23.01.2017

Literature Review

This literature focuses primarily on usability studies and criteria for choosing a web-scale discovery service, with emphasis on search performance. There are few studies specifically reviewing the search performance of web-scale discovery services with each other of those, most base their evaluation of search performance on a very small sample of searches (e.g., Timpson & Sansom, 2011; Zhang, 2013). The studies below represent more extensive attempts to compare the search performance of these products. Asher, Duke, and Wilson (2012).

Study was based on searches performed by test subjects, assessments had been made using studies carried by eminent authors of useful research papers. Rochkind (2013) compared user preference for search results produced by EDS, Summon, EBSCO host "Traditional" API and Ex Libris Primo. His survey tool allowed subjects to enter search terms of their own choosing and view side-by-side results, within the survey window, for two randomly chosen products. Each product was configured to exclude non-scholarly content. The study found no significant difference in preference between products.

Implementation

Implementation phases can run from 3-12 months or more, depending on the size of the library and the number of content providers (Stone, 2010). The core set of content providers established in the evaluation phase should serve as the first build of the discovery data. In the case of e-book content providers, be aware that local e-book subscriptions may not be complete sets, and e-books may have to be identified at the title level rather than the set level. Identify next the OAI-compliant services that can also provide metadata for simultaneous searching in the discovery tool, such as Content DM, DSpace or other image or document repositories. These are easily integrated into the discovery tool, usually by providing a single URL for each OAI server. Objects containing metadata should be reviewed for completeness; depending on customization of the metadata in local systems, additional mapping may be required similar to the ILS database at the discovery tool level (Allison, 2010; Luther, 2011).

Discovery tools, by design, are extraordinarily capable of searching multiple fields of data quickly, whereas existing ILS data is derived from specific, limited MARC fields indexed in one or more narrowly defined indexes. Notes fields, local holdings fields,

even item-level information including faculty course reserves or acquisition pricing, could conceivably be included if library chose to harvest that information for inclusion in the discovery tool. In some cases, even the weighting and precedence of certain fields of data can be determined by the library in the discovery tool. The ILS database is not a stagnant data file, so once the entire ILS database is uploaded, synchronization scripts need to report additions, changes, and deletions of records from the ILS database and convey these to the discovery tool, daily, if possible (Marcin & Morris, 2008). Full database reloading, once scripts are in place, should not need to occur more than twice a year. The discovery tool synchronization intervals should be flexible to accommodate the needs and work patterns of the library, including rapid additions of large sets, or only seasonal discards for weeding. Note that this synchronization step may not be needed if the ILS and discovery tool are on the same vendor platform; or, more likely, the scripts will be provided and executed automatically, if on the same vendor platform.

EBSCO

EBSCO Information Services is involved in three main activities, offering subscription services to libraries for print and electronic journals, developing tools for libraries to manage their electronic resources based on the EBSCONET platform, and producing database and discovery products on its EBSCO host platform. It not only provides a search interface for accessing e-books and full-text journals that are native to the EBSCO platform, but it also indexes and provides links to most of the electronic content that the library has purchased or leased from other information suppliers. Furthermore, it indexes the catalog of the library's physical inventory the books, journals, CD-Roms, maps, etc., Ebsco provides a link to the content straight from the citation. In the case of the library catalog, real-time availability is supplied showing the location and status of the material. It provides superior relevancy ranking, most comprehensive collection of full text for searching and capable of leveraging controlled vocabularies for key subject indexes. Flexibility, including interface customizations and seamless interaction with most third-party vendors such as document delivery services, ILS etc. Like the ability to search all databases at once. There were several databases that were not able to be loaded into EDS due to EBSCO might not have agreements with them which made having a discovery service less than ideal and the user interface is confusing. It is a challenge to make it work with non-EBSCO partners/products. It searches

well, and results are reasonably relevant. However, it presents a pedagogical problem in the subjects or discipline- and need-specific databases. It is a great step forward for improving access to journal articles, but it lacks in searching for known items in the catalog. It is useful for academic libraries, but public, school and special libraries are generally facing confusion on which search tool to use and sometimes it does not integrate very well with current ERMS.

Worldcat Local

OCLC launched a pilot for WorldCat Local in April 2007. Libraries using World share Management services operate directly with the WorldCat database; those using a local ILS might need to synchronize their holdings on WorldCat through a process called reclamation. OCLC’s WorldCat Local is the web-based discovery service that delivers access through a single search box to more than two billion items. It also offers syndication services through partnerships and collaborations with more than 200 partners that make library collections accessible to searchers through leading search engines and other websites, More than 977 million articles with easy access to full text, 37 million digital items from trusted sources like Google Books, OAIster and HathiTrust, 15 million e-books from leading aggregators and publishers, More than 30 million pieces of evaluative content. Some issues with display and wording that are confusing to users and librarians, and links to items are not consistently available. Its strength is in delivering records for any book imaginable, but naturally it is not strong for journals. Have some Functionality issues needs urgent fixing and support has been less than helpful. OCLC’s WorldCat Local discovery tool provides access to local, consortia, and worldwide content in a single user interface. The limited scope of the search makes it mostly useless for advanced students and faculty, but the unsorted hodgepodge of search results is not ideal for novice.

Primo + Primo Central

Primo was developed by Ex Libris Group, Primo was launched as a discovery interface in 2006. Primo was designed to provide a more modern and sophisticated interface for library collections, based

on a local index created from records imported from a library’s ILS or other local or remote repositories for which the library can load copies of metadata. Save search queries and use them again without having to reformulate the query. Change the number of results that appear in the brief result display. Change the interface language based on the interface language made available to the users by the library, Save (“push-to”) items to a personal e-Shelf as well as third-party applications such as Connotea, del.icio.us, RefWorks, and EndNote Web. Export in RIS format is supported to enable the user to work with the client versions of various citation managers such as EndNote. Print and e-mail results or have results sent to their cell phone (by SMS) so that items can be easily located on the library shelf. Configure links from a record to other queries, to individual records or to other systems, based on a URL template and information in the record. Problems with relevance ranking and support and generally works OK if you search like you’re using Google and use the facets to narrow your search. Less effective for advanced searchers who want to do Boolean searches, etc. Primo is a step forward from MetaLib, but it has definite shortcomings.

Summon

The Summon discovery service was created by Serials Solutions, a division of ProQuest. Serials Solutions was founded in 2000 to help libraries manage their e-journal holdings. Current Serials Solutions products include 360 Core, its base package for e-resource management, The company is currently developing a new library services platform called Intota. A preliminary product, Intota Assessment, provides tools for data-driven collection management. Like other web-scale discovery products, Summon provides a pre-harvested central index allowing users to search across a library’s book and journal holdings through a single search box and provides critical feedback to the vendor. The Summon API and RSS feeds potentially allow deeper, more customized integration into library services; the use of Unicode throughout Summon means that it can be searched in multiple languages (Arabic; Chinese; etc). Summon Topic Explorer – Highlights relevant reference resources and provides recommendations to related topics. Database recommender directs users to

Table 1: Example of how discovery service displays the results

Examples of Known-item and Topical Search Queries	Topical search queries:
Religious and cultural background of north eastern part of India	Religion and culture
National Cultures and Work Related Values The Hofstede Study	solar power coating nanoparticle
Economics contribution of Koutiliya	Koutilyana Arthashathra
How wind and solar energy contribute to Global warming	Solar and wind energy

specialized databases through a combination of library controlled and community-sourced tags and relevance based recommendations.

Content spotlighting, visually distinguishing valuable content by type, content spotlighting dynamically groups newspaper and image content into distinct visual elements within search results to ease navigation and evaluation of these results and promote unique content in the library's collections. Facet category, Discipline—Supported Authoritative, item-level discipline mapping in the Summon index allows users to zoom in on discipline specific content or combine disciplines for interdisciplinary searching. Direct linking to resources has improved access and usability. Option to exclude citation records has also improved usability and leads to full-text download. Good improvement in coverage of OA content and linking problems mainly caused by withholding of metadata by competitors. It is difficult to integrate with ILS. For example, clear holdings availability and ability for patrons to view their own account using the same interface and Clunky interface. Ongoing problems implementing new titles and the relevance ranking, level of duplication, interface, and index coverage could use improvement and main difficulty with Summon has been in ingesting and displaying catalog records. Issues with article linking through to full text and Provenance of search results (database) unknown until native interface is reached, and also eBook linking is problematic.

Indian Scenario

In India, research Libraries like University Libraries, IIMs, IITs, AIIMS etc are subscribers of electronic journals, electronic books, and databases and their own digital repositories and OPACs and many of them are part of different consortiums. Libraries similar to this array are desperate for Discovery Solution. In this scenario, users are in advantageous position regarding access of resources but often in the confusion, from where to start and which resource to be used to get their information. This forces users to depend on Google like search engines to get their information. Many studies show that many of the resources are underutilized due to lack of information to users regarding the availability of such resources in library. Web Scale discovery solutions eliminate this confusion and provide Single Search Box environment to users to retrieve all the relevant information from multiple sources that are subscribed by the Library. Discovery solution can make a difference in reducing the valuable search time of researchers and also better chance in utilizing the near optimal utilization of Library subscribed

resources.

Still web discovery service in its early stage in a country like India, while randomly surfing the websites of premier Institutes especially IITs and Central Universities, EDS is the most commonly used discovery service in India. Indian Institute of Spices Research, Kozhikode is one of the best examples for Research Libraries in India who adopted Discovery Solution. They have only around hundred FTE., still they subscribed EBSCO Discovery Service in the year 2012, considering the value to researchers and retain the subscription for the third consecutive year. IIT Bombay, Delhi, Guwahati, IIMs of Raipur, Ahmadabad, Kozhikode, Ranchi and JNU (Delhi) is another example which are successfully implemented EBSCO and providing single window search facility for their users.

Conclusion

This study presented the various implementation procedures of web service discovery methods. It is observed from the study that different approaches are used to measure and to estimate the accuracy of the discovered web services. This study compares the search performance of web-scale discovery services of ProQuest's Summon, EBSCO Discovery Service (EDS), OCLC and Primo + Primo Central. There was found to be no significant difference in performance of above mentioned discovery service providers. Because there was no significant difference in the search performance of available discovery service, any decision to purchase one product or the other should be based upon other considerations (e.g., technical issues, cost, customer service, or user interface). Web Scale services are still in its initial stages of development and lots of developments in the features, functionality, level of integration with other systems, scope of content, and soundness of metadata, flexibility of the interface are all evolving and it is expected, will continue to evolve in meeting the needs and expectations today's next generation users. The comparative analysis shows that all the major service providers are extending competitive features and services, but varies in some features and the choice is depends on the concerned library's preference, selecting, evaluating and implementing a web-scale discovery products has taught us much about project effectiveness, communication strategies, implementation processes and ongoing challenges. Such lessons will stand us in good stead not only in this instance but also in the future as we continue to grapple with the ever-increasing rate of technological change and innovation.

References

1. Susan Johns-Smith, Evaluation and Implementation of a Discovery Tool, Kanas Library Association, 2012; 2(1):17-23.
2. Parnab Chatterjee, Debasis Das, Web Scale Discovery Services and Tools with special reference Central Library, IIT Kharagpur, International Journal of Innovative Science, Engineering & Technology, March 2015; 2(3):677-686.
3. Kwon O, Im GP, Lee KC. MACE-SCM: A multi-agent and case-based reasoning collaboration mechanism for supply chain management under supply and demand uncertainties. *Expert Systems with Applications*. 2007; 33(3):690-705.
4. Grigori D, et al. Ranking bpel processes for service discovery. *IEEE Transactions on Services Computing*. 2010; 178-92.
5. Jordan D, et al. Web services business process execution language version 2.0. OASIS standard. 2007; 11:10-13.
6. Marshall Breeding, Library Resource Discovery Products: Context, Library Perspectives, and Vendor Positions, Major Discovery Product Profiles, January 2014; 50(1):27-41.
7. Karen Ciccone, John Vickery, Evidence Based Library and Information Practice 10.1, Summon, EBSCO Discovery Service, and Google Scholar: A Comparison of Search Performance Using User Queries, 2015.p.34-49.
8. Suchithra M, Ramakrishnan M. Enhanced web service ranking approach based on non-functional QoS criterion with user preference. National Conference on Computing Techniques (NCCT'15); India. Mar 2015.p.67-72.
9. Guo W-Y, Qu H-C, Hong C. Semantic web service discovery algorithm and its application on the intelligent automotive manufacturing system. 2010 The 2nd IEEE International Conference on Information Management and Engineering (ICIME); IEEE; 2010.p.601-4.
10. Klusch M, Fries B, Sycara K. Automated semantic web service discovery with OWLS-MX. Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems and ACM; 2006.p.915-22.
11. Zhou B, et al. Using inverted indexing to semantic WEB service discovery search model. 5th IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCom'09); 2009.p.1-4.
12. Aiello M, et al. Web service indexing for efficient retrieval and composition. E-Commerce Technology. The 8th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services; 2006.p.61-4.
13. Rajendran T, P. Balasubramanie. An optimal agent-based architecture for dynamic web service discovery with QoS. 2010 IEEE International Conference on Computing Communication and Networking Technologies (ICCCNT); 2010.p.1-7.
14. Chukmol U, Benharkat AN, Amghar Y. Enhancing web service discovery by using collaborative tagging system. IEEE 4th International Conference on Next Generation Web Services Practices (NWESP'08); 2008: 54-9.
15. Implementing web scale Discovery Services Special reference to Indian libraries context. By nikesk narayanan, <http://www.slideshare.net/nikeshn/implementing-web-scale-discovery-services-nikesh>.