

WEB SCALE DISCOVERY SERVICE: DESIGN AND DEVELOPMENT OF 'SINGLE WINDOW SEARCH BOX' AT CENTRAL LIBRARY, IIT KHARAGPUR

Parnab Chatterjee

School Librarian, Janardandih High school Email:parnabchatterjee@gmail.com

Debasis Das

Library Professional, IIT Kharagpur, E mail: debdas12312@gmail.com

Abstract

The purpose of this paper is to discuss Web Scale Discovery Service and tools and find out how much discovery tools have achieved towards becoming the next generation catalog. The paper summarizes the evolution of Discovery service and tool. The paper also discusses the service provided by Central Library of IIT Kharagpur i.e. "eSearch Software". Discovery tools have many next generation catalog features, but only a few can be called real next generation catalogs. Federated searching and relevancy based on circulation statistics are the two areas that both open source and proprietary discovery tools are missing. This paper gives information about discovery service provide by Central Library of IIT Kharagpur. It will provide information as to exactly where discovery tools stand in light of the much desired next generation catalog.

Keywords: Web Scale Discovery Services, Discovery Service- evaluation, Discovery Tools, Federated search, Next-Generation cataloguing, e-Search Software.

Introduction

The impact of Internet and Google like search engines radically influenced the information behavior of Next Generation users. They expect same environment in library services such that all their required information make available in a single set of results through unified search across all the available resources. Libraries have been striving to respond to this challenge for years. Until recently, federated search technology of the past decade was the better attempt in this area to meet these user expectations. But federated search solution is marked by the drawbacks of its slowness as it searches each database on the fly. New Generation cloud based Library Web scale discovery technology is a promising entrant in this landscape.

What Is Web Scale Discovery Service

'Discovery services' is a new name for the traditional information retrieval system. A bit of market confusion is created by the new discovery product, vendors, projecting

their products as a potential substitute for many search services that the libraries have been using for decades like card catalogue, OPAC, Web Opac, Bibliographic database search etc. Two things are projected as key benefits of these new age discovery services. First they act as a single point search interface for all the contents (subscribed, in-house and open access) that the library wants to present to its user community for searching. It is mimicking Google, and second they operate on web scale. Most of libraries have different types of electronic resources and each e-resources database has their own search engine therefore users need to visit different search engines to get their required information for which it is required services a single point search interface for all the content has emerged, which is more effective and useful for the users.

Web Scale Discovery services are certainly making waves of revolution in Library and information search arena. Case studies

show that these services are getting wide acceptance both among Library staff and also from patrons. Google like simplicity and efficiency in providing relevant result attracts users and thus bringing back them to Library from internet search engines. The success stories of Web scale Discovery services is the evidence of a notable happening of an emphasis shift in the library world from the in-house installed software to cloud based services. 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 net 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 and the cost involved.

Web scale discovery services are those integrated web based services with major potential to transform the nature of library systems. These services are offered as cloud computing model and have the capacity to more easily connect researchers with the library's vast information repository including remotely hosted resources and local content. It provides a unified platform for library users to access and search from all the library resources to get single set of results by providing a Google like environment with the following basic features.

- Unified platform to search all the resources including licensed, open and local collections
- Pre-harvested central index of metadata
- Google like single search box

- Single results list for all collections
- Relevancy ranking across entire results
- Full featured user interface
- Facets and tools for narrowing results
- Holdings and status information for library catalogue items
- Connections to full text
- Infrastructure, processing and indexing provided and maintained remotely by the vendor.

Evolution of web scale discovery

Evolution of information discovery tools within the library context started with card catalogs. As a next step, those card catalogs transferred to online integrated library systems (ILS). These were only available in the library. In the 1990s, with the development of the web, HTTP web-based online catalogs were created. Electronic journal content, e-text and e-book content, abstracting and indexing databases also appeared in 1990s, but firstly in CD-ROM format. In 1998, a new search engine developed by two Stanford graduate students. It was Google which is often first and mainly the last stop for today's' users who have grown up with Google for information discovery. Google became the standard for users with its simple interface as easy as entering keywords in a single search box, speed, accessibility wherever a searcher can get Internet access, broad content and quality results.

In contrast to Google, libraries generally subscribe about 100-400 databases each of which have different interfaces. Federated searching also known as metasearching, identified by Roy Tennant and others during late 1990s and early 2000s, was supposed to be simple interfaces providing seamless searching across logically clustered databases of information and being a way of meeting the expectations and

needs of Google generation, that they would allow users to search, retrieve and display content from multiple resources such as abstracting and indexing databases and full-text databases simultaneously as easy as entering keywords in a simple search box. In other words, they could allow libraries to become "one-stop shop" for their users. It is one of the major advantages of federated search to access library resources without having to select a specific database and without repeating the search that users do not want to jump from one interface to another when they are searching for information. The biggest players in this market are MuseGlobal, Serials Solutions WebFeat, and Fretwell- Downing.

The development of federated searching had become one of the main growths in academic libraries at those years that many libraries thought federated searching instead of Google because of including more scholar sources than Google. Federated search tools have not been able to achieve to provide a Google-like search box that can quickly retrieve information for all library resources for several reasons. Over time, it has been clear that federated searching are not meeting the needs of users. Issues raised on the capabilities of federated search mainly as difficulty, complication, slowness, merging of resources, reduplication (merging results) and ranking of retrieved results. Beside these, the number of individual resources that can be simultaneously searched was limited with federated search tools. It may be better for smaller libraries or public libraries to use federated search tools for providing access to a selected group of databases.

In 2000s, library web catalogs changed as "next-generation" library catalogs. Next-generation library catalogs such as Encore, Aquabrowser, Exlibris Primo were more functional that they had "Web 2.0" features

like tagging, submission of reviews, facets etc. and a user interface with popular sites like Amazon. They provided harvesting of records from various locally hosts; catalog records from one host, digital collection records from another host. They search, retrieve and present results in a single next-generation catalog interface. It was believed that these modern looking interfaces would retake some of the users from Amazon and maybe from Google but they have failed so far, users continued to use Amazon, Google. Library Catalog was used more to check whether an item they found in Amazon or Google was available in the library. Next-generation catalogs are still new technology for many libraries.

Development of Google Scholar next to these issues about federated search and next-generation catalogues, a quest began for a new kind of resource that would compete with Google Scholar both in terms of speed, scope, harvesting and preprocessing of information. Beside these, the new kind of resource is supposed to suggest better search terms, spell check, suggest other terms based on the entered search terms, easily live help, relevance ranking as default display, helping links, include less jargon such as Boolean, ISSN etc..

After the failure of Federated search engines and Next generation catalogs, libraries began to suffer against Google and Google Scholar. A few months after Google Scholar start in November 2004, Marshall Breeding discussed that federated searching could not compete with Google Scholar's speed and power. He called for a "centralized search model". Web discovery tools became the latest attempt to solve this problem, providing a Google-like single search box that has access to all library resources. Different from federated search engine which searches multiple databases and

aggregates the results, they search a unified index and presents search results and a single interface Federated search engines results rely on each tool's search algorithm and relevance rankings as well as the federated search engines. Web discovery tools import metadata into one index and apply one set of search algorithms to retrieve and rank results. First web discovery service appeared in late 2007 was OCLC WorldCat Local. In the middle of 2009 Serials Solutions announced its web scale discovery tool Summon. Others followed with similar products, such as Ebsco's Discovery Service in early 2010, finally Encore Synergy and Exlibris Primo Central in mid 2010. The approach of Encore Synergy to web-scale information discovery is a bit different from the other four. It doesn't create a preharvested control index. Ebsco Discovery Service and Innovative Interfaces Encore Synergy were released in 2010. There are not any studies on these Web Discovery Services yet. Also some open sourced Web-scale discovery services, such as eXtensible Catalog (XC) Project by University of Rochester were developed.

Summon was announced in January 2009 with Dartmouth College Library and Oklahoma State University Libraries as beta sites. University of Sydney, and the University of Liverpool were the other two beta sites. Summon launched commercially in July 2009 and recognized as Best Enterprise Search Solution at 2011 CODIE awards by the Software & Information Industry Association (SIIA). It has the capability to search more than 800 million records that is more than even the size of the largest traditional library database. It contains more than 35 Open Access repositories and more than 90 institutional repositories, providing the most full-text searchable content. The service is openly available for searching without

authentication. It includes content and metadata from more than six thousand publishers, database producers and content providers. The content consists of different type of sources such as manuscripts, archival materials, journal articles, monographs and sound recordings. For harvesting content, Serials Solutions' Summon has been contracting with dozens of content producers, publishers including the American Institute of Physics, the American Psychological Association, Cambridge University Press, Oxford University Press, and Springer. Summon is also capable of utilizing OpenURL and Digital Object Identifiers. Summon is able to search better than even the best federated search tool. Unlike from Google Scholar, Summon is tied to a library's resources. Finally, unlike federated search or Google Scholar, Summon's normalized data also allows for a greater level of purification of data before and after the search. John Law who is the director for Serial Solutions Summon noticed that Summon had a very significant impact on library usage. From the publisher side, libraries using Summon have increasing use of their content.

Sam Brooks indicated that the main aim for developing the EBSCO Discovery Service (EDS) was helping libraries compete with Google and Wikipedia that one of the most important content of EDS is academic encyclopedia. Millersville University's integration to EDS was easy because of common using of EBSCO products in this university. Usage of some subject databases increased after EDS, it also prevented the cancellation of some. Carl Grant pointed out that the libraries using Primo have increased average number for search sessions, on the other hand, decreased session length. They thought this situation as the effect of quickly finding of what is looked for by users. Chip Nilges states that OCLC WorldCat Local

came along for integrating library collections that provides a single search of all library collections. University of Delaware (UD) was the first place of production for WorldCat Local and it was a major conceptual change for the library staff of UD.

There are different criteria to consider while selecting a web discovery service for your library/institution. Some of the important criteria are scope and depth of content, richness of metadata, simplicity of interface, customizing of interface, supporting of mobile access and of course cost. Diane Bruxvoort, summarized main issues for selecting discovery services for an academic library as prices, top criteria of the institutions, discovery service providers, time you need to get ready for a new product etc.. Carl Grant think that the next steps for discovery services are “personal relevancy ranking”, “improved mobile interfaces” and “addressing of growth in e-book usage”.

Implementation of e-Search Software for providing Discovery Services at Central Library, IIT Kharagpur

Design and development of 'Single Window Search Box' by Central Library started in the year 2010 when library professionals want to search all the journal titles which are having the phrase “Computer Architecture”. Library realized need of a single window search box, which will give the names of all the journal titles having the phrase “Computer Architecture”. The digital resources that libraries makes available to its users are in varied formats like text, audio, video, or a mixture of various formats i.e. multimedia. Libraries do a great job for selecting the best e-resources for their users. A library has also collection of subscribed e-resources, Physical books, open access digital collections, as well as collection from other sources such as other academic or public libraries, primary or secondary database, aggregator's databases, IDR contents of scanned document of rare books, out of print books faculty publication and PHD theses which need to access through single window search interface.

Full Text Databases Portal @ Central Library, IIT Kharagpur

The screenshot displays the Central Library IIT Kharagpur website. At the top, there is a header with the library logo and name, along with navigation links for downloading Android apps, contacting the library, and visiting the IIT Kharagpur website. Below the header, there are several milestones and achievements listed, such as the 1951 Journey Begins at Sahel Bhavan, 2000 Golden Jubilee, 2009 RFID Pilot Project Implemented, 2010 World-Class Modernisation, and 2013 Technology Sharing.

The main content area is divided into several sections:

- CENTRAL LIBRARY IIT KHARAGPUR**: Includes links for Library Web Mail and Archived Books.
- OPAC SEARCH**: Provides an Online Public Access Catalogue (OPAC) Search.
- INSTITUTIONAL DIGITAL REPOSITORY**: Offers online access of institute's Ph.D theses (with a note to disable proxy while accessing).
- LIBRARY MENU**: A sidebar menu with links to Home, About Us, Ask the Librarian, Downloads, Our Team, Photo Gallery, and Archives on Technical Education.
- E RESOURCES**: A sidebar menu with links to E - Books, E - Journals Subscription 2014, Full Text Databases, Bibliographic Databases, Dissertations & Theses, Previous Year Question Papers, Video Lectures, Open Access e-Resources, and Book Accompany CD.
- Full Text Databases [e-Subscriptions 2014]**: The main content section, dated Thu, 03/27/2014 - 18:29 by cllk@iitkgp. It lists e-Resources and Print Journals subscribed by INDEST-AICTE Consortium and Central Library IIT Kharagpur. It includes a table of Science, Technology, Literature, Business and Humanities Databases with links to view lists for various journals like ACM Digital Library, ACS Journals, APJ Journals, APS Journals, ASCE, ASME, ASTM Standards & Journals, Annual Reviews, and Cambridge University Press. It also lists EBSCO Database Subscriptions: Business Source Complete, Communication & Mass Media Complete, and MagIOnLiterature Plus.
- DOWNLOAD LIBRARY LATEST UPDATES**: A section with links to Arrival of New Books (Weekly Updates), Latest Submission of PhD Theses (Monthly Updates), and 9th Technology Book Fair Photo Album.
- LIBRARY SECTIONS**: A sidebar menu with links to Acquisition, Back Volume, Circulation section, Current Periodical, Digital Library, and Technical Processing.

So the central Library has decided to design and develop the “eSearch Software” a single window search box to access e-resources. Library made an evaluative study of various user interface of e-resource search software mainly MIT “VERA”, Stanford University “e-Journal Search” and Stanford University “Search Works”.

After evaluative study of e-Search software Central Library decided to develop in-house software similar to VERA using open source software. The next version of this software will be developed with a feature to search full text.

Feature of e-Search Software

1. Platform Independent: Server Search Engine may be installed in any type of computer.
2. Develop Using FOSS
3. eJournal Title Search
4. eBook Title Search
5. Boolean Title Search: Present in Advanced Search mode
6. Full Text Search: Will be available in the next version
7. Open Access Journals are Included
8. Use of Wild Cards: % is used to match any number of characters:
9. Integration of other Social Tool or Databases: SCOPUS, WOS, Library Web Opac, Standards etc. are included and gets opened in the same container application.
10. Feedback Mechanism

The user GUI is very easy and handy. Users need to click on the list of databases mention in the left hand side of the windows i.e. e-Book, Online Journals, Conference Proceedings or Standards. Based upon the selection the search window activates then only the key word needs to be entered, the search engine gives the desired result.

The search engine also has an advanced Boolean search where user may put a search based on criteria. Also user may give wild cards like “%” to search varied types of words.

Hardware Requirements:

Processor: Dual/ Quad Core Intel Pentium Processor, base upon the number of clients that will be access the web Server.

RAM: 4 GB- 8 GB, HDD – 300GB

Software Requirements:

OS: Any Platform along with WAMP/ LAMP with latest PHP version (5.1 or above)

Database: MySQL

How to Install: The Installation Process is very easy and will be finished by following the three steps mentioned below.

Step 1: Install the WAMP Server present in the CD WAMP is installed in case the Operating System is Windows LAMP installed in case the Operating System is Linux variant

Please Note: Find you're the windows bit support i.e. 32 bit / 64 bit OS and installed the correct WAMP bit

Step 2: Copy eSearch folder from the CD and paste it into the “WWW” folder of the WAMP

Step 3: Restart WAMP and Browse the website with the URL <http://localhost/eSearch>.

The eSearch software is web application software therefore any well configured computer will support to install eSearch software. Most of the Feedback received from the faculty and the students are mentioning that the eSearch is very handy and useful. The only shortfall of the software is that it does not perform any full text search which is needed by today's information seekers. The forth coming version of this software will have a full text search facility.

Those library users who are searching books on title name this tool is very helpful for them. The search engine performs a search from the local database; therefore, the search result is very quick and precise. Central Library, IIT Kharagpur provides Discovery services, but this is a trial access. Central Library of IIT Kharagpur provides Informatics-Federated Discovery Service, Ebsco- Discovery Service and Summon-Discovery Service.

Conclusion

New services and products go through an evolutionary development phase – it is important to critically evaluate whether a service or product is mature enough to warrant purchase and implementation.

Implementing a product still under development has advantages and disadvantages: the ability to influence development of the product vs. helping the

vendor identify and fix bugs.

A “preview phase” prior to an official launch has the same benefits as beta testing software, e.g. testing functionality with a wider group, gathering feedback for customizations and enhancements, and allowing stakeholders to introduce themselves to the new service at their own pace. This could be useful for many library services.

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.

Discovery Service Portal @ Central Library, IIT Kharagpur

The screenshot displays the Central Library website interface. At the top, there are two main sections: "ARCHIVED BOOKS" with a link to "List of Archived Books" and "INSTITUTIONAL DIGITAL REPOSITORY" with a link to "Online access of institute's Ph.D theses (Disable Proxy while accessing)". Below these is a "LIBRARY MENU" with links to Home, About Us, Ask the Librarian, Downloads, Our Team, Photo Gallery, and Archives on Technical Education. To the right of the menu are three photographs of library events, with the text ":shops" written vertically. Below the menu is the "E RESOURCES" section, listing items like E-Books, Journals Subscription 2014, Full Text Databases, Bibliographic Databases, Dissertations & Theses, Previous Year Question Papers, Video Lectures, Open Access e-Resources, and Book Accompany CD. The "ABSTRACTS OF PHD THESES" section lists abstracts for August 2012 to July 2013 and July 2011 to August 2012. The "LIBRARY UPDATES" section, dated Tuesday, 02/11/2014, includes news items such as "Grammarly@edu [Trail Access]", "Central Library News Published in Eenadu Telugu News Paper Dated 13 Feb 2014", and "Web-Scale Discovery Services - Trial Access (Subscribed E-resources Search Window)" with a list of services: Informatics - Federated Discovery Service, EBSCO - Discovery Service, and Summon - Discovery Service. A "Feedback" button is provided. Below the updates is a link for "Document Delivery Service and Retrieval click to know more". On the right side, there is a "DOWNLOAD LIBRARY LATEST UPDATES" section with links to "Latest Submission of PhD Theses Monthly Updates", "9th Technology Book Fair Photo Album", and "Highest Usage award During 9th Annual meet of INDEST-AICTE Consortium". At the bottom right, there is a "LIBRARY SECTIONS" button.

References

1. Anita K, F., & Popp, M. P. (Eds.). (2012). "Early Adoption: EBSCO Discovery Service at Illinois State University: Planning and Implementing Resource Discovery Tools in Academic Libraries". IGI Global. Retrieved from <http://www.igi-global.com/chapter/evaluating-selecting-library-web-scale/67814>
2. Asher, A. D., Duke, L. M., & Wilson, S. (2012). "Paths of Discovery: Comparing the Search effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources". College & Research Libraries. Retrieved from <http://crl.acrl.org/content/early/2012/05/07/crl-374.short>
3. Axford, M. A. (2012). "Ultimate Debate Program on Web-Scale Discovery Services". (A Report of the LITA Internet Resources and Services Interest Group Meeting, American Library Association Annual Conference, New Orleans, 2011.). Breeding, M. (2010). "The State of the Art in Library Discovery." *Computers in Libraries*, 30(1), pp. 31-34.
4. Breeding, M. (2013). "Beyond the ILS: A New Generation of Library Science Platforms". IGI Global. Retrieved From www.igi-global.com/chapter/beyond-ils-new-generation-library/76457
5. Caplan, P. (2012). "On discovery tools, OPACs and the motion of library language". *Hi Tech*, 30(1), pp. 108-115
6. Comeaux, D. J. (2012). "Usability Testing of a Web-Scale Discovery System at an Academic Library". <http://dx.doi.org/10.1080/10691316.2012.695671>
7. Dallis, D. (Ed.), "Planning and Implementing Resource Discovery Tools in Academic Libraries". IGI Global. Retrieved from <http://www.igi-global.com/chapter/evaluating-selecting-library-web-scale/67814>
8. David, B., & Popp, M. P. (Eds.). (2012). "Designing an Evaluation Process for Resource Discovery Tools. Planning and Implementing Resource Discovery Tools in Academic Libraries". IGI Global. Retrieved from <http://www.igi-global.com/chapter/designing-evaluation-process-resource-discovery/67818>
9. Fagan, J. C., Mandernach, M., Nelson, C. S., Paulo, J. R., Saunders, G. (2012). "Usability Test Results for a Discovery Tool in an Academic Library". Information Technology and Libraries. Retrieved From ejournals.bc.edu/ojs/index.php/ital/article/view/1855
10. Fagan, J.C. (2012). "Top 10 Discovery Tool Myths." *Journal of Web Librarianship*, 6(1), pp. 1-4.
11. Graves, T., & Dresselhaus, A. (2012). "One Academic Library: One Year of Web-scale Discovery:." *Serials Librarian (SERIALS LIBR)*.
12. Guthrie, A., and McCoy, R. (2012). "A Glimpse at Discovery Tools within the HBCU Library Landscape." *College & Undergraduate Libraries*, 19(2-4), pp. 297-311.
13. Hoepfner, A. (2012). "The Ins and Outs of Evaluating Web-Scale Discovery Services." *Computers in Libraries*, 32(3), pp. 6-10, 38-40.
14. Hoy, M. B. (2012). "An Introduction to Web Scale Discovery Systems". *Medical Reference Services Quarterly (MED REF SERV Q)*. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/02763869.2012.698186>
15. Kaufmann, K., Larsen, J., and DeSalvo,

- P. (2012). "Discovering the Discovery Tool: The Introduction and Impact on Research and Instruction at Seminole State College of Florida." *College & Undergraduate Libraries*, 19(2-4), pp. 278-296.
16. Leebaw, D., College, C., Conlan, B., College, S. O., Sinkler-Miller, C., College, C., Wilson, N., et al. (2012). "Improving Library Resource Discovery: Exploring the Possibilities of VuFind and Web Scale Discovery in a Consortial Environment". *Library Technology Conference*. Retrieved from http://digitalcommons.macalester.edu/cgi/viewcontent.cgi?article=1263&context=libtech_conf
17. Library Technology Guides. Retrieved from librarytechnology.org.
18. Narayanan, N., & Mukundan, R. (n.d.). Cloud web scale discovery services landscape an overview. Retrieved March 15, 2015, from <http://www.slideshare.net/.../cloud-web-scale-discovery-services-landscape-an->
19. Pitts, J., & University, K. S. (2012). "The Advent of Web Scale Discovery Tools: What it Means for Undergraduate Research". (Presented at *SIDLIT Conference*). Retrieved from http://scholarspace.jccc.edu/c2c_sidlit/2012/Thursday/6
20. Sutradhar, B. & Guha Roy, S . (2014). "Design and Development of e-Search software for providing Discovery services at Central Library, IIT Kharagpur". (Presented at the 9th Technology Book Fair & Workshops. Indian Institute of Technology, Kharagpur)
21. Thomsett-Scott, B., and Reese, P.E. (2012). "Academic Libraries and Discovery Tools: A Survey of the Literature." *College & Undergraduate Libraries*, 19(2-4), pp. 123-143. DOI: 10.1080/10691316.2012.697009.
22. Vaughan, J. (2011). "Web Scale Discovery" What Is Web Scale Discovery? Why Web Scale Discovery? *Library Technology Reports: Expert Guides to Library Systems and Services*, 47(1), pp. 5-11.
23. Vaughan, J. (2011). Web scale Discovery Services. *Library Technology Reports*, 47(1), pp.5-11.
24. Vaughan, J., & University of Nevada, L. V. (2011). "Investigations into library web scale discovery services". Retrieved from http://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=1043&context=lib_articles
25. Way, D. (2010). "The Impact of Web-Scale Discovery on the Use of a Library Collection." *Serials Review*, 36(4), pp. 214-220.
26. Yang, S.Q., and Wagner, K. (2010). "Evaluating and Comparing Discovery Tools: How Close Are We Towards Next Generation Catalog?" *Library Hi Tech*, 28(4), pp.690-709.