

Web-Scale discovery System in Medical Libraries

Lakshana kumar BM

Senior Librarian

Shimoga Institute of Medical Science

Email:lakshanbm@gmail.com

Abstract: In the present day, Libraries are Crossroads of development and Librarians need to adopt new technologies, At present situation Information scientist is searching a tools to solve their user needs. Libraries at web scale explore the impact of the web on our rapidly changing information landscape and presents on overview of the opportunities and challenges that operating in a web connected world provides for libraries and library users, Web scale system is a best method to sharing a document from different libraries throughout the world. It will explain the Web scale Management system, web scale management service, Librarian Challenges in resources sharing through the system, benefits and some live examples Web-scale providers, so on.

Key word: Web scale Management system and services, Librarian Challenges, Resource sharing, Web Scale Discovery Services.

Introduction

We've been hearing a lot about a web-scale" lately. It has become one of the buzzwords of the library technology arena, and it's often used rather loosely. While the term has taken on something of a marketing bent, it also characterizes some important trends and strategies for libraries to capitalize on today's large-scale technology platforms. Though web-scale is necessarily a term that lends itself to precise. Mean while,, it represents an important shift in the way that libraries engage with technology. Web-scale is a concept worth exploring as one of the new alternatives for Information communication technology. For Example,, OCLC will work with libraries that are interested and prepared to implement Web-based services for acquisitions and circulation. This will be followed by successive updates for subscription and license management, and cooperative intelligence—analysis and recommendations based on statistics and workflow evaluation among participating libraries.

OCLC members have made it clear that new, innovative responses are needed to meet these challenges. For the past eight months, OCLC has worked with an Advisory Council and six libraries and library groups as pilots for Web-scale management services. These groups have provided advice to OCLC on an overall direction, offered new ideas that were not in the original development plan, and validated strategic positioning for the service. OCLC Web-scale Management Services offer a next-generation choice for traditional, back-office operations. Moving these functions to the Web alongside cataloging and discovery activities allows libraries to lower the total cost of ownership for management services, automate critical operations, reduce support costs and free resources for high-priority services.

What Is Web-Scale?

As it stands in the library community, the creators of a certain number of discovery products have described their offerings as web-scale. As a result, the term has also been applied to a certain class of library management products. But more than an architecture applied to specific products, web-scale represents a new alternative paradigm for libraries to operate, both in terms of their internal operations and in the way that they provide access to their collections and services for their users by leveraging current internet technologies and concepts.

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 .

Unified platform to search all the resources including licensed, and open and local collections including Library Catalog and Institutional Repository.

- 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
- Connections to full text
- Infrastructure, processing and indexing provided and maintained remotely by the vendor.

Characteristics :

- Large-scale technology platforms
- Applications delivered through multi tenant software as a service
- Massively aggregated approaches to data
- Highly cooperative arrangements among participating libraries

Web-scale things combine cloud computing, highly shared data models, and expansive aggregation of library-related resources.

Web-Scale Enters the Library Vocabulary

January 2007, he wrote a blog posting that noted how the term had already been used by organizations such as Amazon to describe its infinitely scalable infrastructure services such as the Elastic Compute Cloud (EC2) or Simple Storage Service (S3). Dempsey noted, "

'Web-scale' refers to how major web presences architect systems and services to scale as use grows.

OCLC has since taken web-scale as the fundamental concept behind the way that it provides its products and services through its global platform. OCLC's WorldCat, originally created as a cataloging utility, has spawned an expanding suite of services, such as World-Cat Local and, more recently, Web-Scale Management Services, which folds the conceptual label into its product name. OCLC clearly sees itself operating services at "web-scale."

Web-Scale in Product Strategies

OCLC does not hold exclusive claim to the concept of web-scale. Other organizations, such as Serials Solutions, have explicitly adopted the term, initially to its Summon discovery service and, more recently, for its planned offering for library management. While Serials Solutions states that it plans to give the product a name, initially, it describes it as its "web-scale management solution." Even though OCLC and Serials Solutions clearly both embrace these core concepts, we can expect each to have distinctive and different functional approaches in their product offerings. in the partner institutions rather than through a multi tenant software-as-a-service model.

Web-Scale as a Scalable Technology Platform

We can easily recognize that web-scale implies large-scale systems. It carries a connotation of massively large scope, size, or extent. We all understand the enormous expansiveness of the web. Web-scale uses this point of reference as we think about the qualities of services that operate within different domains, such as those related to libraries.

Web-scale also has a connection with cloud computing, though the terms are not synonymous. Any web-scale service would, almost by definition, be deployed using some kind of cloud-computing infrastructure. In practical terms, we see all of the products within the web-scale category offered through software as a service rather than as software for local implementation.

But not only does web-scale mean large, it also implies the ability to expand without constraint. By leveraging cloud-computing technologies such as software as a service and infrastructure as a service, products based on this model will be able to grow in proportion to the libraries that adopt the service, with the expansion of the scope of content managed by libraries,

Cooperative Data Models and Workflow Management

Software as a service relieves the library of the need to expend its limited resources on maintaining server hardware and operating systems. A web-scale system takes the benefit a step farther, layering on data models and workflows that enable new levels of efficiency.

Web-scale computing stands in direct contact with the traditional model of library automation based on a server housed and operated on premises, based on an isolated self-contained database. Web-scale computing offers the potential to bring together the collective efforts of many libraries to create systems more powerful than possible through many separate and independent implementations.

Expansive Scope of Search

One of the main trends in the area of discovery interfaces involves the extension to an ever-wider scope of search. The first round of online catalog replacements, sometimes characterized as next-generation library catalogs, brought more modern interface techniques and more powerful relevancy-based search technologies for access to a library's local collections. Some of these products also integrated metasearch technologies to extend access to at least some of the materials represented in a library's subscriptions to ejournal databases or other scholarly content from external providers.

Web-scale evokes a more expansive view of the scope of search in library discovery products. The online catalogs of the integrated library systems or even the initial wave of next-generation catalogs address content managed directly by the local library. A new genre of web-scale discovery solutions attempt to provide instant access to the broadest view of library content, including not only local collections but also the vast amount of materials represented in externally provided resources. A comprehensive view of library collections today consists of many components: physical print and media collections

erspective Among Competing Alternatives

While web-scale products may be an interesting and growing segment of the library automation scene, we're at a very early stage of the development, marketing, and adoption cycles. Discovery products based on this model have gained quite a bit of traction, with a steady expansion of their underlying indexes, broader participation by the publishers and other content providers, and maturing software platforms. The competition among web-scale management products has just begun. We can expect things to heat up considerably in this arena over the next year or so.

The emergence of these so-called web-scale products does not necessarily entail a widespread or rapid decline of products based on localized computing. Not only do the cycles of change turn slowly in the library technology arena, but many libraries will continue to prefer the control and other characteristics of locally implemented automation systems. Some libraries may not have the same attraction to web-scale products as others based on a variety of factors

Web Scale Library Management Systems

In web scale system following steps are important

Current Pain Points for Libraries

- Assessment: need better data analysis tools for decision making. Need to prove library contribution to organization's mission.
- System maintenance: costly and time consuming. Ongoing hardware and software costs. IT expertise tending to ILS.
- Workflows: split between electronic and print. Silos for print, digital, electronic materials resulting in duplicate data and duplicate workflows. Split among jobbers. Metadata from multiple vendors. No tool to track requests. No interoperability.

Advantages of Web-Scale Library Management Systems

- Reduces overall costs. Eliminates the need to locally host multiple servers and equipment. Eliminates constant hardware and software upgrades. System changes do not require local rewrites.
- Enables library staff to focus on serving users, and enables library IT staff to redirect focus to innovating at the local level.
- If done right, simplifies processes
- Better support. Eliminates the need for institutional IT support for the system.
- Data as a Service (DaaS). Depending on the architecture of the cloud computing solution, if data is part of the hosted offering and it's enabled for sharing, data becomes a service rather than a function replicated library by library.
- Better analytics. Greater ability to utilize analytic tools with shared data, analyze user trends and potential click streams.

Disadvantages or Concerns of Web-Scale Library Management Systems

- Security of data
- Loss of software configuration options which enable library to meet unique, local needs.
- Branding. End users do not know where the information appearing on the screens is originating.
- Loss of control. Library must maintain sufficient control to support "value-add" of their local collections and services.

Web-Scale Library Management System Marketplace

OCLC WMS (WorldShare Management Services)

- Open and extensible platform built on an extended view of WorldCat. Designed from the ground up. Core infrastructure underlying all OCLC applications. Shareable applications gallery. Some built by OCLC; others by customers. Libraries can build applications to meet local needs, while benefiting from the innovation of the broader global library community.

- Status: 150 committed academic, public and special libraries; of those, 40 are in full production
- Current features
 - Acquisitions for physical and electronic collections
 - Circulation and patron management
 - Discovery via WorldCat Local
 - Cataloging tools for data creation and enrichment
 - Service Configuration manger
 - OCLC Worldshare License manager
 - OCLC WorldShare platform (data, tools, and services that developers can use to create and share custom applications)

Serials Solutions Intota

- Shared data model utilizing KnowledgeWorks knowledgebase. Authority control to be done by Serials Solutions via Knowledgebase eliminating need for local authority control management. Designed for Summon, but will allow for use with other discovery systems. Supports multiple formats (MARC, Dublin Core, MODS, etc.). Shared customer union database; local holdings to be attached programmatically based on profile. 360 ERM, Counter, some parts of Core, MARC Update will be rolled into Intota. Knowledgebase, 360 Core and Summon will not be a part of Intota, but will work well with it.
- Status: in design and development
 - Phase 1. Data Management – 2012
 - Data Management (Resource Manager, Cataloging)
 - Overlap analysis
 - Acquisitions (print & electronic)
 - Phase 2. Selection – 2012/2013
 - Selection
 - External system integration
 - Phase 3. Fulfillment – 2013 (Fulfillment = Circulation.)
 - Full circulation
 - Network authority control
 - Data harvesting by campus warehousing services

Next Steps

- Assess costs expended in supporting local SirsiDynix platform - servers, sys admin support, software, development support, maintenance, etc.
- Explore options with others in PSUL (Rebecca Macintosh, Ann Snowman, Bob Alan, and Dace Freivalds to meet in mid-April as informal ‘web scale’ study group to investigate the web scale library management system landscape and discuss how to move forward.)
- Educate Libraries faculty and staff about options for moving beyond SirsiDynix ILS
- Benchmark with peers in CIC and ARL.
- Define functional and system requirements. Determine PSUL library system needs that are not being met currently that must be met in a new system. Break into components (selection, acquisition, cataloging, discovery, fulfillment inventory,

circulation) and determine requirements for each across all resource types (print, electronic, digital).

Assess pros and cons of web scale vs open source library management systems, and determine appropriate direction for PSUL (web-scale or open source).

Choice of Discovery: Web Scale Discovery Services

Web Scale Discovery Services are becoming the most sought after solution for Libraries to Connect its patrons with the relevant information they seek, This state of the art technology Solution holds the potential to be the evolution that Libraries have long sought for information discovery¹. Web Scale Discovery solution is getting wide acceptance from library community which is evident from the fact that many Libraries are replacing their Federated search Solutions of the past decade and adopting this state of the art technology.

1.0 What is Web Scale Discovery ?

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 ². 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, and open and local collections including Library Catalog and Institutional Repository.
- 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
- Connections to full text
- Infrastructure, processing and indexing provided and maintained remotely by the vendor.

Web Scale Discovery Vs Federated Search

Federated Search Solutions are the technology Utilized by Libraries for resources retrieval in the past decade. The inherent weaknesses of federated solutions are its slowness and associated problems as the searches are conducted on the fly against each and every

databases, Federated system needs to have separate connectors for searching different databases. There are some distinctive features which delineate the reward of Discovery over Federation.

Discovery Service	Federated Search Engines
Search is very fast as retrieval is done in pre-harvested index	Slow (Longer time for search completion) Federated searching is performing the meta search on the fly from different resources
Standardized unified index	Many indexes : Individual indexes and different database structures of various publishers makes it difficult for metadata retrieval.
Robust Relevancy Ranking as retrieval is from Unified index	Relevancy ranking is a major issue in retrieval of quality data.
Enhancement is possible on the harvested Metadata	Metadata enhancement is not possible
Comprehensive results	Shallow results in many cases and eventually users will miss much relevant content.
Performance quality is very high	Many times important information from relevant resources are missed out due to connection error.
As employing harvesting method, Separate connectors are not needed for each and every publisher. Even small publishers data can be added in central index.	Many small publishes and societies with one or two journals normally, do not have federated search connectors and eventually out of the scope of federated searching.

Users approaches in federated search technology :

Search for information in multiple information resources through a single query. View search results in a single list. In short, an Information Portal uses federated search technology to enable users to:

- Search for information in multiple information resources through a single query.
- View search results in a single list.
- Link directly to each resource to expand the search.

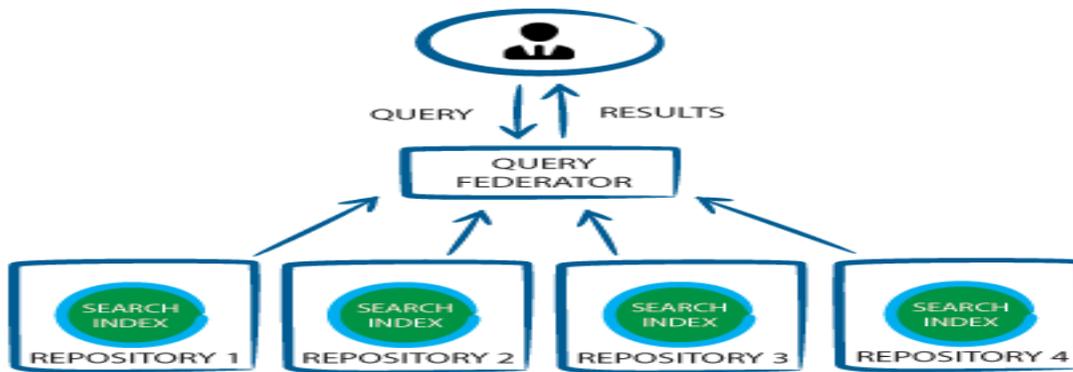
There are two distinct approaches to federated search, which can be labeled as *index-time merging* and *query-time merging*. The pros and cons of these are outlined below.

SEARCH-TIME MERGING

In most circumstances, this is the faster and easier solution to implement.

A *query federator* intercepts the query, and passes it to multiple search engines. The federator then waits to hear replies from the search engines, and when received, merges or concatenates the results into a results list.

This model relies on data repositories to provide a search function.



PROS:

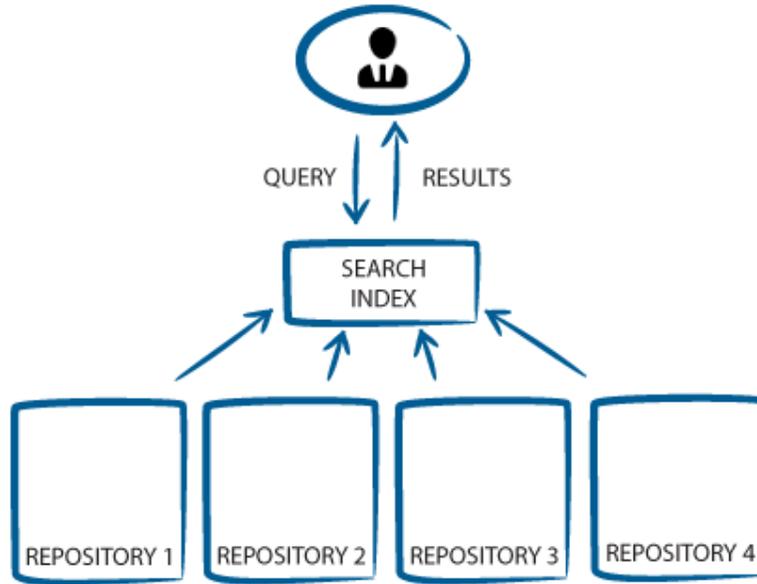
- The primary advantage of this approach is ease of implementation, because no additional indexing of content is necessary. The query federation system simply taps into existing systems and extracts results, which are then merged.
- In some cases, query-based federation is the only viable option. For example:
 - Federating to large-scale Web content via a major search engine such as Google
 - Federating to a private data set, held behind a pay-wall and therefore not available to be indexed locally

CONS:

- Performance issues can occur if the federator waits for the slowest remote search engine to respond
- The merging of search results into a sensible hit list is difficult if based on relevancy, as each search engine called will score relevancy in a different way. Often, it is better not to attempt merge on relevancy but instead; either present separate results lists (behind tabs for example) or use a more deterministic data item to merge on, such as date, location or price, or present results from different sources in blocks
- Search engines provide varying levels of query sophistication. Federation at query time usually implies a "dumping down" to suit the least capable search engine, however, this need not always be the case.
- Document level security is a potential cause of performance issues, but this depends on the complexity of the security environment

INDEX-TIME MERGING

This approach requires content to be acquired into a central index, and it is typical of traditional enterprise search systems.



PROS:

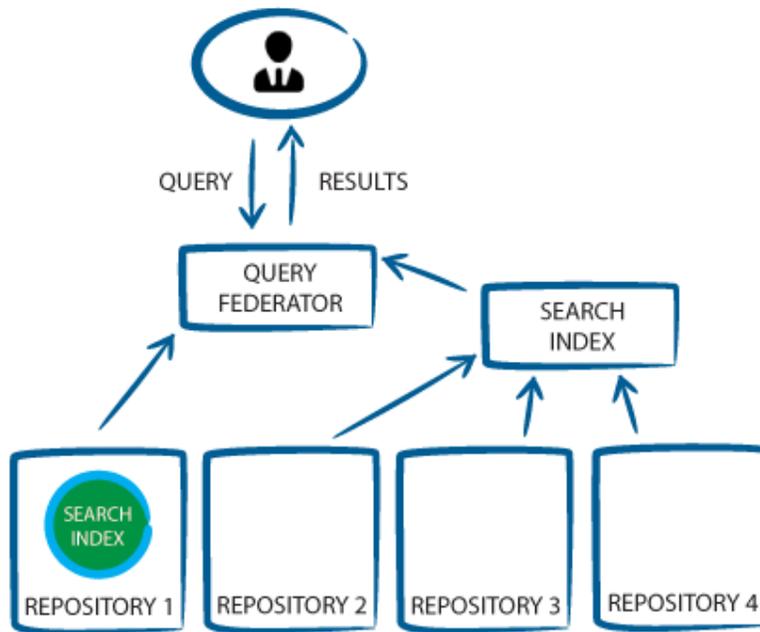
- Most search engines default to ranking by relevancy, which is what most users expect. Through acquiring all data into a central index, sophisticated query enhancement and relevancy algorithms can be applied, providing the user with excellent search results.

CONS:

- The effort needed to acquire the content from the various repositories can be substantial. This is done via read-only processes. The content of remote repositories is not moved or changed, but the indexing process must read each item, and re-read it every time a change occurs. In some cases, for example where private content behind pay wall is involved, this is not possible

HYBRID FEDERATED SEARCH

Sometimes, the optimum solution is a hybrid approach. Where practical, content is indexed centrally. Repositories for which that are not cost effective (or simply not possible) are federated to at query time. If this approach is used, careful thought is needed about results presentation, to make sure that users understand how the system is set up, and how to navigate and interpret results efficiently.



WHICH APPROACH WORKS BEST?

The approach that works best all depends on your data environment and your user needs. Start by looking at the data environment, user requirements and business drivers, then informed decisions can be taken. In our engagements, this process usually begins with a Search Assessment.

Many academic and research libraries are making significant investments in the relatively new and still imperfectly understood Web-scale discovery systems—the four leading ones being EBSCO Discovery Service (EDS) (2010), Primo from Ex Libris (2010), Summon from Serials Solutions (2009), and WorldCat Local from OCLC (2007).

There is great hope that these rapidly maturing products will not only promote information literacy strategies but also deliver what meta search (or federated search) has failed to achieve—a Google-like interface that provides a fast, single point of entry to an institution’s relevant and vetted scholarly content.

2.0 Web Scale Discovery Architecture

Web Scale Discovery service constitutes two important components Content or resources coverage is the prime factor and the second factor is appropriate technologies to make available the relevant information to the library users from available content. This include technologies that facilitate to harvest, index, relevance ranking search and retrieve the content and user interface platform features to provide a user friendly environment to users.

3.0 **Content**:-Normally, a Web Scale Discovery system covers all informative contents that scholarly users are interested Web Scale discovery services are able to index variety of

content, whether hosted locally or remotely. Such content can include library ILS records, digital collections, institutional repository content, and content from locally developed and hosted databases.

4.0 Technology

Web Scale Discovery Systems make use of mash ups of many technologies and tools to harvest, index, store, search, and retrieve the content in response to user queries through a unified web interface.

- Harvester
- Metadata mapping
- Platform Blending of subject indexes
- Central index
- Link Resolvers
- Relevancy Algorithms
- Interface

Conclusion

Web Scale services are fairly new entrant in to the Library and Information Retrieval and still in its initial stages of development with regard to its features, functionality, level of integration with other systems, interoperability, scope of content, soundness of metadata, flexibility of the interface etc. Different Discovery Services vary in its offerings. The best way to evaluate and experience WSD system is to request fully customized trial including Catalogue and Institutional repository integration from the Discovery Vendors. The final choice always depends on the concerned Library's Preferences.

References :-

1 . Breeding Marshall (January 2012) Library Technology Guides Library Web Scale.

<https://www.google.co.in/webhp?sourceid>

[Library%20Technology%20Guides%20Library%20web%20scale](#)

2 . **Overview of Web Scale Library Management System 2012**

<https://www.libraries.psu.edu/psul/groups/webscale/overview.html>

3. OCLC's Web Scale Library Management Services, Copyright © 2011,

<https://oclc.org/content/dam/oclc/reports/worldshare-management-services/libraries-at-webscale.pdf>

4. Choice of Discovery : Web Scale Discovery Services (2011)

<https://www.google.co.in/choice%20of%20discovery%20evaluating%20web%20scale%20Discovery%20services>

5. Vaughan, J. (2011) Web scale Discovery Services, *Library Technology Reports*, 47(1), 5-11.

6. Vaughan, J. & University of Nevada, L.V. (2011). Investigations into library web scale discovery services. Retrieved from