Searching Online Resources in New Discovery Environment: A State-of-the-Art Review

Dinesh R Pradhan   **Kruti Trivedi**   **Jagdish Arora**

Abstract

The universe of libraries has evolved and expanded from repository of books, journals and other non-print materials in physical format to knowledge repositories that contain collections in physical as well as electronic and multimedia format. With this changing environment of collection of library resources, the greatest challenge in front of libraries is to find out the best way to present, promote and make accessible the growing collections of heterogeneous resources. Libraries are required to look for search and discovering tools to provide seamless access to their resources. This article introduces basic philosophy of federated search and discovery tools. It elaborates on technological tools, techniques and protocols that are used for developing federated search and discovery tools. The article narrates disadvantages of federated search solutions and discovery tools as compared to conducting a search directly on the interface of primary database. The Concept of open URL as a technology used for link-resolving and finding "appropriate copy" is discussed briefly. Lastly, the article describes contemporary federated search solutions and discovery tools that are available commercially or as open source solution with their merits, demerits and comparative performance.

Keywords: Resource Discovery, Federated Searching, Web-scale Discovery, OpenURL

1.   Introduction

The universe of libraries has evolved and expanded from repository of books, journals and other non-print materials in physical format to knowledge repositories that contain collections in physical as well as electronic and multimedia format with search and browse Web-based interfaces enabling users to interact with its diverse collections. With remarkable increase in availability of content in electronic format, libraries are increasingly committing their funds for purchase and subscription to full-text e-journal and e-journal databases, bibliographic databases, e-books, etc. Moreover, libraries are also developing their own electronic content through institutional repositories consisting of e-prints of research articles, manuscripts, electronic theses and dissertations, institutional publications, digital images, photographs and images and artifacts of cultural heritage. Broadly, the collection in a library may include items that the library acquires and manages, items that are managed by third parties and are available for access from remote locations, items that are held locally, items that are subscribed or are available for a fee and items that are free at their point of use[1]. One of the major challenges that the modern libraries faces is to find ways and means to select and offer seamless and coherent access to growing collections of heterogeneous resources to the users in a fashion that enriches learning and research experience, provides timely and convenient access to relevant and appropriate resources, exposes potentially valuable resources that otherwise might have been overlooked and enables users and the library to focus on a fruitful use of collections rather than dealing with different aspects of access, navigation and manipulation of the result sets.[2]

With growing number and variety of electronic resources available in the market, libraries are required to look for search and discovery tools that provide seamless and coherent access to all resources that a library subscribes to as well as those that can be accessed through open access channels. This article introduces basic philosophy of federated search and discovery tools. It elaborates on technological tools, techniques and protocols that are used for developing federated search and discovery tools. The article narrates disadvantages of federated search solutions and discovery tools as compared to conducting a search directly on the interface of primary database. The Concept of open URL as a technology used for link-resolving and finding "appropriate copy" is discussed briefly. Lastly, the article describes contemporary federated search solutions and discovery tools that are available commercially or as open source solution with their merits, demerits and comparative performance.

2.   Resource Discovery: Towards a Definition

The Meaning of "Discovery" as per Dictionary.com is to see, get knowledge of, learn of, find, or find out; gain sight or knowledge of (something previously unseen or unknown). According to Eddie Clarke, resource discovery in libraries means the process of identifying and accessing the information relevant to the user.[3] As defined by Andrew Nagy, a "Resource Discovery" system implies the discovery of resources that might be unknown or new to the user.[4] It may be noted that the scope of the library collection has been widened from resources available in the library to the resources available on web. As such, discovery tools available in conventional environment have to be reoriented and repurposed not only to include conventional content but also the contents that are available on the web.

The following three approaches are being deployed for developing federated search solutions and discovery tools:

- ♦ Federated search, wherein is a search query is sent to the multiple databases selected by the users, search is conducted simultaneously on multiple databases and search results are displayed. The users gets an experience of single gateway to all online resources subscribed by an institute;

- ♦ F   Web search engine approach, wherein discovery services maintain a database of structured and standardized metadata obtained directly from the publishers with link to respective full-text databases. Subscribers to discovery services search through pre-harvested metadata. This approach provides faster access with multiple search, retrieval and display options.

3.   Advantages of Federated Search

There are certain advantages of using Federated Searches. Federated search is affordable, proven discovery technology. In a number of situations, federated search has been linked with the discovery layer to provide a more unified user experience. Federated search tools present the user with a single, simple interface, often very similar to what a user would find in Web search engines eliminating

the visit to various databases with varied interfaces. The federated search tools give an option for the libraries to group various resources to provide a customized subject specific search to the users. It redirects the users directly to the native source searched for accessing the resource.

4.   Disadvantages of Federated Search

The biggest constraint of conventional real-time federated search engines is the time it takes to fetch search results from multiple resources it searches in comparison to search engines like Google and Yahoo. Since federated search engines send the search query to different databases and fetch results in a real-time environment, the speed with which search results are delivered depends upon the response time taken by multiple databases being searched. At times, a few databases from amongst the databases being searched are too slow to respond, as such the user get timeout message or zero hits for those databases causing frustration in users. As the federated search solutions are mostly provided on-campus networks, for off-campus users the searching becomes much slower as the users have to pass through authentication procedures of the library. Besides, with increase in number of subscribed databases in bigger institutions, time taken in executing a search through federated search interfaces increases dramatically. In fact, a library that has access to 100 databases available to participate in a federated search, in reality, usually chooses to group them into smaller subject-wise subsets. These smaller subsets serve to fragment the collection and lead to search box proliferation."[5]

Moreover, federated search solution may also cause information overload including irrelevant or marginally relevant information. Simultaneous searching through large number of databases often loads the users giving thousands of results with the relevant information buried in randomly scattered topics.[6] The users have to spend a lot of time and effort to browse and analyze through the results and finding the results relevant to their research. This process leads to a negative impact on the quality of research. It also prevents discovery of new content and relationships that are relevant to users' research.[7]

The other limitations of federated search include the coverage of the resources it searches, inability of inclusion of some of the databases, such as SciFinder Scholar, that require specialized interface. Besides, there are issues with the ability of these interfaces to merge, de-duplicate and rank retrieved search results appropriately.

5.   Technology behind Federated Search and Discovery Services

With the multiple numbers of open access and subscribed electronic resources available to users in research and education organizations, searching for relevant literature becomes a challenging task for a user. A user has to visit a number of individual databases, understand different types of user interface and retrieve results. If a users gets sufficient number of results from first few preferred databases, he / she tends to ignore remaining databases at the risk of missing out on important research work published in them. The new-generation web-savvy users prefer single search box approach offered by search engines like Google, Yahoo, Bing, etc that serve as a single-window

access interface for all web-based resources. Design and approach of federated search and discovery tools are essentially inspired by the Web search engines. The federated search solutions facilitate users to search multiple full-text and bibliographic databases subscribed by the library as well as open access e-resources, library OPAC, IRs, and other library resources through a single search box. Important search and retrieval protocols that make the foundation of federated search or meta-search solutions are as follows:

i) **Z39.50**: Z39.50 is an ANSI / NISO standard for information storage and retrieval. It is a protocol which specifies data structures and interchange rules that allow a client machine to search databases on a server machine and retrieve records that are identified as a result of such a search. Z39.50 protocol is used for searching and retrieving bibliographic records across more than one library system. This protocol is not used by the Internet search engines. It is more complex and more comprehensive and powerful than searching through http. Z39.50 has been extended to allow system feedback and inter-system dialogue. Like most applications working under client-server environment, Z39.50 needs a Z39.50 client program on one end, and a Z39.50 server program on the other end.

The name Z39 came from the ANSI committee on libraries, publishing and information services which was named Z39. NISO standards are numbered sequentially and Z39 is the 50th standard developed by the NISO. The current version of Z39.50 was adopted in 1995 superseding earlier versions adopted in 1992 and 1988.

ii) **SRU/SRW**: Search and Retrieval via URL (SRU) and Search and Retrieval Web Service (SRW) are Web Services-based protocols for querying Internet indexes or databases and returning search results. The web services are two types, i.e, REST (Representational State Transfer) and SOAP (Simple Object Access Protocol). The SRW uses SOAP protocol and the SRW uses REST protocol for information retrieval. [8]

iii) **NISO Metasearch XML Gateway (MXG):** MXG is proposed as an alternate to Z39.50 protocol and is based on the SRU protocol. The NISO MXG is a low-barrier-to-entry method for content providers to expose their content to metasearch application.[9]

iv) **Proprietary XML Gateways:** Some database vendors provide proprietary XML gateways to search through their database. These gateways vary from vendor to vendor.

v) **HTML Parsing** or **Screen Scraping**: HTML parsing or screen scraping is similar to the process of searching the database by humans by filling out search forms and process the search result (summaries of documents generated by the remote search engines). Screen scraping is the most 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.[10]

vi) **OpenSearch:** It is a way for websites and search engines to publish search results in a standard and accessible format suitable for syndication and aggregation. The OpenSearch is

built on XML and supports a mechanism for telling a deep web search engine how to query it and the search results data are retrieved in a highly structured format. So the search results are easy to process and display by a federated search service.[11]

5.1 Reference Linking: DOI and OpenURL

The availability of web-based resources has opened-up yet another route for discovering resources by linking references given at the end of article to their full-text. The web being a hypermedia-based system, allows linking amongst electronic resources stored on servers geographically dispersed on distant locations. For enabling the linking facilities the early steps include the adoption for DOI systems, an extension of handle system for publishers, that is used for providing unique identity to each article or document available on the web. DOI and Crossref system has brought the publishers, aggregators and libraries to a single platform for linking between various resources.

5.1.1 Digital Object Identifier (DOI)

Unique identifier, called Digital Object Identifier (DOI), is being used to support links to full-text or to bibliographic databases. The Digital Object Identifier (DOI) (http://www.doi.org/) is a system for interoperability identifying and exchanging intellectual property in the digital environment. It provides framework for linking customers with content suppliers, facilitating electronic commerce and enabling automated copyright management for all types of media. The DOI was developed to ensure precise identification of content and thereby enable precise management of rights. More than 200 publishers have adopted the technology behind DOI. One of the initiatives to support and promote concept of DOI is CrossRef.org. The CrossRef.org (http://www.crossref.org/) is a collaborative reference linking service using DOI and associated metadata for linking citations. Crossref has more than 3200 publishers and societies that participate accounting for over 23,000 journals with more than 44 million content items, including journal articles, conference proceedings, books, book chapters, reference entries, technical reports, standards, and data sets in the database. It functions as a sort of digital switchboard. While the site does not hold full text content, but rather effects linkages through Digital Object Identifiers (DOI), which are tagged to article metadata supplied by the participating publishers. A researcher clicking on a link will be connected to a page on the publisher's Web site showing a full bibliographical citation of the article, and, in most cases, the abstract as well. The reader can then access the full-text article through the appropriate mechanism. While subscribers will generally go straight to the text, others will receive information on access via subscription, document delivery, or pay-per-view.

The DOI registry contains the information about the provider of the article and it links the users directly to the URL of the article even if the publisher of the content changes. However, with libraries subscribing to resources from primary publishers and aggregators, articles are available to the library users from multiple sources, i.e, directly from the primary publisher, through aggregators and through library holding in print form library. The availability of copies of articles from multiple sources has led to the problem of appropriate copy in the library subscribing to multiple databases. The OpenURL technology evolved as a solution to answer the problem of appropriate copy.

5.1.2   OpenURL and Linkresolver

OpenURL (Wikipedia, 2006) is a versatile linking scheme that uses metadata (instead of an object identifier such as DOI) for generating dynamic link by passing metadata about a resource to a resolver program. It consists of two components, i.e. the URL of OpenURL resolver followed by a description of the information object consisting of a set of metadata elements (e.g. author, journal issue no., volume, year, etc.).

For OpenURL to work, a library is required to setup a resolution server with information on full-text journals accessible to the library with their link as well as how to link to local print holdings and other local services. The information provider (or publisher) must also be OpenURL-enabled to redirect the linking request to the local resolution server. A "link resolver" or "link-server", parses the elements of an OpenURL and provides links to appropriate services as identified by a library. OpenURL link allows access to multiple information services from multiple resources, including full-text repositories, abstracting, indexing, and citation databases, online library catalogues, document delivery service and other web resources and services.

When a user clicks at an OpenURL link, he / she is directed to OpenURL resolver. The resolver, based on the services availed by the library provides him an HTML page consisting of a sets links to resources from where user can access the resource (full-text from publisher's site, DDS, Aggregators, etc). The user selects an appropriate service, clicks on the link that takes him to the site of service provider. OpenURL was developed by Herbert van de Sompel, a librarian at the University of Ghent. His link-server software, SFX, was purchased by the library automation company Ex Libris which popularized OpenURL in the information industry. Many other companies now market link server systems, including Openly Informatics (1Cate -acquired by OCLC in 2006), ExLibris (SFX), Endeavor Information Systems, Inc. (Discovery: Resolver), SerialsSolutions (ArticleLinker), Innovative Interfaces, Inc. (WebBridge), EBSCO (LinkSource), Ovid (LinkSolver), SirsiDynix (Resolver), Fretwell-Downing (OL2), TDNet (TOUR Resolver), Bowker (Ulrichs Resource Linker), SFU(Godot), and KINS (K-Link). Besides, several libraries have implemented home grown link resolvers.

The National Information Standards Organization (NISO) has developed OpenURL and its data container (the ContextObject) as international ANSI standard Z39.88.

5.2   Web 2.0 & Next Generation Library Catalogue for better Discovery

The new generation users think, process, and manage information differently from their predecessors, all leading to changed (and higher) needs and expectations. They are no longer treated as patrons, and behaving as customers and taking active role in choosing their information providers. As customers, they:

♦   expect and want more personalization and instant gratification;

♦   are collaborative and multitask;

♦   learn experimentally through trial and error rather than by formal learning or reading;

- ◆ prefer non-linear access to information;

- ◆ respond better to graphic than text; and

- ◆ expect highly intuitive interfaces and convenience[12]

As such the traditional OPACs do not meet the user requirement in better discovery of the library resources which gave birth to the next generation catalogues which provides faceted browsing of the resources and integration of web 2.0 tools. The facets are a good tool in refining the large no search results to narrow down to specific interest of the users. This also ensures the users that there is no null result for the user.[13] A discovery tool is defined as a third party component whose purpose is to "provide search and discovery functionality and may include features such as relevance ranking, spell checking, tagging, enhanced content, search facets"[14] The next generation catalogues (NGC) are meant in view of integrating the above features. Several open-source initiatives and also commercial initiatives are there for providing these new interfaces for discovery of resources. Few of such products developed in the open source domain are LibraryFind (Oregon State University Libraries), Vufind (Villanova University), Scriblio, Blacklight, SOPAC (Social OPAC), Summa, Fac-Back OPAC, Rapi, eXtensible Catalogue Project etc. Many proprietary software have also come up to provide such discovery feature to their ILS customers like, SirsiDynix's Enterprise, Ex Libris's Primo, Innovative Interfaces' Encore, VTLS' Visualizer, Aquabrowser, Worldcat Local, Summon etc.

Yang & Wanger (2010)[15] have made a comparison of some of the above discovery tools. According to the comparison, LibraryFind tops among the discovery tools and the only product which enables discovery of articles as well as other resources. It is followed by Vufind and Scriblio both with second position, and SOPAC & Blacklight, Fac-Back OPAC, Rapi followed. Among the proprietary products, only Summon is having a single interface covering articles.

A resource discovery product is not complete without the integration of federated search results with the catalogue records and other library materials like repositories, theses etc.

6.   Web Scale Discovery (Integration of Federated search & Next Generation Library Catalogue)

The biggest advantage of Search engines over the federated search is the speed with which it retrieves the results. The reason behind the first retrieval is that the search engines search the indexes instead of the databases. To overcome this alternate approach for federated searching is the harvesting service or maintaining a centralized index of relevant sources of information with a normalized metadata schema. Some major standards for harvesting of metadata are as:

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides a mechanism for repository interoperability. Using OAI-PMH, the service providers can make request to the repositories to harvest metadata. This is mostly used by the repositories and some e-journals providers also expose their content using it. [16]

METS – similar to OAI-PMH in purpose and function, METS supports XML-encoded metadata harvesting, but unlike OAI-PMH, METS can harvest both metadata and object.[17]

With the advent of the indexing engines like Lucene etc, another revolution is coming in the field of library resource discovery. In the current era where most of contents are born digital, it is possible to create index of the full-text journals in these indexers to make them searchable.[18] With these indexers it is possible to create index of the library catalogues and integrate the index of openly accessible contents like open access journal contents, open access digital books, openly available contents in institutional repositories etc. The searching from indexes retrieves faster results and the ranking of the search results based on the relevance ranking algorithm which determines the relevance ranking based on the occurrences of search term is an advantage to the users for resource discovery.

These indexes gave birth to the Web-scale discovery tools which could give the libraries an option to attract the users to the library once again, who are more acquainted with Google Scholar for discovery of scholarly content. But Google Scholar only helps to find out the content and the ultimate goal of the user, the access of the valuable content is provided by the library and for this they again have to come back to the library webpage. The Web-scale discovery tools will provide a one stop solution for discovery of the resources that the users can access whether as part of the library subscription or available as open content on the web with a speed and ease similar to Google Scholar.

For providing such discovery, the libraries contents e.g, the articles, book chapters for online books, and metadata of the library catalogues are to be indexed. This need the metadata to be collected from the publishers for authenticity of the content and the content is to be updated frequently for providing the discovery of most current content to the users. In short, Web Scale Discovery can be termed as the "Googlisation" of the library's collections.[19]

The foremost product in Web Scale Discovery is the Worldcat Local followed by Summon from Serials solution and later on other competitors in library software & services like, EBSCO Discovery Service, Primo Central from ExLibris have launched their products.

6.1   WorldCat Local

WorldCat Local is most easily thought of as a localized version of WorldCat.org. The Library Technology reports defines as the WorldCat Local as the first product in the field of Web-scale discovery products. In 2009, it entered into partnerships with various providers to include substantially greater amounts of article-level content into its centralized index.[20] The service is provided in two versions, Worldcat Local and WorldCat Local "quick start". For the libraries with subscription to WorldCat Catalogue and FirstSearch, the service is provided free of costs.

It provides single-search access to 1,000 databases and collections used in libraries today. The central index contains more than 400 million items from hundreds of content sources, including over 200 million article-level records. As on Dec 2010, more than 120 million article-level records from OAIster, JSTOR, ArticleFirst, PapersFirst, ProceedingsFirst, Medline, ERIC, British Library Inside Serials

and Elsevier indexed in WorldCat. The WorldCat Local covers 164 million books, 14 million Theses/ Dissertations, 4.5 million eBooks, 1,000 databases, 200 million articles from 50,000 periodicals and 25 million items from IRs. [21]

The WorldCat Local does not require authentication for searching through the database and the local libraries have to configure the authentication level of the users, whether to authenticate at the point of search or at the time of retrieval of full-text.

The WorldCat Local interface is presently provided in six languages and gives options to the users to refinement of search results based on facets, export of records into various citation formats, RSS feeds for searches, spelling suggestions etc. It gives social networking features like sharing search results on facebook, Delicious and many other social networking websites. It also provides option to for visibility of books in Google Books search.

6.2   Serial Solutions Summon

The Summon™ by Serial Solutions, which is one of the early entrants in the library web-scale discovery services, with its general release in the mid 2009. It is offered as a hosted software-as-a-service solution providing the Summon service and index. As on end of Dec 2010, the Summon includes content from from 6,800+ publishers and 94,000+ journal and periodical titles, with over 500 million items indexed in the centralized index.[22] The items in Summon include library catalog records, e-journal articles, databases, newspaper articles, e-books, dissertations, institutional repositories, conference proceedings, grey literature, cited references, reports, digital library, etc. It has made agreements with major content providers and aggregators like ProQuest, LexisNexis Academic and Gale. The academic publishers include Springer, IEEE, Emerald, Ingentaconnect, Sage, Taylor & Francis and ScienceDirect freedom collection and 90% of Scopus. It also includes content from Web of Science and ABC-CLIO. The index also covers content from many open access repositories which include DOAJ, Hindawi Publishing, arXiv.org e-Prints, and the HathiTrust materials. [23]

The index is updated through an automated process and the content providers provide the content on a regular basis varied on the content publication, like daily for a newspaper and on monthly basis for a monthly journal.

One of the interesting things is that it is open to all on the web and does not require any user authentication for searching as in case of most federated search solutions. It works with the library's link resolver to guide the user to the relevant content subscribed by the library. By default the users get search results for the content accessible in the library which includes the subscribed content of the library and the local holding available in the catalogue of the library. The search results can also be expanded to include all the content from the Summon index.

The search results are sorted by relevance ranking based on various parameters like peer-reviewed journal articles, citation count of the articles from Web of Science, current-ness of the article and relevance ranking algorithms from different index fields.

Since Summon is built on Web-based open API, it provides lots of flexibility in customization of the interface to give look and feel of the library website. Like the most discovery layer products it also provides faceted navigation for search refinement, spelling suggestions, RSS feeds for searches, export of search results to various bibliographic tools and customized interface for mobile devices.

6.3   EBSCO Discovery Services

EBSCO Discovery Service™ (EDS) launched in early 2010 provides web-scale discovery through creating a unified, customized index of an institution's information resources, and an easy, yet powerful means of accessing all of that content from a single search box.[24] It is offered as a hosted service and required user authentication for searching the index. Later it released a guest mode with limited search capabilities for unauthorized users. The EDS content include metadata from Journal & magazines, index of all the EBSCO host databases that the library subscribes to, other databases that library subscribes like Alexander Street Press, LexisNexis, Cambridge Unviersity Press, IEEE, Ingenta, Springer, Elsevier, Taylor & Francis, Wiley, NewsBank, Readex etc. It also indexes institutional archives/repositories, catalogue records of the library and other additional contents like book jacket images, book records, entertainment records, annotations, family keys, subject headings, demand information, awards, review citations, etc. It also indexes content from various open access repositories like DOAJ, OAISTER, and arXiv.org. As on Dec 2010, the EDS Base Index contains nearly 50,000 Magazines & Journals, 6 million Books, 20,000 Conference Proceedings, 825,000 CDs & DVDs, and Hundreds of thousands of additional information sources from various source-types which include Biographies, Health Reports, and Newspapers etc.[25]

The EDS platform users the institute's link resolver for resolution to the full-text of the articles that library subscribes. The search results are ranked by relevancy based on the frequency of term in document, which field the term appears in and the uniqueness of a word in the overall index. The relevancy is also determined based on the currency, number of times cited, type of document etc.

The EDS provides an administrative interface for customization of the interface as per the library requirement, like layout, colour, logo of the library etc. The interface also provides faceted navigation, search refinement, export to various citation formats, printing, e-mailing and adding to cart, RSS feeds, spelling suggestions, option for adding widgets and also a mobile interface for searching on mobile devices. [26]

6.4   ExLibris Primo Central

The ExLibris Primo which is its next generation discovery layer product by Ex Libris provides a common platform to search the local library collections, such as bibliographic records, digital collection materials, and items within institutional repositories by harvesting the contents into a single index. It also provides an option to configure for searching remote resources and provide results blended with local library collection. In mid 2010, Ex Libris released its Primo Central Index which provides a pre-harvested index of hundreds of millions of journal articles, e-books, reviews, legal documents etc harvested from primary and secondary publishers and aggregators, and from open-access repositories.[27]

The Primo Central Index is just an extension of the Primo discovery layer and the discovery layer is offered as a hosted service as well as local installation and the Central index is hosted and maintained by the Ex Libris Group in a cloud environment.

The search results in Primo are listed by relevance which calculated based on Primo's proprietary relevancy- ranking algorithm that includes but is not limited to factors such as term frequency, field weighting, number of times a record has been accessed, and currency. It also considers peer-review status of journal articles for giving relevance.

The Primo hosted service provides customization of the user interface with change in colour, library branding etc. For libraries with appropriate staff & skill can completely redesign it to give localized interface using the APIs and Web services layer of Primo Central. The search results provide option for refinement by facets, did you mean suggestions, exporting of records to various bibliographic management tools, tagging and commenting the records and RSS feed for searches and a mobile interface for mobile users. It also provides option for sharing items on sites like delicious, Connotea etc. An optional scholarly recommender service, the bX Recommender service can also be integrated into it to provide recommendation of related items —"Users interested in this article also expressed an interest in the following" generated from the analysis of extensive SFX link resolver usage logs. The results provides additional information like Details, Reviews and Tags, Additional Services, availability and also a preview of the webpage in case of the online resources.[28]

6.5   Google Scholar: An Alternate

While libraries and other scholarly resources are trying to provide Google-like single user interface to encourage usage of resources subscribed by the libraries, Google has started providing scholarly content to the users through the Google Scholar that searches through scholarly full-text databases, theses, books, pre-prints and other grey literature available in open access. Since Google Scholar searches through a precompiled database of scholarly content, users get the results within a fraction of seconds. Google uses its link analysis algorithms for relevance ranking of the results. Recognizing potential and perception of the new generation users, several publishers have signed agreement with Google to supply metadata of their scholarly resources including books and journals to Google for enabling better discovery and usage of their resources through Google Scholar. With integration of online databases and online books in Google books programme with other information resources, Google Scholar provides almost 100% discovery of the resources. However, discovery of resources is only half of the process and the users have to return to the library / library catalogue to gain access to resources in full-text found through search engines like Google Scholar. This is where; libraries need to position themselves as a critical part of the information-seeking process. Search engines may enable users to discover information resources, but it is the library that provides the "last mile" service, enabling users to gain access to full-text resources.[29] To minimize the intermediate interface of library OPACs or any other gateways, Google Scholar has come up with its Library linking programme that provides direct full-text linking to the library resources through the OpenURL resolver.

However, there are a number of apprehensions regarding Google scholar, its scope, coverage, and accuracy of the content, completeness of the data, focus on specific disciplines and the lack of control over it, which is perhaps the most important concern of all.[30]

7. Conclusion

The very problem of the libraries, the discovery of resources by the users has changed with changing library resources and the library user perspective. When the generation Y, who are more web-savvy and refered the web search engines as a replacement to the libraries, the federated search solutions have once provided the library to give a single search box to the users to come back to the library for access to the scholarly and authentic information sources. But it lack with the coverage of the content of the resources. But the success of discovery product always lies in the coverage of the content which the library subscribes. With the web-scale discovery tools which provides an integrated discovery for all the resources available in library, whether directly held by the library subscribed e-content, or print material in form of catalogue, or other information sources like repository, ETDs held in the library or other openly available contents. The users who always assume the search box on library webpage for searching everything might be right now with this new technology.

References

1. Walker, J, (2006). New resource discovery mechanisms. The e-resources management handbook, 78-89

2. Walker, J., ibid.

3. Clarke, Eddie. (). Resource Discovery Tools Guide and Evaluation. Available: http://www.staffs.ac.uk/COSE/DICE/ResDisToolsandEval.pdf. Last accessed 30th Dec 2010.

4. Nagy, Andrew. (2009). A Conversation With Karen Schneider: VuFind and Summon. Available: http://www.wils.wisc.edu/events/wworld09/sessions/anagy.pdf. Last accessed 30th Dec 2010.

5. Wisniewski, Jeff. (July-Aug. 2010). Web Scale Discovery: The Future's So Bright, I Gotta Wear Shades. Available: http://findarticles.com/p/articles/mi_hb3328/is_201007/ai_n54716433/. Last accessed 30th Dec 2010.

6. Bedord, John. (2008, May). Federated Search For Your Library and in Your Enterprise. Computers in Libraries, 28.5, Available: http://www.infotoday.com/cilmag/may08/FederatedSearch_WhitePaper_CIL0508.pdf. Last accessed 30th Dec 2010

7. Bedord, John. ibid

8. Morgan, Eric Lease. (2004). An Introduction to the Search/Retrieve URL Service (SRU).Ariadne 40. Available: http://www.ariadne.ac.uk/issue40/morgan/. Last accessed . 30 Dec. 2010

9. NISO RP-2006-02, NISO Metasearch XML Gateway Implementers Guide. Available: http://www.niso.org/publications/rp/RP-2006-02.pdf. Last accessed 30th Dec 2010.

10. Content access basics - Part I - screen scraping. Available at http://federatedsearchblog.com/2007/12/27/content-access-basics-part-i-screen-scraping/. Last accessed 30th Dec 2010

11. Content access basics - Part III – OpenSearch. Available at http://hederatedsearchblog.com/2008/01/04/content-access-basics-part-iii-opensearch/. Last accessed 30th Dec 2010

12. Mercun, Tanja, and Maja, Dumer. 2008. New Generation of Catalogues for the New Generation of Users: A Comparison of Six Library Catalogues. Program: Electronic Library and Information Systems 42.3. pp. 243-61.

13. Walker, J., ibid.

14. 14Yang, Sharon Q. and Wagner, Kurt. (2010). Evaluating and comparing discovery tools: how close are we towards next generation catalog?. Library Hi Tech. 28 (4), pp.690-709.

15. Yang, Sharon Q. and Wagner, Kurt, ibid.

16. Open Archives Initiative Protocol for Metadata Harvesting. Available: http://www.openarchives.org/pmh/. Last accessed on 30th Dec 2010

17. Gibson, Ian., Goddard, Lisa and Gordon, Shannon. (2009). One box to search them all Implementing federated search at an academic library. Library Hi Tech. 27 (1), pp. 118-133.

18. Morgan, E. L. (2009). Web-scale discovery indexes and "next generation" library catalogs. Available: http://infomotions.com/musings/web-scale/. Last accessed on 30th Dec 2010

19. Burke, J. (2010). Discovery versus disintermediation: the new reality driven by today's end-user. VALA 2010. Available: http://www.vala.org.au/vala2010-proceedings/vala2010-session-1-burke. Last accessed on 30th Dec 2010

20. Vaughan, Jason. (2011). OCLC WorldCat Local. Library Technology Reports. 47(1). pp. 12-21.

21. WorldCat® Local. Available: http://www.oclc.org/worldcatlocal/. Last accessed on 30th Dec 2010

22. Summon™ Content & Coverage. Available: http://www.serialssolutions.com/summon-content-and-coverage/. Last accessed on 30th Dec 2010

23. Vaughan, Jason. (2011). Serials Solutions Summon. Library Technology Reports. 47(1). pp. 22-29.

24. EBSCO Discovery Service™ (EDS). Available: http://www.ebscohost.com/discovery/eds-about. Last accessed on 30th Dec 2010

25. EDS Content - EBSCO Discovery Service™ (EDS). Available: http://www.ebscohost.com/discovery/eds-content. Last accessed on 30th Dec 2010

26. Vaughan, Jason. (2011). Ebsco Discovery Services. Library Technology Reports. 47(1). pp. 30-38.

27. Ex Libris the bridge to knowledge. Available at http://www.exlibrisgroup.com/category/PrimoOverview. Last accessed on 30th Dec 2010

28. Vaughan, Jason. (2011). Ex Libris Primo Central. Library Technology Reports. 47(1). pp. 39-47.

29. Pomerantz, J. (2006). Google Scholar and 100% Availability of Information. Information Technology and Libraries, 25(1), 52-56

30. Sadeh, Tamar. (May 31-June 4, 2005). Google Scholar Versace Metasearch Systems. Available: http://elag2005.web.cern.ch/elag2005/media/workshop8.pdf. Last accessed 30th Dec 2010.

About Authors

Mr. Dinesh R Pradhan, Scientific Technical Officer (LS), INFLIBNET Centre, Ahmedabad - 380009
E-mail: dinesh@inflibnet.ac.in

Miss Kruti J Trivedi, Scientific Technical Assistant (LS), INFLIBNET Centre, Ahmedabad -380009
E-mail: kruti@inflibnet.ac.in

Dr. Jagdish Arora, Director, INFLIBNET Centre, Ahmedabad - 380009.
E-mail: director@inflibnet.ac.in