Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.3 No.1, 2013, pp.48-52

DOI: https://doi.org/10.51983/ijiss.2013.3.1.373

Federated Search Engines for Online Journals: A Study

M. Tamizhchelvan[1] and S. Dhanavandan[2]

[1]Deputy Librarian, [2]Assistant Librarian, Gandhigram Rural Institute, Deemed University, Gandhigram - 624 302, Dindigul District, Tamil Nadu, India

Email: tamizhchelvan@gmail.com, dhanavandan@gmail.com

(Received on 22 February 2013 and accepted on 18 April 2013)

Abstract – **Search strategy in Information retrieval is important task to get proper results without wasting time. There were many developments in searching the articles or information from the internet. There are various search engines available to search e-journals, online databases. Federated search is new search technique to get one query with results from different online databases and online journals, etc. This paper made an attempt to identify the various tools available over the Internet.**

Keywords: Search Engine, Federated Search, Knimbus

I. Introduction

The searches are quick search, combined search or advanced search in a particular website, information retrieval systems, OPACs, online databases, etc. There may be many inter-related database / RDBMS for searching the keywords and getting results. It is extended to put the bibliographic record to hold up to abstract level not the full record or full text. In this method, subject wise (LISA), publisher wise (SCOUPUS - Elsevier), Google scholar for scholar articles irrespective all subjects, and then search aggregator which is meta search engine gathers results from multiple search engines simultaneously through RSS search results, Deep web, now federated search. The Library professionals develop or assembling (subject gateway) related websites into one place for easy accessing of various sites in a particular field. Federated searching technique is a hot topic that seems to be gaining traction in libraries now-a-days. There are many technologies coming up, there will be some misconceptions about what it does. The federated search engines (FSEs) made even easy for the users to search the required information at one search and get results from various websites / portals. It is aggregated the search results from several e-resources of one's preference.

A. Definition

According to Wikipeadia1, Federated search is an information retrieval technology that allows the simultaneous search of multiple searchable resources. A user makes a single query request which is distributed to the search engines participating in the federation. The federated search then aggregates the results that are received from the search engines for presentation to the user.

Federated search2 is the technology of simultaneously searching multiple content sources from one search form and aggregating the results into a single results page. Federated search engines sometimes perform additional functions such as removing duplicates from the results lists and ranking documents against one another.

As described by Peter Jacso in 2004, federated searching consists of (1) transforming a query and broadcasting it to a group of disparate databases or other web resources, with the appropriate syntax, (2) merging the results collected from the databases, (3) presenting them in a succinct and unified format with minimal duplication, and (4) providing a means, performed either automatically or by the portal user, to sort the merged result set. Development of Federated Search engine.

There are other terms 'search aggregator', 'Deep web', 'Google scholar', 'metasearch', 'Z39.50', etc.

B. Search Aggregator

The RSS help the search aggregator for searching different search engines with their latest updates to aggregate the search results. A search aggregator3 is a type of metasearch engine which gathers results from multiple search engines simultaneously through RSS search results. It combines user specified search feeds (parameterized RSS feeds which return search results) to give the user the same level of control over content as a general aggregator.

C. Deep Web

The deep web is usually defined as the content on the Web not accessible through a search on general search engines. This content is sometimes also referred to as the hidden or invisible web.4

The Web is normally mixer of different file types and different media used to have the pages. The web pages are not at all static and changes by their administrator frequently and some of the pages are adding contents every and then for example the publishing house always the content will be added to their sites. The publishing houses also have systematic additions with help databases. In those database could not access by the users. In such database may be called as Deep web content. This type of information can be accessed only through proper search forms with username and password or authorized users.

D. Google Scholar

Google Scholar5 is a freely accessible web search engine that indexes the full text of scholarly literature across an array of publishing formats and disciplines. The Google Scholar index includes most peer-reviewed online journals of Europe and America's largest scholarly publishers, plus scholarly books other non-peer reviewed journals. It is similar in function to the freely available Scirus from Elsevier, CiteSeerX, and get CITED. It is also similar to the subscription-based tools, Elsevier's Scopus and Thomson ISI's Web of Science.Now they provide H-Index of authors' articles.

E. Metasearch Engine

A metasearch engine is a search tool6 that sends user requests to several other search engines and/or databases and aggregates the results into a single list or displays them according to their source. Metasearch engines enable users to enter search criteria once and access several search engines simultaneously. Metasearch engines operate on the premise that the Web is too large for any one search engine to index it all and that more comprehensive search results can be obtained by combining the results from several search engines. This also may save the user from having to use multiple search engines separately.7

F. Z39.50

The other concepts which is used for search strategy is Z39.50 is a client–server protocol for searching and retrieving information from remote computer databases. It is covered by ANSI/NISO standard Z39.50, and ISO standard 23950. The standard's maintenance agency is the Library of Congress. Z39.50 is widely used in library environments and is often incorporated into integrated library systems and personal bibliographic reference software. Interlibrary catalogue searches for interlibrary loan are often implemented with Z39.50 queries.8

II. Federated Search - Brief Note

In traditional search engines such as google, yahoo, msn, etc are using search engine's crawler technology for retrieving information from service provider. Whereas the searching information from the computer through internet and that to subscribed online journals is a great task. The information pooled in different online journals with different search modes and techniques used to search in each publisher's portal. If one database, one search engine is better to get results from one publisher. It is like one library having online catalogue searching the library collection is easy. Whereas more than one library or integrated with more libraries online catalogues. There will be so many problems raised to solve them. The information centres / libraries are subscribing more than one database or online journals. The federated search technique is coming up with the solution for the searching different portals in single query and searching many portals and get results in integrated manner.

There is more software available for federated search engines. Federated search engines use software "connectors" to access information sources. The federated search engines designed in such a way that the user's query will be transformed the search terms to match each content source's requirements. It submits the query to each and every source marked in the federated search tools simultaneously and gets the results from the original sources. All the results will be integrated into singly set of results page with unique display styles.

The actual databases9 results are not directly allow a user to enter the information source's application. More sophisticated ones will de-dupe the results list by merging and removing duplicates. There are additional features available in many portals, but the basic idea is the same: to improve the accuracy and relevance of individual searches as well as reduce the amount of time required to search for resources.

This process allows federated search some key advantages when compared with existing crawler-based search engines. Federated search need not place any requirements or burdens on owners of the individual information sources, other than handling increased traffic. Federated searches are inherently as current as the individual information sources, as they are searched in real time.

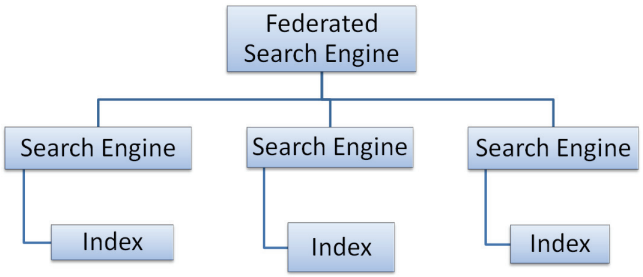

Fig.1. Federating across three search engines7
(Source: http://en.wikipedia.org/wiki/Federated_search)

The Fig.1 shows that how federated search works in different search engines and getting results. The query gets in and translates into the search query to different search engines and gets back the results to federated search engines and this display the results in a single display format instead of different display format.

Jacso (2004)10 described federated searching consists of

(1) Transforming a query and broadcasting it to a group of disparate databases or other web resources, with the appropriate syntax;

(2) Merging the results collected from the databases;

(3) Presenting them in a succinct and unified format with minimal duplication; and

(4) Providing a means, performed either automatically or by the portal user, to sort the merged result set.

III. Purpose of The Study

a) It helps to search multiple databases at one in real time and arrange the results from the various databases into a useful form and then present the results to the user;

b) It searches different content sources with one query;

c) To have an single entry point for searching;

d) It saves the time of the users for searching different portals instead of one portal.

IV. Federated Search Engines

The various portals / tools have been identified for this study, and they are discussed below:-

- Sehrch.com;
- Knimbus;
- Dbwiz;
- LibraryFind.

A. Sehrch.Com

This search engine is new approach and is a new breed of search engine; the search queries can be put across the document web and the semantic web harmoniously. The simple query language method is allowed a user to intertwine between both of these worlds. Sehrch.com is to complete a users overall information retrieval task in as short time as possible by providing the most informative entity centric result. This is accomplished by both accepting an unstructured query and applying conceptual awareness or by making structured queries, something all current mainstream search engines are incapable of doing (as they only concern themselves with the document web/ not the semantic web), which adds a whole new dimension to information retrieval systems[11].

B.Worldwide Science

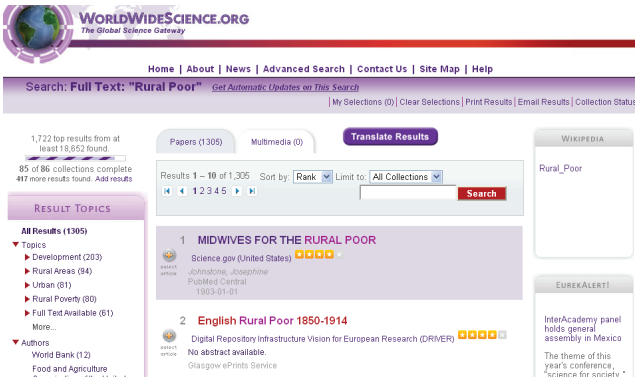

Fig. 3 WorldWideScience portal
(source: http://worldwidescience.org/wws/result-list/fullRecord:Rural+Poor/preferredLanguage:en/)

WorldWideScience is a global science gateway that combines national and international scientific databases into a search engine. From a single search form, a scientist, researcher, or curious citizen can search over fifty databases in English and now 22 multilingual sources (with translation to the searcher's native language) and seven multimedia sources. WorldWideScience is the brainchild of the director of the DOE Office of Scientific and Technical Information (OSTI). The gateway is maintained and hosted by OSTI and governed by the WorldWideScience Alliance. In this screen, authors made search on "Rural Poor" and results shown in the figure 3.

Deep Web Technologies is proud to have developed the federated search technology behind Worldwide Science. And, with the cooperation of the Microsoft Translation services team, Deep Web Technologies also implemented

the multilingual technology. It was a major undertaking but a worthwhile one for the science community, whose members can now greatly expand their reach to scientific papers in languages beyond their own.

C. Knimbus

Knimbus13 (K=Knowledge + nimbus=cloud) is a collaborative knowledge platform that connects you to relevant content and communities. It combines the best of Search, Social features and Library tools to enable higher research productivity and valuable insights. Knimbus is a collaborative research platform that enables researchers to find and share knowledge globally. It discovered the valuable and relevant information from subscribed library resources and open/web content. It provides to select the online databases by giving drop down menu and also from simple search to advanced search. The total list shows by sources in the left side of the screen. It has facility of profile option and work space. In the main page shows the search strategy of the query and results such as key term searched and the number of database chosen, etc.

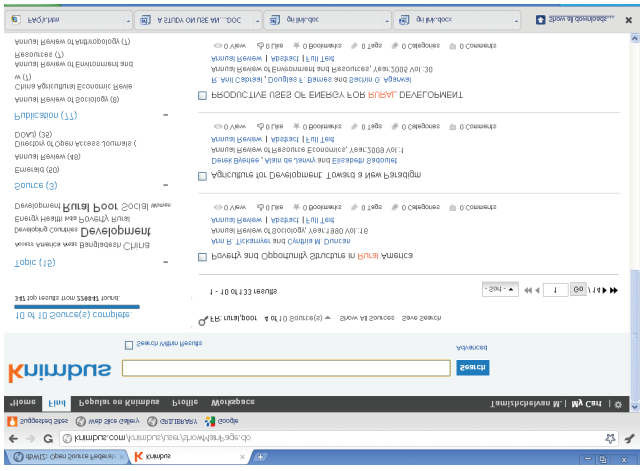

Fig. 4 Knimbus search screen
(Source: http://knimbus.com/knimbus/user/showMainPage.do)

D. Dbwiz

dbWiz14 is an open source federated search engine developed at the Simon Fraser University Library. dbWiz provides library users with a single interface for searching a wide range of library resources, and returns records in an integrated result listing. The dbWiz screencast shows the power of federated searching. Researchers can be searched multiple databases, online databases, websites, online catalogues and other online resources at the same time and get results in an integrated list. It saves the time of researchers. It can be embed the search box into the library web page from there the users can be searched the federated search.

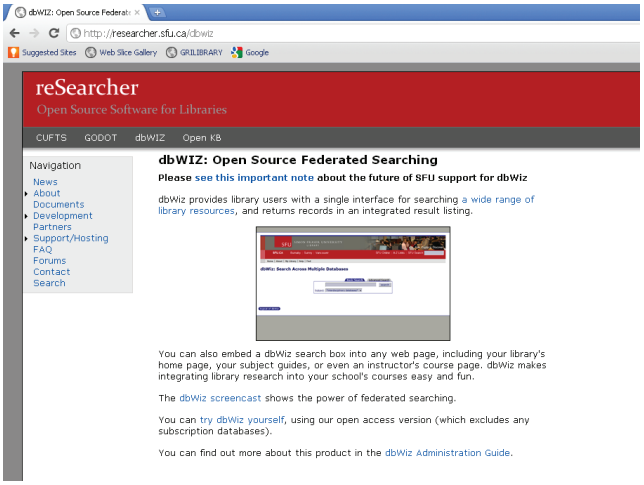

Fig. 5 dbWiz Open Source Federated Searching
(Source: http://researcher.sfu.ca/dbwiz)

E. Library Find

LibraryFind15 is an open source metasearch application developed by librarians for libraries, built with Ruby on Rails. The Features are as follows:-

- Built-in Open URL resolver;
- 2-click find workflow;
- Ability to locally index collections;
- Web-based administration;
- 3-tiered caching system (to improve speed of searches);
- Customizable user interface.

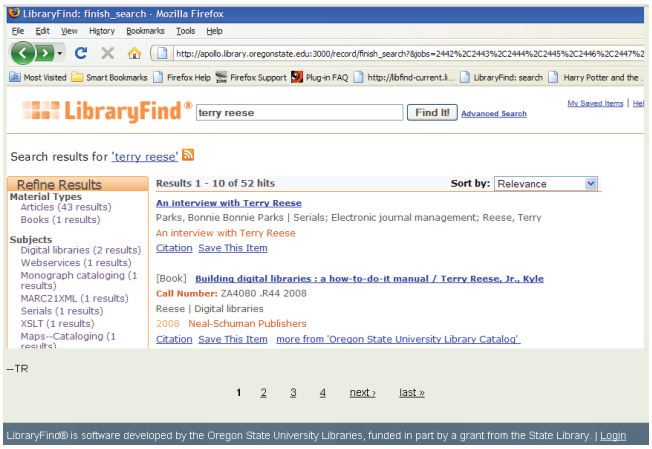

Fig. 6 Library Find federated search screen
(Source: http://libraryfind.org/)

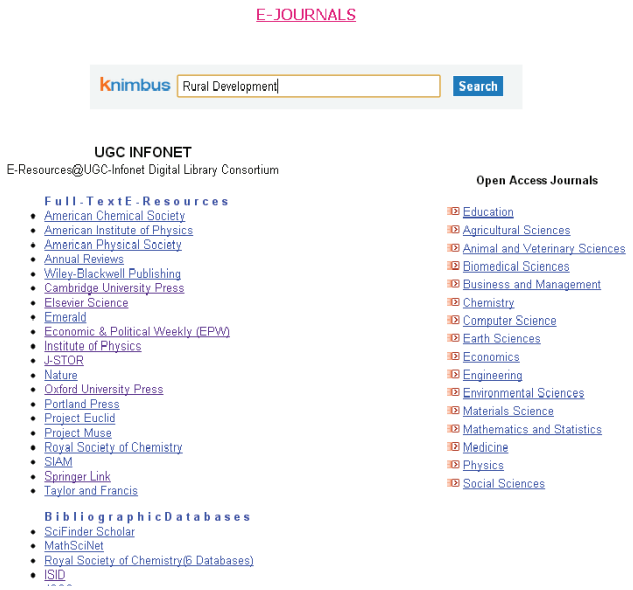

Fig. 7 Knimbus search buttons integrated with library online subscribed page

V. Problems and Prospects of Federated Search

a) Simultaneous search with heterogeneous resource types;

b) The federated search can be customized in the locally developed library website;

c) It has facility to choose user choice from subscribed journals;

d) Along with search history, user can define article alert option like individual publishers site;

e) Searching and retrieving the performance is reasonably slow;

f) As the collections and number of databases or journals grow and naturally speed of the search and retrieval comes down;

g) Searching with query is another problems using natural or controlled vocabulary.

VI. Conclusion

Federated search method is a new search method to save the user time. The various tools are available to get Federated search methods. If the federated search is properly works then there is no need of subscribing the indexing and abstracting journals for the searching the articles. Further, it brings lot of convenience to the users for their research and there is no chance to miss the articles. The user has to find out their related journals and databases for their research and bring into federated search mode and get everything in a single search query. There are ideal federated searching using Google universal searching helps more for the users. Many times Google searching ends the query instead of online journals search. Google scholar alert is another help for the users for getting their related research article published immediately client get alerted.

References

[1] Federated Search, http://en.wikipedia.org/wiki/Federated_search

[2] The Royal Hospital Medical Library, http://www.rhcml.com/FAQ

[3] Search Aggregator, http://en.wikipedia.org/wiki/Search_aggregator

[4] The Deep web, http://www.internettutorials.net/deepweb.asp

[5] Google Scholar, http://en.wikipedia.org/wiki/Google_Scholar

[6] Metasearch engine, http://en.wikipedia.org/wiki/Metasearch_engine#cite_note-0

[7] Sandy Berger's Great Age Guide to the Internet By Sandy Berger. Que Publishing, ISBN 0789734427, 2005.

[8] Z39.50, http://en.wikipedia.org/wiki/Z39.50

[9] Federated Search, http://en.wikipedia.org/wiki/Federated_search

[10] Jacso, Peter, "Thoughts about Federated Searching. Information Today", Vol. 21, Issue 9, 2004.

[11] sehrch.com

[12] Worldwide Science receives warm welcome at the UN http://deepwebtechblog.com/worldwidescience-receives-warm-welcome-at-the-un/

[13] http://knimbus.com/

[14] Dbwiz – http://researcher.sfu.ca/dbwiz

[15] LibraryFind – http://libraryfind.org/

[16] Peter Jacso,Thoughts about Federated Information Retrieval. Information Today, Vol.21, No. 9, pp.17-19, October, 2004.