Federated Search: An Information Retrieval Strategy for Scholarly Literature

Kakoli Gogoi   *Jyotika Borthakur*   *Mukut Sarmah*

Abstract

Due to rapid growth of scholarly information, need of federated searching arises. The article describes the concept of federated search and different federated search engines, features of the search engines and search strategies of some of the federated search engines. It also highlights the advantages of federated search along with its drawbacks. Various technologies used for federated searching have been also discussed. This paper also tries to give the difference between popular search engines and federated search engines.

Keywords: Federated Search, Information Retrieval

1.   Introduction

Federated search also known as metasearch or distributed information retrieval system provides one portal for searching information from multiple sources such as fee based databases, library catalogues (OPAC), internet resources, user specific digital sources etc. It has become increasingly popular for libraries to improve library services.

Due to exponential growth of databases it is very difficult to search information in various databases one by one. It is a very tedious and time consuming process for the research scholars to retrieve necessary documents. Modern libraries subscribe various databases to fulfil the information requirements of their clientele. But due to different database structure and searching methodology of each database clientele has to search each database separately. Federated search engine can help research scholar to retrieve information in a single list from different databases by putting single query. Behind the scenes, the federated search software transforms the single query into specific queries for each information source, which are then executed simultaneously. The results of those queries are returned to the needy in a single list with minimal or no item duplication. The results may be displayed to the information seeker in a list or clustered by categories.

2.   What is Federated Search

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 search is the simultaneous search of multiple online databases and is an emerging

feature of automated, Web-based library and information retrieval systems. It is also often referred to as a portal, as opposed to simply a Web-based search engine. It is sometime termed as broadcast search, parallel search, cross reference search etc. Various terms are used to refer to these tools in the literature including: meta-searchers, cross-searchers, cross-database searchers, portals, broadcast searchers or parallel searchers [9].

The term federated searching is the function of search tools, which search a number of databases, particularly subscription databases, simultaneously with one interface. The content searched by federated search tools is content that could not normally be searched via a web search engine [9].

3. Features of Federated Search:

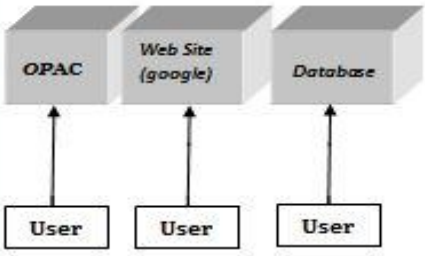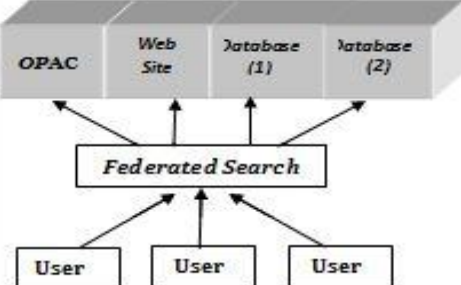Following are some important features of federated search:

❖ Federated search has ability to search multiple   databases concurrently.

❖ It has simple and advanced search facilities.

❖ It searches databases in real time.

❖ It gives importance on de duping of records.

❖ Merge search facility is also available.

❖ Sorting of records is also there.

❖ Faceted –topically clustered search results.

❖ Save, print and e-mail results.

❖ Exporting of results to a file.

❖ Extensive patron authentication support.

❖ Display databases by categories.

❖ Compliance with an Open URL resolver.

❖ Search status report.

❖ Ability to search local and remote collections as well as internet resources and search engines using three modes –http or screen scraping, Z39.50 and XML.

❖ Personalised access to resources.

❖ Ability to access to electronically available content without further authentication.

❖ Relevance ranking

❖ Unlimited simultaneous users

❖ Ability to link into interlibrary loan system.

❖ Extensive search statistics

4. Search Engine Vs Federated Search Engine   An Overview

The following Table-1, Figure-1 and figure-2 indicate the differences between the search engines and the federated search engines:

Table 1: Difference between the Search Engines and Federated Search Engines

Search Engine	Federated Search Engine
But in general engine user has to put different search query to different databases to get required information.	In federated search engine user can search multiple databases
It is time consuming	It is time less consuming. So it saves the time of researcher.
Only simple search	It has both simple and advanced search features
But it is not feasible for general search engine.	User get more scholarly information in federated search engine
But it is not possible for general search engines.	Federated search engine search real time data which is essential for researchers to get up to date information.
General search engines don't give much importance on quality.	In federated search engine targeted searches are filtered for quality.
This facility is not available in popular search engines.	In federated search engine user has get the opportunity to select the sources from where data should be extracted.
	In federated searches qualify the authenticity of information.
	Technologies used in federated search engines to get precise information are Z39.50, screen scraping or http, zing SRU protocol, XML.

Figure 1: Old Search Model

Figure 2: Federated Search Model

5. Different Federated Search Engines

There are different types of federated search engines amongst them few are free of cost and some are commercial search engines. Following are the search engines available on the web:

Free Federated Search Portals

- ❖ Scitopia
- ❖ Science Accelerator
- ❖ Science.gov
- ❖ WorldWideScience.org, etc.

Commercial Federated Search Engines

- ❖ Deep web Technology
- ❖ GIST Find- (Used by Delhi University Library)
- ❖ EBSCO Host integrated search
- ❖ Exlibris metalib
- ❖ Summon
- ❖ 360 search- (used by IIM Ahmedabad)
- ❖ Knimbus etc.

Most familiar federated search engines and their facilities are described below:

Science Accelerator

It is a gateway to science, including R&D results, project descriptions, accomplishments, and more, via resources from the Office of Scientific and Technical Information (OSTI), U.S. Department of Energy is made available as a free public service. It implements federated search to provide its encompassing coverage of important databases and collections.

5.1.1 Facilities

❖ Search within the deep Web where most search engines cannot reach.

❖ It provides facilities of both simple search and advance searches.

❖ Search diverse science databases and resources from a single entry point

❖ View early search results as the search is being performed

❖ Sort the search results by rank, source, date, or   title

❖ Select and display selected search results

5.1.2   Search Strategy

Science Accelerator is a search engine which allows both simple search and advance search strategy. The basic search form is easy to use and provides an adequate means of developing a search topic and narrowing it down to get the desired results. The advanced search (Figure 3) is available for researchers looking for specific objects when they know the title and author's name. It allows the searcher to limit the search to a date range and select which sources available in Science Accelerator.
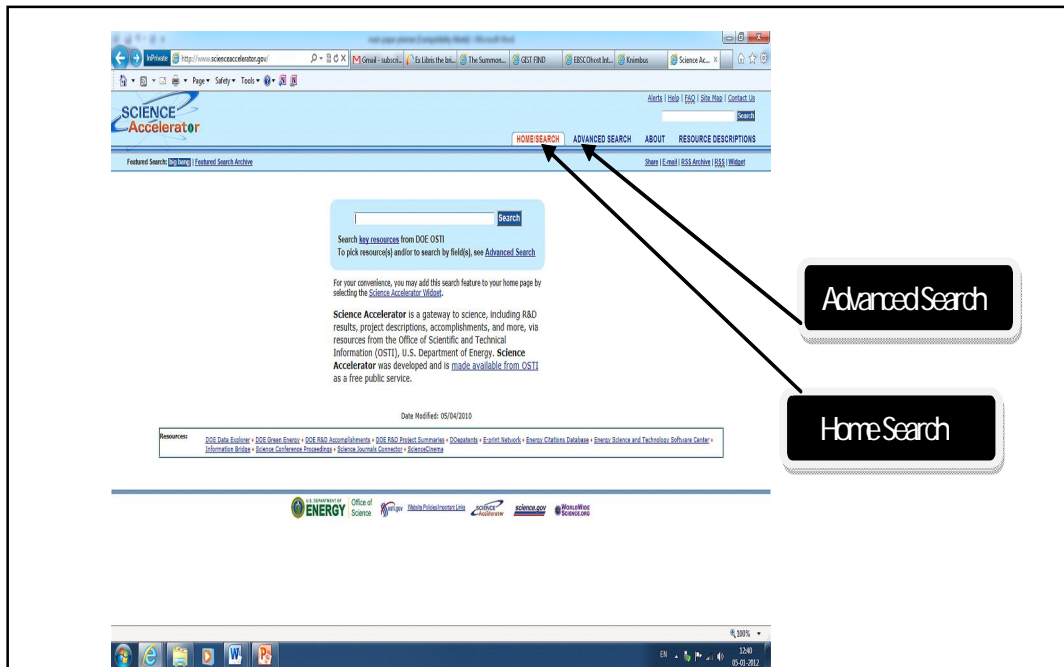
Source: http://www.scitopia.org/scitopia/ (Accessed on 30-12-2011)

Figure 4: Web Page of Scitopia.

5.3 Science.gov

Science.gov is a gateway to government science information and research results. Currently in its fifth generation, Science.gov provides a search of over 50 scientific databases and 200 million pages of science information with just one query, and is a gateway to over 2100 scientific Websites (see Science.gov fact sheet).

Science.gov is an interagency initiative of 18 U.S. government science organizations within 14 Federal Agencies. These agencies form the voluntary Science.gov Alliance which governs Science.gov.

5.3.1 Facilities

Science.gov 5.0 provides the ultimate science search through a variety of features and abilities, including:

❖ Accessing over 50 databases and 200 million pages of science information via one query

❖ Clustering of results by subtopics, authors, or dates to help you target your search

❖ Wikipedia results related to your search terms

❖ Eureka News results related to your search terms

❖ Mark & send option for emailing results to friends and colleagues

❖ Download capabilities in RIS, provide more citations download options.

❖ Enhanced information related to your real-time search

❖ Aggregated Science News Feed, also available on Twitter

❖ Updated Alerts service

❖ Image Search.

5.3.2 Search Strategy

In Science.gov initially three databases are being searched from one search box. More image databases will be added to Science.gov in the
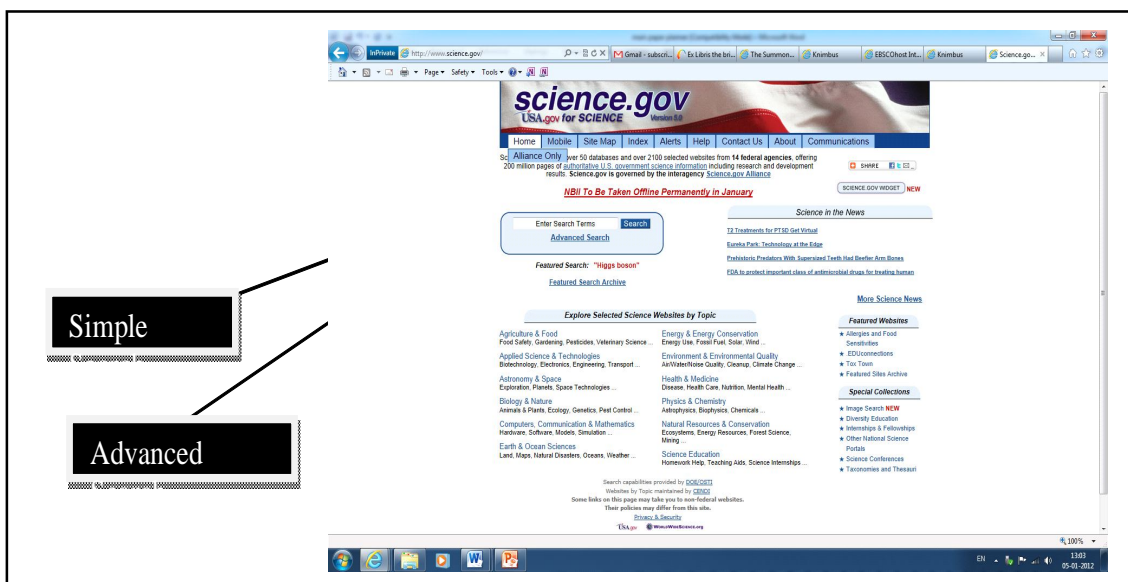

http://www.science.gov/about.html (Access on 6-1-2012)

Figure 5: Web Page of Science.gov

coming months. It allows both simple search and advance search strategies (figure 5). The researcher can explores selected science websites by topic example, Agriculture and food; applied science and technologies; astronomy and space; biology and nature; computers, communication and mathematics; earth and energy; conservation, environment and environmental quality; health and

medicine; physics and chemistry; natural resources and conservation and science education.

5.4 WorldWideScience.org

It is a global science gateway consists of national and international scientific databases and portals. The Worldwide Science Alliance, a multilateral partnership, consists of participating member
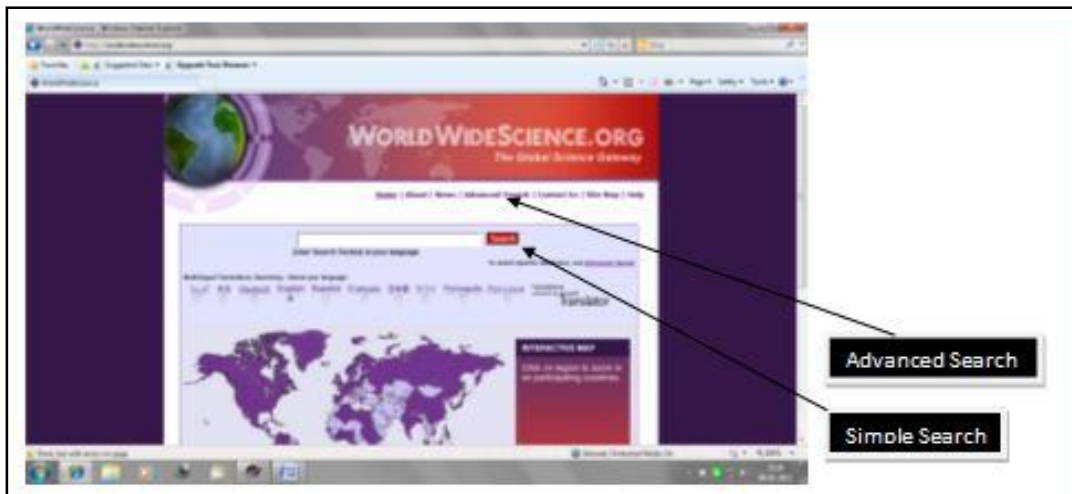
Source: http://worldwidescience.org/about.html (Accessed on 5-1-2012)

Figure 6: Web Page of Worldwide science.org

countries and provides the governance structure for **WorldWideScience.org.**

On behalf of the Worldwide Science Alliance, *WorldWideScience.org* was developed and is maintained by the Office of Scientific and Technical Information (OSTI), an element of the Office of Science within the U.S. Department of Energy [6].

5.4.1 Facilities

❖ WorldWideScience.org accelerates scientific discovery and progress by providing one-stop searching of databases from around the world.

❖ Multilingual WorldWideScience.org provides real-time searching and translation of globally-dispersed multilingual scientific literature.

❖ It provides both simple and advance searching.

5.4.2 Search Strategy

The Worldwide Science is a federated search engine which provides both simple and advance searching. Simple search technique is easy to use and provides adequate sources to satisfy the need of researchers. The advance search is the deep searching which allows the researchers to search full text of books

and journal articles when they are familiar with the title and author's name. There are three types of sources available in advance searching of Worldwide Science i.e., English sources, Multilingual sources and Multimedia sources (figure 6).

5.5 Knimbus

It is a collaborative research platform that enables researchers to find and share knowledge globally. Knimbus combines the best of search, social features and Library tools to enable higher research productivity and valuable insights. In a nutshell, Knimbus provides a smarter way to do research.

It allows Researchers to discover valuable and relevant information from subscribed library resources and open/web content. It also helps to collaborate with other researchers to build collective intelligence and insight.

6. Misconceptions and the Realities of Federated Search

Most commonly repeated misconceptions about federated searching are:

❖ **Misconception:** Federated Search Engines can search all data.

Reality: Not all federated search engines can search all databases, although most can search Z39.50 and free databases.

❖ **Misconception:** De-dupe Really Works

Reality: For federated search engines, true de-duplication is virtually impossible. In order to de-dupe, the engine would have to download all search results and compare them.

❖ **Misconception:** Relevancy rankings are totally relevant.

Reality: It's impossible to perform a relevancy ranking that's totally relevant. A relevancy ranking basically counts the occurrence of words being searched in a citation. Based on this frequency of occurrence, items will be moved closer to the top or farther down the results list.

❖ **Misconception:** Federated searching is software.

Reality: It certainly is software, but its best consumed as a service. A federated search engine searches databases that update and change an average of 2 to 3 times per year.

❖ **Misconception**: We don't make your search engine. We make your search engine better.

Reality: Federated searching cannot improve on the native databases' search capabilities. It can only use them.

7. Advantage of Federated Search

Following are the advantages of federated searching:

❖ One stops shopping like popular search engine Google.

❖ Enhanced information discovery.

❖ The ability to search multiple repositories without having to learn the specific search strategy of each repository.

❖ The ability to create one portal for all library content.

8. Disadvantages of Federated Search

There are some limitations of federated searching which are mentioned below:

❖ Federated search doesn't emulate native search. But native search is a more sophisticated than federated search.

❖ It is harder to reach deeper into a collection through federated search.

❖ Complete de duping (removing duplication) is also not feasible.

❖ It takes more time to retrieve required information than popular search engine like Google.

9. Conclusion

In this information explosion age the need of the users becomes very specific and pinpointed day by day. In such situation federated searching provides a great opportunity to information community to access more scholarly information. Now a days, federated search have received much importance amongst information community. There are many limitations while using this federated search strategies as a path of information retrieval but the main advantages of federated search engines are its use of citation indexing, multidisciplinary coverage and scholarly search features. It is also very much helpful for users to learn other information literacy basics, such as the ability to interpret bibliography or to tell the difference between books, book chapters, and periodical articles, in order to make better use of the federated search engine offered by their libraries.

References

1. Chen, X. Metalib, Webfeat and Google: The Strength and Weaknesses of Federated Search Engines compared with Google. *Online Information Review.*2006, 30(4), pp. 413-427.

2. http://knimbus.com/web/home.jsp (Access on 5-1-2012).

3. http://www.science.gov/about.html (Access on 5-1-2012).

4. http://www.scienceaccelerator.gov/ (Accessed on 5-1-2012).

5. http://www.scitopia.org/scitopia/ (Accessed on 30-12-2011).

6. http://worldwidescience.org/about.html (Accessed on 5-1-2011).

7. Jesco, P. Google Scholar: the pros and cons. *Online Information Review.*2005, 29(2), pp. 208-214.

8. Joshi, H; Vyas, M. Framework for a Federated Digital Library. . Proceedings of the 2nd International CALIBER 2004, New Delhi, 11ᵗʰ-13 Feb 2004. Ahmedabad: INFLIBNET Centre, 2004.pp.346-350.

9. Kumar, S; Sanaman, G and Rai, N. Federated Search: New Option for Libraries in the Digital era. Proceedings of the International 6ᵗʰ International CALIBER 2008, University of Allahabad, Allahabad 28ᵗʰ-29ᵗʰ February & 1ˢᵗ March, 2008. Ahmedabad: INFLIBNET Centre, 2008. pp. 267-285.

10. Maharana, Bulu and Mahapatra, Sasmita. Google Scholar: A Tool to Search Scholarly Information on the Web. Proceeding of the 4ᵗʰ Convention PLANNER-2006, Mizoram University, Aizawl, 09-10 November 2006. Ahmedabad: INFLIBNET Centre, 2006. pp.475-481.

11. Pradhan, D R; Trivedi, K and Arora, J. Searching Online Resources in New Discovery Environment: A State of the Art Review. Proceedings of the 8th International CALIBER 2011, Goa University, Goa, 2nd-4th March 2011. Ahmedabad: INFLIBNET Centre, 2011. pp. 132-145.

12. Vastrad, G; Bharathy and J and Kumar, P D. Federated Search and Discovery Tools. Proceedings of the 8th International CALIBER 2011, Goa University, Goa, 2nd-4th March 2011. Ahmedabad: INFLIBNET Centre, 2011. pp. 47-55.

13. Verma, S; Mahawar, K and Verma, R. Federated Searching: A Myth or a Reality n Web based System. Proceedings of the PLANNER 2007, Gauhati University. Guwahati, 7th-8th Dec 2007. Ahmedabad: INFLIBNET Centre, 2007. pp.

14. Zhang, S.; Tang, R and Yee, IH. User Perceptions of a Federated Search System. Digital Libraries, 2006. JCDL '06. Proceedings of the 6th ACM/IEEE-CS Joint Conference on 2006-01-01. pp. 338.

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4119154 (Accessed on 25/12/2011)

About Authors

Kakoli Gogoi, Indian Statistical Institute North East Centre, Tezpur University , Assam
E-mail: kakoli781@gmail.com

Borthakur Jyotika, Swahid Maniram Dewan College, Charing, Sivasagar, Assam
E-mail: sumki.jyoti@gmail.com

Mukut Sarmah, DLISc, Assam University, Silchar, Assam
E-mail: drmsarmah@gmail.com