

Federated Search Systems Employed in Different Libraries: An Overview

*Taufat Hussain**

Department of Library and Information Science, University of Kashmir,
Jammu and Kashmir, India

Abstract

The purpose of this paper is to provide an overview of federated model of information search and retrieval, and to explore various federated search products employed in different libraries. Limited search of literature was carried out by employing search terms as “federated search”, “cross-search”, “metasearch”, interoperability and digital libraries-federated search in various nationally and internationally reputed databases and a collection of literature on federated model and search products was identified and retrieved. The recognition of literature related with the aim of study was followed by a thorough analysis of it in order to draw inferences regarding the purpose of this study. Federated search tools can be sustainable alternatives to general search engines in searching relevant information and filling a gap established by general search engines like Google, if equilibrium is maintained between the challenges and promises presented by these systems by both vendors and clients. The study is an attempt to contribute various aspects of federated model of information search and retrieval in libraries.

Keywords: Federated search, cross-search, metasearch, interoperability, academic libraries

***Author for Correspondence** E-mail: taufat.pathan@gmail.com

INTRODUCTION

In the present knowledge driven society, World Wide Web is one of the most widely used platforms for searching and retrieving information. The searching and retrieving of some general information are achieved with the help of common information search and retrieval tools like Google, Yahoo!, etc. However, to search and retrieve some specific information from online databases need some special tools like federated search tools, provided to the user community by libraries especially academic libraries. The knowledge industry is mushrooming very fast, and thereby accelerating the production of knowledge and information which is becoming a challenge for knowledge and information management industries. Different users search disparate information databases individually, therefore spending a lot of time in accessing and retrieving reliable and relevant information. Moreover, general search engines like Google, Yahoo!, etc. cannot crawl all web pages resulting in resource accessibility issues [1]. To overcome these problems federated search systems were introduced and nowadays

“federated searching” and “meta-searching” (the outcome of federated search system) have become more popular especially in the research world. Boyd et al. comments on “federated searching” to a single box that searches everything [2]. Noerr (2006) also comments on federated search platforms as “middle ground of search” [3]. Federated searching was enunciated in 1998 when a basic concept of establishing a user friendly and common interface by Web-feat to search multiple databases simultaneously with a single query was turned into a reality [4]. Federated search also known as federated information retrieval or distributed information retrieval can be simply referred to as a mechanism for finding information from a select or group of information resources by accessing them through a single and common search interface [5]. They provide integrated access to resources of all types at one point [6].

Federated search system is an information retrieval technology that permits an information seeker to search multiple

resources on a single query simultaneously. A web patron seeking information opens a portal search tool like "Science.gov" and enters a query, just as he or she would do at general search engine like Google. But, while the patron's experience looks like Google, the architecture behind federated search is entirely different. The query is transmitted to a central server and then it is fanned out to each of a suite of databases geographically spread out across the entire world. At each database, the query causes a search to be executed and produces a hit list of search result summaries. The hit list is then transmitted back to the central server, where the hits are ranked on relevancy and sent on to the web patron. Different phrases are used to denote federated search tools in the literature including "meta-searchers", "cross-searchers", "cross-database searchers", "broadcast searchers" or "parallel searchers". Among the above mentioned phrases, the phrase meta-searcher is widely well known and agreed to by the United States National Information Standards Organization (NISO), which has taken a meta-search initiative [7].

However, the phrase "federated searching" is more desired than "meta-searching" in order to explain the utility of search tools - searching a number of databases, especially subscription databases, concurrently with a common search interface. Normally, a general web search engine cannot search the content which is searched by federated search tools [8]. The approach that general search patterns like Google uses is to "crawl" the web. Crawling, as a searching mechanism, has a major flaw of not being able to find everything. It might be a common belief that one would be able to find all web pages by means of an adequate crawling. Indeed, only a meager amount of information available on the web is reachable through crawling. The phrase "deepweb" recognizes a huge section of the web that cannot be accessed via the typical "surface web" crawlers. Surface web search engines like Google can't easily fathom the deep web because most deep web content has no links to it and further it is not designed to fill out search forms and click "submit" the way humans do. In most cases, for instance, Google doesn't fill out search forms, this is exactly what federated search applications (also

known as federated search engines) do. Why doesn't Google fill out forms? Google encounters a trouble in filling out forms as it has a general way of crawling links from any website. Unlike Google, federated search engines are customized for each web form encountered by them, and are programmed with sufficient information about each search form.

LITERATURE REVIEW

A number of studies have been conducted on federated search tools. However, selective studies on federated search tools by different researchers are enlisted as:

Chen finds that MetaLib and WebFeat (library federated search tools) have fundamental differences between them [9]. They cannot compete with Google in speed, simplicity, ease of use, and convenience, nor can they be truly one-stop shopping. Further explored their strengths lie in the contents they search as well as in the objective way they retrieve and display results, adding that the federated search engines have still relevancy with information literacy education. George, 2008 [10] reveals some important issues related with MetaLib interface like problems with the login, problems with primary and secondary navigation, confusing terminology, and inconsistency with the site design and user expectations.

Warraich et al. depicts that users perform the required tasks and have found Electronic Library Information Navigator (ELIN) to be a useful tool for searching [11]. They, however, have had some technical limitations and faced difficulties with the choice of search options and were frustrated by the display of results. The authors also reveal that the full text availability and relevancy of subject results have been as major problems. Ruddock and Hartley found that choice processes employed to choose federated/metasearch tools were repeated across libraries [12]. They also found that a prior/existing relationship with vendors has had a strong influence on how libraries choose metasearch systems. Mohamed and Hassan found that none of the available federated search tools provide error messages for model queries; most of the scholars had short queries; Boolean operators were used with half of the total queries; federated search

tools did not provide techniques for query reformation; the optimal days for system maintenance were the non-weekend vacations; and early morning was the best time for maintenance [13]. Boyd et al., reveals that one-box federated search of databases brings as many challenges as promises to database searching, especially in terms of adapting these systems to user needs and the effects [2]. This new mode of searching will have on users' research behaviours, and thereby this issue of balance between the challenges and promises presents several librarians with strong interests in reference and instruction, who tell the story of adapting a federated search system for their libraries and reflect upon how federated searching can transform the approach of students towards research, and on the repercussions federated searching has on information literacy skills and the relevancy of results found. Liu explored the changing user needs and expectations like a single point of discovery (federated search tool) for all resources from anywhere at any time; a simple, intuitive user interface and good results; seamless services to ensure resources immediately available; open platform and easy user communication; convenient self-services and personalized services; and embedding library services into users' workflows. Further it is concluded that libraries, particularly those serving academic communities, need to reassess their roles to offer services in innovative ways like introducing federated search systems into library services [14].

Gibson et al., recommended active development of the current federated search systems, re-assessment of the need of a federated search tool on a consortial basis, continuous assessment of the current federated search marketplace with an eye to choose a next-generation federated search tool that includes effective de-duping, sorting, relevancy, clustering and faceting [10]. Si, O'Brien and Proberts found the large variety of terminology resources (like thesaurus) distributed throughout the web, the proposed middleware service was essential to integrate technically and semantically the different terminology resources in order to facilitate subject cross-browsing [15]. Further, researchers made a set of recommendations,

outlining the important approaches and features that support such a cross-browsing middleware service. Joint found the success of introducing a federated search engine was dependent on a fundamental approach which subsidiaries other active library search tools to the novel one-stop search interface which necessarily be devised for ease without imitating existing, complicated library information search and retrieval tools [16]. Armstrong found that users exhibit a slight partiality for federated searching over a unified database searching, based on quality of results retrieved and probability of future use [17]. Woods revealed that neither federated search products nor enterprise search solutions like the Google Search Appliance could answer the needs of libraries to provide a single search box with a single, integrated result set because they had a fundamental flaw—they searched in real time [18]. Warnick found WorldWideScience.org that filled a unique niche in discovering scientific material in an information landscape that includes search engines such as Google and Google Scholar [19]. Shokouhi and Si found that federated search is preferred over centralized search alternatives in many environments like enterprise environments [5].

Georgas investigated that students prefer the federated search tool to Google for doing research. He further adds that despite federated searching's limitations, students see the need for it, libraries should continue to offer federated search (especially if a discovery search tool is not available), and librarians should focus on teaching students how to use federated search more effectively [20]. Baker and Gonzalez concluded that students were highly satisfied with federated searching being most effective when looking for very specific information and article abstracts and least effective when searching general terms and resources other than journal articles [21]. Caplan, finds prior to the recognition of the term "discovery tool" as a "unified discovery product" in 2009, federated searching was a little component of the search interface provided by a discovery tool [22]. However, post 2009 period federated search systems, searching databases individually without making a unified index of all available

federated databases, were distinguished to a great extent from “discovery tools”- as being able to establish a unified index of accessible resources like Google. Federated searching was found to be one of the fragile areas in the OPACs, keeping in consideration the adaptability of 12 advanced functionalities of the Next Generation Catalogues (NGC) into the contemporary OPACs [23]. The purpose of choosing and incorporating federated search component into the library gateway of University of Northern Iowa (UNI) was to liberate users from cumbersome activity of knowing databases and search strategies to let them begin searching earliest without going through multiple steps from the homepage of the library, and to enable them to search various relevant databases concurrently [24]. Prior to implementation of federated search at UNI, the searching process usually needed users to go away with library regulated web space to do their job. However, post implementation of federated search the library was capable of keeping users in library regulated web space to a much greater extent while performing the information search and retrieval process [24]. Federated searching benefitted libraries by increasing the use of infamous databases, liberating librarians from educating users about specific interfaces in order to make them available for imparting knowledge about search strategies to users, and allowing new users to fan a single search to manifold databases [24].

Boyd et al., reveal that one-box federated search of databases brings as many challenges as promises to database searching, especially in terms of adapting these systems to user needs and the effects this new mode of searching will have on users' research behaviors, and thereby this issue of balance between the challenges and promises presents several librarians with strong interests in reference and instruction, who tell the story of adapting a federated search system for their libraries and reflect upon how federated searching can change the way students do research and on the implications federated searching has on information literacy skills and the quality of results found.

FEDERATED MODEL

Federated search products/tools are definitely embedded with certain protocols like Z39.50 International Next Generation (ZING), Simple Object Access Protocol (SOAP), and XML, etc. Library has made use of Z39.50 protocol for quite a long time for searching and retrieving information from bibliographic and full-text databases. This protocol can be acknowledged as a US National Standard and is promoted by the Library of Congress, U.S.A. [25].

A client and server architecture is used in federated search systems. The server is authenticated with updation and responding to queries. The client part is responsible for connecting with end users, receiving queries from end-users; and sending, receiving and mixing received responses from the server, finally presenting them to the end-user. Indeed, the relationship can be ascertained by certain protocols.

OVERVIEW OF SOME POPULAR FEDERATED SEARCH TOOLS

There are (Ruddock & Hartley) several common features in these systems as summarized below:

- Synchronized searching of multiple resources.
- The source of the search results is located as and when any search hit is clicked on by an information seeker
- Navigation to the original source of the search result.
- Selected resources are on hand-meaning the federated resources are always available.
- Search results can be saved by print, e-mail or download.
- Search results are sortable and reorderable.

At present, MuseGlobal and Webfeat have come out as most popular choices in the market. Webfeat and Metalib have become focus of most of the literature, although other vendors like Deep Web Technologies, Millennium Access Plus and Innovative are discussed, too. Besides vendors, libraries have also developed their own meta-search systems as done by California Digital Library and the Fedlemur project [12].

A noteworthy change has taken place in the market, possibly most particularly; Webfeat has been got hold of by Proquest, a supplementary of the Cambridge Information Group, and will be amalgamated with Serials Solutions being already as one of the leaders in the market. Moreover, novel federated search systems are progressing to come up, such as V-spaces by Infor Library Solutions and Liblime's Masterkey Meta-search. Another name SirsiDynix has also been a focus of library administrators, by Vista Equity Partners.

OJAX is an instance of a rich and open source (OS) federated search system for metadata harvesting from Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) compatible electronic repositories. It was developed by employing Web 2.0 open source (OS) design methods and usability testing within an agile software development framework. Agile software development engages a series of incremental software releases or iterations. Coding is started at an early stage of a project, and each release, produced periodically over time, necessarily creates a "Working Version". Agile methodology not only involves developers, but also users and customers. Every working version is presented to customers and users for an assessment, and their opinion is incorporated into the design for the next release or iteration.

CONCLUSION

As the employability of federated search tools in digital libraries and academic libraries has increased over time, it can be deduced that they are playing an indispensable part in effective information search and retrieval processes. In terms of cost-benefit analysis, it is an investment of privileged returns. It also implies that developmental research should be continued with regard to federated search tools in order to improve their functionalities and long-term permanency. As per the analysis done on identified and retrieved literature, the following inferences were drawn about federated search tools and their employability: The power of federated search tools like Metalib and Webfeat are possessed by the content they search. In the scientific world, the

discovery of scientific material is powered by Worldwidescience.org. However, some federated search tools like Electronic Library Information Navigator (ELIN) have a problematic issue with full-text availability and relevancy of subject results. The success of federated search technology is the result of strong architectural base behind its technological framework. Federated searching technology is capable of searching networked resources/sites using their native search interface, unlike crawlers (general search engines). This feature of searching networked resources/ sites through their native search interface adds to their usability and popularity. However, the implementation of federated search tools in digital libraries and academic libraries has a considerable impact on information literacy education.

It has been seen that how libraries choose (i.e., activities involved in procuring federated search systems) federated search systems is repeated by libraries. Moreover, active relationship with vendors has had a great manipulation on how libraries pick these systems.

Federated search tools can be sustainable alternatives to general search engines in searching relevant information if equilibrium is maintained between the challenges and promises presented by these systems by both vendors and clients. It is also concluded that libraries especially academic libraries are required to reconsider their roles to provide services in ground-breaking ways like introducing federated search systems into libraries.

REFERENCES

1. Xu F. Implementation of a Federated Search System: Resource Accessibility Issues. *Serials Review*. 2009; 35(4): 235–41p.
2. Boyd J, Hampton M, Morrison P, Pugh P, Cervone F. The One Box Challenge: Provides a Federated Search That Benefits Research Process. *Serials Review*. 2006; 32(4): 247–54p.
3. Noerr KT. Muse Metasearch: Beyond Federated Searching. *Serials Review*. 2006; 32(3): 186–9p.

4. Pesch O. Re-inventing Federated Searching. *Serials Review*. 2006; 32(3): 183–5p. DOI: 10.1016/j.serrev.2006.06.007
5. Shokouhi M, Si L. Federated search. *Foundations and Trends in Information Retrieval*. 2011; 5(1): 1–102p.
6. Boss SC, Nelson ML. Federated Search Tools. *The Reference Librarian*. 2005; 44(91–92): 139–60p.
7. NISO. *NISO Metasearch Initiative*, 2014.
8. Cox CN (ed.). Federated Search: Solution or Setback for Online Library Services, 2007.
9. Chen X. Metalib, Webfeat and Google: The strengths and weaknesses of federated search engines compared with Google. *Online Information Review*. 2006; 30(4): 413–27p.
10. Gibson LG, Goddard L, Gordon S. One Box To Search Them All. *Library Hi Tech*. 2009; 27(1):118–33p.
11. Warraich NF, Ameen K, Tahira M. Usability study of a federated search product at Punjab University. *Library Hi Tech News*. 2009; 26(9): 14–5p.
12. Ruddock B, Hartley D. How UK libraries choose metasearch systems. *Aslib Proceedings: New Information Perspectives*. 2010; 62(1): 85–105p.
13. Mohamed KA, Hassan A. Web usage mining analysis of federated search tools for Egyptian scholars. *Program*. 2008; 42(4): 418–35p.
14. Liu G. The Application of Intelligent Agents in Libraries: A Survey. *Program: Electronic Library And Information Systems*. 2011; 45 (1): 78–97p.
15. Si LE, O'Brien A, Proberts S. Integration of distributed terminology resources to facilitate subject cross-browsing for library portal systems. *Aslib Proceedings*. 2010; 62 (4/5): 415–27p.
16. Joint N. Managing the Implementation of a Federated Search Tool in an Academic Library. *Library Review*. 2009; 58(1): 10–6p.
17. Armstrong AR. Student perceptions of federated searching vs. single database searching. *Reference Services Review*. 2009; 37(3): 291–303p.
18. Woods RF. From Federated Search to the Universal Search Solution. *The Serials Librarian (From the Printed Page to the Digital Age)*. 2010; 58 (1-4): 141–8p.
19. Warnick W. Federated search as a transformational technology enabling knowledge discovery: the role of WorldWideScience.org. *Interlending and Document supply*. 2010; 38(2): 82–92p.
20. Georgas H. Google vs. the Library: Student Preferences and Perceptions When Doing Research Using Google and a Federated Search Tool. *Portal: Libraries and the Academy*. 2013; 13(2): 165–85p.
21. Baker S, Gonzalez AC. Graduate Students and Federated searching. *Internet References Services Quarterly*. 2012; 17(1): 13–31p.
22. Caplan P. On discovery tools, OPACs and the motion of library language. *Library Hi Tech*. 2011; 30(1): 108–15p.
23. Yang SQ, Hofman MA. Next Generation or current generation?: A study of the OPACs of 260 academic libraries in the USA and Canada. *Library Hi Tech*. 2010; 29(2): 266–300p.
24. Caswell JV, Wynstra JD. Improving the search experience: federated search and the library gateway. *Library Hi Tech*. 2010; 28(3): 391–401p.
25. Ingram C, Awre C. *Contextual Resource Evaluation Environment (CREE): Literature Review*, 2005.

Cite this Article:

Hussain T. Federated Search Systems Employed in Different Libraries: An Overview. *Journal of Advancements in Library Sciences*. 2015; 2(1): 57–62p.