

FEDERATED SEARCHING: A MYTH OR A REALITY IN WEB-BASED SYSTEM

Shilpi Verma

K L Mahawar

Rekha Verma

Abstract

In today's information explosion age the needs of users' becoming complex day by day. The task of information professionals is also very complicated and difficult to meet out the needs of the different users. This paper deals with concept of Federated searching that provide facility to search over different databases in one interface. The challenges of today's searching and need of federated searching is also dealt. The paper also describes the process and hub as well as local database relationship along with point to point model and broker model architecture. The factors affecting the federated searching are also covered. The myths of the federated searching are also dealt in this paper in detail.

Keywords : Web Bases System, Federated Search

1. Introduction

As we know the importance of computer increases our capacity to store, search and retrieve for information externally. Over the past decade strong advances have been made in the development of technologies and system that enable the effective management of digital resources .The users' needs are becoming very specific and pin pointed. No one database or search service had all relevant information. Digital resources multiply organizational aspect at present. Such resources are like a new kind of 'Island', document delivery, delivery services, data archive, data service, collection management of different sources have created a challenge in front of information organizers and professionals. The increases in such an island have various drawbacks also. For users, the use of several databases is really a frustrating and daunting task, So Federated searching is a very new concept that is gaining importance in libraries everywhere.

2. Concept of Federated Searching

Federated searching that is known as meta- searching or cross database searching that provides facility to user to search many networked information resources from one interface. "The types of resources that can be searched include local and remote library catalogs, abstracting and indexing databases, full-text aggregator databases, and digital repositories. From a

technical standpoint, this software uses a distributed search method across heterogeneous databases using multiple search protocols. "Some specialized federated search engines are limited to metadata harvesting, searching homogenous repositories, or using a limited number of protocols. Because of the special nature of these applications, they have limited value for general library purposes." [1]

"Federated search technology is an integral component of an Information Portal, which provides the interface to diverse information resources. Once the user enters his or her search query in the search box of the Information Portal, the system uses federated search technology to send the search string to each resource that is incorporated into the Portal. The individual information resources then send the Information Portal a list of results from the search query. Users can view the number of documents retrieved in each resource and link directly to each search result." [2]

Today's Search Challenges

As the users are becoming more pinpointed the queries are getting more complex, so we can say these below major search challenges have to be faced by the information Professionals:

- Getting the data for centralized and vertical search services
- Keeping search quality high for these large databases
- Answering hard search queries

Why Federated Searching

There are several needs for evolvement of federated searching. Some of them are as follows:

- The increased need for pin-pointed and information
- It is very difficult to find out which database to search
- It is very difficult to search all the databases even if you know the databases.
- It saves the time of the user, as faster searching is possible.
- The need to learn one simple interface rather than many complex interfaces.
- The search quality is high.
- The databases have simplest technical search.
- It helps in getting answers of quite complex and typical queries.
- The Search environment is robust type and centralized.
- It helps in locating best documents with the help of ranking.

It means federated searching helps users in expanding there queries to multiple information resources by a single hit and it also presents

information in a single ranked list. It also enables users to have link directly to each resources to expand their queries.

Process of Federated Searching

- User gives the search query;
- Software uses a distributed search method across heterogeneous databases using multiple search protocols;
- Sources that can be searched include local and remote library catalogue, abstracting & indexing databases, digital repositories and full text also;
- Provides a ranked list for the user's query.

The process involves gathering of data from numerous databases into a single database. XML/GML format is used to read and publish the data. Federated search software uses standardized protocols to access databases such as Z39.50. "Gathering of data from numerous disparate databases into a single database from which it can be re-published in a unified manner. Since the easiest way to read and publish data is in XML / GML format, the term also incorporates the concept of "schema translating" this data into XML / GML formats and delivering this data according to a particular schema." [3]

Local & Hub Database Relationship in Federated Searching

The hub collects the data from multiple sources and presents the same in a unified manner for whole world.

Local Database

Local database hold the information of particular organization and fed it to the hub database.

"Defining a common model is ideally suited to XML Schema, which in turn makes XML / GML the ideal format for harvesting. Once the mapping to common model has occurred, the easiest way to respond to harvesting demands from the Hub Database is to make the data available to the Internet (in some cases, intranet) via a Web Feature Server.

As such, the supply is achieved via these three easy steps :

Step 1) Understanding of the common model

Step 2) Mapping to the common model

Step 3) Publishing the data via a Web Feature Server" [4]

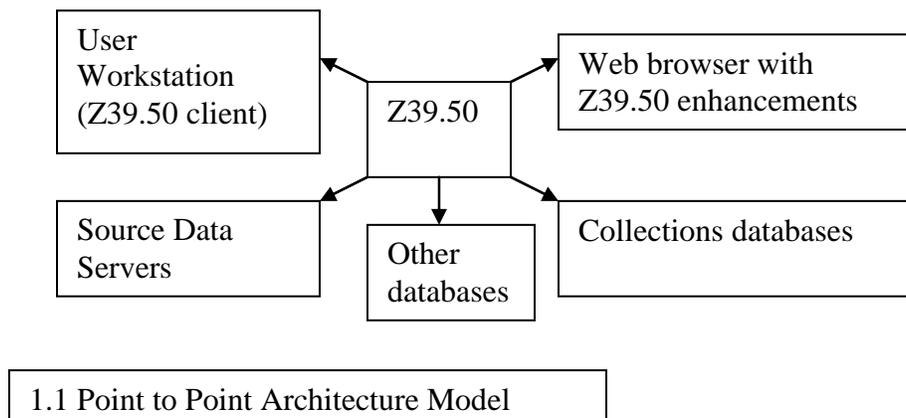
Hub Database

Hub database periodically collects the data from different databases and after collating the same present it to the external users. The mapping at local databases is necessary to do the process.

There are 2 kinds of Hub & Local databases architecture :

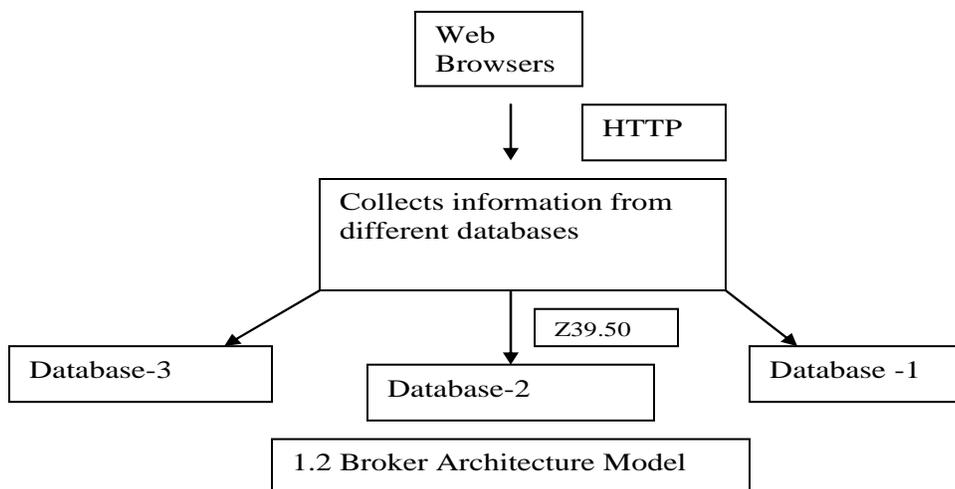
(1) Point to Point Model Architecture

In this approach for each client to have information about the servers it wants to interoperate with and contact them directly.



(2) Broker Model Architecture

In this system a central component, broker operates independently. It provides access to multiple resource databases. Broker manages the community of the users.



Factors which make federated searching more powerful

There are certain factors that make the federated searching more powerful .Some of these are as follows :

- Compatibility with other standards
- Database compatibility
- Display of full text in true native interfaces
- Unlimited number of database searching at a time
- Open URL compatibility with all database results
- Parse citation for all databases
- Sorting by relevancy, date, author, title & Publication

Myths about Federated Searching:

There are some myths about the Federated searching:

- (1) Most of the Federated searching providers don't search all the databases available for local and remote users, because it is very difficult to manage authentication for subscription databases.
- (2) "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. The limiting factor is not federated search engine technology, but the way databases return results: 10 or 20 records at a time. Completing a true de-dupe operation would take hours because a single search might produce 100,000 hits. These hits or citations typically come back 10 to 20 at a time. If it takes 5 seconds to download 20 hits, it would take hours to download them all. And the same citation may appear in different places in results sets from different databases. So to completely de-dupe search results, it's necessary to download all results from all databases" [5]
- (3) "Federated search engines are limited in what they know about the documents they find, because they don't actually crawl and index those documents—the underlying one-index search engines do. So, while Google's spider looks at billions of documents across the Internet, Dogpile does not look at any—it merely gets the list of results from Google (and other search engines) and stitches together a list of search results." [6]
- (4) It can not provide totally relevant ranking based searching. The indexing that content providers are providing with full text is unavailable to federated search engines. They only have citations for searching.

3. Conclusion

There is great deal more information available now days-of this there can be no doubt. At the same time lot of junk information is available and that waste the time of the users. Federated searching provides facility to user to search several databases in same interface. Federated searching provides

several opportunities for information professionals, but at the same time it has several complexities also. The paper provides an overview of federated searching, its need, process etc. The local database and hub database relationship is also very important at the same time. The myths about the federated searching are dealt in the paper. No doubt federated searching provides a great opportunity for information professionals to provide better services to the users.

References

- (1) <http://www.ala.org/ala/alctscontent/alctspubsbucket/webpublications/cataloging/researchtopics/federated.htm>.
- (2) "What is federated search"
(http://www.library.geac.com/page/fedsearch_LIB.html)
- (3) <http://vectorlmedia.com/resources/white-papers/data-harvesting/>
- (4) <http://www.snowflakesoftware.co.uk/news/papers/dataharvesting.pdf>
- (5) "The truth about Federated searching"
(<http://www.infotoday.com/IT/oct03/hanel.shtml>)
- (6) "Federated Search always seems just around the corner" Biznology Blog : Where Business and Technology Come Together (by Mike Moran), April 28, 2006.

ABOUT AUTHORS

Shilpi Verma is Lecturer, Dept. of Library & Inf. Sci., Babasaheb Bhimrao Ambedkar University, Lucknow (UP). E-mail- shilpi_lko@rediffmail.com

K L Mahawar is Reader & Head, Dept. of Library & Inf. Sci., Babasaheb Bhimrao Ambedkar University, Lucknow (UP). E-mail- mahawar_67@yahoo.co.in

Rekha Verma is Librarian, MBP Govt. Degree College, Lucknow (UP). E-mail- rekhaverma_lib@yahoo.co.in