

Federated search and discovery solutions

Arunadevi S Lingam

Librarian, Oriental College of Pharmacy, Library, India

*Corresponding Author: Arunadevi S Lingam

Email: librarian@ocp.edu.in

Abstract

This paper represents Federated search tool which brings together the resources to a library subscribes and allows cross-searching of these resources. They work using the Z39.50 protocol used by database suppliers to enable communication between computers. It also covers search provider. Advantages & disadvantages of federated search have been discussed.

Keywords: Federated Search, Search Engines, Discovery tools.

Introduction

It is a fact that with the advent of the World Wide Web, the information "search" has grown to be a significant business sector of a global, competitive and commercial market. Libraries are only one player within this market. In the electronic information environment one of the responses to the problem of bringing large amounts of information together has been for libraries to introduce portals. A portal is a gateway, or a point where users can start their search for information on the web. Federated search tool" which brings together the resources to a library subscribes and allows cross-searching of these resources. They work using the Z39.50 protocol used by database suppliers to enable communication between computers. Federated search compliments the Z39.50 facilities, it offers the ability to extend search beyond the online catalogue virtually to any electronic resources that can be authenticated at the point of access. This means that users no longer need to jump from one search interface to another.

History

Federated searching started in 1998 when WebFeat Team (originally founded in 1992 as an information technology consultancy) took the initiative to connect any or all of their databases at the same time through a simple, common user interface and turned this idea into a product called Webfeat. WebFeat can search any database, including licensed databases, free databases, catalogues, Z39.50, Telnet, proprietary databases, anything! And with SMART, WebFeat's next generation usage tracker, you can report detailed database usage within your library with just a few mouse clicks.

Definition

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, databases or other query engines participating in the federation. The federated search then aggregates the results that are received from the search engines for presentation to the user.

Peter Jasco defines federated search as, "Transforming a query and broadcasting it to a group of disparate databases with the appropriate syntax, merging the results collected from the databases, presenting them in a succinct and unified format with minimal duplication, and allowing the library patron to sort the merged result set by various criteria".

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. Discovery Services are capable of searching quickly and seamlessly across a wide range of periodically indexed content provided by participating publishers and may include a library's catalogue and other local content in the index.

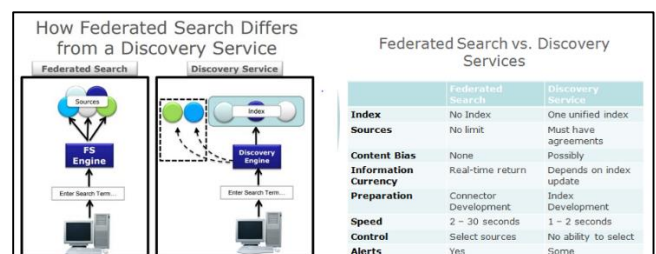


Fig. 1: Difference between Federated Search and Discovery Services

Need and Purpose: The need and purpose of the federated searching are as follows: -

1. The growth of different types of databases, produced by different suppliers, with numerous interfaces and logins means that library users can find it confusing when attempting to access information;
2. Library OPACs and web-pages have been alienating users with their use of library terminology and by including long lists of databases that users find it difficult to select from and search;

3. The needs and expectations of library users, particularly students using academic libraries. The growth of different types of databases, produced by different suppliers, with numerous interfaces and logins means that library users can find it confusing when attempting to access information.

In industrial search engines, such as LinkedIn, federated search is used to personalize vertical preference for ambiguous queries. For instance, when a user issues a query like "machine learning" on LinkedIn, he or she could mean to search for people with machine learning skill, jobs requiring machine learning skill or content about the topic. In such cases, federated search could exploit user intent (e.g., hiring, job seeking or content consuming) to personalize the vertical order for each individual user.

Features of Federated Search Engine: The features of Federated Search Engines are as follows:

1. **Limiters:** These were restricted to keyword and a few other choices. Today, many more are available and include subject, keyword/descriptor, author, title, date (range), full text, peer reviewed, and format (book, article, image, and audio).
2. **Simple and Advanced Search:** Simple search is more like Google's one-box search, usually the default, while the advanced search allows for further options such as the setting of limiters.
3. **Clustering:** With this feature, results can now be grouped together by subject. So, Jaguar, the animal, would be grouped separately from Jaguar, the car.
4. **Visual Search Interface:** This gives the user the added option of viewing results through a visual interface, similar to Aqua Browser.
5. **Faceted Results:** Conceptually similar to clustering, this groups results by source (database, OPAC), subject, or format/type (article, book, image, audio).
6. **RSS Feeds/Search Alerts:** These tools allow users to follow the same search over a period of time and keep track of any changes or updates.

Federated Search Technologies

There are mainly four technologies used for federated searching.

They are as follows:

Screen Scrapping or HTTP

"Hyper Text Transport Protocol" - HTTP is the single most important technology that drives the web and yet remains virtually transparent. Without this protocol HTML and XML via the web would not be able to perform the myriad of tasks that we put them to daily. The HTTP is an application-level protocol for distributed, collaborative, hypermedia information systems.

HTTP communication usually takes place over TCP/IP connections. TCP guarantees that packets arriving to and from the web server are error free and in the right order. It

doesn't however guarantee that packets arrive no matter what the network conditions are. When communications are congested or unavailable web page delivery is slow and can time-out.

Z39.50 Protocol

Z39.50 is an American national standard for information retrieval. It is an open standard that enables communication between systems that run on different hardware and use different software. The Z39.50 standard was developed to overcome the problems associated with multiple databases searching such as having to know the unique menus, command language, and search procedures of each system accessed. It simplifies the search process by making it possible for a searcher to use the familiar user interface of the local system to search both the local library catalogue as well as any remote database system that support the standard.

In libraries, the Z39.50 protocol is most often used for searching OPAC sources. The important facilities offered by Z39.50 are as follows:

1. **Browse:** Allows the client to scan the contents of wordlists or indexes on the server. This can be particularly useful in the case of controlled keyword lists or facets.
2. **Access and resource control:** Allows authentication of users, and cost control and online charging for commercial services.
3. **Sort:** Allows the client to request different orderings of query results, eg. relevance ranking, sorting by date or version number, etc.
4. **Explain:** Allows the client to interrogate the server about a number of details about its contents and its level of support for the application profile.
5. **Item Order:** Allows offline ordering of materials in cases where they cannot be delivered electronically, or where per-unit charging (eg. online charging) is required. Such services are being supplied in an ad-hoc fashion by online Web-based component repositories such as ASSET. The item order service provides a ready-made and semantically standardized version of this service.
6. **Item Update:** Permits an authorized client to update the contents of the remote database.

SRW (Search/Retrieve Web Service)

Search/Retrieve Web Service is a new HTTP-based information retrieval protocol providing broadly the same facilities as Z39.50, but by means of very different technology. SRW is designed to be a low barrier to entry solution to performing searches and other information retrieval operations across the internet. It uses existing, well tested and easily available technologies such as SOAP and XPath in order to perform what has been done in the past using proprietary solutions. The protocol has two ways that it can be carried, either via SOAP or as parameters in a URL. This second form is called SRU – Search Retrieve by URL. Other transports would also be possible, for example simple

XML over HTTP, but these are not defined by the current standard.

The primary function of SRW is to allow a user to search a remote database of records. This is done via the search Retrieve operation, in which the client sends a search Retrieve Request and the server responds with a search Retrieve Response. The request has several parameters, most of which are optional. The response is primarily a list of XML records which matched the search, along with the full count of how many records were matched.

XML (Extensible Markup Language)

XML stands for Extensible Markup Language. XML is a markup language much like HTML and was designed to carry data, not to display data. XML is a software and hardware independent tool for carrying information. It is used both to encode documents and serialize data. It supports Unicode, allowing almost any information in any written human language to be communicated.

It is the most common tool for data transmissions between all sorts of applications, and becomes more and more popular in the area of storing and describing information. XML simplifies data sharing as in the real world; computer systems and databases contain data in incompatible formats.

Advantages and Disadvantages

Advantages

1. With federated searches, not as many results come up with a specific search related to their topic compared to the abundance of unrelated results with Google. More is not always better. In addition to filling out forms and combining documents from multiple sources, another important benefit of federated search engines is that they search content in real time. Real time data is crucial for researchers who are searching for up-to-the-minute content or for content that changes frequently.
2. It is difficult for most students to choose appropriate, relevant sites from hundreds of thousands of hits. Using a federated search engine can be a huge time saver for researchers. It performs the many searches on the user's behalf.
3. Targeted searches are usually filtered for quality. Federated search engines show their value best in environments in which the quality of results matters, such as libraries, corporate research environments, and the federal government.
4. Federated searches qualify the authenticity of the information. For example, anyone can write a report on a topic and post it on the Internet. That does not mean that information was checked for accuracy. By using this new add-on feature to the school's library's automation system, students can better ensure the information they use for their research is accurate. With a federated search engine, the information has been checked and verified by educators and professionals.
5. The federated search includes books and other materials that already may exist in the school library. Thus, the

federated search engine acts as a helpful librarian does, directing users to excellent quality.

Drawbacks of Federated Search

Limitations of the current generation of federated search engines are as follows:

1. The lack of a uniform authentication standard means that some databases are inaccessible to federated search engines.
2. True, full, duplication is impossible because databases download results in small sets and metadata standards vary by resource.
3. Relevancy ranking is limited by the quality of the metadata, which usually does not include abstracts or full-text information.
4. Although federated search systems are fundamentally software, they must be implemented and managed as a service, which takes a great deal of resources.
5. Federated search engines cannot improve on the native interface in terms of search accuracy and precision.
6. Federated searching is not for power searching needs. Just as with searching meta search engines, only basic Boolean commands can be used.
7. The federated search has some other issues as well. First, it cannot cover all online library resources.
8. The goal of one-stop shopping cannot be achieved completely by any federated search. There are various reasons for this: Some databases do not work with any federated search at all, such as SciFinder Scholar. SciFinder Scholar does not use a web browser but rather requires its own internet client. Neither MetaLib nor WebFeat can cover SciFinder Scholar.
9. If databases require a login, they will not work with the federated search.
10. Some databases work with one federated search product but do not work with the other. MetaLib cannot search LexisNexis databases because LexisNexis does not allow Z39.50 or XML gateway access. WebFeat cannot search databases that do not have a search box on their front page because WebFeat counts on the search box on the native interface to search. □ Many libraries have databases on a pay-per-search basis, and libraries normally do not want them to be searched by a federated search for budgetary reasons.

It may not make sense to add to a federated search menu the very specialized databases that most general users would not be interested in, or the databases that require special software. One example is Inter-university Consortium for Political and Social Research (ICPSR) that requires statistics software such as SPSS to view data.

Conclusion

Search-engines are among the most used resources on the internet. Federated search technology is an integral component of an Information Portal, which provides the interface to diverse information resources. Individual end users will benefit from federated search technology. This

blends e-journals, subscription databases, electronic print collections, other digital repositories, and the Internet. Federated searching reduces the time it takes to search and usually displays results in a common format. It cannot replace information literacy education or the learning process either, partly because it cannot make searching as easy as a Google search, as serious research may require selecting various information sources beyond Google results.

Source of Funding

None.

Conflict of Interest

None.

References

1. Federated Search, Wikipedia. Retrieved on 18/07/2020 from https://en.wikipedia.org/wiki/Federated_search
2. Dhruv A. (2015). Personalized Federated Search at LinkedIn. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM). Retrieved on 17/07/2020 from <https://arxiv.org/pdf/1602.04924.pdf>
3. Kumar S, Sanaman G, Namrata R, (2008). Federated Search: New Option for Libraries in the Digital Era. International CALIBER – 2008, 267-285p. Retrieved on 17/07/2020 from <http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/1233/1/27.pdf>
4. Gayatri V, Bharathy J, Kumar D, (March, 2011). Federated Search and Discovery Tools. Conference paper presented at 8th International CALIBER - 2011, Goa University, Goa. Retrieved from <https://www.researchgate.net/publication/320107410>
5. Fryer D. (March-April, 2004). Federated search engines: federated searching aggregates multiple channels of information into a single searchable point, Online, 28 (2), p.16.
6. Milad S, Luo S, (2011). Federated Search Retrieved on 18/07/2020 from <https://www.microsoft.com/en-us/research/wp-content/uploads/2011/01/now.pdf>
7. Abe L, (2013). Understanding Differences Between Federated Search and Discovery Services. A Slide share Retrieved on 10/01/2019 from <http://deepwebtechblog.com/wp-content/uploads/2013/11/ANKOS-SWETS.2013Presentation.pdf>

How to cite: Lingam AS. Federated search and discovery solutions. *IP Indian J Libr Sci Inf Techno* 2020;5(1):39-42.