Federated Search Engines, 2001–2003

Federated searching (also known as meta-searching or cross-database searching) is a technology that allows users to search many networked information resources from one interface. Despite this seemingly simple definition, the technology is quite complex, and the implementation of the technology in the context of libraries is still relatively young. This essay will provide an overview of federated searching, summarize recent publications on the topic, and discuss potential future research.

“Meta-search engines” that cumulate results from multiple search engines have been on the Web since the mid- to late-nineties. However, early systems only searched publicly accessible Web sites and used relatively simple technologies to retrieve information. The new generation of federated search engines possesses qualities that allow sophisticated in-house implementation for library purposes. Most importantly, the software provides the ability to search across multiple information resources from one user interface (Tennant 2003). 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. They are not addressed in this essay.)

Federated search software uses standardized protocols to access databases. The most common protocol used is Z39.50. Some target databases that do not comply with the Z39.50 standard can still be searched using “translator” programs that convert the query format of the federated system into the format of the native system. However, many information resources do not make their query protocols public, and thus they cannot be searched using a federated search engine. The search results that are retrieved from various targets may be deduplicated to reduce extraneous results. Some systems also rank results by relevancy or permit some other type of sorting. User authentication is another necessary technology for federated search systems. This stems from the use of licensing agreements that libraries sign with vendors. These agreements typically limit access to certain groups or numbers of users affiliated with an institution or consortia. Luther provides a list of the software vendors for federated search systems (2003).

When a federated search engine is implemented at a particular library, it then becomes a unique service (Tennant 2001). Federated-searching software allows customization, so no two implementations are exactly the same. For example, a library may choose to include all of its online resources as targets for a federated search engine, or it may choose to create subject groupings first, each of which leads to a federated search service for a narrow topic. Gerrity, Lyman, and Tallent discuss implementing a federated search system at Boston College, where they promoted the new service as “MetaQuest” (2002).

A common theme in the literature is the difficulty of implementing a federated search engine. A great deal of work must go into installing one, especially integrating the system with heterogeneous databases. For example, the Z39.50 protocol is implemented differently by various services, and each target requires individual configuration. While setting up the system, many decisions must be made in terms of how to present the service to the public. The maintenance of the service is ongoing, because target databases change continuously and libraries add and drop databases. Training for staff is required to implement the service and learn the administrative interface. There also has to be a certain level of orientation and instruction for patrons to learn the new service. In order to facilitate a movement towards standardization for federated search systems, the National Information Standards Organization (NISO) is sponsoring a Metasearch Standards Initiative. Information about this initiative is available on the NISO Web site (NISO 2003).

Future Research
There are many areas for future research. Lewis notes that the most important issues facing the federated search system at the University of East Anglia are difficulties with authentication, ongoing maintenance of the system, and monitoring user satisfaction (2002). Lewis also anticipates that integration of systems will lead to new services, such as the ability to request items from interlibrary lending through the federated search system. Hane notes the limitations of the current generation of federated search engines (2003). These include:

- the lack of a uniform authentication standard means that some databases are inaccessible to federated search engines.
- true, full, deduplication is impossible because databases download results in small sets and metadata standards vary by resource.
- relevancy ranking is limited by the quality of the metadata, which usually does not include abstracts or full-text information.
- although federated search systems are fundamentally software, they must be implemented and managed as a service, which takes a great deal of resources.
- federated search engines cannot improve on the native interface in terms of search accuracy and precision.

There is much potential for ongoing research to address current shortcomings and future services of federated search engines.

**Works Cited**


Prepared by Brian Surratt, Electronic Resources Cataloger, Texas A&M University Libraries, bsurratt@lib-gw.tamu.edu