

# **Special Topic: Federated Searching**

**Prepared for:**

**LIS 672:**

**Technology for Libraries & Information Centers**

**Dr. Luz M. Quiroga**

**University of Hawai'i at Mānoa**

**Fall 2009**

**By:**

**Peter Kanehiro**

**November 25, 2009**

## Table of Contents

What is Federated Searching?.....	1
Selecting the Right Solution .....	3
Hawaii State Public Library System, University of Hawaii Libraries .....	5
Two Representative Products .....	6
Comparison of WebFeat and LibraryFind .....	7
User Experience: LibraryFind.....	10
Conclusion .....	12
References.....	13

## What is Federated Searching?

Dr. Péter Jácso wrote in an article, “[Database searchers] don’t go to see if another database might have more and/or better results.... Most give up and angrily leave.... They’ll go to Google...” (Jacso, 17). This is often the case when people have to spend their precious time struggling with multiple unfamiliar interfaces and query languages only to find a minimal number of useful results, if any.

Searching for information using electronic databases can be tedious and time-consuming. We may start by searching one database and, after dead-ending a few times, find a few useful items. We know that we should searching a different database but need to move on in our busy schedules or are simply too frustrated and fearful of the experience that might ensue if we were to continue with that second database whose interface and query language we might not be familiar with (Jacso, 17).

If only there were a way to gather resources from multiple sources simultaneously and deal with a single interface, the job would be so much faster and easier.

Federated searching is a solution to this plight of many searchers. The following definition of “federated search” highlights the aggregating function of the federated search:

**“[a] search for information using software designed to query multiple networked information resources via single interface” (Reitz).**

From this definition, one can infer the following characteristics that identify federated searching:

1. *Search scope*: the federated search inspects multiple resources simultaneously for relevant resources. In fact, depending on the capabilities of the tool, a federated search can include multiple databases (abstracts, indexes, or full text), and digital repositories (Reitz). Some

sources mention that even the “deep Web” can be a search target (“Federated Search.” *Wikipedia*).

By way of background, conventional search engines (e.g., Google) cannot find deep-web items. A conventional search engines find new websites by following links from the sites that it already “knows” about, but many deep web items do not have links from other pages (i.e., it “crawls” the web) (Lederman).

The deep web has been defined as follows: **“publicly accessible information available via the World Wide Web but not retrievable using search engines that rely on crawlers of spiders, for example, data in file formats such as PDF, database content accessible only by query,...etc. The number of documents...is estimated to be 400 to 500 greater than the amount...retrievable via conventional search engines...” (Reitz).**

2. *Software*: it needs software in the form of a federated search engine. These search engines use a protocol standard such as Z39.50 (Reitz). (Z39.50 is “a set of rules allowing different computer systems to communicate” (Kochtanek and Matthews, 106). Another standard, SRW/SRU is a “more standardized XML based gateway protocol [that] was developed as an “advancement on the Z39.50 protocol” to work with XML-based gateways that e-journal vendors (e.g., EBSCO, Elsevier) have developed to improve the passing of search requests from federated search software to their databases (Webster, 360). Another standard, MXG, is a standard exchange protocol that is described as a “refinement of SRW/SRU (Webster, 360).
3. *Presentation*: results are presented in a uniform interface, that of the federated search engine. Depending on the particular program’s capabilities, the results can be ranked and “deduped” (i.e., duplicates items are removed if they appeared in more than one source) (Reitz). The

“relevance ranking feature of some search software weights the records it retrieves by the degree to which they meet the requirements of a query” and typically presents them in descending order of rank (Reitz). “Deduping” refers to the removal of all but one occurrence of a record from a list of search results (Reitz).

### **Selecting the Right Solution**

There are a number of federated search products on the market. Linoski and Walczyk offer some helpful suggestions for selection. The program selection criteria should consider both the library administrator’s and the end-users’ perspectives.

First, from the administrative standpoint, areas that the library should consider include the following:

1. *Organizational considerations:* a) the reason for implementing federated searching should be clarified before selecting a product, b) staff buy-in and support are essential in getting library users to utilize federated searching (opportunities will be sub-optimized if library staff members do not introduce it to users and teach them how to use it) (Linowski and Walczyk).
2. *Administration module’s capabilities:* the administration module enables the library to customize the search engine, brand it, add or delete connectors, etc. (Linowski and Walczyk). Some products may offer the alternative of having the vendor make these adjustments, and the library will need to compare the relative costs and benefits of vendor and in-house adjustments (Linowski and Walczyk). Customizable items may include: a) predefined subject search categories: these can alleviate the burden on the user of determining which databases to have the search engine search, b) visual customization (simplicity to facilitate ease of use,

branding), and c) statistical reports (search terms, sessions, etc.): these may be available from the module directly or may need to be ordered from the vendor (Linoski and Walczyk).

3. *Hosting*: a vendor-hosted solution would be beneficial if the library lacks in-house technical support. In this case, the library may still be able to make changes via the administrative module, but the vendor would handle all maintenance (Linoski and Walczyk). Local hosting is a widely available option that might be viable if the library has the sufficient in-house support and technical knowledge to perform maintenance and install upgrades (Linoski and Walczyk). Locally hosted open source software might be a cost-effective alternative if in-house support is very strong (Linoski and Walczyk).
4. *Pricing*: the price will be directly dependent on the number of resources (or “connectors”) that the search engine will be configured to search, so the library should determine which resources to include (e.g., OPACs of neighboring libraries might be a useful resource to include) (Linoski and Walczyk).

From the users’ standpoint, the library should consider points such as the following in order to provide an optimal user experience:

1. *Interface*: the library should evaluate the user-friendliness of the interface of each product being considered. The user should be able to “log on and immediately know what to do” (Linoski and Walczyk).
2. *Search progress indicator*: of course, the search engine should give some indication when it is searching. Other things to consider include whether the users can begin to view partial results even while the search engine is still searching (Linoski and Walczyk).
3. *Results display*: the library should consider the search engine’s ability to configure search results: this includes the method of relevancy ranking of results retrieved from disparate

databases, de-duping capabilities, user-initiated re-ordering of results (Linoski and Walczyk).

### **Hawaii State Public Library System, University of Hawaii Libraries**

The federated searching systems of two Hawaii library systems' will now be discussed. They are the University of Hawai'i (UH) libraries and the Hawai'i State Public Library System (HSPLS). These two library systems were chosen because their locale and familiarity made them of particular interest to this writer, a library science student at the University of Hawai'i at Mānoa. Also, for purposes of research for this paper, the resource people at the University of Hawai'i were easy for this writer to access and the already-operational HSPLS software was accessible by using the writer's HSPLS library card.

The UH libraries plan to implement LibraryFind, and HSPLS already has WebFeat in place (Jordan, November 20, 2009; Hodges, November 20, 2009). User impressions of LibraryFind as implemented at the Oregon State University libraries will be summarized in the next section as it is of interest to this writer's classmates and instructor, being affiliates of UH.

WebFeat is available through the HSPLS website ([librarieshawaii.org](http://librarieshawaii.org)) under the online databases section. The user must login by entering a valid HSPLS card barcode and PIN. Both basic and advanced search modes are available. The user can limit the search to one or more of 63 databases (HSPLS subscriptions only)—including the HSPLS OPAC—one of ten categories (e.g., biographies, business, careers and testing, etc.), and year (1900 to 2009) (Hawaii State Public Library System <http://wfxsearch.webfeat.org/wfsearch/menu?cid=10134>). In the advanced search mode, the user can enter search terms in search boxes and specify for each the search parameter (abstract, author, keyword, subject, or title).

## **Two Representative Products**

WebFeat and LibraryFind are presented on the following pages as a tabular comparison to provide a sampling of two representative federated search products. WebFeat, represents a commercial federated searching solution, and LibraryFind is an open-source solution that will be addressed in detail later in this report.



### Comparison of WebFeat and LibraryFind

No.	Criterion	Search Engine	
		WebFeat (tested at: Hawaii State Public Library System (HSPLS))	LibraryFind (tested at: Oregon State University)
General:			
1	No. of compatible databases	~9,000 (HSPLS implementation: 63 HSPLS-subscription databases including HSPLS OPAC)	(not available)
2	Principal products	a WebFeat Express: single library/user profile	LibraryFind 0.9.2
		b WebFeat Enterprise: multiple libraries, library-specific user profile/access privileges/screen views, usage tracking	
Administrative:			
3	Administration module	a Administration: database usage, individual-library usage (WebFeat Enterprise)	a Administration: website claims “web-based administration” (probably indicates that institution can generate reports, etc.). Ability to index collections locally.
		b Customizability: appearance can be tailored to match customer institution’s look.	b Customizability: customizable interface (probably includes tailoring to match customer-institution’s look)
4	Hosting	a Local hosting: implied in vendor website.	a Local hosting. No indication of vendor hosting.
		b Vendor-hosted: Hosted on “redundant, load-balanced server cluster network.” Includes installation, configuration, and maintenance at no additional charge.	
5	Pricing	Not available.	Free

No.	Criterion	Search Engine			
		WebFeat (tested at: Hawaii State Public Library System (HSPLS))		LibraryFind (tested at: Oregon State University)	
6	Online support	Not available. Customer probably needs to negotiate support contract.		a	General discussion: blog-like forum. User registration required for posting privileges. Open archives are indexed by month (January 2007 through present). Each month's posts are sortable by author, date, subject, and thread.
				b	Documentation: installation, customization, and PowerPoints on these subjects. (Note: browser could not find webpage or displayed warning "this connection is untrusted.")
				c	Downloads: appear to be program files (could not access)
User-Focused:					
7	Interface	a	Search box driven: blue "search" button starts search.	a	Search box driven: "find it" button starts search.
		b	Layout: simple and logical. Search box, category, and list of resources are set off from each other in separate light blue areas. Can select subject area with check box or select desired databases manually.	b	Layout: simple and logical. <b>Left:</b> "Refine results" <b>Center:</b> search results <b>Right:</b> sort options (drop box)
		c	Navigation: buttons (View, PDF, Check with A-Z to see if full-text is available) allow for viewing of PDF-format article or access of item through article's resident database.	c	Navigation: hypertext links (search results: blue; sort options: gray)
8	Search progress indicator	a	Blue bar (right part of screen): blue segments flow from left to right.	a	Solid horizontal blue bar (left part of screen).
		b	Results count (right part of screen): updates number of hits by database.	b	Rotating pinwheel to right of search box.

No.	Criterion	Search Engine	
		WebFeat (tested at: Hawaii State Public Library System (HSPLS))	LibraryFind (tested at: Oregon State University)
9	Results display	a	<i>Summary list:</i> each item has abstract (if available), article title, publication title, page numbers, language, subject terms, ISSN, etc.
		b	<i>Native results:</i> full-text items presented in fully functional interface of resident database (e.g. EBSCO Host).
		c	<i>Sort options:</i> author, clustered, date, grouped, relevancy, title
		d	<i>Exportability:</i> citations can be exported to citation management applications (e.g., EndNote)
Source:		www.webfeat.org	libraryfind.org
		www.librarieshawaii.org	osulibrary.oregonstate.edu

## User Experience: LibraryFind

I tried out the LibraryFind implemented at Oregon State University libraries. The basic-search box is very easy to identify, near the top of the libraries' home page and marked prominently with the LibraryFind name (Oregon State University Library).

To perform a basic search, the user simply enters the search terms into the search box and clicks "go." The results are displayed in a list similar to that displayed in most article databases. Items are numbered, and each item has basic bibliographic information and a short sentence about the item. The results can be reconfigured by using the following functions:

1. *Refine results*: this function allows the searcher to limit the search results to one of four clickable parameters—material types (books, articles), subjects, authors, and database.
2. *Sort by*: this function allows the searcher to re-order the search results according to one of five ordering schemes—relevance, author (alphabetical order by surname), reverse chronological order, and author (reverse alphabetical order by surname).

Each item summary in the list has the following information: format (e.g., book), title, author(s), content description, publication year, and facts of publication.

1. *Book Record*: each list item also contains these hypertext links which function in the following manner when clicked:
  - a. Title: the OCLC WorldCat (web-based version) record showing the nearest libraries that have copies entered in WorldCat.
  - b. Citation: clicking this link displays a box that contains full bibliographic information.
  - c. Save This Item: clicking this link resulted in the message "item saved." It was not apparent where the item was saved.

- d. More from 'OCLC WorldCat': this shows a blank WorldCat search screen. The relevance of this to the selected item was not apparent.
2. *Article Record*: the summary record for an article has information roughly analogous to that of a book. In addition to the items of information found in the record for a book, the article record contains also a short sentence about the article's content, availability, etc.

The article record's hypertext links are:

- a. Title: this link retrieves an item record that has information similar to that of article index records found in a database (title, author(s), citation URL, abstract, keywords, issue date, publisher, remainder of citation information, ISSN, etc.).
- b. Full Abstract: the record for a book does not contain this. This link superimposes on the record a gray box containing an abstract. The abstracts I viewed were incomplete, and some were truncated in mid-sentence. Moreover the abstracts do not necessarily describe the article's content. Both of these idiosyncrasies make it difficult at best for the user to determine whether the article is useful to his or her purposes Example: "Available from Haworth Press at:" (The phrase ended in mid-sentence.)
- c. Citation: this link superimposes on the record a box containing bibliographic information for the article (author(s), title, publication title, volume, issue, page(s), and date).

Notably the information is not all present in every record.

- d. Save This Item: as with the book record, it is not apparent where the item is saved, but the phrase "Item Saved" in green characters replaces "Save This Item" when the latter is clicked.

There is no link for "More From 'OCLC WorldCat.'"

## **Conclusion**

Federated searching offers a lot of promise and hope to information seekers. It has the convenience of simplicity since the user can search with just one interface. It should be noted, however, that the user may have to deal with the “native” interface (e.g., EBSCO) once he or she clicks an item in the federated search results list, so using the search results may ultimately require dealing with multiple interfaces. As I observed, the search results do not seem to have the same quality as those from the individual databases (incomplete sentences, abstracts that do not provide insight into the resource’s content, etc.). The advantage of time savings, however, is indisputable, assuming that the software searches exhaustively.

The products on the market seem to offer solutions that are varied enough to fit an institution’s individual needs. In terms of installation and maintenance, the institution can choose between vendor-supplied or in-house alternatives. For example, multiple-library systems with varying user-access levels can make use of products such as WebFeat Enterprise, which can be tailored to meet such needs. Also, the software cost can also be as low as zero for an open-source solution if the institution has the requisite in-house IT capabilities to install it. Overall the federated searching seems to be headed in the right direction, and library users can appreciate one more information retrieval tool that has the potential to boost their work efficiency.

## References

- “Documentation” <http://libraryfind.org/documentation> (accessed November 21, 2009).
- “Download” <http://libraryfind.org/download> (accessed November 21, 2009).
- “Federated Search.” *Wikipedia* [http://en.wikipedia.org/wiki/Federated\\_search#cite\\_note-0](http://en.wikipedia.org/wiki/Federated_search#cite_note-0) (accessed November 19, 2009).
- Hawaii State Public Library System <http://wfxsearch.webfeat.org/wfsearch/menu?cid=10134> (accessed November 22, 2009).
- Jacso, Peter. “Thoughts about Federated Searching.” *Information Today* 21, issue 9: 17-20 <http://eres.library.manoa.hawaii.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=14656508&site=ehost-live> (accessed November 19, 2009).
- Kochtanek, Thomas R. and Joseph R. Matthews. *Library Information Systems: From Library Automation to Distributed Information Access Solutions*. Westport, CT: Libraries Unlimited, 2002.
- Lederman, Sol. “Federated Search Finds Content that Google Can’t Reach Part I of III.” *Alt Search Engines* <http://www.altsearchengines.com/2009/01/11/federated-search-finds-content-that-google-cant-reach-part-i-of-iii/> (accessed November 19, 2009).
- “LF General – General Discussion of LibraryFind” <http://lists.oregonstate.edu/mailman/listinfo/LF-General/> (accessed November 21, 2009).
- LibraryFind <http://libraryfind.org/home> (accessed November 21, 2009).
- Linowski, Alexis and Tine Walczyk. “Federated Search 101.” *Library Journal* Volume 133 (Summer 2008) <http://www.libraryjournal.com/article/CA6571320.html> (accessed November 19, 2009).
- Oregon State University. <http://osulibrary.oregonstate.edu/> (accessed November 21, 2009).
- Reitz, J. M. *ODLIS: Online Dictionary for Library and Information Science* [http://lu.com/odlis/odlis\\_f.cfm](http://lu.com/odlis/odlis_f.cfm) (accessed November 20, 2009).
- WebFeat <http://www.webfeat.org/index.htm> (accessed November 22, 2009).
- Webster, Peter M. “Challenges for Federated Searching.” *Internet Reference Services Quarterly* 12, issue 3 / 4 (2007): 357-68 <http://eres.library.manoa.hawaii.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=27265315&site=ehost-live> (accessed November 21, 2009).