Metasearch Technologies in Reference Work, OAI, and Search Navigation Assistance

William H. Mischo, Mary C. Schlembach

**Grainger Engineering Library Information Center
University of Illinois at Urbana-Champaign**

Introduction

Libraries and librarians will continue to develop and collaborate on access services that better address user information needs. It is important for libraries to focus on providing mechanisms to transform the many distributed digital collections into true “digital libraries” with the essential services that are required to make these digital libraries useful to and productive for users.¹ To accomplish this, libraries will focus on providing enhanced gateway and navigation services to guide users through the present distributed, heterogeneous information environment. This distributed scholarly information environment is populated by: silos of full-text repositories maintained by commercial publishers, professional societies; open preprint servers; Abstracting and Indexing (A & I) Services and publisher search and discovery sites; local, regional, and national online catalogs; and publisher, vendor, and library vertical portals, institutional repository systems, and learning management systems. Navigation and database selection functions are extremely important in providing users with more effective and efficient retrieval assistance for these multiple, discrete information resources.²

One technique to provide enhanced access to multiple, distributed information resources is metasearching.³ Metasearch technologies provide simultaneous, asynchronous search capabilities over multiple information resources. Metasearch systems typically employ either broadcast or federated search methods. In a broadcast search, the user search arguments are sent, or “broadcast” asynchronously (all at the same time) to remote, distributed systems and the search results are collected and displayed to the user. In a federated search system, the heterogeneous remote information resources are typically imported or “harvested” (sometimes using Open Archives Initiative or OAI protocols) into a local, central site and the normalized results are placed into a homogeneous database system for search and discovery. The broadcast and federated approaches are not mutually exclusive. Some systems perform broadcast searching over a combination of distributed, heterogeneous systems and systems that are already federated.

Technologies

Various metasearch system solutions have been developed by libraries, vendors, and information providers for simultaneous searching over online databases, A & I Services, Online Catalogs (OPACs), OAI federated systems, and other information resources.^{4,5,6,7} Libraries have

implemented a number of vendor-based metasearch systems utilizing broadcast-search techniques over the major A & I Services and OPACs.⁸ These systems are offered within integrated library systems often in conjunction with local link servers. These metasearch systems utilize Z39.50, HTTP “screen-scraping”, and XML gateway broadcast search techniques. These standard techniques are utilized in vendor cross-database search systems and in other custom metasearch applications. Most articles on metasearching describe vendor systems or report on custom library applications, such as SearchLight in the California Digital Library and FlashPoint at the Los Alamos Laboratory. The vendor systems are well-described in the above cited articles and within their Web sites.

By moving the resource selection task to a higher level, metasearch systems over A & I Services have served to shift the focus from interface designs that display lists of discrete information resource titles to interfaces that emphasize function or needs-based approaches to retrieval. However, metasearch technologies offer additional opportunities and applications beyond simultaneous search of discrete A & I Services. Broadcast-based metasearch systems can provide a powerful mechanism for training reference staff and assisting in both reference and patron point-of-need services by performing multiple searches within OPACs and other resources and simultaneous searches over OPACs, e-resource registries such as TDNet or Serials Solutions, and local topic-specific databases. Metasearch techniques can provide the parallel processing approach that Richardson suggested to help solve problems relating to reference transactions.⁹ This paper will report on several applications, both broadcast and federated, that utilize metasearching techniques to provide enhanced reference assistance for staff and end-users. These applications address several primary point-of-need reference activities such as verification of specific journal and conference references and identification and retrieval of the full-text of articles. These types of metasearch-based services also allow for the exploration of weighted searching and best-match searching techniques that allow the metasearch system to interpret the results of multiple searches and present custom result displays.

Metasearch vendors and system developers are exploring best practices and system designs that utilize information technologies such as local link resolving, OpenURL, OAI protocols, and Digital Object Identifiers (DOIs). The OpenURL standard provides a streamlined mechanism for expressing and transmitting metadata (data describing an information resource—authors, title, journal name, volume, publication year, URL of full-text resource, etc.) between different systems. The OAI is based on the preprint server movement and provides protocols for harvesting remote metadata. Libraries and information centers can then build services, such as search and discovery, over the harvested metadata records. DOIs are unique, persistent labels or identifiers for journal articles and conference papers that are used to link to the full-text of the specific article or paper.

Projects

The Digital Library team based at the Grainger Engineering Library Information Center at the University of Illinois at Urbana-Champaign (UIUC) has been involved in OAI and metasearch activities as part of the Digital Library Initiative-I grant program, a Andrew Mellon Foundation grant to produce OAI harvesting tools, a National Science Foundation NSDL (National Science

Digital Library) grant, and a grant from the federal Institute of Museum and Library Services (IMLS).

The Grainger group has developed a number of custom broadcast and federated search systems that operate over a number of OPACs (local, regional, and national), A & I Services, OAI information providers, the CrossRef metadata database, local databases, and a Library-wide E-Resource Registry based on TDNet data feeds. Several of these custom metasearch systems perform multiple searches within a single resource in order to determine the optimum result set. Together they form a suite of modules that make up a system for training and assisting reference staff and for providing end-users with a function-based or point of need interface. In the course of testing the system with reference staff and patrons at the Reference Desk, it became clear that the system could provide a valuable service for end-users. The resulting services assist in training Grainger Library Graduate Assistants (Library Science students who work 15-20 hours per week in the Grainger and other libraries) and also assist library users in meeting their information needs.

Several example custom metasearch applications are described below. These include a federated OAI-based search/discovery system and three demonstration systems that focus on improved reference services. These three reference applications provide expert system-like searching for: (1) library journal holdings (both print and electronic); (2) the full-text of a specific article; and (3) conference proceeding holdings. All of these modules feature the capability of performing multiple searches within a specific resource in order to arrive at the optimum result set. The paper will conclude with a short discussion of issues and further work needed.

OAI Service Provider

The employment of federated search systems that utilize a common search engine operating over collected, central data such as A & I databases or full-text metadata is becoming more common. For example, the ISI Web of Knowledge and EI Village provide federated simultaneous searching over multiple databases. The Grainger Library has implemented a federated search and discovery service over 47 harvested OAI provider sites in the subject areas of engineering, physics, and computer science. This system includes item-level metadata for over 800,000 documents from the ArX physics preprint server, the NCSTRL computer science report collection, journal articles from IOP (Institute of Physics), NASA and NACA documents, MIT theses and dissertations, California Institute of Technology reports and conferences, a UIUC engineering documents collection database, the Wolfram Research MathWorld and Functions sites, and other OAI provider sites. These metadata records are harvested from OAI provider sites and are indexed in a central search engine. Figure 1 shows the search menu for the Grainger OAI Search System. The Grainger OAI search system and other science and engineering OAI-based aggregation services, such as the National Science Digital Library (NSDL) are described by Brogan.⁷ The records from OAI aggregation sites complement A & I Services and the open Web and provide a rich resource for academic research.

Figure 1. Grainger OAI Search System Menu

In addition, the Grainger Library has for several years provided an engineering simultaneous search system that offers broadcast searching over several A & I Services (Compendex, INSPEC, Current Contents, Applied Science and Technology Index), the statewide Online Catalog, the Grainger Library New Books database, Google, and the Grainger OAI search service. This system helps assist users in selecting the best A & I service for a particular search need and allows the searching of periodical index databases along with the online catalog and a major Web search engine. It is anticipated that the simultaneous search module of the Grainger Search Aid will be replaced by a vendor system that is presently out for bid in a statewide RFP.

Journal Finder

The Journal Finder system is a custom metasearch application designed to facilitate identification of local print and electronic journal holdings by performing asynchronous broadcast searches over several resources. This site is designed for the user seeking to determine if the Library has a print or electronic subscription to a specific known journal. Users with a citation to a specific article (i.e., they have the author, article title, journal title, volume, pages, year, etc.) are asked to go to the Full-Text Article Locator site described below.

The resources searched in the Journal Finder include: the UIUC Library's e-resource registry database (taken from a TDNet data feed), the OPAC (via two searches: a start of title search and a keyword title search limited to serial format), an engineering serials check-in database, a local IEEE and ACM journals holding database, and a difficult-to-find citations database. The e-resource registry includes licensed e-journals that may not be listed in the OPAC or may not have electronic links in the OPAC. The difficult-to-find citations database allows easier identification of titles where the search string may result in high numbers of retrieval matches in an OPAC search.

The Journal Finder and the other applications described here have the ability to interpret the user-entered search argument and then route the query to selected resources based on elements in the search argument. For example, in the Journal Finder and Conference Locator systems, if the user search term includes the word 'IEEE' or 'ACM', the term will be searched against a local IEEE or ACM holdings database. The Journal Finder also takes the user-entered title search string and checks against an abbreviation database at the title level. If the title argument fails to match in the full abbreviations database, each individual word is checked to determine if it is a common abbreviation and, in certain cases, appropriate prepositions or intervening words ('of', 'of the', etc.) are also added. Figure 2 presents the main search menu for the Journal Finder.

Full-Text Finder - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail

Address <http://g118.grainger.uiuc.edu/jnlfinder/jnlsearch.asp> Go Links

Search: Web Reference Shopping Travel Games

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Journal Finder (Print or Electronic) in UIUC Library

To find the full-text of a specific article (if you know author name, journal title, specific volume, pages, year, etc.) [click here.](#)

Look for:

Journal Title
(either title keywords or beginning of title)

OR

Journal ISSN:
International Standard Serial Number

☒ Show Search Results as they Return
☐ Interpret Results for Best Matches

To search for a Conference proceeding, [click here.](#)

Internet

Figure 2. Journal Finder User Entry Screen.

For the displayed results, the Journal Finder offers users a choice of having search results (the number of matches for each search) presented as they are returned or having the aggregate results interpreted and presented in rank order with accompanying explanations. Figure 3 shows the results of an 'Interpret Results for Best Matches' search.

Figure 3. Journal Finder Results Page.

In this example, the journal *Genetic Programming and Evolvable Machines* is not held in print but is available in its entirety through a consortial agreement.

Full-Text Article Locator

The second metasearch system presented here is designed to facilitate access to the full-text of specific journal articles. This metasearch system is similar to modules appearing in local link servers such as Ex Libris' SFX. The full-text article locator site performs simultaneous searches of the CrossRef metadata database to obtain a specific article DOI, the e-resources registry (for journals not covered by CrossRef or for articles in e-journals that are too new or too old for DOIs), and the online catalog for print holdings (and e-resource links if they are present). If the DOI is found, the user is immediately redirected to the article full-text. At UIUC, we regard this as the best option, given the Library's wide e-journal subscriptions.

Figure 4 shows the Full-Text Locator entry menu.

Figure 4. Full-Text Article Locator User Entry Page.

Conference Locator

Holdings for conference proceedings are particularly problematic for reference staff and users to retrieve, in part because of the main entry cataloging issues and the fact that some conferences are cataloged as monographs and some as serials. For these reasons, we have implemented a conference locator metasearch system to assist reference staff and users in locating conference holdings. This conference search system provides broadcast search capabilities that mimic the search behavior of an expert reference librarian. The system performs four different searches of the OPAC (conference name, author, title, and keyword limited to serial), a search of a locally maintained IEEE and ACM publications database, and a search of the EI Village database (over Compendex and INSPEC) to try to identify or verify conference proceeding holdings. Here again, the system offers users the choice of having search results presented as they are returned or interpreted for best match candidates. Figure 5 shows the Conference Locator main menu.

Full-Text Finder - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail

Address <http://g118.grainger.uiuc.edu/linker/confei.asp> Go Links

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Locate a Conference Proceeding in the Online Catalog and in Compendex/INSPEC

A typical conference citation may look like this:
International Truck and Bus Meeting and Exposition, Toledo, Ohio, November 16-19, 1992.

Please fill out the Form below. Include the Conference Name, editor if known, and year of publication.

Conference Name Keywords:

Editor's Last Name (if known):

Number (if applicable), e.g 22, 22nd, etc.

Year:

Place (city is preferable):

Internet

Figure 5. Conference Locator User entry Page.

Issues and Further Work

Broadcast-based metasearch systems operating over A & I Service databases and OPACs have become important tools in the library's public service offerings. All metasearch applications are designed to recommend and rank specific information resources for users by presenting them with the results of multiple searches. This paper reports on a federated OAI search system and three non-A & I broadcast applications that address custom reference and information needs. These applications search relevant information resources in an algorithmic manner similar to the actions of a trained reference librarian. They are designed to facilitate reference work and provide end-users with enhanced access to important and difficult to retrieve information resources such as journal holdings, the full-text of journal articles, and conference proceeding holdings.

As noted in the ALA MOUSS site, "there is a dearth of research and literature published on user satisfaction with library metasearch tools."⁶ We are in the process of performing several user studies based on user interviews and transaction log analysis to try to determine the efficacy of our metasearch systems. There are several technical issues within broadcast search systems that need to be examined, including the need for logging off of resources after search execution to

minimize the number of simultaneous connections, the impact on vendors of multiple metasearch hits when their resource is secondary to the search request, the use of OpenURL or some other metadata search query standard, and the usefulness to users of being able to return to the native interface of the resource being searched. These questions, and many others, are being examined by the NISO Metasearch Initiative.¹⁰

Bibliography

1. Lynch, Clifford. "Digital Collections, Digital Libraries, and Digitization of Cultural heritage Information," Web-Wise 2002: A Conference on Libraries and Museums in the Digital World, March 20-22, 2002. Available at <http://www.ims.gov/pubs/webwise2002/wbws02cp2.htm>.
2. Ma, Wei. "A Database Selection Expert System based on Reference Librarian's Database Selection Strategy: a Usability and Empirical Evaluation," *Journal of the American Society for Information Science and Technology* 53(7): 567-580, 2002.
3. Tennant, Roy. "The Right Solution: Federated Search Tools," *Library Journal* 128(11): 28-30, (June 15, 2003). Available at: <http://www.libraryjournal.com/index.asp?layout=article&articleid=CA302427>
4. Mischo, William H. "Library Portals, Simultaneous Search, and Full-Text Linking Technologies," *Science and Technology Libraries* 20(2/3): 133-147, 2001.
5. Mischo, William H., Habing, Thomas G., and Timothy W. Cole. "Integration of Simultaneous Searching and Reference Linking across Bibliographic Resources on the Web," *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, 2002, Association for Computing Machinery, pp 119-125.
6. Moore, Anne C. "One-Stop Metasearch Tools: Friendly Time Saver or Confusing Quagmire," ALA RUSA MOUSS Section Bibliography, last revised February 2004. Available at: <http://www.ala.org/ala/rusa/rusaourassoc/rusasections/mouss/moussection/mousscomm/cataloguse/2003annualprogrambibliography.htm>.
7. Brogan, Martha L. "A Survey of Digital Library Aggregation Services," The Digital Library Federation and Council on Library and Information Resources, 2003. Available at: <http://www.diglib.org/pubs/brogan/>.
8. Luther, Judy. "Trumping Google? Metasearching's promise," *Library Journal* 128(16): 36-39, (October 1, 2003). Available at: <http://www.libraryjournal.com/index.asp?layout=article&articleid=CA322627>.
9. Richardson, John V., Jr. "Understanding the Reference Transaction: a Systems Analysis Perspective," *College and Research Libraries* 60(3): 211-222, May 1999.
10. National Information Standards Organization (NISO) Metasearch Initiative. http://www.niso.org/committees/MS_initiative.html

WILLIAM H. MISCHO

William H. Mischo is Head of the Grainger Engineering Library Information at the University of Illinois at Urbana-Champaign and Professor of Library Administration.

MARY C. SCHLEMBACH

Mary C. Schlembach is an Assistant Engineering Librarian in the Grainger Engineering Library Information at the University of Illinois at Urbana-Champaign and Associate Professor of Library Administration.