WEB SERVICES APPROACH TO LIBRARY FEDERATED SEARCH: BANGALORE UNIVERSITY ACADEMIC LIBRARY NETWORK(BAL-NET) INITIATIVE

P V Konnur   T A Mohan

Abstract

The objective of this project is to setup a co-operative framework and develop a central index system for accessing the collections of all Affiliated College Libraries of Bangalore University. This is ought to be achieved by applying a new approach for search and retrieval via SRU/W combined with MARC and Dublin core meta-data paradigm; harvest meta-data using Open Archives Initiative – Protocol for Meta-data Harvesting. The project includes a server, client and portal, functionality partly running on Server and client browser resulting in a low implementation barrier, maximum scalability, and browser independence, as well as giving users control over the search interface and what collections to search.

Keywords : Federated Search, Library Networks, Web Services

1.   Introduction

Bangalore University Library has formed a network of all Academic Libraries affiliated to Bangalore University and named it as BAL-NET (Bangalore University Academic Library Network). All academic, scientific, and special libraries will also be networked with Bangalore University Library. The primary objective of this network is to Share functions, resources and services. The primary challenge faced by the university library is:

1.   Build a union catalogue of conventional as well as digital resources alike?

2.   Union catalogue be accessed alike by End users and librarians?

Coordinated by the Bangalore University Library, the project presently consisting of all Constituent Colleges of Bangalore University as Under:

1.   Bangalore University Library

2.   University Vishwaraiah college of Engineering

3.   University Law College

4.   Canara School of Management

5.   University College of Physical Education

In the second phase all the 412 affiliated colleges will be added to the network by creating sub-networks based on disciplines like Management Library Network, Law Library Network etc,. In the final Phase all the Academic and Research Libraries will be added to the Network.

2.   Development of search and retrieval

Although the use of Z39.50 is widespread, it was anticipated that an obvious alternative would be to search using Standardized URL parameters and to provide responses using XML. Z39.50 Implementer's

Group has developed the concept: Search/Retrieve Web Service (SRW). SRW was announced as an alternative to Z39.50 under the umbrella of the Z39.50 international Next Generation (ZiNG). Two access mechanisms were proposed:

1. SRW (Search and Retrieve via the Web), based on the use of SOAP(Simple Object Access Protocol)

2. SRU (Search and Retrieve via URLs), based on the use of URLs.

Both approaches offer a lower implementation barrier than Z39.50 and are more amenable to implementation in modern web based systems.

SRU is the simpler of the two mechanisms. With SRU, a search request takes the form of a 'base-URL' and associated parameters, such as the query, the start record, the maximum number of returned records and the record schema. By varying the base-URL the same SRU search request can be sent to different targets.

The SRW approach uses SOAP as an extra protocol layer; the main difference in the protocol being that a search is defined using XML. SRW can therefore provide for more complex search parameters. An advantage of SRW is that it can be integrated into the service framework provided by SOAP web services. On the other hand, a benefit of SRU is that it is very similar to most web-based search and retrieval applications, with the exception that those interfaces respond with HTML rather than XML.

3. The BAL-NET Portal

The BAL-NET Portal runs in a standard browser using JavaScript and XSLT. A set of service and collection descriptions encoded in XML is loaded into the browser either from a URL or from a local file. An XSLT style sheet presents the collection descriptions to the user and the user selects the services or collections required for the search.

The search query is defined using CQL (Common Query Language) [CQL]. CQL may vary from simply searching for one or more words to complex Boolean expressions using different index sets. An example of a more complex CQL query is:

query=dc.title exact "Christmas Carol" and dc.creator="Charles Dickens"

A search form is used to create the query, which is human readable. The bold words in the example above are CQL keywords and the unbold text is entered by the user. Truncation, Boolean expressions, fuzzy searching can all be expressed in CQL. The prefix "dc." defines an index set. Servers may use different index sets, which are defined together with other information describing the server configuration in a record called the "explain" record.

Using the query, a JavaScript function initiates parallel searching across the services for all the selected collections. The results are received by the browser as ynchronously as independent SRU responses in XML. The XML is transformed using a stylesheet selected by the user. The stylesheets contain the logic to present the XML elements using characteristics from the registry and to offer the user specific functionality for each type of element, such as providing access to digital objects, document ordering, translation and further searching. The detail of this functionality is described later in this article.

4.   OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)

One of the primary objectives of BAL-NETis to build a union catalogue. BAL-NETserver harvests the meta-data of digital as well as conventional resources. The greatest challenges in building an automated union catalogue is the quality of cataloguing at satellite nodes. However intense training as well as awareness has been carried out by Bangalore University Library

The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH in the remainder of this document) provides an application-independent interoperability framework based on metadata harvesting. There are two classes of participants in the OAI-PMH framework:

•   Data Providers administer systems that support the OAI-PMH as a means of exposing metadata

•   Service Providers use metadata harvested via the OAI-PMH as a basis for building value-added services.

Future Scenarios

When SRU is more widely adopted by the Internet community we expect there will be a need for further development. An important feature of the BAL-NETconcept—the combination of SRU, the metadata registry and the implementation of the portal in the user's browser—is to simplify development to meet new functional requirements. We will illustrate this by presenting two scenarios.

5.   Handling extensions to Dublin Core and updates of MARC21

We expect that many applications will use Simple Dublin Core or—to meet specific functional requirements—Dublin Core with local extensions. The schemas defining such extended records are generally unknown to other applications. Client applications using the SRU and OAI-PMH protocol can only request known record schemas from the server and will probably never be aware of the potentially useful extensions to the standard DC record schema.

Now, imagine a scenario in which a client application could request a record that comprises Dublin Core plus unknown local extensions. The target system responds with a message saying, "Here is DC, plus some additional fields." The client (or portal) displays the known terms and indicates that there are also data fields unknown to the client. These fields may be of interest, and the user may want to investigate the exact meaning of these new terms. As a result, he may choose to use these extra fields to trigger extra functionality that is not offered by the original application but which can be realized by selecting another style sheet.

6.   SRU protocol extensions

The second scenario is an example related to the SRU protocol. When a user receives "0 hits" in response to a search, it is often very frustrating, as the user has little help on how to improve the search. Services might provide special responses in this case, such as guidance on similar searches, suggestions for correcting a typing error, a list of index entries close to the search terms, or a list of terms that are similar to the search term based on some matching process. Although SRU contains operations that would allow the portal to investigate the server's capabilities in supporting this type of guidance, it would considerably complicate the portal's functionality (bearing in mind the portal is implemented by a stylesheet and associated JavaScript functions). It would be more helpful if the user could request that the server respond to this situation directly. SRU version 1.1 has an extension mechanism in which extra parameters may be supplied; these are identified by means of the "x-" prefix. A server is free to ignore

them, but servers that do understand such parameters can respond to them. For example, there may be a parameter "x-nohits=dosomething" that requests the server to respond with extra data to assist the user. If there are no records found by the search request, then the response to the search might include a list of index entries close to the search query (a scan operation). Servers that do not support particular parameters can be recognised by the client because they echo these unknown parameters in the search response.

This example shows how SRU can be extended without conflicting with the current standard. When a larger community adopts the extensions, the standard itself can be extended. It allows data and service providers to use the SRU protocol without being limited by the protocol: one can take part in the standard and at the same time add local extensions.

7.   Conclusions

The BAL-NETapproach and technical solution described in this article—consisting of a SRU portal running in the browser, a gateway and a data model based on the concept of Dublin Core and MARC21 Application Profiles—offers a number of advantages compared to more conventional approaches. Using NewGenLib, a product of charitable, non-profitable organization namely Kesavan Institute of Information and Knowledge Management(KIIKM) the above concept is realized. NewGenLib will be supporting more meta-data standards to cater to newer domains. The advantages include scalability, functionality, low barrier of entry into BALNET, and increased control of functionality for users, data providers and service providers. Last but not least, with the BAL-NETapproach there is no longer a need for a central portal. With the combination of concepts like DCX and the SRU extension mechanism, the BAL-NETapproach can undergo further development and could become an example for similar projects.

8.   References

1.   The home page for SRW/U is http://www.loc.gov/z3950/agency/zing/srw/

2.   CQL is fully described at http://www.loc.gov/z3950/agency/zing/cql/

3.   The home page of MARC::Record is http://marcpm.sourceforge.net/

4.   Xsltproc is an application written against two C libraries called libxml and libxslt described at http://xmlsoft.org/

5.   Swish-e is an indexer/search engine application with a C as well as a Perl interface. See: http://swish-e.org/

About Authors

Dr. P.V.Konnur is presently working as Librarian of Bangalore University, Bangalore 560 056. Earlier to this he has also served as University Librarian in Goa University, Goa and as Reader in Dept. of Library and Information Science, Karnatak University, Dharwad.
Email : konnur@bub.ernet.in

T.A. Mohan is presently working as Information Scientist of Bangalore University, Bangalore 560 056. Previously working as Technical Staff, National Aerospace Laboratories, Bangalore 560 017 and served for British Library, Bangalore and Indian Institute of Science Bangalore 560 012
Email : mohan@bub.ernet.in