Improving Visibility of Libraries through SRU

A.K. Tyagi, Madhavi M. Dhanwantari, Aparna Raghuraman, and Priyanka D. Kalbhor

Defence Institute of Advanced Technology (Deemed University), Girinagar, Khadakwasala, Pune-411 025
E-mail: draktyagi@yahoo.co.in

ABSTRACT

Library users nowadays expect digital libraries to be searchable through a single search form. The ultimate goal is to provide a high ranking search quality to enable access to documents containing relevant information from all participating libraries. The Information Centre and Library at the Defence Institute of Advanced Technology (DIAT), Deemed University (DU), at Pune, aims to ensure wide dissemination of knowledge, through an innovative pilot project and achieve a gradual change towards indigenous knowledge storage. The present work is intended to develop a client gateway using search and retrieval via URL (SRU) protocol for searching SRU-compliant databases, keeping in mind trials done in information retrieval in the 1980s and 90s. The client has been developed, tested and implemented successfully with some limitations. Efforts are on to extend the facility for searching DSpace that conforms to industry standard for developing IR's. We are in the process of developing and implementing a server for DSpace to make it SRU compliant. The client part however, has been developed, and tested successfully, though with some limitations.

Keywords: Information retrieval standards, digital libraries, client gateway, databases

1. INTRODUCTION

Over the past few years, significant and important changes have been seen in the field of published knowledge. Using the latest information and communication technology, digital libraries have become a crucial component of global information infrastructure. Scholarly information is now brought out in a variety of forms, formats, and standards in which relevant information is published. It is important to provide challenging opportunities to the library and information professionals, so that they have ample scope to prove their capabilities, both in theory and in practice. The wealth of information in different societies and institutions is managed using heterogeneous systems, e.g., different search engines, database management systems, and platforms. This makes the task of information users tedious as they have to visit different sites and learn different search mechanisms of each system. Therefore, it was strongly felt that there should be a mechanism that can allow users to submit queries through a common search form to various databases and get results in a common format. A number of efforts have been made in this direction and a sincere attempt has been made to find a common solution for the same. We have achieved breakthrough success using information retrieval standards in general, and search and retrieval via URL (SRU) in particular, which can be accessed from our website. This would eventually help information users as well as library professionals in developing local databases by harvesting the already available metadata.

2. INFORMATION RETRIEVAL STANDARDS

DIAT strives to provide network access to information resources, including services as well as documents and data to all users. The main issue was how to integrate and optimise information management so that users can quickly and easily locate what they seek amidst resources numbering in the millions and beyond. Presently, two systems namely, *Suchika* and DSpace are in operation at DIAT. Our users have to visit various databases separately, to gather relevant information. Keeping their problems and requirements in mind, we have attempted to provide a solution in the form of an SRU gateway. The Z39.50 standard was introduced when internet technologies were not so developed and hence implementation was rather difficult. The standard was implemented for a number of database services with drawbacks, which implied that people aware of the importance of information retrieval standards. Z39.50 defines a standard way for two computers to

communicate for the purpose of information retrieval[1]. The computers may be different types of platforms. A Z39.50 implementation enables one particular interface to access multiple systems, providing the end user with nearly transparent access to other systems. The primary goal of the protocol was to allow the user to access remote database records by specifying criteria to identify appropriate records, and then requesting the transmission of some or all of the identified records. This standard is intended for systems supporting information retrieval services for organisations such as information centres, universities, libraries, and union catalogue centres. It addresses connection-oriented program to program communication and does not specify a user interface.

2.1 Search and Retrieve Web Services

Search and retrieve web service (SRW) and search and retrieve URL service (SRU)[2] are web services-based protocols of the ZNG family that are built on the Z39.50 for querying internet indexes or databases and returning search results. Despite the differences in implementation, both SRW and SRU are similar, since they both define a similar set of commands (known as "operations") and responses. SRU is actually SRW without the SOAP-ful service. SRU adds a response schema parameter and is intended for thin clients, where the browser is the application. As such operations sent by SRU clients can only be transmitted via HTTP GET requests. The results of the SRU requests are XML streams.

3. SRU AND SRW-PROBLEMS AND SOLUTIONS

SRW and SRU define a standard platform for Internet search queries as well as the structure of the responses. Even though the queries are the same, the syntax implementing the queries is different. What is worse is the structure of the responses. Each response not only contains the search result but also a lot of formatting. SRW and SRU address these shortcomings by specifying the syntax for the queries and the results. Such specifications open up Internet-accessible search functions and allow for the creation of tools to explore the content of the hidden web more efficiently and effectively. SRW/U allows people and HTTP user agents to query Internet databases more seamlessly without the need of more expensive and complicated meta-search protocols.

3.1 Differences between SRW and SRU

The differences between SRW and SRU lie in the way operations are encapsulated and transmitted between the client and the server as well as how results are returned. SRW is essentially a SOAP-ful web service. Operations are encapsulated by clients as SOAP request and sent to the server. Likewise, responses by servers are encapsulated using SOAP and returned to clients. On the other hand SRU is essentially a REST-ful web service. Parameters are encoded as name/value pairs in the query string of a URL. As such operations sent by SRU clients can only be transmitted via HTTP GET requests. The results of the SRU requests are XML streams, the same streams returned via SRW requests sans the SOAP envelope.

3.2 Search and Retrieve URL (SRU) Service

SRU is a web service-based protocol for querying databases and returning results. Its requests and results are similar to SRW but different in the way that the queries are encapsulated and transmitted between client and server applications.

4. SRU CLIENT GATEWAY

In this study, we have attempted to probe SRU in detail and develop a simple client gateway to search the SRU compliant servers using it. During Dr A.K. Tyagi's Fulbright Scholarship tenure at the Virginia Tech. University, Blacksburg, Virginia, USA, he worked toward developing a client gateway using SRU protocol for searching SRU-compliant databases (Fig. 1).

The effort yielded fruitful results, where Dr Tyagi could search heterogeneous databases including *Suchika*, but only to a certain level. SRU-compliant databases through common search form[3] can be accessed from our website. A new service can also be added and configured using the same website link. As most institutes included are still in the process to make them fully SRU-compliant, the output in some of the databases can only be displayed in XML format. Efforts are still on to apply a common style sheet so that the XML output received from different databases can be uniformly displayed. A number of SRU test servers are available at the official website maintained by the Library of Congress. During this, he found that
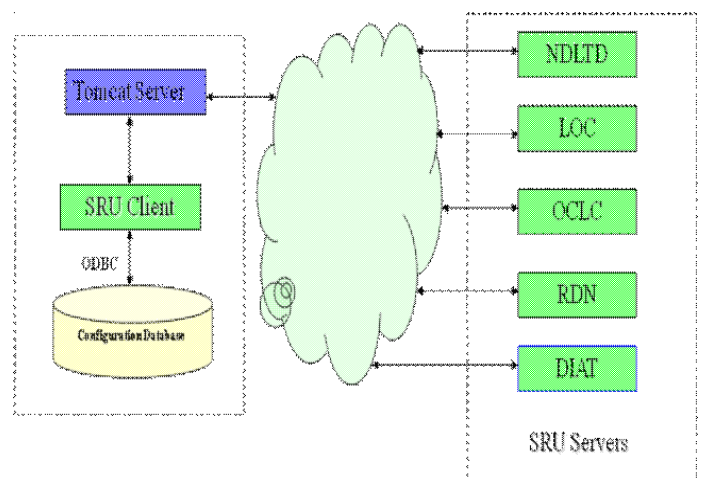

Figure 1. SRU client gateway at DIAT.

some servers worked partially and some not at all may be because these are still in the implementation stage. The comparison of various facilities available has been given in the Table 1.

Table 1. Comparison of facilities and features

Name of the Service	Features available	Record Schema
NDLTD Union Catalog	Title, field, person, subject, date, identifier, description	Dublin Core, Marc XML
Library of Congress	Free text, Title, person, Subject, date, ISBN	Dublin Core, Marc XML
University of Toronto, Canada	Free text, title, person, subject, Date, ISBN	Dublin Core, Marc XML
DIAT SRU Test Server	Title, Person, Subject, ISBN	Dublin Core, Local
The European Library	Title, Person, Subject, Date	Dublin Core
British Library SRU gateway Pilot Service	Title, Person, Subject, Date	Dublin Core
DSpace@DIAT	Title, Person, Subject	Dublin Core

4.1 SRU and *Suchika*

Encouraged by the success of SRU client gateway, we tried to make the *Suchika* server SRU complaint so that it could also be searched using the SRU client gateway. The *Suchika* database contains the bibliographic details of DIAT library holdings. The database is based on the Common Communication Format (CCF) schema. Microsoft Access 2003 has been used as backend database management system. *Suchika* has its own search engine for storing, processing and retrieval of records. It provides the facility to search the database over network using its web OPAC module. As shown in the Fig. 2, when a user from the SRU client gateway sends an SRU request, it is received by the SRU server. It serves the SRU query parameters into local SQL query and passes it on to the search engine. The search engine submits the query to the database and returns the search results. The results are then sent back to the SRU Client as per the SRU parameters for display at the local client machine. We have successfully tested the SRU server and made it available to others for testing SRU
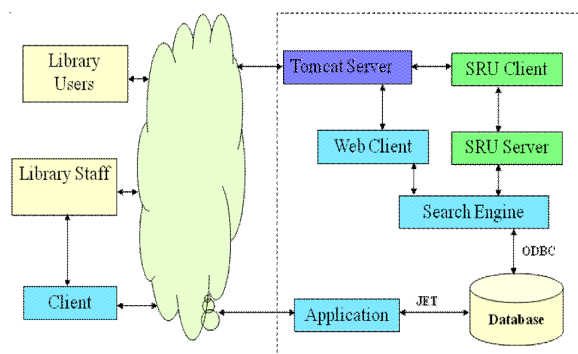

Figure 2. Suchika implementation.

client gateway facilitates like adding a new SRU compliant server[3] and its supported parameters, and using them at the time of search. So, once configured properly, users in a library need not worry about the functions supported by the service. We have not been able to implement the complete protocol particularly when the protocol itself is in the implementation stage. Still, we are in the process of providing a facility to the users for easy retrieval of required relevant information and in addition to share our work. Those who have even a little knowledge of any programming language can easily modify the present SRU client gateway or develop their own client with enhanced features using the available code[4]. Presently almost all the major as well as smaller libraries are computerised. SRU can play a major role in improving their quality of records. The libraries that want to improve their catalogue records can help themselves through this gateway and make full use of this opportunity for further development. The first implementation seen was that records retrieved from an SRU server using SRU clients could be manipulated and converted into local schema and harvested into the local database with little more efforts. This not only helped in improving the quality of records, but also brought consistency in the databases of more libraries around the globe.

The SRU server presently at DIAT has the following limitations:

⌘ It returns the records in Dublin Core Schema in XML packing and does not use any style sheet.

⌘ It supports only three field relations (all, any, and exact).

⌘ It returns zero hits even if there is an error in the query.

⌘ It accepts only words or phrases as search statements. Parentheses in search statements are not supported.

⌘ Only three search statements can be used in a single search. These can be combined using any of the three Boolean operators.

5. RECENT DEVELOPMENTS

At IC&L, attempts are always towards assessing and finding solutions for problems in varying denominations. We set out to re-examine the way in which we delivered almost complete computerised services to our users. Our main objectives were to:

✄ Provide users with dynamic access to the library resources

✄ Offer better navigation of large sets of search results

✄ Enhance the visibility and awareness of our available resources, and

✄ Facilitate the introduction of interoperable services

To ensure that we met the users growing demands and focused on our objectives, a research project titled Interoperable Defence Academic Institutional Repository (IDAIR), was started wherein, all the institutional documents of archival value were digitised and uploaded. We have successfully completed the initial phase of the project that includes the selection and installation of suitable software for institutional repository (IR). Keeping in view the available expertise and by elaborate comparative study, it was decided that DSpace would be suitable in a Windows 2003 environment. Presently, 300+ documents under different communities are available and we aim to upload all the dissertations and research papers (1000+ approx.) into it by the end of current fiscal year. This is presently hosted on a test server[5] and can be accessed using our website. The IR is meant to allow greater participation from the users. They could directly submit their full-text documents in any form and help add to the development of scholarly information. The repository made the storage and retrieval procedure much less strenuous and time consuming, yet it could not act as the single source to fulfill all informational requirements of the user. In case of DIAT, we presently have two sources of information, i.e., *Suchika* database and DSpace. Both are available through different search mechanisms. Hence, the user will have to visit different sites and use different search forms to gather relevant information.

5.1 SRU and DSpace

The success of SRU client gateway and its extension to Suchika greatly boosted our confidence and helped us achieve breakthrough in implementing the same to DSpace (Fig.3). Presently, the IR at DIAT that uses DSpace can be searched using the SRU Client Gateway maintained at DIAT server, however, with some limitations. Now, we can implement a simple search facility in DSpace. With this users can search DSpace@DIAT using the client gateway. However, the results are still displayed in DSpace format.
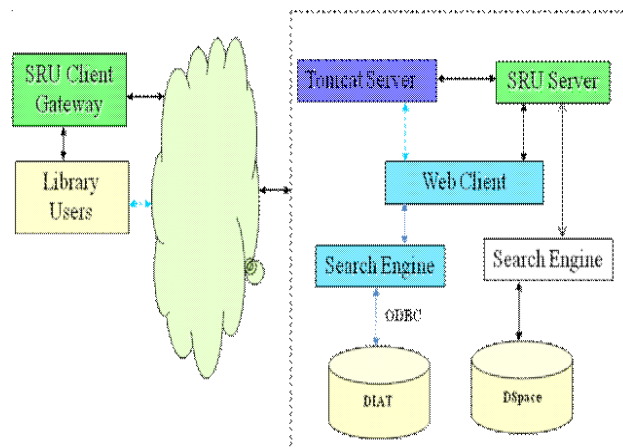

Figure 3. SRU Implementation on DSpace.

6. FUTURE DEVELOPMENTS

The development and implementation of SRU client and server protocol not only cleared our thoughts about the implementation of a protocol, but also gave us the confidence in building the application. We are sure that we shall be able to make it fully searchable using the client gateway and the results will also be displayed in a common format. Furthermore, our attempts are now focused towards finding a single channel to harvest available metadata to form a catalogue. This would approximately involve a period of one year. Whatever the method, our goal continues to be the same; and that is to ensure that we create a culture of lifelong learning and take necessary steps to boost literacy levels in our country.

REFERENCES

1. www.loc.gov/z3950/

2. http://www.loc.gov/standards/sru/

3. http://121.241.65.37/sru/sruServices.jsp

4. http://121.241.65.37/sru.zip

5. http://121.241.65.37/dspace