World Applied Sciences Journal 29 (Computer Sciences, Engineering and Its Applications): 279-284, 2014
ISSN 1818-4952
© IDOSI Publications, 2014
DOI: 10.5829/idosi.wasj.2014.29.csea.2266

# SRW/U as a Lingua Franca in Managing the Diversified Information Resources

*V. Khanaa, K.P. Thooyamani and R. Udayakumar*

School of Computing Science, Bharath University-73, India

**Abstract:** The rapid evolution of information technology and internet has made explosive growth of information resources. The results from different resources are provided in different formats, which make them difficult to search and subsequently merge for providing a federated framework. eXtensible Markup Language (XML) is the 'Lingua Franca': a common language that two computers can use to efficiently speak with one another. Z39.50 protocol was initially developed for bridging the search of the diversified heterogeneous information resources. As the advancement in the technology, Z39.50 protocol has evolved into Search Retrieval Webservices/Uniform resource identifier (SRW/U) protocol. In this paper an SRW/U interface is discussed for providing a common search platform in the diversified heterogeneous applications. The methodology used is the Contextual Query Language (CQL) in the searching technique. The diversified information resources discussed in this paper are KOHA (library automation software), GreenStone Digital Library Software (GSDL) (digital library software) and the pool of information resources created by WAMP (Windows, Apache, Mysql, Php) search.

**Key words:** Z39.50 · SRW/U · CQL · Digital Library · KOHA · GSDL

## I.INTRODUCTION

Due to the explosive growth of information technology and resources, the information are diversified in different formats. Every information has its own format, standard, interface and protocol for its access. Information retrieval the process of finding the required material of an unstructured nature from large collection based on the need of the user Database management systems (DBMS) is an implementation tool for information retrieval functionality [1]. Metadata has been used in various applications as a mean for cataloguing archived information and many search engines have been developed based on metadata. Every search engine has its own metadata format, syntax, interfaces and searching techniques for retrieving the information. In this context, in order to retrieve the required information, the user must be aware of the meta data formats and searching techniques of the respective information resources. This will become the bottleneck and may reduce the information availability. Protocol binding defines mapping between the search interface and the heterogeneous information store. This binding is implemented using the web services. The protocol being used in this paper is SRW/U [6]. SRW/U is based on the services defined over the Hyper Text Transport Protocol (HTTP) binding. Simple Object Access Protocol (SOAP) is used for information exchange in the web service defined by SRW/U. [2] This paper focuses the informational resources from the heterogeneous diversified applications and the customized repositories of Scientific Information and Research Development (SIRD) developed with WAMP and provide a common search gateway for federating these resources.

**Search and Retrieve Protocols:** A search protocol is a series of messages between a client (called user) and a server (called heterogonous application). It is generally used for the management of the messaging and the retrieval based on the query that expresses the search criteria. The messages provide the context for the query, information on user preferences, information about handling of the result and the vehicle to convey the records retrieved. The advantages of using a search-and-retrieve protocol are rich enough to have a semantic interpretation induced on it. Although HTTP is

**Corresponding Author:** V. Khanaa, School of Computing Science, Bharath University-73, India.

standardized and can be used for searching and retrieval, it does not have the standard semantics attached to the Uniform Resource Locator (URLs) requested and the data POSTed or the documents returned. What is needed is a protocol which also specifies standard semantics for structured queries and responses. At present, there are two realistic contender protocols for this role: ANSI/NISO Z39.50 [3] and SRW/U [4].

**Z39.50:** American National Standards Institute/National Information Standards Organization (ANSI/NISO) Z39.50 defines a standard way for two computers to communicate for the purpose of information retrieval. Z39.50 makes it easier to use large information databases by standardizing the procedures and features for searching and retrieving information. Figure 1 shows the protocol architecture of Z39.50 where Z39.50 supports information retrieval in a distributed, client and server environment where a computer operating as a client submits a search request (i.e., a query) to another computer acting as an information server. Software on the server performs a search on one or more databases and creates a result set of records that meet the criteria of the search request [7][8]. The server returns records from the result set to the client for processing. The power of Z39.50 is that it separates the user interface on the client side from the information servers, search engines and databases. Z39.50 provides a consistent view of information from a wide variety of sources and it offers client implementers the capability to integrate information from a range of databases and servers [9].

Even though Z39.50 protocol have seen to be the wise search and retrieve protocol, this protocol standard is seen as complicated and difficult to implement[10]. Z39.50 requires connection-based sessions, it uses binary encoding.Z39.50 interoperates with the web but is not yet fully compatible according to some definitions. While interoperating with the web, it is possible for Z39.50 to deliver the following error,

- "Cannot accept the query with those particular attributes or combination of attributes".
- "Cannot return records in the format required".

**ZING:** ZING (Z39.50 International: Next Generation) is the umbrella name used to describe several experiments. It should be strongly stressed, however, that ZING is currently only a set of experiments intended to get some experience with what future versions of Z39.50 might look like and its functionality [11]. The ZING Experiments are

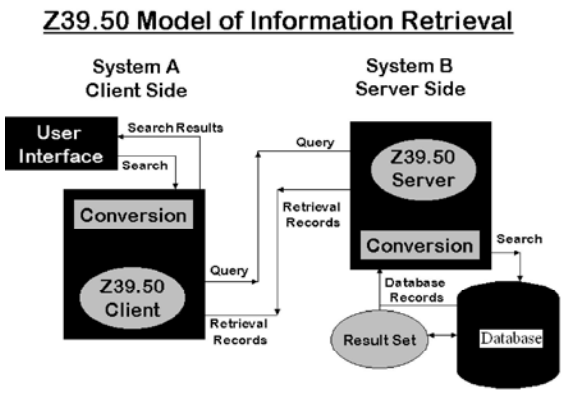


Fig. 1: Z39.50 Information retrieval

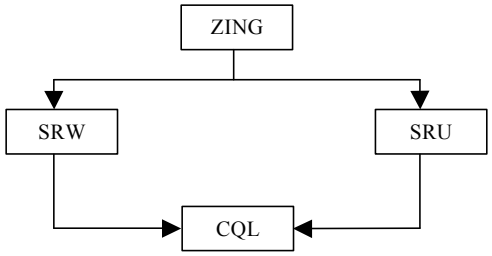


Fig. 2: ZING Categorization

- Search/Retrieval Web Service (SRW)
- Search/Retrieve URI (Uniform Resource Identifier) Service (SRU)

The Figure 2 shows the ZING experiments categorization,

**SRW/U PROTOCOL:** SRW uses SOAP to deliver structured query payloads from client to server and responses including zero or more records from server to client. The requests and responses are both expressed in eXtensible Markup Language XML. Several operations are defined, each consisting of a request-response pair [12][13]. These including "search Retrieve", "explain" (in which a client asks a server to describe its capabilities) and "scan" (for browsing index entries).

Table 1: Simple Search Request Retrieval from SRW:

```
<searchRetrieveRequest>
 <version>1.1</version>
 <query>dc.title all "IEEE transanctions
 on computers" </query>
 <maximumRecords>1</maximumRecords>
 <startRecord>1</startrecord>
 <recordSchema>dc</recordSchema>
 </searchRetrieveRequest>
```

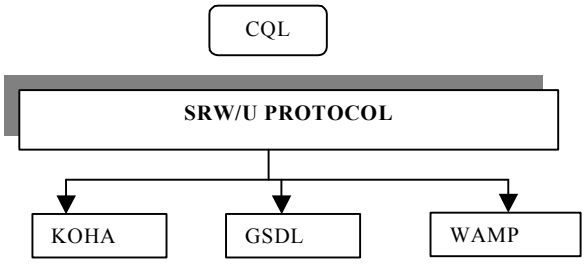


Fig. 3: Architecture of SRW/U protocol

SRU is semantically equivalent to SRW, but uses a simpler mechanism as its transport. SRU requests are expressed as URLs with query parameters that carry information equivalent to that in the corresponding SRW request XML documents. SRU response payloads are identical to those of SRW, but are returned directly as the content of the HTTP response rather than being wrapped in a Simple Object Access Protocol (SOAP) envelope as in SRW.

Table 3: Example of SRU

| |
|---|
| *http://z3950.loc.gov:7090/voyager?version=1.1 and operation=searchRetrieve and query=digitallibrary and maximunRecords=1 and recordSchema=dc* |

The TABLE III illustrates the search for the term "digitallibrary", requesting that at most one record be returned, according to the 'dc' schema in SRU format.

The Figure 3 illustrates the architecture of SRW/U protocol for accessing the heterogeneous application viz., KOHA, GSDL and WAMP search.

**CQL:** CQL[8] stands for Common Query Language. It is a formal language for representing queries to Information Retrieval systems such as web indexes, bibliographic catalogues and museum collection information. It is being developed by theZ39.50 Maintenance Agencyas part of itsZINGinitiative [14]. Based on thesemanticsof Z39.50, its design objective is that queries be human readable and writable and that the language be intuitive while maintaining the expressiveness of more complexquery languages. It is being developed and maintained by the Z39.50 Maintenance Agency, part of the Library of Congress.

A CQL query consists of either a single search clause [example 1], or multiple search clauses connected by boolean operators [example 2]. It may have a sort specification at the end, following the 'sortBy' keyword [example 3].

Table 5: Example of CQL

| | |
|---|---|
| 1. | dc.title any *IEEE journal* |
| 2. | dc.title any *IEEE journal* or dc.creator *any sanderson* |
| 3. | dc.title any *IEEE journal* sortBy *name* |

**Library Automation Softwares:** Libraries have been looking forward for the better technologies even before the onset of the computers. The introduction of the type writer into libraries was a revolutionary concept in the late 1800's. Later stages of modernization witnesses the introduction of unit record equipment, the move of offline computerization, use of online systems. By the mid 1960's, computers were being used for the production of Machine Readable Catalogue Records (MARC) [9] by the Library Of Congress (LOC). Between 1965 and 1968, LOC began the MARC I project, followed quickly by MARC II. MARC was defined as way of "tagging" bibliographic records using 3-digit numbers to identify fields. In 1974, the MARC II format became the basis of a standard incorporated by NISO. This was a significant development because the standards created meant that a bibliographic record could be read and transferred by the computers between different library systems.

A digital library is a library in which collections are stored in digital formats and accessible by computers. The digital content may be stored locally or accessed remotely via computer networks. A digital library is a type of information retrieval system and the software used to enable this functionality in DLS. There are many open source DLS available namely GSDL and Dspace.

**Overview of Koha:** Koha [5] is the first open source Integrated Library System (ILS).In 1999 when the Horowhenua Library Trust (HLT) in New Zealand, was looking for a Year 2000 (Y2K) compliant replacement for their library system, Katipo Communications proposed a new system, using open source tools to be released under the GPL. Koha (the Maori word for 'gift' or 'donation') went live at HLT in January 2000 and was the world's first open source ILAP and is distributed under GNU GPL license. Latest version is Koha-3.0.2 (Linux platform only) and Koha 2.2.9 (for Windows and other platforms) (http://koha.org). It runs on different platform like Linux, MacOSx, FreeBSD, Solaris and Windows. Originally developed on the Linux operating system, is written in Perl, uses Apache web server, has better support for multi-RDBMS Relational DataBase Management System like MySQL, PostgreSQL. OPAC interface is in Cascading Style Sheets (CSS) with eXtensible HyperText Markup Language (XHTML). It supports all major library

standards such as MARC record import/export, Z39.50 and SRW/U feature. Records are stored internally in an Standard Generalized Markup Language -like format and can be retrieved in MARC XML, Dublin Core (DC), MODS, Really Simple Syndication (RSS), Atom, SRWDC and the OPAC can be used by citation tools such as Zotero. Koha's default installation supports running Zebra which is configured to support SRU queries on bibliographic and authority data. Zebra itself is capable of detecting Z39.50 or HTTP and responding with SRU if the incoming request is HTTP.

**Overview of GSDL:** Greenstone[6] is a suite of software tools for building and distributing digital library collections on the Internet or CD-ROM. It is open-source, multilingual software, issued under the terms of the GNU General Public License. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato and has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium.

Greenstone may be used to create large, searchable collections of digital documents. In addition to command line tools for digital collection building, Greenstone has a graphical Greenstone Librarians Interface (GLI) used to build collections and assign metadata [15].

Through user selected plugins, Greenstone can import digital documents in formats including text, html, jpg, tiff, MP3, PDF, video and Word, among others. The text, Portable Document Format (PDF), HTML and similar documents are converted into Greenstone Archive Format (GAF) which is an XML equivalent format.

**Implementation of SRW/U AT SIRD:** To perform the unified search, first the heterogeneous environment must be created, here the heterogeneous environment is created with the help of KOHA, GSDL and WAMP[7] and by integrating all these applications, the full text search option is provided in KOHA. To implement the full text search and retrieval feature in kohav2.2.9 (KOHAV2.2.9 is the stable release for windows), the software is downloaded from www.koha.org and installed in windows platform. Now, the downloaded KOHA contains four components viz., Online Public Access Catalogue (OPAC), Intranet, Daemons and Database. OPAC is used to search and locate library holdings. There are two components in the OPAC package viz., the main-opac and the html-opac. The Main-OPAC contains all the Perl scripts and the HTML-OPAC contains all the HTML templates for all the HTML pages. In OPAC, search can

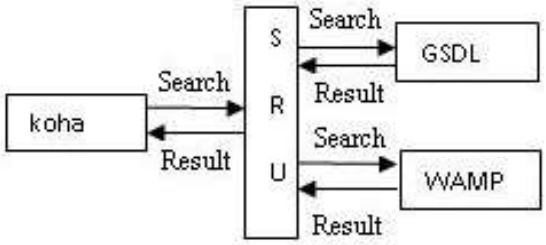


Fig. 4: Implementation of SRU with GSDL and WAMP

be done in different ways: advance search, search by subject and basic search. Koha allows field based searching by making use of field tags such as Title, Author, ISBN. Intranet is the back office and the front desk side of the system. There are 9 components in the intranet package. The main intranet component contains all the main Perl scripts to handle the navigation, login, logout and provide connection to the other components. The HTML-intranet component contains all the HTML templates for all pages for the intranet side of the system. The navigation of all these applications with the help of SRU protocol is illustrated in the Fig 4.

GSDL v2.83 is downloaded from www.greenstone. org and is also installed. After indexing for each collection (a set of documents form a collection) 9 folders are created viz., archives, building, index, etc, import, perllib, images, macros, tmp. The documents to be indexed are kept in import. The collect.cfg file in etc folder determines the look and feel of the collection.

WAMP server 2.1 is downloaded from www.wamp server.com/en/. The downloaded package contains Apache 2.2.11,Php 5.3.0 MySQL 5.1.36. SIRD makes use of WAMP for creating institutional repositories. Now this WAMP search must be indexed from KOHA.

The integration of these applications is done by providing full text search option and wamp indexing in koha. The results from GSDL and WAMP will then be displayed in OPAC of KOHA. For this purpose,six different perl scripts are written, viz, fulltextsearch.pl, fulltextsearch1.pl, wampsearch.pl, fulltextsearch.tmpl, fulltextsearch.tmpl and wampsearch.tmpl. In fulltextsearch.pl and in wampsearch.pl query term is obtained from the user and passed to GSDL and WAMP respectively.

Query string passed through SRU from koha to GSDL is as follows:

First part: (GSDL Location)

Table 5: GSDL location

| http://localhost:1025/gsdl?uq=11540421 |
|---|

Second part: (Collection Name)

Table 6: Collection name

| e=p-01000-00---off-0koha--00-1--0-10-0---0---0prompt-10---4-------0-1l--11-en-50---20-about---00-3-1-00-0-0-11-1-0utfZz-8-00 |
|---|

Third part: (Query Term)

Table 7: Query Term

| a=q and r=1 and hs=1 and fqf=TX and t=0 and q=java |
|---|

Fourth part: (Do option ): perform

The above mentioned table V,VI,VII, are the parts url of GSDL which is passed to koha and the full text search option of GSDL is accessed from koha[10] by changing the coding mentioned in table VIII.

Table 8: Koha Modified HTML code

```
<form name="QueryForm1" method="get"
action=”http://localhost:1025/gsdl?uq=11540421”
id="searchform">
<input type="hidden" name="a" value="q">
<input type="hidden" name="r" value="1">
<input type="hidden" name="hs" value="1">
<input type="hidden" name="e" value="p- 01000-00---off-0koha--00-1--0-10-0---0--0prompt- 10---4--0-1l--11-en-50--20-about---00-3-1-00-0-0- 11-1-0utfZz-8- 00”>
<input type="hidden" name="fqf" value="TX">
<input type="hidden" name="t" value="0">
<input type="hidden" name="q" value="">
</form>
```

Query string passed through SRU from koha to WAMP is as follows:

First part: (WAMP Location)

Table 9: WAMP Location

| http://10.1.6.153/koha/search.php |
|---|

Second part: (Keyword Searched)

Table 10: Keyword Searched

| Search = “keyword searched” |
|---|

The WAMP search is integrated with KOHA by modifying the necessary perl scripts of KOHA. For instance, the integration of WAMP search with KOHA in SIRD is given in the table XI.

Table 11: WAMP modified code

```
#!/usr/bin/perl
use strict;
use warnings;
use CGI;
my $cgi = new CGI;
my @params = $cgi->param();
my $value;
my $q = CGI->new( );
foreach my $parameter (sort @params)
{ print $q-> redirect('http://10.1.6.153/dc_fed/koha/search.php?searchh='
. $cgi->param($parameter));}
```

**Implementation of SRW AT SIRD:** SRW uses SOAP to deliver structured query payloads from client to server and responses including zero or more records from server to client. SOAP is an XML-based messaging protocol. It defines a set of rules for structuring messages that can be used for simple one-way messaging but is particularly useful for performing RPC-style (Remote Procedure Call) request-response dialogues. The web services are defined by the Web Service Description Language (WSDL), here the client server, request – response are defined in XML, The SRW service is used by the WAMP search to retrieve and to access the CQL query. The php script is written for accessing the SOAP message.

In the SRW implementation, WSDL file is created for describing the services between client and the server. The WSDL script contains the port bindings between the client and server. The client requests for the service, the WSDL describes the services and passes the request to the server. The server responds for the requests and displays the result to the client. The WSDL file specifying the web service operation is shown in the table IX, the following WSDL file is accessed by the WAMP search by the method of RPC with the help of php script.

## CONCLUSION

The retrieval system presented in this paper conceals the underlying details of the querying process through the user interface. Users do not need to know where the databases reside, the structure of the system, Database Management System used to build databases and the programming behind it. The information retrieval is similar to a system accessing a local database. Besides that, users do not need much effort in understanding the functionality of the information retrieval system, thus will suit a wide spectrum of users. The query was also designed to support search from multiple heterogeneous databases simultaneously. Heterogeneous here means the databases are developed using a variety of Database Management Systems and in heterogeneous data format.

```
<message name='bookRequest'>
 <part name='symbol' type='xsd:string'/>
</message>
<message name='bookResponse'>
 <part name='Result' type='xsd:string'/>
</message>

<portType name='booksearchPortType'>
 <operation name='book'>
 <input message='tns:bookRequest'/>
 <output message='tns:bookResponse'/>
 </operation>
</portType>
```

The information retrieval was developed to retrieve the library collections like books, journals, audio and video from various digital libraries like koha, GSDL and from WAMP search.

## REFERENCES


1. Frank, P. McCreedy and David B. Marks, 2009. "The Naval Research Laboratory's ongoing implementation of the Open Geospatial Consortium's Catalogue Services specification", ieee transactions on oceanic engineering, 26-29 Oct. 2009, ISBN: 978-1-4244-4960-6.
2. Mike Taylor and Marc Cromme, 2005. "Searching Very LargeBodies of Data Using A Transparent Peer-To-Peer Proxy" proceedings of the 16th international workshop on database and expert systems applications(dexa'05), 22-26 Aug. 2005.
3. The ANSI/NISO Z39.50 Protocol: Information Retrieval in the Information Infrastructure http://www.cni.org/pub/NISO/docs/Z39.50-brochure/50.brochure.toc.html.
4. Library of congress, "SRU:search/retrieval via url website: http://www.loc.gov/zstandards/ sru/simple. html.
5. koha, 1999. "koha:features", website: www.koha.org
6. Greenstone Digital Library Software, 2005. "about greenstone", website: http://www.greenstone. org.
7. Anuradha, K.T., 2009. "sru/w: a digital 'lingua franca' in bridging heterogeneous applications: a case study", national conference on recent advances in information science and technology, December 29-30, 2009.
8. CQL standards: http://www.loc.gov/cql/
9. The home page of MARC:: Record is http://marcpm.sourceforge.net/
10. Anuradha, K.T. and R. Sivakaminathan, 2009. "Enhancing Full Text Search Capability in Library Automation Package: A Case Study with Koha and Greenstone Digital Library Software" proceeding of 2009 international conference on computer science and information technology, October 9-11, 2009, ISBN: 979-1-84626-xxx-x.
11. MARC21 Standards. http://www.loc.gov/marc.
12. Thooyamani, K.P., V. Khanaa and R. Udayakumar, 2013. Detection of Material hardness using tactile sensor, Middle-East Journal of Scientific Research, ISSN:1990-9233, 15(12): 1713-1718.
13. Thooyamani, K.P., V. Khanaa and R. Udayakumar, 2013. Blue tooth broad casting server, Middle-East Journal of Scientific Research, ISSN:1990-9233, 15(12): 1707-1712.
14. Saravanan, T., V. Srinivasan and R. Udayakumar, 2013. Images segmentation via Gradient watershed hierarchies and Fast region merging, Middle-East Journal of Scientific Research, ISSN:1990-9233, 15(12): 1680-1683.
15. Udayakumar, R., V. Khanna, T. Saravanan and G. Saritha, 2013. Cross Layer Optimization For Wireless Network (Wimax), Middle-East Journal of Scientific Research, ISSN: 1990-9233, 16(12): 1786-1789.