The Potential for Web Services to Enhance Information Access to Legacy Data: An Exploratory Study and Application

Fatih Oguz
Research Associate, Texas Center for Digital Knowledge
University of North Texas
PO Box 311068 Denton, TX 76203
foguz@lis.admin.unt.edu

William E. Moen
Associate Professor, Texas Center for Digital Knowledge
University of North Texas
PO Box 311068 Denton, TX 76203
wemoen@unt.edu

Abstract

This paper presents an overview of an exploratory research project to identify, describe, and investigate the applicability of the Web services (WS) approach to access legacy data. In the Z Texas Implementation Component of the Library of Texas (ZLOT) project, the ZLOT technical team has implemented a multi-purpose Texas Library Directory Database (TLDD) that is used as a back-end database to support the Library of Texas (LOT) Resource Discovery Service (RDS). The researchers developed and implemented a prototype WS application to show how a legacy system can be accessed and its data can be searched and retrieved. This study focused on understanding how requests and responses between software applications are encoded in Extensible Markup Language (XML).

Introduction

The TLDD is a robust MySQL relational database that includes vital information about Texas academic and public libraries. It would be useful if the TLDD can be made available to the Texas library community in a way that librarians can query the database and get results in structured XML documents for reuse in other applications. Currently this service is not available. A WS application appears to offer a potentially useful approach for such interaction with the TLDD.

The RDS was purchased by the Texas State Library and Archives Commission (TSLAC) and source code is available to TSLAC. This allows reuse of the RDS by local libraries to provide a single search interface to locally licensed resources and other distributed or local resources. The main problem, however, is how locally installed RDS applications can access the TLDD which currently powers the LOT RDS. One solution is to have the local libraries access the TLDD via SQL calls to the database. However, the TLDD has security requirements that can limit table level access.

The assumption for this proposed project is that a WS application can offer a reliable, flexible, and standards-based solution for accessing to the TLDD by the library community to search and retrieve structured and reusable data.

The elegant simplicity and flexibility of the XML made it a definitive standard for data transmission and storage. XML is an open standard and can be accessed and processed by any tool capable of reading and writing American Standard Code for Information Interchange (ASCII) text.

XML, Simple Object Access Protocol (SOAP) and Web Services Description Language (WSDL) are emerging tools to create a WS. WS provide a framework for creating the next generation of distributed systems by which organizations can encapsulate existing business processes, publish them as services, search for and subscribe to other services, and exchange information throughout and beyond the enterprise. Besides recognizing heterogeneity of networked resources and applications as a fundamental ingredient, WS are independent of platform and the development environment can be packaged and published on the Internet. Also WS enable just-in-time integration and interoperability of legacy applications.

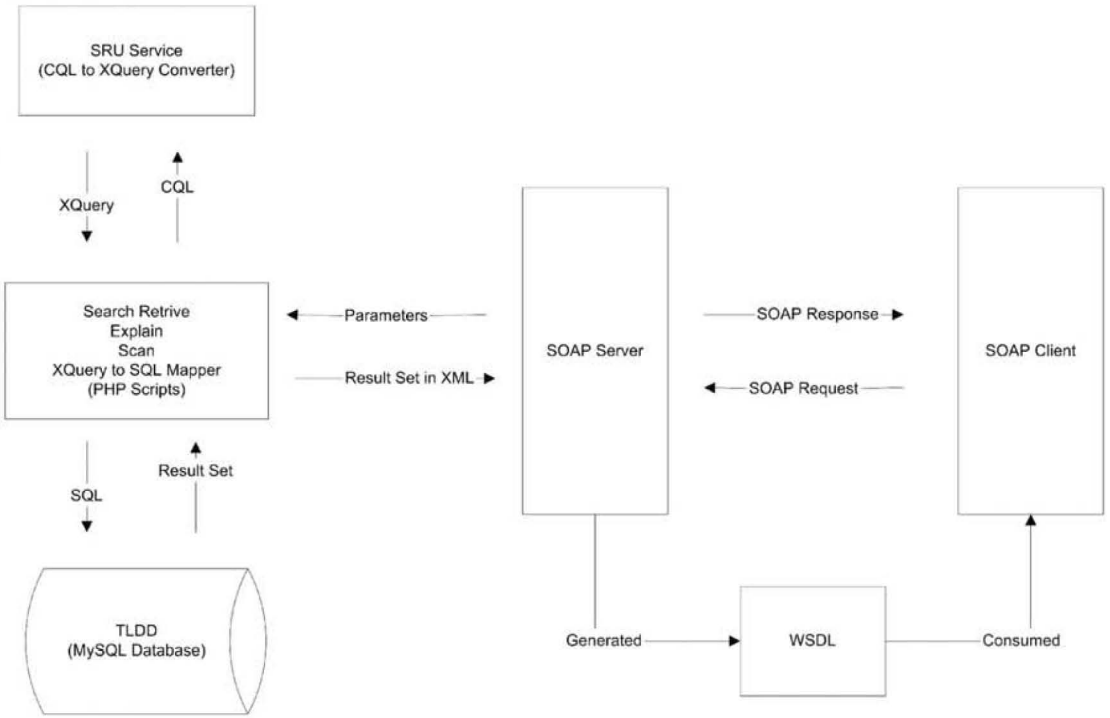
Search/Retrieve Web service (SRW) is a WS standard to perform searches and other information retrieval operations on the Internet using SOAP and XML Path Language (XPath), for example. SRW has been built on 20 years of experience with the Z39.50 information retrieval protocol. SRW uses various schemas such as Dublin Core and it allows developers to define their own schemas as well. SRW enables developers to implement a standards-based search interface to information retrieval systems easier than with the more complex Z39.50. SRW uses a query language called Common Query Language (CQL) that offers simplicity and intuitiveness of Common Command Language (CCL) as well as power and expressiveness of Structured Query Language (SQL). CQL is a formal query language to express searches on Web indexes, bibliographic catalogs and museum collection information. SRW version 1.1 has been released with CQL version 1.1 in 2004.

Researchers adopted the SRW 1.1 protocol in the implementation of Texas Library Directory Web service.

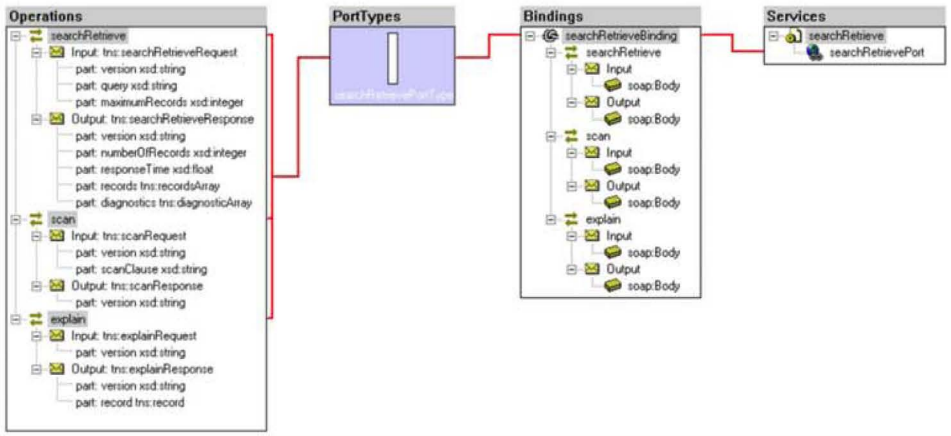
Web Services are software systems designed to support interoperable machine-to-machine interaction over a network by using XML for sending and receiving messages.

SOAP is an XML based communication protocol that enables different programs running on different platforms able to communicate.

WSDL is an abstract description of a WS in a machine processable format.



Schema/WSDL View



SearchRetrieve SOAP Request

```
<SOAP-ENV:Envelope xmlns:SOAP-
ENV="http://schemas.xmlsoap.org/soap/envelope/" xmlns:SOAP-
ENC="http://schemas.xmlsoap.org/soap/encoding/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
    <SOAP-ENV:Body>
        <m:searchRetrieve
xmlns:m="http://web2.unt.edu/tsa/WS/SRW/:SRW" SOAP-
ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/">
            <version xsi:type="xsd:string">1.1</version>
            <query xsi:type="xsd:string">ds.city=denton not
ds.organization=public</query>
            <maximumRecords
xsi:type="xsd:integer">10</maximumRecords>
        </m:searchRetrieve>
    </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

SearchRetrieve SOAP Response

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<SOAP-ENV:Envelope SOAP-
ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:SOAP-
ENC="http://schemas.xmlsoap.org/soap/encoding/"
xmlns:si="http://soapinterop.org/xsd"
xmlns:ds="http://web2.unt.edu/tsa/WS/SRW/"
xmlns:tns="http://web2.unt.edu/tsa/WS/SRW/">
    <SOAP-ENV:Body>
        <ns1:searchRetrieveResponse
xmlns:ns1="http://web2.unt.edu/tsa/WS/SRW/:SRW">
            <version xsi:type="xsd:string">1.1</version>
            <numberOfRecords
xsi:type="xsd:integer">2</numberOfRecords>
            <records xsi:type="SOAP-ENC:Array" SOAP-
ENC:arrayType="tns:records[2]">
                <item xsi:type="tns:records">
                    <recordPacking
xsi:type="xsd:string">XML</recordPacking>
                    <recordSchema
xsi:type="xsd:string">http://web2.unt.edu/tsa/WS/SRW/DS.xs
</recordSchema>
                    <recordData xsi:type="tns:dsData">
                        <orgID xsi:type="xsd:integer">359</orgID>
                        <orgName xsi:type="xsd:string">Texas
Woman&apos;s University - Blagg-Huey Library</orgName>
                        <orgType xsi:type="xsd:string">Academic
Library</orgType>
                        <orgAddress xsi:type="xsd:string">P. O. Box
425528</orgAddress>
                        <orgCity xsi:type="xsd:string">Denton</orgCity
                        <orgZip xsi:type="xsd:string">76204</orgZip>
                        <orgPopulationServed
xsi:type="xsd:string">7887</orgPopulationServed>
                    </recordData>
                    <recordPosition
xsi:type="xsd:integer">1</recordPosition>
                </item>
                <item xsi:type="tns:records">
                    <recordPacking
xsi:type="xsd:string">XML</recordPacking>
                    <recordSchema
xsi:type="xsd:string">http://web2.unt.edu/tsa/WS/SRW/DS.xs
</recordSchema>
                    <recordData xsi:type="tns:dsData">
                        <orgID xsi:type="xsd:integer">390</orgID>
                        <orgName xsi:type="xsd:string">UNT
Libraries</orgName>
                        <orgType xsi:type="xsd:string">Academic
Library</orgType>
                        <orgAddress xsi:type="xsd:string">P.O. Box
305190</orgAddress>
                        <orgCity xsi:type="xsd:string">Denton</orgCity
                        <orgZip xsi:type="xsd:string">76203</orgZip>
                        <orgPopulationServed
xsi:type="xsd:string">27858</orgPopulationServed>
                    </recordData>
                    <recordPosition
xsi:type="xsd:integer">2</recordPosition>
                </item>
            </records>
            <diagnostics xsi:nil="true"/>
        </ns1:searchRetrieveResponse>
    </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```