



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

Fatih Oguz
Assistant Professor
Valdosta State University
1500 N. Patterson St. Valdosta, GA 31698
foguz@valdosta.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-protocol 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 that shows 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 their applications. Currently this service is not available.

When data can be exposed in XML, it appears that WS provide an opportunity to expose legacy data to applications and services. In the context of this project, legacy data are defined as those that need to be converted to a different format when another application or service will ingest, consume, or otherwise use the data. For this project, a relational database containing directory information about Texas academic and public libraries was used as a target for the implementation of Search and Retrieve URL/Web Service (SRU/SRW) which is a WS for search and retrieval based on XML semantics. The target database, the TLDD, was developed as part of a larger application, but it contains useful information about Texas libraries that could be reused and repurposed if the TLDD information could be provided in a standard structure using XML. The goal of this project was to investigate the extent to which the SRU/SRW WS could be used to search the relational database and return records formatted according to an XML schema to enable reuse of those data.

The Texas State Library and Archives Commission (TSLAC) developed and implemented a metasearch application called the LOT in 2002-2003. The LOT comprises several basic components. One component is the Resource Discovery Service (RDS) that provides users a mechanism to search multiple online library catalogs, commercial databases, and other digital resources from a single search interface. Another component is the TLDD, which contains directory information about Texas libraries. Data from the TLDD is used by the RDS to dynamically customize the RDS search interface. The TLDD is a robust and complex MySQL database. TSLAC contracted with Index Data (<http://www.indexdata.com>) to implement the RDS using open source tools and code. The RDS source code is available to TSLAC, which is making it available to Texas libraries that choose to use it as a single search interface to locally licensed resources and other distributed or local digital resources. However, since the TLDD was a key component for powering the RDS, local libraries wanting to use the RDS faced the challenge of how their locally installed RDS applications could access the TLDD in a standardized and interoperable manner. Our project assumed that a WS application could offer a reliable, flexible, and standard-based solution for accessing the TLDD, not only by local RDS applications but also by the library community, to search and retrieve structured and reusable data. Specifically, if the TLDD data could be presented under other applications in an XML format, reuse and repurposing of the TLDD data could be enhanced. The SRU/SRW WS would provide the means for querying the TLDD, and a XML schema specific to the TLDD data would provide a standard structure for encoding the data to return to processing applications.

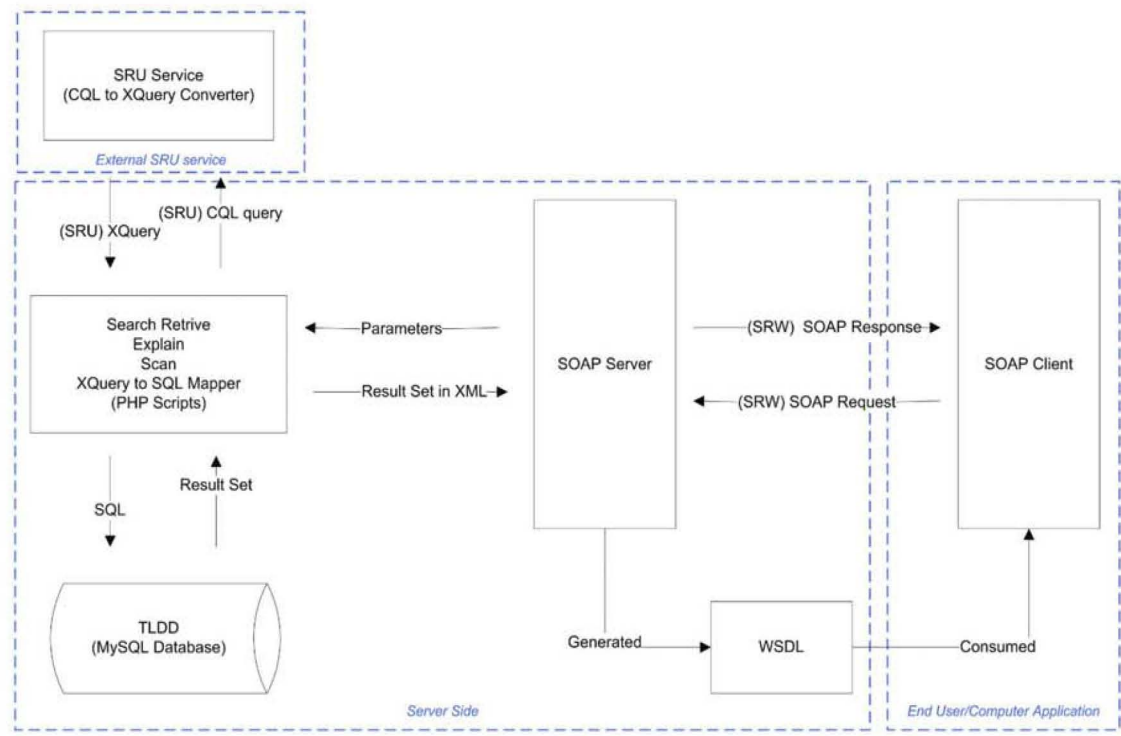
WS enables developers to implement a standardized search interface to information retrieval systems easier than with the more complex Z39.50. SRU/SRW uses a query language called Common Query Language (CQL). CQL is a query language used to express queries on Web indexes, bibliographic catalogs and museum collection information.

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

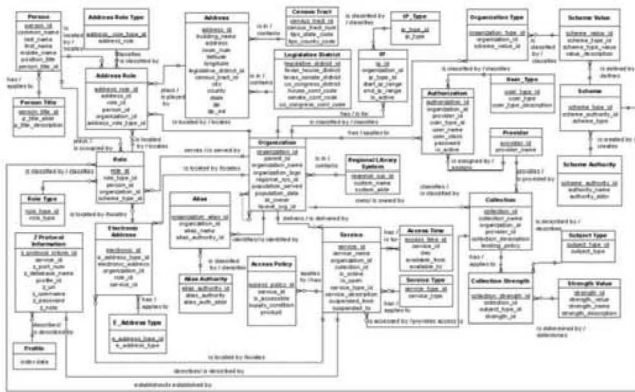
Web Services are software systems designed to support interoperable machine-to-machine interaction over a network by using XML for sending and receiving messages.

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.



Texas Library Directory Database (TLDD)



Library of Texas (LOT) Resource Discovery Service (RDS)



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/" SOAP-ENC:arrayType="xsd:string" xmlns:SOAP-ENC="http://schemas.xmlsoap.org/soap/encoding/">
      <version xsi:type="xsd:string">1.1</version>
      <query xsi:type="xsd:string">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-ENC:arrayType="xsd:string" xmlns:SOAP-ENC="http://schemas.xmlsoap.org/soap/encoding/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:SOAP-ENC="http://schemas.xmlsoap.org/soap/encoding/" xmlns:si="http://soainterop.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/" SOAP-ENC:arrayType="tns:records[]">
      <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[]">
        <record xsi:type="tns:records">
          <recordSchema xsi:type="xsd:string">XML</recordSchema>
          <recordData xsi:type="tns:recordsData">
            <recordID xsi:type="xsd:integer">359</recordID>
            <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>
            <recordPosition xsi:type="xsd:integer">1</recordPosition>
          </recordData>
          <recordPacking xsi:type="xsd:string">XML</recordPacking>
          <recordSchema xsi:type="xsd:string">http://web2.unt.edu/tsa/WS/SRW/DS.1.1</recordSchema>
          <recordData xsi:type="tns:recordsData">
            <recordID xsi:type="xsd:integer">390</recordID>
            <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>
            <recordPosition xsi:type="xsd:integer">2</recordPosition>
          </recordData>
          <diagnostics xsi:nil="true"/>
        </records>
      </SOAP-ENV:Body>
    </ns1:searchRetrieveResponse>
  </SOAP-ENV:Envelope>
```

