

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

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 client server networks were dominant in the 1980s and early 1990s are evolving into a new kind of network built around the Web using open protocols (Coyle, 2002). Web services exchange data between a server and a client using a standard XML format. According to Coyle (2002), the WS technical infrastructure ensures that services from different vendors will interoperate to complete one or more business processes.

The Merriam-Webster dictionary defines legacy as “something transmitted by or received from an ancestor or predecessor or from the past” (“Merriam-webster online”, n.d.). Legacy data can be defined as the data that need to be converted to a different format when another medium will be utilized to process the data or the non-XML data (Dyson *et al.*, 1998). However, legacy data is often defined as the data left from earlier technologies and that may even be in use.

RDS is implemented by using the PHP scripting language and XML. The TLDD is a robust MySQL relational database that includes vital information about Texas academic and public libraries (Lopatovska *et al.*, 2004). 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 access locally licensed resources and other distributed or local resources. The main problem, however, is how locally installed RDS applications can access the TLDD in a standardized and interoperable manner which currently powers the LOT RDS.

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 not only by local RDS applications but also 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.

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

XML, Simple Object Access Protocol (SOAP) and Web Services Description Language (WSDL) are emerging tools to create a WS. SOAP is an XML-based communication protocol that enables different programs running on different platforms able to communicate with each other and WSDL is an abstract description of a WS in XML format. 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” (Adam, 2002). 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. SRW has been built on 20 years of experience with the Z39.50 information retrieval protocol (Sanderson, 2004). 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 (Sanderson, 2004). 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 the Texas Library Directory Web service (TLDD WS).

Project Evaluation

The purpose of this study was to investigate the applicability of WS technology in accessing legacy data. The application itself is not robust and provides limited functionality (e.g., CQL), however, this was not due to WS architecture but programming limitations. The application implementation process showed that developing such services does not require allocation of vast resources and proprietary toolkits. NuSOAP has been selected not only because of its simplicity and flexibility but also because of its popularity among the WS developers (De, 2005). However, researchers were not able to discover sufficient documentation for the NuSOAP toolkit; found instead many sample scripts and limited number of short articles. The NuSOAP toolkit actually offers more features than researchers used in the pilot implementation due to limited time and resources. The researchers focused on the features needed to develop the application.

Even though the development cycle of TLDD WS is short, researchers spent a great deal of time figuring out how to use PHP in WS development because of the limited documentation (i.e., books). Once, the researchers were comfortable with the technology, development was easy and quick.

Although simplicity and flexibility of NuSOAP enabled researchers to implement TLDD WS in a short period of time, other mature SOAP toolkits (e.g., Apache Axis) should also be considered in future studies. The current NuSOAP implementation may fall short in meeting required security and interoperability standards for mission critical WS applications especially for the ones that are made available to third parties. NuSOAP may be the best candidate for WS to be used within the organization internally where organizational network security could be ensured with additional software.

Conclusion

The results of this exploratory study demonstrated the applicability of WS protocols to operational systems in the library community. In this study, a pilot application of WS incorporated with the TLDD has been implemented in an open source environment such as Linux Operating System and Apache Web Server. Further research should be carried out to ensure pilot software is scalable and interoperable with other SOAP implementations.

Additional WS implementations to access more data from TLDD would enable decision makers at TSLAC to deploy RDS application across Texas quickly and smoothly, since local libraries do not have to allocate additional resources while adopting RDS application for their searchable library catalogs. Significant gains should be expected by using RDS application to search local library

catalogs within the physical library, because users are already using LOT RDS from their homes or local libraries to access TexShare databases and online catalogs of other academic and public libraries since March 2004 ("Library of Texas", 2004).

The relevant literature suggests that WS technology is likely to be adopted by organizations at various sizes and is expected to play a key role in information systems, but it will not replace expensive, proprietary systems like Electronic Data Interchange (EDI) in near future because of their large user base (Siegrist, 2005). WS technology can be used to as a complementary tool for current EDI implementations. According to a Yankee Group survey conducted in 2004, forty eight percent of the US enterprises have already deployed WS and another thirty nine percent planning to do so even though most of these WS implementations are at the experimental level (Barlas, 2004). The user base of WS technology is getting larger day by day, the library community should not hesitate to adopt this technology for different services they offer (e.g., Inter Library Loan). WS technology would be the best candidate for accessing the legacy data and application integration due to its low learning curve, short development cycle, and quick return on investment. Thanks to Library of Congress' SRW initiative, libraries will be able to preserve their current technology investments (i.e., Z39.50) while adopting this new technology.

Finally, this study may provide new perspectives for future projects on RDS and TLDD where WS could be used as an underlying technology in ensuring interoperability.

References

- Adam, C. (2002). Why web services? Retrieved October 27, 2003, from <http://www.webservices.org/index.php/article/articleview/75/>
- Barlas, D. (2004). Web services update. Line56.com, Retrieved January 1, 2005, from <http://www.line56.com/articles/default.asp?ArticleID=6028&TopicID=4>
- Coyle, F. P. (2002). Xml, web services and the changing face of distributed computing. *ACM Ubiquity*, 3(10). Retrieved October 15, 2003, from http://www.acm.org/ubiquity/views/f_coyle_1.html
- Dyson, P., Chang, D., Godwin, R., & Weinberg, P. (1998). Repurposing legacy data: Everybody's business. Retrieved January 7, 2006, from http://seminars.seyboldreports.com/1998_san_francisco/ETAPE_4_8.html
- Library of Texas. (2004). Retrieved February 18, 2005, from <http://www.tsl.state.tx.us/lot/overviewlib.html>
- Lopatovska, I. V., Oguz, F., & Moen, W. E. (2004, November 13 - 18). Design, development, and implementation of a Texas library directory database: A multipurpose database for the library of Texas. Paper presented at the ASIST 2004 Annual Meeting; "Managing and Enhancing Information: Cultures and Conflicts", Providence, RI.
- Merriam-webster online. (n.d.). Retrieved March 12, 2005, from <http://m-w.com/cgi-bin/dictionary?book=Dictionary&va=legacy&x=0&y=0>
- Sanderson, R. (2004). *Srw: Search/retrieve webservice version 1.1*. Retrieved February 2, 2005, from <http://srw.cheshire3.org/SRW-1.1.pdf>
- Siegrist, J. L. (2005). An evaluation of web services with emphasis on small hotels. Unpublished Master Thesis, Bournemouth University, Dorset, United Kingdom.