Introduction to Web Services: An ALA Annual Meeting Pre-Conference presentation

This essay outlines what Web Services are and why you, as a librarian, should care.

Web Services is a computing technique for systematically disseminating XML content, usually over a network. In its simplest form, one computer sends another computer a request for information in the shape of an HTTP request or an XML stream. The second computer uses the request or XML stream as input, does some sort of processing, and returns a stream of XML as output. The originating computer is then expected to process the returned XML for some desired purpose. This foundation provides for many opportunities:

1. Since the shapes of the HTTP requests and XML streams are expected to be similar from service to service it is easy to create services that do similar things, such as provide the definitions of words, search indexes, display weather information, etc. Because these service should be implemented similarly, it should be easy to swap out one index search for another index search, for example. It is easy to create standards-compliant services.
2. Since the input of Web Services are HTTP requests and XML streams, Web Services computing does not favor any particular computer language or operating system.
3. Since the output of Web Services includes just information and no presentation layer, the output can be transformed for a wide variety of uses. The most easily understood is HTML, but the output could just as easily be transformed into PDF, email, RSS, an integer, a word, provide the input to update a relational database, or a channel for a portal application.
4. Since the goals of libraries are to collect, organize, archive, and disseminate data, information, and knowledge, it makes a lot of sense for libraries to exploit the Web Service technique in order to accomplish their goals, especially in a globally networked computer environment.

The balance of this essay elaborates on the points outlined above.

REST and SOAP, WSDL and UDDI

Web Services is a computing technique for systematically disseminating XML content, usually over a network, and there are two well-defined ways of implementing this technique. One is called REST and the other is called SOAP.

REST (Representational State Transfer) is considered the simpler of the two techniques. It exploits the World Wide Web (HTTP) protocol to communicate between computers, and initiating requests are usually in the form of URL's. From the library world, the most popular Web Service fitting this description is the Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH). Another very good library example is called Search and Retrieve via URL (SRU).

SOAP (Simple Object Access Protocol) is not dependent on the World Wide Web as a transport mechanism. Requests can be made directly from one computer program to another, via telnet or SSH, via HTTP, or even through email. Unlike REST, SOAP requests as well as responses are encoded within a specific XML syntax called a SOAP envelope. For all these reasons, SOAP is seen as being more complicated and at the same time more flexible when compared to REST. SOAP, unlike REST, is formally supported by the World Wide Web Consortium.

WSDL (Web Services Description Language) is an XML vocabulary used to describe specific Web Services. It describes things like the location and shape of REST-ful Web Services. The WSDL XML file will list the various name/value pairs expected on a URL, and it will define the elements and types of content the XML response will contain. In other words, WSDL documents describe the inputs and output of specific Web Services. Think of them sort of like MARC records for Web Services.

UDDI (Universal Description, Discovery and Integration) is a specific Web Service used to query and identify WSDL files. It a standardized directory service, implemented as a Web Service, for institutions hosting Web Services. If the WSDL records are compared to MARC records, then UDDI is compared to a library catalog.

Examples

There are may relevant Web Service examples applicable to Library Land.

SRW/U

My favorite examples are SRW/U (Search and Retrieve via the Web/Search and Retrieve via URL). Apply the following query -- dogs and cats -- to popular Internet search engines and your browser will return a number of differently shaped URLs:

- http://www.google.com/search?hl=en&ie=ISO-8859-1&q=dogs+and+cats&btnG=Google+Search
- http://search.yahoo.com/search?fr=fp-pull-web-t&p=dogs+and+cats
- http://search.msn.com/results.aspx?FORM=MSNH&q=dogs%20and%20cats

These queries will return lists of results, all formatted differently.

SRW/U are designed to provide a consistent and standard way to query Internet-accessible indexes as well as define how search results should be returned. Here is the same query applied against an SRU interface:

- http://example.org/?operation=searchRetrieve&query=dogs+and+cats&version=1.1

The search result would in the form of an XML file/stream, and it would look a lot like this:

```
<searchRetrieveResponse>
  <version>1.1</version>
  <numberOfRecords>2</numberOfRecords>
  <records>
    <record>
      <recordSchema>info:srw/schema/1/dc-v1.1</recordSchema>
      <recordPacking>xml</recordPacking>
      <recordData>
        <dc>
          <title>The bottom dog</title>
          <identifier>http://example.org/bottom.html</identifier>
        </dc>
      </recordData>
    </record>
    <record>
      <recordSchema>info:srw/schema/1/dc-v1.1</recordSchema>
      <recordPacking>xml</recordPacking>
      <recordData>
        <dc>
          <title>My Life as a Dog</title>
          <identifier>http://example.org/my.html</identifier>
        </dc>
      </recordData>
    </record>
  </records>
</searchRetrieveResponse>
```

Using the SRW/U approach to Internet search, queries could be entered once, sent to an index, and results could be analyzed and displayed depending on the needs of end-user. The same query could then be applied to other indexes simply by changing the root of the SRU URL. SRW works in exactly the same way as SRU except it uses SOAP. SRU is a REST-ful Web Service.

OAI-PMH

OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting) is Web Service complimentary to SRW/U. Where SRW/U provide searching mechanisms, OAI-PMH provides browsing services. OAI-PMH is a REST-ful Web Service. It can use either the GET or the POST methods for sending requests. Here is a simple GET query used get a list of all the items in an OAI repository:

- http://infomotions.com/gallery/oai/index.pl?verb=ListIdentifiers&metadataPrefix=oai_dc

Here is a typical response:

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2005-06-21T19:21:48Z</responseDate>
  <request verb="ListIdentifiers"
metadataPrefix="oai_dc">http://www.infomotions.com/gallery/oai/index.pl</r
equest>
  <ListIdentifiers>
    <header>
```

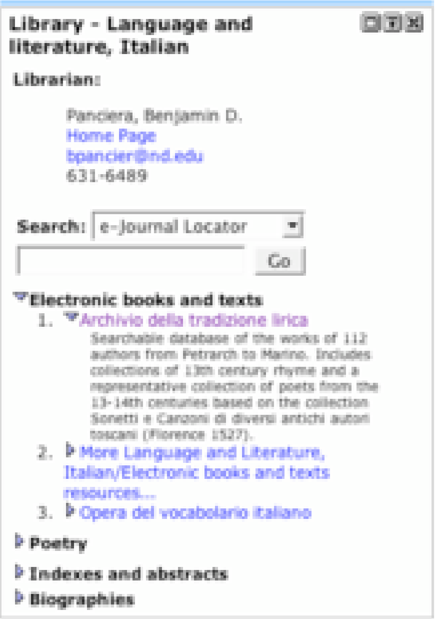
```
      <identifier>oai:images:DSCN0001_3-173</identifier>
      <datestamp>2004-12-25T21:56:02Z</datestamp>
      <setSpec>heart-of-texas</setSpec>
    </header>
    <header>
        <identifier>oai:images:DSCN0288-1</identifier>
        <datestamp>2004-12-25T21:56:02Z</datestamp>
         <setSpec>heart-of-texas</setSpec>
    </header>
    <header>
        <identifier>oai:images:DSCN0290-3</identifier>
        <datestamp>2004-12-25T21:56:02Z</datestamp>
        <setSpec>heart-of-texas</setSpec>
    </header>
  </ListIdentifiers>
</OAI-PMH>
```

MyLibrary, SCT Luminus, and uPortal

Here at Notre Dame we took advantage for our database-driven website application -- MyLibrary -- and created an informal Web Service providing content to our campus-wide portal application -- SCT Luminus (a.k.a. uPortal). In this scenario Luminus sends MyLibrary a URL containing the unique identifier of a MyLibrary subject term, like this:

- http://dewey.library.nd.edu/morgan/mylibrary2portal/mylibrary2portal.cgi?disc_id=4536

A simple CGI script reads the value of disc_id, queries the underlying database for suggested resources with this identifier, creates an RDF/XML file, and returns it to Luminus. Luminus then transforms the RDF/XML into snippets of HTML to be displayed as channels. Each channel looks much like the image to the right.



Library - Language and literature, Italian

Librarian:

Panciera, Benjamin D.
Home Page
bpancier@nd.edu
631-6489

Search: e-Journal Locator

Go

Electronic books and texts

1. Archivio della tradizione lirica
   Searchable database of the works of 112 authors from Petrarch to Marino. Includes collections of 13th century rhyme and a representative collection of poets from the 13-14th centuries based on the collection Sonetti e Canzoni di diversi antichi autori toscani (Florence 1527).
2. More Language and Literature, Italian/Electronic books and texts resources...
3. Opera del vocabolario italiano

Poetry

Indexes and abstracts

Biographies

WordNet

WordNet is a thesaurus designed by the Cognitive Science Department of Princeton University. By querying the WordNet database you can get sets of additional, related words, phrases, and definitions. This interface has been modified into a REST-ful Web Service. Queries look like this:

- http://jws-champo.ac-toulouse.fr:8080/wordnet-xml/servlet?word=origami&recurse=0

Results look a lot like this:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE wordnet SYSTEM "WordNet.dtd">
<wordnet>
```

```
  origami
  <word word-id="94206">origami</word>
  <pos name="noun">
    <category name="act">
      <sense number="1">
        <synset size="1" synset-id="100834453">
          <word word-id="94206">origami</word>
            <definition>the Japanese art of folding paper into shapes
              representing objects (e.g., flowers or birds)</definition>
        </synset>
        <links>
          <hypernym>
            <synset size="3" synset-id="100831589">
              <word word-id="8379">artistic production</word>
              <word word-id="8071">art</word>
                <definition>the creation of beautiful or significant
                  things</definition>
                <sample>art does not need to be innovative to be good|I
                  was never any good at art|he said that architecture is
                  the art of wasting space beautifully</sample>
            </synset>
         </hypernym>
        <domain.region>
          <synset size="3" synset-id="108351073">
            <word word-id="90220">nippon</word>
            <word word-id="90105">nihon</word>
            <word word-id="71660">japan</word>
            <definition>a constitutional monarchy occupying the Japanese
              Archipelago; a world leader in electronics and automobile
              manufacture and ship building</definition>
          </synset>
        </domain.region>
        </links>
      </sense>
    </category>
  </pos>
</wordnet>
```

By exploiting WordNet-like Web Services in our many index interfaces, libraries could seamlessly enhance queries and search results making the interfaces "smarter."

Summary

Just as we can decreasingly expect people to come to our physical libraries in order to use our collections and services, we can decreasingly expect people to come to our websites. Web Services provide the means for you to make your content available in the user's space, as opposed to the other way around. The ability make this happen comes at a cost -- the understanding and implementation of an additional skill set.

Finally, when requesting Web Service functionality from integrated library system vendors, insist on the implementation of standard Web Services such as SRW/U and OAI. Not vendor-enhanced standards. Also, insist on thorough documentation. Thorough documentation will describe how to request the XML output as well as provide DTDs and/or XML schema for validating it.

Links

1. Web Services - http://www.w3.org/2002/ws/
2. REST - http://www1.ics.uci.edu/%7Efielding/pubs/dissertation/rest_arch_style.htm
3. SOAP - http://www.w3.org/TR/soap/
4. WSDL - http://www.w3.org/TR/wsdl
5. UDDI - http://www.uddi.org/
6. SRW/U - http://www.loc.gov/z3950/agency/zing/srw/
7. OAI-PMH - http://www.openarchives.org/
8. MyLibrary2Portal - http://dewey.library.nd.edu/morgan/mylibrary2portal/
9. WordNet - http://wordnet.princeton.edu/

Eric Lease Morgan
University Libraries of Notre Dame

June 23, 2005