

Web Services For Translation

Mike Roche,
IBM Software Group, Dublin,

MikeRoche@ie.ibm.com

Introduction

The paper will discuss the ways in which Internet 'Web Services' technology can be used to enable the producers and consumers of translation services to find and connect to each other. It will show how translation services (including Human and Machine translation) can be:

- Found or sourced across the internet (using UDDI discovery servers and WSDL/WSFL)
- Connected to (using SOAP & HTTP/HTTPS)
- And how translation jobs may be submitted and retrieved (using XLIFF packages)

The research into this technique and the resulting solution arose from real customer needs and requests, and these will be discussed along with the problems that this solution solves. I will show how the task of locating a translation services vendor may be automated and how the whole process of managing translation jobs may be automated through clever use of automated Workflow technology as part of a translation process. The embedding of such services into translation tools and technology permits solutions to be more user friendly and easier to use particularly for first time localisers and also allows greater control for project managers.

Also discussed is a working web service co-developed by IBM and Berlitz which uses IBM Lotus Notes using WSDL, SOAP & HTTP to connect to the Berlitz translation server and submit and retrieve translation jobs. This prototype service is serving as the basis for an industry standard version being worked on by a wider group of companies.

I will also discuss the XLIFF standard ("XML Localisation Interchange File Format") which is a standard created by a working group of representatives from many companies including, IBM, Oracle, Berlitz, LionBridge, Novell. The standard is now managed by OASIS and is at the core of Web Services for translation. And finish up with where we can go from here.

Description of problem

For companies faced with translation requirements and projects for the first time it can be a daunting task to source and set up the required business relationships and processes.

For companies or individuals with ad-hoc immediate translation requirements it can be difficult to source and arrange translation. You then have the problems associated with packaging and transmitting the files for translation as well as checking the status of work in progress so you can schedule around it.

We frequently hear comments on a recurring theme from our customers, along the lines of, "You provide good solutions to help us prepare, manage and serve up multilingual content but we have not worked with translators before, can you help us find them and set up

processes to work with them". While we could often point them in the right directions to locate the required help we thought it would be nicer if the technology and solutions we ship could include this functionality.

And the problem is not restricted to those new to the translation process. Similar problems are also faced by companies trying to automate their translation process - in the case of web content management systems for example.

An Ideal Solution

As a project manager or basic end user requiring any of the above it would be great if software could automate this process for me, requiring minimal input from me. This Utopian scenario would have been very difficult to develop until recently due to lack of inter operable standards in many of the areas concerned but two key developments have made it possible: The first is Web Services technology and the second is a relatively new lesser known XML standard for translation file interchange - XLIFF. Specific translation web services would be possible without XLIFF but it provides a way of sharing information and jobs between many different processes, vendors and translation tools. So XLIFF enables us to share translation jobs due to a common packaging format being supported, meaning that solutions that mix and match different components are now possible.

What are Web Services ?

Web services are "self-contained business functions that operate over the Internet" - essentially services offered over the web. They are written to strict specifications using standards to enable them to work together and to work with other similar kinds of components.

Because Web services are built on standards, they make it possible for many solutions developers to enter the market and offer the services, which helps increase competition and ultimately bring down costs. Basing systems on standards helps avoid being tied to a specific supplier or vendor as it is possible to easily switch to any vendor that supports the specific service. This freedom helps to increase competitiveness which in turn helps drive innovation, value added services and efficiency in the services supplied.

Standards and technology allowing interoperation and connection of services have existed before now but they were nowhere near as flexible as Web Services. Previous standards such as Internet Inter-ORB Protocol (MOP) and Remote Procedure Calls (RPC) achieved some success but suffered from many short comings such as; requiring specialised software to be installed on the client side, or problems working across fire walls or they didn't pass information around in easily readable form - by either machine or human consumers.

The key differences between Web Services technology and these previous generations of interoperability standards lie in the fact that Web Services are built on existing proven flexible standards - the standards that are at the core of how we use the web. They require no special software to be installed and more importantly they allow connection between systems regardless of the operating systems or programming languages being used. Web services can be developed for example using Java, VB, C++, Perl, or C# , running on any mix of Windows, Unix, MacOS, PDA's, Mainframes etc.

So what are the standards used ?

- Web Services clients communicate with the web service server using the HTTP protocol - the standard protocol used to browse the web.
- XML is used to carry the request and response parameters between the client and server.
- SOAP - the Simple Object Access Protocol is used to package the messages communicated to and from the web service.

And these three standards are all that is required to develop a basic web service. Using only the above and any of the programming languages mentioned it is possible to develop both the client and server sides of a web service. Such a service however would not be easily sharable with others who may wish to connect to it, what we need is a way to publish services and allow them to be discovered and connected to. Enter UDDI and WSDL...

- UDDI - the Universal Description, Discovery and Integration initiative was created to facilitate discovery of web services over the Internet. It effectively provides a way of providing directories for web services. The standard was initially developed by IBM, Microsoft and Ariba. It allows web services providers to register service definitions and response specifications with the registry. Others can then 'find' these services and their descriptions.
- WSDL - Web Services Description Language is the standard that makes it possible to share the description and interface mechanism to a web service. The newer versions of web development tools can even automatically generate code to interface/connect to the service based on the description specified in WSDL held in the UDDI database.

How can Web Services technology be used to construct solutions for translation services

Several translation vendors have implemented web based translation services over the past few years. These solutions provide a web portal that allows users to register on-line, submit jobs for translation, view the status of jobs in progress and retrieve completed jobs. These services are a step in the right direction but suffer from a few major drawbacks:

- They have to be located manually by a potential user either searching for them or being directed to them.
- The format of the files or packages submitted and retrieved need to be agreed up front and/or adhere to one of the formats the service accepts.
- The jobs need to be manually packaged, submitted and retrieved

The advent of web services technology means connections between clients and these web based services can be automated but the missing piece was a standard way of packaging translation jobs for exchange between services. So with the development of the XLIFF specification we now have all the pieces required to exchange jobs between any client requiring translation and any translation web service.

XLIFF explained

The "XML Localisation Interchange File Format" (XLIFF) is an extensible specification for the interchange of localisation information. The main aim is to support a way for submitting localisable content in a standard way. Additional information such as translation suggestions and glossaries can be included in the XLIFF to help translators when working on a job they have received or downloaded. The specification is also designed to be extensible.

The group working on the XLIFF specification includes representatives from Microsoft, Sun, Novell, IBM, SAP, & Oracle. The specification for XLIFF 1.0 is available from:
[Http://www.oasis-open.org/committees/xliff/documents/xliff-specification.htm](http://www.oasis-open.org/committees/xliff/documents/xliff-specification.htm)

The specification is registered with the OASIS open standards organisation through which the committee work and they are currently working on an extended specification for XLIFF version 2 which is being worked on by a wider group of companies.

So using a combination of the web services technology and XLIFF a translation vendor can create a simple connector to their on-line translation services allowing many and varied clients to connect. Obviously this is simpler to set up if you already have a web based system for accepting and returning translation jobs that you can just expose through a web service. A solution could also be developed without this, it would just be a little more work. And as I will discuss later it is possible to connect web services into a workflow based solution which would allow you to use web services in a content management based system without a translation portal.

Why would I do this ?

For a translation vendor developing a translation service based on web services technology and XLIFF provides the ability to serve a large potential market comprising clients from anywhere on the web.

For a client requiring translation services it provides the ability to source vendors through UDDI registries allowing wider choice and the ability to work with multiple vendors without any additional overhead.

A full solution would include the following functionality:

- A standard translation vendor web service described in WSDL and registered with the UDDI servers
- Account set up/registration
- Job rate negotiation
- Job submission
- Job status
- The ability to tell the user when a job is complete/ready for pickup
- Job retrieval
- Account administration
- Statistics retrieval

Such a comprehensive solution would be great but as with everything we have to start somewhere and I will now discuss the early system IBM and Berlitz co-developed.

Description of IBM/Berlitz solution

Working with our customers we specified and designed a solution to connect our Domino Multilingual Content Management solution (Domino Global WorkBench) to the Berlitz on-line translation vendor service. The solution had to contain the following features:

- It should require minimum or no set-up up front
- The packaging and transportation of translation jobs to the service should be automatic
- The retrieval and reintegration of translation jobs should be automatic
- Status should be available on request through a user initiated automatic response mechanism
- It should support translations into a large number of possible languages
- It should be extensible for the future and entirely based on standards

We co-developed a web service that connects IBM Lotus Domino clients to the Berlitz web based translation service. The service currently exposes three interfaces to the client:

- SubmitJob
- GetJobStatus
- RetrieveJob

The server side was developed in VB using Microsoft dot net technology and the client side was developed in Java using IBM Lotus Domino. The server side runs on Windows platforms and the client side runs on Windows, Linux or AIX, so you can see how this demonstrates the connectivity and interoperability of different platforms and development environments.

For our first release the service requires that a user be registered with the Berlitz service. Prior to using the service an account needs to be created and a user name and password obtained. This user name and password will be passed to the web service every time a call is made to one of the functions.

To use the service all a user has to do is choose the Berlitz option in the Domino Glossary application select the source and target languages and chose send. The job is packaged as an XLIFF document and sent to the Berlitz web service along with the user name and password supplied. If the server accepts the job then a job ID is returned. This ID can be used to retrieve job status and the retrieve the completed job in the future. At any one time there may be many jobs in progress for an account so the job ID provides a way of identifying individual jobs for any account or user.

The function GetJobStatus can be used at any time to query the progress of a job.

For the first release the Berlitz server will e-mail the registered user when a job is ready for pickup, and the user can then retrieve the job using the corresponding function in the Domino Glossary. Alternatively if the job is reported as complete when the status is queried, then the user can retrieve it.

Where we go from here - "Web Services for Translation"

Our initial solution has some short comings, notably:

- We have not published a WSDL definition and registered it with a UDDI server.
- First time registration/account set-up is not supported.
- Billing and account administration is not supported through the web service

The first is by choice as we wanted to prove the concept before working on a more general industry standard implementation.

The second and third shortcomings are mostly due to a lack of adequate security standards in the web services arena. Solutions are now starting to appear in this space so this will be possible shortly.

Our current implementation is also user initiated for job submission and retrieval although everything else is automated. Tying the web service into a translation process workflow would allow for complete automation and no user involvement unless problems or exceptions were encountered.

Automating the process & connecting to other translation related web services

The following diagram shows an example of a fully automated solution.

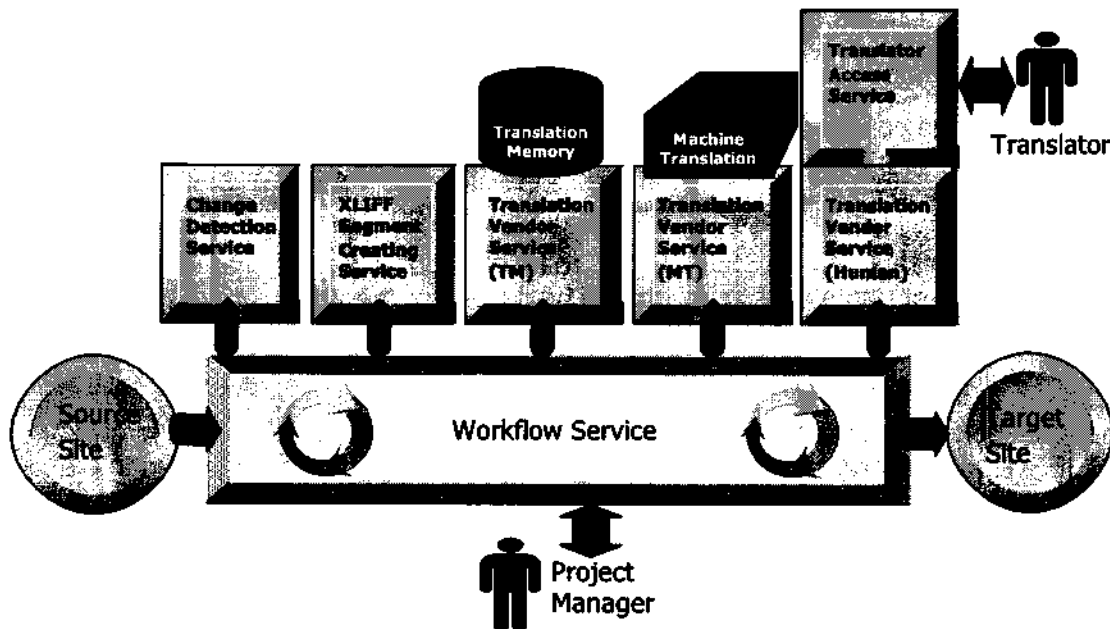


Figure 1. Utopian service-based localisation process

In this system a translation workflow sits at the heart of all operations connecting to the various services as required. This obviously offers very high flexibility.

The diagram also suggests other potential web services including:

- Services to connect to Machine Translation systems
- Services to connect to Translation Memory systems
- Services to enable remote translators to connect to the translation server to claim, pick up jobs or drop off completed jobs

In addition to services directly related to translation, you can also see additional supporting services in the diagram:

- A change detection service to detect changes in content requiring translation attention
- An XLIFF packaging service
- There are many other possibilities and other services could be plugged in to such a solution as they became available.

The above list represents the core services required to build flexible translation solutions. There are many other possibilities including data access services, services to connect to existing legacy systems, analysis and reporting services etc. As more vendors adapt their solutions and technology to embrace web services and XLIFF we will be able to plug applications and processes together to produce flexible translation systems.

Connecting to these different services from a translation workflow opens up endless possibilities to create flexible processes. As the technology is so simple it can be linked easily to either proprietary or standard workflow systems - such as IBM MQ series, Domino Workflow or the workflow of specific content management systems such as Intranet.

We have formed an early working group to further refine the requirements for a general web service for translation vendors. This group has started to draft a specification for a full industry standard solution and once the ground work has been done the group will extend an invitation for more widespread involvement.

Conclusion

Web services can add a lot of value and flexibility to producers and consumers of translation services and technology. Building systems that support access to or access from web services opens your solutions up to all these, and to whatever future services may be developed and published. Over time the work effort can be shifted away from the technology and onto the service and the value that different vendors and applications provide.

List of standards and links to further information

SOAP - Simple Object Access Protocol. Allows applications to 'call' each other over the internet. Based on XML.

UDDI - Universal Description Discovery and Integration. A web based distributed directory that allows businesses to list, and describe, themselves and (optionally) their web service offerings

WSDL - Web Service Definition Language - used to define the interface (generally SOAP) that a web service allows/uses.

WSFL - Web Services Flow Language - used to describe business processes as compositions of web services.

XLIFF - XML Localisation Interchange File Format - A format for submitting localisable content in a standard way. It can also include translation suggestions and glossary information.

XML - extensible Markup Language - a subset of SGML designed especially for web documents. It allows designers to 'create' their own language.

Useful Web Links

Web services at IBM -

<http://www.ibm.com/webservices> for documentation, toolkits, discussion

<http://www.ibm.com/services/uddi> for information about UDDI.

More general web service information -

<http://www.uddi.org/> official UDDI initiative site

<http://www.webservices.org/> independent, free membership, white papers, discussion area

XLIFF

<http://www.oasis-open.org/committees/xliff> - XLIFF site on OASIS

Translation Services

<http://www.berlitzit.com> - Berlitz on-line translation service

Bibliography

Looby, B. 2002. "Web services - The future of localisation", Localisation Focus, The International Journal for Localisation

Savourel, Y. 2001. "XML Internationalization and localization", SAMS

OASIS. 2001. XLIFF specification, www.oasis-open.org/committees/xliff

Figure 1. Reproduced by kind permission of Bill Looby, IBM