Web Services for the Translation Industry

A Whitepaper from the Translation Web Services group

Table of Contents

Introduction

Many translation and localisation projects involve a multiple of complex tasks carried out in different countries by different companies. The management of this is a very complex business. To find out what the cost of a project will be involves sending files, data, and other information to different companies to get information on their charges. As the project progresses a lot of time is spent by project managers chasing up status information from the various vendors. Even receiving the completed translated files can be cumbersome with the project manager having to collect files from different ftp servers.

The need for 'Joined-up localisation' has long been a dream of many in the translation industry. Why is it not possible to automate much of these processes? Why can't the Web be used to transfer data and information effectively? Why can't Project Managers see everything they want on one screen regardless of which vendor is used? There have even been efforts to try and get some way near 'Joined-up localisation'. One of these was the Transrouter project which was pioneered by the Localisation Resource Centre. This tool was designed to know information about the project, and possible tools and services and use this to find the best route for a particular localisation project.

However these efforts remained dreams as the technology was not there to provide 'Joined-up localisation'. Web Services has changed that. By allowing computer to computer communication over the Internet automation of many localisation processes is now possible. A publisher of software or documentation could use Web Services to get not just one quote but quotes for different tasks and from different vendors. Web Services could be used as the backbone to a workflow automating the links between the different tasks. Reporting information from a number of sources could be consolidated at the customer's computer with the aid of Web Services.

This document is the outline of the agenda of the Translation Web Services group. The report has been prepared by this steering group to outline what we intend to do to establish localisation Web Services and some of the issues surrounding this. It will not provide a blueprint for this but be a starting point for the conversations that will establish this blueprint. The report will begin with a business justification for a localisation Web Service and an illustration of what has been done already. This will be followed with a description and an outline of the issues that need resolving for this to become a reality.

The business case for a localisation Web Service

Automation of translation services has attracted attention of the academic and business world for some fifty years. The ambition to have a machine produce fully automatic high quality translation appeared to be a dream that could not be fulfilled any time soon. Nevertheless several companies dedicated years of research and development to building software that could either translate automatically or assist the human translator by maintaining glossaries and databases of translated segments. Since the last twenty years computer-assisted-translation software has become very common in the translation market, in particular in the software localization industry. Clients and vendors report that the use of this CAT software helps them to reduce translation volumes by double digit percentages. But at the same time the translation and localization market has turned into a much more complex environment. Volumes of translation work keep growing as global corporations face the challenge to not just localize their products but also their human resources Intranets, their customer support databases, their training materials and even email and instant messaging. Time-to-market seems to become more critical. And cost of course remains an issue. In this complex environment it has become apparent that the cost of the actual translation production work amounts often to not more than 50% of the total attracted cost of translation. The other 50% is spent on managing thousands of transactions and interactions in an ever more complex and more distributed supply chain. No matter whether you are dealing with a translation of one sentence or of an entire manual, the customer must coordinate with the multi-lingual vendor who must coordinate with the in-country single language vendors (for 25 target languages) who in turn must coordinate with the free-lance translators. This supply chain gets activated back and forth multiple times for any given project or client change request. It is no surprise that the localization industry since a few years started looking at the possibility of automating these transactions in workflow systems. However the offering of the first of these kind of workflow systems forced customers to use exclusively the services of the provider of the workflow automation system. Without a standard for the definition of translation services it was hard to imagine an open translation procurement system. And without web services even an open procurement system could not prevent that project managers would have to respond to every email request for price details.

Web services now create the best opportunity for the translation market to become a professional services industry. Web services help us to overcome the dilemma that has suppressed the growth of the translation market. The mergers and acquisitions of recent years and the launch of new market offerings have only reinforced this dilemma. The supply chain has become more complex, with more vendors and more tools …. and more incompatibility. Any gain possibly booked by the use of translation memory software is lost again because of the increase in overhead of managing multiple translation databases. Web services connect tools and suppliers automatically without the need for direct communication. Articles or sections from web sites will find their way to the translator fully automatically. No phone conversations, no faxes, no FTP, no emails (except perhaps automatically generated emails). Translation memory matches will be retrieved fully automatically from anywhere in the network. And for real-time translation the web services architecture

will fully automatically route texts to a machine translation engine. No pre- and post-processing, no translation folders and no so-called translation engineering. Web services allow us to build the intelligence into the infrastructure, in standard definitions that communicate openly with any tool and any supplier that comply with the standard.

Case history - Domino Global Workbench/ BerlitzIT integration

IBM and Bowne Global Solutions (formerly Berlitz GlobalNET) have come together to offer a technological solution for translating notes databases which uses Web Service technology. IBM have developed a technology called Domino Global Workbench which has been integrated with the BerlitzIT online translation service.

Users of Lotus Domino from IBM will be already aware of the functions offered by Domino Global Workbench (DGW). This is a tool which is used to manage the translation of notes databases. The BerlitzIT integration links this technology with a pipeline to professional human translators, who offer reliable high-quality translation. DGW Glossaries are converted to an XML format called XLIFF. This was created by a number of companies working through Oasis as a standard localisation interchange file format. The XLIFF file is sent to the BerlitzIT WSDL document using SOAP technology. Here, the sender and other information included in the file is also automatically identified. The file is then translated using the BerlitzIT process. The translated XLIFF file is then returned to DGW using the Web Services technology and the glossary is updated.

The diagram below gives an overview of how the Domino Global Workbench/ BerlitzIT process works.
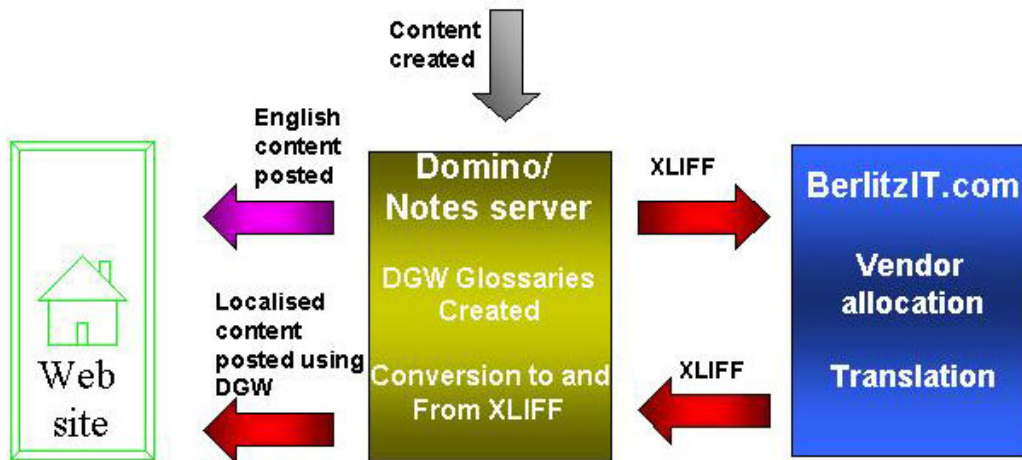

Figure 1 – Domino Global Workbench/ berlitzIT integration

Towards an industry standard

There are many companies using or planning to use Web Services to add automation over to the Internet to their processes. This is very good exciting work but there is a great danger of the wheel being constantly re-invented as companies discover the benefits of Web Services. The TWS group is proposing that those involved in the translation industry agree an industry standard for Web Services. Publishers and Vendors will be able to build software that uses Web Services. The TWS group will work towards agreed terminology which is used throughout the industry. For example it is possible that a publisher would like a document translated and wants to use Web Services to achieve this. One vendor could use the text 'translation' for the translation task, another might use 'trans' another 'trns'. The work of this group is to ensure that when a publisher would like a document translated they use one set of terminology which is understood by vendors providing that service.

Web Services are essentially, server based applications that can be fully described using WSDL. This WSDL definition can be used to easily create a client, and if UDDI is used to publish the availability of the service, the Web Service story is complete.

In this case the service application might be something such as service providing online translation. The interface to this application is opened over the Internet using SOAP (Simple Application Access Protocol) and details of the interface are published in a .WSDL (Web Services Definition Language) file. If the publisher of the service wants it possible for their service to be found through the UDDI registry it is published to there. The client application could be a software application which searches the UDDI for an appropriate service and finds the published .WSDL. This has information about how to access the service over the Internet which the client application can now do.

The building of client and service applications is something that is up to individual companies to do in a way that suits their business needs. WSDL documents explain the service which is being offered and as long as everyone uses the same WSDL they can all interoperate. The TWS group is proposing that the participants in the translation industry agree a single standard WSDL services definition which can then be registered as a UDDI tModel. UDDI tModels provide a way that several providers can all reference the exact same agreed upon service definition. Companies providing services will then use these definitions when creating their own Web Services.

The process for doing this is as follows:
- Form an industry wide group capable of defining a set of services types for the translation industry. This will be the TWS group.
- Agree the scope of work which will be covered by this group.
- Define the most meaningful transactions that can take place between a translation vendor and customer, as part of the process of translation.
- Define service types and describe them with service interface definition of the common WSDL documents.
- These WSDL documents will then be published as UDDI tModels with the overviewDoc field in each of these tModels pointing to the relevant WSDL documents.

Client and service applications can then be developed which use these agreed service definitions.

Defining a set of service types

At the core of a localisation Web Service is the ability for publisher A to request quotes and other services from vendors 1, 2 and 3 and for each party to understand what the other is asking of them. To do this the terms which are used within a localisation Web Service must be agreed. One of the main intentions in setting up the Translation Web Services group is to establish these sets of terms and establish a body which the localisation and translation industry will agree its definition of terms.

It is strongly recommended that where standards already exist such as ISO standards that these are used. For example language codes are defined by the ISO committee.

Conclusions/ Next Steps

This document is intended to give an indication of the work that needs to be done in order to have Web Services used widely within the translation industry. On November 13th the first meeting of the TWS group will take place as part of the Localisation Resource Centre 2002 conference in Dublin. A Technical Committee is being formed at Oasis to do the work required for the creating of the WSDL service definitions.

If you are interested in joining this group please contact:
Peter Reynolds peter.Reynolds@bowneglobal.ie
Tony Jewtushenko tony.jewtushenko@oracle.com
Reinhard Schaler reinhard.schaler@ul.ie

Glossary

Term	Definition
SOAP	Simple Object Access Protocol – The protocol for sending business data over the Internet.
WSDL	Web Service Definition Language - An XML language for defining Web Services.
UDDI	Universal Description, Discovery and Integration – The registry where you register a Web Service if you are offering one or search for a particular service if you are looking for one.
businessService	businessService - Services are represent within UDDI by the businessService data structure.
tModels	tModels – These provide the ability to describe compliance with a specification, concept or shared design within the UDDI
XLIFF	XML Localisation Interchange File Format – is an XML standard created by the localisation industry for the interchange of localisation information.