D4.1.3: LINGUASERVE ONLINE SYSTEM MODIFICATION

Pablo Nieto Caride, Giuseppe Deriard

Distribution: Public

MultilingualWeb-LT (LT-Web)
Language Technology in the Web

FP7-ICT-2011-7

Project no: 287815

Document Information

Deliverable number:	4.1.3
Deliverable title:	Linguaserve Online System Modification
Dissemination level:	PU
Contractual date of delivery:	31st January 2013
Actual date of delivery:	15th February 2013
Author(s):	Pablo Nieto Caride, Giuseppe Deriard
Participants:	Linguaserve
Internal Reviewer:	Linguaserve
Workpackage:	WP4
Task Responsible:	Giuseppe Deriard
Workpackage Leader:	Pedro L. Díez Orzas

Revision History

Revision	Date	Author	Organization	Description
1	14/02/2013	Pablo Nieto, Giuseppe Deriard	Linguaserve	Final version
2	18/02/2013	Pablo Nieto, Giuseppe Deriard	Linguaserve	Revision
3	2/07/2013	Pedro L. Díez Orzas	Linguaserve	Description of the basic ITS 2.0 functionality

CONTENTS

1. LINGUASERVE ONLINE SYSTEM MODIFICATION

This use case demonstration illustrates how ITS via a Real Time Translation System connected to different MT Service Providers allows to:

1. Communicate instructions on language, domain and translation, and convey information about the translation to a content editor.
2. Communicate instructions and information to a post-editor regarding the state of a process, and also to inform the user of the progress of the same process.
3. Communicate the identity of agents that have been involved in the revision and the translation of the content, and to allow translation quality reviewers, to evaluate how the performance of these agents affects the quality of the translation.
4. This scenario may involve the following product classes: Content Authoring Tool; Content Editor; Content Management System (CMS), MT Systems and Web Browsers.

1.1 Link to prototype or downloadable

http://atlaspw1-pre.linguaserve.net:10080/its2_prototype/home.html

User: mlw-lt

Password: its2-wp4

1.2 Description

ATLAS PW1 is a real-time solution for multilingual publications through the Internet. ATLAS PW1 allows the user to navigate a website in a completely transparent way.

When ATLAS PW1 receives a translation request, it downloads the original document, sends it to the MT System and finally delivers the translated document to the user's browser.

Although ATLAS PW1 acts as a reversed Proxy, and needs to have previously configured a website in order to navigate it, for testing purposes, a testing page is available and accessible from https://atlaspw1-pre.linguaserve.net/admin/ (under above credentials).

The showcase:

- enables an HTML5 content reviser (language editor, translation post-editor) to add ITS 2.0 metadata to the contents of web documents.
- uses the ITS 2.0 metadata to control the behavior of different Machine Translation (MT) Systems and Multilingual Publication System.
- covers post-editing of translations generated by MT.

This implementation demonstrates how an Online MT System can automatically:

- Translate HTML5 documents from an ITS-conformant Web CMS.

- Control a process depending on its progress, convey information to editors and indicate the state of the content to the user.
- Communicate the identity of agents that have been involved in the revision and the translation of the content, and to allow translation quality reviewers, to evaluate how the performance of these agents affects the quality of the translation.

1.2.1 ITS 2.0 FUNCTIONALITY

In this use case ITS meta-data is used to solve the following problems, and the ITS 2.0 markup provides:

- key information to drive the reliable extraction of translation-relevant content from HTML5;
- helps to control workflow dimensions such as selection of domain-specific vocabulary to improve the Machine Translation results;
- provides information for post-editing.

The interoperability behavior has the following design assumptions:

- After clicking in the language selector the user will send a request to the real time multilingual publishing system (RTMPS) to translate the input file.
- Some of the metadata of the input will be deleted in the output after the process.
- The input file example is based on the HTML5 files of the Test Suite.

Some key functionalities and its corresponding benefits are:

ITS 2.0 Functionality	ITS 2.0 Benefits
Informing the RTMPS of precisely which sentences or sentence fragments should or should not be translated and which is the source language. Uses the **Translate** data category and the **Language Information** data category.	Allows the user to block automatically the machine translation of certain parts of the Web page that are not required to be translated or must not be machine translated because of its difficulty or provenance, i.e. a technical essay or constitutional laws. Avoids automatically the machine translation of parts of the Web page that are in various languages and must remain that way, i.e. a language selector. Specifies automatically to the RTTS the source language of the text and whether it applies to the whole text or not.
Informing the RTMPS, at a paragraph, sentence or word level, of the appropriate training corpora or glossary (depending on the MT System) that should be used on the translation by the MT Systems. Uses the **Domain** data category.	Improves the accuracy and quality of the machine translation.

Providing the editor with the necessary information to review the text in order to help him with the disambiguation and to improve the quality and accuracy of the revision. Uses the **Localization Note** data category.	Benefit: Help to improve the accuracy and quality of the a review machine translation after the post-edition process.
Informing the editors of the priority of a determined production process, if the content has been updated since the last process, and the date expected to be completed. Uses the **Readiness** data category (Extension for CMD, out of ITS 2.0).	Provides version control. Allows the editor to establish a planning based on priorities and deadlines concerning the different processes that has assigned.
Informing the user of the progress of different processes to which the content is submitted. Uses the **Readiness** data category (Extension for CMD, out of ITS 2.0).	Helps the user to extract information about the processes and allows him to create statistics and plan future processes.
Informing the RTTS of the progress of different processes in order to take different course of actions. Uses the **Readiness** data category (Extension for CMD, out of ITS 2.0).	Allows the RTTS to manipulate the content and present it to the user in a determined way, depending on the behaviour selected.
Informing the translation consumers of how a content was translated and subsequently revised. Uses the **Provenance** data category.	Provides information on the identity of the translation and revision agents whether they are a person, an organization or a piece of software.
Providing the translation consumers information concerning what language the original source text was in. Uses the **Readiness** data category (Extension for CMD, out of ITS 2.0).	Helps the revisers and post-editors to discern the possible cultural nuances of the original text so as to perform a better revision
Passing feedback on the severity of the errors detected during a language-oriented quality assurance (QA) process, to the translation consumer. Uses the **Localization Quality Issue** data category.	Allows the content author and the consumer of the translations to have a better understanding of the common errors made during the translation and revision process and to help them in making decisions for the improvement of these processes quality.

Locale Filter (Implementers: DCU)	Reveals that content is only relevant for certain locales (useful in localization).
MT Confidence (Implementers: DCU)	Assesses the confidence in the quality of the translation generated by the MT system.

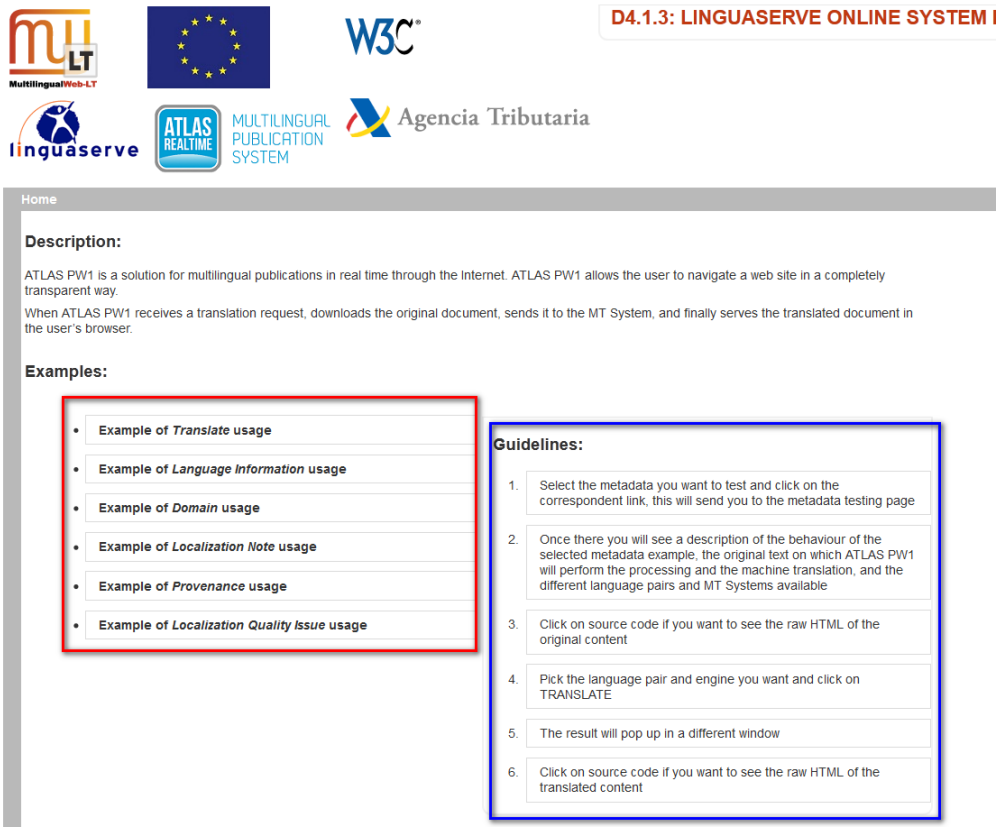
1.2.2 DATA CATEGORY USAGE

The selected data categories have been used for the following purposes and implemented by the different partners as indicated:

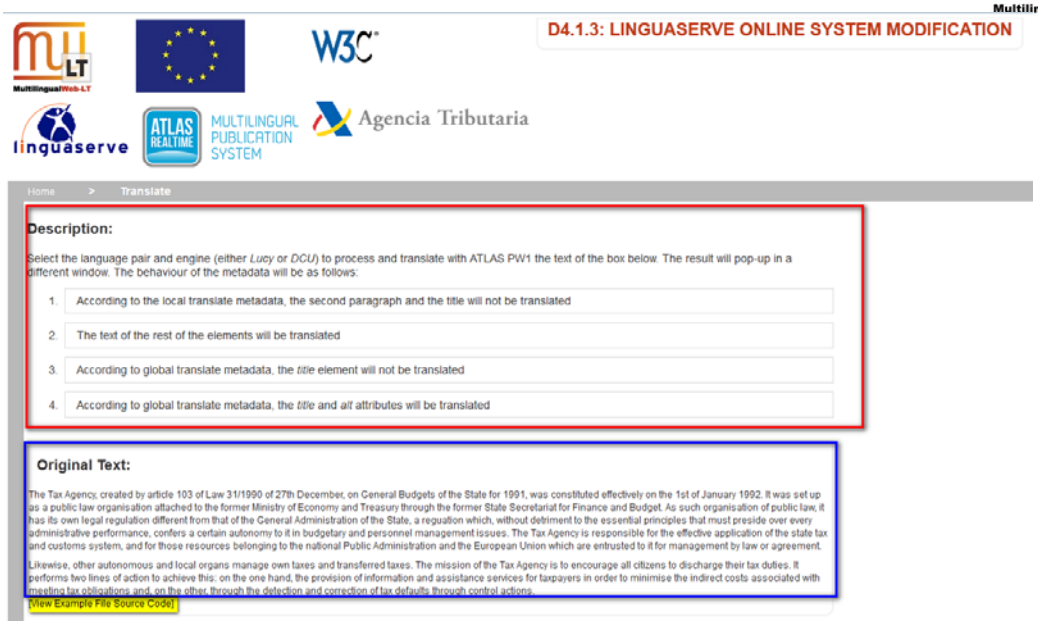
- **Translate** - Parts that are not translation-relevant are marked (and protected).
  - Implementers: Linguaserve, DCU, LucySoftware.
- **Localization Note** - Provides additional information for language or translation editors to facilitate translation.
  - Implementers: Linguaserve.
- **Language Information** - Controls workflow dimensions such as setting the source language, and the target language (via the lang attribute of the output), it also protects the translation of contents where the lang attribute is different from the source language.
  - Implementers: Linguaserve, DCU, LucySoftware.
- **Domain** - Domain values are mapped to the domains used by the individual MT systems, and used to select the appropriate vocabulary.
  - Implementers: Linguaserve, DCU, LucySoftware.
- **Provenance** - Allows tracking of human agents (language or translation editors) or software agents (MT systems) that processed the content.
  - Implementers: Linguaserve.
- **Localization Quality Issue** - Can be provided for the translated content by the reviser. Can be utilized for example by MT developers to improve the MT System.
  - Implementers: Linguaserve.
- **Locale Filter** - Reveals that content is only relevant for certain locales (useful in localization).
  - Implementers: DCU.
- **MT Confidence** - Assesses the confidence in the quality of the translation generated by the MT system.
  - Implementers: DCU.

1.3 User guide

Click on the link provided above and enter the username and password to access the prototype. You will access the home page.



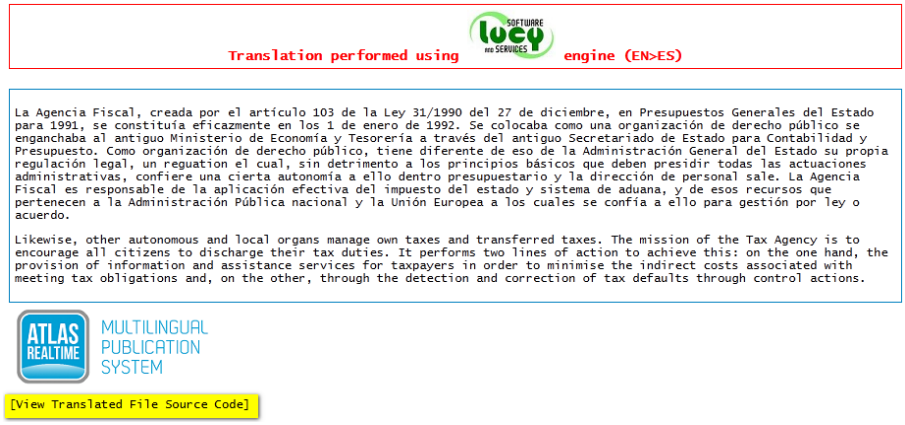
Read the instructions (blue textbox) and select a metadata to test (red textbox). You will access the data category home page.



You will first see a description of how the example will work (red textbox) as well as the source text of the example (blue textbox). There will also be a link to the source code of the source example file (highlighted).

Select a language pair (available pairs are: English to Spanish, English to French or English to German –only Lucysoftware) and the engine, and click on the *"TRANSLATE"* button.

Then, a new window will open with the resulting translated text.



You can also view the source code of the translated file by clicking on the bottom link (highlighted).

As for the testing page provided, it can be configured to show a normal output or a test suite output, and allows the user to load HTML5 documents or to type the input in a textbox. It also allows the user to select the desired MT engine. The only constraint is that the source code must be a valid HTML5 + ITS 2.0 document.