

The MultilingualWeb-LT Working Group



Commercial translation is a \$21 to 26 billion-per-year industry enabling hundreds of billions of dollars of cross-border business around the world. Governments, NGOs and multinational companies often need to translate Web content into dozens of languages simultaneously and individuals often need to access content published in languages they do not know.

The lack of standards for exchanging information about translations raises translation costs by billions of dollars every year, and barriers to distributing content in more than one language mean lost business.

51% of European retailers sell via the Internet, but only 21% support cross-border transactions.

The Web is becoming more linguistically diverse –Chinese, Russian and Arabic have grown the most over the past decade– and translation demands are increasing rapidly.

All this means that building support for multilinguality into the core of Web and content management technologies is rapidly becoming a critical issue for business, government, and other sectors.

We Need your Input

The MultilingualWeb-LT project (<http://www.multilingualweb.eu>), funded by the European Commission and administered by the World Wide Web Consortium, is defining the standards to support multilingual content in the Web. To ensure that these standards meet your requirements, we welcome your feedback. (A technical draft of our work is available for feedback at:

<http://www.w3.org/International/multilingualweb/lt/wiki/Requirements>

We encourage you to review it and contribute your experience to help us improve it).



Administered by the World Wide Web Consortium



Sponsored by the European Commission

Type of project: Support Action • **Duration:** 1.1.2012 – 31.12.2013 • **Goal:** filling in gaps for using language (related) technologies in the Web
Project partners: Cocomore • Dublin City University • ENLASO • German Research Center for Artificial Intelligence (DFKI) • Institut Jožef Stefan • Linguaserve • Lucy Software and Services • Microsoft • Moravia • Trinity College Dublin • University of Economics, Prague • University of Limerick • Vistatec

The MultilingualWeb-LT WG

Defining Metadata for the Multilingual Web



The goal of MLW-LT is to set the foundation for the integration of language technologies into core Web technologies, via the creation of a standard, defining three kinds of metadata on:

- 1) Information in Web content relevant for language technology processing;
- 2) Processes for creating Web content via localization and content management workflows;
- 3) Language technology applications and resources used in these applications.

The expected uptakes are:

- The 'translate' attribute in HTML5 (already achieved);
- Part of the HTML5 draft;
- Implementation in online MT systems;
- Basic solutions to problems with metadata;
- More to come: close cooperation with the HTML Working Group in W3C;
- RDF 1.1 will allow us to use 'our' metadata;
- MLW-LT metadata will provide hooks for localizing the semantic web;
- Goal: re-using existing implementations and best practices in the localization industry;
- Important: we focus on hooks for integrating LR standards into the Web:

ISO 37 / SC4 and others provide the standards

- Under consideration:
 - Example: TAUS / Let'sMT! categories for 'subject domain' and 'text type';
 - Example: language identifiers JA-t-IT ('Japanese transformed from Italian');
 - Close liaisons: XLIFF, Unicode consortium.

The organisations participating in MLW-LT joined the World Wide Web Consortium (W3C) and created a group to define the standard on metadata within the framework of the Internationalization Activity of W3C. Using the credibility and outstanding position of W3C, MLW-LT will create broad consensus across communities, involving producers of content, localization workers, language technology experts, browser vendors and users.

W3C will help to establish a network of stakeholders in these communities, to ensure the long-term adoption of the metadata standard as well as future opportunities for integrating language technologies and emerging Web technologies. Finally, MLW-LT will lay the technical foundations for new business opportunities, to reach out to a growing linguistic and cultural diversity of Web users in the European Union and its trade partners, and to respond to their specific needs in a timely and cost-effective manner.

The expected effects are the **building of a strong community** and the **creation of tangible outcomes** (reference implementations): metadata in Content Management systems and Localization systems, Online Machine Translation systems and Machine Translation training. We also foresee exploring future interactions between MultilingualWeb and Linked Open Data.

Relevant links:

MultilingualWeb-LT web page	www.multilingualweb.eu
MultilingualWeb-LT Working Group Home Page	www.w3.org/International/multilingualweb/lt/
Press release	www.w3.org/2012/02/mlw-lt.html
Requirements	www.w3.org/International/multilingualweb/lt/wiki/Requirements

Contact information: Felix Sasaki (fsasaki@w3.org) • David Filip (david.filip@ul.ie) • Dave Lewis (dave.lewis@cs.tcd.ie)