



4th workshop The Multilingual Web The Way Ahead

15-16 March 2012

European Commission - Jean Monnet Building
rue Albert Wehrer L-2920 - Luxembourg

15 March 2012

WELCOME

9:00

Piet Verleysen
*European Commission - Resources Director,
responsible for IT at the Directorate-General for
Translation*
Brief welcome address

**Kimmo Rossi - European Commission – DG
INFO E1**
EC language programs – the challenge for the future

**Richard Ishida - W3C Internationalization
Activity Lead & MultilingualWeb Project
Coordinator**
Workshop logistics



KEYNOTE

9:30

Iván Herman – Semantic Web Activity Lead, W3C
What is Happening in the W3C Semantic Web Activity?

This talk gives a short overview of the current work done at the W3C related to the Semantic Web, Linked Data, and related technical issues. The goal is not to give a detailed technical account but, rather, to give a general overview and use this as a basis for further discussions on how that particular technology can be used for the general issue of Multilingual Web

Q&A

10:45

Jan Anders Nelson – Microsoft
*Support for Multilingual Windows Web Apps
(WWA) and HTML 5 in Windows 8*

This talk looks at support for web apps running on Windows8 and steps through the workflow of how a developer can optimize creation of multilingual applications, organizing their app projects to support translation and related tools available that make creation of multilingual apps easy for anyone considering shipping in more than one market, perhaps for the first time.

Tony Graham – Mentea
XSL-FO meets the Tower of Babel

XSL Formatting Objects (XSL-FO) from the W3C is an XML vocabulary for specifying formatting semantics. While it shares many properties with CSS, it is most frequently used with paged media, such as formatting XML documents as PDF. XSL-FO 2.0 is currently under development, and one of its top-level requirements is for further improved non-Western language support. However, the requirement for improved support in XSL-FO 2.0 is actually less specific than the 1998 requirements for XSL 1.0 since we recognised that we didn't have the knowledge and expertise to match our ambitions. For that, we would need more help -- either from individual experts or from the W3C forming more task forces along the lines of the Japanese Layout Task Force to capture and distill expertise for use by all of the W3C and beyond.

Richard Ishida – W3C

Jirka Kosek - University of Economics, Prague
*Ruby & translate flags: Two hot multilingual topics
in HTML5*

This talk will brief attendees on developments related to ruby annotations and a translate flag in HTML5. The Internationalization Working Group at the W3C has been proposing changes to the current HTML5 ruby markup model. Ruby are annotations used in Japanese and Chinese to help readers recognize and understand ideographic characters. It is commonly used in educational material, manga, and can help with accessibility. The Working Group has also proposed the addition of a flag to indicate when text in a page should not be translated. This talk will deliver an up-to-the-minute status on the progress made in these two areas.



Q&A

11:45

Brian Teeman – Joomla!Shack University
Building Multi-Lingual Web Sites with Joomla! the leading open source CMS

Joomla is the leading Open Source CMS used by over 2.8% of the web and by over 3000 government web sites. With the release in Jan 2012 of the latest version (2.5) building and creating truly multi-lingual web sites has never been easier. This presentation will show how easy it is to build real multi-lingual web sites and not to rely on automated translation tools.

Loïc Dufresne de Virel – Intel Corp.
How standards (could) support a more efficient web localization process by making CMS - TMS integrations less complicated.

As Intel just deployed a new Web Content Management system, which we integrated into our TMS, We had to deal with multiple challenges, and also faced a great deal of complexity

and customization. In this 15-min talk, I will look into what we did well, what we did wrong, what we could have done better, and will try to put a dollar figure on the cost of ignoring a few standards...

Gerard Meijssen – Wikimedia Foundation
Translation and localisation in 300+ languages ... with volunteers. The best practices.

There are over 270 Wikipedias and over 30 languages are requesting one. These languages represent most scripts and represent small and large populations. The WMF enables the visibility of text with web fonts, it supports input methods, there is a big multi application localisation platform at translatewiki.net and we are implementing translation tools for our “pages” for documentation and communication to our communities. To do this, we rely on standards. Standards that get more relevance as they are implemented in more and more places in our software. Some standards are not supporting all the languages that we support.

CREATORS

Q&A

14:00

Spyridon Pilos – Directorate General for Translation - European Commission
The Machine Translation Service of the European Commission

The Directorate General for Translation (DGT) has been developing since October 2010 a new data-driven machine translation service for the European Commission. MT@EC should be operational in the second semester of 2013. One of the key requirements is for the service to be flexible and open: it should enable, on one hand, the use of any type of language resources and any type of MT technology and, on the other, facilitate easy access by any client (individual or service). The speaker will present the approach taken and highlight problems identified, as pointers to broader needs that should be addressed.

Matjaz Horvat – Mozilla
Live website localization

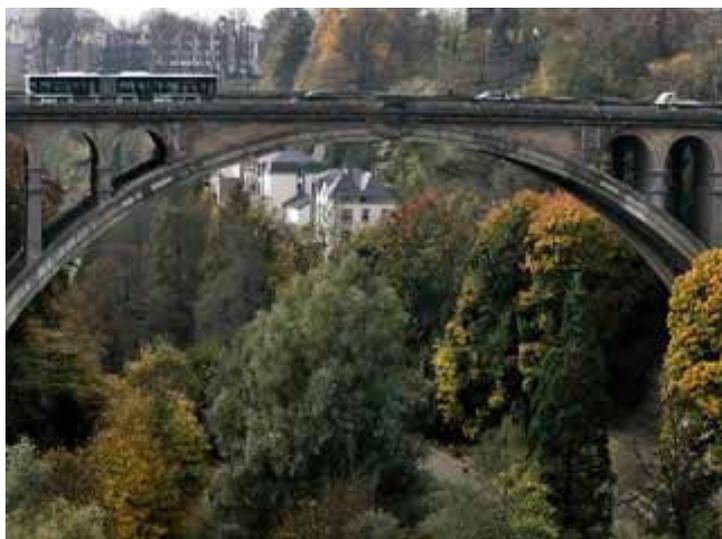
Pontoon is a live website localization tool developed at Mozilla. Instead of extracting website strings and then merging translated strings back, Pontoon can turn any website into editable mode. This enables localizers to translate websites in-place and also provides context and spatial limitations.

David Lewis – Centre for Next Generation Localisation at Trinity College Dublin
Meta-data interoperability between CMS, localisation and machine translation: Use Cases and Technical Challenges

January 2012 saw the kick off of the MLW-LT (“Multilingual Web - Language Related Technologies”) Working Group at the W3C as part of the Internationalization Activity. This WG will define meta-data for web content (mainly HTML5) and “deep Web” content (CMS or XML files from which HTML pages are generated) that facilitates content interaction with multilingual language technologies such as machine translation, and localization processes. The WG brings together localisation and content management companies with content and language meta-data research expertise, including strong representation from the Centre for Next Generation Localisation. This talk will present three concrete business use cases that span CMS, localisation and machine translation functions. It discusses the challenges in addressing these cases with existing metadata (e.g. ITS tags) and the technical requirements for additional standardised metadata. This talk will be complemented by a breakout session to allow attendees to voice their comments and requirements in more detail, in order to better inform the working group.

LOCALISERS

Q&A



15:00**Peter Schmitz – Publications Office**
Common Access to EU Information based on semantic technology

The Publications Office is setting up a common repository to make available at a single place all metadata and digital content related to public official EU information (law and publications) in a harmonised and standardised way in order: to guarantee to the citizen a better access to law and publications of the European Union; to encourage and facilitate reuse of content and metadata by professionals and experts. The common repository is based on semantic technology. At least all official languages of the EU are supported by the system, thus the system is a practical example of a multilingual system accessible through the WEB.

Paul Buitelaar –DERI - National University of Ireland, Galway
Ontology Lexicalisation and Localisation for the Multilingual Semantic Web

Although knowledge processing on the Semantic Web is inherently language-independent, human interaction with semantically structured and linked data will be text or speech based – in multiple languages. Semantic Web development is

therefore increasingly concerned with issues in multilingual querying and rendering of web knowledge and linked data. The Monnet project on 'Multilingual Ontologies for Networked Knowledge' provides solutions for this by offering methods for lexicalising and translating knowledge structures, such as ontologies and linked data vocabularies. The talk will discuss challenges and solutions in ontology lexicalisation and translation (localisation) by way of several use cases that are under development in the context of the Monnet project.

Tadej Štajner – Jožef Stefan Institute
Cross-lingual named entity disambiguation for concept translation

The talk will focus on our experience in developing an integrated natural language processing pipeline, consisting of several distinct components, operating across multiple languages. We will demonstrate a cross-language information retrieval method that enables reuse of the same language resources across languages, by using a knowledge base in one language to disambiguate named entities in text, written in another language, as developed in the Enrycher system (enrycher.ijs.si). We discuss the architectural implications of this ability on the development practices and its prospects as a tool for automated translation of specific concepts and phrases in a content management system.

Q&A

16:30**Annette Marino – European Commission**
Web translation, public service & participation

For most Europeans, the internet provides the only chance they have for direct contact with the EU. But how can we possibly inform, communicate and interact with the public if we don't speak their language on the web? With the recent launch of the European Citizens' Initiative website in 23 languages, there's no doubting the role of web translation in participatory democracy, or the Commission's commitment to a multilingual web presence. But as well as enthusiasm, we need understanding – of how people use websites and social media, and what they want from us – so that we can make best use of translation resources to serve real needs. As the internet evolves, the Commission is on a steep learning curve, working to keep up with the possibilities - and pitfalls - of web communication in a wide range of languages.

Murfah Hossari – University College Dublin
Localizing to right-to-left languages: main issues and best practices

Internationalization and Localization efforts need to take extra care when dealing with right-to-left languages due to specific features those languages have. Many localization issues are specific to right-to-left languages. The talk will try to categorize those issues that face localizers when dealing with right-to-left languages with special focus on text direction part and handling bidirectionality. The talk will also mention best practices and areas for improvements.

Nicoletta Calzolari – CNR-ILC
The Multilingual Language Library

The Language Library is a quite new initiative – started with LREC 2012 – conceived as a facility for gathering and making available, through simple functionalities, all the linguistic knowledge the field is able to produce, putting in place new ways of collaboration within the Language Technology community. Its main characteristic is to be collaboratively built, with the entire community providing/enriching language resources by annotating/processing language data and freely using them. We exploit today's trend towards sharing for initiating a collective movement that works also towards creating synergies and harmonisation among different annotation efforts that are now dispersed. The Language Library could be considered as the beginning of a big Genome project for languages, where the community will collectively deposit/create increasingly rich and multi-layered linguistic resources, enabling a deeper understanding of the complex relations between different annotation layers.

Fernando Serván – Food and Agriculture Organization of the United Nations
Repurposing language resources for multilingual websites

This presentation will address lessons learned regarding the reuse of language resources (translation memories in TMX, terminology databases in TBX) to improve content and language versions of webpages. It will address the need for better integration between existing standards, the needs for interoperability and the areas where standards and best practices could help organizations with a multilingual mandate.

Q&A

19:30**At the Hotel Parc Belle-Vue**

To further promote networking among attendees, there will be a reception from 7.30pm to 9.30pm in the Hotel Parc Belle-Vue, about a 5 minute walk from the city centre. Entry is

free to workshop participants.

The location is 5 Avenue Marie-Thérèse, Luxembourg.



16 March 2012



SET UP

9:00

**Jaap van der Meer
TAUS**

Explanation of the format for the morning, and selection of discussion topics. Topics are suggested by participants, and the most popular are allocated to breakout groups. A chair is chosen for each group from volunteers.



OPEN
SPACE

9:30

Break-out discussions.

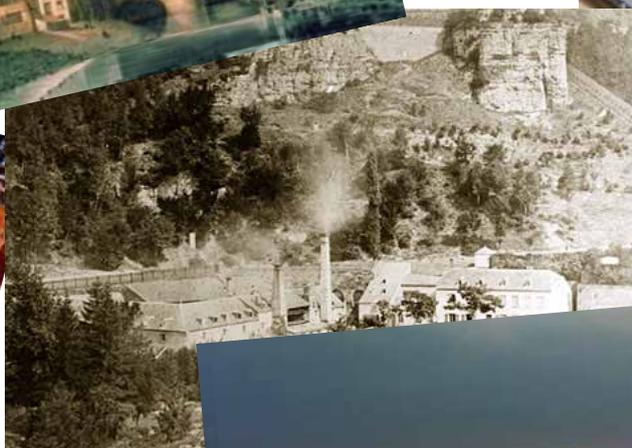
Various locations are available for breakout groups. Participants can join whichever group they find interesting, and can switch groups at any point. Group chairs facilitate the discussion and ensure that notes are taken to support the summary to be given to the plenary.

11:30

Group reports and discussion.

Everyone meets again in the main conference area and each breakout group presents their findings. Other participants can comment and ask questions.

**12:45 - 13:00
Wrap up & Close**



Photos source: Wikipedia, Hotel Parc Belle-Vue, R. Ishida, M. T. Carrasco-Benitez

Organised by:



Sponsored by:



Brochure design: Patrizia Andronico,
Scientific Secretariat, IIT-CNR, Pisa, Italy