

Open architecture for multilingual social networking

Position paper for the W3C Workshop on the Future of Social Networking [FSN], 15-16 January 2009, Barcelona
M.T. Carrasco Benitez

1. Abstract

The purpose of this paper is to sketch an open architecture for all the multilingual aspects of social networking. This architecture should be comprehensive and address well-trodden fields such as localization, and more advanced multilingual techniques to facilitate the communication among users (e.g., translation). Also this architecture should elucidate what mechanisms are specific to social networking and what mechanisms should be addressed at a lower layer such as *multilingual web sites*.

2. Common experience

There should be one common experience for all stakeholders:

- **Users** should expect the same multilingual behavior when using different client programs or visiting different sites.
- **Webmasters** should expect the same data structure or mechanisms when building systems.

Facilitating and enjoying this common experience entails standardization of data structures and mechanisms. Lets consider two examples:

- **Browsers** could contain a *language button* (in the row with File, Edit, etc) that becomes enabled when other languages are available. The language button could also be a more general *metadata button* with all the *variants*; i.e., media type, charset and encoding.
- **Servers** should expect the same multilingual data structure. The absence of this data structure means that each application has to craft this facility; having the same data structure means that servers (or other programs) would know how to process this data structure directly.

3. Break nothing

Stating the obvious, adding these facilities must not break any of the existing standards. Indeed, sometimes it just means implementing mechanisms already in the standards such as *transparent content negotiation* [TCN]. For example, in the case of the language button it means that servers have to return the available languages in the HTTP header, and browsers must process this data and activate the language button, if necessary.

4. Framework

There should be one comprehensive architecture encompassing all the multilingual aspects of social networking:

- **Wider context:** the relation among *multilingual parallel texts* [OAMPT], multilingual web sites [OAMWS] and multilingual social networking.
- **Localization:** traditional *internationalization* (I18N) and *localization* (L10N).
- **Translation:** more advanced techniques to facilitate the communication among users with different languages.

4.1. Wider context

The wider context might be viewed as the following hierarchy:

- **Multilingual parallel texts:** the larger framework.
- **Multilingual web sites:** a particular case of multilingual parallel texts.
- **Multilingual social networking:** a particular case of multilingual web sites.

The relation might be more complex. In general it might be as the above hierarchy, though with extensions. The point is to identify the appropriate layer in which to address the different mechanisms.

4.2. Localization (L10N)

This field is well developed and understood. One should develop internationalized software so it can be localized for different cultural environments; even if the original needs are for one culture. Anyhow, in our world today there are high chances that the software would have to be adapted to other cultures and languages; the first time that this is needed would largely pay for taking into account L10N.

4.3. Translation

This refers to techniques to facilitate the communication (e.g., chat or email) among the users with different languages in the social networking. It should allow many different translation engines and translation engines techniques.

Some of this translation engines could be very simple, such as a table with multilingual parallel texts. For example, a user selects a sentence from the table with parallel text in Spanish because it is his mother tongue and the system will send the corresponding English sentence to his contact.

There is an overlapping with L10N as multilingual parallel texts are also used for L10N.

5. Legal and miscellaneous

5.1. Disclaimer

This document represents only the views of the author and not necessarily the views of any other parties. In particular, it does not necessarily represent the opinion of the European Commission, his employer.

5.2. Comments

Comments are very welcome: <ca{{AT}}dragoman{{DOT}}org>. For follow-ups of this paper see: <http://MultilingualWebSites.org>.

6. References

[DC] Dublin Core Metadata Element Set, Version 1.1
<http://dublincore.org/documents/dces>

[DITA] Darwin Information Typing Architecture
<http://dita.xml.org>

[FSN] W3C Workshop on the Future of Social Networking
<http://www.w3.org/2008/09/msnws>

[ISG] Interinstitutional style guide
<http://publications.europa.eu/code/pdf/370000en.htm>

[ISO-639-1] Codes for the Representation of Names of Languages -- Part 1: Alpha-2 code
http://www.loc.gov/standards/iso639-2/php/code_list.php

[LISA] Localization Industry Standards Association
<http://www.lisa.org>

[MED] Multilingual Electronic Dossier, M.T. Carrasco Benitez.
<http://dragoman.org/med>

[MWP] Panel, The Multilingual Web: anything missing ?, 9th International WWW Conference
<http://www9.org/w9-panels.html>

[OAMPT] Open architecture for multilingual parallel texts, M.T. Carrasco Benitez
<http://arxiv.org/pdf/0808.3889v1>

[OAMWS] Open architecture multilingual web sites, M.T. Carrasco Benitez
<http://MultilingualWebSites.org/oamws.pdf>

[PARTEXT] Generation of Multilingual Parallel Texts with XML, M.T. Carrasco Benitez
<http://www.dragoman.org/partext>

[SEIBUD] Système d'édition informatisé du Budget
<http://budget.opoce.cec.eu.int>

[SMT] Moses
<http://www.statmt.org/moses>

[TB] Text boundaries, Unicode Standard Annex #29, M. Davies
<http://www.unicode.org/reports/tr29>

[TCN] Transparent Content Negotiation in HTTP
<http://www.ietf.org/rfc/rfc2295.txt>

[TEC] Travaux en cours, M.T. Carrasco Benitez
<http://dragoman.org/tec.tiff>

[TMX] TMX 1.4b Specification, Y. Savourel
<http://www.lisa.org/standards/tmx/tmx.html>

[UNICODE] The Unicode Consortium
<http://unicode.org>

[WIMS] Web Internationalization & Multilinguism Symposium
<http://www.w3.org/International/Sevilla-96>

[WINTER] Winter (Web Internationalization and Multilinguism), Lapsed Internet-draft, M.T. Carrasco Benitez.
<http://www.w3.org/International/tomas.carrasco.benitez.html>

[XEP-LT] XEP-0171: Language Translation, B. Fletcher et al.
<http://www.xmpp.org/extensions/xep-0171.html>

[XMPP] Extensible Messaging and Presence Protocol
<http://xmpp.org>

Ello va seco y sin llover

