



Web as a medium

M.T. Carrasco Benitez

Dagstuhl, 6 September 2012



Three aspects

- 1 Common ground**
- 2 Multilingual web sites (MWS)**
- 3 Multilingual linked open data (MLOD)**



1 Common ground - crossroad

- **Web technology**
- **NLP - RBMT, SMT**
- **Human translation - CAT**
- **Multilingual parallel texts**

Common terminology



Common ground - layers

A Web technology

- **URI**

B Multilingual web

- **Multilingual web sites (MWS)**

C NLP web

- **MSW**



Common ground - interoperability

- **Standards**
- **Using existing standards**
- **Gap: fill with a new standard**
 - **Scope**
 - **Break nothing**
 - **Make it as simple as possible**
 - **Lower the barrier**
 - **Non-specific tools**
 - **Implementation - IETF two**



2 Multilingual web sites - views

- **Users**
 - **Same behaviour of the sites**
 - **Monolingual**
- **Webmaster**
 - **Easy creation of the sites**
 - **Multilingual**



Multilingual web sites - techniques

- **Addressing with URI**
- **Browser side**
 - **Language button**
 - **Machine clients**
- **Server side**
 - **Variant page**
 - **<http://example.com/foo/var>**



3 MLOD - overview

- **Structure the data, how is secondary**
 - **Granularity**
 - **Description**
- **Multilingual parallel texts**
 - **Small data - record**
 - **Big data – database**
 - Multilingual Dataset Format (**muset**)
- **Addressing with URI**
 - **<http://example.com/foo/de>**



MLOD - nature

- **Data is imperfect**
 - **Quality**
 - **Legal**
 - **Accept all**
 - **Be aware of the imperfections**
- **Data is hard work**
 - **Pâtissier - cakes**
 - **Miller - flour**



Summary

Multilingual web standards

<http://dragoman.org/dagstuhl>