



ITS 2.0

Foster Multilingual Schema (and data?)
creation

Felix Sasaki

DFKI / W3C Fellow

March 2014

“WHY?”

1. Mail from A to public list: “I would like to translate ontology xyz into language ... – what parts have to be translated”
2. Answer from ontology editor
3. Clarification question - mail from A
4. Answer from ontology editor
5. Acknowledgement from A: “Thank you, everything clear to me!”

- Two persons time lost
 - Mail writing & reading, ontology file “reading”, ...
- In total up to ½ a day
- How could this be avoided?
- Provide machine readable information on what parts of the ontology should be translated

“HOW?”

- Machine readable metadata to foster automated content creation
- Input content: XML or HTML5 – or RDF in RDF/XML
- Metadata: so called ITS “data categories”
- Example just uses ITS “Translate”

- Translate
- Localization Note
- Terminology
- Directionality
- Ruby
- Language Information
- Elements Within Text
- Domain
- Text Analysis /
Disambiguation
- Locale Filter
- Provenance
- External Resource
- Target Pointer
- Id Value
- Preserve Space
- Localization Quality Issue
- Localization Quality Rating
- MT Confidence
- Allowed Characters
- Storage Size

```
<messages ...>
```

```
<msg num="123">Click Resume Button on Status  
Display or
```

```
<panelmsg its:translate="no">CONTINUE</panelmsg>  
Button on printer panel</msg>
```

```
</messages>
```



```
<its:rules version="2.0" ...>  
  <its:translateRule translate="no"  
    selector="//code"/>  
</its:rules>
```

- Relies on [XPath](#) or CSS selectors
- Global and local ITS work like CSS
 - Local ITS overrides global ITS
 - Global ITS rules: the last one wins

- RDF (ontologies and instances) can be stored as RDF/XML
- ITS 2.0 global rules provide a machine readable approach to specify what is translatable and what not
- The following slide shows how the mail conversation could have been avoided

```

<its:rules ...>
  <its:translateRule selector="//*" translate="no"/>
  <its:translateRule
    selector="//rdfs:label[lang('en')] |
    //dct:title[lang('en')] | rdfs:comment[lang('en')]"
    translate="yes"/></its:rules>
  
```

- General global rule: **all elements** are not translated
- Exception: **various elements with human readable text in English** (= source language) are to be translated

- Problem solved 😊
- A lot of ITS 2.0 tool support for XML & HTML5 content

http://www.w3.org/International/its/wiki/ITS_Implementations

- Example: Okapi framework
 - Open source modules and WYSIWYG tools for various localization related tasks
 - Screenshot of ITS 2.0 aware workflow: 1) segmentation 2) translation memory lookup 3) machine translation 4) quality check 5) XLIFF generation
 - RDF content can make use of all above steps right away!

- Raw Document to Filter Events
- Segmentation**
- Leveraging
- Microsoft Batch Translation
- Quality Check
- Rainbow Translation Kit Creation

- 1)
- 2)
- 3)
- 4)
- 5)

Segment the source text using the following SRX rules:

{rootDir}/data/defaultSegmentation.srx



Edit...

Segment existing target text using the following SRX rules:



Edit...

Options

Behavior if input text is already segmented:

- Keep existing segmentation
- Overwrite existing segmentation (resegment)
- Keep existing segmentation, segment further against the SRX rules

- Copy source into target if no target exists
- Verify that a target segment matches each source segment when a target content exists
- When possible force the output to show the segmentation
- Renumber code IDs

Add Step...

Remove Step

Move Up

Move Down

Load...

New

Save

Save As...

Step Help

Apply SRX segmentation to the text units content of a document. Expects: filter events. Sends back: filter events.

Help

Cancel

Close

Execute

- Problem solved only for RDF/XML serialization
- What about Turtle, RDFa, N3, ...?
- Option 1: Recommend RDF/XML conversion for localization
 - Benefit: existing tools can be used
 - Drawback: formulation of ITS 2.0 rules relies on XPath
- Option 2: Define ITS 2.0 metadata syntax for other serializations
 - Benefit: no need for XPath in non XML syntaxes
 - Drawback: tool support not likely to happen soon
 - Statements in RDF about strings don't work – “ITS to NIF” conversion might work, but adds another layer



ITS 2.0
Foster Multilingual Schema (and data?)
creation

Felix Sasaki
DFKI / W3C Fellow
March 2014
