The Journey of the W3C Internationalization Tag Set
– Current Location and Possible Itinerary –

Christian Lieske (SAP AG), Felix Sasaki (DFKI), Yves Savourel (ENLASO),
Richard Ishida (W3C), Jirka Kosek (University of Economics in Prague)

W3C Workshop: A Local Focus for the Multilingual Web, 20-21 September 2011, Limerick
Agenda

1. W3C Internationalization Tag Set (ITS) in a Nutshell
2. W3C ITS in Commercial and Open Source Tools
3. W3C ITS for/in popular Formats
4. Suggested Enhancements to W3C ITS
5. Outlook
Knowledge Architect

- Content engineering and process automation (including evaluation, prototyping and piloting)
- Main field of interest: Internationalization, translation approaches and natural language processing
- Contributor to standardization at World Wide Web Consortium (W3C), OASIS, the Unicode Consortium the European Commission and elsewhere
- Degree in Computer Science with focus on Natural Language Processing and Artificial Intelligence

This presentation is **purely personal** — our employers have no responsibility for any information contained here
The W3C Internationalization Tag Set (ITS) is a W3C Recommendation.

ITS helps to internationalize XML-based Contents by allowing standardized statements.

Content that has been internationalized with ITS can be more easily processed by humans and machines.

ITS is an important ingredient to the W3C Note „Best Practices for XML Internationalization“
W3C ITS in a Nutshell – It’s Abound Annotations

Which parts have to be translated?

Anything I need to know when working on this?

Does the “x” element split a run of text into two linguistic units?
W3C ITS in a Nutshell – Why and How (Details)?

Scenario: Configure a spell checker or communicate to a translator, so that only natural language content is being considered. Answer a couple of questions for getting the configuration right.

Language of the content?
Terms?
Codes?
Footnotes?
Foreign language expressions?
Annotations for readers?

Adapted from Yves Savouel http://www.opentag.com/xfaq_charrep.htm#char_nonasciitag
### W3C ITS in a Nutshell – What (Details)?

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate</td>
<td>Mark whether the content of an element or attribute should be translated or not” (other processes than translation can use Translate)</td>
</tr>
<tr>
<td>Localization Note</td>
<td>Communicate notes to localizers about a particular item of content</td>
</tr>
<tr>
<td>Terminology</td>
<td>Mark terms and optionally associate them with information, such as definitions</td>
</tr>
<tr>
<td>Directionality</td>
<td>Specify the base writing direction of blocks, embeddings and overrides for the Unicode bidirectional algorithm</td>
</tr>
<tr>
<td>Ruby</td>
<td>Provide a short annotation of an associated base text, particularly useful for East Asian languages</td>
</tr>
<tr>
<td>Language Information</td>
<td>Express the language of a given piece of content</td>
</tr>
<tr>
<td>Elements Within Text</td>
<td>Identify how an element behaves relative to its surrounding text, eg. for text segmentation purposes</td>
</tr>
</tbody>
</table>
W3C ITS in Tools – Okapi

Set of libraries and components for writing translation- and localization-related tools

Applications, Tools, Scripts

Steps

Filters
Connectors (TM, MT, etc.)
Other Components (Segmenter, Tokenizer, etc.)

Core Libraries (Resource model, Event model, APIs, Annotations, etc.)

Web Resources

Downloads, source, bugs, etc: http://code.google.com/p/okapi

Help and documentation: http://okapi.opentag.com

Users Group: http://code.google.com/p/okapi
W3C ITS in Tools – ITS2XLIFF Tool, More …

http://fabday.fh-potsdam.de/~sasaki/its

http://gitorious.org/itstool

The W3C Internationalization Tag Set (ITS) defines data categories and their implementation as a set of elements and attributes called the Internationalization Tag Set (ITS). ITS is designed to be used with schemes to support the internationalization and localization of schemas and documents. Please take a look at the collection for more ITS-related information (amongst others presentations, and references to implementations).

Aside: The W3C ITS Interest Group gives ample opportunities to get involved.

The XML Localization Interchange File Format (XLIFF) gives any documentation or software provider a single interchange file format that can be understood by any localization provider. This page allows you to generate XLIFF 1.2 from XML files for which W3C ITS rules is available (local markup or global rules). Currently, only the ITS ITS 1.9 Translate is supported.

The central functionality behind the page is the ITS General Detector, an XSLT-based ITS processor which decorates an input document with information for ITS data categories. To be learned, take a look at the XLIFF Excision and Merging.

For the generation of an XLIFF file, use this form:

```
XML File
```

Generate

For the integration of translated content into the original XML file, use this form:

```
XLIFF File for conversion back to XML
```

Generate
W3C ITS for Popular Formats

- Android Strings resources
- DITA
- DocBook
- Glade XML
- Java XML Properties
- Office Open XML (OOXML) 1.0
- POWDER
- OpenDocument (ODF) 1.0
- XHTML 1.1
- XML Spec

http://www.w3.org/International/its/wiki/RulesRepository
W3C ITS in Popular Formats

Several well-known XML document types

- [http://www.w3.org/TR/xml-i18n-bp/#Modularization](http://www.w3.org/TR/xml-i18n-bp/#Modularization)

W3C ITS-enabled schemas for DocBook do already exist

- [http://www.docbook.org/xml/5.0/rng/dbits.rnc](http://www.docbook.org/xml/5.0/rng/dbits.rnc)
- [http://www.docbook.org/xml/5.0/rng/dbits.rng](http://www.docbook.org/xml/5.0/rng/dbits.rng)

The *translate* data category is under discussion for HTML5

- [http://www.w3.org/Bugs/Public/show_bug.cgi?id=12417](http://www.w3.org/Bugs/Public/show_bug.cgi?id=12417)
Suggested Enhancements to W3C ITS

| **targetPointer** | • Where is the translation or where should it go? |
| **idValue**      | • How can I identify things? |
| **Local „Elements within Text“** | • How to say „Do not segment“ in place? |
| **Whitespaces**  | • Can I do something to whitespace? |
| **„Context“**    | • What can you tell me in addition? |
| **localeSpecificContent** | • Is this only for one country? |
| **Automated Language Processing** | • Does this lend itself to automatic processing? |

http://www.w3.org/International/its/wiki/IssuesAndProposedFeatures
Outlook

! Additional tool support

! Further usage scenarios (forthcoming MultilingualWeb-LT project)

? Continue to run ITS IG

? ITS 2.0
Digging Deeper

Specification
http://www.w3.org/TR/its/
Best Practice Note
http://www.w3.org/TR/xml-i18n-bp
W3C ITS Interest Group
http://www.w3.org/International/its/ig/

Internationalization and Localization of XML: Introducing "ITS"

Christian Lieske
Sebastian Rhatz
Felix Sasaki

Slides:
http://www.w3.org/2006/Talks/0518-xtech-its/
Thank You!

Contact information:

Christian Lieske  
christian.lieske@sap.com  
www.sap.com

Felix Sasaki  
felix.sasaki@dfki.de  
www.dfki.de

Yves Savourel  
ysavourel@translate.com  
www.translate.com

Richard Ishida  
ishida@w3.org  
www.w3.org

Jirka Kosek  
jirka@kosek.cz  
www.kosek.cz
Disclaimer

All product and service names mentioned and associated logos displayed are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

This document may contain only intended strategies, developments, and is not intended to be binding upon the authors or their employers to any particular course of business, product strategy, and/or development. The authors or their employers assume no responsibility for errors or omissions in this document. The authors or their employers do not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

The authors or their employers shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials. This limitation shall not apply in cases of intent or gross negligence.

The authors have no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages.