

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

Presenter



Christian Lieske

**SAP Language Services
Globalization Services
SAP AG**

- 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

W3C ITS in a Nutshell – What and Why (Overview)?

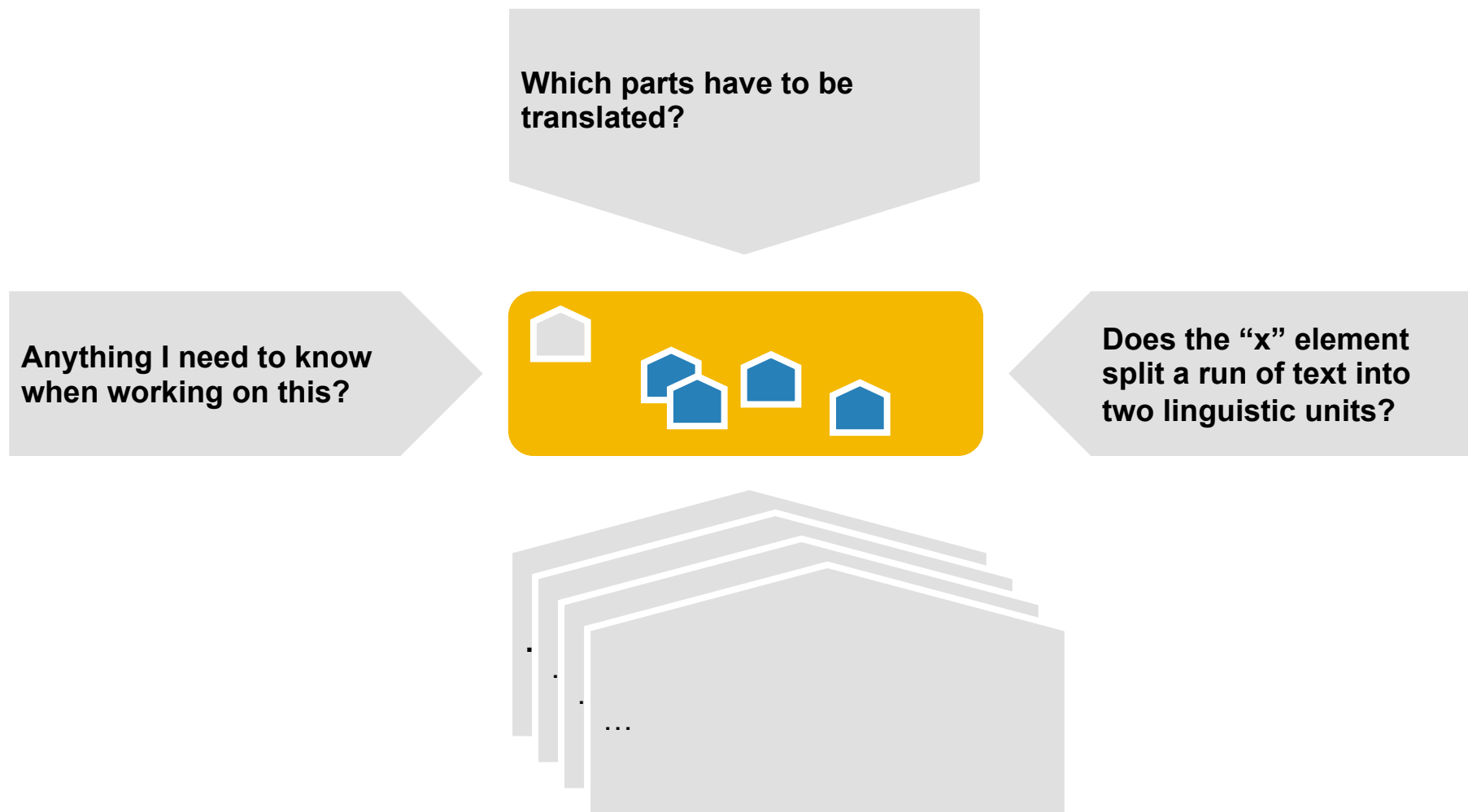
The W3C Internationalization Tag Set (ITS) is a W3C Recommendation.

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 About Annotations



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.



Adapted from Yves Savouel http://www.opentag.com/xfaq_charrep.htm#char_nonasciitag

W3C ITS in a Nutshell – What (Details)?

Translate

Mark whether the content of an element or attribute should be translated or not". (other processes than translation can use Translate)

Localization Note

Communicate notes to localizers about a particular item of content

Terminology

Mark terms and optionally associate them with information, such as definitions

Directionality

Specify the base writing direction of blocks, embeddings and overrides for the Unicode bidirectional algorithm

Ruby

Provide a short annotation of an associated base text, particularly useful for East Asian languages

Language Information

Express the language of a given piece of content

Elements Within Text

Identify how an element behaves relative to its surrounding text, eg. for text segmentation purposes

W3C ITS in Tools – SDL Trados Studio, XTM

The image is a collage of three screenshots related to W3C ITS in tools.

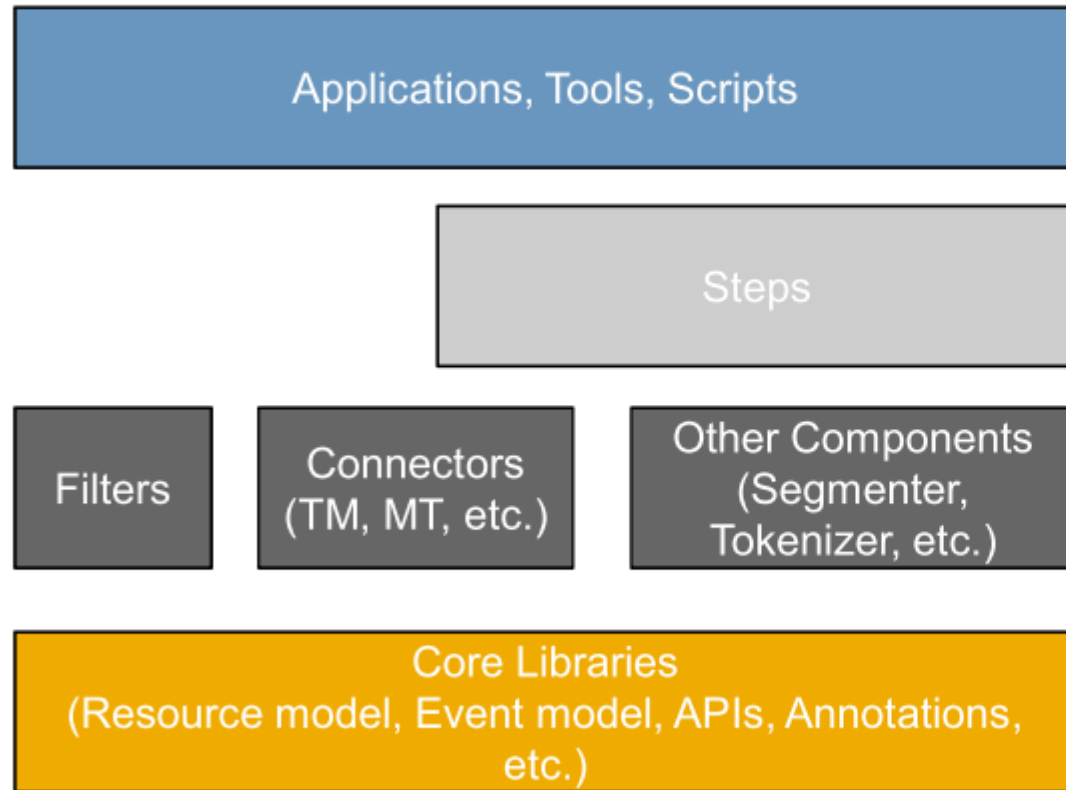
Top Left: SDL Trados Studio - Project 1
 This screenshot shows the 'Project Settings' dialog box. The 'File type information' section is active, showing 'File type name: XML: W3C ITS Compliant'. The 'File type icon' is set to 'assembly://Sdl.FileTypeSupport.Native.Xml_1.1/Sdl.FileTypeSupport.Native'. The 'Settings file' is 'C:\Documents and Settings\vd025418\My Documents\SDL Trados Studio\Project'. The 'File type identifier' is 'XML: ITS 1.0 v 1.1.0.0'. The 'Name of individual document' is 'ITS XML Document'. The 'Name of document category' is 'ITS XML Document'. The 'File dialog wildcard expression' is '*.xml'. The 'Description' field contains: 'XML documents using local files, which do not have any dependencies and processing of localizable content...'. The 'Files' pane on the left shows a list of file types including SDL XLIFF, TRADOS tag, Microsoft Word, Microsoft PowerPoint, Microsoft Excel, OpenOffice, Rich Text Format (RTF), HTML, Adobe FrameMaker, Adobe InDesign, XLIFF, PDF, Comma Delimited Text (CSV), Java Help content, XML Microsoft .NET Resources, XML OASIS DITA Compliant, XML OASIS DocBook 4.5 Compliant, Duplicates, Parser rules, Namespaces, Entity conversion, File detection, Writer settings, Whitespace handling, Preview, Tag verification, XML validation, XML Any XML, Text, Verification, and Language Pairs.

Top Right: XTM International Website
 This screenshot shows the XTM International website. The header includes the XTM logo, contact information (Phone: +44 1753 480 469, Email: sales@xtm-int.com), and navigation links (Home, Company, Products, Customers, Partners, Demos, Resources). The main content area features the heading 'XTM International Standards Support'. Below this, there is a paragraph stating: 'XTM International is totally committed to the development and evangelization of Open Standards for the localization industry. We are passionate about Open Standards and have worked tirelessly within USA, W3C and OASIS to provide a comprehensive Open Architecture for Authoring and Localization based on Open Standards. The culmination of this activity has been the creation of a new OASIS Technical Architecture Committee called OAXAL (Open Architecture for XML Authoring and Localization)'. A list of supported Open Standards is provided: W3C ITS, Unicode TR29, LISA OSCAR SRX, LISA OSCAR xmftm, LISA OSCAR TMB, LISA OSCAR GMX, OASIS XLIFF, and OASIS OAXAL. There are also buttons for 'Free Trial', 'Register Here', 'XTM Cloud', and 'Buy Here'. A 'Case Studies' section is visible with a 'Select a Case Study' dropdown.

Bottom: Translation Comparison Table
 This screenshot shows a translation comparison table for the text 'Frequently Asked Questions about the Darwin Information Typing Architecture'. The table compares the source text (left) with the target text (right) and includes a quality score (rightmost column). The source text is: '1 Frequently Asked Questions about the Darwin Information Typing Architecture. 2 The Frequently Asked Questions for DITA have been collected from questions asked on the DITA Forum and among other communities. 3 Some of the answers reflect DITA Best Practices. 4 General DITA questions Why is "Darwin" in the name of this architecture? 5 The entire name of the architecture has this combined explanation: 6 The Darwin one: it uses the principles of specialization and inheritance'. The target text is: 'Häufig gestellte Fragen zur "Darwin-Information-Typing-Architecture" Die-h häufig gestellte Fragen'. The quality scores are: T+, SHD+, T+, FQ+, FA+, and U+.

W3C ITS in Tools – Okapi

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



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>

ITS2XLIFF generation version 0.6

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 schemas to support the internationalization and localization of schemas and documents. Please take a look at [link 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 1.0 Translate is supported.

The service/functionality behind the page is the [ITS General Decorator](#), an XSLT-based ITS processor which decorates an input document with information for ITS data categories. To be able to use the [XLIFF Extraction and Merging](#).

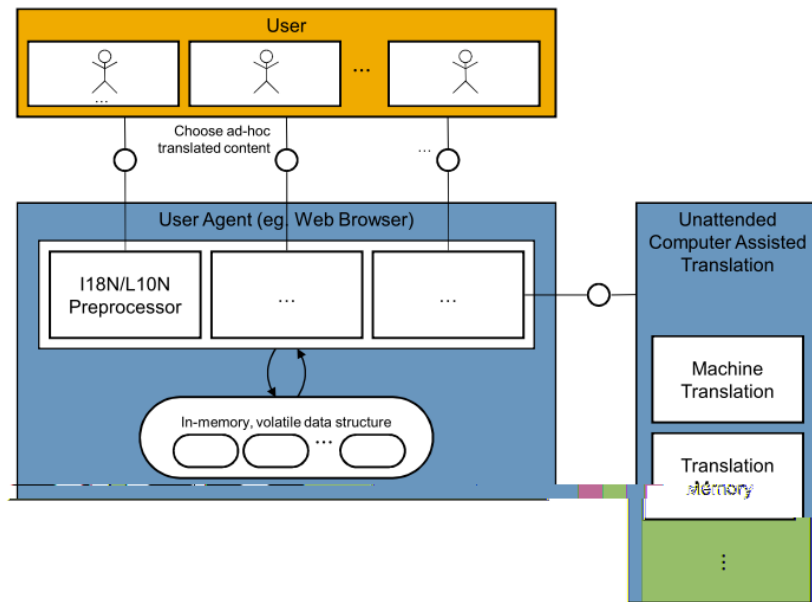
For example use the samples in the [ITS Test Suite](#).

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

XML File

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

XLIFF File for conversion back to XML



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>

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.rng>

The *translate* data category is under discussion for HTML5

- 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/>

W3C Internationalization (I18n) Activity
Making the World Wide Web world wide!

Home About Groups Topics Techniques Re...

Internationalization Tag Set Interest Group Home Page

Key information: Mission - News - Work in p...

About the Internationalization Tag Set

The Internationalization (I18N) Interest Group Activity.

Mission: To foster a community of users of the 1.0, by promoting its adoption within and outside further development.

For more details, see the Interest Group Chart...

How to Join

Subscribe to the public-i18n-its-ig@w3.org mailing list. You have no obligations if you participate via the mailing list (no subscription necessary).

Meetings

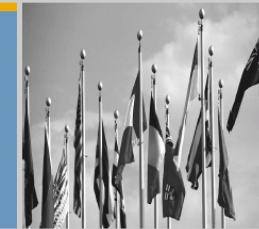
The Interest Group is having a regular call once a year. See the list of previous meetings...

The Internationalization Tag Set

Richard Ishida
W3C Internationalization Activity Lead

<http://www.w3.org/2006/Talks/10-lrc-its/slides/Slide0010.html>

Standards-based Translation with W3C ITS and OASIS XLIFF



Christian Lieske (SAP AG)
Felix Sasaki (Fachhochschule Potsdam)
Yves Savourel (Enlaso)
Bryan Schnabel (Tektronix)

tcworld
conference 2009

Rhein-Neckar-Hallen Wiesbaden
Thursday, 5th November 2009
8:45 - 10:30 am, Room 1A/3

<http://www.tekom.de/upload/2913/>

http://www.tekom.de/upload/2913/12_Sasaki_Lieske.pdf





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.