Best Practices and Standards for Improving Globalization-related Processes

Christian Lieske (SAP AG)

W3C Workshop: The Multilingual Web - Where Are We? 26-27 October 2010, Madrid
Agenda

1. Globalization-related Processes
2. Best Practices and Standards
3. Reality Check

for two types of audiences

… introduce to currently available best practices and standards
   … begin to identify gaps
   … Localization standards & tools
   TM and terminology databases
   MT; Crowd-sourcing; Cloud based issues

Workshop/Session Objectives
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 and elsewhere
- Degree in Computer Science with focus on Natural Language Processing and Artificial Intelligence

This presentation is purely personal — my employer has no responsibility for any information contained here
Globalization – Making This Happen

SAP SOLUTIONS
Transforming your business into a break-out business. Turn inefficiencies into opportunity.

SAP SERVICES
At SAP, providing superior services and support throughout your solution lifecycle is just part of our standard operating procedure.

SAP ECOSYSTEM/PARTNERS
SAP's industry-focused ecosystem enables customers, partners, and individuals to accelerate innovation and improve business performance.

ABOUT SAP
With more than 120,000 customers, it’s no wonder SAP is the world leader in collaborative business software.

MEHR ERREICHEN
FINDEN SIE HERAUS, WIE UNTERNEHMEN MIT SAP IHRE BESTE GEHEN, UND MEHR.

RUN BETTER
SEE HOW SAP HELPS THE BEST-RUN BUSINESSES DO WHAT THEY DO BEST, EVEN BETTER.

ECOSYSTEM & PARTNER
Gemeinsam mit unseren Partnern können wir unseren Kunden einzigartige und höchst innovative Lösungen bieten.

DAS UNTERNEHMEN SAP
Mehr als 112.500 Kunden und 27 Jahre Erfahrung haben wir Ihnen zu einem einzigartigen Marktführer für Unternitmenssoftware geworden.
Globalization Tripod

**Internationalization**
- Allow any character to be entered and rendered correctly
- Ensure that collation/sorting works for any script/language

**Translation**
- Create proper terminology
- Find adequate expression for target language

**Localization**
- Adapt functionality to a locale
- Adapt non-translatable content

The following slides will often use the term globalization and not refer to internationalization/translation/localization.
Globalization Size, Impact, and Prospects

82% of online shops only in one language

2/3 of consumers prefer e-shop in own language

1.8 million pages translated

202 million words translated

$6.5 billion revenues for language services market

4500/$450 million employees/revenue for large Language Service Provider

1/3 goes to the translator
Globalization Vulnerability
Globalization Best Practices

Some links will be provided at the end of the presentation.

Use a well-supported source format

- XHTML, DocBook, Darwin Information Typing Architecture (DITA), Open Document Format (ODF), Office Open XML (OOXML), …

Consider using a framework (e.g. related to Cascading Stylesheets)

- Yet Another Multicolumn Layout (YAML), YUI Grids, Blueprint, …

Describe your resources

- Provide general annotations (e.g. batch) with standardized metadata

Internationalize

- Internationalization Quick Tips for the Web
- XML Internationalization Best Practices

Pseudo-Translate

- S_rêt_E

Get Terminology in Order

- Wiggle

Assure Linguistic Quality Automatically

- ret, yellow, blu

Some links will be provided at the end of the presentation.
Globalization Standards

Some links will be provided at the end of the presentation.

Assets

- Terminology – TermBase eXchange (TBX)
- Former Translations – Translation Memory eXchange (TMX)

Canonicalized Content

- XML Localization Interchange File Format (XLIFF)

Resource Description related to Internationalization

- Internationalization Tag Set (ITS)
OASIS XLIFF – Unify

Format 1 → Format 2 → Format 3 → Format 4 → Format ... → Format n → XLIFF
W3C ITS – Explain

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?
Virtues of Standard Formats
Insights

You are Wrong! 😞

You are Right! 😊
## Reality Check

<table>
<thead>
<tr>
<th>Category</th>
<th>Metrics</th>
</tr>
</thead>
</table>
| Standards/Formats | • Scope  
                    • Maturity |
| Implementations  | • #  
                    • Completeness  
                    • Quality |
| Interoperability | • 10% loss  
                    • 100% loss |
| Deployment       | • Here and there |
Bird’s Eye View on Accidents

**Creators of Standards**
- Pretension
- Missing Reuse/Orchestration,
- Misconception,
- Means to support conformance

**Solution Providers**
- Pretension
- Ambition

**Many**
- Disrepect for the virtues of Standards
The 5 M Safety System

Meetings
- Liaisons

Models
- Content
- Processes
- Actors

Metas
- Standardized Vocabularies
- Resource Description Framework
- Linked Data Principles

Modules
- Language Identification
- State
- Conformance Clauses
- Reference Implementations

Mashes
- Services
Example Virtue – Models and Modules
Example Virtue – Metas

1. Switch from a proprietary general encoding scheme to a generalized one such as RDF
2. Switch from proprietary encoding for data categories and values to standardized ones such as Dublin
### Example Virtue – Mashes

#### Services Framework

<table>
<thead>
<tr>
<th>Tool/Infrastructure/Service 1</th>
<th>Tool/Infrastructure/Service 2</th>
<th>…</th>
<th>Tool/Infrastructure/Service 1</th>
</tr>
</thead>
</table>

Learn More/Get Involved

Tutorial Standards-based Translation with W3C ITS and OASIS XLIFF

- [http://www.tekom.de/upload/2913/LOC12_Sasaki_Lieske.pdf](http://www.tekom.de/upload/2913/LOC12_Sasaki_Lieske.pdf)

Internationalization Quick Tips for the Web

- [http://www.w3.org/International/quicktips/Overview.en](http://www.w3.org/International/quicktips/Overview.en)

OASIS XLIFF 1.2 Specification (note: in addition representation guides exist)

- [http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html](http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html)

OASIS XLIFF Technical Committee


W3C ITS 1.0 Specification

- [http://www.w3.org/TR/its/](http://www.w3.org/TR/its/)

Best Practices for XML Internationalization

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

W3C ITS Interest Group

- [http://www.w3.org/International/its/ig/](http://www.w3.org/International/its/ig/)

Translation Memory/Term Base eXChange

- [http://www.lisa.org/Translation-Memory-e.34.0.html](http://www.lisa.org/Translation-Memory-e.34.0.html)
- [http://www.lisa.org/Term-Base-eXchange-TBX.32.0.html](http://www.lisa.org/Term-Base-eXchange-TBX.32.0.html)
<table>
<thead>
<tr>
<th>The Story that Has Been Told</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content that is available in several languages/adapted to several locales is an important ingredient of the Web</td>
</tr>
<tr>
<td>The creation of this kind of content can be challenging (e.g. due to the size and complexity of the corresponding processes)</td>
</tr>
<tr>
<td>Often, the processes involve core (e.g. terminology creation and translation) and context activities (e.g. billing)</td>
</tr>
<tr>
<td>Fortunately, there exist standards and Best Practices for mastering the challenges</td>
</tr>
<tr>
<td>The standards and Best Practices pertain to different entities related to the processes (e.g. the translatable content and assets – such as Translation Memories)</td>
</tr>
<tr>
<td>Not all standards and Best Practices are related to formats – some are related to processes (e.g. pseudo-translation, and automated linguistic quality checks)</td>
</tr>
<tr>
<td>Some standards are or can be combined (notably ITS and XML-based formats, as well as ITS and XLIFF)</td>
</tr>
<tr>
<td>Current gaps pertain to several areas (e.g. the use of ingredients to the semantic web - RDF, linked data principles)</td>
</tr>
<tr>
<td>The gaps should be bridged by surveying the overall processing needs, and then creating standards „modules” that cover one specific aspect (cf. for example BCP 47 or Dublin Core)</td>
</tr>
<tr>
<td>The standard development and implementation processes benefit from conformance clauses, test suites, and reference implementations/libraries</td>
</tr>
</tbody>
</table>

Thank you!
Contact

Christian Lieske

cristian.lieske@sap.com
www.sap.com
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.