Making the Multilingual Web work: with Open Standards and CMS

Bryan Schnabel
bschnabel@bschnabel.com
@bryanschnabel
A few things about me

- Content Management Architect at Tektronix (Trisoft, Drupal, . . . )
- Chair OASIS XLIFF TC (https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff)
- Creator of xliffRoundTrip tool (http://sourceforge.net/projects/xliffroundtrip/)
- Creator of DITA-XLIFF plugin for DITA OT (http://sourceforge.net/projects/ditaxliff/files/)
- (modest) Contributor to XLIFF Tools Drupal Module (http://drupal.org/project/xliff)
Our good fortune today . . .

The creator of the XLIFF Tools module, Gábor Hojtsy is here with us today.
Multilingual Web needs multilingual content
My use cases

• Web CMS: Drupal

• Component CMS (CCMS): Trisoft
Proprietary solutions vs. open standards-based solutions

• Drupal
  – Out of the box (or core) translation (http://drupal.org/documentation/modules/translation) vs.
  – Translation using the XLIFF Tools module (http://drupal.org/project/xliff)

• Trisoft
  – Out of the box (good, but not XLIFF) translation vs.
  – Translation using DITA-XLIFF Roundtrip plugin (http://sourceforge.net/projects/ditaxliff/files/)
Drupal: Translation Core Module

http://drupal.org/documentation/modules/translation
Translator logs in (assuming proper permissions and access)
Clicks the “Translate” tab
Picks the translation language

Translations of a piece of content are managed with translation sets. Each translation set has one source post and any number of translations in any of the enabled languages. All translations are tracked to be up to date or outdated based on whether the source post was modified significantly.

<table>
<thead>
<tr>
<th>Language</th>
<th>Title</th>
<th>Status</th>
<th>Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>English (source)</td>
<td>BitAlyzer BA</td>
<td>Published</td>
<td>edit</td>
</tr>
<tr>
<td>Czech</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>French</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>German</td>
<td>BitAlyzer BA</td>
<td>Published - outdated</td>
<td>edit</td>
</tr>
<tr>
<td>Hungarian</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Italian</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Korean</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Polish</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Portuguese, Brazil</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Portuguese, Portugal</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Russian</td>
<td>BitAlyzer BA</td>
<td>Published</td>
<td>edit</td>
</tr>
<tr>
<td>Spanish</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Swedish</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Turkish</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Chinese, Simplified</td>
<td>广谱仪 BitAlyzer</td>
<td>Published - outdated</td>
<td>edit</td>
</tr>
<tr>
<td>Chinese, Traditional</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
<tr>
<td>Japanese</td>
<td>n/a</td>
<td>Not translated</td>
<td>add translation</td>
</tr>
</tbody>
</table>

Select translations for BitAlyzer BA

Alternatively, you can select existing nodes as translations of this one or remove nodes from this translation set. Only nodes that have the right language and don't belong to other translation sets will be available here.
Translate subtitle, good
Translate details title, good

What is all of this? I’m confused

Ah, Unique Name must mean German name, I’ll translate this

BOOM! Just blew up every node of this content type!

Oh, and I’m accidentally adding German to the Japanese page
Pros and Cons of using Drupal out of the box translation

• Advantages:
  – Translation can be nimble

• Disadvantages:
  – Translator is not using Computer Aided Translation (CAT) tool
  – No clear way to leverage Translation Memory (TM)
  – Translator needs access to Drupal Admin interface
  – Translator needs Drupal training
  – Translator can easily do harm to web site
Translating Drupal with XLIFF

http://drupal.org/project/xliff

(my company’s last version http://drupal.org/node/881620#comment-7041716 )
Fundamental difference the XLIFF Tools module brings

• All Drupal steps are done by a Drupal administrator

• Translator does all his/her work in CAT tools
Drupal Admin picks the XLIFF module
Selects the node type to translate
Exports the nodes to XLIFF
Saves the XLIFF file locally
Post processing is usually necessary

<trans-unit id="field_warranty-13">
  <source>One-year Warranty</source>
  <target>One-year Warranty</target>
</trans-unit>

<trans-unit id="field_inventory_status-15">
  <source>Active</source>
  <target state="final">Active</target>
</trans-unit>

For example, translating "inventory status" would corrupt the Drupal database. We apply automated post processing that adds the ‘state="final"’ attribute to protect that field
XLIFF sent to LSP
LSP leverages TM and translates the XLIFF
Translate only desired text

<trans-unit id="field_warranty-13">
  <source>One-year Warranty</source>
  <target>1-Jahres-Garantie</target>
</trans-unit>

<trans-unit id="field_inventory_status-15">
  <source>Active</source>
  <target state="final">Active</target>
</trans-unit>
XLIFF sent back to content owner
In some cases, non-content text strings require additional processing

For example, in Drupal, menu strings are not typically processed in the Drupal XLIFF module

These strings are stored as “Portable Object” (PO) files
Drupal admin navigates to the Translate interface
Exports the menu PO file
Saves the PO file locally
This provides a non-XLIFF method to translate the menu text

Post processing can/should be done to transform the PO files to XLIFF files

Note: This is not part of the XLIFF Tools module; we custom-built this for our workflow
LSP translates Drupal PO XLIFF, then returns translated XLIFF
Transform translated XLIFF to PO

<trans-unit id="d_16">
  <source>Home</source>
  <target>Startseite</target>
</trans-unit>

<trans-unit id="d_20">
  <source>About Us</source>
  <target>Über uns</target>
</trans-unit>

<trans-unit id="d_24">
  <source>Service</source>
  <target>Service</target>
</trans-unit>

#: item:1195:title
msgid "Home"
msgstr "Startseite"

#: item:19232:title
msgid "About Us"
msgstr "Über uns"

#: item:19250:title
msgid "Service"
msgstr "Service"
Import the translated PO file
The site is translated
Pros and Cons of using open standards and the XLIFF tools module

• Advantages:
  – Translator uses CAT tool
  – TM is easily leveraged
  – No need for translator to access to Drupal Admin interface
  – No need for translator to have Drupal training
  – Mitigated risk to web site

• Disadvantages:
  – Translation can be less nimble
Component Content Management System (CCMS)

• Sometimes the content is owned by the Web CMS, but . . .

• Increasingly it is owned by the CCMS
DITA and CCMS

• The strength of a CCMS is its ability to manage rich topics
• The open standard for topic-base authoring is DITA
• With DITA, translating a “document” often means translating hundreds, thousands, millions of files
For example, our datasheets are managed in our CCMS.
They are made up of many topics (205 in this example)
Under the covers, topics, maps, images live in the repository
Translating Trisoft DITA using out of the box features
Pick the “publication”
Assign language(s)

[Image of software interface showing file properties]

- File name: MSO/DPO3000Datasheet
- In Folder: /General/Tektronix/Oscilloscopes/MSO_DPO3000/01-PubObjects
- Identifier: GUID-5ADOC875-7CBB-4ED8-87CB-ED58A426684A
- Publication type: Data Sheet
- Created by: kirkw
- Product type: Instrument
- Product category: Oscilloscopes
- Description: MSO3000-DPO3000-Mixed-Signal-Oscilloscope-Datasheet
- Translation management:
  - Requested languages: EN-US, ES-ES, FR-FR
- Date created: 10/23/2011 11:56:44 AM
- Date modified: 1/22/2013 12:58:11 PM
Initiate translation

Select Language for translation cycle

Add metadata

Click “OK”
Generate translation report
## Push translation

<table>
<thead>
<tr>
<th>CREATIONDATE</th>
<th>ACTION</th>
<th>STATUS</th>
<th>EVENT DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013/03/03</td>
<td>Request queued</td>
<td>SUCCESS</td>
<td>Request &quot;PUSHTRANSLATIONS&quot; will start asynchronously. [ISAuthor.CTranslationMgmt]</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>PushTranslations started</td>
<td>SUCCESS</td>
<td>Push translations started for 'GUID-03B582F6-2BC7-44A7-9C59-0FE7681E5BFF'. See auditdata for more details.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Request queued</td>
<td>SUCCESS</td>
<td>Request &quot;PUSHTRANSLATIONS&quot; will start asynchronously. [ISAuthor.CTranslationMgmt]</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Private translation management</td>
<td>SUCCESS</td>
<td>Start translation management for 'GUID-DB387DB3-A8A1-4F3A-BBF8-0B39EA4D82DD' in version '2' and language 'EN-US'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>Start translation management for 'GUID-DB387DB3-A8A1-4F3A-BBF8-0B39EA4D82DD' in version '2' and language 'EN-US'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>For GUID-DB387DB3-A8A1-4F3A-BBF8-0B39EA4D82DD the language cards from 'EN-US' to 'DE-DE' are created.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Private translation management finished [PID:5920]</td>
<td>SUCCESS</td>
<td>Finish translation management for 'GUID-DB387DB3-A8A1-4F3A-BBF8-0B39EA4D82DD'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Request queued</td>
<td>SUCCESS</td>
<td>Request &quot;PUSHTRANSLATIONS&quot; will start asynchronously. [ISAuthor.CTranslationMgmt]</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>Start translation management for 'GUID-509E5F41-51E9-40B3-8C94-75F120C50509D' in version '2' and language 'EN-US'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>For GUID-509E5F41-51E9-40B3-8C94-75F120C50509D the language cards from 'EN-US' to 'DE-DE' are created.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Private translation management finished [PID:5920]</td>
<td>SUCCESS</td>
<td>Finish translation management for 'GUID-509E5F41-619E-40B4-8C94-75F120C50509D'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Request queued</td>
<td>SUCCESS</td>
<td>Request &quot;PUSHTRANSLATIONS&quot; will start asynchronously. [ISAuthor.CTranslationMgmt]</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>Start translation management for 'GUID-1C525BA1-F50E-4B33-AB3C-D489A7F6DCA4' in version '2' and language 'EN-US'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>For GUID-1C525BA1-F50E-4B33-AB3C-D489A7F6DCA4 the language cards from 'EN-US' to 'DE-DE' are created.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Private translation management finished [PID:5920]</td>
<td>SUCCESS</td>
<td>Finish translation management for 'GUID-1C525BA1-F50E-4B33-AB3C-D489A7F6DCA4'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Request queued</td>
<td>SUCCESS</td>
<td>Request &quot;PUSHTRANSLATIONS&quot; will start asynchronously. [ISAuthor.CTranslationMgmt]</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>Start translation management for 'GUID-2AC4DA16-D23B-4669-8991-F10132A2A483' in version '2' and language 'EN-US'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Translation management for language</td>
<td>SUCCESS</td>
<td>For GUID-2AC4DA16-D23B-4669-8991-F10132A2A483 the language cards from 'EN-US' to 'DE-DE' are created.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>Private translation management finished [PID:5920]</td>
<td>SUCCESS</td>
<td>Finish translation management for 'GUID-2AC4DA16-D23B-4669-8991-F10132A2A483'.</td>
</tr>
<tr>
<td>2013/03/03</td>
<td>PushTranslations finished [PID:5920]</td>
<td>SUCCESS</td>
<td>Push translations finished</td>
</tr>
</tbody>
</table>
Next steps

• A zip file with all (100s, 1000s, 1000000s) DITA files arrives on the server.

• Send the zip file to the LSP.

• Make sure your translation agency does NOT modify the filenames nor the DITA structure. Only the content can be translated.

• When the translations come back, unzip them and put on the server.

• Import translated topics, maps, pictures into Trisoft.
100, 1000, 1000000 DITA files sent to LSP; translated; returned
Pros and Cons of using out of the box

• Advantages:
  – CCMS skillfully manages, tracks, audits, complex set of files and languages
  – CCMS *knows* how many topics per master map are translated to which languages

• Disadvantages:
  – LSP must manage 100s, 1000s, 1000000 files per transaction
  – LSP must know DITA
  – LSP manages the conversion of DITA to and from CAT tool
XLIFF and the DITA-XLIFF OT Plugin for CCMS

http://sourceforge.net/projects/ditaxliff/files/
Note about the DITA-XLIFF plugin and our integration with Trisoft

- I wrote the DITA-XLIFF plugin for the DITA Open Toolkit as an open source, standalone

- We adapted it to work with Trisoft, and augmented its integration with Visual Studio
We’ve added our “Trisoft Publishing Assistant” custom application

- Pick the publication part number
- Generate an XLIFF of all topics, maps and images
- Add a PFD file to the translation package

The writer essentially follows the same steps as out-of-the-box.

But now an XLIFF shows up instead of DITA files.
1 XLIFF file (not 100, 1000, 1000000 DITA files) sent; translated; returned
Pros and Cons of using XLIFF plugin

• Advantages:
  – LSP manages an XLIFF file, not 100, 1000, 1000000 DITA files
  – LSP does not need to know DITA
  – XLIFF opens natively in CAT tools (enabling nimble workflow and leveraging of TM)
  – CCMS (still) skillfully manages, tracks, audits, complex set of files and languages
  – CCMS (still) *knows* how many topics per master map are translated to which languages

• Disadvantages:
  – Complexity of plumbing the plugin to CCMS is nontrivial
Thank you

Bryan Schnabel

bschnabel@bschnabel.com

@bryanschnabel