

//XUGS22[@date='20022019']/*

speaker: Stefan Fisahn / Informatica

topic: XLIFF (Helicopter View)

location: HdM Stuttgart



XML Localization Interchange File Format (OASIS-Standard)

Agenda

- XLIFF ?
- XLIFF-Versionen
- Andere Formate ...
- Ein erster Blick ...
- Workflows
- Ein zweiter Blick: `//xliff/file/unit/segment/source`
- Play around
- Resümierende Notizen
- Weiterführende Ressourcen

XLIFF?

XML Localization Interchange File Format (OASIS-Standard)

- Datenaustausch für Übersetzungen / Lokalisierungen (Unterschied)
- Über den Austausch hinaus: Workflows
- Extract und Merge statt nur Transformation
- Erweiterbar
- Lob des Standards und: XML .. es ist XML!

XLIFF-Versionen

- Aktuell: XLIFF Version 2.1 (2/2018) ... Abwärtskompatibel zu 2.0
 - In Verwendung: XLIFF Version 2.0 (8/2014) ... Nicht kompatibel zu 1.2
 - Legacy - aber auch noch in Verwendung (2/2008) XLIFF Version 1.2
- [XLIFF 2.0 vs. XLIFF 1.2](#)

Andere Formate

- TMX
 - Austauschformat von Sprachpaaren (XML)
 - Keine Hierarchie-Info
 - Kein Mechanismus für die Rekonstruktion des Original Files
 - Multilingual (XLIFF = Bilingual)
 - Wird parallel zu XLIFF verwendet um TMS Daten zu transportieren.

```
<tmx version="1.4">
  <header
    creationtool="XYZTool" creationtoolversion="1.01-023"
    datatype="PlainText" segtype="sentence"
    adminlang="en-us" srclang="en"
    o-tmf="ABCTransMem"/>
  <body>
    <tu>
      <tuv xml:lang="en">
        <seg>Hello world!</seg>
      </tuv>
      <tuv xml:lang="fr">
        <seg>Bonjour tout le monde!</seg>
      </tuv>
    </tu>
  </body>
</tmx>
```

Andere Formate

- PO

Übersetzungsdateien für Poedit (GNU gettext System). Kontext: Programm-Lokalisierung - kann direkt als ProgrammLibrary eingebunden werden (Stichwort: Developertool)

The screenshot displays the Poedit PO editor interface. The main window is divided into three sections: a translation table, a source text field, and a target text field. The translation table has three columns: 'Quelltext', 'Übersetzung – Deutsch (Deutschland)', and 'ID'. The row for '&Preferences...' is highlighted in blue. Below the table, the 'Quelltext:' field contains '&Preferences...' and the 'Übersetzung:' field contains '&Einstellungen ...'. On the right side, a sidebar titled 'Übersetzungsvorschläge:' (Translation suggestions) lists several suggestions with their respective keyboard shortcuts: '&Einstellungen ...' (Strg+1 + 100X), 'Voreinstellungen' (Strg+2 + 96X), 'Den ...Einstellungen entsprechend' (Strg+3 + 96X), '...Einstellungen ...' (Strg+4 + 95X), and 'Einstellungen ...' (Strg+5 + 92X).

Quelltext	Übersetzung – Deutsch (Deutschland)	ID
Poedit uses a large online translation me...	Poedit verwendet einen großen Online-Über...	322
POT Translation Templates	POT-Übersetzungsvorlagen	141
powered by	unterstützt von	327
&Preferences	&Einstellungen	415
&Preferences...	&Einstellungen ...	383
Preparing migration...	Migration wird vorbereitet ...	532
Preserve formatting of existing files	Formatierung vorhandener Dateien beibehalten	351
Previous source text:	Bisheriger Quelltext:	500
&Previous translation	&Vorherige Übersetzung	444
&Previous Translation	&Vorherige Übersetzung	445
P&revious unfinished	V&orherige unfertige	448
P&revious Unfinished	V&orherige unfertige	449

Übersetzungsvorschläge:

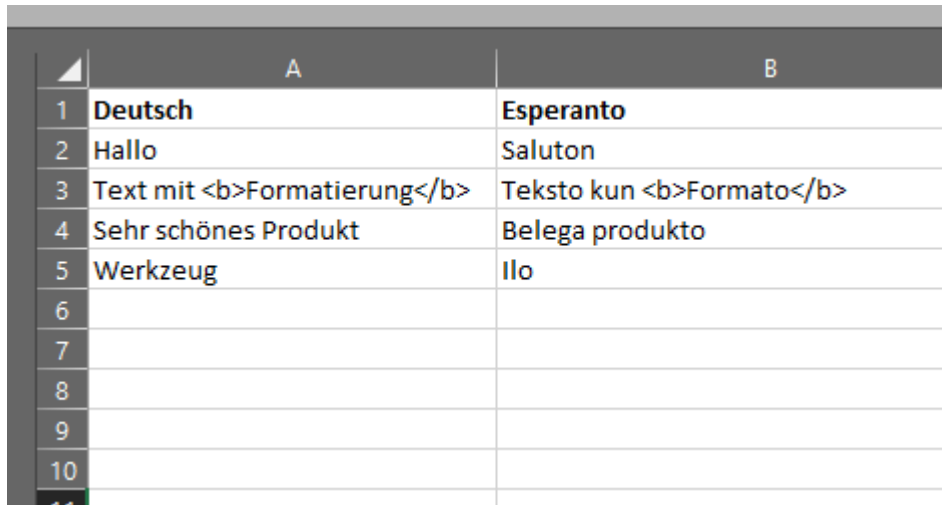
- &Einstellungen ...
Strg+1 + 100X
- Voreinstellungen
Strg+2 + 96X
- Den ...Einstellungen entsprechend
Strg+3 + 96X
- ...Einstellungen ...
Strg+4 + 95X
- Einstellungen ...
Strg+5 + 92X

Quelltext:
&Preferences...

Übersetzung:
&Einstellungen ...

Andere Formate

- Excel, ernsthaft jetzt?



	A	B
1	Deutsch	Esperanto
2	Hallo	Saluton
3	Text mit Formatierung	Teksto kun Formato
4	Sehr schönes Produkt	Belega produkto
5	Werkzeug	Ilo
6		
7		
8		
9		
10		
11		

... und es gibt mehr, es geht um unterschiedliche Aufgaben ...

Ein erster Blick

The screenshot displays an XML editor window with the file `*xliiff2.xml` open. The XML content is as follows:

```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <!DOCTYPE xml>
3 <xliiff xmlns="urn:oasis:names:tc:xliiff:document:2.0"
4     version="2.0"
5     srcLang="de"
6     trgLang="eo"
7     xsi=" https://docs.oasis-open.org/xliiff/xliiff-core/v2.0/os/schemas/xliiff_core_2.0.xsd">
8
9     <file id="f1">
10
11     <skeleton href="myFile.skl"/>
12
13     <unit id="u1">
14     <segment state="initial">
15     <source>Dieser Text ist noch nicht übersetzt</source>
16     </segment>
17     </unit>
18
19     <unit id="u2">
20     <segment state="translated">
21     <source>Dieser Text ist schon übersetzt.</source>
22     <target>Ĉi tio teksto jam estas tradukita.</target>
23     </segment>
24
25     </unit>
26 </file>
27 </xliiff>
```

On the right side, the `XPath` panel shows the expression `string(//unit[@id="u2"]/segment/target)` and the resulting location `Ĉi tio teksto jam estas tradukita.` Below the XPath panel, there are tabs for `Sheet 1` through `Sheet 5`, and a `Terminal` tab. The `Outline` panel shows a tree view of the XML structure:

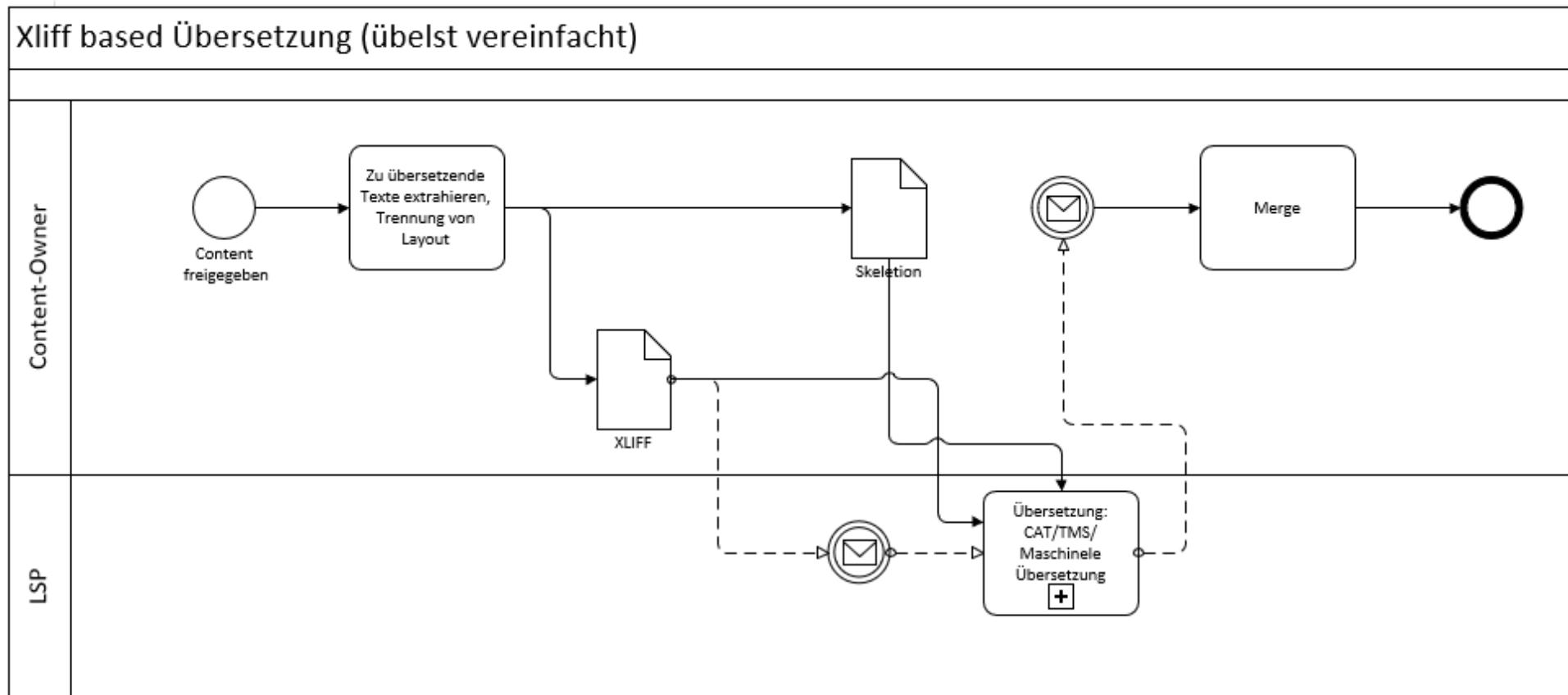
- xml
- DOCTYPE:xml
- xliiff xmlns=urn:oasis:names:tc:xliiff:dc
 - file id=f1
 - skeleton href=myFile.skl
 - unit id=u1
 - segment state=initial
 - source
 - unit id=u2
 - segment state=translated
 - source
 - target

Ein erster Blick .. ins Schema

<http://docs.oasis-open.org/xliff/xliff-core/v2.1/os/xliff-core-v2.1-os.html>

```
<xliff>
|
+---<file> +
|
+---<skeleton> ?
| |
| +---<other> *
|
+---<other> *
|
+---<notes> ?
| |
| +---<note> +
|
+---At least one of (<unit> OR <group>)
| |
| +---<unit>
| |
| | +---<other> *
| |
| | +---<notes> ?
| | |
| | | +---<note> +
| |
| | +---<originalData> ?
| | |
| | | +---<data> +
| |
| | +---At least one of (<segment> OR <ignorable>)
| | |
| | | +---<segment>
| | | |
| | | | +---<source> 1
| | | | |
| | | | | +---<target> ?
| | |
| | | +---<ignorable>
| | | |
| | | | +---<source> 1
| | | | |
| | | | | +---<target> ?
| |
|
+---<group>
|
+---<other> *
|
+---<notes> ?
| |
| +---<note> +
|
+---At least one of (<unit> OR <group>)
```

Workflows



Ein zweiter Blick - von innen nach außen ...

segment, source, target

<http://docs.oasis-open.org/xliff/xliff-core/v2.1/os/xliff-core-v2.1-os.html#segmentation>

```
e xliff xmlns=urn:oasis:names:tc:xliff:
v e file id=f1
  e skeleton href=myFile.skf
  > e unit id=u1
  v e unit id=u2
    e segment state=translated
      e source
      e target
```

Übersetzen: yes|no: file, group, source, target, mrk

Example 10.6 – Scope of the translate attribute

```
<file translate="no" id="f1">
  <group id="g1" translate="yes">
    <unit translate="no" id="u1">
      <segment id="s1">
        <source><mrk translate="yes" type="term" id="m1">Crawl</mrk></source>
      </segment>
    </unit>
  </group>
  <group id="g2">
    <unit id="u2">
      <segment id="s2">
        <source><mrk id="m2" type="term">Walk</mrk></source>
      </segment>
    </unit>
  </group>
  <group id="g3">
    <unit id="u3">
      <segment id="s3">
        <source><mrk translate="yes" type="term" id="m3">Run</mrk></source>
      </segment>
    </unit>
  </group>
  <group id="g4">
    <unit translate="yes" id="u4">
      <segment id="s4">
        <source><mrk id="m4" type="term">Fly</mrk></source>
      </segment>
    </unit>
  </group>
</file>
```

Quelle: A practical Guide to XLIFF 2.0

//xliff22[@topic='XLIFF']

Ein zweiter Blick...

unit und group

<http://docs.oasis-open.org/xliff/xliff-core/v2.1/os/xliff-core-v2.1-os.html#group>

- group = Strukturinformation für Units (Optional)
- unit = Repräsentiert einen zu übersetzenden "Abschnitt" und enthält mind. 1 Segment

Ein zweiter Blick...

skeleton und file

<http://docs.oasis-open.org/xliff/xliff-core/v2.1/os/xliff-core-v2.1-os.html#file>

<http://docs.oasis-open.org/xliff/xliff-core/v2.1/os/xliff-core-v2.1-os.html#skeleton>

- file = Container pro Quelldokument; wirklich die "größte Einheit"..
Subelemente (Kaptiel usw.) -> Group
- skeleton ...

Ein zweiter Blick...

skeleton?

Container für nicht zu übersetzende Einträge im Kontext zum parent-Element /file.

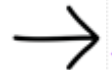
Enthält entweder

- Nicht zu übersetzende Inhalte
- Elemente anderer Namespaces

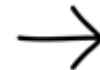
Ein zweiter Blick...

skeleton?

```
<html>
  <head>
    <title>Ein Titel</title>
  </head>
  <body>
    <p>Ein Abschnitt</p>
  </body>
</html>
```



```
<html>
  <head>
    <title>skl_id1</title>
  </head>
  <body>
    <p>sklid2</p>
  </body>
</html>
```



```
<? xml version="1.0" ?>
<xliff version="2.1" srcLang="de" trgLang="eo">
  <file id="beispiel.html">
    <skeleton href="beispiel.skl"/>

    <unit id="skl_id_1">
      <source>Ein Titel</source>
    </unit>
    <unit id="skl_id_2">
      <source>Ein Abschnitt</source>
    </unit>
  </file>
</xliff>
```

html → skeleton → XLIFF

Anwendung: html, DITA, DocBook, SVG, ...

Play around



- Produktdaten in Informatica MDM Product 360 (PIM) und ergänzende Daten in CSV-Datei
- BaseX: Per REST Call Daten aus Informatica MDM Product 360 (PIM) holen, JOIN mit CSV-Daten
- Ausgabe per Xquery als XLIFF.

- Demo: 

Resümierende Notizen

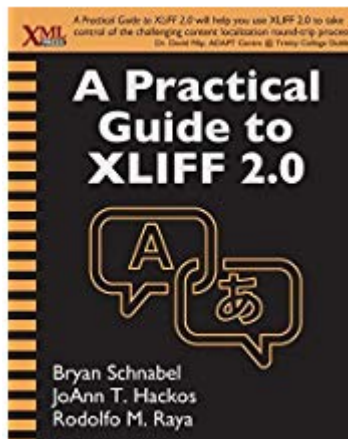
- Extract & Merge ... nicht nur Transformieren & Transformieren
- Um einfach nur Sprachpaare zu übermitteln, vielleicht ... overdone
- Direkt Anbindung über REST-API kann auch interessant sein
- Und Json? #ToCheck: JLIFF [XMLprague 2019 proceedings p.295](#)
- XLIFF zusammen mit BPM rocks!
- Achja: XLIFF ist erweiterbar. (Namespaces)

Notiz die es fast nicht in die Folie geschafft hätte

- xsd not valid ?? ... sagt Eclipse-Validierung

Weiterführende Ressourcen

- Yves Savourel: An introduction to XLIFF 2.0 - Multilingual Magazine
- Jesús Torres: XLIFF and the Translator: Why Does it Matter?
- XLIFF 2.0 vs. XLIFF 1.2
- OASIS v2.1 Specs
- integrated Metadata and Custom Metadata (Extensions)
- Ein Buch, ein Buch: Bryan Schnabel: A Practical Guide to XLIFF 2.0



fin.