

# Workaround for misused XLIFF



 $\rightarrow$ 

zaac

1



### XLIFF XML Localization Interchange File Format

### • What it is:

- Format for translation
  - Bilingual setup
- Original file format gets converted to XLIFF
  - Segmentation
- What it is NOT:
  - It is NOT supposed to contain other file formats, like HTML (as you can do with XML itself).
  - It is NOT format where you can customize attributes to contain user-defined comments, IDs or length information.
  - It is NOT format that gets segmented when importing it into a CAT tool.
    - The segmentation is supposed to happen when creating the XLIFF.



### XLIFF XML Localization Interchange File Format

- A lot of developers who create XLIFF don't seem to be aware of what XLIFF is supposed to be.
- We have to deal with "misused" XLIFF.
- Most of the time, you can use an XML filter instead <sup>(C)</sup>.
  - Copy source to target
  - Lock the content between the source tags.
  - Use the content between the target tags for translation.



## **1. XLIFF with HTML**

<?xml version="1.0" encoding="utf-8" standalone="no"?> <xliff xmlns="urn:oasis:names:tc:xliff:document:1.2" version="1.2">

### <source>

. . .

• • •

5

<target></target>

</trans-unit></body></file></xliff>



### **XLIFF Filter**

Source

Target

This is some text for a website. The text is written in HTML and therefore contains. several HTML-specific tags. For example <strong>formatting<strong>..In.order.to. embed the text in XML the user has chosen. the CDATA element (which signifies that everything that follows is pure text. Tags are not:to:be:interpreted:as:such).'It:also: contains several sentences within one source segment:as:the:user:has:not:done:any: segmentation while producing the XLIFF. file.<br/>Linebreaks<br/>are<br/>created<br/> through Prand BR tags. And the user has also used an entity, here the non-breaking space.



## Importing as XML (+ HTML)

• Use an XML filter where you lock the content of the < source > ... < / source > sections, effectively importing the text between the <target>...</target> tags as source text. • Add a cascading HTML filter and define whether a <br/> tag should beak a segment or not. • This text will get overwritten with your translation.



Document import settings							×
Filter	XML filter						~
Filter configuration	<new con<="" td=""><td>figuration&gt;</td><td></td><td></td><td></td><td>~</td><td>2</td></new>	figuration>				~	2
Add cascading filter	Remove	Remove					
Encoding and reference files	s General	General Tags and attributes Entities Subtitles					
Handled tags			Та	ig settin	gs		
Name	Info			Inline	_		
body	Str;Inh		$\checkmark$	Not tran	slated		
file	Str;Inh		R	equired:	Nor	mal	$\sim$
houder	Str;Inh		Whitespace handling:				
source	Str;NT;Inh		In	herit.			$\sim$
torget	Str;Inh		in ment			4.6	
trans-unit	Str;Inh			l ag con	itent is con	text ID for	siblings
xliff	Str;Inh			_			

### XML filter + HTML filter

🖳 Select filter to cha	in	×
Filter	HTML filter	~
Filter configuration	default (read-only)	~
	OK Cance	l Help
Document import setting	s	×
Filter	Cascading filter	~
Filter configuration	<new configuration=""></new>	~ 📃 🛃
Add cascading filter XML filter > HTML fi Select import type	L Remove Iter	
Import markup as inlin	e tags	
O Import markup as men	noQ {tags}	
Import settings		
Ignore content of dr	op-down lists	
✓ Import non-breaking	spaces as entities	
Enforce empty <img.< td=""><td>/&gt; and   tags</td><td></td></img.<>	/> and tags	
Treat tags as	inline	
Import access keys		
Import HTML entities	s as characters	

### XML filter + HTML filter

ource	Target		🔍 📮 👎 🍭 s	ort No sorting	
This is some text for a website.					0%
The text is written in HTML and there several HTML specific tags.	efore contains.				0%
For example strong formatting strong					0%
In order to embed the text in XNL the chosen the CDATA element (which si everything that follows is pure text.	ne·user·has· gnifies·that·				0%
Tags are not to be interpreted as su	ch).				0%
It also contains several sentences wi segment as the user has not done a while producing the XLIFF file.	thin one source ny segmentation				0%
Linebreaks are created through P an	d·BR·tags.	Use	ed for segmentation	on	0%
And the use has also used an entity non-breaking here space.	, here the				0%
				Ch	nanged never
/iew pane	>			□ <del>7</del> ×	
<pre><source/><ucdata[this a="" for="" is="" some="" text="" w<br=""><strong>formattingtags are xLIFF file   inebreaks are created as such). It als xLIFF file   inebreaks are created as such). It als xLIFF file   inebreaks are created as such as such as a such as such a</br></br></br></br></br></strong></ucdata[this></pre>	ebsite. The text is written in mbed the text in XML the us to contains several sentence ated through P and BR tags or a website. The text is i>. In order to embed t ags are not to be interp y segmentation while put the non-breaking space	n HTML and therefore con ser has chosen the CDATA es within one source segm . And the user has also us s written in HTML and the text in XML the us preted as such). It als roducing the XLIFF fil ce.]]>	tains several HTML-specific tags. element (which signifies that even nent as the user has not done any sed an entity, here the non-break d therefore contains several ser has chosen the CDATA el so contains several sentence le.Linebreaks are created the	For example arything that follows is pur- segmentation while prod in   pace.]]>HIML-specific tags. F lement (which signifie as within one source rough P and BR tags.	re text. lucing the urce> For es that And

</trans-unit>



### 2. Length Restriction in Attribute

• Use the XML filter to specify the content of an attribute as a comment, for example a number as length restriction.



Length Restriction in Attribute



### Length Restriction in Attribute

ocument import settings	;						×	i
ilter	XML filter						~	
ilter configuration	<new confi<="" td=""><td>guration&gt;</td><td></td><td></td><td></td><td>~</td><td></td><td></td></new>	guration>				~		
Add cascading filter	Remove							
Encoding and reference fil	es General	Tags and	attribut	tes Entities	Subtit	es		
Handled tags				Tag settin	gs			
Name	Info			Inline				
body	Str;Inh			Not tran	slated			
file	Str;Inh			Required:		Normal	$\sim$	
group	Str;Inh			Whitespace	e handli	na:	_	
source	Str;NT;Inh			Inherit		.9.	~	-
target	Str;Inh						4.6	
trans-unit	Str;Inh			lag cor	ntent is o	context ID for	siblings	
XIIIT	Str;Inh			Tag cor	ntent is (	comment for s	iblings	
				Siblings as:	Only	following	$\sim$	
			t 🛏	Clear li	ist	Popula	ate	
Tag attributes				Attribute s	ettings	;		
Name	Info			Translat	able			
id				Require	d			
maxwidth				Filtered				
size-uniit								
translate					Non-trar	nslation		
		-	t 🛏	Context	t	Commer	nt	



Comment settings for attribute

×

hoose how memoQ will import comments from the value of the selected attribute.

### Not a comment source

memoQ will not import the attribute's value as comment.

### Comment for tag's children

The attribute's value will be imported as a comment for the tag's children.

### Comment for tag's children and following siblings

The attribute's value will be imported as a comment for the tag's children and the tags that follow, until the parent tag is closed.

### Comment for tag's children and preceding and following siblings

The attribute's value will be imported as a comment for the tag's children and for tags that follow afterwards until the parent tag is closed and for preceding tags from the point the parent tag was

OK	Cancel
----	--------

12



## Number in Comment for Length Check

Sourc	e	Target	] <b>Ş</b>	\$ *	0	Sort	No sorting			```	~
1.	Main Switch 1								0%	×	Î
2.	MS1								0%	<b>F</b>	
з.	Aaster Control 2								0%	×	
4.	4C·2								0%	×	~
									Change	1 never	
View	pane							0 <del>4</del> ×			
Activ	e comments					HTM	L preview	Review	Active co	mment	5
Crea Appl 18	ted by XML filter Date/time 06.06.2023 14:34 Category ( ies to Row	Information								Û	

Edit QA settings									
Segments and terms	Consistency	Numbers	Punctuation	Spaces, capitals, characters	Inline tags	Length	Regex	Severity	
Character-based le	ngth check								
Check length of	Check length of translation based on value stored with the row or in its comment								



### 3. Normalization

- If the XLIFF file contains line breaks or indents, make sure to set up the XML filter to preserver whitespace!
- Otherwise the XML filter will "normalize" all whitespace within the text (i.e. reduce several whitespaces to just one).

### Normalization

### Source text with line breaks and indents

memoqfest



Document import setti	ngs	
Filter	XML filter	
Filter configuration	<new configuration=""></new>	~

Х

 $\sim$ 

.



### Document import settings Х XML filter Filter $\sim$ ~ 🖃 . <New Configuration> Filter configuration Add cascading filter... Remove Encoding and reference files General Tags and attributes Entities Subtitles Associated DTD or namespace DTD or namespace URL Root tag name Content XML comment handling Break segment $\sim$ Translate XML comments Detect encoding if possible Input encoding if not detected Unicode (UTF-8) $\sim$ Output encoding Same as original $\sim$ Write BOM to Unicode-encoded files on export Normalize whitespace by default Observe xml:space attribute in file Change xml:lang attribute value at export Break segments at newlines if whitespace is preserved Import processing instructions as inline tags

Add cascading filter... | Remove

Encoding and reference files	General	Tags and attributes	Entities	Subtitles
------------------------------	---------	---------------------	----------	-----------

Handled tags		Tag settings
Name	Info	
body	Str;Inh	Not translated
file	Str;Inh	Required: Normal 🗸
header source	Str;Inh Str;NT;Inh	Whitespace handling:
target trans-onit xliff	Str;Pres Str;Inh Str;Inh	Tag content is context ID for siblings
		Tag content is comment for siblings

### XML filter + HTML filter

### Normalization

### 2nd step: HTML filter

17

So	urce Target Target Sort No sorting		~
1.	Dear {salutation}	0%	×
2.	Thank you for your interest in our newsletter.	0%	×
3.	Please enter your email address by clicking on the following link: 3rd step: Regex Tagger	0%	×
4	a href='{uri}' Complete registration a	0%	×
	If the link does not work, please copy/paste	0%	×
5.	the link address into the address field of your browser.		
6.	Registrierung abgeschlossen	0%	×
			~
		hanged n	ever
Vi	ew pane		
	<source/> <![CDATA[</td> <td>1</td> <td>^</td>	1	^
	Dear {salutation},		
	Thank you for your interest in our newsletter. Please enter your email address by clicking on the following link:		
	<a href="{uri}">Complete registration</a>		
	If the link does not work, please copy/paste the link address into the address field of your browser.		
	<pre><target state="translated">&lt;![CDATA[ Source and target text with same line&lt;/pre&gt;</target></pre>		
	Dear {salutation}, breaks and indents.		
	Thank you for your interest in our newsletter. Please enter your email address by clicking on the following link:		

<a href="{uri}">Complete registration</a>

If the link does not work, please copy/paste the link address into the address field of your browser. ]]></target>

</trans-unit>



### 4. Source Text in Target Section

# The <source>...</source> tags do not contain the source text.

<trans-unit id="1"> <source xml:lang="en-US"> <![CDATA[company.articleOverview.all]]> </source>

<target xml:lang="de-DE"><![CDATA[All Articles]]></target></trans-unit>

Use an XML filter as previously described.



### 5. Translation in the Note

# The client wants to have the translation in a different area.

<trans-unit id="1">

<source xml:lang="en-US">This is the source segment.</source>

<note>TRANSLATE TO HERE</note>

</trans-unit>

Use a **multilingual XML** filter to define the source text area and the target text area.



## 5. Translation in the Note

### × XML import rules XML file (click on row for its XPath expression) XPath for selected row <xliff version="1.2" xmlns1="urn:oasis:names:tc:xliff:document:1.2"> <file original="demo.txt" source-language="en-US" note-language="de-DE" datatype="plaintext" date="2"</p> Content rule ⊿ <body> Fill from selected row Content XPath ⊿ <group id="I/O-Texts"> ✓ <trans-unit id="001"> //note source xml:space="preserve"> memoQ language German (Germany) Source segment one. Context (ID) XPath for content Fill from selected row xml:space="preserve"> TRANSLATE TO HERE </note> </trans-unit> Fill from selected row Length restriction XPath for content <trans-unit id="002"> strans-unit id="003"> b <trans-unit id="004"> Comment XPath for content Fill from selected row </group> </body> </file> memoQ comment type Information </xliff: Save to rule set

Import rule set

eng-US://source;;;;0; ger-DE://note;;;;0;

Please note that this filter does not allow a cascading HTML filter (at the moment).



### 5. Translation in the Note

Sou	urce		Target	♀ ♦
1.	Source se	gment <sup>.</sup> one.		TRANSLATE·TO·HERE
2.	Source'se	gment <sup>.</sup> two.		TRANSLATE TO HERE
3.	Source'se	gment <sup>.</sup> three.		TRANSLATE TO HERE
4.	Source se	gment four.		TRANSLATE TO HERE
Vie	ew pane			
Sc ge	ource er-DE	Source segment one. TRANSLATE TO HERE		
		1		
Sc	ource	Source segment two.		
ge	er-DE	TRANSLATE TO HERE		
ID	1			



## 6. Non-HTML Tags

22

# The text contains custom tags to signify a line break.

😑 color info .xlf	
13	<pre>Source&gt;The names of the following buttons are customizable:</pre>
14	<pre>ctarget&gt;The names of the following buttons are customizable:</pre>
15 -	<pre>x/trans-unit&gt;</pre>
16 🛱	<pre>%trans-unit xml:space="preserve" id="chapter:2_1_0_0_6-page:0-object:Interaktion.0_0-type:inter_tit4.tit"</pre>
17	<pre>Source&gt;Button small</pre>
18	<pre>ctarget&gt;Button small</pre>
19 -	<pre>x/trans-unit&gt;</pre>
20 白	<pre>(trans-unit xml:space="preserve" id="chapter:2_1_0_0_6-page:0-object:Interaktion11_0-type:inter_tit4.tit")</pre>
21	<pre>Source&gt;Button long</pre>
22	<pre>%target&gt;Button long</pre>
23 -	<pre>x/trans-unit&gt;</pre>
24 🛱	<pre>% (trans-unit xml:space="preserve" id="chapter:2_1_0_0_6-page:0-object:Interaktion12_0-type:inter_tit4.tit")</pre>
25	<pre>(source&gt;Button large/multi-line</pre>
26	<pre>ctarget&gt;Button large/multi-line <x ctype="lb" id="lb1"></x></pre>
27 -	<pre></pre>
28	<pre>(trans-unit xml:space="preserve" id="chapter:2_1_0_0_0_6-page:1-object:Text1_0-type:text"&gt;</pre>
29	<pre>Source&gt;Edit the button properties via a right-click. &lt;x id='lb1' ctype='lb'/&gt;</pre>
	Type the new name and save it.
30	(target>Edit the button properties via a right-click. <x ctype="lb" id="lbl"></x>
	Type the new name and save it.
31 -	<pre>(/trans-unit&gt;</pre>



### Non-HTML Tags

Tags show up as text when imported with XLIFF or XML filter.

	Project home 💿 color info.xlf ×
Sour	ce Target
1.	Template:
2.	Buttons
3.	The names of the following buttons are customizable:
4.	Button·small
5.	Button·long
6.	Button large/multi-line
7.	Edit:the:button:properties:via:a:right-click. <x:id='lb1':ctype='lb'></x:id='lb1':ctype='lb'> Type:the:new:name: and:save:it.
8.	Template:



## Non-HTML Tags

Document import settings X				
Filter Filter configuration Add cascading filter   XML filter > XML filter	Cascading filter  V (New Configuration> V E Remove			
Encoding and reference files Handled tags Name body file beader source target trans-unit xliff	General       Tags and attributes       Entities       Subtitles         Info       Inline       Inline         Str;Inh       ✓       Net translated         Str;Inh       Required:       Normal         Str;Inh       Whitespace handling:         Str;Inh       Inherit         Str;Inh       Tag content is context ID for siblings         Str;Inh       Tag content is comment for siblings			

Use 2 XML filter steps
to convert the x tag text to a "real" tag
to use the x tag as a structure tag

Document import settings					×	
Filter     Cascading filter       Filter configuration <new configuration=""></new>						
Add cascating filter   Remove XML filter > XML filter  Encoding and reference files   General   Tags and attributes   Entition   Subtition						
Handled tags	Info Str;Inh	T C	ag settin	gs slated		
		F	Required:	Normal	$\sim$	

## Non-HTML Tags

x tag used for segmentation

	color info.xit ×	
Sou	rce Target	📃 🔤 🔍 🗘 📌 📌
1.	Template:	
2.	Buttons	
3.	The names of the following buttons customizable:	is are
4.	Button small	
5.	Button·long	
6.	Button·large/multi-line	
7.	Edit the button properties via a right	Jht-click.
8.	Type the new name and save it.	
9.	Template:	
10.	Panel 1 base colors	
11.	Blue, Red and Black.	
12.	In addition, 3 shades of gray are a	available.
Vie	<pre>//trans-unit&gt;</pre>	
	ins-unit xml:space="preserve" id="chapter:	r:2_1_0_0NU_6-page:1-object:Text1_0-type:text">
	<pre><source/>Edit the button properties via a rig it.</pre>	'ight-click. <x <="" ctype="lb" id="lb1" th=""></x>
	<target>Edit the button properties via a rig </target>	ight-click.Type the new name and save it.

0

. . . . ..

**Project home** 



## 6. G-Tags for Tags in the Original File

XLIFF has a tag called <g> which can be used to act as a placeholder for tags in the source file format.

### **XLIFF filter**

large segmentsa lot of tags

### Here the <g> tag could be defined as a structure tag.

But here the <g> tag acts as an inline tag as well. Source

g g g Title<sup>o</sup> g g g g g g Use of this document **a a a b** This document is intended for the end user of the product.° g g g g It contains information on all editions of the product. therefore recommend that you print out the section that deals with the product edition that you have purchased. Generation Contains all the information you need for your product does-not-include-a-certain-feature, thisfeature will not be described. If you have any questions, please contact<sup>·</sup>us<sup>·</sup>through<sup>·</sup>our<sup>·</sup>online<sup>·</sup>support<sup>·</sup> platform. g g g

Target



## XML filter g tag is inline tag

Segmentation is slightly better, segments are no longer so big.

Source

Target

### Title of the document

- g g g Title<sup>o</sup> g g g g g g g Use of this
  <sup>2.</sup> document g g g g g This document is intended for the end user of the product.
- g g g g It contains information on all editions of the product. g g g g g g We
   therefore recommend that you print out the section that deals with the product edition that you have purchased. g g

### Cascading filter Regex Textfilter 1 Regex Textfilter 2 Regex Textfilter 3 XML

Segmentation is even better and the amount of tags has been reduced in segments where tags are not used as internal tags. g

It contains information on all editions of the product.

5.

6.

We therefore recommend that you print out
 the section that deals with the product edition that you have purchased.

Each section contains all the information you need for your product edition.

If your product edition does not include a

9. certain feature, this feature will not be described.

10. g

<sup>11.</sup> If you have any questions, please contact us through our online support platform.

Our products combine style , 

design , 
and and Innovation

into one highly creative package.

<sup>13.</sup> See for yourself!

## Cascading filter • memoqfest

General	Paragraph	Include/exclude	Preview	

Rules define content to be excluded (external tags)

Rules define imported content (nothing else is imported)

#	Before	Rule	After
1	<target></target>	.+?	

Regex Textfilter 1 to extract text between <target> tags

Paragraph separator	C Empty line
Line with whitespace only	
<ul> <li>Custom regex</li> <li>Paragraph end</li> </ul>	()+
Paragraph start	( <g.+?>)+</g.+?>

Regex Textfilter 2 to use several g tags as segmentation point

ragraph Include/exclude	Preview						
<ul> <li>Rules define content to be excluded (external tags)</li> </ul>							
<ul> <li>Rules define imported content (nothing else is imported)</li> </ul>							
fore	Rule	After					
	<g.*?>(<g.*?>)+</g.*?></g.*?>	•					
	()+						
	ragraph Include/exclude fine content to be exclude fine imported content (not) fore	ragraph     Include/exclude     Preview       efine content to be excluded (external tags)       efine imported content (nothing else is imported content (nothing el					

Encoding and reference files	General	Tags and attribute	s Entities	Subti
Handled tags			Tag settin	gs
Name	Info		🗸 Inline	
g	ln;lnh		Not tran	slated
			Required:	

Regex Textfilter 3 to exclude g tag sequences

> XML filter to show remaining g tags as internal tags



### More...

- XLIFF files that show the \*.XML extension instead of \*.XLIFF
  - Rename the file extension
- XLIFF files that do not import, but show an error message that the language pair of the XLIFF is not the same as the project's language settings
  - Check the XLIFF in a text editor and search for "sourcelanguage=" and "target-language=" and type in the correct language codes
  - Sometimes XLIFF files do not show language codes, but the information "default"



### More...

• XLIFF files that show the \*.XLIFF extension, but once opened with a text editor you can see that they are not XLIFF files at all, because the XLIFF header is missing.





# Thank you!

Any questions?

zerfass@zaac.de