

FEISGILTT 2015 Public Info sessions

Draft meeting minutes

Wed June 3, 2015

Participants:

[participation in particular sessions was not tracked, this list is based on FEISGILTT 2015 participation]

XLIFF TC members

David Filip (ADAPT Centre), Ryan King (Microsoft), Kevin O'Donnell (Microsoft), Felix Sasaki (Individual), Yves Savourel (ENLASO), Bryan Schnabel (Individual), Joachim Schurig (Lionbridge), Uwe Stahlschmidt (Microsoft)

Public

Evren Ay (MotaWord), Jan Bareš (Moravia), David Čaněk (Memsorce), Daniel Chin (Spartan Software), Loïc Dufresne de Virel (Intel), Fabio Fantino (Symantec), Roger Fienhold Sheen (infotexture), Andrew Gibbons (Welocalize), David Lewis (ADAPT Centre), Christian Lieske (SAP), Arle Lommel (DFKI), Patrik Mazánek (SDL), Marc Mittag (MittagQI), Phil Ritchie (Vistatec), Chase Tingley (Spartan Software), Andzej Zydrón (XTM International)

XLIFF 2 Implementations: Lessons Learnt

Panelists: Bryan Schnabel, Chase Tingley, Andrzej Zydrón

Chair: dF

Panelists and Ryan confirmed that the XLIFF 2.0 spec is clearly written and provides much more and more useful implementation guidance than the 1.x specs. Confusion about an erroneous inline example clarified with Andrzej. Chase glossed on difference between the general 1.2 and 2.0 approach in OKAPI framework.

XLIFF 2.1 status

Chair: Bryan Schnabel

Panel: all delegates

Bryan informed the audience about the delay of the XLIFF 2.1 progress. Progress achieved during the f2f before FEISGILTT has been reported by dF and Felix. New expected date for publishing XLIFF 2.1 is by the end of October 2015.

Discussion of modules at extension points

TC currently does not have consensus whether the XLIFF 2.0 spec should be interpreted as forbidding modules at extension points where the modules are not explicitly listed in the prose specification.

Tentative consensus has been reached that 2.1 should have 2 provisions that would disambiguate the situation.

- 1) Explicit Constraint disallowing module namespaces on extension points where those are not explicitly allowed.
- 2) Allowing mda module on all extension points. Meaning adding mda explicitly on module extension points where mda hasn't been explicitly allowed in XLIFF 2.0. Core is not affected by this change. Issue will be further discussed by TC in regular teleconferences.

Also the ctr module should be amended within 2.1 to allow for tracking inline code changes.

Apart from the approved new features (ITS mapping and Advanced Validation), the above amendments will be included plus a number of editorial fixes and clarifications, e.g. erroneous examples of inline codes will be fixed in 2.1 (no errata for 2.0 is planned).

Glossary Module <-> TBX Basis mapping will be published as a Committee Note. The mapping has been so far published in Localisation Focus.

XLIFF 2.2 Requirements Gathering

Chairs: Bryan Schnabel & dF

Reporters: Chase Tingley, Christian Lieske et. al.

XLIFF:doc coverage

Chase Tingley reported on XLIFF:doc features coverage by XLIFF 2.

The conclusion was that all relevant XLIFF:doc needs represented by XLIFF:doc features were covered by XLIFF 2.0 spec. Except the tracking of matches usage and preserving match origin when used in core target. This should be largely covered by XLIFF 2.1 between ctr and ITS Provenance mapping. If the functionality is not fully covered by 2.1, the option is to submit an official 2.2 feature request through comment list roughly by October 2015.

Chase and other IN! representatives in the audience (Andrzej) endorsed XLIFF 2 as TIPP payload instead of XLIFF:doc, though both XLIFF:doc and XLIFF payload types should be legal for a transition period.

ULI Requirements

The rationale for having ULI represented in XLIFF 2.2 requirements gathering was that possibly there could be synergies or dependencies between work/concepts/resources between ULI and the XLIFF Technical Committee amongst others in the area of "segmentation". Christian captured his reflections of the discussion as follows:

1. XLIFF 2.0 addresses amongst others a number of processes related to segmentation
2. It would be valuable if XLIFF 2.x could capture information on segmentation-related processes such as

- a. This file has been segmented with the following ICU segmentation configuration/resources..
- b. This file has to be segmented with the following ICU segmentation configuration/resources..

Christian provided some basic information on ICU and ULI to facilitate the discussion. This was welcomed and the participants asked for the following:

1. Christian or David to distribute most relevant ICU and ULI TC information to XLIFF community; candidates:

- a. For ULI: conference presentations from Steven or Uwe, or <http://www.meta-net.eu/events/meta-forum-2013/posters/uli-presentation.pdf>
- b. For ULI: Online segmentation tool: <http://demo.icu-project.org/icu-bin/icusegments#3/en>
- c. For ICU: <http://site.icu-project.org/>

2. A Webinar of ULI for the XLIFF community

ULI related Conclusions:

Although U extension is valid in BCP 47 syntax that is prescribed for source and target language tags within XLIFF and users cannot be prevented from using the U extension to indicate segmentation, it does not seem possible to standardize usage and semantics of the U extension in XLIFF facilitated exchange. BTW, Mark Davis (at a W3C MultilingualWeb workshop couple of years ago) advised against usage of BCP 47 extensions in structured mark up environments such as XLIFF. This said markup mappings and processing expectations can be specified when extracting from a source environment when the existing segmentation is Some specific reasons against trying to convey segmentation information via extension U within XLIFF 2 core:

- 1) The source and target language can only be set at root level <xliff>. There are use cases for setting different segmentation rules for different files, groups, or even units, e.g. postal address fields.
- 2) It would not be clear whether the segmentation rule indicated by the U extension has been applied, is to be applied.
- 3) Clash of canResegment flag with the instruction/information on segmentation set at the root level via the U extension.
- 4) Target segmentation in an XLIFF file is driven by the source segmentation, and re-segmenting target target is not possible without resegmenting source and or reordering target segments within a unit.
- 5) Although mappings and processing expectations for extraction can be specified, XLIFF principally does not normatively prescribe extraction behavior, plus segmentation is expected to be changed during an XLIFF facilitated roundtrip AND is generally NOT expected to be performed at extraction.

The reporter and the audience concluded that a small specialized module would be required in XLIFF 2.2 that would be able to carry the segmentation information/instruction.

The vehicle could be the U extension but in a clear context (has been used/must not be changed/has not been used/has to be used in the next segmentation step etc.). Another candidate vehicle for carrying the instruction/information (and actually the one preferred by the OASIS and ETSI communities) would be embedding or referencing of SRX rules. This would require some ULI break <-> SRX mapping work on the ULI side before submitting to XLIFF TC as a feature request.

If ULI stakeholders decide to sponsor work on such a module, this feature request needs to be formally recorded via the XLIFF comment list (if requestor is XLIFF TC non-member) or on the XLIFF TC working list if requestor is an XLIFF TC member. XLIFF TC will transfer such a requirement into its public working wiki.

Andrzej Zydrón (XTM International) offered to contribute SRX rules for 16 languages to ULI. Needs to be checked for compliance with contribution rules.

Layout handling as module in XLIFF 2.2?

Joachim introduced this as a possible 2.2 feature. Joachim believes that it is an opportunity and inaction threatens the standard by someone else standardizing rendering of layout during the roundtrip. The underlying idea is that of fs, BUT in this case the handling would be strictly standardized and unlike fs this proposed feature would guarantee roundtrip of the layout information including relevant transforms.

dF opined that layout roundtrip is currently possible using resource data module. Resource data module and skeleton are both expected to guarantee roundtrip. Resource data of the mime type html with reference set to "no" would do exactly what Joachim is suggesting. The TC should look into this in more detail if the feature is officially proposed to the mailing list by October and has a TC sponsor to elaborate on it.

XLIFF 2 Object Model and other serializations

Panelists: Patrik Mazánek (SDL), Yves Savourel (ENLASO), Ryan King (Microsoft)

Chair: dF

Material discussion

Object model approaches by ENLASO, Microsoft and SDL (BCM) were briefly introduced and discussed.

Panelists and delegates agreed that XLIFF prose specification actually hints to a serialization independent object model. In some sense the XML serialization of the objects defined in the spec (Ryan) is arbitrary and there is no material reason not to have XLIFF serializations other than XML and not to have a general object model underlying not just the XML serialization, but for instance a JSON

serialization etc. The possible standardization targets are 1) other serializations, 2) the underlying object model as prose or maybe an abstract syntax such as the browser DOM as basis of an XLIFF DOM for not only browser implementations 3) maybe also a standard API for retrieving XLIFF fragments in a number of possible serializations, prominently XML, JSON, or RDF.

Institutional discussion

dF explained that the [OASIS XLIFF TC](#) under its current [Charter](#) cannot work on a generalized serialization independent object model or other than XML serializations.

The options at hand appear to be:

1) Recharter XLIFF TC so that the Object Model and non-XML serializations are in scope.

a. Pros

Not losing focus, preserving critical mass, major stakeholders already on the committee

b. Cons

Legally risky, all XLIFF TC participants since inception would need to recommit their IPR to the re-chartered TC. This could incapacitate the TC for an indefinite period and endanger progress of 2.1 and 2.2 versions.

2) Charter a parallel OASIS committee – “Sister Committee”

a. Pros

All current TC members can join w/o admin hassle. The sister committee would have a large membership overlap and work could be easily coordinated, see DITA and DITA Adoption committees. Red tape on setting up the sister committee does not directly affect progress and continuity of work on XLIFF 2.1 and 2.2. Overall less risk and cleaner cut than option 1). Has PR potential for announcements etc.

b. Cons

Requires relatively heavy weight red tape to convene the brand new TC. Active participation requires OASIS membership, so harder to engage e.g. the web community, standardization savvy folks who are W3C members but not OASIS members.

3) Charter a W3C community group

a. Pros

Low admin burden, easily accessible to all interested stakeholders at no cost.

b. Cons

Insufficient IPR protection, participating in the group does not oblige to grant essential patents on RF terms. The final Recommendation work requires chartering of a follow up Working Group that requires W3C membership

Joachim Schurig, Chair of ETSI ISG LIS offered ETSI ISG LIS as a publication channel. The object model and other serializations however need to be first developed and only then could be co-published under the ETSI OASIS MOU.

Conclusion:

No one was in favor of option 1), options 2) and 3) received both support. With option 2) having little more traction and perceived as better feasible in terms of development and the publishing of the final standard.