XSLT Conversion between XLIFF and RDF

Dimitra Anastasiou
SFB/TR8
Computer Science/Languages Science
University of Bremen
Bremen, Germany
anastasiou@uni-bremen.de

Abstract. This paper focuses on the conversion between the open standard XML Localisation Interchange File Format (XLIFF) and the Resource Description Framework (RDF). XLIFF is a localisation standard supported by proprietary and free and open source software (FOSS) localisation tools, while the latter is a standard model, basic ingredient in Semantic Web. We developed a converter based on Saxon XSLT Processor which translates XLIFF to RDF.

Keywords: Conversion, Localisation, Semantic Web, Standards.

1 Introduction

Generally speaking, standards incorporate a solid body of knowledge and provide a unified framework. In addition, when metadata is standardised, resources can be identified, catalogued, and processed faster and more efficiently. Although standards as such are a benefit for information management, in the last years we have seen too many standards evolving in information science. In our opinion, the existence of too many standards in tandem with their inflexible structure (of some standards) adds complexity and leads to lack of interoperability; interoperability between Web resources is crucial for communication between application components.

This paper focuses on XLIFF[1] and RDF[2] and the conversion based on Saxon from the former to the latter. Our work is motivated by the insight that Web resources should be multilingual and XLIFF as a localisation standard is capable to help localise ontologies and thus create multilingual linked data. A wider target range of users and applications will then be reached. The automatic conversion from XLIFF into RDF can be used as an API both by localisation tools and Semantic Web applications.

In section 2 we describe some related work about combining multilinguality with Semantic Web. In sections 3 and 4 some examples of XLIFF and RDF are provided. Section 5 discusses the XLIFF-RDF interoperability and then we conclude the paper.

---

[1] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff, 12/09/11
[2] http://www.w3.org/RDF/, 12/09/11

2 Related Work

In 2004 [1] stated that Human Language Technology faces new multilingual and multicultural challenges for the Semantic Web and presented relevant ongoing initiatives. One year later, [2] pointed out the usefulness of a multilingual Semantic Web, particularly to help translate websites through the use of ontologies, manage group knowledge in multilingual form, and create international communication base for industry and commerce. [3] used the Universal Networking Language (UNL) as a step between the process of acquiring knowledge from textual sources and translating it into one of the state-of-the-art knowledge representation formalisms for building multilingual ontologies.

The Multilingual Semantic Web workshop started in 2010 and continues with annual workshops; the same holds for the XLIFF International Symposium. Some research projects: the Multilingual Web[3], Flarenet[4], META-NET[5], and Monnet[6] see the symbiotic relationship between multilingual resources and Semantic Web.

As far as the conversion between XLIFF and other standards is concerned, the Okapi Framework provides XLIFF conversion utilities, e.g. to Translation Memory eXchange (TMX). [4] describes how to convert documents to XLIFF and back to the original format through text extraction, pre-translation, translation, reverse conversion, and translation memory improvement. A framework which combines many localisation standards is the MultiLingual Information Framework (MLIF) [5]; an overview about localisation standards can be found in [6]. A model that has been proposed to associate linguistic data to ontologies is the 'Linguistic Information Repository' (LIR) [7], designed to account for cultural and linguistic differences among languages. Lemon[7] is another model sharing lexical information on the Semantic Web; noteworthy is the converter between lemon and the Lexical Markup Framework (LMF).

Our main motivation for XLIFF2RDF conversion is the concept of 'ontology localization', a term coined by [8]: *"Ontology Localization is the adaptation of an ontology to a particular language and culture"*. [9] state that ontology localisation is an activity with both pragmatic and economic goals. The former can be seen in the fostering reuse of ontologies already available for the domain in question instead of building them from scratch, and the latter, a result of the former, is seen in the stage of cost reduction compared to building a completely new ontology.

3 XLIFF

XLIFF is an open localisation standard supported by proprietary and FOSS localisation tools. It is under the auspices of OASIS and is understood by many

---

[3] http://www.multilingualweb.eu/en, 12/09/11
[4] http://www.flarenet.eu/, 12/09/11
[5] http://www.meta-net.eu/, 12/09/11
[6] http://www.monnet-project.eu/Monnet/Monnet/English?init=true, 12/09/11
[7] http://lexinfo.net/, 12/09/11

actors: software providers, localisation service providers, and localisation tools providers. Semantic localisation metadata is very important in a localisation workflow to distinguish between the responsibilities of each stakeholder (project manager, engineer, translator, proofreader), between translatable and non-translatable content, annotate (in the case of translatable content) the status of the strings and so on. Particularly in software localisation, coordinates of menus dialogue boxes, version control, count of screenshots belong to the most important metadata. The following example contains an XLIFF file with three translation units (TUs). TU elements include a <source>, <target> and associated elements.

```
1.   <?xml version="1.0" encoding="UTF-8" ?>
2.   <xliff version="1.2" xmlns="urn:oasis:names:tc:xliff:document:1.2">
3.   <file original="minimal_XLIFF.html" source-language="en-us" target-language="de-de"
     datatype="html">
4.   <body>
5.   <trans-unit id="#1">
6.   <source>book</source>
7.   <target>Buch</target>
8.   </trans-unit>
9.   <trans-unit id="#2">
10.  <source>book publisher</source>
11.  <target>Buchverlag</target>
12.  </trans-unit>
13.  <trans-unit id="#3">
14.  <source>This book is good!</source>
15.  <target>Dieses Buch ist gut!</target>
16.  </trans-unit>
17.  </body>
18.  </file>
19.  </xliff>
```

Example 1. XLIFF file with three translation units. *Line 1*: XML declaration, *Line 2*: XML schema, *Line 3*: file metadata, *Lines 5-16*: file data (three TUs).

4 RDF

RDF is family of W3C specifications which describe Web resources. Here is a brief explanation of *Resource*, *Property*, and *Property value* by means of the XLIFF Ex.1:

- A *Resource* is anything that can have a URI, e.g. `minimal_XLIFF.html`;
- A *Property* is a Resource that has a name, such as `trans-unit`, `source`;
- A *Property value* is the value of a Property, such as `This book is good!`

The example 1 can be represented in an RDF graph as follows:
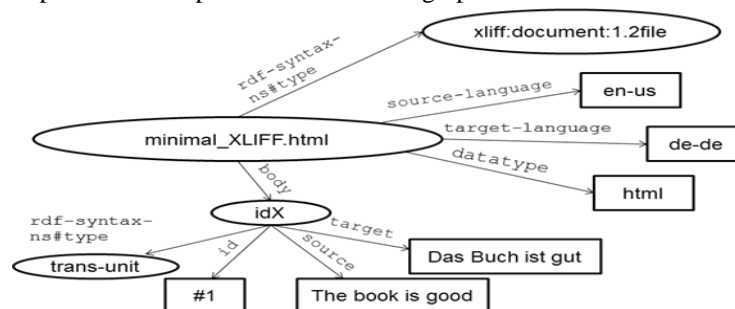
Diagram 1. RDF graph of Example 1

Accordingly, every XLIFF file can be represented in an RDF graph. The circles are the *resources,* the labels on the arrows are the *properties*, and the content of the rectangles are the *property values*. `idX` is a placeholder for a resource representing the `body`.

Building a bridge for interoperability between RDF and other standards is something common: WSDL-RDF, RDF-Topic Maps, OWL-RDF, and others. However, these standards, which RDF can be converted from and into, also come from the Semantic Web world and not from the localisation scene.

As far as the representation of multilingual information in RDF is concerned, RDF used the RFC 3066 standard (published in 2001) for language tags for literals in natural languages. The revision RFC3066bis included productive use of language, country and script codes. [10] suggested a small change to the RDF model theory to permit access to the language tag in the formal semantics, giving this ontology a precise formal meaning; their approach defined a new property called `rdflg:lang`.

5   Interoperability

The greatest contribution of XLIFF is the nature of its content, i.e. the capture of translation pairs, rather than the formalisation vehicle of the knowledge, be it XML or RDF. We do not intend to reify XLIFF, but to make XLIFF portable to RDF. The reasons why an XLIFF2RDF mapping and conversion are useful follow:

  i. Any file format which can be converted into XLIFF can be then converted to RDF;
 ii. RDF ontology labels can be translated using XLIFF;
iii. Web resources can be described by XLIFF metadata.

A practical implementation of standards' interoperability between XLIFF and RDF(S) is distinguished between two parts: mapping XLIFF elements and attributes to RDF and automatically converting from XLIFF into RDF. The mapping of three XLIFF files has been described in [11]. In order to cover more than three use cases, automatic conversion is needed. We created different types/use cases of XLIFF files and accordingly incremental EXtensible Stylesheet Language Transformations (XSLTs) to translate various XLIFF files: a file with 3 translation units, with file processing metadata, with alternative translations, a document containing two files, and a modularised file containing a lot of metadata and inline markup.

A sample of an XSLT follows:

```
1.  <xsl:stylesheet version="1.0"
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:a="urn:oasis:names:tc:xliff:document:1.2"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xliff="http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html#">
2.  <xsl:template match="/">
3.  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
4.  <xliff:file>
5.  <xsl:attribute name="rdf:about">
6.  <xsl:value-of select="a:xliff/a:file/@original"/>
7.  </xsl:attribute>
8.  <xsl:attribute name="source-language">
9.  <xsl:value-of select="a:xliff/a:file/@source-language"/>
```

Example 2. Sample of the XSLT

It should be mentioned that there is discrepancy between interoperability between data based on standards and interoperability between standards. Conversion between

standards plays a small part within the wider scope of interoperability which includes, among others, supporting relevant standards and conforming with specifications.

5.1 Converter

The development of a conversion tool to translate from XLIFF into RDF automates and thus accelerates the process. We used NetBeans IDE to create a GUI of the conversion tool (see Screenshot 1). For our conversion utilities we used the Saxon home edition 9.3 version[8]. The home edition is an open source product available under the Mozilla Public License. It provides implementations of XSLT 2.0, XQuery 1.0, and XPath 2.0 and is available for both Java and .NET. The user can input one or more XLIFF file(s) to the tool, convert them to RDF and preview them.
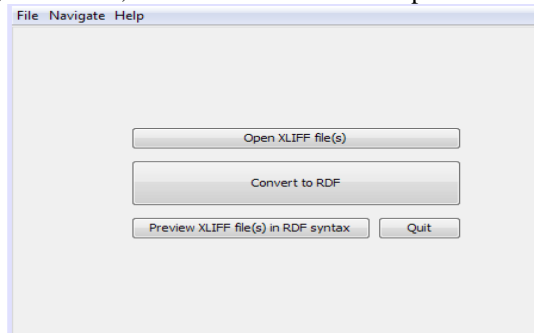
Screenshot 1. XLIFF2RDF conversion tool

The converter is under Google code hosting[9] website. There users can freely get a local copy of the tool or create their own clone.

6 Discussion and Conclusion

In this paper we discussed the interoperability between the localisation standard XLIFF and RDF. We showed ongoing initiatives, projects, and tools combining multilinguality with Semantic Web. We developed a converter from XLIFF to RDF by using and adapting the Java API of the XSLT processor Saxon. We wrote some sample XLIFF files and adopted a modular transitional file provided in the XLIFF latest specifications in order to create corresponding XSLTs.

In our opinion, localisation is often regarded only as a business strategy to increase return on investment and not as a research field which can both enrich and gain from the Semantic Web and Linked Data. Localisation standards and particularly XLIFF has received little attention although it covers many actors' needs.

In Semantic Web context, it is an arbitrary decision in which natural language the ontology labels are provided, and thus many researchers see the need for multilingual ontologies; challenges, like cross-lingual mapping and translation follow the existence

---

[8] http://saxon.sourceforge.net/, 12/09/11
[9] http://code.google.com/p/xliff-rdf/, 28/03/11

of multilingual ontologies. Our conversion tool is a contribution to build a bridge between localisation and Semantic Web resources, so that localisation tools can localise ontologies and Semantic Web resources are populated with localisation-related metadata. After the XLIFF2RDF conversion, metadata can be reused in the Semantic Web to represent multilingual ontologies. The XLIFF2RDF conversion tool is hosted on Google code hosting website. There other users can freely get a local copy of the tool; thus replication of the tool is allowed. The conversion tool fulfills its basic requirements, i.e. XLIFF files are represented in RDF. Not only minimal XLIFF examples with one TU, but with more TUs and also with file processing metadata, alternative translations, etc. can be successfully converted. Five use cases have been successfully tested, however more quantitative and qualitative examples are planned to be converted. We plan to extend the conversion API for other standards. At first place, we plan to translate from XLIFF into OWL. Also interoperability between other localisation and internationalisation standards is also among future prospects. In terms of quality assurance, existing validation tools will be part of our tool.

References

1. Declerck, T., Buitelaar, P., Calzolari, N. & Lenci, A. Towards A Language Infrastructure for the Semantic Web. *Proceedings of LREC* (2004)
2. Hahn, W. and Vertan, C. Challenges for the Multilingual Semantic Web. *Proceedings of the International MT Summit X* (2005)
3. Cardeñosa, J., Gallardo, C., Iraola, L., & De la Villa, M. A New Knowledge Representation Model to Support Multilingual Ontologies. A Case study. *Proceedings of the International Conference on Information and Knowledge Engineering*, 313-319 (2008)
4. Raya, R. XML Localisation Interchange File Format as an intermediate file format. *IBM developerWorks* (2004) http://www.maxprograms.com/articles/xliff.html
5. Cruz-Lara, S., Bellalem, N. Ducret, J. & Kramer, I. Standardizing the management and the representation of multilingual data: the MultiLingual Information Framework. *International Workshop on Language Resources for Translation work, Research and Training*, 35--38 (2006)
6. Anastasiou, D., Morado Vázquez, L. Localisation Standards and Metadata. *Proceedings of the 4th Metadata and Semantics Research Conference (MTSR 2010),* Communications in Computer and Information Science, Springer, 255--276 (2010)
7. Peters, W., Montiel-Ponsoda, E. & Aguado de Cea, G. Localizing Ontologies in OWL. *Proceedings of the ISWC07 OntoLex workshop* (2007)
8. Suarez-Figueroa, C., M. and Gomez-Perez, A. First attempt towards a standard glossary of ontology engineering terminology. *Proceedings of the 8th International Conference on Terminology and Knowledge Engineering* (2008)
9. Cimiano, P., Montiel-Ponsoda, E., Buitelaar, P., Espinoza, M., & Gomez-Perez, A. A note on ontology localization. *Journal of Applied Ontology* (JAO), 5(2), 127--137 (2010)
10. Carroll, J.J., & Phillips, A. Multilingual RDF and OWL. *The Semantic Web: Research and Applications*, *Lecture Notes in Computer Science*, Vol. 3532/2005, 15--19 (2005). doi: 10.1007/11431053_8
11. Anastasiou, D. XLIFF Mapping to RDF. *JIAL (The Journal of Internationalisation and Localisation),* to appear (2011)