

RDF/XHTML: Ontology Editing in HTML

Kun Ji and Lauri Carlson

Department of Modern Languages, University of Helsinki, Helsinki, Finland

Keywords: Ontology Engineering, Semantic Web, Ontology Sharing and Reuse, Natural Language Processing.

Abstract: Although global, the web and its standards are not language independent. Semantic Web document standards remain skewed toward Western languages and alphabets. This causes problems for truly multilingual terminology work in the Semantic Web. Some of the problems are discussed and remedied by the proposals in this paper. In particular, we present a format for representing RDF triple sets in XHTML and software that supports multilingual editing of RDF/OWL ontologies in this format. The format was developed for and is used in TermFactory, an ontology based multilingual terminology management environment.

1 INTRODUCTION

Resource Description Framework (RDF 2004) is a modeling language originally meant for annotating semantic meta data in the web. RDF is a graph based data model that allows making statements about resources (in particular Web resources, identified by URIs) and their relationships in the form of subject-predicate-object triples. RDF Schema (RDFS 2004) adds vocabulary and axioms for defining classes and their instances, and primitives for defining further vocabulary. One such vocabulary is Web Ontology Language (OWL 2009), a Web version of description logic. OWL has a back translation to RDF in which one OWL predicate or construct may go over to a RDF graph composed of several triples.

Ontology languages like RDF and OWL have become a significant format for describing complex concept systems in areas such as natural sciences and medicine. The main focus in ontology work has been in the concept systems as such. Many large scale domain ontologies are language neutral, in practice, biased toward English. To the extent multilingual or indeed natural language terminology is included, it is provided as simple string labels. Terms are usually not described as ontological resources in their own right.

But nothing prevents describing natural language terms as ontology resources as well (Buitelaar et al. 2011). TermFactory (Kudashev et al, 2010, Carlson, 2012) is an ontology-based terminology management system that does just that. TermFactory

(TF) comprises an ontology schema, web API, and platform for collaborative terminology work that is based on explicit ontological representation of both concepts and their designations.

2 TERMFACORY

TermFactory (TF) is an architecture and a workflow for distributed and collaborative terminology work in a global multilingual context. TF applies Semantic Web technologies to the representation of specialised multilingual terms and related concepts. It also provides a workflow by which terminologies can be collected, updated and agreed about by professionals in relevant fields all over the globe, during their everyday work, using virtual work platforms over the web.

TF can be considered a semantic web framework for multilingual terminology work. It provides ontology and terminology formats, format conversions, query and edit tools, repositories, web services. TF enables people to do professional quality terminology work jointly or separately, building their work on others' efforts, maintaining quality and consistency of the jointly developed terminology.

With Semantic Web techniques TF aims to achieve:

- Openness and conformance: both conceptual and linguistic content can be globally identified and mechanically validated.

- Flexible reuse of content: different ontologies and terminologies should be able to coexist, complement one another, and co-develop on separate sites.
- Ease of implementation and deployment: contents usable by third party tools, help divide and conquer big ontologies.

For TermFactory web ontology based terminology management, we found a need for editing tools for non-ontologists to edit term ontologies that would be simpler and more accessible to terminologists than mainstream ontology editors. We designed RDF/XHTML and its Web API as an answer to this need. As the format and tools are quite generic with little that is specific to term ontologies in particular, we propose to present them here to the ontology developer community at large.

3 RDF FORMATS

RDF has several serialization formats (file formats) which vary in the way in which resources and triples are encoded. For historical reasons, XML is the official syntax for RDF. For multilingual ontologies, RDF/XML is not a good choice. A concrete flaw of RDF/XML for multilingual (or non-Latin, in general) ontologies is that there is no provision for coding property names containing non-Latin characters. The only RDF/XML representation for property names is XML element name (QName), which has a restricted character repertoire. Non-Latin property names require lengthy and unreadable character encodings. A simple expedient would be to extend RDF/XML with property elements identified by full URIs, e.g.

```
<ex:label>term</ex:label>
```

could be written as

```
<rdf:Property
  rdf:about="http://example.com#label">
  term
</rdf:Property>
```

A simpler alternative to RDF/XML is Turtle (2010), a textual format for RDF graphs close to the triple format. Turtle is terse and human readable. Yet it too has limitations. Resource names cannot be abbreviated with prefixes if they contain Turtle reserved characters. It would be as well for Turtle not to reserve punctuation characters, since Turtle punctuation is conventionally separated by whitespace anyway.

4 ONTOLOGY EDITING

Syntax editing of ontology triples can yield unexpected results. Deletion of facts in general involves difficult problems of nonmonotone reasoning or belief revision. The best one can do is avoid redundancies by using some normal form.

A normal form is a unique choice among equivalent representations. Reduction to normal form by term rewriting is what many reasoners in effect do. RDF/OWL databases are supposed to keep graphs in a nonredundant form to support updates.

The standard serializations of RDF do not provide a unique normal form. Textual normal forms for RDF have been proposed (Carroll/Stickler 2004, Dau 2006, Gutierrez et al. 2011). Semantic normal forms for some description logics have been proposed (Hitzler/Eberhart 2007, Bienvenu 2008).

We have argued that the standard serializations of RDF are not well suited for multilingual ontology editing as such. Special purpose ontology editors avoid problems by building their own graphical editing interfaces often borrowing from Eclipse. Many standalone ontology editors exist, both open source and commercial.

5 EDITING IN RDF/XHTML

In designing TF, we did not want to build yet another application. Instead, we wanted to choose or adapt a serialization format for the web that is familiar to users, has support in general purpose web editing tools, without yet compromising machine processability.

XHTML seems to best fill the bill. As an extension of XML, it supports Unicode and can be manipulated with common XML tools. As the native representation format of browsers, can be depended on to provide good support for display. XHTML can be edited with a wide range of standalone tools and browser extensions. The HTML 5 standard (2012) is to merge with XHTML and provide built in support for direct editing.

The RDF/XHTML format represents RDF models in the form of a sorted HTML list of tree-structured entries, isomorphic with Turtle. It supports WYSIWYG editing through a user definable XHTML skin (Figure 1). The idea is similar to that applied in XML editors like XMLmind (2012).

The layout of the XHTML document can be customised with templates also written in RDF. The output of the XHTML writer can be varied with a number of parameters:

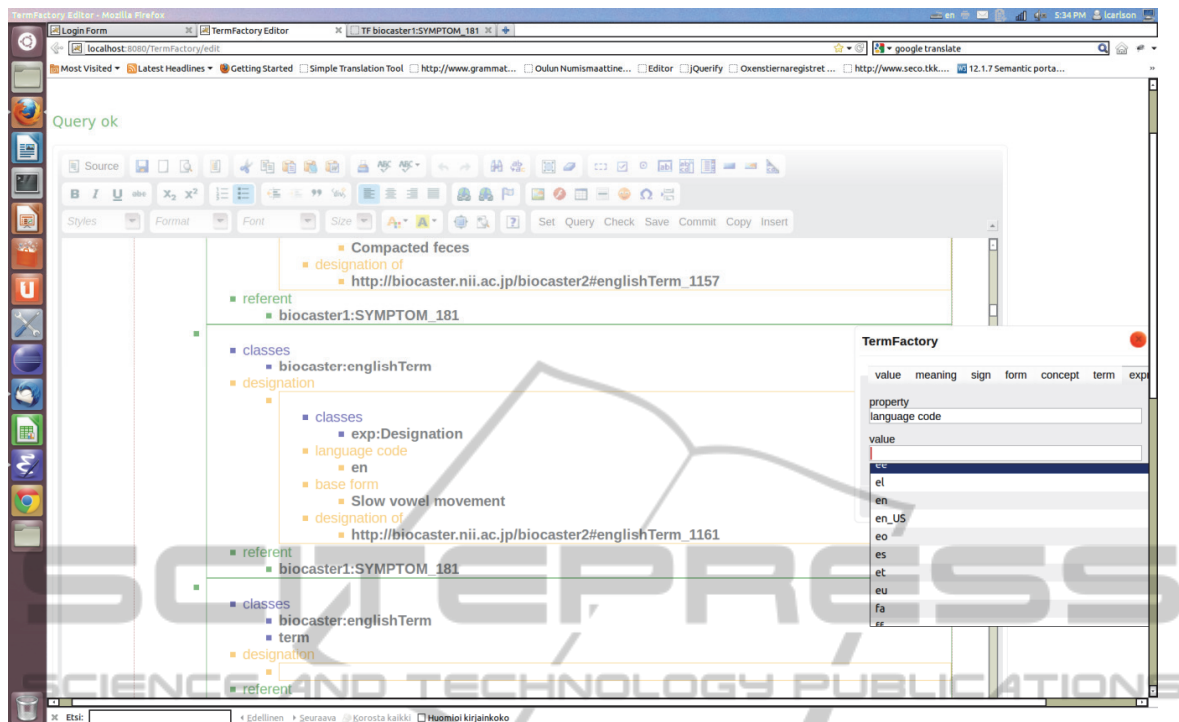


Figure 1: Editing TF/XHTML with a textarea editor.

Table 1: RDF/XHTML parameters.

template=<URI>	RDF template to define the structure of the serialization
schema=<URI>	RDF/OWL schema to bridge between TermFactory and user ontologies
root=<URI>	output filter (list of instances/classes to include in the output)
active=<URI>	active ontology (editable triples)
locals=<URI>	localization vocabulary (a multilingual term ontology)
lang=<ISO langcode>	localization language
links=<URI>	hyperlink mapping for redirecting resource URIs

The XHTML writer writes these options it used in the header of XHTML document head element and they are stored in the RDF model by the RDF writer. As a result, an XHTML entry can be roundtripped through RDF without having to bother about the settings, and RDF/XHTML entries complete with layout can be stored in a RDF database.

TF tries to facilitate multilingual terminology editing by making terms communicable across locales. TF does this by way of localization, using URIs as an interlingua. It converts local terms to global URIs and localizes them back as terms

familiar to users in another locale (Figure 2).

TF meta classes and properties like *Concept*, *Term*, or *hasDesignation* are also described in TF as (instances of) concepts, terms, and properties. It means that TF is capable of *reflection*: it can document and localise itself. A TF schema localisation ontology provides definitions and translations of TFS descriptive classes and properties in TF itself. This information is then used to change language in the TF front end tools. Figure 2 shows an entry localized into Chinese.

Ontology edit operations, such as actual modification of the active ontology with the edits supported by a webservice API implemented as an Axis2 webservice **EditService** using the Jena library, mediated by a Java servlet **EditForm**.

6 CONCLUSIONS

RDF/XHTML has no editing support for multiple-RDF triple constructs like OWL axioms. Currently, it is best suited for editing RDF or OWL instance bases (ABox).

HTML 5 defines primitives to support online editing of HTML documents. We expect to be able to generalize our approach further come HTML 5.

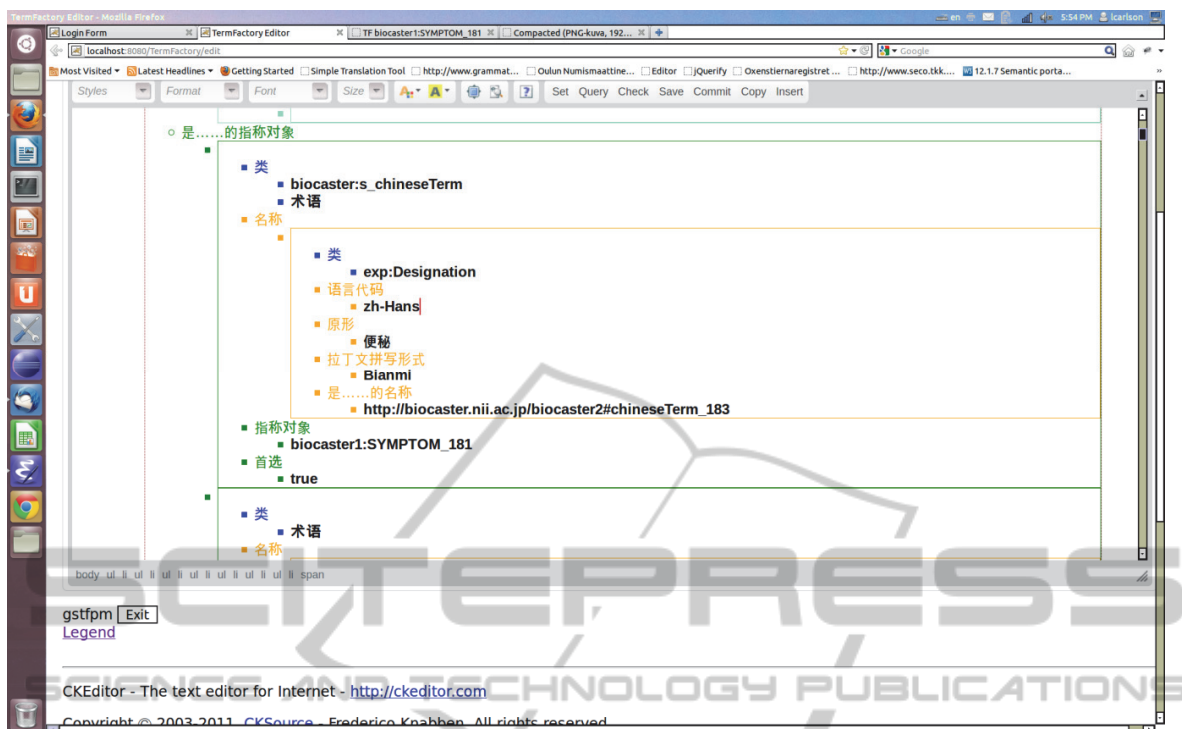


Figure 2: TF/XHTML entry localized into Mandarin.

REFERENCES

- Buitelaar, P., Cimiano, P., McCrae, J., Montiel-Ponsoda, E., Declerck, T. 2011. Ontology Lexicalization: The lemon perspective, TIA 2011, November 8-10 2011. http://oa.upm.es/9772/1/Ontology_Lexicalisation.pdf (accessed 4 July 2012)
- Carlson, L., 2012. TermFactory Manual. http://tfs.cc/doc/TFManual_en.xhtml (accessed 30 April 2012)
- Francopoulo G., Bel N., George M., Calzolari M., Pet M., Soria C., 2008. Multilingual resources for NLP in the lexical markup framework (LMF). Language Resources and Evaluation (revue) ISSN 1574-020X (print) + 1572-0218 (online) Springer Netherlands
- Gutierrez, C., Hurtado, C., Mendelzon, A., Pérez, J., 2011. Foundations of Semantic Web databases. J. Comput. Syst. Sci. 77(3): 520-541 (2011)
- HTML5, 2012: A vocabulary and associated APIs for HTML and XHTML. Editor's Draft 28 April 2012.(Accessed 30 April 2012)
- Kudashev, I., Carlson, L., Kudasheva, I., 2010. TermFactory: Collaborative Editing of TermOntologies. In: Bhreathnach, Ú., Barra-Cusack, F. (eds.) Terminology and Knowledge Engineering Conference 2010, pp. 479–500. Fiontar, Dublin City University, Dublin (2010)
- OWL, 2009. OWL 2 Web Ontology Language Document Overview.W3C Recommendation 27 October 2009. <http://www.w3.org/TR/owl2-overview/> (accessed 30 April 2012)
- RDF, 2004. Resource Description Framework (RDF). <http://www.w3.org/RDF/> (accessed 30 April 2012)
- RDFS, 2004. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/rdf-schema/> (accessed 30 April 2012)
- Turtle, 2010. Turtle – Terse RDF Triple Language. <http://www.w3.org/TeamSubmission/turtle/> (accessed 30 April 2012)
- XMLmind, 2012. XML Editor. <http://www.xmlmind.com/> (accessed 4 July 2012)