Representing multilingual lexical and terminological information in RDF vocabularies

BabelData project - TIN2010-17550

Elena Montiel-Ponsoda

Ontology Engineering Group – Universidad Politécnica de Madrid

Event: Vocabulary Day – DCMI 2013
Place: Lisbon, Portugal
Date: September 2013
• Representing lexica (a quick overview)
  • RDF(S), OWL
  • SKOS, SKOS-XL
  • LMF
  • ...

• lemon
  • Main features
  • lemon core
  • Representing lexical variants, terminological variants and translations
  • Tools

• W3C Ontology Lexica Community Group
RDF(S), OWL

dct:Abstract \rdfs:label \rightarrow \text{“Abstract”}@en
**RDF(S), OWL Example**

- **source**

- **provenance**

- **provenance**
  "Curso de agua principal - Catalogo de fenomenos. Proyecto GEOALEX"@es

- **comment**
  "A body of inland water flowing for the most part on the surface of the land but which may flow underground for part of its course."@en

- **comment**
  "Masa de agua continental que fluye en su mayor parte sobre la superficie del suelo, pero que puede fluir bajo tierra en parte de su curso"@es

- **label**
  "River"@en

- **label**
  "Curso de agua principal"@es

- **label**
  "Curso fluvial"@es
SKOS—Simple Knowledge Organization System—model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary.\(^1\)

1 http://www.w3.org/2004/02/skos/specs

2 http://www.mkbergman.com/date/2007/05/
RDF(S), OWL

dct:creator

```
rdfs:label
```

“Creator”@en

SKOS labels: prefLabel, altLabel & hiddenLabel.

SKOS

dct:creator

```
rdfs:SubPropertyOf
```

```
skos:prefLabel
```

“Creator”@en
SKOS enables a simple form of **multilingual labeling**:

```
dct:creator rdf:type rdfs:Property;
   skos:prefLabel "creator"@en;
   skos:prefLabel "creador"@es.
```

How can we create **explicit links** between labels?

What if we need to define **two preferred labels** in the same language?
SKOS-XL

dct:creator

skosxl:labelRelation

rdfs:subPropertyOf

ex:isTranslationOf

skosxl:prefLabel

dct:CreatorLabel1

skosxl:literalForm

“Creator”@en

ex:isTranslationOf

“Creador”@es

skosxl:literalForm

dct:Creator2

rdfs:subPropertyOf

skosxl:Label

rdf:type

skosxl:Label

rdf:type
What is the evolution?
LMF

Lexical markup framework
Limitations

- **SKOS**
  - Limited for richer linguistic descriptions

- **LMF**
  - Conceptualisation not ontology-driven
  - Not RDF-native
  - Some data categories hidden inside literal values

- **LingInfo, LIR (Linguistic Information Repository), LexInfo, LexOnto...**
An RDF-based lexicon model for ontologies
Main features:

- Semantics by **reference**
- Rich **lexical** and **terminological** description of ontology elements
- **Concise** (i.e., trade off between complexity and expressivity)
- **Descriptive** not prescriptive (i.e., uses data categories)
- **Modular** and extensible

**Details in lemon cookbook:** [http://lexinfo.net/lemon-cookbook.pdf](http://lexinfo.net/lemon-cookbook.pdf)
ISOcat recommends linking using the dcr:datcat property

- i.e., Create a property mylexicon:partOfSpeech.
- Add triple relation to ISOcat identifier, e.g., DC-396
• Main features:
  • Semantics by reference
  • Rich **lexical** and **terminological** description of ontology elements
  • **Concise** (i.e., trade off between complexity and expressivity)
  • **Descriptive** not prescriptive (i.e., uses data categories)
  • **Modular** and extensible

*Details in lemon cookbook:* [http://lexinfo.net/lemon-cookbook.pdf](http://lexinfo.net/lemon-cookbook.pdf)
Linguistic Description

Morphology

Variation

Syntax and Mapping

Phrase Structure

depends

may use

may use
Ontology

: Lexicon
  language="en"

entry

: LexicalForm
  writtenRep="animal"@en

form

: LexicalEntry

sense

: LexicalSense

reference

: OntologyReference
  http://example.org/ontology#animal
Lexical variants

- **Lexical Form**: writtenRep="creador"@es
  - property: isocat:masculine

- **Lexical Form**: writtenRep="creadora"@es
  - property: isocat:feminine

- **Lexical Entry**: form
  - sense
  - reference: dct:creator

- **Ontology Reference**:
  - dct:creator
Terminological variants

Lexical From
writtenRep="tuberculosis"@en

Lexical From
writtenRep="phthisis"@en

Ontology reference
agrovoc:tuberculosis

Ontology

Lexical Entry
form

Lexical Sense
sense

Lexical Entry
form

Lexical Sense
diachronic variant

Lexical Entry
form

Lexical Sense
reference

Lexical Entry
form

Lexical Sense
reference

Lexical Entry
form

Lexical Sense
reference
Downloads

- Download the model as RDF/XML
- Download the model as Turtle
- View the model as HTML+RDFa

Other Tools

- Java Lemon API
- Lemon 2 GF Converter
- Lemon Source
**author**

**noun**

Instance of: lemon:LexicalEntry

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ns1:partOfSpeech</td>
<td>lexinfo:commonNoun</td>
</tr>
<tr>
<td>lemon:canonicalForm</td>
<td>canonicalForm</td>
</tr>
<tr>
<td>rdf:type</td>
<td>lemon:Form</td>
</tr>
<tr>
<td>lemon:writtenRep</td>
<td>+ “author”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>lemon:sense</th>
<th>sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdf:type</td>
<td>lemon:LexicalSense</td>
</tr>
<tr>
<td>lemon:isA</td>
<td>subject</td>
</tr>
<tr>
<td>rdf:type</td>
<td>lemon:Argument</td>
</tr>
<tr>
<td>lemon:reference</td>
<td><a href="http://dbpedia.org/ontology/Author">http://dbpedia.org/ontology/Author</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>lemon:sense</th>
<th>sense1</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdf:type</td>
<td>lemon:LexicalSense</td>
</tr>
<tr>
<td>lemon:objOfProp</td>
<td>adpositionalObject</td>
</tr>
<tr>
<td>rdf:type</td>
<td>lemon:Argument</td>
</tr>
<tr>
<td>lemon:reference</td>
<td><a href="http://dbpedia.org/ontology/writer">http://dbpedia.org/ontology/writer</a></td>
</tr>
<tr>
<td>lemon:subjOfProp</td>
<td>subject1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ontology-Lexica Community Group

The mission of the Ontology-Lexicon community group is to: (1) Develop models for the representation of lexica (and machine readable dictionaries) relative to ontologies. These lexicon models are intended to represent lexical entries containing information about how ontology elements (classes, properties, individuals etc.) are realized in multiple languages. In addition, the lexical entries contain appropriate linguistic (syntactic, morphological, semantic and pragmatic) information that constrains the usage of the entry. (2) Demonstrate the added value of representing lexica on the Semantic Web, in particular focusing on how the use of linked data principles can allow for the re-use of existing linguistic information from resources such as WordNet. (3) Provide best practices for the use of linguistic data categories in combination with lexica. (4) Demonstrate that the creation of such lexica in combination with the semantics contained in ontologies can improve the performance of NLP tools. (5) Bring together people working on standards for representing linguistic information (syntactic, morphological, semantic and pragmatic) building on existing initiatives, and identifying collaboration tracks for the future. (6) Cater for interoperability among existing models to represent and structure linguistic information. (7) Demonstrate the added value of applications relying on the use of the combination of lexica and ontologies.
• [http://www.w3.org/community/ontolex/](http://www.w3.org/community/ontolex/)
• Work begun in December 2011
• Monthly telcos
• Chaired by Paul Buitelaar (DERI, Galway) and Philipp Cimiano (University Bielefeld)
• This work has been supported by the BabelData (TIN2010-17550) Spanish national project
Representing multilingual lexical and terminological information in RDF vocabularies

BabelData project - TIN2010-17550

Elena Montiel-Ponsoda

Ontology Engineering Group – Universidad Politécnica de Madrid

Event: Vocabulary Day – DCMI 2013
Place: Lisbon, Portugal

Date: September, 2013