

Ontology-Lexica Community Group: Mission Statement

Motivation

Ontologies have numerous applications and they represent the conceptual backbone of the Semantic Web. Significant efforts have gone into standardization efforts under the auspices of the W3C to produce recommendations for ontology languages, i.e. the Resource Description Framework (RDF)¹ and the Web Ontology Language (OWL)².

While such ontology languages enable the definition of logical theories and corresponding axioms by use of formalized symbols, a linguistic grounding of ontologies is crucial in order to render them useful for human interaction with ontology-based systems. For instance, it seems reasonable to assume that Semantic Web access will for a large extent be mediated by natural language as this is the preferred means of expression and communication employed by human beings.

However, current web-based knowledge representations languages such as OWL and RDF(S) lack the rich linguistic grounding that is required for language-mediated access to ontologies. OWL and RDF(S) rely on the `rdfs:label` property to capture the relation between a vocabulary element and its lexicalization in a given language. These lexicalizations in some sense provide a lexical anchor that makes the concept, property, individual etc. understandable to a user. While SKOS³ distinguished between further types of labels, i.e. preferred, alternative and hidden, it does not increase the expressivity of the label system from a linguistic point of view. The mechanisms for linguistic grounding available in OWL, RDF(S) and SKOS can thus be seen at best as rudimentary. They are far from enabling the capture of necessary linguistic information that ontology-based NLP applications need. Such applications include for example:

- ontology-based natural language generation systems that produce coherent discourses by verbalization of a set of triples
- ontology-based question answering systems that interpret user questions with respect to one or more ontologies
- ontology-based text mining systems that extract triples from texts with respect to a given ontology
- ...

Mission and Milestones

The mission of the ontology-lexica working group is to produce a specification for a lexicon-ontology model that can be used to provide a rich linguistic grounding for (domain) ontologies, including the representation of morphological and syntactic properties of ontology terms (labels), but most importantly to “go beyond the label”, stating alternative lexicalizations of the vocabulary elements in question as well as their linguistic properties. An important issue herein will be to clarify how existing language resources can be leveraged and reused for this purpose.

¹ <http://www.w3.org/RDF/>

² <http://www.w3.org/2004/OWL/>

³ <http://www.w3.org/2004/02/skos/>

As a byproduct of our work on specifying a lexicon-ontology model, we intend that such a model will become also the basis for a web of lexical linked data (LLD): a network of lexical and terminological resources that are linked according to Linked Data principles forming a large network of lexical knowledge.

Six important meta-requirements can be already advanced:

- Requirement 1: The actual model will be an OWL ontology, while a specific lexicon instantiating the model will be a plain RDF document.
- Requirement 2: (“Modularity”): the models should be organized in dependent modules that can be combined to fit the needs of a specific application.
- Requirement 3: (“Multilingualism”): The model should support the specification of the linguistic grounding with respect to any language, including scripts, dialects, etc.
- Requirement 4: (“Semantics by reference”): The meaning of lexical entries will be specified through a principle we call “semantics by reference” by which the semantics of a linguistic unit with respect to a given ontology will essentially be specified by referencing the URI of a class or property of the ontology in question.
- Requirement 5: (“Openness”): the ontology-lexicon model will be “open” in two ways; first, it will be extensible by new vocabulary elements which extend existing elements, e.g. as required by a certain application. Second, it will not make unnecessary choices with respect to which linguistic theory to use, leaving open the possibilities to have very different instantiations of the ontology-lexicon model.
- Requirement 6: (“Reuse of relevant standards”) We will aim to reuse as many standards as possible, in particular lexicon or terminology models such as LMF⁴ or TMF⁵ as well as linguistic data categories such as defined by ISO⁶.

Specifying requirements and use cases for such a model will be an important milestone for this working group. Roughly, the six milestones for the ontology-lexica group are as follows:

- Milestone 1: Specification of ontology-lexicon use cases, i.e., identify and describe real use cases that require such a model in the context of natural language generation, ontology-based information extraction etc.
- Milestone 2: Specification of requirements on the ontology-lexicon model derived from the above use cases
- Milestone 3: Specification of the core ontology-lexicon model
- Milestone 4: Development of a reference implementation
- Milestone 5: Development of specific modules (e.g. morphology, translation, syntax, etc.)
- Milestone 6: Release of the final specification including the core ontology-lexicon model and modules on which consensus has been reached.

Scope

⁴ <http://www.lexicalmarkupframework.org/>

⁵ <http://www.loria.fr/projets/TMF/>

⁶ <http://www.isocat.org/>

While we have mentioned above what the main mission of the ontology-lexica group is supposed to be, we find it equally important to explicitly exclude certain goals. The mission of the group will not be:

1. to create an annotation format allowing to annotate text with ontological concepts (see recent work on RDFa⁷)
2. to develop specific NLP tools that make use of ontology-lexica
3. to develop yet another lexicon model as various models exist here, e.g. LMF
4. to develop yet another terminology standard: see TMF etc.

⁷ <http://www.w3.org/TR/xhtml-rdfa-primer/>