The Ontology-Lexicon Model

The Ontolex Community Group

June 7, 2013

1 Motivation

Ontologies have numerous applications and they represent the conceptual backbone of the Semantic Web. In fact, significant efforts have gone into standardization efforts under the auspices of the W3C to produce recommendations for data and knowledge representation languages, i.e. the Resource Description Framework (RDF) and the Web Ontology Language (OWL). While such ontology languages allow us to define logical theories consisting of ungrounded symbols and corresponding axioms, a grounding in language is crucial in order to render such ontologies for human consumption and thus support meaningful interaction with them by human users. Going further, it seems reasonable to assume that access to the Semantic Web will be to a large extent mediated by language as this is the natural means of expression and communication employed by humans. 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 a property rdfs:label to capture the relation between a vocabulary element and its (preferred) lexicalization in a given language. This lexicalization in some sense provides a lexical anchor that makes the concept, property, individual etc. understandable to a user. The mechanisms for linguistic grounding available in OWL and RDF(S) can be seen at best as rudimentary. They are far from being able to capture the necessary linguistic and lexical information that NLP applications working with a particular ontology need. Such NLP applications are for example:

- Natural language generation systems that produce coherent discourses verbalizing a set of triples.
- Question Answering systems that interpret user questions with respect to ontologies.
- Text interpretation systems that interpret texts with respect to a given ontological vocabulary, extracting triples with respect to this vocabulary
- Information retrieval systems

2 Mission and Goal

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 particularly focusing on how the use of linked data principles can allow for the re-use of existing linguistic information from resource 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.

3 General Requirements on the Model

Five important meta-requirements can be already advanced:

- R1: The actual model will be an OWL ontology, while a specific lexicon instantiating the model will be a plain RDF document.
- R2: (Multilinguality): The model should support the specification of the linguistic grounding with respect to any language
- R3 (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 lexical entry with respect to a given ontology will essentially be specified by referencing the URI of the concept or property in question.
- R4: (Openness): the lexicon-ontology model will be open in two ways; first, it will also be extensible by new constructs as needed, e.g. by a certain application. Second, it will not make unnecessary choices with respect to which linguistic data categories to use, leaving open the possibilties to have very different instantiations of the lemon model. In this sense lemon can thus be called a lexicon-ontology meta-model.
- R5: (Reuse of relevant standards) We will aim to reuse as many standards as possible, in particular lexicon models such as LMF as well as terminology models such as LMF as well as linguistic data categories

4 Rationale for the design of the model

The main purpose of the lexonto model is to capture the meaning of lexical entries with respect to a given domain ontology in a so called *ontology lexicon*. Let the ontology lexicon entry contain a set of lexical entries L. Let the ontology introduce a vocabulary V consisting of predicates of any arity as well as constants.

The main relation between a lexical entry and a vocabulary elements of the domain ontology is the relation `denotes`. We define the relation `denotes` as follows:

Definition 1 (denotes)
A lexical entry $l \in L$ denotes a vocabulary element $v \in V$ iff and only if for all situations in which l is used to refer to referent r, r is contained in the extension of v.

pci: the above needs to be refined still, but as a first approximation it should suffice to get discussions started.

The `denotes`-relation thus specifies the possible meaning of a given lexical entry in the context of a given ontology by referring (pointing) to all the concepts in the ontology that the lexical entry can refer to (in some context/situation).

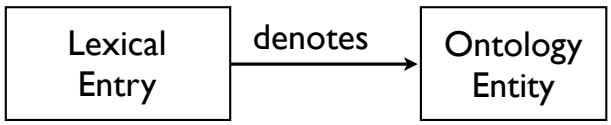
The domain of `denotes` is then clearly a Lexical Entry, while the range of `denotes` is what we informally call an *extensional entity*, i.e. an ontological entity that has an extension in some interpretation/model/world.

$$\forall x, y \; ontolex : denotes(x, y) \rightarrow ontolex : LexicalEntry(x)$$

Further, a lexical entry is a semiotic expression (a form!), i.e.

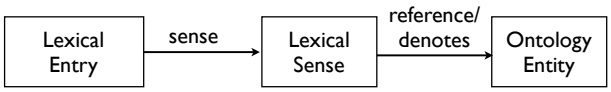
$$\forall x \; ontolex : LexicalEntry(x) \rightarrow semio : Expression$$

This modelling can be graphically depicted as follows:



Ontolex allows the relation between a lexical entry and the ontology entity it denotes to be mediated by a `Lexical Sense`. While the ontological concept represents the meaning of the lexical entry in a language-independent fashion, a Lexical Sense represents the specific (lexical) meaning of the lexical entry in the particular language. This lexical sense can be used to attach pragmatic constraints and contextual conditions under which it is legitimate to interpret lexical entry l as referring to an element in the extension of v. We define what a lexical sense as follows:

Definition 2 (Lexical Sense)
A lexical sense represents the disambiguated, lexical meaning of lexical entry l when interpreted as referring to some entity e contained in the extension the vocabulary v that it denotes.



The relation between a lexical entry and its denotation can be further mediated by what we call a *lexical concept*:

Definition 3 (Lexical Concept)
A lexical concept v is an intensional entity representing the meaning of one or several lexical entities.

A lexical concept in this sense is not an extensional entity as introduce above. Examples of lexical concepts are WordNet synsets, which represent the lexical meaning in a particular language of a set of (synonym) words. A skos:concept in this sense can also be a lexical concept.