The Metadata Module for OntoLex

Manuel Fiorelli¹, Armando Stellato¹, John P. McCrae², Philipp Cimiano², Maria Teresa Pazienza¹

¹ ART Research Group, University of Rome “Tor Vergata”, Rome, Italy
{fiorelli, stellato, pazienza}@info.uniroma2.it

² Cognitive Interaction Technology Center of Excellence, University of Bielefeld
{jmccrae, cimiano}@cit-ec.uni-bielefeld.de
Linguistically Grounded Semantic Web

A plethora of models for the lexicalization of ontologies.

W3C Standards
- RDFS $\rightarrow$ human readable labels
- SKOS $\rightarrow$ distinction between preferred, alternative and hidden labels
- SKOS-XL $\rightarrow$ reification of labels to allow for their description in RDF

Academic Proposals
- LingInfo (Buitelaar et al., 2006) $\rightarrow$ internal structure of labels (morphosyntactic description)
- LexOnto (Cimiano et al., 2007) $\rightarrow$ external structure of labels (mapping between syntactic and semantic arguments)
- Linguistic Watermark (Pazienza et al., 2008; Oltramari & Stellato, 2008) $\rightarrow$ a software library and a suite of ontologies for describing lexical resources and the lexical assets of ontologies
- LexInfo (Cimiano et al., 2011) $\rightarrow$ LingInfo + LexOnt mediated by LMF
- LIR (Montiel-Ponsoda et al., 2011) $\rightarrow$ deals with lexico-cultural aspects of ontologies
- lemon (McCrae et al., 2012) $\rightarrow$ a modular specification based on the models above
One Model to Rule Them All

The W3C Community Group Ontology-Lexica (OntoLex)
http://www.w3.org/community/ontolex/

• Gathered the creators of various lexicalization models
• In particular, it was informed by the original lemon model
• It benefits from lessons learned and experimentations with previous models
• will deliver an agreed-upon lexicon model for ontologies (Ontolex-Lemon)
OntoLex: Semantics by reference

person
OntoLex: Semantics by reference

person

people
OntoLex: Semantics by reference

Person : NN

person

people
OntoLex: Semantics by reference
OntoLex: Semantics by reference
OntoLex: Semantics by reference

A lexicalization

Person : NN

ontolex:denotes

foaf:Person

Manual Fiorelli
fiorelli@info.uniroma2.it  art.uniroma2.it/fiorelli
OntoLex: Semantics by reference

Individual : NN

- individual
- individuals

Person : NN

- person
- people

ontolex:denotes

foaf:Person

A lexicalization
OntoLex: Semantics by reference

myLexicon : ontolex:Lexicon

Individual : NN

- individual
- individuals

Person : NN

- person
- people

Linguistic knowledge

ontolex:denotes

A lexicalization

ontolex:denotes

foaf:Person
OntoLex: Semantics by reference

myLexicon : ontolex:Lexicon

Individual : NN

Person : NN

Linguistic knowledge

individuai

individuals

person

people

ontolex:denotes

ontolex:denotes

A lexicalization

Describes the semantics of a domain

FOAF : owl:Ontology

OntoLex: Semantics by reference

18/06/2015

Manuel Fiorelli
fiorelli@info.uniroma2.it  art.uniroma2.it/fiorelli
OntoLex: many-to-many relation between ontologies and lexica

English Lexicon for FOAF

Italian Lexicon for FOAF

Princeton WordNet

English for DBpedia Ontology

FOAF

DBpedia Ontology
Problems: discovery, understanding and exploitation

• How can we discover relevant resources?

• How are they encoded?

• How can we exploit their linguistic content?
Our Approach

• A metadata-based approach
• Metadata about ontologies, lexica and lexicalization set
• Provides metadata about the linguistic content:
  – Content (which natural languages?)
  – Interlinking (which lexicon has been used to lexicalize this ontology?)
  – Data summaries (to which extent an ontology has been lexicalized?)
• Metadata can be gathered in remotely accessible catalogs
• Metadata can be used to reason upon resources:
  – Understanding where we are (e.g. the linguistic compatibility of two ontologies)
  – Deliberate how to achieve our goal (e.g. how do we match them?)
Overview

• The OntoLex Lemon model
• VoID
• LIME
The Lexicon Model for Ontologies (lemon) consists of:

- A core model (ontolex)

Together with additional modules for specific use cases and applications:

- Syntax and Semantics (synsem)
- Decomposition (decomp)
- Variation and Translation (vartrans)
- Linguistic Metadata (lime)
Vocabulary of Interlinked Datasets

VoID (Vocabulary of Interlinked Datasets)

[http://www.w3.org/TR/void/]

• W3C Interest Group Note 03 March 2011

• A dataset consists of triples published, maintained or aggregated by a single provider

• Description of RDF datasets and their interconnections
An excerpt of VoID

void:Dataset

void:partition

void:subset

void:target

void:Linkset
Describing the Ontology-Lexicon Interface

• Recognize the nature of a dataset with respect to the conceptual model of the ontology-lexicon interface defined by lemon.
• Provide each dataset with metadata specific to its nature.
Different Publication Options

void:Dataset

void:partition

void:target

void:Linkset

ontolex:Lexicon

lime:LexiconDataset

lime:LexicalizationSet

lime:referenceDataset

void:subset
Each component is published separately
Different Publication Options
A dataset published together with its lexicalization using an existing general-purpose lexicon.
Different Publication Options

void:Dataset

void:partition

void:target

void:Linkset

ontolex:Lexicon

lime:LexiconDataset

lime:referenceDataset

lime:LexicalizationSet

void:subset
Lexicalization set combined with a lexicon independently developed for a given reference dataset
Describing the Reference Dataset

- LIME introduces no specific term for this purpose
- Terms and conventions from VoID apply
- Reuse of existing metadata vocabularies

<http://xmlns.com/foaf/0.1/> a voaf:Vocabulary ;

  foaf:homepage <http://xmlns.com/foaf/0.1/> ;
  dct:title “The Friend of a Friend (FOAF) Vocabulary”@en ;
  void:dataDump <http://xmlns.com/foaf/spec/index.rdf> ;
  void:entities 75 ;
  voaf:classNumber 13 ;
  voaf:propertyNumber 62 .

18/06/2015
Manuel Fiorelli
fiorelli@info.uniroma2.it            art.uniroma2.it/fiorelli
Since a ontolex:Lexicon is a void:Dataset, then VoID terms do apply:

ex:myLexicon a void:Dataset ;

void:dataDump <http://example.org/lexicon/dump.rdf> ;
void:sparqlEndpoint <http://example.org/lexicon/sparql> ;
void:triples 10000 .
We can extend the base description with linguistic characteristics:

ex:myLexicon a ontolex:Lexicon ;


ontolex:language “en” ;

lime:lexicalEntries 13 ;

lime:linguisticModel <http://www.lexinfo.net/ontology/2.0/lexinfo> .
A lexicalization set provides the natural language realizations of the formal symbols defined in a reference dataset, possibly by means of an OntoLex lexicon.

```
ex:LexicalizationSet a lime:LexicalizationSet ;
    ontolex:language "en" ;
    lime:lexicalizationModel <http://www.w3.org/ns/lemon/ontolex#> ;
    lime:referenceDataset <http://xmlns.com/foaf/0.1/> ;
    lime:lexiconDataset ex:myLexicon .
```
We want to support also lexicalizations provided with different models than OntoLex, such RDFS and SKOS(-XL). 

```turtle
ex:embedLexSet a lime:LexicalizationSet ;
    ontolex:language "en" ;
    lime:lexicalizationModel <http://www.w3.org/2000/01/rdf-schema#> ;
```
• The supported language provides a coarse-grained description of what information is there
• Specific statistics provides us with a more specific insight on the nature of the available information.
lime: lexicalizations = |{(e, r) : e ontolex: denotes r}|

lime: references = |{r ∈ referenceDataset : ∃e ∈ lexiconDatset . e ontolex: denotes r}|

lime: lexicalEntries = |{e ∈ lexiconDataset : ∃r ∈ referenceDatset . e ontolex: denotes r}|

lime: percentage = \frac{lime: references}{|\{r ∈ referenceDataset\}|}

lime: avgNumOfLexicalizations = \frac{lime: lexicalizations}{|\{r ∈ referenceDataset\}|}

These definitions have been expressed in terms of the OntoLex lexicalization model
A lexicalization set can be related to a partition that provides lexicalizations for instances of a given class

The class can be either:

- A meta-class so that we can distinguish between lexicalizations of classes and properties
- A domain class so that we can distinguish between lexicalizations of different types of domain objects

```
A : lime:LexicalizationSet
lime:lexicalizations = 13
...
...

B : lime:LexicalizationSet
lime:resourceType = owl:Class
lime:lexicalizations = 13
...
...```

18/06/2015

Manuel Fiorelli
fiorelli@info.uniroma2.it  art.uniroma2.it/fiorelli
Describing wordnets (1/2)

A dataset defining Domain Concepts

void:Dataset ← lime:LexicalizationSet ← ontolex:Lexicon

ontolex:ConceptSet ← lime:Conceptualization ← ontolex:Lexicon

A dataset defining Lexical Concepts
Describing wordnets (2/2)
Lexical Linkset (1/2)

• One interesting use of wordnets is to enrich ontologies with links to lexical concepts.

• Links are represented via the property ontolex:evokes.

• The statistics about linksets are analogous to the ones associated to lexicalization set.
Lexical Linkset (2/2)

lime:referenceDataset

void:target

void:Dataset

void:partition

void:subset

lime:referenceDataset

void:Linkset

lime:LexicalLinkset

ontolex:ConceptSet

lime:conceptualDataset

ontolex:Conceptualization

ontolex:Lexicon

lime:LexicalizationSet

lime:lexiconDataset

lime:lexiconDataset

lime:conceptualDataset

lime:conceptualDataset
Conclusion

• Linguistic grounding of ontologies is valuable
• Open ecosystem of independently published resources
• Need for discovering relevant resources and assessing their suitability for given tasks
• LIME extension of the existing metadata vocabularies to support these exigencies


That’s all…

..thanks for your attention