INTERCONNECTING AND MANAGING MULTILINGUAL LEXICAL LINKED DATA

Ernesto William De Luca
Overview

- Motivation
- EuroWordNet
- RDF/OWL EuroWordNet
- RDF/OWL LexiRes Tool
- Conclusions
Overview

- Motivation
- EuroWordNet
- RDF/OWL EuroWordNet
- RDF/OWL LexiRes Tool
- Conclusions
Language Resources

- Contain a wide range of linguistic information
- Vary from simple lists to complex resources
  - lexical resources: word list, machine readable dictionary, thesauri, ontologies, glossaries, etc…
- Can be used for language and knowledge engineering

WordNet/EuroWordNet: Synonym Sets (SynSets)

- Each representing one constitutional lexicalized concept

EuroWordNet format

- defined by the EuroWordNet Database Editor Polaris that uses a proprietary specification.
Motivation
Related Work

- **VisDic**
  - Browsing and editing multilingual EuroWordnet information
  - BUT, here users can browse static information on text blocks.

- **MultiSemCor Web Interface**
  - Bilingual information browsing
  - MultiWordNet-annotated Parallel corpus
    - Browsing of words with their related annotated word senses,
    - BUT, the corpus is very restricted.
    - All accessible information is static.
    - Bilingual search in a closed domain.
Overview

- Motivation
- EuroWordNet
- RDF/OWL EuroWordNet
- RDF/OWL LexiRes Tool
- Conclusions
General (WordNet) Problems

- “Expressivity lack” (concepts vs. instances)
  - Solution: Named Entity Recognition Approaches

- Too fine grained distinction of word senses
  - Solution: Redesign/Merging concepts/SynSets.

- Not all word senses are covered
  - Solution: WordNet (and OWL) extension
Overview

- Motivation
- EuroWordNet
- RDF/OWL EuroWordNet
- RDF/OWL LexiRes Tool
- Conclusions
EuroWordNet and RDF/OWL WordNet

- Conversion of EuroWordNet into RDF/OWL
  - To have a standardized OWL format.
  - To access it with standard OWL query tools.
  - To enrich it with additional domain-specific ontologies.

- WordNet already been converted into an OWL format from (van Assem et al., 2004) using the OWL-DL sublanguage.

WordNet (1/3) - Main classes

- Word
- Synset
  - NounSynset
  - VerbSynset
  - AdjectiveSynset
    - AdjectiveSatelliteSynset
  - AdverbSynset
- WordSense
  - NounWordSense
  - VerbWordSense
  - AdjectiveWordSense
    - AdjectiveSatelliteWordSense
  - AdverbWordSense
- Collocation
Three kinds of properties in the schema:

- Connection of instances of the main classes together.
  - SynSet linked to WordSenses (property containsWordSense)
  - WordSense linked to Word (property word)

- Connection of WordNet relations. Three kinds:
  - between two SynSets (e.g. hyponymOf)
  - between two word senses (e.g. antonymOf)
  - Miscellaneous sets (e.g. gloss)

- Information about entities
  - XML Schema datatypes (e.g. xsd:string).
Use of URIs to provide some information about the entity meaning, built with pattern similar to:

wn20instances: + synset- + lexical form- + type- + sense number.

Example Code

```xml
<wn20schema:NounSynset rdf:about="&wn20instances;synset-bank-noun-4" rdfs:label="bank">
    <wn20schema:synsetId>102690337</wn20schema:synsetId>
</wn20schema:NounSynset>
```
Adaptation to the RDF/OWL Schema of WordNet (van Assem et al., 2004) and extension with new relations.

These steps can be subdivided into:

- Analysis of the requirements for EuroWordNet
- Adaptation of WordNet RDF-Schema to EuroWordNet
- Multilinguality
- OWL Property Conversion
- OWL Domain Extension
To avoid redundancy:

- Only relations in transitive direction are listed (e.g. hyponymOf and not hypernymOf)
- Others can be retrieved with the `owl:inverseOf` property implemented in the RDF Schema.

Instances of all classes and properties (separated in several data files)

- one for the SynSets
- one for the WordSenses and Words
- one for each relation
van Assem et al. (2004)

- focused on staying close to the original source (i.e. reflect the original data model without interpretation)
- direct use in Semantic Web applications, or as a source for modified WordNet versions
- Agreement with their assumptions also for a multilingual task.
RDF/OWL
EuroWordNet - Examples

```xml
<ewn20schema:NounSynset rdf:about="&ewn20instances;synset-bank-noun-1"
    rdfs:label="bank">
  <ewn20schema:synsetId>102690337</ewn20schema:synsetId>
</ewn20schema:NounSynset>
<ewn20schema:Word rdf:about="&ewn20instances;word-bank"
    ewn20schema:lexicalForm="bank"/>
<ewn20schema:NounWordSense rdf:about="&ewn20instances;wordsense-bank-noun-1"
    rdfs:label="bank">
  <ewn20schema:word rdf:resource="&ewn20instances;word-bank"/>
</ewn20schema:NounWordSense>
<rdf:Description rdf:about="&ewn20instances;synset-bank-noun-1">
  <ewn20schema:containsWordSense rdf:resource="&ewn20instances;wordsense-bank-noun-1"/>
  <ewn20schema:containsWordSense rdf:resource="&ewn20instances;wordsense-bank_building-noun-1"/>
</rdf:Description>
```

Eurowordnet-english-synset.rdf

```
<ewn20schema:hyponymOf rdf:resource="&ewn20instances;synset-deposit-noun-1"/>
</rdf:Description>
```

Eurowordnet-english-hyponymOf.rdf
WordNet already converted into OWL [van Assem et al., 2004] using the OWL-DL sublanguage.

Conversion of EuroWordNet into RDF/OWL
- To have a standardized OWL format.
- To access it with standard OWL query tools.
- To enrich it with additional domain-specific ontologies.

Extensions
- Adding domain ontologies (Pizza and Travel, EN) [GLDV2007]
- Adding the Hamburg Metaphor Database (DE, FR) [LREC2008]
- Adding the Basic Multilingual Dictionary MEMODATA (37655 entries, EN, DE, FR, ES, IT) [IGI-Book2012]
RDF/OWL
Lexical Linked Data Cloud

Hamburg Methaphor DB
Eurowordnet
Pizza
Travel

Basic Multilingual Dictionary Memodata
But how can we manage, interconnect and navigate this Lexical Linked Data Cloud?
Overview

- Motivation
- EuroWordNet
- RDF/OWL EuroWordNet
- RDF/OWL LexiRes Tool
- Conclusions
Navigation of ontology hierarchy for:
- restructuring it, by manual merging
- using automatic merging functions
- organizing the ontology/lexical resource structure
- visualizing the ontology hierarchy

Support the author in:
- Using EuroWordNet
- Exploring the lexical resource ontology hierarchy
- Disambiguating the word senses of the search word
- Giving translations of search word in different languages
- Creating individual lexical collections
- Adding and deleting meanings
- Merging meanings
- Importing OWL ontologies
RDF/OWL
LexiRes Tool - Browsing and Merging Word Senses
Overview

- Motivation
- EuroWordNet
- RDF/OWL EuroWordNet
- RDF/OWL LexiRes Tool
- Conclusions
Conclusions

RDF/OWL EuroWordNet

- Need of tools supporting the user in creating data
  - RDF/OWL EuroWordNet Browsing
  - Maintainance of Multilingual Resources
  - Customization of Multilingual Resources
  - Addition / deletion of word senses
  - Addition domain-specific ontologies
    - directly under its new hyperonym
  - Manual and Automatic Merging methods

- Discussion about licenses of LR
Conclusions

The RDF/OWL LexiRes Tool

- RDF/OWL EuroWordNet Browsing
- Maintainance of Multilingual Resources
- Customization of Multilingual Resources
- Addition / deletion of word senses
- Addition domain-specific ontologies
  - directly under its new hyperonym
- Manual and Automatic Merging methods
Conclusions
Current and Future Work
Conclusions
Current and Future Work