

Mapping Multilingual Lexical Material on the Ontology in LT4eL Project

Kiril Simov - Bulgarian Academy of Sciences
Cristina Vertan, University of Hamburg

Contents

- The Project LT4eL
- Role of the Ontology
- Mapping lexical material on ontology
- Using ontology and lexical material

LT4eL - Language Technology for eLearning

- EU-IST-FP6 Project 2005 - 2008
- The LT4eL project uses multilingual language technology tools and semantic web techniques for improving the retrieval of learning material. The developed technology will facilitate personalized access to knowledge within learning management systems and support decentralisation and co-operation in content management.

LT4eL - Partners

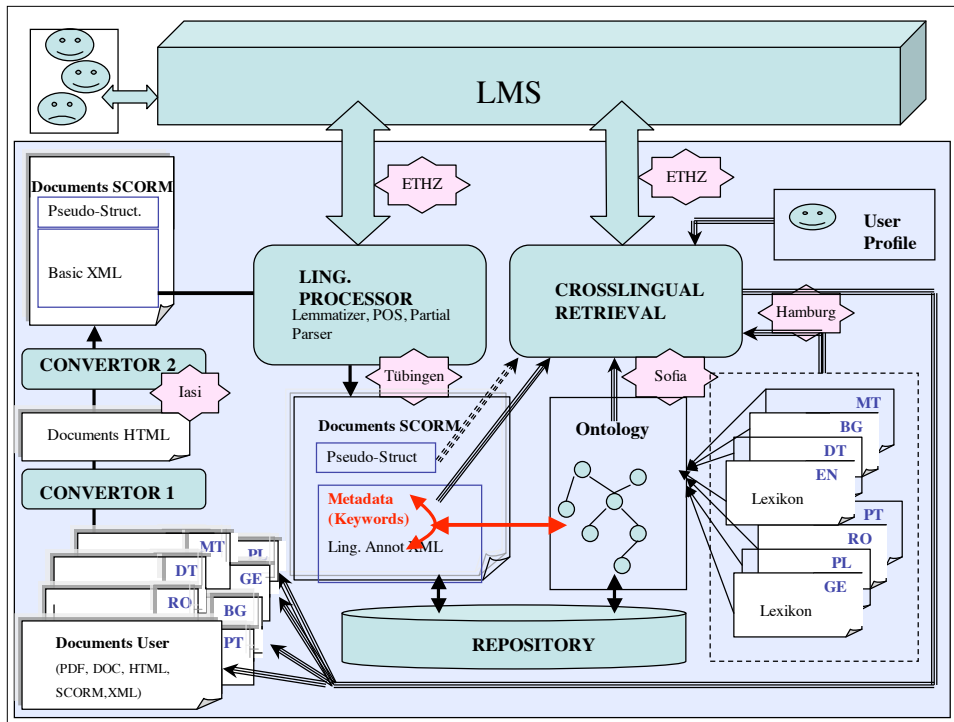
- Utrecht University (UU), The Netherlands
- University of Hamburg (UHH), Germany
- University "Al.I.Cuza" of Iasi (UAIC), Romania
- University of Lisbon (FFCUL), Portugal
- Charles University Prague (CUP), Czech Republic
- Institute for Parallel Processing, Bulgarian Academy of Sciences (IPP-BAS), Bulgaria
- University of Tübingen (UTU), Germany
- Institute of Computer Science, Polish Academy of Sciences (ICS-PAS), Poland
- Zürich University of Applied Sciences Winterthur (ZHW), Switzerland
- University of Malta (UOM), Malta
- University of Cologne (UCO), Germany
- Open University (OU), United Kingdom

LT4eL - Languages

- Bulgarian
- Czech
- Dutch
- German
- Maltese
- Polish
- Portugese
- Romanian
- English

LT4eL - Workpackages

- [WP1](#) - *Setting the scene* - WP leader: University Al. I. Cuza of Iasi
- [WP2](#) - *Semi-automatic metadata generation driven by Language Technology resources* - WP leader: University of Tübingen
- [WP3](#) - *Enhancing eLearning with semantic knowledge* - WP leader: IPP, Bulgarian Academy of Science
- [WP4](#) - *Integration of the new functionalities in the ILIAS Learning Management System* - WP leader: University of Cologne
- [WP5](#) - *Validation of new functionalities in the ILIAS Learning Management System* - WP leader: Open University (England)
- [WP6](#) - *Dissemination, transfer and broad take-up of programme results* - WP Leader: Utrecht University
- [WP7](#) - *Project management* - WP Leader: Utrecht University
- [Multilinguality](#) - Leader: University Hamburg



Ontologies and Lexicons in the LT4eL Project

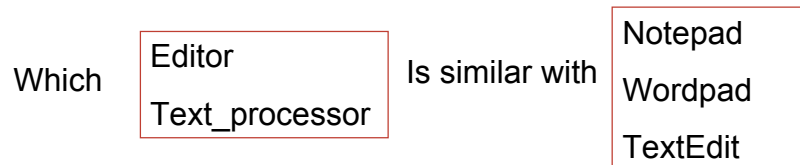
Scenario -1-

- The ontology will be used by
 - Students in order to retrieve LOs which:
 - are available on the eLearning platform and
 - are annotated following the LT4eL approach.
 - Teachers in order to prepare LOs, which means:
 - Extract keywords
 - Link LO's to the ontology
- Materials relevant for a certain topic can be retrieved in a number of languages specified by the user
- The ontology will be visible for the teachers but not for the students
- The retrieved materials will be classified according to the corresponding concepts and their relations in the ontology
- Relations: it would be useful to give some indications which notion depends (has to be studied after) on what: i.e. introduce such relations in the ontology

Generic approach

- For each domain :
 - Using computers for beginners
 - Impact of eLearning in Society
 - we built a domain ontology
 - For consistency reasons we consider also an upper ontology
 - Lexical material in all 9 languages is mapped on the ontology and on the upper ontology
 - According to :
 - types of relations in the ontology and
 - Uses cases
 - Similarity (predefined ontological chunks)
- we define some search patterns for the user interface

Search Patterns



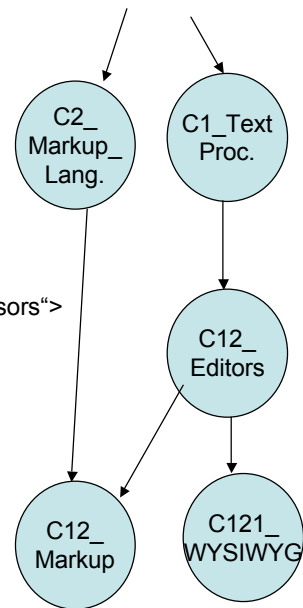
Similarity defines as sum of movements inside the ontology

Domain Ontology

- First built starting with English documents
- Concepts are based on :
 - Extracted keywords in WP2 and
 - Glossaries for the given domains
- Concepts have generic names with parts in english (for readability reasons) e.g: C11_editors
- For each concept we provide labels with explanation of the concept in english and ideally in all other languages
- Types of relations:
 - Is_a
 - Part_of
 - Here we need some informations about what people are searching
- The ontology will be encoded in OWL- DL

OWL encoding - Hierarchy definition

- `<owl:Class rdf:ID=„C1_Text_processors“>`
- `<owl:Class rdf:ID=„C2_Markup_Languages“>`
- `<owl:Class rdf:ID=„C12_Editors“>`
 `<rdfs:subclassOf rdf:resource=„#C1_Text_processors“>`
`</owl:Class>`
- `<owl:Class rdf:ID=„C121_WYSIWYG“>`
 `<rdfs:subclassOf rdf:resource=„#C12_Editors“>`
`</owl:Class>`
- `<owl:Class rdf:ID=„C121_Markup“>`
 `<rdfs:subclassOf rdf:resource=„#C121_Editors“>`
 `<rdfs:subclassOf`
 `rdf:resource=„#C2_Markup_Languages“>`
`</owl:Class>`



Lexical resources -1-

- For each language there are several existent lexicons (SIMPLE, WordNet-like, particular formats)
- Problems:
 1. different encoding formats (problem for the retrieval engine)
 2. Different levels of specification of morphological information. Best available resource for almost all languages is WordNet but there is no morphological information
 3. Some domain specific words do not exist or some domain specific senses are not present
 4. Semantic lexicons (especially WordNet) introduce to high polysemy which we do not need but induce problems for the mapping on the ontology.

Example of non desired polisemy

- In Wordnet one may find word „cell“ with 3 or 4 senses but none referring to a cell in the sense of a Table processor !

Lexical resources -2-

- Solution to 1:
 1. Look at all the formats and decide a common standard format including only the information we need. Full-form to avoid problems with tools.
 2. Each language provider converts his resources to this format or
 3. We use the MANAGELX Tool (Uni Hamburg) to convert. This would be preferable because of 2 reasons:
 1. We ensure also the conversion of the linguistic structure
 2. We consider offer a long term solution for the inclusion of further lexicons
- Solution to 2 :
 1. For the words which do not have morphological information provide first empty templates and then add the morphological part manually or
 2. Insert lemma in the morpho-syntactic part of the entries in the full-form lexicon.

Lexical resources -3-

- Solution to 3 :
 1. Manual insertion
- Alternative solutions to 4:
 - Map the not local useful senses to the upper ontology
 - Leave these senses unmapped at all or
 - Before mapping reduce the polysemy according to the approach described in Buitelaar P., "Ontology-based Semantic Lexicons: Mapping between Terms and Object Descriptions":
 - Domain specific tuning & determine domain -specific relevance of WordNet synsets (senses) on the basis of the relevance of their constituent synonyms that co-occur within domain specific corpora
 - We preserve only the senses having a domain relevance over a certain threshold.

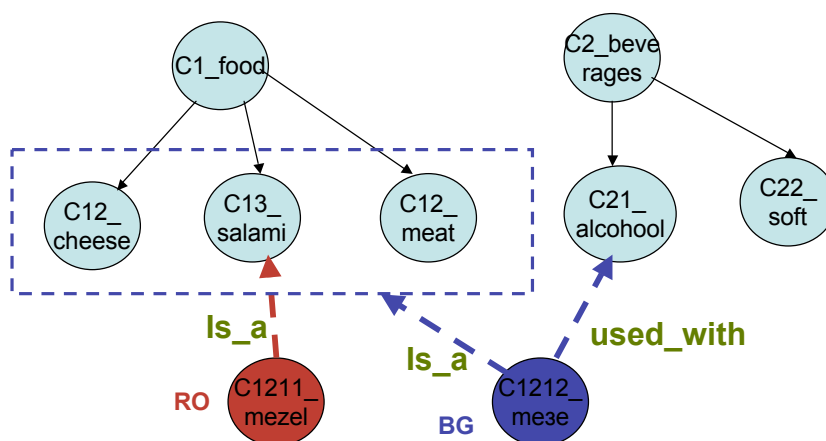
Mapping lexical resources on the domain ontology

- We will concentrate on nouns and verbs
- The meaning Common nouns and verbs are classes
- NEs are mainly instances
- Why "mainly" :
 - Following the above criteria "VW" is an instance but
 - "I changed my VW" refers at VW as car and
 - "VW are very good and un-expensive cars" refers to VW as Company and
 - We cannot define 2 individuals with the same name
 - Is CAPSLOCK a particular key or defines any combination with which one may write capital Letters ?

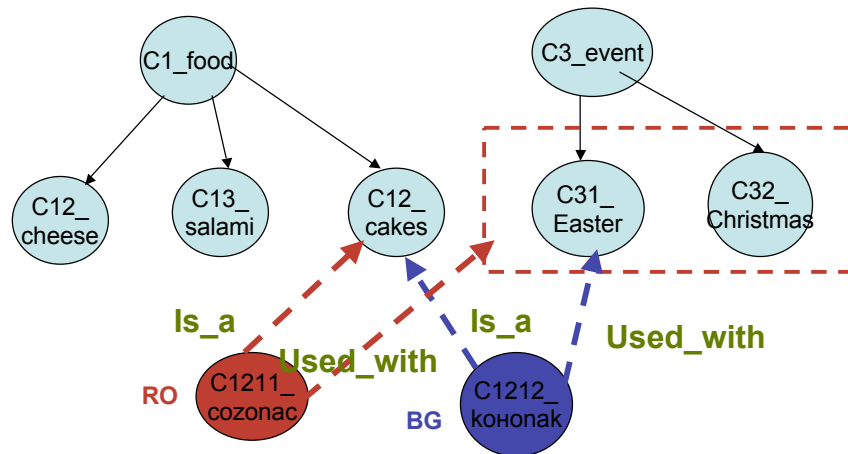
Mapping multilingual resources on the domain ontology -1-

- Trivial for words having exact a correspondent in the ontology
- Problems appear when:
 1. One word in a language sub-sums two or more concepts in the ontology
 2. One word in a language sub-sums two or more concepts in an ontology but only in relations with some other concepts
 3. One word has a much restrictive meaning not present in the ontology

Example of different / new meanings induced by lexical terms -1-



Example of different / new meanings induced by lexical terms -2-



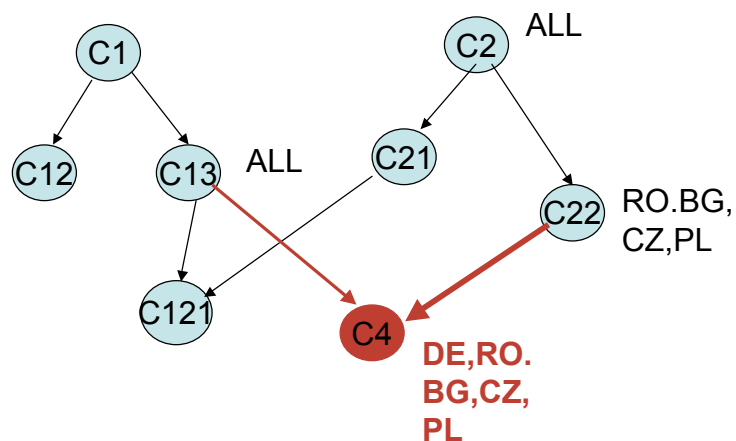
Mapping multilingual resources on the domain ontology -2-

- Solution to 1:
 - Express the lexical items in OWL-DL expressions: disjunction, conjunctions of classes (give example)
- Solution to 2:
 - Express the lexical items in OWL-DL using together with operations on classes also relations between the involved concepts
- Solution to 3:
 - Insert new concept in the ontology

Ontology enrichment

- If one word cannot be mapped directly on the ontology look if a similar meaning can be retrieved in some other languages.
- If this seems to be not an isolated case insert the new concept in the ontology.
- In any case assign to each concept a label indicated the languages in which this concept is lexicalised
- The insertion of a new concept will be done with FACT or RACER

Ontology Enrichment -Example



Linking lexicon, domain ontology and upper ontology

- Domain ontology concepts will be mapped on the upper ontology.
- This will ensure that all important properties of main classes are considered.
- Not relevant senses of some lexical items could be also mapped directly on the upper ontology

Annotated LOs

- Each document will have attached a list of keywords mapped on the ontology
- Possible extension: Given the concepts in the ontology we could provide some RDF statements, at least for showing some interconnections between Los (see EU-Project METOKIS). We could do this for demonstration purposes at least for 2-3 languages.

User's view on the ontologies

- The user (teacher) will view:
 - The list of lexical terms
 - A part of the ontology containing concepts lexicalised for the language of the document

Steps

1. Ontology development according to the English documents and retrieval engine V1.
2. Mapping of lexical items from:
 1. Bulgarian
 2. Dutch
 3. German
 4. Romanianand retrieval engine v2
1. Mapping of lexicons from other languages: PL, PT, CZ, MT. and retrieval engine v3