Applying Ontology-Based Lexicons to the Semantic Annotation of Learning Objects

Kiril Simov, Petya Osenova
Linguistic Modelling Laboratory, IPOI,
Bulgarian Academy of Sciences
Acad. G.Bonchev 25
Sofia 1113, Bulgaria
{kivs|petya}@bultreebank.org

Abstract

This paper discusses the role of the ontology in the definition of domain lexicons in several languages and its usage for the semantic annotation of Learning Objects (LOs). We assume that the ontology has the leading role and the lexicons are created on the basis of the meanings defined within the ontology. The semantic annotation requires the construction of special partial grammars connected to the terms in the lexicons. These special grammars are used for automatic annotation of domain texts. The ambiguous cases are resolved manually on the base of the context. The process of semantic annotation plays a twofold role: first, it produces semantically annotated texts (gold standard corpus), and second, it helps in checking the coverage of the lexicon as well as the precision of the ontology.

Keywords

Ontology, Ontology-based lexicon, Semantic annotation.

1. Introduction

LT4eL European project[1] [1] aims at demonstrating the relevance of the language technology and ontology document annotation for improving the usability of learning management systems (LMS). This paper discusses the role of the ontology in the definition of domain lexicons in several languages and its usage for the semantic annotation of Learning Objects (LOs). The relation between the domain ontology and the domain texts (the learning objects) is mediated by two kinds of information (or layers of information as they are called in [2]) – domain lexicons and concept annotation grammars. In our model the lexicons are based on the ontology. We assume that the ontology is defined first in a formal way and then the lexicons are built on the basis of the concepts and relations defined in the ontology. The terms in the lexicons are mapped to grammar rules for partial analyses of texts. These rules constitute annotation grammars for recognizing the ontological concepts in the texts. The last component of a complete set of knowledge sources for semantic annotation – the disambiguation rules – is not implemented within the project. For the experiments within the project and for the creation of gold standard for concept annotation we have disambiguated the LOs manually.

The structure of the paper is as follows: in the next section we give a short overview of ontology creation process in the project; then we present in detail the lexicon model that we exploit within the project; section 4 discusses the annotation of learning objects with concepts; section 5 outlines the main problems of the current annotation process; the last section concludes the paper and points to the future work.

2. LT4eL Domain Ontology

The domain of LT4el Project is "Computer Science for Non-Computer Scientists". It covers topics like operating systems; programs; document preparation – creation, formatting, saving, printing; Web, Internet, computer networks; HTML, websites, HTML documents; email, etc. The main application of the ontology has to do with the indexing of LOs within the domain.

The creation of the ontology was done on the basis of manually annotated keywords in the eight languages of the project (Bulgarian, Czech, Dutch, German, Maltese, Polish, Portuguese, Romanian). The annotated keywords were translated into English. Then by search on the Web we collected definitions for the keywords. The set of definitions of a keyword highlights the various meanings of the keyword and the relations between its meanings and other concepts. After the determination of the keywords meanings we created concepts corresponding to these meanings.

These concepts constitute the backbone of the domain ontology. The next step of the ontology development was to map the domain concepts to an upper ontology (in our case we used DOLCE – [3], [4]) in order to inherit some knowledge already encoded in the upper ontology (relations, for instance) and to ensure right concept classification with respect to concept metaproperties defined in the ontology creation methodology – OntoClean [5]. The mapping to the upper ontology was done via

[1] http://www.lt4el.eu/ – the LT4eL (Language Technology for eLearning) project is supported by the European Community under the Information Society and Media Directorate, Learning and Cultural Heritage Unit.

OntoWordNet [6] – a version of WordNet restructured in accordance to DOLCE.[2]

The ontology was extended with additional concepts taken from: the restriction on already existing concepts (for example, if a *program* has a *creator*, the concept for *program creator* is also added to the ontology); superconcepts of existing concepts (if the concept for *text editor* is in the ontology, then we added also the concept of *editor* (as a kind of program) to the ontology); missing subconcepts (if *left margin* and *right margin* are represented as concepts in the ontology, then we add also concepts for *top margin* and *bottom margin*); from the annotation of the learning objects (if a concept is represented in the text of a learning object and it is relevant for the search within the learning material, we add the concept to the ontology). After having applied these steps we built a domain ontology (the current version) with about 750 domain concepts, about 50 concepts from DOLCE and about 250 intermediate concepts from OntoWordNet. We also have about 200 new concept candidates to be added to the ontology. They are extracted from LOs in the different languages on the basis of their ontological annotation.

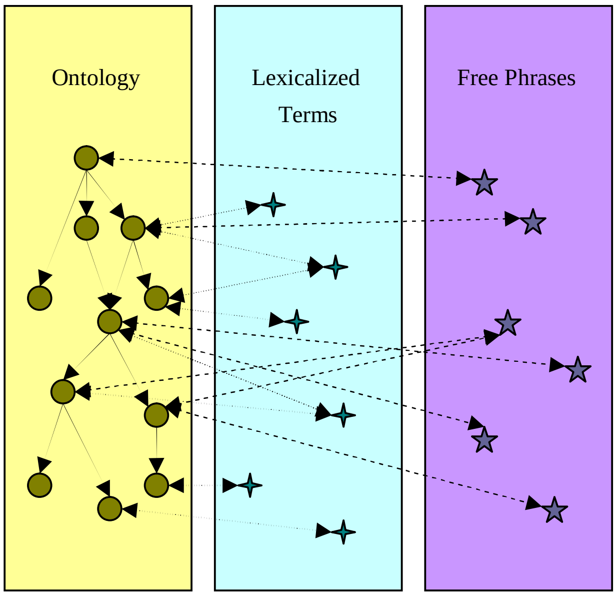
Although we have started with keywords annotated within the LOs, they are not enough to connect the concepts in the ontology with their explications within the texts. The incompleteness is due to the following facts: (1) not all keywords are shared by the annotations in all languages; (2) some concepts from the extension of the first set of concepts (created on the basis of the keywords) appear in the texts, but were not annotated as keywords. Thus, in order to connect the concepts within the ontology (later also the relations) to the text of the LOs we need also a lexicon for each language aligned to the ontology; a mechanism for recognition of the lexical items in the texts; and a mechanism for selection of the right concept for a phrase in the text when the phrase is ambiguous with respect to several concepts.

3. Ontology-Based Lexicon Model

In this section we present the model and the creation of the lexicons for each project language on the basis of the existing ontology. The lexicons represent the main interface between the user's query, the ontology and the ontological search engine. The annotation of the learning object is facilitated by an additional language tool – annotation grammars for concepts which will be discussed later.

There exist various attempts to approach this mapping task. Most of them start from lexicon compilation for different languages, and then try to establish the connection to the concept. Such initiatives were WordNet [7], EuroWordNet [8], SIMPLE [9]. In spite of the fact that we employ the experience from these projects (mapping to WordNet and Pustejovsky's ideas in SIMPLE), we also suggest an alternative in connecting the ontology and the lexicons. Our model is very close to LingInfo model (see [11]) with respect to the mapping of the lexical items to concepts, but also with respect to the other language processing tools we connect to the ontology – the concept annotation grammars and concept disambiguation tools. We will discuss LingInfo below.

The terminological lexicons were constructed on the basis of the formal definitions of the concepts within the ontology. By this approach of construction of the terminological lexicon we escaped the hard task of mapping different lexicons in several languages as it was done in EuroWordNet Project [8]. The main problems with this approach are that (1) for some concepts there is no lexicalized term in a given language, and (2) some important term in a given language has no appropriate concept in the ontology which to represent its meaning. In order to solve the first problem we allow the lexicons to contain also non-lexicalized phrases which have the meaning of the concepts without lexicalization in a given language. Even more, we encourage the lexicon builders to add more terms and phrases to the lexicons for a given concept in order to represent as many ways of expressing the concept in the language as possible. These different phrases or terms for a given concept are used as a basis for construction of the regular grammar rules for annotation of the concept in the text. Having them, we could capture in the text different wordings of the same meaning. The following picture shows the mapping varieties:



The picture depicts the realization of the ontological concepts in a natural language. The concepts are language independent and they might be represented within a natural

[2] We also a mapped the domain concepts to WordNet 2.0 and in this way we ensured a mapping to SUMO [10].

language as form(s) of a lexicalized term (or item), or as a free phrase. In general, a concept might have a few terms connected to it and a (potentially) unlimited number of free phrases expressing this concept in the language. Some of the free phrases receive their meaning compositionally regardless their usage in the text, other free phrases denote the corresponding concept only in a particular context. In our lexicons we decided to register as many free phrases as possible in order to have better recall on the semantic annotation task. In case of a concept that is not-lexicalized in a given language we require at least one free phrase to be provided for this concept.

In order to solve the second problem we modify the ontology in such a way that it contains all the important concepts for the domain. However, this solution requires a special treatment of the "head words" in the lexicons, because such phrases allow bigger freedom with respect to their occurrences in the text. Variability is a problem even with respect to the lexicalized cases and our idea is to represent the most frequent (based on the learning objects we already processed) variants for each concept. We are not able to solve this problem in general within the project, but we hope to demonstrate some approaches to it.

The specific solutions for the lexical terms without appropriate concept in the ontology are the following:

More detailed classes in the ontology. In cases where it was possible, we created more specific concepts in the ontology. For example, the concept of *shortcut*, as it was initially defined, was the general one, but the lexical items in English to some extent depend on the operating system, because each operating system (MS Windows, Linux, etc) as a rule introduces its own terminology. When the notion is borrowed in other languages, it could be borrowed with different granularity, thus, we introduce more specific concepts in the ontology in order to ensure correct mapping between languages.

More complex mapping between the ontology and some lexicons. Our initial idea was that each meaning of a lexical item in any language is mapped to exactly one concept in the ontology. If for some lexical item this one-to-one mapping is not appropriate or it requires very complicated changes in the ontology, we realize a mapping based on OWL expressions. This mechanism allows us to keep the ontology simpler and more understandable, and to handle cases that do not allow appropriate mappings. Currently, such cases are not detected in our domain.

We could summarize the connection between the ontology and the lexicons in the following way: the ontology represents the semantic knowledge in form of concepts and relations with appropriate axioms; and the lexicons represent the ways in which these concepts can be realized in texts in the corresponding languages. Of course, the ways in which a concept could be represented in the text are potentially infinite in number, thus, we could hope to represent in our lexicons only the most frequent and important terms and phrases. Here is an example of an entry from the Dutch lexicon:

```
<entry id="id60">
   <owl:Class rdf:about="lt4el:BarWithButtons">
     <rdfs:subClassOf>
       <owl:Class rdf:about="lt4el:Window"/>
     </rdfs:subClassOf>
   </owl:Class>
   <def>A horizontal or vertical bar as a part of a window,
      that contains buttons, icons.</def>
   <termg lang="nl">
     <term shead="1">werkbalk</term>
     <term>balk</term>
     <term type="nonlex">balk met knoppen</term>
     <term>menubalk</term>
   </termg>
 </entry>
```

Each entry of the lexicons contains three types of information: (1) information about the concept from the ontology which represents the meaning for the terms in the entry; (2) explanation of the concept meaning in English; and (3) a set of terms in a given language that have the meaning expressed by the concept. The concept part of the entry provides minimum information for formal definition of the concept. The English explanation of the concept meaning facilitates the human understanding. The set of terms stands for different wordings of the concept in the corresponding language. One of the terms is the representative for the term set. Note that this is a somewhat arbitrary decision, which might depend on frequency of term usage or specialist’s intuition. This representative term will be used where just one of terms from the set is necessary to be used, for example as an item of a menu. In the example above we present the set of Dutch terms for the concept *lt4el:BarWithButtons*. One of the term is non-lexicalized - attribute *type* with value *nonlex*. The first term is representative for the term set and it is marked-up with attribute *shead* with value 1.

Here we present a (part of) DTD for the lexicon:

```
<!ELEMENT LT4ELLex (entry+)>
<!ELEMENT entry
   ((owl:Class|rdf:Description|rdf:Property), def, termg+)>

<!ELEMENT def (#PCDATA)>
<!ELEMENT termg (term+,def?)>
<!ATTLIST termg
          lang     (bg|cs|de|en|mt|nl|pl|pt|ro)  # REQUIRED
>
<!ELEMENT term (#PCDATA)>
<!ATTLIST term
          type     (lex|nonlex)      "lex"
          shead    (1|0)             "0"
          gram     CDATA             #IMPLIED
>
```

The lexicon consists of entries. Each *entry* consists of a *concept*, *relation* or *instance* (partial) definition, followed by a *definition* of the concept content in English and one or several *term groups*. Each *term group* represents all the available lexical terms or free phrases for the corresponding concept (relation or instance) in a given natural language (determined by the attribute *lang*). Optionally, the term group for a given language could contain a definition of the content of the concept in that language. Each *term* represents a normalized form of the term. Additionally, we could state whether: the term is a lexicalization of the concept in the language or it is a free phrase (attribute *type*); the term is representative for the concept in the language (the attribute *shead*) or not; and which grammar rules recognize this term (related to the concept (relation or instance) of the entry) in text. The format of the currently implemented grammars is given below.

As it was mentioned above, we envisage two more language tools to help in the annotation of text with ontology information, namely – concept annotation grammars and (sense) disambiguation rules. The first kind of information could be seen as a special kind of partial parsing tool which for each term in the lexicon contains at least one grammar rule for recognition of the term. The second kind of information in our project is still in a very preliminary state of development and it is not discussed here.

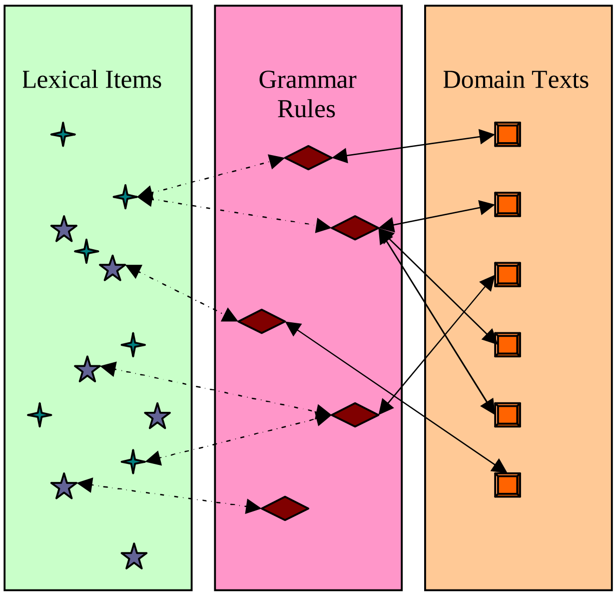
For the implementation of the annotation grammar we rely on the grammar facilities of the CLaRK System[3]. The structure of each grammar rule in CLaRK is defined by the following DTD fragment:

<!ELEMENT *line* (*LC*?, *RE*, *RC*?, *RM*, *Comment*?) >
<!ELEMENT *LC* (#PCDATA)>
<!ELEMENT *RC* (#PCDATA)>
<!ELEMENT *RE* (#PCDATA)>
<!ELEMENT *RM* (#PCDATA)>
<!ELEMENT *Comment* (#PCDATA)>

Each rule is represented as a line element. The rule consists of regular expression (RE) and category (RM = return markup). The regular expression is evaluated over the content of a given XML element and could recognize tokens and/or annotated data. The return markup is represented as an XML fragment which is substituted for the recognized part of the content of the element. Additionally, the user could use regular expressions to restrict the context in which the regular expression is evaluated successfully. The *LC* element contains a regular expression for the left context and the *RC* for the right one. The element *Comment* is for human use. The application of the grammar is governed by Xpath expressions which provide additional mechanism for accurate annotation of a given XML document. Thus, the CLaRK grammar is a good choice for implementation of the initial annotation grammar.

The creation of the actual annotation grammars started with the terms in the lexicons for the corresponding languages. Each term was lemmatized and the lemmatized form of the term was converted into regular expression of grammar rules. Each concept related to the term is stored in the return markup of the corresponding rule. Thus, if a term is ambiguous, then the corresponding rule in the grammar contains reference to all concepts related to the term.

The following picture depicts the relations between lexical items, grammar rules and the text:



The relations between the different elements of the models are as follows. A lexical item could have more than one grammar rule associated to it depending on the word order and the grammatical realization of the lexical item. Two lexical items could share a grammar rule if they have the same wording, but they are connected to different concepts in the ontology. Each grammar rule could recognize zero or several text chunks.

In the next two sections we present the process of annotation of LOs with the grammars constructed in the way, explained above. Also, we discuss the problematic cases. Because of lack of any disambiguation rules at this stage of our work we did the disambiguation manually (see Figure 1 at the appendix).

4. Semantic Annotation of Learning Objects

From the perspective of the Learning Management System, the semantic (ontological) annotation concerns only the metadata section of the learning objects. In the metadata,

[3] http://www.bultreebank.org/clark/index.html

according the Learning Object Metadata [12] standard, some ontological information can be stored and used later on to index the learning objects for retrieval. The annotation needs not be anchored to the content of the learning object. The annotator of the learning object can include in the annotation all concepts and relations he/she considers to be important for the classification of the learning object. However, in order to accurately link a learning object and/or its parts to the proper places in the conceptual space of the ontology, the *inline annotation* of the content of learning objects becomes an obligatory intermediate step in the meta-annotation of the learning objects with ontological information. The *inline annotation* is done by regular grammar rules attached to each concept in the ontology reflecting the realizations of the concept in texts of the corresponding languages (as it was explained in the previous section). Additionally, rules for disambiguation between several concepts are applied when a text realization is ambiguous between several concepts. Recall that at the current stage of the project we do not have a great progress on disambiguation rules.

Within the project we performed both types of annotation, *inline* and *through metadata*. The metadata annotation is used during the retrieval of learning objects from the repository. The *inline annotation* will be used in the following ways: (1) as a step to metadata annotation of the learning objects; (2) as a mechanism to validate the coverage of the ontology; and (3) as an extension of the retrieval of learning objects where, except for the metadata, we could use also coocurrences of concepts within the whole LO or its subparts (paragraphs or sentences).

Let us consider in more detail the strategy behind the semantic annotation. The annotation was done through a version of CLaRK System that includes the appropriate DTDs, layouts, grammar and constraints. The process included the following phases: (1) preparation for the semantic annotation and (2) actual annotation. The former refers to the compilation of appropriate regular grammars that explicate the connection between the domain terms in some natural language and the ontological concepts. It also considers the construction of a DTD, layouts and support semi-automatic tools for assigning and disambiguating concepts, namely the constraints. The annotation phase envisages the execution of the above-mentioned tools. The regular grammar finds the occurrences of terms in the text and it assigns all the possible concepts per term. As it was explained in the previous section, the regular grammars were constructed automatically on the basis of the lemmatization of the terms in the lexicons. Thus, in some cases the grammar can under- or over-generate.

The constraints, on the other hand, aim at making the annotation process more accurate. The constraints support the manual annotation. They work in the following way – if there is no ambiguity, the unique concept is assigned as an annotation. If the term is ambiguous, then the constraint proposes to the annotator the possible options and he or she has to select the right choice. The annotator has two possibilities: using Constraint 1 (*Select Concept*) or using Constraint 2 (*Select LT4eL Concept*). The Constraint 1 stops at each recognized term despite being ambiguous or non-ambiguous one and suggests artificial ambiguity via the options ERASE and EXTENDED. The option ERASE is chosen when a concept was assigned to a common word, not term. The option EXTENDED is chosen when a concept is recognized partially. This option covers two basic cases: occurrence of general vs. specific notions (e.g. *Internet* vs. *Wireless Internet*), or notions that can be expressed by single word as well as multiwords (*disc* vs. *hard disc*; *user* vs. *end-user*). There is a third option, which is incorporated into both constraints, namely – adding a correction over a concept. This happens when the term is used in broader or a narrower sense, which lacks in the assigned concept (e.g. *Insert* concept in narrowing sense of the term *Paste*, and in the broader sense of the term *Insert*). The usage of the Constraint 1 is recommended at the beginning of the annotation process, when the annotation grammar is not considered to be very precise, and when its automatically compiled versions rely only on lemmas.

All these repairing techniques (although subjective and depending on the annotator) lead to the improvement of the regular grammar, which assigns the concepts.

The usage of Constraint 2 can be relied upon at later stage, when the grammar for a language has been improved at least to some extent (as a result from the previous constraint). This constraint does not introduce artificial ambiguity choices. It stops only at real ambiguities in the texts. For example, the term 'word' might be assigned two concepts depending on the context: either being common words (*WordLang*), or elements of computer memory (*WordMemory*).

We have annotated all the learning objects in all languages in this way. A reasonable part (at least one third) of all learning objects was annotated by using the first constraint. On the basis of the problems found during this annotation we have improved the annotation grammars for some of the languages (Bulgarian, Dutch, English, Portuguese) and have introduced some disambiguation rules for the same languages. In the next section, we present some of the more frequent problems with the current concept annotation.

5. Problematic Cases

In this section we discuss some typical problems of the semantic annotation process, and point out to the possible solutions.

The sources of the problems with assigning the appropriate concepts to the domain terms are of various kinds. We can sum them up in the following way: (1) some concepts in the ontology might be quite specific or rather broad with respect to the term which they were assigned to; (2) some general terms in a connected text refer to specific entities (in anaphoric relations); (3) the boundaries of the terms might need extension; (4) the ambiguity within concepts might be fake; (5) the detected term has also a common-use meaning and hence – in the context this common meaning is triggered; (6) do verbs receive semantic annotation.

Let us consider each of the above-mentioned issues in more detail.

(1) Some concepts in the ontology might be quite specific or rather broad with respect to the term. For example, the concept Size is defined as 'number of unique entries in a database'. However, it has broader sense. The same holds for the concept TableOfContents. Thus, an additional concept is suggested (Content), which covers the whole content of an information object, not just the reference to it.

(2) Some general terms in a connected text refer to specific entities (in anaphoric relations). For example, the term is Systems for Personalization, but further in the text they are called just Systems.

(3) The boundaries of the terms might need extension. For example, within the term Personalized system, only system is detected, and hence – the more general concept is assigned, namely – ComputerSystem. Extension of terms helps in adding more domain specific concepts to the ontology. The same holds for the term Agent technology, in which only technology was detected and tagged.

(4) The ambiguity within concepts might be fake. This problem contributes to the precision of the ontology, since the annotators detected ambiguous-like concepts which practically coincide, such as: Metadata and DescriptiveMetadata; Search and Searching. This issue arose due to several reasons: the naming of the concept is conventional, and thus language-independent. The real pointer is the definition. Consequently, sometimes several conventional namings compete, resulting in spuriousness.

(5) The detected term has also a common-use meaning. For example, Help as command and as request. Another example is the word Sector. In the sequence 'educational sector' it is not in its computer domain meaning. Also, point in 'point of view' is definitely not the format-oriented concept Bullet (at least in Bulgarian there is homonymy).

(6) Do verbs receive semantic annotation? This problem concerns preferably the commands and buttons. Button 'Save' and the command '(to) Save'. We decided not to treat the commands as verbs, and also not to assign real verbs semantic concepts. This was done for practical reasons.

To sum up, most of the problematic issues might be repaired via the annotator intervention. Later on, the ontology and lexicon builders could use these observations for extending the lexicons and making the ontology more precise. Needless to say, the specificities of the above cases vary with respect to the language. However, the problem abstraction is cross-lingual.

6. Conclusions

In this paper we have presented the current state of ontology, lexicons and ontology annotation within the LT4eL project. We have built a lexicon model which relates ontology concepts, relations and instances to lexical items, grammar rules and disambiguation rules. In this model the central role is played by the ontology which determines the content of the other components of the model. Another advantage of our model is that it supports the work in multilingual environment. The mappings between lexicons and ontology are performed with the aim each term in each language to have a corresponding concept, and vice versa – each concept in the ontology to be exemplified by at least one term expression (be it lexicalized or a free phrase). Thus, the ontology itself as well as specific language lexicons are verified in a cross-lingual context. We have created lexicons and annotation grammars for all the languages in the project. The mapping between the language specific lexicons was facilitated by the ontology. Our model shares common features with other lexicon models: with WordNet-like ([7], [8]) lexicons we share the idea of grouping lexical items around a common meaning and in this respect the term groups in our model correspond to synsets in WordNet model. The difference in our case is that the meaning is defined independently in the ontology. With SIMPLE model ([9]) we share the idea to define the meaning of lexical items by means of the ontology, but we differ in the selection of the ontology which in our case represents the domain of interest, and in the case of SIMPLE reflects the lexicon model. With the LingInfo model ([11]) we share the idea that grammatical and context information also needs to be presented in a connection to the ontology, but we differ in the implementation of the model and the degree of realization of the concrete language resources and tools.

In other approaches to semantic annotation usually information extraction systems are adapted. This choice is appropriate as much as more instance information is expected to be annotated within the domain text. Thus, in such cases recognition of new instances and their classification with respect to the ontology information is of great importance. In our application in addition to the instance information we also have to annotate mentionings of ontological information where the properties of the concepts are stated in general and not as on a particular instance. Thus, the term lexicons are a good way to create the necessary resources for implementing the model for many languages.

In future, more work has to be done on the extension of the annotation grammars, on the implementation of disambiguation rules, on the connection of the lexicon and the grammars to non-domain dependent concepts. Also, we need to add domain relations.

7. Acknowledgements

This work has been supported by LT4eL (Language Technology for eLearning) project of the European Community under the Information Society and Media Directorate, Learning and Cultural Heritage Unit.

We would like to thank our partners from the project for their useful feedback. Especially to Paola Monachesi, Cristina Vertan, Lothar Lemnitzer, Corina Forascu, Claudia Borg, Rosa Del Gaudio, Jantine Trapman, Beata Wójtowicz, Eelko Mossel.

8. Appendix

The following screenshot depicts manual disambiguation within CLaRK System. The grammars were used to annotate the texts which refer to some concepts. In case of ambiguities the user is asked to select the relevant concept from a list of possible ones.

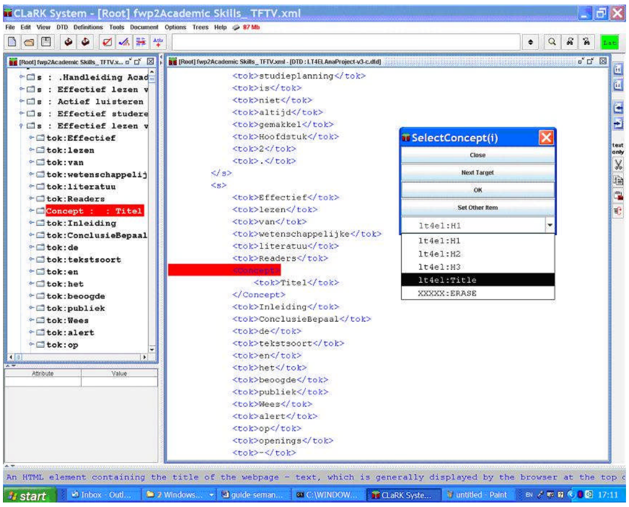


Figure 1: Here the concept *Title* is chosen as the correct one for the term *Titel* in Dutch.

9. References

[1] Paola Monachesi, Lothar Lemnitzer, Kiril Simov. Language Technology for eLearning. Proceedings of EC-TEL 2006, in Innovative Approaches for Learning and Knowledge Sharing, LNCS 0302-9743, pp. 667-672. 2006

[2] Paul Buitelaar, Michael Sintek, Malte Kiesel A Lexicon Model for Multilingual/Multimedia Ontologies In: Proceedings of the 3rd European Semantic Web Conference (ESWC06), Budva, Montenegro, June 2006.

[3] Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, Alessandro Oltramari and Luc Schneider. 2002. The WonderWeb Library of Foundational Ontologies. WonderWeb Deliverable D17, August 2002. http://www.loa-cnr.it/Publications.html.

[4] Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino and Alessandro Oltramari. 2002. Ontology Library (final). WonderWeb Deliverable D18, December 2003. http://www.loa-cnr.it/Publications.html.

[5] Nicola Guarino and Christopher Welty. 2002. "Evaluating Ontological Decisions with OntoClean." Communications of the ACM, 45(2): 61-65.

[6] Aldo Gangemi, Roberto Navigli, Paola Velardi. The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet. Meersman R, et al. (eds.), Proceedings of ODBASE03 Conference, Springer, 2003.

[7] Christiane Fellbaum. 1998. Editor. WORDNET: an electronic lexical database. MIT Press.

[8] Piek Vossen (ed). EuroWordNet General Document. Version 3, Final, July 19, 1999, http://www.hum.uva.nl/~ewn

[9] Alessandro Lenci, Federica Busa, Nilda Ruimy, Elisabetta Gola, Monica Monachini, Nicoletta Calzolari, Antonio Zampolli, Emilie Guimier, Gaëlle Recourcé, Lee Humphreys, Ursula Von Rekovsky, Antoine Ogonowski, Clare McCauley, Wim Peters, Ivonne Peters, Robert Gaizauskas, Marta Villegas. 2000. SIMPLE Work Package 2 - Linguistic Specifications, Deliverable D2.1. ILC-CNR, Pisa, Italy.

[10] Ian Niles and Adam Pease. Towards a standard upper ontology. In: FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems, ACM Press (2001)

[11] Paul Buitelaar, Thierry Declerck, Anette Frank, Stefania Racioppa, Malte Kiesel, Michael Sintek, Ralf Engel, Massimo Romanelli, Daniel Sonntag, Berenike Loos, Vanessa Micelli, Robert Porzel, Philipp Cimiano LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In: Proc. of OntoLex06, a Workshop at LREC, Genoa, Italy, May 2006.

[12] Standard for Learning Object Metadata, Final Draft, IEEE, 2002 -- P1484.12.1