Semantic Enrichment of Legal Ontologies with Linguistic Content

Sorina CORNOIU

Consiliul Legislativ, Bucureşti
sorina_c_2000@yahoo.com

Abstract. The amount of legal information is growing faster and faster. Archiving and access to legal documents has much improved. Yet, it is becoming more and more difficult to search and select the appropriate information. Legal ontologies provide formal models for representing domain knowledge. The huge legal information systems require semantic enrichment of legal ontologies with linguistic content. The Semantic Web offers the most appropriate scenario for exploiting ontologies' potentialities, due to the large amount of information which is to be exposed and accessed [2].
In this paper I propose (using Ontoling - a plugin for the popular ontology editing tool Protege) linguistic enrichment of legal ontologies, which has run through the identification of different categories of linguistic resources and planning their exploitation to augment the linguistic expressivity of legal ontologies.

Keywords: Ontology , Semantic Web , Knowledge Representation, Semantic Search, Natural Language Processing
Math. Subject Classification 2000: 03B99

1 Introduction

Semantic Web is characterized by huge quantities of documents and by users will to access them. Though machine readability is a primary aim for allowing automatic exchange of data, several SW services like Intelligent Q'still need to understand and expose knowledge expressed in the sole way humans can easily understand it: natural language [2]. These premises suggest that legal ontologies, should be linguistic enriched to identify a series of different processes sharing the common objective of augmenting the linguistic expressivity of an ontology through the exploitation of existing Linguistic Resources [1].
In this paper I propose (using Ontoling - a plugin for the popular ontology editing tool Protege) linguistic enrichment of legal ontologies, which has run through the identification of different categories of linguistic resources and planning their exploitation to augment the linguistic expressivity of legal ontologies.

2 Ontology for legal framework modelling(Lex-is)

We use the "Lex-is" [8], the Ontology for Legal Framework Modelling, an ontology that organise and structure legislative information with the objective to enable improve access in the legislative process. To facilitate this, Lex-is ontology capture and structure interrelated legal information about the legislative process, specifically :
- Structure, type and content of EU legislation (treaties, regulation, directives, decisions) organized in thematic areas along with their interrelations
- Structure, type and content of national legislation (constitutions, laws, presidential degrees, etc) along with their interrelation to EU legislation - for example legal acts that incorporate EU directives in national law .

The core entities of the Lex-is ontology are :
- Legal Elements - incorporate the EU directives, recommendations and directives, as well as national laws, decrees and constitutions
- Preparatory Acts - creating new or updating legal elements
-Legal Frameworks - embrace the EU and national legislation on particular thematic areas and context
- Legal Rules - represent the interpretation of the laws by law experts or as adopted by common practice
- Arguments - recognize the issues at stake of the legal elements or the preparatory acts and pass through deliberation activities
- Activities - identify the various stages of the deliberation process (i.e. preparation activities, participative activities and decision-making activities) and reflect their anticipated inputs and outputs

3 Semantic enrichment of Lexi-is ontology with WordNet

Objective of semantic enrichment task is to identify semantic pointers from ontological objects to semantic elements (e.g. synsets, for WordNet) of a linguistic resource. The most ambitious feature of WordNet, however, is its attempt to organize lexical information in terms of word meanings, rather than word forms [4].

WordNet is a large lexical database of English,. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations [6]. For semantic enrichment of Lex-is ontology with WordNet, I propose using Ontoling - a plugin for the Protege , the popular ontology editing tool .

For semantic enrichment of Lexi-is ontology with WordNet, we use Protege 3.2. and OntoLing Tab. Protege is a popular ontology editor developed by

Stanford Medical Informatics at the Stanford University School of Medicine, allowing the linguistic enrichment of ontologies created within this working environment[9].

OntoLing [5] is a framework for a semi-automatic linguistic enrichment of ontologies. It has been developed at the AI Research Group, Department of Computer Science, Systems and Production of the University of Rome, Tor Vergata. OntoLing has been designed as a plug-in for Protege.

The ontology can be enriched with:
- Additional labels for the selected class, i.e., synonyms
- Glosses as descriptions for the selected class
- IDs of the selected senses as additional labels for the selected class

OntoLing accesses the selected linguistic resources. The term linguistic resources refers to sets of language data and descriptions in machine readable form, to be used in building, improving, or evaluating natural language and speech algorithms or systems [3].

In particular, this definition includes lexical databases, bilingual dictionaries and terminologies. In the current version of OntoLing(3.2) , two linguistic resources are available for the linguistic or multilingual enrichment, WordNet, for the linguistic enrichment of ontologies with English labels, and DICT dictionaries, for the linguistic and multilingual enrichment of ontologies [7].

For example,in our case, with this Linguistic Interface the user visualizes the linguistic information for the "Article" class in the Linguistic Browser Panel embedded in the Protege framework (Figure 1) .

For enrichment "Article" concept with "clause", we add terms selected from linguistic browser (in our case "clause") as additional labels for the selected frame("Article") (Figure 2).

We can use SPARQL for quering the enriched LEX-is ontology. SPARQL is a W3C Candidate Recommendation towards a standard query language for the Semantic Web. SPARQL can be used to query an RDF Schema or OWL model to filter out individuals with specific characteristics [12]. SPARQL can be used to express queries across diverse data sources and the results of SPARQL queries can be results sets or RDF graphs [10].

The SPARQL query processor will search for sets of triples that match these four triple patterns, binding the variables in the query to the corresponding parts of each triple. Important to note here is the "property orientation" (class matches can be conducted solely through class-attributes / properties [11].

For example, if I want to find all article labeled with "clause" for regulation with Celex number 32006R1052, I use the following query in SPARQL (according with Figure 3) :

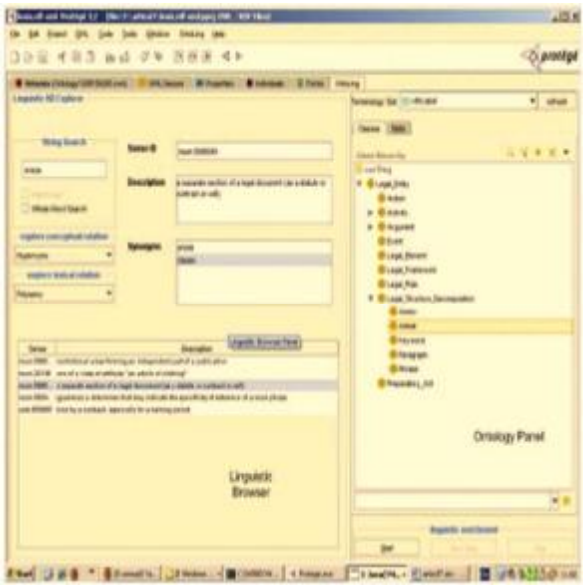

Fig. 1. A screenshot of the Ontoling PluginFigure

```
SELECT DISTINCT ?Article
WHERE
{
?Article rdfs:label "clause" .
?LegalElement ?hasIdentifier "32006R1052"
}
```

4 Conclusions

The amount of legal documents which are available on the web are mostly written in natural language; at the same time, people like to interact with computers using even more friendly interfaces using the natural language[3]. It appears evident the role of linguistic resources in legal ontology development. On the other hands, a more linguistic awareness could also help semantic search engines in augmenting the retrieval of proper results, or, at least, in excluding information which is not pertinent to the intention behind the submitted query.

In this paper I propose (using Ontoling - a plugin for the popular ontology editing tool Protégé) linguistic enrichment of legal ontologies, which has run through the identification of different categories of linguistic resources and planning their exploitation to augment the linguistic expressivity of legal ontologies.

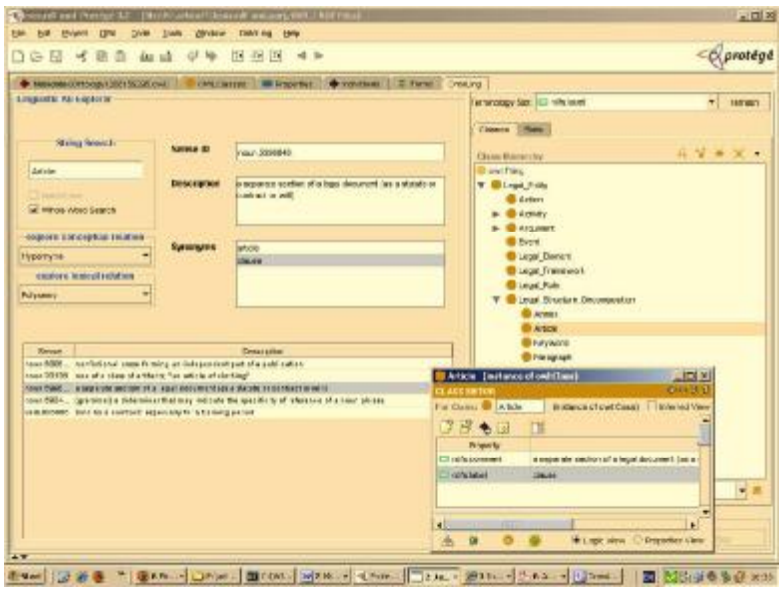

Fig. 2. Enrichment "Article" concept with "clause"

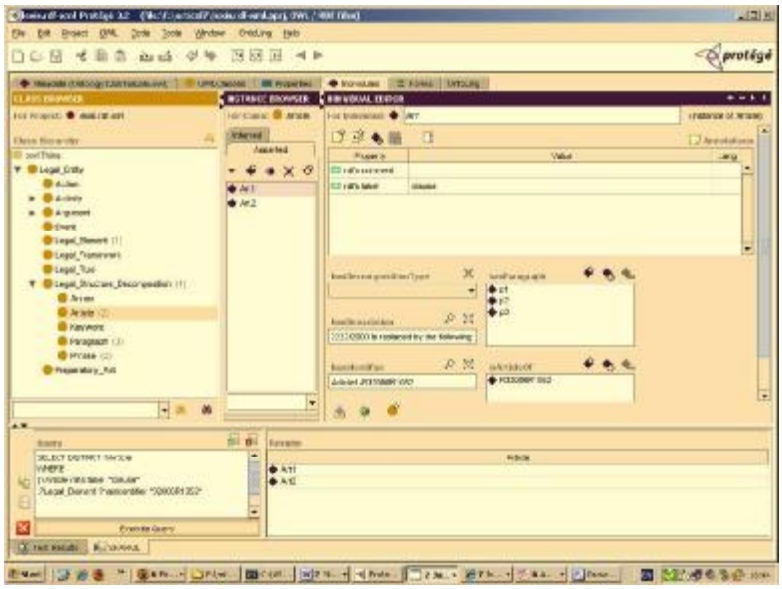

Fig. 3. Using SPARQL for quering the enriched ontology

References

[1] **Pazienza, M., Stellato, A**., Linguistic Enrichment of Ontologies: a methodological Framework, Workshop on Interfacing Ontologies and Lexical Resources for Semantic Web Technologies (OntoLex2006), Genoa, Italy (2006)
[2] **Pazienza, M., Stellato, A**., Exploiting Linguistic Resources for building linguistically motivated ontologies in the Semantic Web,Workshop on Interfacing Ontologies and Lexical Resources for Semantic Web Technologies (OntoLex2006), Genoa, Italy (2006)
[3] **Pazienza, M., Stellato, A.**, An open and scalable framework for enriching ontologies with natural language content, The 19th International Conference on Industrial, Engineering 'Applications of Applied Intelligent Systems (IEA/AIE'06), special session on Ontology '(2006)
[4] **Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D.**, 'K. Tarantello, G.:Introduction to WordNet: An On-line Lexical Database 1993
[5] ***, Lifecycle Support for Networked Ontologies, NeOn-project: Semantic-based knowledge and content systems", D 2.4.1 Multilingual ontology support 2007
[6] ***, WordNet : http://wordnet.princeton.edu/
[7] ***, OntoLing Tab : http://ai-nlp.info.uniroma2.it/software/OntoLing/
[8] ***, "Lex-is" Project - D1.3 - Ontology for Legal Framework Modelling, 2007:http://www.lex-is.eu/upload/deliverables/0402F10-LEX-IS-OntologyforLegalFrameworkModelling.pdf
[9] ***, Protege : http://protege.stanford.edu/
[10] ***, Protocol and RDF Query Language SPARQL Query Language for RDF : http://www.w3.org/TR/rdf-sparql-query/
[11] ***, Protocol and RDF Query Language SPARQL, Wikipedia : http://en.wikipedia.org/wiki/SPARQL
[12] ***, Using the sparql query engine with protege-owl : http://protege.stanford.edu/doc/sparql/