OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection

Paul Buitelaar, Thomas Eigner, Thierry Declerck

DFKI GmbH, Language Technology Dept.
Stuhlsatzenhausweg 3,
66123 Saarbruecken, Germany

paulb@dfki.de

The OntoSelect Ontology Library

OntoSelect provides an access point for ontologies on any possible topic or domain that will be updated continuously, organized in a meaningful way and with automatic support for ontology selection in knowledge markup. Unlike the DAML[1] and SchemaWeb[2] ontology libraries, OntoSelect is not based primarily on a static registration of published ontologies, but includes a crawling procedure that monitors the web for any newly published ontologies in the following representation formats: RDF/S, DAML or OWL.

Collected ontologies are analyzed using the OWL API (Bechhofer et al., 2003) that allows for the extraction of structure and content of any RDF/S, DAML or OWL ontology. There are currently around 745 ontologies in the OntoSelect library, covering a wide range of topics and domains. Ontologies are stored in a database and are organized according to: format; ontology-, class- and property-names; class- and property-labels. In the following two tables we present some statistics for the ontologies collected so far.

Table 1 gives an indication of the distribution of the three representation formats used. Here, it is interesting to see that the OWL format already shows a clear advance over the other two formats, even quite shortly after the finalization of its definition[3].

Format	***Unknown***	**RDF/S**	**DAML**	**OWL**
Percentage	*65 (8.72%)*	155 (20.81%)	213 (28.59%)	312 (41.88%)

Table 1: Percentage of Ontologies by Format

Table 2 gives an indication of the distribution of human languages used in the definition of labels for classes and properties. Labels are important for the use of ontologies in knowledge markup of text documents (in various languages). The table gives percentages of collected ontologies with labels in one or more languages (if explicit language-identification has been provided). The advance of English over other languages is not surprising as most ontologies still originate mainly from English speaking countries (UK, USA), although some start to appear with labels also in different languages (e.g. French).

[1] http://www.daml.org/ontologies/

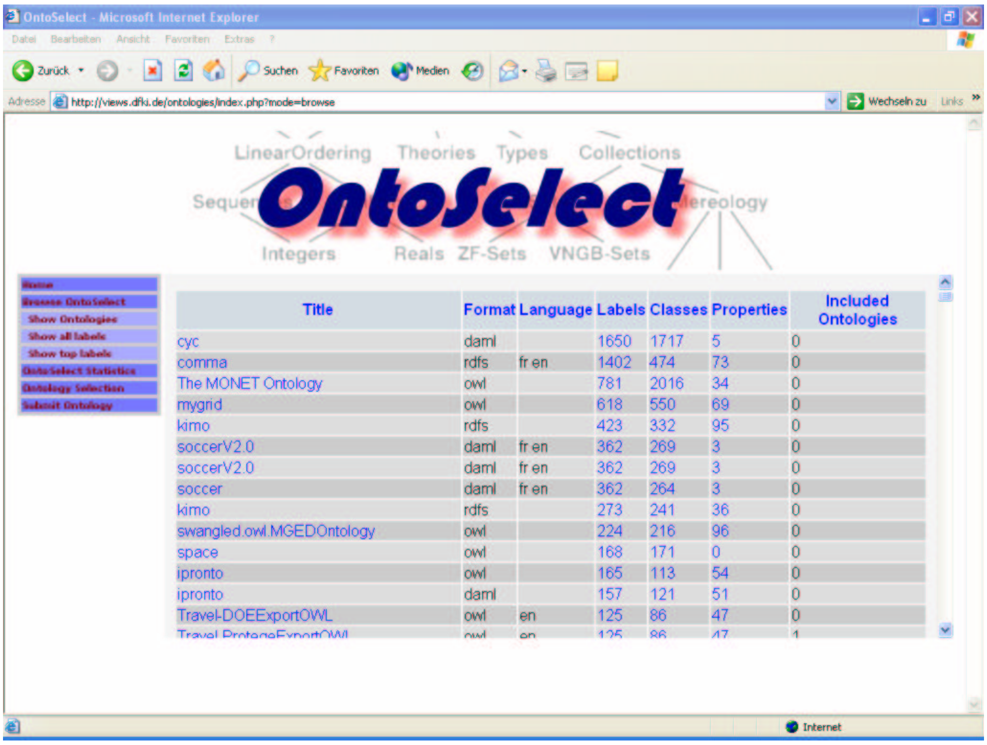
[2] http://www.schemaweb.info/

[3] The contents of OntoSelect are constantly updated. Therefore, any of the statistics mentioned in this paper may not correspond with the current situation.

Language(s)	Percentage
English	64 %
French + English	19 %
German + English	13 %
German	2 %
Dutch	2 %

Table 2: Percentage of Ontologies with Labels in one or more Language

OntoSelect provides the user with a detailed and up-to-date overview of web-accessible ontologies. The collection can be browsed by: ontology name (derived from `owl:Ontology/rdfs:comment` or from the ontology URL); format (from the ontology URL); human language (from `rdfs:label`); number of labels, classes, properties, or included ontologies (`owl:imports`). To illustrate this, Figure 1 presents the OntoSelect library with an overview of included ontologies sorted according to the number of labels:



Adresse: http://views.dfki.de/ontologies/index.php?mode=browse

Title	Format	Language	Labels	Classes	Properties	Included Ontologies
cyc	daml		1650	1717	5	0
comma	rdfs	fr en	1402	474	73	0
The MONET Ontology	owl		781	2016	34	0
mygrid	owl		618	550	69	0
kimo	rdfs		423	332	95	0
soccerV2.0	daml	fr en	362	269	3	0
soccerV2.0	daml	fr en	362	269	3	0
soccer	daml	fr en	362	264	3	0
kimo	rdfs		273	241	36	0
swangled.owl.MGEDOntology	owl		224	216	96	0
space	owl		168	171	0	0
ipronto	owl		165	113	54	0
ipronto	daml		157	121	51	0
Travel-DOEExportOWL	owl	en	125	86	47	0
Travel-ProtegeExportOWL	owl	en	125	86	47	1

Figure 1: View of the OntoSelect library with ontologies organized by number of labels

As mentioned above, the assignment of labels is important from the perspective of knowledge markup, as the automatic annotation of documents with ontological knowledge crucially depends on the availability of terminology for classes and properties. In order to provide ontology developers with an insight in actual terminology used in ontologies, the OntoSelect library stores all labels together with the ontologies in which they are used. In this way, the user can browse ontologies by terminology and keep track of labels that are used for classes or properties in different ontologies. For instance, `<rdfs:label>person</rdfs:label>` is used in 35 different ontologies, which indicates a repeated (but mostly slightly different) definition of the corresponding concept.

Ontology Selection with OntoSelect

As there is a rapidly increasing number of published ontologies available, it is becoming a more and more difficult task to select the most appropriate one(s) in knowledge markup. To provide semi-automatic support for this, OntoSelect includes a functionality for selecting ontologies for a given knowledge markup task, based on the following criteria that address ontology content and structure:

- *Coverage*: How many of the terms in the document collection of the particular knowledge markup task are covered by the classes and properties in the ontology?

  Coverage is measured by the number of labels for classes and properties that can be matched in the document collection[4]. This is implemented by a combination of ontology preprocessing (normalization of labels or class- and property-names), linguistic analysis (part-of-speech tagging, morphological analysis of the document collection) and statistical processing (pre-selection of statistically relevant terms from the document collection).

- *Structure*: How detailed is the knowledge structure that the ontology represents?

  Structure is measured by the number of properties relative to the number of classes of the ontology. This parameter is based on the observation that more advanced ontologies generally have a large number of properties. Therefore, a relatively large number of properties would indicate a highly structured and hence more advanced ontology. -- The *Structure* parameter is considered only for the top ranked ontologies according to *Coverage*.

- *Connectedness*: Is the ontology connected to other ontologies and how well established are these?

  Connectedness is measured by the number and quality of imported ontologies and by the number and quality of ontologies that import the ontology under consideration. The basic idea here is that more qualified ontologies will be based on other qualified ontologies and/or will be included in other qualified ontologies. This parameter is still very preliminary as only a small number of ontologies do in fact import other ontologies. -- The *Connectedness* parameter is considered only for the top ranked ontologies according to *Coverage*.

To illustrate this process, we show the results of selecting appropriate ontologies for knowledge markup of a set of documents from the Knowledge Media Institute on research projects and visits (PlanetNews[5]). These news stories have been used in knowledge markup, for instance in the Magpie[6] project at KMI on ontology-based hyperlinking. The results for this document collection are presented in Figure 2 below.

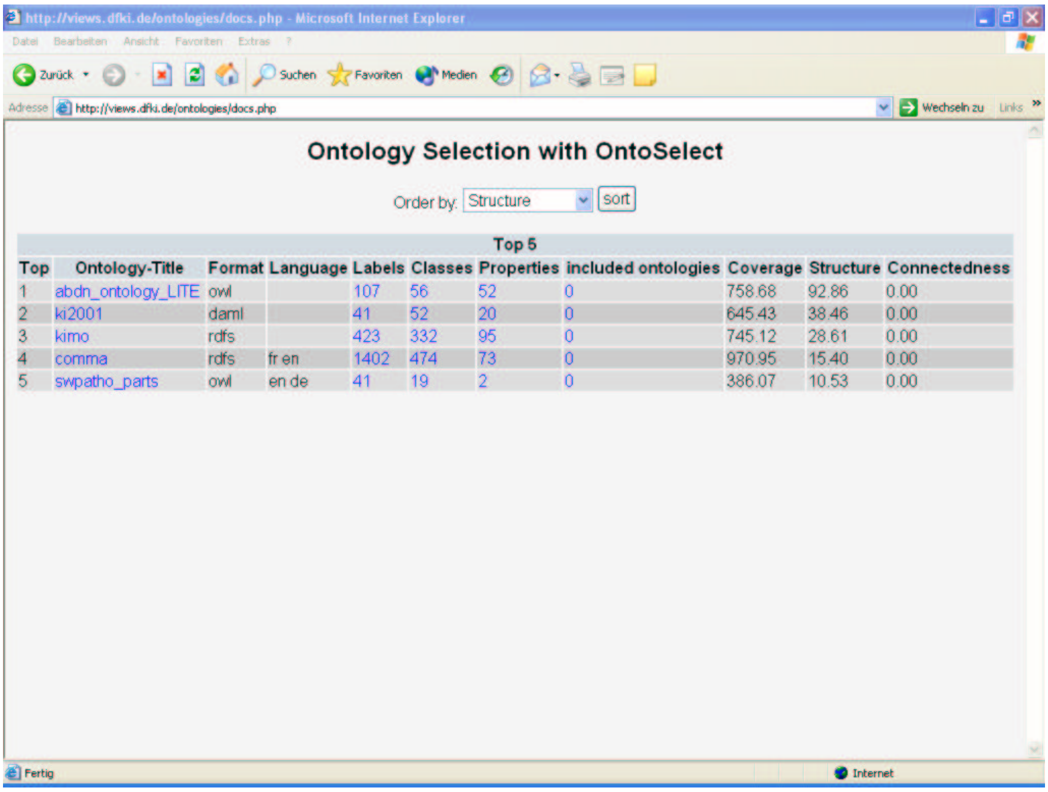
A combination of the results of the three parameters will indicate the most appropriate ontology for a particular knowledge markup task. We are currently investigating how best to combine results into one score. In the PlanetNews case as shown in Figure 2, the "abdn_ontology_LITE"[7] ontology seems most appropriate for knowledge markup of this document collection as it has a high coverage of document terms (labels of classes and/or properties occurring in PlanetNews) and a rich structure (ratio of properties to classes).

[4] A more elaborate approach to coverage is discussed by (Brewster et al., 2004).

[5] http://kmi.open.ac.uk/news/planetnews.html

[6] http://kmi.open.ac.uk/projects/magpie/main.html

[7] A version of the reference ontology on Research Management developed in the AKT project.


http://views.dfki.de/ontologies/docs.php - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Zurück
Suchen
Favoriten
Medien
Adresse
http://views.dfki.de/ontologies/docs.php
Wechseln zu
Links
Fertig
Internet

Ontology Selection with OntoSelect

Order by: Structure sort

Top 5										
Top	Ontology-Title	Format	Language	Labels	Classes	Properties	included ontologies	Coverage	Structure	Connectedness
1	abdn_ontology_LITE	owl		107	56	52	0	758.68	92.86	0.00
2	ki2001	daml		41	52	20	0	645.43	38.46	0.00
3	kimo	rdfs		423	332	95	0	745.12	28.61	0.00
4	comma	rdfs	fr en	1402	474	73	0	970.95	15.40	0.00
5	swpatho_parts	owl	en de	41	19	2	0	386.07	10.53	0.00

Figure 2: Ontology selection results in OntoSelect for the PlanetNews document collection

References

S. Bechhofer, Phillip Lord, Raphael Volz. Cooking the Semantic Web with the OWL API. 2nd International Semantic Web Conference, ISWC, Sanibel Island, Florida, October 2003.

Ch. Brewster, H. Alani, S. Dasmahapatra and Y. Wilks. Data Driven Ontology Evaluation. International Conference on Language Resources and Evaluation, Lisbon, Portugal, 2004.

Availability

The OntoSelect ontology library is available at http://views.dfki.de/ontologies/

Acknowledgements

OntoSelect is supported by grants for the projects VIeWs (by the Saarland Ministry of Economic Affairs) and SmartWeb (by the German Ministry of Education and Research: 01 IMD01 A).