Evaluating Ontology Search

Paul Buitelaar, Thomas Eigner

German Research Center for Artificial Intelligence (DFKI GmbH)
Language Technology Lab & Competence Center Semantic Web
Stuhlsatzenhausweg 3
Saarbrücken, Germany
paulb@dfki.de

Abstract. As more and more ontologies are being published on the Semantic Web, selecting the most appropriate ontology will become an increasingly important subtask in Semantic Web applications. Here we present an approach towards ontology search in the context of OntoSelect, a dynamic web-based ontology library. In OntoSelect, ontologies can be searched by keyword or by document. In keyword-based search only the keyword(s) provided by the user will be used for the search. In document-based search the user can provide either a URL for a web document that represents a specific topic or the user simply provides a keyword as the topic which is then automatically linked to a corresponding Wikipedia page from which a linguistically/statistically derived set of most relevant keywords will be extracted and used for the search. In this paper we describe an experiment in evaluating the document-based ontology search strategy based on an evaluation data set that we constructed specifically for this task.

1 Introduction

A central task in the Semantic Web effort is the semantic annotation or knowledge markup of data (textual or multimedia documents, structured data, etc.) with semantic metadata as defined by one or more ontologies. The added semantic metadata allow for automatic processes (agents, web services, etc.) to interpret the underlying data in a unique and formally specified way, thereby enabling autonomous information processing. As ontology-based semantic metadata are in fact class descriptions, the annotated data can be extracted as instances for these classes. Hence, another way of looking at ontology-based semantic annotation is as ontology population.

Most of current work in ontology-based semantic annotation assumes ontologies that are typically developed specifically for the task at hand. Instead, a more realistic approach would be to access an ontology library and to select one or more appropriate ontologies. Although the large-scale development and publishing of ontologies is still only in a beginning phase, many are already available. To select the most appropriate ontology (or a combination of complementary ontologies) will therefore be an increasingly important subtask of Semantic Web applications.

Until very recently the solution to this problem was supposed to be handled by foundational ontology libraries [1,2]. However, in recent years, dynamic web-based ontology libraries and ontology search engines like OntoKhoj [3], OntoSelect [4], SWOOGLE [5] and Watson [6] have been developed that enable a more data-driven approach to ontology search and retrieval.

In OntoSelect, ontologies can be searched by keyword or by document. In keyword-based search only the keyword(s) provided by the user will be used for the search. In document-based search the user can provide either a URL for a web document that represents a specific topic or the user simply provides a keyword as the topic which is then automatically linked to a corresponding Wikipedia page from which a linguistically/statistically derived set of most relevant keywords will be extracted and used for the search. In this paper we describe an experiment in evaluating the document-based ontology search strategy based on an evaluation data set that we constructed specifically for this task.

The remainder of the paper is structured as follows. Section 2 gives a brief overview of the content and functionality of the OntoSelect ontology library. Section 3 presents a detailed overview of the ontology search algorithm and scoring method used. Section 4 presents the evaluation benchmark, experiments and results. Finally, section 5 presents some conclusions and gives an outlook on future work

2 The OntoSelect Ontology Library

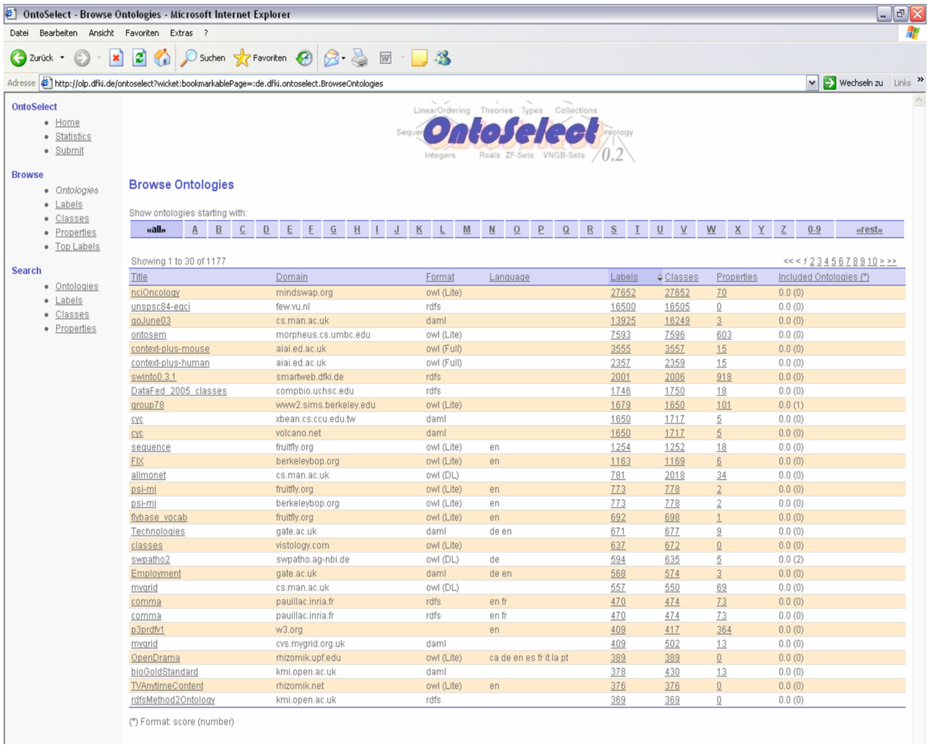
OntoSelect is a dynamic web-based ontology library that collects, analyzes and organizes ontologies published on the Semantic Web. OntoSelect allows browsing of ontologies according to size (number of classes, properties), representation format (DAML, RDFS, OWL), connectedness (score over the number of included and referring ontologies) and human languages used for class- and object property-labels. OntoSelect further includes an ontology search functionality as described above and discussed in more detail in the following sections.

OntoSelect uses the Google API to find published ontologies on the web in the following formats: DAML, OWL and RDFS. Jena is used for reading and analyzing the ontologies. In the case of OWL, OntoSelect also determines its type (Full, DL, Lite) and indexes this information accordingly. Each class and object property defined by the ontology is indexed with reference to the ontology in which it occurs. Correspondingly, each label is indexed with reference to the corresponding ontology, class or object property, the human language of the label (if available), and a normalized label name, e.g. TaxiDriver is normalized to “taxi driver”. Object properties are handled similarly as classes except that also information on their type (functional, transitive, symmetric) is indexed. Finally, a separate index is build up in which we keep track of the distribution of labels over all of the collected ontologies. In this way, a ranked list of frequently used labels can be maintained and browsed by the user.

3 Ontology Search

3.1 Ontology Search Measures and Criteria

The ontology search problem is a very recent topic of research, which only originated with the growing availability of ontologies on the web. A web-based ontology, defined



Title	Domain	Format	Language	Labels	Classes	Properties	Included Ontologies (*)
nciOncology	mindswap.org	owl (Lite)		27652	27652	70	0.0 (0)
unspsc84-eqci	few.vu.nl	rdfs		16500	16505	0	0.0 (0)
goJune03	cs.man.ac.uk	daml		13925	16249	3	0.0 (0)
ontosem	morpheus.cs.umbc.edu	owl (Lite)		7593	7596	603	0.0 (0)
context-plus-mouse	aiai.ed.ac.uk	owl (Full)		3555	3557	15	0.0 (0)
context-plus-human	aiai.ed.ac.uk	owl (Full)		2357	2359	15	0.0 (0)
swinto0.3.1	smartweb.dfki.de	rdfs		2001	2006	918	0.0 (0)
DataFed_2005_classes	compbio.uchsc.edu	rdfs		1746	1750	18	0.0 (0)
group78	www2.sims.berkeley.edu	owl (Lite)		1679	1650	101	0.0 (1)
cyc	xbean.cs.ccu.edu.tw	daml		1650	1717	5	0.0 (0)
cyc	volcano.net	daml		1650	1717	5	0.0 (0)
sequence	fruitfly.org	owl (Lite)	en	1254	1252	18	0.0 (0)
FIX	berkeleybop.org	owl (Lite)	en	1163	1169	6	0.0 (0)
allmonet	cs.man.ac.uk	owl (DL)		781	2018	34	0.0 (0)
psi-mi	fruitfly.org	owl (Lite)	en	773	778	2	0.0 (0)
psi-mi	berkeleybop.org	owl (Lite)	en	773	778	2	0.0 (0)
flybase_vocab	fruitfly.org	owl (Lite)	en	692	698	1	0.0 (0)
Technologies	gate.ac.uk	daml	de en	671	677	9	0.0 (0)
classes	vistology.com	owl (Lite)		637	672	0	0.0 (0)
swpatho2	swpatho.ag-nbi.de	owl (DL)	de	594	635	5	0.0 (2)
Employment	gate.ac.uk	daml	de en	568	574	3	0.0 (0)
mygrid	cs.man.ac.uk	owl (DL)		557	550	69	0.0 (0)
comma	pauillac.inria.fr	rdfs	en fr	470	474	73	0.0 (0)
comma	pauillac.inria.fr	rdfs	en fr	470	474	73	0.0 (0)
p3prdfv1	w3.org		en	409	417	364	0.0 (0)
mygrid	cvs.mygrid.org.uk	daml		409	502	13	0.0 (0)
OpenDrama	rhizomik.upf.edu	owl (Lite)	ca de en es fr it la pt	389	389	0	0.0 (0)
bioGoldStandard	kmi.open.ac.uk	daml		378	430	13	0.0 (0)
TVAnytimeContent	rhizomik.net	owl (Lite)	en	376	376	0	0.0 (0)
rdfsMethod2Ontology	kmi.open.ac.uk	rdfs		369	369	0	0.0 (0)

(*) Format: score (number)

Fig. 1. Browsing ontologies in OntoSelect

by representation languages such as OWL or RDFS, is in many respects just another web document that can be indexed, stored and retrieved. On the other hand, an ontology is a highly structured document with possibly explicit semantic links to other ontologies. The OntoSelect approach is based on both observations by ranking ontologies by coverage, i.e. the overlap between query terms and index terms; by structure, i.e. the ratio of class vs. property definitions; and by connectedness, i.e. the level of integration between ontologies.

Other approaches have similarly stressed the importance of such measures, e.g. [7] describe the “Class Match”, “Density”, “Semantic Similarity” and “Betweenness” measures. The Class Match and Density measures correspond roughly to our coverage and structure measure, whereas the Semantic Similarity and Betweenness measure the semantic weight of query terms relative to the different ontologies that are to be ranked. These last two measures are based on the assumption that ontologies are well-structured with equal semantic balance throughout all constitutive parts, which unfortunately is only seldom the case and we therefore do not take such measures into account.

Another set of measures or rather criteria for ontology search has been proposed by [8]. The focus here is more on the application of found ontologies and therefore includes such criteria as: ‘modularization’ (can retrieved ontologies be split up in useful modules); ‘returning ontology combinations’ (can retrieved ontologies be used in combination); ‘dealing with instances’ (do retrieved ontologies include instances as well as classes/properties).

These criteria are desirable but are currently not central to the OntoSelect approach and to this paper. Our focus is rather on providing data-driven methods for finding the best matching ontology for a given topic and on providing a proper evaluation of these methods.

3.2 Ontology Search in OntoSelect

Ontology ranking in OntoSelect is based on a combined measure of *coverage*, *structure* and *connectedness* of ontologies as discussed above. Further, OntoSelect provides automatic support in ontology ranking relative to a web document instead of just one or more keyword(s). Obviously this allows for a much more fine-grained ontology search process.

For a given document as search query, OntoSelect first extracts all textual data and analyses this with linguistic tools (i.e. ‘part-of-speech tagger’ and ‘morphological analysis’) to extract and normalize all nouns in the text as these can be expected to represent ontology classes rather than verbs, adjectives, etc. The frequencies of these nouns in the query document is then compared with their frequencies in a reference corpus - consisting of a large collection of text documents on many different topics and covering a large section of the English language - to estimate the relevance for each noun based on how often it is expected to appear in a more general text of the same size. Chi-square is used to estimate this relevance score (see also Coverage score below). Only the top 20 nouns are used further in the search process as extracted keywords.

To calculate the relevance of available ontologies in OntoSelect, the set of 20 extracted keywords is used to compute three separate scores (*coverage*, *structure*, *connectedness*) and a combined score as described below:

Coverage: How many of the terms in the document are covered by the labels in the ontology?

To estimate the coverage score, OntoSelect iterates over all ontologies containing at least one label (either the original label name or the normalized label name) occurring in the top 20 keyword list of the search document. For each label occurring in the document, OntoSelect computes its relevance, with which the coverage score of an ontology O is calculated.

$$
\begin{aligned}
QD &= \text{Query Document} \\
KW &= \text{Set of extracted keywords of } QD \\
OL &= \text{Set of labels for ontology } O \\
RefC &= \text{Reference Corpus} \\
Exp_k &= \tfrac{RefC_k}{|RefC|} \times |QD| \\
\chi^2(k) &= \tfrac{QD_k - Exp_k}{Exp_k} \\
coverage(O, QD) &= \textstyle\sum_{k \in KW(QD) \cap OL(O)} \chi^2(k)
\end{aligned}
$$

Connectedness: Is the ontology connected to other ontologies and how well established are these?
Similar to the Google PageRank algorithm [9], OntoSelect checks how many ontologies import a specific ontology, but also how many ontologies are imported by that one. The connectedness score of an ontology O is calculated accordingly.

$$\begin{aligned}
cIO(O) &= \text{number of imported Ontologies for } O \\
cIRO(O) &= \text{number of imported Ontologies} \\
&\quad \text{(that could be parsed) for } O \\
cIFO(O) &= \text{number of Ontologies importing } O \\
IO(O) &= \{x \mid x \text{ imports the Ontology } O\} \\
iS(O, level) &= \tfrac{cIFO(O)}{2^{level}} + \textstyle\sum_{O' \in IO(O)} iS(O', level+1) \\
connectedness(O) &= \begin{cases} cIO(O) > 0 : & \frac{iS(O,0)*cIO(O)}{iS(O,0)cIO(O)} \times \frac{countIRO(O)}{countIO(O)} \\ else : & 0 \end{cases}
\end{aligned}$$

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 O. 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.

$$structure(O) = \frac{\text{\# of properties in ontology O}}{\text{\# of classes in ontology O}}$$

Combined Score: Since the ranges of coverage, connectedness and structure are very discrepant these values have to be normalized. In other words, all coverage values are divided by the maximum coverage value, all connectedness values by the maximum connectedness value and all structure values by the maximum structure value, giving rise to final values between 0 and 1. Because each type of score has a different significance, the final score is a weighted combination of the three individual score.

$$score = \frac{3 \times coverage_{norm} + 2 \times connectedness_{norm} + structure_{norm}}{6}$$

3.3 An Example of Ontology Search in OntoSelect

The application of the ranking and search algorithm discussed above can be illustrated with an example of ontology search on the topic ‘genetics’, which may be represented by the Wikipedia page on ‘Gene’:

http://en.wikipedia.org/wiki/Gene

The results of the keyword extraction and ontology ranking process for this query document are reported by OntoSelect in two tables, one that shows the top 20 keywords extracted from the query document and one with the ranked list of best matching ontologies according to the computed score (see Figure 2). Combined and individual scores - connectedness, structure, coverage - are shown as well as the matching labels/keywords and their relevance scores. Extracted and top ranked keywords include "gene", "molecule", "transcription", "protein", etc., all of which are indeed of relevance to the 'genetics' topic.

Retrieved and top ranked ontologies include a large number that are indeed of relevance to the 'genetics' topic, e.g. "nciOncology", "bioGoldStandard", "mygrid", "sequence", etc. Only some of the ontologies are not or less relevant, e.g. "swinto" (which is mainly on football but also includes all of SUMO that does in fact cover many terms that are relevant to genetics), "gold" (which is mainly on linguistics but includes some terms that have also some relevance to genetics), "dolce" (which is a foundational top ontology that includes some terms with relevance to genetics).

Best ontologies found in input document:

Score	Title	Matches (*)	Domain	Format	Language	Labels	Classes	Properties	Connectedness	Structure	Coverage
3.48	nciOncology	gene (946.01), height (862.27), research (191.91), variation (132.32), evolution (127.72), plant (113.12), level (86.73), mutation (84.08), environment (81.67), outcome (66.25)	mindswap.org	owl (Lite)		27652	27652	70	0.0	0.0	0.7
1.5	bioGoldStandard	organism (2834.07), gene (946.01), structure (88.93)	kmi.open.ac.uk	daml		378	430	13	0.0	0.0	1.0
1.5	mygrid	organism (2834.07), gene (946.01), structure (88.93)	cvs.mygrid.org.uk	daml		409	502	13	0.0	0.0	1.0
1.5	mygrid	organism (2834.07), gene (946.01), structure (88.93)	cs.man.ac.uk	owl (DL)		557	550	69	0.0	0.0	1.0
1.17	swinto0.3.1	organism (2834.07), plant (113.12), environment (81.67)	smartweb.dfki.de	rdfs		2001	2006	918	0.0	0.0	0.78
0.75	OBI	organism (2834.07), environment (81.67)	fugo.sourceforge.net	owl (Full)	en	153	161	9	0.0	0.0	0.75
0.7	umlssn	organism (2834.07)	swpatho.ag-nbi.de	owl (DL)	de en	75	87	65	1.0	0.0	0.73
0.5	group78	height (862.27), square (326.85), plant (113.12)	www2.sims.berkeley.edu	owl (Lite)		1679	1650	101	0.0	0.0	0.34
0.43	psi-mi	gene (946.01), interaction (91.92), mutation (84.08)	fruitfly.org	owl (Lite)	en	773	778	2	0.0	0.0	0.29
0.43	psi-mi	gene (946.01), interaction (91.92), mutation (84.08)	berkeleybop.org	owl (Lite)	en	773	778	2	0.0	0.0	0.29
0.37	dolce2.0-lite-v3	organism (2834.07)	coli.lili.uni-bielefeld.de	owl (DL)		81	79	75	0.0	0.0	0.73
0.37	context-core	organism (2834.07)	aiai.ed.ac.uk	owl (Full)		29	31	15	0.0	0.0	0.73
0.37	MGEDOntology	organism (2834.07)	mged.sourceforge.net	daml		228	437	10	0.0	0.0	0.73
0.37	context-plus-human	organism (2834.07)	aiai.ed.ac.uk	owl (Full)		2357	2359	15	0.0	0.0	0.73
0.37	loggerhead_nesting	organism (2834.07)	fruitfly.org	owl (Lite)	en	308	314	4	0.0	0.0	0.73
0.37	context-core-protege	organism (2834.07)	aiai.ed.ac.uk	owl (Full)		29	31	15	0.0	0.0	0.73
0.37	obi	organism (2834.07)	berkeleybop.org	owl (Full)	en	198	211	15	0.0	0.0	0.73
0.37	context-plus-mouse	organism (2834.07)	aiai.ed.ac.uk	owl (Full)		3555	3557	15	0.0	0.0	0.73
0.26	comma	research (191.91), journal (110.38), structure (88.93)	pauillac.inria.fr	rdfs	en fr	470	474	73	0.0	0.0	0.13
0.2	russiaA	square (326.85), plant (113.12), level (86.73)	aifb.uni-karlsruhe.de	owl (Lite)	en	150	151	60	0.0	0.0	0.14

(*) Format: matching keyword in ontology (significance)

Fig. 2. Ranked list of retrieved ontologies for Wikipedia page 'Gene'

4 Evaluation

In order to test the accuracy of our approach we designed an evaluation experiment with a specifically constructed benchmark of 57 ontologies from the OntoSelect library that were manually assigned to 15 different topics, each of which represented by one or more Wikipedia pages. In this way we were able to define ontology search as a regular information retrieval task, for which we can give relevance assessments (manual assignment of ontology documents to Wikipedia-based topics) and compute precision and recall for a set of queries (Wikipedia pages). In the following we describe the evaluation benchmark in some more detail as well as the evaluation process and results.

4.1 Evaluation Benchmark

The evaluation experiment is based on a benchmark that consists of 15 Wikipedia topics and 57 out of 1056 ontologies that have been collected through OntoSelect. The 15 Wikipedia topics covered by the evaluation benchmark were selected out of the set of all class/property labels in OntoSelect - 37284 in total - by the following steps:

- Filtering out labels that did not correspond to a Wikipedia page - this left us with 5658 labels (i.e. topic candidates)
- Next, the 5658 labels were used as search terms in SWOOGLE to filter out labels that returned less than 10 ontologies (out of the 1056 in OntoSelect) - this left us with 3084 labels / topics
- We then manually decided which of these 3084 labels actually expressed a useful topic, e.g. we left out very short labels (‘v’) and very abstract ones (‘thing’) - this left us with 50 topics
- Finally, out of these 50 we randomly selected 15 for which we manually checked the ontologies retrieved from OntoSelect and SWOOGLE - in this step we checked 269 ontologies out of which 57 were judged as appropriate for the corresponding topic

The resulting 15 Wikipedia topics with the number of appropriately assigned ontologies are: Atmosphere (2), Biology (11), City (3), Communication (10), Economy (1), Infrastructure (2), Institution (1), Math (3), Military (5), Newspaper (2), Oil (0), Production (1), Publication (6), Railroad (1), Tourism (9) For instance, the following 3 ontologies could be assigned to the topic (Wikipedia page) City:

- http://www.mindswap.org/2003/owl/geo/geoFeatures.owl
- http://www.glue.umd.edu/ katyn/CMSC828y/location.daml
- http://www.daml.org/2001/02/geofile/geofile-ont

4.2 Experiment and Results

Based on the evaluation benchmark we defined an experiment that measures how accurate the OntoSelect ontology ranking and search algorithm returns results for each of the topics in the benchmark and compare results with SWOOGLE. Average precision for OntoSelect and SWOOGLE is shown in Figure 3 with detailed results presented in Table 1. The first two columns present the benchmark, against which the experiment is evaluated. The third and fourth columns show recall, precision and F-measure computed over the top 20 retrieved ontologies in OntoSelect and SWOOGLE respectively.

Results unfortunately show that OntoSelect on average performs worse than SWOOGLE, although for selected topics OntoSelect does give better results. In current work we are therefore improving our search algorithm in various ways, e.g. by introducing a centrality score for individual classes - and therefore also for corresponding labels that are to be matched with the search topic and related keywords.

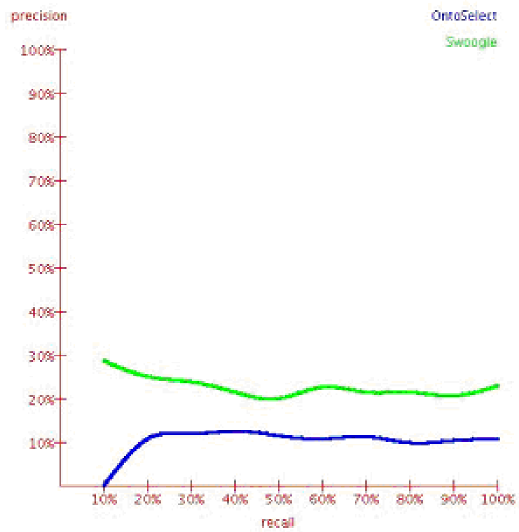


Fig. 3. Average precision for OntoSelect and SWOOGLE

Benchmark		**OntoSelect**			**SWOOGLE**		
Topic	**Assigned Ontologies**	**Rec.**	**Prec.**	**F**	**Rec.**	**Prec.**	**F**
Atmosphere	2	0.5	0.1	0.2	1.0	0.2	0.3
Biology	11	0.7	0.8	0.8	0.1	0.1	0.1
City	3	0.3	0.1	0.2	0.3	0.1	0.2
Communication	10	0	0	0	0.6	0.6	0.6
Economy	1	0	0	0	1.0	0.1	0.2
Infrastructure	2	0.5	0.1	0.2	1.0	0.2	0.3
Institution	1	0	0	0	0	0	0
Math	3	0.3	0.1	0.2	1.0	0.3	0.5
Military	5	0	0	0	0.6	0.3	0.4
Newspaper	2	0.5	0.1	0.2	0.5	0.1	0.2
Oil	0	0	0	0	0	0	0
Production	1	1.0	0.1	0.2	0	0	0
Publication	6	0.2	0.1	0.1	0.3	0.2	0.3
Railroad	1	0	0	0	1.0	0.1	0.2
Tourism	9	0	0	0	1.0	0.9	0.9

Table 1. Detailed results over all 15 topics

More in general however, we see our contribution in establishing an evaluation benchmark for ontology search that will enable us to improve the OntoSelect search service in a systematic way. As we intend to make this evaluation benchmark (the 'OntoSelect data set') publicly available, we hope this will also be of use to the Semantic Web community and will allow for better comparison between different systems and methods.

5 Conclusions and Future work

We discussed the OntoSelect search algorithm and described an experiment in evaluating this against an evaluation benchmark (the 'OntoSelect data set') that we constructed specifically for this task. The benchmark consists of 15 topics (represented by Wikipedia pages) that were manually assigned to 57 ontologies from a set of 1056 that were collected automatically through OntoSelect. The evaluation experiment has shown that OntoSelect on average performs worse than SWOOGLE, although for selected topics OntoSelect does give better results. In future work we will further investigate the reasons for this, e.g. we currently investigate the influence of centrality of classes relative to an ontology which may be used to reduce the relevance of general ontologies such as SUMO (as included in the SWIntO ontology). We also intend to extend the evaluation benchmark towards 50 topics and make this resource publicly available.

Demonstration

The OntoSelect ontology library and ontology search is available at:

http://olp.dfki.de/OntoSelect/

Acknowledgements

We thank Michael Velten for implementing the current version of OntoSelect and Bogdan Sacaleanu for providing us with useful comments and insights on the evaluation experiments. This research has been supported in part by the SmartWeb project, which is funded by the German Ministry of Education and Research under grant 01 IMD01.

References

1. G. van Heijst, A.T. Schreiber, and B.J. Wielinga. Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 46(2/3):183–292, 1997.
2. Y. Ding and D. Fensel. Ontology Library Systems: The key to successful Ontology Re-use. *Proceedings of the First Semantic Web Working Symposium. California, USA: Stanford University*, pages 93–112, 2001.
3. C. Patel, K. Supekar, Y. Lee, and EK Park. OntoKhoj: a semantic web portal for ontology searching, ranking and classification. *Proceedings of the fifth ACM international workshop on Web information and data management*, pages 58–61, 2003.

4. P. Buitelaar, T. Eigner, and T. Declerck. OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. *Proceedings of the Demo Session at the International Semantic Web Conference. Hiroshima, Japan*, 2004.
5. L. Ding, T. Finin, A. Joshi, R. Pan, R.S. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs. Swoogle: a search and metadata engine for the semantic web. *Proceedings of the Thirteenth ACM conference on Information and knowledge management*, pages 652–659, 2004.
6. M. d'Aquin, M. Sabou, M. Dzbor, C. Baldassarre, S. Gridinoc, L. Angeletou, and Motta E. WATSON: A Gateway for the Semantic Web. *In Proceedings of the 5th International Semantic Web Conference (ISWC), Georgia, USA*, 2005.
7. H. Alani, C. Brewster, and N. Shadbolt. Ranking Ontologies with AKTiveRank. *Poster session of the European Semantic Web Conference, ESWC*, 2006.
8. M. Sabou, V. Lopez, E. Motta, and V. Uren. Ontology Selection: Ontology Evaluation on the Real Semantic Web. *Proceedings of the Evaluation of Ontologies on the Web Workshop, held in conjunction with WWW*, 2006, 2006.
9. L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web, 1998.