OLS2OWL. A repository management facility.

Alexander García-Castro[2], Leyla García-Castro[1], José María Villaveces[3], Guillermo Calderón[3], Martin Hepp[1]

[1]E-Business and Web Science Research Group, Bundeswehr University Munich, Germany
leyla.garcia@ebusiness-unibw.org, mhepp@computer.org
[2]Bremen University, Germany
cagarcia@uni-bremen.de
[3]Universidad de La Sabana
jose.villaveces, guillermo.calderon@unisabana.edu.co

Abstract

This paper presents OLS2OWL, an ontology repository manager plug-in developed for Protégé 4.0. The tool (1) facilitates search and retrieval over collections of ontologies, (2) supports comparison across terms from different ontologies by providing integrative views; the latter supports weaving and maintaining a dense network of semantic relationships between multiple ontologies, and (3) facilitates reusing entire ontologies as well as portions of them—i.e., slices of ontologies. We argue that these three functionalities are essential for building good ontologies without redefining conceptual elements that already exist. This is particularly relevant to broad collaborative settings in which redundant and partly overlapping definitions are hard to control with current tooling.
Availability: http://ols2owl.sourceforge.net/

1. Introduction

As the Semantic Web (SW) envisions a metadata-rich Web where human-readable content will have machine-understandable semantics, there has been an increasing number of OWL ontologies [1] responding to these knowledge representation requirements. *Wang et al* collected 1275 files, both OWL and RDF schemas, in 2005; a more recent count, based on web crawling, gave an impressive result of over 6000 validated OWL ontologies (*Backer et al*, unpublished data); by the same vein Swoogle [2] hosts 2,563,125 Semantic Web Documents (SWD) [3]. These growing numbers, which reflect the intrinsic need of the SW for ontologies, have fostered a number of research projects aimed at supporting re-usability, better modularization as well as repositories for ontologies that, in principle, should support intelligent storage and retrieval for the encoded knowledge.

Repositories, within the context of the SW, should offer more than just data storage. The Ontolog community, a virtual community of practice of ontology experts, propose that the purpose of an Open Ontology Repository (OOR) should be to provide an architecture and an infrastructure that supports: a) the creation, sharing, searching, and management of ontologies, and b) linkage to databases, XML Schema structured data and documents [4]. Currently there are some ontology repositories accessible over the Web, however none of these complies with the requirements agreed upon during the last Ontolog Summit. For instance, Swoogle provides a single entry-point to several semantic web documents (ontologies), but does not offer any validation, as there is no quality control over the exposed material. Swoogle's query approach for finding ontologies is based on (sub) string search and link-based reference counting; once the document has been found it does not support any further operation. Also allowed is the composition of queries via the REST interface. OntoSelect [5] offers a similar approach; it presents the user with a basic overview of web-accessible ontologies. The collection can be browsed by: ontology name (derived from owl:Ontology/rdfs:comment); format (from the ontology URL); human language (from rdfs:label); number of labels, classes, properties, or included ontologies (owl:imports). Currently OntoSelect hosts 1530 ontologies. The TONES repository, developed as part of the TONES project [6], hosts 185 ontologies. It aims to provide a reasonable number of ontologies for testing purposes, emphasizing reasoning techniques. This repository also supports the REST interface for programmatic access. Ontologies can be selected and sorted by means of metrics for expressivity, class and property restrictions and axioms, logics, and individuals. A more novel approach is presented by Rubin *et al* [7] with Bioportal. Not only does this provide access to several ontologies, but it also facilitates online editing operations such as annotation of ontologies in the form of marginal notes –currently only available for classes. Finally, Pan *et al* [8] present a lightweight metadata ontology supporting an ontology repository based on a multiagent system; this repository is also accessible *via* a REST interface.

Although existing ontology repositories aim to provide access to semantic web documents by means of similar query facilities, they diverge in the methods provided by the different APIs. Furthermore, most of the investigated repositories are not integrated with ontology editors; this makes it difficult for users to gather several ontologies from one or many repositories in one operation. Queriability across multiple repositories should be facilitated within an ontology editor enviroment so that the subsequent manipulation of results is possible; it should also be made straightforward to include new repositories with no changes to the client or ontology editor. To realise these goals, we present OLS2OWL, a plug-in for Protégé 4.0 that facilitates search and retrieval operations against ontology repositories; the plug-in also facilitates successive operations over those retrieved ontologies.

2. OLS2OWL

Currently, ontology engineers often search repositories, one at a time, and retrieve those ontologies they find interesting on a one-by-one basis. Neither repositories nor ontology editors facilitate the extraction of segments/slices of the ontologies. For instance, after having searched over WATSON and SWOOGLE, two independent operations, an ontology engineer has to download those ontologies he/she has found and, once stored locally, manipulate the selected ontologies with an ontology editor. This manipulation may require slicing the ontologies, extracting only those portions we would like to reuse, and integrating them into a new ontology. In a nutshell, available tooling support for managing large quantities of ontologies in the engineering process is very limited, particularly with regard to maximizing reuse of existing definitions and semantic links between related conceptual entities.

OLS2OWL is a plug-in for Protégé 4.0 that allows users to define local and external repositories and to navigate through ontologies; it facilitates the execution of queries across ontology repositories. In order to increase the set of repositories known by OLS2OWL a web service has been developed; this web service can be easily extended so that OLS2OWL can access new repositories.

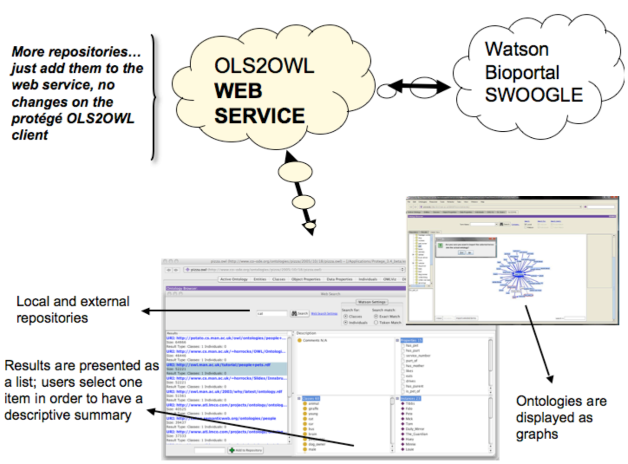

Fig 1. Plug in architecture and basic functionalities

Users select those repositories they wish to include in the query; they are also able to specify if they want to limit the query to instances, properties, classes or a combination of the above OWL constructs. Results include a description of the ontologies based on available metadata and on-the-fly illustrative statistics such as number of classes, properties, and instances. Figure 1 illustrates the architecture of the plug-in as well as some of the functionalities available. OLS2OWL also delivers a direct manipulation interface over a visualization layer; ontologies are displayed as graphs by means of GrOWL (http://www.uvm.edu/~skrivov/growl/).

OLS2OWL makes it easy for users to compare classes by visually inspecting their surroundings. A side-by-side view, as seen in Fig 2, is provided. Finally, the slicing operation allows users to select portions of active ontologies and "import" (facilitated by a drag and drop operation) them into a new blank ontology. This facilitates reusing both entire ontologies as well as sections of them.

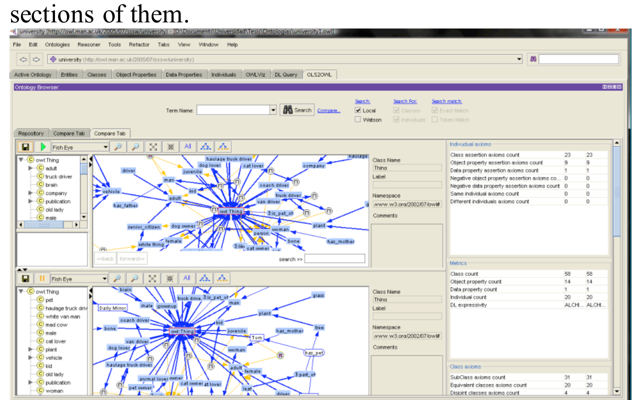

Fig 2. Side-by-side comparison. Here, the pizza ontology (http://www.co-ode.org/ontologies/pizza/2007/02/12/pizza.owl) and the pet ontology (http://www.atl.lmco.com/projects/ontology/ontologies/people+pets/people+petsB.owl)

Discussion and conclusions

Ontology Lookup Services (OLS) are not new. For instance the European Bioinformatics Institute (EBI) OLS [9] facilitates search operations against the EBI Open Biomedical Ontologies (OBO) repository. The Watson Plug-in [10] also facilitates similar operations against the Watson repository. Although these examples allow users to query a specific repository, they do not facilitate querying several repositories in one operation, nor do they support in any special way succesive manipulation operations. Moreover, these tools do not facilitate reuse in any special way, nor do they provide mechanisms for dealing with more than one ontology; for instance, the side-by-side visual inspection facility.

OLS2OWL facilitates search and retrieval operations over ontology repositories by providing a web service that standardises the access to heterogeneous APIs. This web service queries, retrieves and sends results back to the client; this approach inherits the limitations of available APIs: the more expressive the API the more facilities could be provided by OLS2OWL. Repositories indexing ontologies, not storing them, also pose a significant limitation to services such as ours; it often happens that ontologies are no longer available; as indexes are not frequently updated, false positives are then presented to the users in the resulting set.

Although existing ontology repositories aim to provide access to semantic web documents, interestingly each one of them interprets and uses metadata in a different manner. For instance, Swoogle defines three categories of metadata; (i) basic metadata, which considers the syntactic and semantic features of an ontology, (ii) relations, which consider the explicit semantics between individual ontologies, and (iii) analytical results such as SWO/SWDB classification, and ontologies [2]. Both TONES and OntoSelect also rely on structural metadata; however, the use of this metadata is limited to a subset of it. As Bioportal supports the involvement of communities of practice it makes use not only of structural metadata but also of that metadata describing how the community has engaged with the development: for instance, by describing those who have defined a new relationship by means of a marginal note that facilitates establishing confidence rankings. Having a common metadata framework could improve queriability and comparison across ontologies; for instance, finding similar classes across several ontologies could be possible. More complex queries, such as "which is the most commonly used object property for which there is this X domain and this Y range defined within the domain of transport ontologies across K, L, and M repositories", could be executed if we had such a common metadata framework. Some metadata proposals have addressed this issue, for instance the Ontology Metadata Vocabulary (OMV) [11].

As we envision that several repositories of ontologies will be established, interoperability is an important aspect that needs to be preserved. This can be facilitated by having core metadata that facilitate the development of specialized descriptors while maintaining the coherence of the core—thus enabling interoperability. As better metadata become available, OLS2OWL will facilitate the execution of richer semantic queries against existing repositories; in the meantime we are adopting the minimal common metadata denominator—i.e., metadata shared by all repositories.

We have presented OLS2OWL. Our approach aims to bridge the gap between ontology repositories and ontology editors —specifically Protégé. Our Ontology Repository for Assistive Technologies (ORATE) is based on Protégé, a stand alone editor, Web Protégé, supporting collaboration when building ontologies, and Bioportal technology, providing an ontology repository. Our software infrastructure should support the Ontolog vision for a repository; it should also facilitate the automatic comparison of ontology and ontology elements (*e.g.* classes, properties, etc) because reusability is central to our needs. To meet this end we are currently better defining the metadata for ontologies, as well as improving metadata interoperability.

Acknowledgments

Alexander Garcia acknowledges the financial support of the EU Large-scale Integrating Project (IP) OASIS, Grant Agreement # 215754.

References

1. *Wang, D., B. Parsia, and J. Hendler. A survey of the web ontology landscape. in 5th International Semantic Web Conference. 2006. Athens, GA, USA.*
2. *Ding, L., Swoogle: A Search and Metadata Engine for the Semantic Web, in Thirteenth ACM Conference on Information and Knowledge Management. 2004.*
3. *Swoogle.umbc.edu. Swoogle's Statistics of the Semantic Web. 2008 [cited June 2008].*
4. *OntologySummit. Ontology Summit 2008 Communiqué: Towards an Open Ontology Repository 2008 [cited 2008 June]; Available from: http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008_Communique.*
5. *Buitelaar, P., T. Eigner, and T. Declerck. OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. in Demo Session at the International Semantic Web Conference. 2004. Hiroshima, Japan.*
6. *Calvanese, D., TONES – Thinking ONtologiES, Project overview, in Tones Industrial Advisory Team Workshop. 2006: Edinburgh, U.K.*
7. *Rubin, D.L., et al. BioPortal: A Web Portal to Biomedical Ontologies. in AAAI Spring Symposium Series, Symbiotic Relationships between Semantic Web and Knowledge Engineering. 2008: Stanford University Press.*
8. *Pan, J., S. Cranefield, and D. Carter. A lightweight ontology repository. in Second international joint conference on Autonomous agents and multiagent systems. 2003.*
9. *Cote, R., et al., The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics, 2006. 7(97).*
10. *d'Aquin, M. and M. Suarez-Figueroa. A Quick Guide to Knowledge Reuse with the Watson Plugin for the NeOn Toolkit. [cited 2009 April, 2009]; Available from: http://watson.kmi.open.ac.uk/DownloadsAndPublications_files/WNTP-guide.pdf.*
11. *Palma, R., J. Hartmann, and P. Haase, Ontology Metadata Vocabulary for the Semantic Web. 2008, OMV Consortium.*