How to access and reuse ontologies in real-world scenarios

Ghislain Auguste Atemezing
MONDECA
35 boulevard Strasbourg,
Paris, France
ghislain.atemezing@mondeca.com

Boris Villazón-Terrazas
Expert System Iberia
Campo de las Naciones,
Madrid, Spain.
bvillazon@expertsystem.com

Nuria García-Santa
Expert System Iberia
Campo de las Naciones,
Madrid, Spain.
ngarcia@expertsystem.com

ABSTRACT

Developing ontologies, by reusing already available and well-known ontologies, is commonly acknowledge to play a crucial role to facilitate inclusion and expansion of the Web of Data. Some recommendations exist to guide ontologists in ontology engineering, but they do not provide guidelines on how to reuse vocabularies at low fine grained, i.e., reusing specific classes and properties. Moreover, it is still hard to find a tool that provides users with an environment to reuse terms. This paper presents ProtégéLOV, a plugin for the ontology editor Protégé, that combines the access to the Linked Open Vocabularies (LOV) during ontology modeling. It allows users to search a term in LOV and provides three actions if the term exists : (i) replace the selected term in the current ontology; (ii) add the `rdfs:subClassOf` or `rdfs:subPropertyOf` axiom between the selected term and the local term; and (iii) add the `owl:equivalentClass` or `owl:equivalentProperty` between the selected term and local term. Results from a preliminary user study indicate that ProtégéLOV does provide an intuitive access and reuse of terms in external vocabularies.
{TODO: Find the right one here**}**

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.3.5 [**Online Information Services**]: Data sharing— *Web-based services*

General Terms

Accessibility, Linked Data, Vocabularies

Keywords

ontology engineering, ontology reuse, LOV, Protégé, Plugin

1. INTRODUCTION

So far, Linked Data principles and practices are being adopted by an increasing number of data providers, getting as result a global data space on the Web containing hundreds of LOD datasets [4]. There are already several guidelines for generating, publishing, interlinking, and consuming Linked Data [4]. An important task, within the generation process, is to build the vocabulary to be used for modelling the domain of the data sources, and the common recommendation is to reuse as much as possible available vocabularies [4, 5]. This reuse approach speeds up the vocabulary development, and therefore, publishers will save time, efforts, and resources.

There are research efforts, like the NeOn Methodology [7], the Best Practices for Publishing Linked Data - W3C Working Group Note [5], and the work proposed by Lonsdale et al. [6]. However, at the time of writing we have not found specific and detailed guidelines that describe how to reuse available vocabularies at fine granularity level,i.e., reusing specific classes and properties. Our claim is that this difficulty in how to reuse vocabularies at low fine grained level is one of major barriers to the vocabulary development and in consequence to deployment of Linked Data.

Moreover, the recent success of Linked Open Vocabularies (LOV[1]) as a central point for curated catalog of ontologies is helping to convey on best practices to publish vocabularies on the Web, as well as to help in the Data publication activity on the Web. LOV comes with many features, such as an API, a search function and a SPARQL endpoint.

In this paper we propose an initial set of guidelines for this task, and provide technological support by means of a plugin for Protégé , which is one of the popular frameworks for developing ontologies in a variety of formats including OWL, RDF(S), and XML Schema. It is backed by a strong community of developers and users in many domains. One success on Protégé also depends on the availability to extend the core framework adding new functionalities by means of plug-ins. In addition, we propose to explore, design and implement a plug-in of LOV in Protégé for easing the development of ontologies by reusing existing vocabularies at low fine grained level. The tool helps to improve the modeling and reuse of ontologies used in the LOD cloud.

2. REUSING VOCABULARY ELEMENTS WHEN BUILDING ONTOLOGIES

In this section we describe the procedure of reusing available vocabulary terms when building ontologies. In a nut-

---

[1] `http://lov.okfn.org/dataset/lov/`

shell, the task of building vocabularies by reusing available vocabulary terms consists of

- Search for suitable vocabulary terms to reuse from the LOV repository. The search should be conducted by using the terms of the application domain.

- Assess the set of candidate terms from LOV repository. In this particular case the results coming from LOV repository include a score for each term retrieved.

- Select the most appropriate term taking into the account its score.

- Include the selected term in the ontology that has being developed. The selected term will be the external term. There are three alternatives in this case:

  - Include the external term and use it directly in the local ontology by defining local axioms to/from that term in the local ontology.
  - Include the external term, create a local term, and define the `rdfs:subClassOf` or `rdfs:subPropertyOf` axiom to relate both terms.
  - Include the external term, create a local term, and define the `owl:equivalentClass` or `owl:equivalentProperty` axiom to relate both terms.

3. RELATED WORK

The BioPortal Reference Plugin[2] allows the user to insert into the ontology references to external ontologies and terminologies stored in BioPortal[3]. The plugin allows to generate external reference of a selected term. Additionally, the BioPortal Import Plugin[4] allows users to import classes from external ontologies stored in the BioPortal ontology repository. The user can import entire trees of classes with a desired depth and choose which properties to import for each class. However, those plugins work only with Protégé 3.x releases and are not ported yet to recent versions.

Most closely related to the Protégé LOV plugin are approaches that use semantic search engine to support the process of editing an ontology and make large scale knowledge reuse automatically integrated in the tool. An example is the Watson Plugin [2] for the NeOn Toolkit [3], a plugin supporting the NeOn life-cycle management using the Watson [1] APIs[5].However the similar plugin for Protege[6] is just a proof of concept rather than a real plugin.

4. LINKED OPEN VOCABULAIRES (LOV)

The intended purpose of the LOV [8] is to help users to find and reuse terms of vocabularies in Linked Open Data. For achieving that purpose, the LOV gives access to vocabularies metadata and terms using programmatic access with APIs. LOV[7] catalogue is a hub of curated vocabularies used in the Linked Open Data Cloud, as well as other vocabularies suggested by users for their reuse. Some of the three main features of the LOV are for: (1) searching ontologies according to their scope, (2) assessing ontologies by providing a score for each term retrieved by a keyword search and (3) interconnecting ontologies using VOAF vocabulary[8]. The search function uses an algorithm based on the term popularity in LOD and in the LOV ecosystem. The matched resource set for each term should be the value of the a) `rdfs:label` b) `rdfs:comment`, or c) `rdfs:description` property.

Futhermore, the LOV APIs give a remote access to the many functions of LOV through a set of RESTful services[9]. The APIs give access through three different type of services related to: (1) vocabulary terms (classes, properties, datatypes and instances), (2) vocabulary browsing and (3) ontology's creators.

5. EVALUATION

We have conducted an initial user driven evaluation of the tool. For this preliminary evaluation we keep the number of evaluators low; seven ontology engineers used our plugin and gave us an insight on their experience by filling in a questionnaire [10]. The results indicate that while the tool needs some minor improvements, the majority of the evaluators still consider the tool to be useful.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented (1) initial guidelines that describe how to reuse available vocabularies at low granularity level, i.e., by reusing specific classes and properties, and (2) a tool that provides technological support by means of a plugin for Protégé that access LOV API. As future work we plan to fully evaluate the tool and guidelines.

During the demonstration we will show to potential users how to install the plugin and how to quickly build ontologies by reusing terms from LOV repository in Protégé .

7. REFERENCES

[1] M. D'Aquin, C. Baldassarre, L. Gridinoc, M. Sabou, S. Angeletou, and E. Motta. Watson: Supporting next generation semantic web applications. 2007.

[2] M. D'Aquin and M. C. Suárez-Figueroa. A Quick Guide to Knowledge Reuse with the Watson Plugin for the NeOn Toolkit, 2008. `http://watson.kmi.open.ac.uk/DownloadsAndPublications_files/WNTP-guide.pdf`.

[3] P. Haase, H. Lewen, R. Studer, D. Tran, M. Erdmann, M. d'Aquin, and E. Motta. The neon ontology engineering toolkit. *In WWW*, 2008.

[4] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space: Theory and Technology*, volume 1. Morgan & Claypool Publishers, 2011.

[5] B. Hyland, G. Atemezing, and B. Villazon-Terrazas. Best Practices for Publishing Linked Data. W3C Working Group Note, 2014. `http://www.w3.org/TR/ld-bp/`.

---

[2] `http://protegewiki.stanford.edu/wiki/BioPortal_Reference_Plugin`
[3] `http://bioportal.bioontology.org/`
[4] `http://protegewiki.stanford.edu/wiki/BioPortal_Import_Plugin`
[5] `http://watson.kmi.open.ac.uk/WS_and_API.html`
[6] `http://protegewiki.stanford.edu/wiki/Watson_Search_Preview`
[7] `http://lov.okfn.org/dataset/lov/`

[8] `http://lov.okfn.org/vocab/voaf`
[9] `http://lov.okfn.org/dataset/lov/apidoc/`
[10] `http://goo.gl/H4YgBJ`

[6] D. Lonsdale, D. W. Embley, Y. Ding, L. Xu, and M. Hepp. Reusing ontologies and language components for ontology generation . *Data & Knowledge Engineering*, 69(4):318 – 330, 2010. Including Special Section: 12th International Conference on Applications of Natural Language to Information Systems (NLDB07) - Three selected and extended papers.

[7] M.-C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi. *Ontology Engineering in a Networked World*. Springer, Berlin, 2012.

[8] P.-Y. Vandenbussche, G. A. Atemezing, M. Poveda-Villalónc, and B. Vatant. LOV: a gateway to reusable semantic vocabularies on the Web. *Semantic Web Journal (under review)*, 2015. `http://www.semantic-web-journal.net/system/files/swj974.pdf`.