BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 9 2014, pages 1340–1342
doi:10.1093/bioinformatics/btu008

Databases and ontologies

Advance Access publication January 9, 2014

Ontodog: a web-based ontology community view generation tool

Jie Zheng[1,†], Zuoshuang Xiang[2,†], Christian J. Stoeckert, Jr[1] and Yongqun He[2,*]

[1]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 and [2]Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

Associate Editor: Dr Janet Kelso

ABSTRACT

Summary: Biomedical ontologies are often very large and complex. Only a subset of the ontology may be needed for a specified application or community. For ontology end users, it is desirable to have community-based labels rather than the labels generated by ontology developers. Ontodog is a web-based system that can generate an ontology subset based on Excel input, and support generation of an ontology community view, which is defined as the whole or a subset of the source ontology with user-specified annotations including user-preferred labels. Ontodog allows users to easily generate community views with minimal ontology knowledge and no programming skills or installation required. Currently >100 ontologies including all OBO Foundry ontologies are available to generate the views based on user needs. We demonstrate the application of Ontodog for the generation of community views using the Ontology for Biomedical Investigations as the source ontology.

Availability: http://ontodog.hegroup.org/

Contact: yongqunh@umich.edu

Received on July 24, 2013; revised on November 22, 2013; accepted on January 5, 2014

1 INTRODUCTION

A biomedical reference ontology represents an agreed upon understanding of terms and relations between terms in a specific domain and is often developed collaboratively by various communities. Reference ontologies can be very large and complex. With large ontologies, it is not efficient to work with the entire reference ontology for an application when only a part of the ontology is used. For example, the Gene Ontology (GO) (Ashburner, 2000) contains >30 000 terms to support consistent annotation of gene and gene product attributes across various species. For a particular model organism, only a part of GO relevant to that specified organism is needed. The GO slim mechanism has been developed for meeting this need (http://www.geneontology.org/GO.slims.shtml). Slim is a cut-down version of an ontology containing a subset of terms defined in the ontology according to users' needs. An implementation of the slim mechanism using a Perl script is available. However, the script only works for ontologies using the OBO format. The Ontoviews project aims to build application-oriented ontologies based on an 'ontology view', defined as a set of ontology terms retrieved from input ontologies through a SPARQL-based query, and may not be a subset of one ontology (Detwiler and Brinkley, 2011) . That project has not developed tools for users without programming skills to retrieve a subset of an ontology based on their needs.

Ontology term labels are generally chosen for ontological clarity and may not be end user (i.e. biologist) friendly. Here we define 'ontology community view' as a whole or a portion of the ontology that maintains textual and logical term definitions and provides additional user-specified annotations including user-preferred labels. The community views are dependent on the source ontology. With changes in the source ontology, the views might be affected and require updating to reflect the latest changes in the source ontology which may occur frequently. To facilitate ontology community view generation, we have generated a web-based tool Ontodog (http://ontodog.hegroup.org/).

2 FEATURES AND USAGE

Ontodog is implemented using a three-tier system architecture (Fig. 1A). At the front-end, input data are collected using web forms in HTML format. It collects source ontology information and tagged terms of interest. Terms of interest and user-specified annotations, e.g. user-preferred labels, are provided either in a tab-delimited or Excel format file. Ontodog provides a template for generation of this input term file. The template contains all classes and properties in the source ontology. The input data are then processed using PHP and Java with OWLAPI. SPARQL queries are executed in a web server (middle-tier, application server) against an RDF triple store (back-end, database server) for subset extraction. The default RDF triple store used in Ontodog is the SPARQL endpoint for the RDF triple store hosted by He Group (http://sparql.hegroup.org). This RDF triple store contains all the OBO Foundry ontologies and automatically reloads any updates of ontologies every day. The Ontodog outputs are RDF/XML files that can be viewed and downloaded. As a web-based system, Ontodog is accessible anywhere with internet, and does not require any software installation. Ontodog uses OWL-DL as the default ontology format.

The OntoFox SPARQL-related term-retrieval approach (Xiang, 2010) is adopted in Ontodog for ontology subset extraction to retrieve terms of interest and all relevant terms based on logical axioms. The retrieved subset maintains both textual and logical definitions of terms and preserves reasoning consistency.

New annotation properties, such as community preferred labels, are kept in a separate ontology file. The extra annotation property layers can be imported to the source ontology or a

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

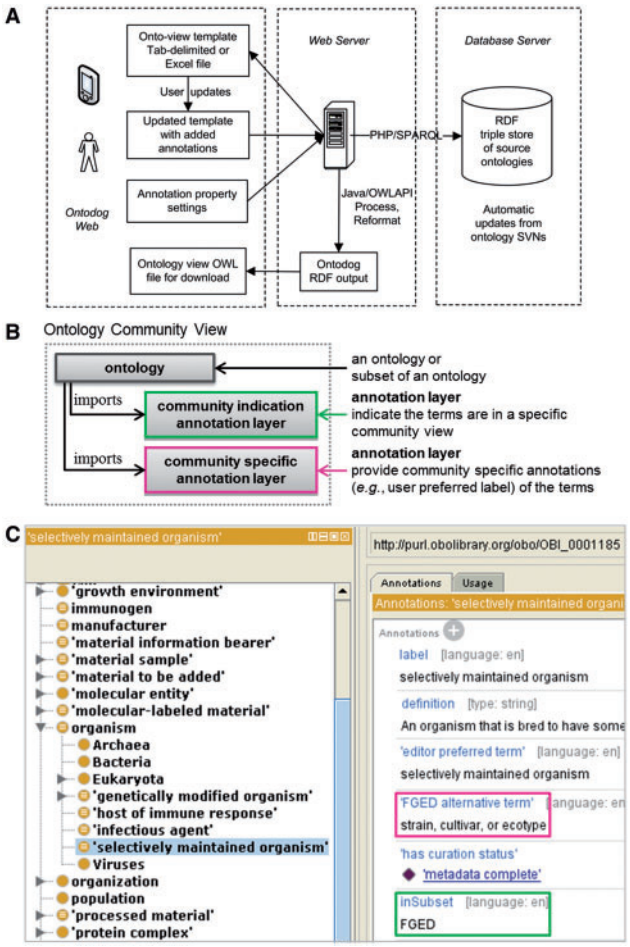

Fig. 1. (A) System architecture of Ontodog. (B) Ontology community view built by importing various annotation layers. (C) FGED community view of OBI illustrating a community preferred label (under the term 'FGED alternative term') and annotation of inSubset

subset of it using the OWL import mechanism (Fig. 1B). The layer approach allows the annotation properties to be reused and provides extra flexibilities based on a user's needs. Ontodog allows users to tag terms of interests using annotation, obo:inSubset, which has been used by GO slim to indicate that a term is a member of a subset of interest to a specific application or community and enable subset extraction using logical queries based on annotation. In addition, the annotation values are not limited to the English language. Ontodog allows users to add labels in four other languages, including Chinese, French, German and Spanish. Details of Ontodog output files and how to use them are provided on the Ontodog tutorial page: http://ontodog.hegroup.org/tutorial/index.php.

Ontodog can be used to generate an ontology subset for new ontology development. For example, Ontodog was recently used to retrieve a subset of the Ontology for Biomedical Investigations (OBI) (Brinkman, 2010) as a basis to build the Beta Cell Genomics Ontology (BCGO) (Zheng, 2013). OBI has been developed collaboratively by >20 different communities to support consistent annotations of all aspects of biological investigations. For each application/community, a portion of OBI is needed with many user-preferred labels (instead of default OBI labels) desired. Ontodog has a unique feature of using an Excel input file.

A more advanced feature of Ontodog is the support of community view generation. For example, Ontodog was used to generate a community view of OBI for the Functional Genomics Data Society (FGED) (http://www.fged.org/). For the FGED community users, only a subset of OBI is needed for description of their experiments. In addition, these users want to have their own labels for terms. For example, 'selectively maintained organism' is used to describe a population of organisms of the same species with distinctive characteristics and generally obtained using breeding. The label given in OBI reflects the consensus choice of developers from different communities. However, it is less meaningful for most biology users than the terms 'strain, cultivar, or ecotype'. The FGED community view of OBI was generated by Ontodog (http://bioportal.bioontology.org/ontologies/43109). In total, 2279 classes were extracted from OBI (which has 3501 classes). The community view was built by importing an inSubset annotation properties OWL file (added inSubset annotation of 'FGED') and a community preferred label properties OWL file (added FGED alternative term for a term) into the retrieved subset of OBI (Fig. 1C). The two OWL files can also be exported to the source ontology OBI to show OBI users which terms and labels the FGED community would prefer to use. It is noted that one term may have many meanings for different communities. The use of community-preferred labels will solve the problem.

Ontodog was also used to generate an OBI community view for the OBI web service community (Guttula, 2011). The community view contains 228 OBI classes. The reduced size and complexity of this community view led to improved efficiency in annotation and semantic similarity computation.

3 SUMMARY

Ontodog is a web system that generates an ontology view with customized annotation. It is easy to use by ontology developers, including those who are mostly domain experts and who might not have deep ontology knowledge or programming skills. Currently, Ontodog only supports OBO Foundry ontologies.

ACKNOWLEDGEMENTS

We thank Drs Carlo Torniai, Bjoern Peters and OBI Consortium for their valuable discussions and feedback.

Funding: National Institute of Health (grants R01AI081062 and 5R01GM93132); Rackham Pilot Research grant in the University of Michigan.

Conflict of interest: None declared.

REFERENCES

Ashburner,M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

Brinkman,R.R. *et al.* (2010) Modeling biomedical experimental processes with OBI. *J. Biomed. Semantics.*, **1** (**Suppl. 1**), S7.

Detwiler,L.T. and Brinkley,J.F. (2011) Value sets via ontology views. In: *Proceedings, Clinical Research Informatics Summit of the American Medical Informatics Association*, p. 100, American Medical Informatics Association, San Francisco, CA.

Guttula,C. *et al.* (2011) Enriching the Ontology for Biomedical Investigations (OBI) to improve its Suitability for Web Service Annotations. In: *International Conference on Biomedical Ontology (ICBO)*. Vol. 833, Buffalo, New York, pp. 246–248.

Xiang,Z. *et al.* (2010) OntoFox: web-based support for ontology reuse. *BMC Res. Notes*, **3**, 175.

Zheng,J. *et al.* (2013) Development of an application ontology for beta cell genomics based on the ontology for biomedical investigations. In: *International Conference on Biomedical Ontologies (ICBO 2013)*. Montreal, Canada. http://ceur-ws.org/Vol-1060/.