

Ontobee: A Linked Data Server and Browser for Ontology Terms

Zuoshuang Xiang¹, Chris Mungall², Alan Ruttenberg³, Yongqun He¹

¹University of Michigan Medical School, Ann Arbor, MI, USA

²Lawrence Berkeley National Laboratory, Berkeley, CA, USA

³Science Commons, Cambridge, MA, USA

Abstract. The Linking Open Data (LOD) community has been working to extend the Web with a data commons by publishing various open datasets as RDF on the Web. To support this effort, we developed Ontobee (<http://www.ontobee.org/>), a web system aimed to serve as a linked data server and browser specifically targeted for ontology terms. Ontobee combines two basic features for one specific ontology term: (1) a user-friendly web HTML interface for displaying the details and its hierarchy of a specific ontology term; and (2) a RDF/XML source code for this ontology term corresponding to the HTML web page, which can be accessed by Semantic Web applications. Ontobee provides an efficient and publicly available method to promote ontology sharing, interoperability, data integration, and Semantic Web applications.

Keywords: Ontobee, ontology visualization, Semantic Web

1 Introduction

Biomedical ontologies are consensus-based controlled biomedical vocabularies of terms and relations with associated definitions, which are logically formulated to promote automated reasoning. The Semantic Web is a web of data that allows machines to understand the meaning – or “semantics” – of information on the World Wide Web. To ensure that computers can understand the semantics of terms, machine-readable ontologies play a central role in Semantic Web development. However, how to present the meaning of ontology terms by its URI is still a challenge. Most ontology URIs do not point to real web pages. While some URIs point to specific pages, the pages shown are often in: (a) pure HTML format (*e.g.*, the Ontology Lookup Service or OLS [1] and NCBO BioPortal) without a RDF source code, or (b) RDF format (*e.g.*, many ontology term URLs such as:

http://purl.org/obo/owl/SO#SO_0001411

which contains the whole ontology instead of individual terms. In both cases these pages do not efficiently support the Semantic Web.

The objective of the Linking Open Data (LOD) community is to extend the Web with a data commons by publishing various open

datasets as RDF on the Web. These RDF links between data items can come from different data sources and accessed anywhere through the web. All of the sources on these LOD diagrams are open data. To support LOD, one basic requirement is to map individual ontology terms to real RDF files through the Web. Many LOD browsers are available *e.g.* Ontology-browser, VisiNav, and others listed at <http://x.co/XjWT>. However, these programs focus on RDF semantic data structure browsing without returning fragmented ontology term information in RDF format.

In this report, we present our development of Ontobee (<http://www.ontobee.org/>), a linked data server and browser for linking and browsing individual ontology terms. Ontobee integrates both ontology content visualization through HTML web pages and Semantic Web content transferring through RDF/XML format.

2 Methods

2.1 Server and Programming Tools

The Ontobee application server is implemented using one Dell Poweredge 2580 server. This server runs the Red Hat Linux operating system (Red Hat Enterprise Linux 5 server).

The open source software program, Apache HTTP Server, is installed as the HTTP application server. PHP is used as programming languages in the web application server.

2.2 Ontology Source Used in Ontobee

The primary ontology source comes from Neurocommons (<http://neurocommons.org>), a knowledge base containing all the ontologies in OBO foundry [2]. Ontobee also supports ontologies from other data sources. For example, Ontobee also maintains a Virtuoso RDF store that contains several ontologies such as the Vaccine Ontology (VO) and the Adverse Event Ontology (AEO).

2.3 Visualization of Ontology Hierarchies and Individual Terms

A software program based on PHP and SPARQL was developed to visualize the hierarchical tree and individual terms from a specific ontology. The ontology tree is assembled from the results of a set of SPARQL queries against the RDF stores. The 'transitive' option in the Virtuoso SPARQL engine was used to minimize the number of SPARQL queries. For individual terms, a HTML page is assembled from the results of another set of SPARQL queries against the RDF stores. The 'transitive' and the 'CBD' (*i.e.*, Concise Bounded Description) options in the Virtuoso SPARQL engine were used to minimize the number of SPARQL queries. In both cases, PHP interacts with the RDF stores by sending SPARQL queries to the stores. The stores then return the results back to PHP scripts in JSON (JavaScript Object Notation) format. The PHP scripts decode information from the JSON formatted results and format it into a user-friendly HTML page.

2.4 RDF Output of Ontology Term Information

In addition to a HTML file, another file is generated to output the ontology term content in OWL (RDF/XML) format. This OWL format is the default page format for the term URI. A link to the HTML file is provided in the OWL format through XSLT (Extensible Stylesheet Language Transformations; <http://www.w3.org/TR/xslt>). When the term URI was visited by

using a web browser, both the OWL file and the HTML content will be retrieved but only the HTML content will be shown in the browser for easy reading. A user can easily access the OWL content by checking the source code of the HTML page. Alternatively, without a human visit to the HTML page, the OWL file can be retrieved by a web application or a Semantic Web system.

3 Results

3.1 Software Design and Statistics

The Ontobee software integrates two basic features: ontology content visualization, and RDF content sharing (Fig. 1). Basically, a user can use a web browser to query Ontobee web page. Ontobee will then issue SPARQL queries to a RDF triple store. The results of the SPARQL queries are returned in JSON format and processed by Ontobee PHP scripts to form a user-friendly HTML page. Simultaneously, RAW OWL output was also returned and used to generate machine-readable RDF output file. A web application can be developed to access the RDF output file directly without access using a web browser.

Currently Ontobee provides access to 88 ontologies, the majority of which come from the OBO Library. These ontologies cover different domains, such as anatomy, health, and experiments. In total, 759,003 ontology terms from 88 ontologies are currently covered.

3.2 Ontobee Features

Fig. 2 is an example of using Ontobee for the term 'vaccine' (VO_0000001) in the Vaccine Ontology (VO). The HTML web page displays the term information (*e.g.*, term equivalents and class hierarchy). Once an ontology term is clicked, the detailed information about the ontology term will be displayed in another web page. The source page is generated using RDF/XML format (Fig. 2B). The second line of the RDF/XML file specifies an XSLT stylesheet in which the HTML code is embedded. The URL of the stylesheet specifies the label abbreviation of a particular ontology (*e.g.*, VO) and an IRI of a specific term (*e.g.*, 'vaccine') in this ontology. Some raw output from SPARQL queries was reformatted by Ontobee to minimize the file size.

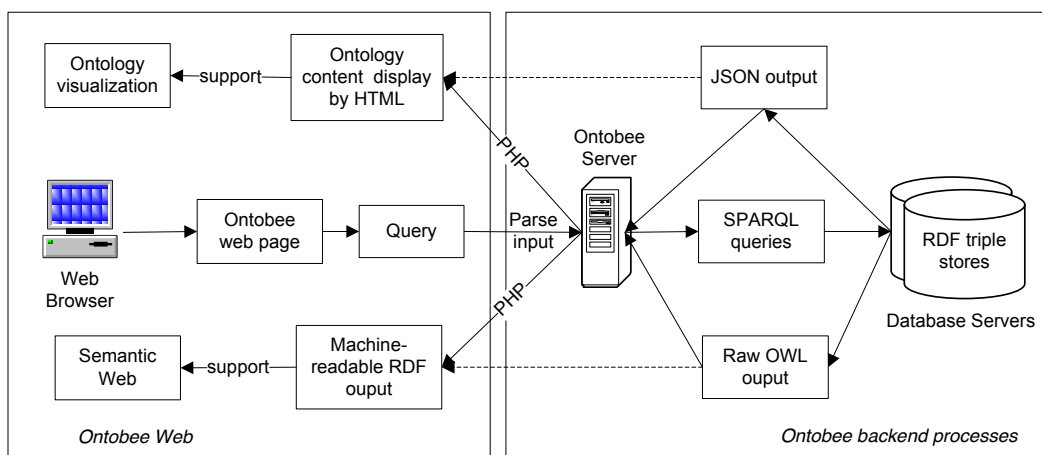


Figure 1. Ontobee architecture design.

(A)

(B)

Figure 2. Screenshots of Ontobee display of a VO term 'vaccine' in Firefox web browser. (A) The web HTML display of this term. (B) Source view of this HTML page.

4 Discussion and Summary

The problem of identity and reference of ontology terms on the web has been discussed for long in and outside WWW circles. Ontobee is an attempt to unify biomedical ontology visualization and ontology term transfer based on RDF/XML format. Ontobee provides a useful tool to support the LOD and the Semantic Web.

Acknowledgments

This project is supported by NIH grant 1R01AI081062.

References

1. Cote RG, Jones P, Martens L, Apweiler R, Hermjakob H. The Ontology Lookup Service: more data and better tools for controlled vocabulary queries. *Nucleic Acids Res* 2008 Jul 1;36(Web Server issue):W372-6.
2. Ruttenberg A, Rees JA, Samwald M, Marshall MS. Life sciences on the Semantic Web: the Neurocommons and beyond. *Brief Bioinform* 2009 Mar;10(2):193-204.