

Ontorat Web Server for Automatic Ontology Term Generation and Annotations

Zuoshuang "Allen" Xiang, Yu "Asiyah" Lin, and Yongqun "Oliver" He



Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, and Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Tel: (734) 615 8231
yongqunh@umich.edu
http://www.hegroup.org

Abstract

Ontorat (<http://ontorat.hegroup.org/>) is a web server that automatically generates new ontology terms and annotates new and existing ontology terms with axioms based on user-provided input in a spreadsheet format (Excel or tab-delimited file). The development of Ontorat is based on the Ontology Design Patterns (ODPs) in the field of ontology engineering and is inspired by the Quick Term Templates (QTT) procedure generated by the developers of the Ontology for Biomedical Investigations (OBI). Ontorat implements the QTT procedure. The use of Ontorat is easy and is supported by an interactive web interface. Ontorat is an efficient method for ontology expansion. As a demonstration, Ontorat has recently been applied for automatically generating terms in the Vaccine Ontology (VO) (1) for 798 licensed animal vaccines.

Introduction

The Web Ontology Language (OWL) has been widely used for ontology development. However, development of a new ontology in OWL is often time consuming and requires specialized knowledge in ontology, OWL, and specific domain science. One frequently observed phenomenon during ontology development is that often we need to create a large number of new ontology terms that follow the same design patterns of logical definitions and axioms. Manual addition of these terms is time consuming, error prone, and often boring. Ontorat is developed to facilitate this process.

Ontorat is developed based on the Ontology Design Patterns (ODPs) in the field of ontology engineering. Ontorat uses the class expressions with variables, which is defined as "OWL class expressions but allowing variables at positions of class expressions. The range of a variable can either be a named class, class expression, or any subtype of it as produced by the corresponding rule in OWL." (2). Ontorat offers a web-based platform for writing up class expressions with variables with the aim to quickly generate a large number of new ontology classes or annotate new and existing classes in a specific target ontology.

The development of Ontorat is also inspired by the Ontology for Biomedical Investigations (OBI) project QTT (Quick Term Templates) (Rocca-Serra, et al., 2011). The OBI QTT procedure is based on the concept of a Quick Term Template (QTT), a template in a spreadsheet format that can be converted into an OWL file. Ontorat implements the QTT procedure.

Development Methods

Ontorat was developed using one HP Linux server. The programming languages PHP and HTML were used to generate the Ontorat web form and analysis program. The Manchester OWL syntax is used for OWL format representation. The OWLAPI library (<http://owlapi.sourceforge.net/>) was used for output reformatting. Ontorat is freely available for public use.

Online Query and Analysis

Ontorat provides a user-friendly and interactive web interface for target ontology specification, data input, and execution (Figure 1). A specific web form is designed for a user to specify different settings for executing Ontorat. The specified settings can also be saved as an Ontorat setting file, which can be used as a setting template and reused in the future.

Figure 1. User-friendly Ontorat web interface.

- The steps using the web form to specify input include:
- 1) Specify a target ontology in OWL format. The target ontology is used to avoid the assignment of repeated ontology identifiers (IDs).
 - 2) Specify the data file (an Excel file or tab-delimited text file). This file provides the input data for high throughput processing. Each column of the data file needs to be very specific. The target ontology and user-provided data files will be parsed by the Ontorat parser.
 - 3) Purpose of the new axiom generation, including generating new ontology classes or providing annotations for existing ontology terms.
 - 4) Annotations. The ontology design pattern is generated using the Manchester OWL Syntax. An internally designed code is used to represent different columns (i.e., variables). Specifically, we use $\{ \$columnA \}$ to represent the first column (or column A), and use $\{ \$columnB \}$ to represent column B, et al. Each column represents a variable.
 - 5) Equivalent classes. The Manchester OWL Syntax is used to express an equivalent class.
 - 6) Superclasses using the Manchester OWL Syntax.
 - 7) Terms used to define annotations, equivalent classes, and superclasses. Those terms may be typed using term labels. In this section, the term labels are specified by ontology URI. This way Ontorat can map the labels with specific ontology URIs.
 - 8) Term URI: Specify a URI as the start of an ontology.
 - 9) For auto-generated term IDs. Three data items are needed: prefix, number of digits, and start number. The information will be used for automatically generating unique term ID.

Use Case: Application of Ontorat in VO development

Ontorat has been used to facilitate the development of the community-based Vaccine Ontology (VO) (He, et al, 2009). Hundreds of licensed animal vaccines are used in the USA. Ontorat is able to import these animal vaccines and their annotations into VO efficiently (Figure 2).

A	B	C	D	E	F	G
1	Vaccine name	USA ID	Parent term label	Parent term ID	Pathogen name	Pathogen Host name
2	Pigeon Pox Live Virus Vaccine	USDA: 1881.10	Pigeonpox virus vaccine	VO_0001481	Pigeonpox virus	10264 Birds
3	Fox Encephalitis Killed Virus Vaccine	USDA: 1635.20	Canine adenovirus type 1 vaccine	VO_0001489	Canine adenovirus 1	10512 Carnivores
4	Feline calicivirus Killed Virus Vaccine	USDA: 15C5.21	Feline calicivirus vaccine	VO_0001509	Feline calicivirus	11978 Cat
5	Anthrax Spore live culture Vaccine	USDA: 1011.00	Bacillus anthracis vaccine	VO_0000013	Bacillus anthracis	1392 Cattle
6	Avian Encephalomyelitis Killed Virus Vaccine	USDA: 1015.10	Avian Encephalomyelitis Virus vaccine	VO_0001522	Avian encephalomyelitis	70796 Chicken
7	Duck Virus Hepatitis Modified Live Virus Vacc	USDA: 1481.10	Duck hepatitis virus 1 vaccine	VO_0001528	Duck hepatitis A virus	69156 Ducks
8	Flavobacterium Columare Avirulent Live Cu	USDA: 1771.00	Flavobacterium columnare vaccine	VO_0000937	Flavobacterium columnar	996 Fish
9	Bordetella Bronchiseptica Avirulent Live Cu	USDA: 1081.00	Bordetella bronchiseptica vaccine	VO_0000994	Bordetella bronchiseptica	518 Gray wolf
10	Equine influenza Killed Virus Vaccine	USDA: 1505.10	influenza virus vaccine	VO_0000642	influenza virus	11309 Horse
11	Erysipelothrix Rhusopathiae Avirulent Live	USDA: 1541.00	Erysipelothrix rhusiopathiae vaccine	VO_0001456	Erysipelothrix rhusiopath	1048 Pig
12	West Nile Virus Killed Virus Vaccine	USDA: 1995.20	West Nile virus vaccine	VO_0000752	West Nile virus	11062 Set whale
13	Bluetongue Modified Live Virus Vaccine	USDA: 1061.00	Bluetongue virus vaccine	VO_0001515	Bluetongue virus	40051 Sheep
14	Bordetella Avium Avirulent Live Culture Vacc	USDA: 1021.01	Bordetella avium vaccine	VO_0001114	Bordetella avium	521 Turkey
15	Escherichia Coli Avirulent Live Culture Vacc	USDA: 1551.02	Escherichia coli vaccine	VO_0000941	Escherichia coli	562 Turkey
16	Porcine Circovirus Type 2, Killed Baculovirus	USDA: 1962.80	Porcine circovirus 2 vaccine	VO_0001034	Porcine circovirus 2	80768 Pig

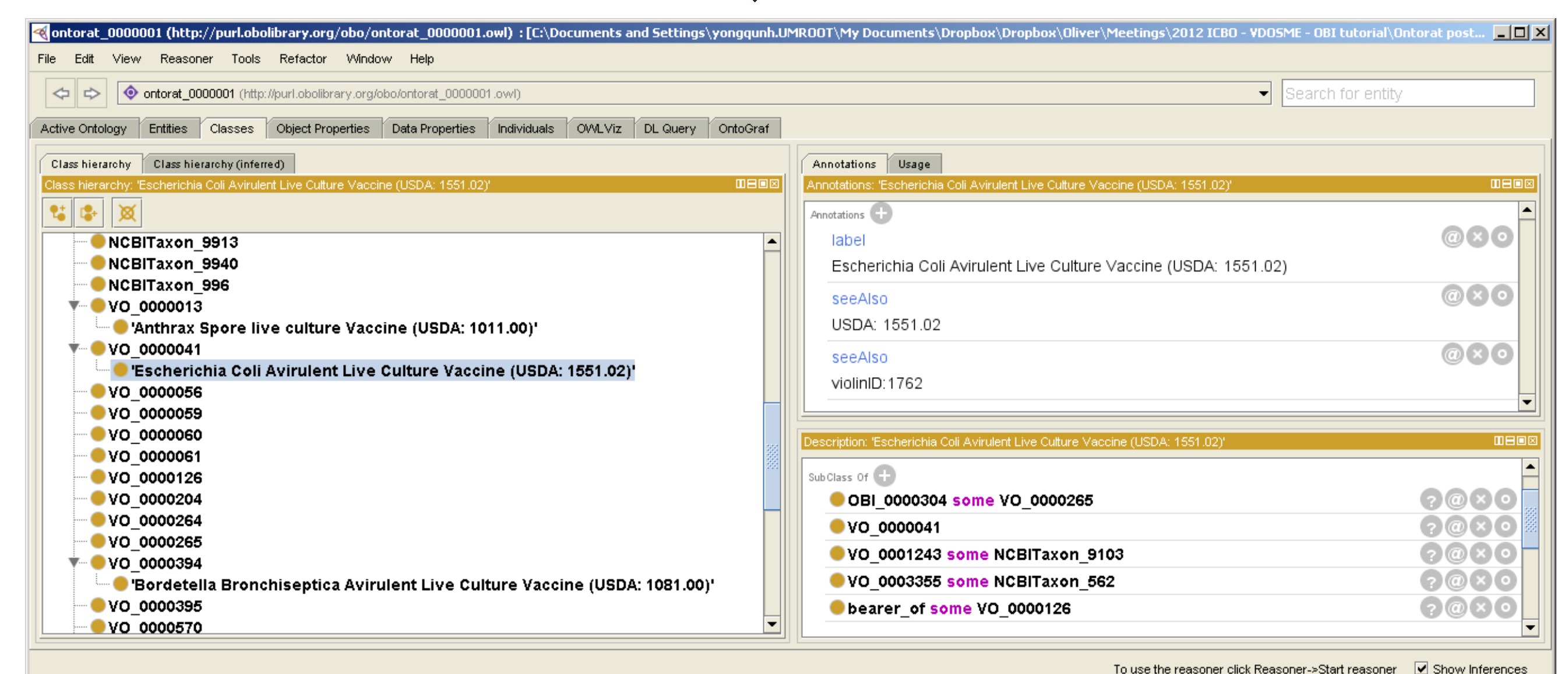
(A) Specify target ontology

(B) Specify input data file (here in Excel format)

(C) Generate settings using Manchester syntax (see Figure 1)

Retrieving Results
Finished the Ontorat execution. Please download the output file.

(D) Execute and retrieve output OWL file



(E) OWL output displayed and merged to target ontology using Protege.

Figure 2. The Ontorat operational steps used in the use case study.

Discussion

The Ontorat web server provides a robust and scalable approach for automatically generating new ontology terms and related axioms or annotating existing ontology terms with new axioms. Ontorat supports efficient ontology expansion with new terms and ontology enrichment with new annotations. The design patterns can be standardized and reused, allowing domain experts and data curators to contribute actively to the ontology development without knowing the specifics of OWL.

The QTT procedure has been implemented in MappingMaster, a plugin program in the Protégé-OWL editor. Similar to MappingMaster, Ontorat uses an Excel or tab-delimited spreadsheet file as input. Ontorat also relies on the pattern program using the OWL Manchester Syntax, a new syntax designed for writing OWL class expressions (<http://www.w3.org/TR/owl2-manchester-syntax/>). Different from MappingMaster, Ontorat is implemented as a web-based application. The MappingMaster plugin works in Protege 3.4 but does not work in Protege 4.x. Ontorat provides a novel feature of automatic assignment of unique ontology identifiers. In addition, Ontorat, but not MappingMaster, can generate annotations for ontology terms.

Acknowledgements

This work is supported by NIH-NIAID Grant 1R01AI081062 to YH.

References

1. Ozgur A, Xiang Z, Radev D, He Y. Mining of vaccine-associated IFN- γ gene interaction networks using the Vaccine Ontology. *Journal of Biomedical Semantics*. 2011, 2(Suppl 2):S8.
2. Noppens Y and Liebig T. Ontology Patterns and Beyond Towards a Universal Pattern Language. *Proceedings of the Workshop on Ontology Patterns (WOP 2009)*, collocated with the 8th International Semantic Web Conference (ISWC-2009), Washington D.C., USA.
3. Rocca-Serra P, Ruttenberg A, O'Connor MJ, Whetzel T, Schober D, Greenbaum J, Courtot M, Sansone SA, Scheurmann R, Peters B. *Journal of Applied Ontology*. 2011, 6, 13-22.