

Easy Extraction of Terms and Definitions with OWL2TL

John Judkins, Joseph Utecht, and Mathias Brochhausen

Department of Biomedical Informatics
University of Arkansas for Medical Sciences
Little Rock, USA
mbrochhausen@uams.edu

Abstract—Facilitating good communication between semantic web specialists and domain experts is necessary to efficient ontology development. This development may be hindered by the fact that domain experts tend to be unfamiliar with tools used to create and edit OWL files. This is true in particular when changes to definitions need to be reviewed as often as multiple times a day. We developed "OWL to Term List" (OWL2TL) with the goal of allowing domain experts to view the terms and definitions of an OWL file organized in a list that is updated each time the OWL file is updated. The tool is available online and currently generates a list of terms, along with additional annotation properties that are chosen by the user, in a format that allows easy copying into a spreadsheet.

Keywords—ontology development; semantic web; software; web application; controlled vocabulary

I. INTRODUCTION

Ontology development requires ongoing collaboration between domain experts from various fields, programmers and ontology developers. After the initial phase of domain analysis quite often an iterative process during which definitions are vetted by ontology developers to fit certain requirements [1] and then reviewed and agreed upon by the domain experts is necessary. We encounter this situation in the NIH funded project "Comparative Assessment Framework for Environments of Trauma Care" (CAFÉ, 5R01GM111324) [www.cafe-trauma.com]. It is frequently the case that domain experts are not familiar with ontology development tools, such as e.g. Protégé [2], if they are not themselves creating OWL files. Hence, the process of optimizing definitions can become tedious as changes implemented in OWL need to be fed back to the domain experts, especially during a project phase where multiple – often minimal – edits per day occur. To allow domain experts to review all terms and definition in an OWL file as soon as updates get available in a format that is familiar to them, we provide a tool called "OWL to Term List" (OWL2TL). OWL2TL is a web application that is part of the CAFE project, hosted at [cafe-trauma.com/owl2tl/], and has source available at [github.com/johnwjudkins/OWL2TL]. While an initial release has been made, we will continue to tweak, optimize, and expand the tool. The goal of OWL2TL is the quick generation of a list of terms and definition from an OWL file. OWL2TL provides a web form to enter the URI for the ontology itself and URIs for additional annotation

properties of terms (e.g. rdfs:label, etc.) and definitions (rdfs:comment, iao:definition, etc.). When the form is submitted, a list of terms and annotation properties is produced.

II. METHODS

Accessing [cafe-trauma.com/owl2tl/] loads a form (Figure 1) for the user to input URIs for: the OWL ontology (required) and five additional annotation properties (e.g. http://purl.obolibrary.org/obo/IAO_0000115 for definitions). Terms are extracted using a SPARQL query that extracts each class by its RDFS label. The query also retrieves the values for annotation properties specified by the user on the form, if any. The array containing the results of the query is posted to a web form – Figure 2 shows as an example term list from the Non-Coding RNA Ontology [3] with definitions (iao:definition) and alternative terms (iao:alternativeTerm).

III. RESULTS

OWL2TL currently generates a list of terms and additional annotation properties for any ontology in which owl:Class introduces a class description, as Figure 2 shows. The list loads in a format that allows easy copying into a spreadsheet while preserving the structure of the table, and even works for ontologies with hundreds of thousands of terms.

Fig. 1. Screenshot of results after form submission.

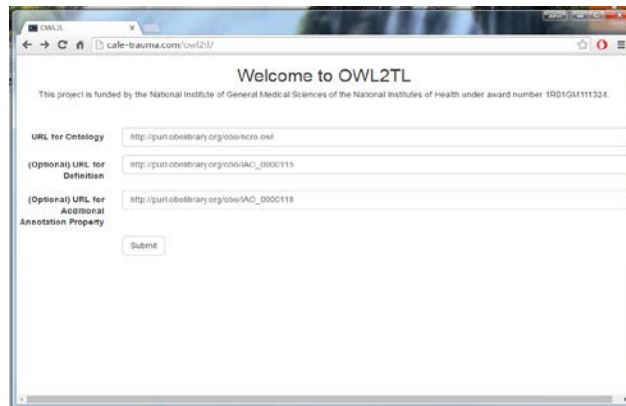


Fig. 2. Screenshot of [cafe-trauma.com/owl2tl/].

Term	Definition	Additional Annotation Property
3'-5'-exoribonuclease activity	Catalysis of the sequential cleavage of mononucleotides from a free 3' terminus of an RNA molecule.	3'-5'-exoribonuclease activity
3'-5'-exoribonuclease activity involved in mature mRNA 3'-end processing	Catalysis of the sequential cleavage of mononucleotides from a free 3' terminus of an RNA molecule that contributes to forming distinct mRNA isoforms from a mature mRNA.	None
CRMs	A regulatory region where transcription factor binding sites clustered to regulate various aspects of transcription activities. (CRMs can be located a few kb to hundred kb upstream of the basal promoter, in the coding sequence, within introns, or in the downstream 3'UTR sequences, as well as on different chromosome). A single gene can be regulated by multiple CRMs to give precise control of its spatial and temporal expression. CRMs function as nodes in large, interrelated regulatory network.	transcription factor module
CRMs	A regulatory region where transcription factor binding sites clustered to regulate various aspects of transcription activities. (CRMs can be located a few kb to hundred kb upstream of the basal promoter, in the coding sequence, within introns, or in the downstream 3'UTR sequences, as well as on different chromosome). A single	cis regulatory module

IV. DISCUSSION AND CONCLUSION

The main goal of OWL2TL has been achieved – the application generates a list of terms and definitions from any ontology that has up to one thousand terms and uses owl:Class for class descriptions. The next step in development is editing the code to redirect the generated term list to a unique URL that can be shared, while maintaining the same functionality and quick response. Efforts to make the code more versatile – for instance, extracting only terms belonging to a particular superclass in an ontology – are underway.

ACKNOWLEDGMENT

The research presented in this demonstration is funded by the National Institute of General Medical Sciences of the National Institutes of Health under award number 1R01GM111324.

REFERENCES

- [1] J. Köhler, K. Munn, A. Rügge, A. Skusa, and B. Smith, “Quality control for terms and definitions in ontologies and taxonomies”, *BMC Bioinformatics*, 2006, 7: 212, PMC1482721
- [2] N. F. Noy, M. Crubézy, R. W. Fergerson, H. Knublauch, S. W. Tu, J. Vendetti, M. A. Musen, “Protégé-2000: An open-source ontology-development and knowledge-acquisition environment”, *AMIA Annu. Symp. Proc.*, 2003, 2003: 953, PMC1480139
- [3] J. Huang, K. Eilbeck, B. Smith, J. A. Blake, D. Dou, W. Huang, D. A. Natale, A. Rutenberg, J. Huan, M. T. Zimmermann, G. Jiang, Y. Lin, B. Wu, H. J. Strachan, Y. He, S. Zhang, X. Wang, Z. Liu, G. M. Borchert, “The non-coding RNA ontology (NCRO): A comprehensive resource for the unification of non-coding RNA biology.” *J. Biomed. Semantics*, 2016, 7:24, PMC4857245