Where does an Ontology start?

Henriette Harmse[a], Melanie Courtot[a], Warren Read[a], Nico Matentzoglu[a], David Osumi-Sutherland[a], Simon Jupp[a]

[a] *The European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK*

Abstract

A good ontology design practice is to define terms based on terms from upper ontologies. Hence, the terms of the current ontology often are found deep down the hierarchy, which users can find confusing. Here we suggest ways to deal with this.

Keywords:

visualisation, tree, roots

Introduction

Biomedical ontologies have become important in the biological and medical domains through their provisioning of common controlled vocabularies. A model of good practice guiding the design of biomedical ontologies is to define terms based on terms from common higher level ontologies (1). In the OBO community the use of upper-level ontologies, such as BFO (1), is recommended. While upper-ontologies bring value for the interoperability of ontologies, they often contain terms that are hard to understand for non-ontologists (2). Yet these terms are often the first thing a user is presented with when exploring an ontology, especially if the ontology class hierarchy is being presented using a standard tree-based visualisation.

To alleviate this, we propose a standard annotation property to indicate to visualisation tools what is the set of classes that should be considered root concepts of the ontology. Online ontology browsers and visualisation tools can use this information to render the ontology and highlight where the root concepts sit in the hierarchy. To illustrate the issues we present an example from the Data Use Ontology (DUO) and support for a new "hasPreferredRootTerm" annotation property for ontologies hosted by the EMBL-EBI Ontology Lookup Service (OLS).

The DUO provides standard codes to indicate the secondary use restrictions and conditions on scientific data sharing. It imports a number of ontologies which are used as the basis on which terms in DUO are defined. DUO is available via OLS and BioPortal, and both browsers present the DUO root classes of "Entity" from BFO and oboInOwl:ObsoleteClass by default. Several users have commented that they find it difficult to find the DUO codes when navigating the ontology as these are buried six levels deep in the hierarchy under the concept of "consent code" and "data use requirement".

Implementation

OLS has been extended support the use of a new OWL annotation property with the IRI http://www.ebi.ac.uk/ols/vocabulary/hasPreferredRootTerm which can be used on the ontology to assert the set of class IRIs that can be considered as roots. OLS will detect this property and change how the class hierarchy is rendered by default to only show the root terms, as specified by the property. The full classification is preserved internally, and users are able to easily switch (via radio button selection) to see the fully expanded classification if they wish.

Discussion

The interest and use of ontologies continues to increase in the life sciences, but we need to be able to effectively present and communicate these ontologies to a wide audience of users. In some cases this means hiding the complexity of the underlying machinery, such as OWL or upper-level ontologies. With over 65 ontologies in OLS alone that start with a root concept of "Entity", our simple proposal could significantly improve the way these ontologies are presented to users.

We are exploring additional ways to simplify the presentation of root terms that include detecting root terms based on OBO namespaces of the presence of terms in the OBO Core[1]. There may also be other more appropriate vocabularies for expressing this information within an ontology, such as the use of SKOS concept schemes. At this stage we are open to working with the community to collect more use-cases and welcome feedback on the approach presented here.

References

1. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: Coordinated Evolution of Ontologies to support Biomedical Data Integration. Nature biotechnology. 2007;25(11):1251-5.

2. Stevens R, Lord P, Malone J, and Matentzoglu N. Measuring Expert Performance at Manually Classifying Domain Entities under Upper Ontology Classes: 2018; abs/1810.05093.

[1] https://github.com/OBOFoundry/Experimental-OBO-Core