A Call for an Abductive-Reasoning Feature in OWL-Reasoning Tools toward Ontology Quality Control

Michael Bada[1], Chris Mungall[2], Lawrence Hunter[1]

[1] University of Colorado Denver, Department of Pharmacology, MS 8303, RC-1 South, 12801 East 17th Avenue, L18-6400A, P.O. Box 6511, Aurora, CO 80045, USA
{mike.bada | larry.hunter}@uchsc.edu
[2] 240C Building 64, Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, CA 94720, USA
cjm@fruitfly.org

It is widely touted that OWL reasoners are able to improve the quality of ontologies by making new inferences from and detecting inconsistencies among the assertions that have been explicitly stated in the ontologies. These reasoner actions are based on standard deductive reasoning, which operates on the principle that assertions inferred from premises that are assumed to be true also must be true. Deductive reasoning is similarly the basis for OBO-Edit, the primary ontology-management tool of the Open Biomedical Ontologies (OBO), the most prominent and highly used set of ontologies in the biomedical domain.

OBOs have typically been created modularly and with informal, natural-language definitions; this is largely due to the fact that they are mostly developed by different groups (and have varying levels of funding, which partly accounts for the varying levels of quality among the OBOs). There have been recent efforts to formalize and link the disparate ontologies, and there is now an extensive effort within the OBO Consortium to create formal definitions of the component terms using more atomic terms. Running a deductive reasoner over these so-called cross-product definitions has resulted in improved ontology quality, usually in the form of inferred `is_a` links among terms. This has the added effect of aligning the linked subject and object terms; thus, the inferred `is_a` links point to what we refer to as *nonalignments*, in which either the subject terms are subsumptively linked and the object terms are not, or vice versa. We have additionally noticed that a form of abductive reasoning not currently widely used has also proven useful in aligning linked terms and thus further improving ontology quality. First, as an example of a nonalignment that can be deductively detected:

```
Class(positive regulation of hydrolase activity
    complete
      biological_process
      restriction(regulates some hydrolase activity))

Class(ATPase activator activity
    complete
      molecular_function
```

```
        restriction(regulates some ATPase activity))
```

and the subsumption:

```
  SubClassOf(ATPase activity
             hydrolase activity)
```

an `is_a` link is inferred between the defined terms, thus pointing to the nonalignment:

```
  SubClassOf(ATPase activator activity
             positive regulation of hydrolase activity)
```

On the other hand, given the initial necessary and sufficient definitions:

```
  Class(coenzyme binding
      complete
        molecular_function
        restriction(results_in_binding_of some coenzyme))

  Class(FAD binding
      complete
        molecular_function
        restriction(results_in_binding_of some FAD))
```

and the subsumption:

```
  SubClassOf(FAD binding
             coenzyme binding)
```

an `is_a` link is *not* inferred between the object terms, thus missing the nonalignment:

```
  SubClassOf(FAD coenzyme)
```

This is, of course, correct OWL behavior. We have found in practice, however, that almost all occurrences of these types of nonalignments point to problems among the linked subject and object terms and that the subject and object terms should be aligned in the large majority of cases. (That is, they should either both be subsumptively linked or both not subsumptively linked.) In a study of nearly 8,000 OBO cross-product definitions, 38.8% of nonalignments detected rely upon this type of abductive reasoning (http://compbio.uchsc.edu/Hunter_lab/Bada/nonalignments_2008_08_08.html). Thus, we assert that a "button" for this type of reasoning is needed in ontology-management tools, analogous to the classifying button that operates in tools such as Protege-OWL. This type of abductive reasoning probably should not by default be enabled, but we assert this analysis of links and identified nonalignments would significantly improve ontology quality.