Post-Coordinating Orthogonal Ontologies for Data Annotation

Landon T Detwiler[1], MS, Marianne Shaw[2], MS, Daniel L Cook[1,4], MD, PhD, John H Gennari[3], PhD, James F Brinkley[1,2,3], MD, PhD
Departments of Biological Structure[1], Computer Science and Engineering[2], Medical Education and Biomedical Informatics[3], Physiology and Biophysics[4]
University of Washington, Seattle, WA

Using an extended SparQL syntax and engine, known as VSparQL, we demonstrate a method for joining concepts from orthogonal reference ontologies to form new concepts on-the-fly for data annotation. We use Skolem functions to produce unique references for each new data annotation instance.

Current efforts such as the Open Biomedical Ontologies (OBO) initiative propose a set of standard non-overlapping (orthogonal) reference ontologies for biomedicine [1]. Two such orthogonal ontologies are the Foundational Model of Anatomy (FMA) [2] and the Ontology of Physics for Biology (OPB) [3]. The FMA ontology represents anatomical entities and their structural relationships. The OPB represents physical properties and the laws of systems dynamics and network thermodynamics.

While some data may be annotated according to a single reference ontology (i.e. anatomical entities seen in an x-ray), often data require references to multiple orthogonal ontologies. For example, measures of the flow (*Fluid_flow* from the OPB) of blood through the coronary artery (*Blood_in_coronary_artery* in the FMA). If *Flow_of_blood_in_coronary_artery* doesn't exist as a pre-coordinated expression in either ontology, what ontology reference should we use for this annotation?

One impractical approach would be to pre-coordinate all possible physical properties of all anatomical entities. This approach does not scale well, particularly as we include additional ontologies. Instead we propose a method for post-coordinating ontologies, generating new ontology classes as needed for data annotation. Our method is query-based and on-demand. Queries are written in an extended SparQL syntax called VSparQL [4]. Our ontologies are represented in OWL.

Query 1 determines all of the portions of blood in the FMA (#2), using a recursive property extension called Gleen [5]. It also determines all of the kinetic fluid properties in the OPB (#3). For each pair it then constructs a new annotation instance (#1). Each annotation takes the form of an RDF statement. Each statement connects an FMA structure to an OPB property using the OPB hasProperty relationship. Figure 1 illustrates 1 such annotation instance for the FMA class *Blood_in_coronary_artery* and the OPB class *Fluid_flow*.

```
CONSTRUCT                                                              (1)
{
 [[annot_view:ann(?pob,?property)]] rdf:type rdf:Statement.
 [[annot_view:ann(?pob,?property)]] rdf:subject ?pob .
 [[annot_view:ann(?pob,?property)]] rdf:predicate opb:hasProperty.
 [[annot_view:ann(?pob,?property)]] rdf:object ?property .
}
FROM NAMED <http://sig.biostr.washington.edu/fma3.0>
FROM NAMED <http://sig.biostr.washington.edu/OPB-01.owl>
WHERE
{
 GRAPH <http://sig.biostr.washington.edu/fma3.0>                       (2)
 {
  ?pob gleen:OnPath ("[rdfs:subClassOf]+" fma:Portion_of_blood).
 }
 GRAPH <http://sig.biostr.washington.edu/OPB-01.owl>                   (3)
 {
  ?property gleen:OnPath ("[rdfs:subClassOf]+" opb:Kinetic_property).
  ?property opb:Physical_domain_class opb:Fluid_domain .
 }
}
```

Query 1: VSparQL query post-coordinates the kinetic properties of portions of blood.

To enable external sources to refer to the newly created annotations, they must have proper URIs. Each statement is given a URI using the skolem functions provided by VSparQL. VSparQL skolem functions have 2 important characteristics:

1. Calling the same function twice with the same arguments always produces the same URI.
2. Calling the same function twice with different arguments always produces different URIs.

These properties are guarantee that a unique URI will be created for each unique set of inputs. The skolem functions shown in Query 1 (delineated by double square brackets in #1) each have 2 arguments, a portion of blood (?pob) and a physical property (?property).

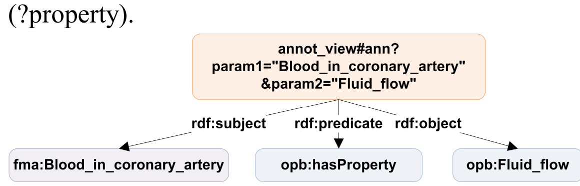

Figure 1: Annotation for flow of blood in the coronary artery.

Supported by NIH grant HL087706

[1] http://www.obofoundry.org/
[2] http://fma.biostr.washington.edu/
[3] http://sigpubs.biostr.washington.edu/archive/00000225/
[4] http://sigpubs.biostr.washington.edu/archive/00000227/
[5] http://sig.biostr.washington.edu/projects/ontviews/gleen/