Post-Coordinating Orthogonal Ontologies for Data Annotation

Structural Informatics Group

Landon T Detwiler[1], MS, Marianne Shaw[2], MS, Daniel L Cook[1,3,4], MD, PhD, John H Gennari[3], PhD, James F Brinkley[1,2,3], MD, PhD

Departments of [1]Biological Structure, [2]Computer Science and Engineering, [3]Medical Education and Biomedical Informatics, and [4]Physiology and Biophysics

University of Washington, Seattle, WA

Introduction:

Using an extended SPARQL syntax and engine, known as vSPARQL [1], 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.

Orthogonal ontologies:

Current efforts such as the Open Biomedical Ontologies (OBO) initiative propose a set of standard non-overlapping (orthogonal) reference ontologies for biomedicine. 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. Both are available in OWL.

Annotations:

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(s) should we use for this annotation?

Simple resource annotation:

Figure 1 shows a simple case of a multi-ontology annotation. In this scenario annotation instances refer to resource nodes in two orthogonal reference ontologies, using specialized properties.

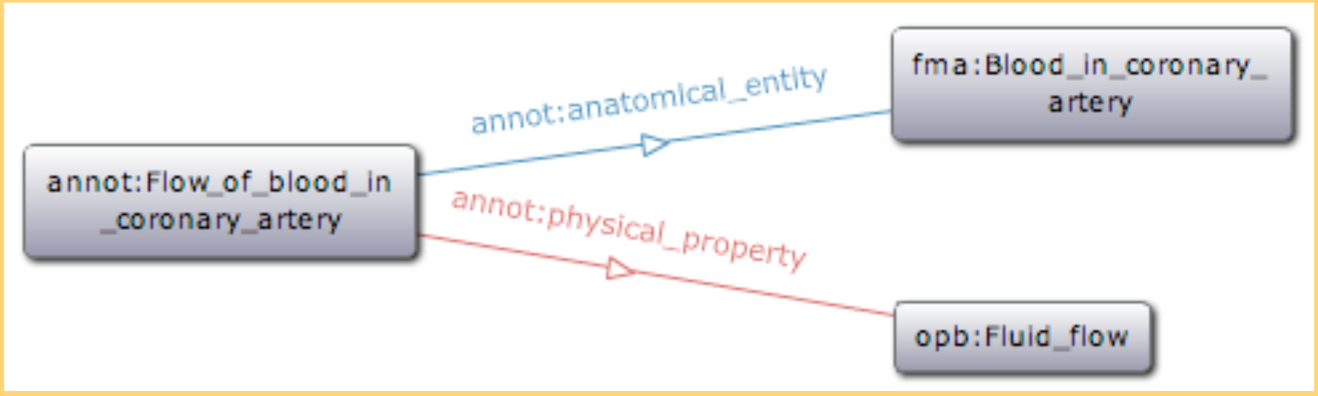

Figure 1: Annotation by resource references

Statement annotation:

For the types of annotations required by [4] and [5], simple resource annotations are insufficient. The relationship between *Blood_in_coronary_artery* and *Fluid_flow* is not captured. We need a method that supports complete statements (or even sets of statements) as annotations. We use RDF reification to create new rdf:Statement instances. This approach is shown in Figure 2.

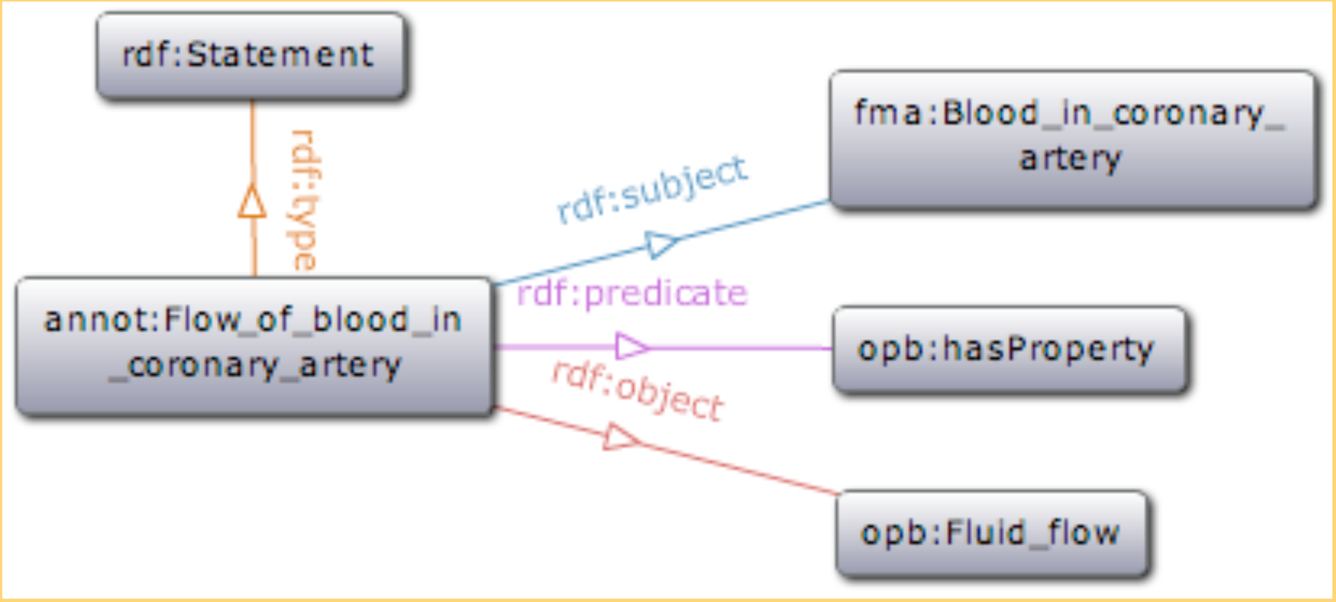

Figure 2: Annotation by statement references

Post-coordination:

It would be impractical to pre-coordinate all possible physical properties of all anatomical entities. Query 1 gathers all portions of blood in the FMA, as well as all fluid kinetic properties from the OPB. From the cross product new statements are created for each pair, using the opb:hasProperty relationship. Reification and anonymous classes enable on-the-fly post-coordination but anonymous classes do not provide a stable reference for annotating data.

```
CONSTRUCT
{
  [] rdf:type rdf:Statement ;
     rdf:subject ?portion_of_blood ;
     rdf:predicate opb:hasProperty ;
     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>
  {
    ?portion_of_blood rdfs:subClassOf+ fma:Portion_of_blood .
  }
  GRAPH <http://sig.biostr.washington.edu/.../OPB-01.owl>
  {
    ?property rdfs:subClassOf+ opb:Kinetic_property .
    ?property opb:Physical_domain_class opb:Fluid_domain .
  }
}
```

Query 1: Post coordination with anonymous classes

Skolem functions:

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 guarantee that a unique URI will be created for each unique set of inputs. The skolem function in Query 2 (in double square brackets) has 2 arguments, a portion of blood (?portion_of_blood)

```
CONSTRUCT
{
  [[annot:annotation(?portion_of_blood,?property)]]
     rdf:type rdf:Statement ;
     rdf:subject ?portion_of_blood ;
     rdf:predicate opb:hasProperty ;
     rdf:object ?property .
} ...
```

Query 2: Post coordination with Skolem functions

and a physical property (?property). For the variable bindings

?portion_of_blood = *fma:Blood_in_coronary_artery*
?property = *opb:Fluid_flow*

it produces the URI:

http://sig.biostr.washington.edu/annotation?
param1="http://.../fma/Blood_in_coronary_artery"&
param2="http://.../opb/Fluid_flow"

Discussion and conclusion:

vSPARQL's Skolem functions enable post-coordination, on-the-fly, from orthogonal reference ontologies. This strategy enables URI generation for any anonymous class, including those representing statements or collections of statements. Defining conventions for uniquely naming statement collections remains as future work.

Supported by NIH grant HL087706

[1] Shaw M, Detwiler LT, Brinkley JF, Suciu D, Mejino JLV. Generating Application Ontologies from Reference Ontologies. AMIA 2008; 672-676.

[2] http://fma.biostr.washington.edu/

[3] Cook DL, Mejino JLV, Neal ML, Gennari JH. Bridging Biological Ontologies and Biosimulation: The Ontology of Physics for Biology. AMIA 2008; 136-140.

[4] Gennari JH, Neal ML, Mejino JLV, Cook DL. Using Multiple Reference Ontologies: Managing composite annotations. ICBO 2009; (in press) 83-86.

[5] Cook DL, Mejino JLV, Neal ML, Gennari JH. Composite Annotations: Requirements for Mapping Multiscale Data and Models to Biomedical Ontologies. EMBC 2009, (in press)