Example Knowledge Resource Illustrating Automated Consistency Validation in RDF vs. XML

John Barkley (jbarkley@nist.gov)

With Semantic Web methods and tools, knowledge resource semantics made explicit can be automatically validated for their consistency. Reducing the need for developing specialized validation tools for each application provides a significant ROI.

This short example illustrates this advantage of using RDF over XML for knowledge resources. For those who may be unfamiliar with the Semantic Web, this note provides an example that can be understood with minimal effort.

This example is a variation of the textbook classic “Student Registration”. Semantics include:

- Each course has only one instructor
- Each course has at least three students.

In XML, one might have:

```
<?xml version="1.0"?>
<Student_Registration>
  <Courses>
    <Course name="PHY499" >
      <Instructor name="Nobel" />
      <Students>
        <Student name="Cornel" />
        <Student name="Hall" />
        <Student name="Phillips" />
      </Students>
    </Course>
  </Courses>
</Student_Registration>
```

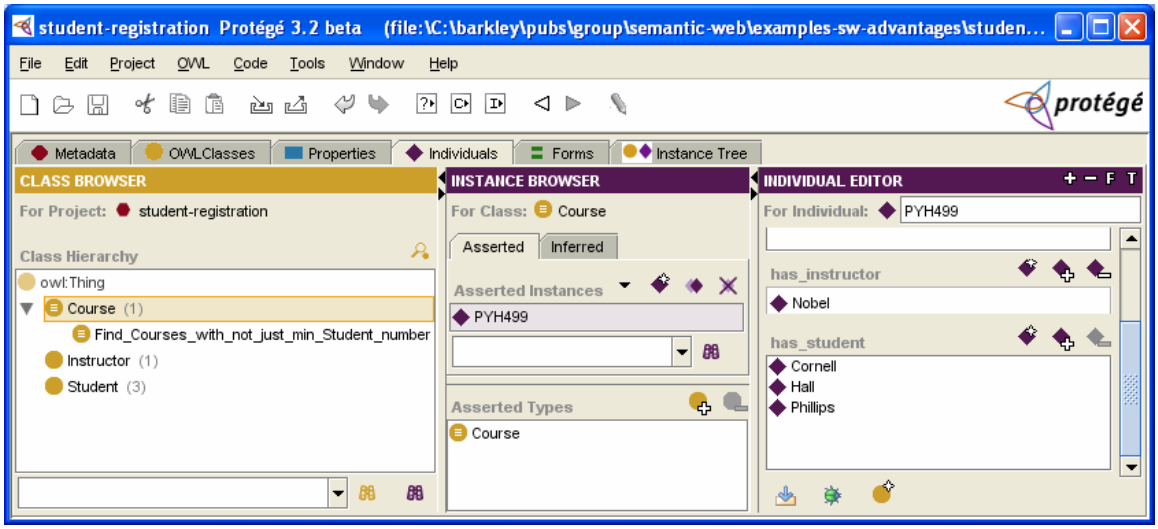
One might use XSLT or XQuery to develop an application to validate the data in this knowledge resource according to the semantics listed above. In order to do this, the semantics must be specified explicitly. For this example, XPath expressions are used:

- Only one Instructor: count(//Course[@name="PHY499"]/Instructor)=1
- At least three Students: count(//Course[@name="PHY499"]/Students/Student)>=3

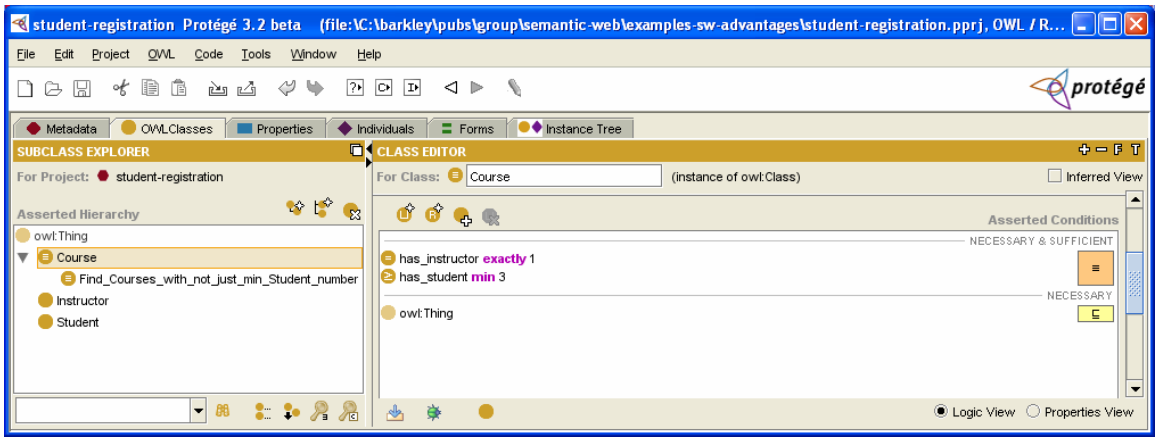
Furthermore, one may want to test queries such as “Find all Courses that do not have just the minimum number of Students” using the XPath expression: //Course[not(count(//Course/Students/Student)>= 3)]. The list of Courses returned by this expression is a subset of all Courses.

Having defined semantics and queries explicitly, how is their consistency to be assured? This is typically done by manual inspection and testing. Since there are no automated tools for checking the consistency of XPath expressions, such tools would have to be developed. These tools would likely be specialized for this application.

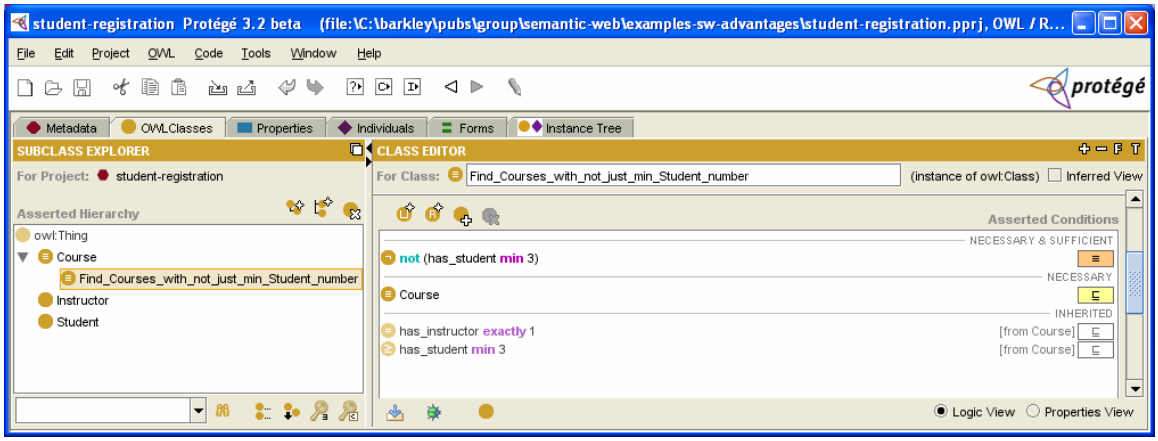
However, semantics and queries can be validated for consistency by existing tools when the knowledge resource is represented in OWL DL. In the Semantic Web tool Protege, this knowledge resource might appear as illustrated below:



There are three OWL Classes: Course, Instructor, and Student. There is one individual in the Class Course: PHY499 that has Instructor Nobel and three Students, i.e., Cornell, Hall, and Phillips.

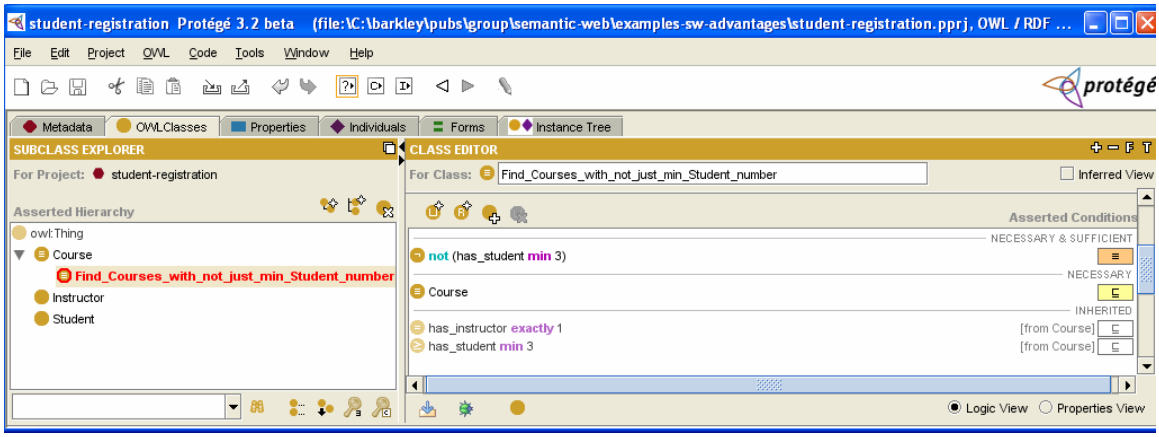


As shown in the display above, the Class Course specifies the semantics: one instructor per Course, and at least 3 Students in each Course.

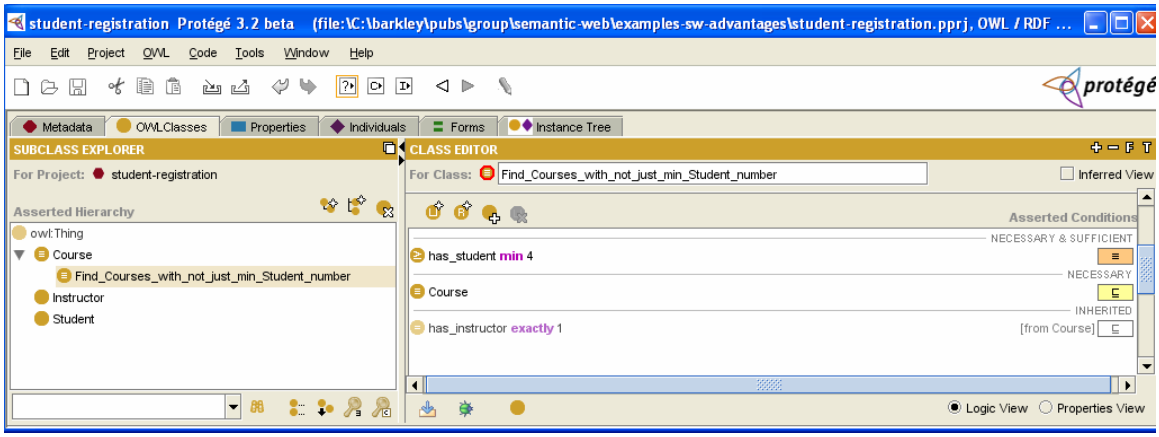


This display shows the specification of the Class Find_Courses_with_not_just_min_Student_number representing the query “Find all Courses that do not have just the minimum number of Students”. This Class is specified as a subclass of Course since all courses which do not have just the minimum number of Students is a subset of all Courses.

In order to check the consistency of the semantics and queries, a reasoner, in this case, Racer, is applied yielding the following result:



As shown in the above, the Class Find_Courses_with_not_just_min_Student_number has been marked in red, indicating an inconsistency. In this case, the reasoner identifies this Class is always empty. A brief examination locates the error in its specification. The Class of all Courses which do not have just the minimum number of students is not the Class of all Courses whose count is not greater than or equal to 3 (i.e., less than 3). It is the Class of all Courses whose count is greater than or equal to 4.



This display shows the results of the consistency check once the error has been corrected.

Note that the XPath expression for the query “Find all Courses that do not have just the minimum number of Students” is also incorrect, and should be:
//Course[count(//Course/Students/Student)>= 4]

In this example, RDF/OWL tools have revealed an error in the specification of semantics and queries which resulted from a mistranslation of the English expression of the query - not an uncommon occurrence.

Additional Notes:

- Generally speaking, knowledge resource semantics represented in either XML or RDF can be explicitly stated in a form that enables automated processing. For XML, the automated processing only includes validating the information in the knowledge resource against the specified semantics which are often called “constraints”. There are no automated tools for validating the specified semantics themselves for their consistency. This is because, in general, validating a collection of expressions of the expressivity of XPath is undecidable, i.e., an automated process cannot exist. For OWL DL representations of knowledge resources, this is not the case. It has been shown mathematically that for semantics expressed in OWL DL, their consistency is decidable.