Two hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Semantic Web

Thursday 22nd January 2009
14:00 – 16:00

Please answer THREE questions from the FIVE provided

The use of electronic calculators is NOT permitted.
1. a) “The Syntactic Web is a place where computers do the presentation and people do the linking and interpreting.” (Goble, 2003)
   
   Briefly elaborate on this statement by giving 3 features that characterise the Syntactic Web. (3 marks)

   b) Give 4 types of task that are hard to accomplish if we are restricted to a Syntactic Web. Give at least one example for each type of task you identify. (4 marks)

   c) What are the fundamental issues preventing us from accomplishing the tasks you identified in 1b via the Syntactic Web? (3 marks)

   d) How does the vision of the Semantic Web, as originally set out in 2001, propose to aid with the tasks you identified in 1b? (4 marks)

   e) Has the Semantic Web Layer Cake (in its different versions) helped or hindered progress towards achieving the goals of the Semantic Web? Discuss, supporting your arguments through reference to appropriate developments and applications in the field. (6 marks)

2. a) Consider the following fragments taken from two XML document instances:

   <course name="Discrete Maths">
     <lecturer>David Billington</lecturer>
   </course>
   <lecturer name="David Billington">
     <teaches>Discrete Maths</teaches>
   </lecturer>

   Why would these fragments represent a problem for processing by machine? (2 marks)

   b) How would you translate the fragments in 2a into Resource Description Framework (RDF) format? You may use a graphic representation, informal triple representation or RDF/XML representation. (4 marks)

   c) When might you consider using items from the XML Schema namespace in RDF documents? (2 marks)

(Question 2 continues on the following page)
d) Consider the following RDF descriptions:

```xml
<?xml version="1.0"?>
<River xmlns="http://www.geodesy.org/river#">
  <name>Yangtze</name>
  <length>6300 kilometers</length>
  <startingLocation>western China's Qinghai-Tibet Plateau</startingLocation>
  <endingLocation>East China Sea</endingLocation>
</River>

<?xml version="1.0"?>
<River xmlns="http://www.geodesy.org/river#">
  <name>Yangtze</name>
  <name>Dri Chu - Female Yak River</name>
  <name>Tongtian He, Travelling-Through-the-Heavens River</name>
  <name>Jinsha Jiang, River of Golden Sand</name>
</River>
```

Explain why an attempt at aggregating from these descriptions would fail, and indicate what changes would be needed to ensure successful aggregation. (4 marks)

e) A multinational manufacturer with a large archive of XML instance documents of many different types, for both internal and external use, seeks advice on whether it should undertake a mass conversion of its archive to RDF format. Set out advantages and disadvantages of conversion, and give your recommendations. Justify your conclusions and recommendations. (8 marks)
3. a) Consider the following (partial) RDF Schema (RDFS) document:

```xml
<?xml version="1.0"?><rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xml:base="http://www.geodesy.org/water/naturally-occurring">
    <rdfs:Class rdf:ID="River">
        <rdfs:subClassOf rdf:resource="#Stream"/>
    </rdfs:Class>
    <rdfs:Class rdf:ID="Stream">
        <rdfs:subClassOf rdf:resource="#NaturallyOccurringWaterSource"/>
    </rdfs:Class>
    <rdf:Property rdf:ID="emptiesInto">
        <rdfs:domain rdf:resource="#River"/>
        <rdfs:range rdf:resource="#BodyOfWater"/>
    </rdf:Property>
    <rdf:Property rdf:ID="length">
        <rdfs:domain rdf:resource="#River"/>
        <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
    </rdf:Property>
    ...
</rdf:RDF>
```

In the following separate RDF/XML document, there is no indication of the class of the resource 'Yangtze' (lines are numbered for ease of reference):

```
(1) <?xml version="1.0"?>
(2) <rdf:Description rdf:ID="Yangtze"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns="http://www.geodesy.org/water/naturally-occurring#">
(3)     <length>6300 kilometers</length>
(4)     <emptiesInto rdf:resource="http://www.china.org/geography#EastChinaSea"/>
(5) </rdf:Description>
```

What inferences can we make from the RDF/XML document using the RDFS document? Explain in detail how each inference you identify is arrived at.

(4 marks)
b) State whether we can make the same inferences you have identified in 3a, if line 6 of the RDF/XML document is changed to:

(6) <emptiesInto>East China Sea</emptiesInto>

Explain your answer. (2 marks)

c) Furche et al. (2006) list the following as types of query of interest for XML/RDF/RDFS querying: Selection, reduction, extraction, combination, restructuring, inference, aggregation. Briefly explain each type of query, giving an English language example of each. Explain to what extent SPARQL, the W3C language for querying RDF, is capable of handling each type of query. (6 marks)

d) Consider the following SPARQL query:

```sparql
PREFIX books: <http://www.books.org/books#>
SELECT ?work
WHERE { ?work books:author ?name .
    OPTIONAL { ?work books:translator ?translator } }
```

which is attempting to query RDF triples regarding works, authors and translators of authors' works.

Explain why this query is not optimal and what modifications you would make to render it optimal. (3 marks)

e) RDFS is described as a “primitive ontology language” by Antoniou & van Harmelen (2008). What grounds are there for this view? To what extent do you agree with this view? Justify your position with appropriate argumentation and examples. (5 marks)
4.  
   a) I am working on an ontology about Pizzas, and add a new primitive class called MargheritaPizza. Using the Manchester syntax, I add restrictions to this class as follows:

   hasTopping some MozzarellaTopping  
   hasTopping some TomatoTopping

   and attempt to use a reasoner to classify this class under an existing defined VegetarianPizza class, which is defined to be a subclass of Pizza and to have toppings that are only VegetableTopping or CheeseTopping. MozzarellaTopping and TomatoTopping are properly specified as subclasses of CheeseTopping and VegetableTopping, respectively. The classification fails. However, MargheritaPizza definitely consists of only vegetarian 'ingredients'. Explain the probable reasons for this failure and what changes I need to make to the class restrictions to get it to classify appropriately.  

   (3 marks)

   b) I am working on an ontology about Pizzas, and add the following OWL class, here labelled OWL1 for reference purposes:

   OWL1:
   <owl:Class rdf:about="#CheesyPizza">  
     <rdfs:subClassOf rdf:resource="#Pizza"/>  
     <rdfs:subClassOf>
       <owl:Restriction>
         <owl:onProperty rdf:resource="#hasTopping"/>
         <owl:someValuesFrom rdf:resource="#CheeseTopping"/>
       </owl:Restriction>
     </rdfs:subClassOf>
   </owl:Class>

(Question 4 continues on the following page)
I then decide to change it as follows, yielding class OWL1a:

OWL1a:
<owl:Class rdf:about="#CheesyPizza">
   <owl:equivalentClass>
      <owl:Class>
         <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about="#Pizza"/>
            <owl:Restriction>
               <owl:onProperty rdf:resource="#hasTopping"/>
               <owl:someValuesFrom rdf:resource="#CheeseTopping"/>
            </owl:Restriction>
         </owl:intersectionOf>
      </owl:Class>
   </owl:equivalentClass>
</owl:Class>

and I further ensure that OWL1a is not disjoint.

Explain what would happen if I ran a reasoner on the original ontology containing OWL1, and what would happen if I then ran the reasoner on the modified ontology containing OWL1a (assuming other appropriate classes and properties exist).

(4 marks)

A principle of ontological engineering is that subclasses should be pairwise disjoint, however I have made OWL1a non disjoint. Why did I do this?

(1 mark)
c) The following is stated in an OWL ontology:

```xml
<owl:ObjectProperty rdf:ID="isTaughtBy">
  <rdf:type rdf:resource="&owl;FunctionalProperty"/>
</owl:ObjectProperty>
```

and also, the following is erroneously stated, which assumes that resources 949318 and 949352 are distinct individuals:

```xml
<course rdf:ID="CT1111">
  <isTaughtBy rdf:resource="#949318"/>
  <isTaughtBy rdf:resource="#949352"/>
</course>
```

Account for the lack of an error being generated by an OWL reasoner applied to the above, and state (informally) what modifications or additions you would make to trap the inconsistency. (2 marks)

d) The main requirements for an ontology language according to Antoniou & Van Harmelen (2008) are: a well-defined syntax; a formal semantics; convenience of expression; efficient reasoning support; and sufficient expressive power. The W3C Web Ontology Working Group has defined however three sub-languages of OWL in response to such requirements. Say what these three varieties are, why they are deemed necessary and describe the relationships among them. (5 marks)

e) You are tasked with coordinating the construction of a large ontology in which numerous subject specialists from the domain at issue will be engaged, who have little knowledge of how to build ontologies. Discuss what steps you would take and what recommendations you would make to ensure a high probability of success of arriving at a reasonable ontology. (5 marks)
5. “Today searching Google for "Toyota used cars for sale in western Massachusetts under $8,000" returns more than 2,000 general Web pages. Once Semantic Web capabilities are added, a person will instead receive detailed information on seven or eight specific cars, including their price, colour, mileage, condition and owner, and how to buy them. [...] specialists and enthusiasts will define taxonomies and ontologies: data sets that describe classes of objects and relations among them. These sets will help computers everywhere to find, understand and present targeted information.” (Shadbolt and Berners-Lee, 2008)

Discuss, giving your views regarding this recent restatement of aspects of the Semantic Web vision. Include discussions of issues of scale and feasibility, and of near-term prospects. In your discussion, justify your views, giving appropriate examples to back up your arguments. (20 marks)