Three hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Semantic Web

Date: Thursday 15th May 2008
Time: 09:45 – 12:45

Please answer any THREE Questions from the FIVE questions provided

This is a CLOSED book examination

The use of electronic calculators is NOT permitted.

Do not answer more than the required number of questions. Additional questions will not be marked. Clearly cross out anything you do not wish to be marked

[PTO]
1. a) Arrange the following unordered “ingredients” into a diagram of the Semantic Web “Layer Cake”:

- Data Interchange: Resource Description Framework (RDF)
- Unifying Logic
- Ontology: Web Ontology Language (OWL)
- Extensible Markup Language (XML)
- Cryptography
- RDF Schema (RDFS)
- XML Schema (XMLS)
- Trust
- Uniform Resource Identifier (URI)
- Internationalized Resource Identifier (IRI)
- Proof
- Query: SPARQL
- Rule: Rule Interchange Format (RIF)
- User interface and applications

Briefly characterise the function of each layer of your diagram in relation to the Semantic Web and explain the relationships between your layers. (4 marks)

What two main principles should be followed in designing a Semantic Web layer? (2 marks)

b) “Today’s Web may be defined as the Syntactic Web.” (Breitman et al., 2007)

What do you understand by this characterisation of the Web? In your answer, discuss how the Syntactic Web and the Semantic Web are related, and assess to what extent this definition might still be true in 10 years' time, in the light of expert judgements you are familiar with. Justify your views and conclusions. (10 marks)
2. a) In XML Schema, when do you use simpleContent and when complexContent? (2 marks)

b) Consider the following XML Schema document (lines numbered for ease of reference):

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://www.books.org"
            xmlns="http://www.books.org"
            elementFormDefault="qualified">
  <xsd:element name="BookStore">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Book" maxOccurs="unbounded">
          <xsd:complexType>
            <xsd:sequence>
              <xsd:element name="Title" type="xsd:string"/>
              <xsd:element name="Author" type="xsd:string"/>
              <xsd:element name="Date" type="xsd:string"/>
              <xsd:element name="Publisher" type="xsd:string"/>
            </xsd:sequence>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Referring to lines 2-5, explain the notion of XML namespaces and how they are used. (2 marks)

Lines 6-21 validly declare an element “BookStore”, containing a sequence of “Book” elements. There is a plan to write another schema, in which the writer wishes to re-use the same information about “Book” to declare a “Document” element. Specify what changes would have to be made in the above schema to allow such re-use. (3 marks)

(Question 2 continues on the following page)
c) Consider the following fragments taken from two XML document instances:

```xml
<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)

How can use of the Resource Description Framework (RDF) help overcome such problems? (4 marks)

d) “We recommend using the RDF format for all your XML documents.” (Costello & Jacobs, 2003)

Do you agree with this recommendation? Discuss, giving appropriate examples, and giving justification for your views. (7 marks)
3. a) In a Resource Description Framework (RDF) document, when would you use “rdf:ID” and when “rdf:about”? (2 marks)

b) In what circumstances are “rdf:subject”, “rdf:object” and “rdf:predicate” used in RDF, and why? (3 marks)

c) For a sports-related web site, which provides also RDF feeds to others, two different companies are each asked to design RDF documents describing the players in sports teams, and you are asked to comment on their designs. One company has used RDF Containers consistently throughout. The other has consistently used RDF Collections. What would be your comment in terms of advantages/disadvantages of these approaches? (2 marks)

d) Consider the following three (informally expressed) queries over RDF/RDFS data:

Q1 Select all data items with any relation to the book entitled “RDF Rocks”.
Q2 Return the transitive closure of the subClassOf relation.
Q3 Return the last year in which an author with name “Ebenezer Snidgewort” published something.

Classify these queries: identify 3 pairs each consisting of a general and its more specific query type from the following list of 7 types, and match each pair with its corresponding query above:

Selection, reduction, extraction, combination, restructuring, inference, aggregation.

Order each pair with the most general type as the first member of the pair. (3 marks)

(Question 3 continues on the following page)
(Question 3 continues from the previous page)

e) Answer EITHER

i) RDF Schema (RDFS) is described as a “primitive ontology language” by Antoniou & van Harmelen (2004). Discuss reasons for this view. Also, drawing on your knowledge of requirements for ontology languages, discuss features not found in RDFS that characterise what are generally considered richer ontology languages (such as OWL, the Web Ontology Language). In your discussion, comment in particular on why an ontology language designer needs to be aware of the relationship between expressive power and efficient reasoning support. (10 marks)

OR

ii) The World Wide Web Consortium (W3C) recently published its Recommendation for the SPARQL Query Language for RDF. Comment on the extent to which SPARQL is able to handle the above queries under 3d; compare its characteristics and capabilities with other RDF/RDFS query languages with which you are familiar; and assess to what extent there exists an “ideal” RDF/RDFS query language. (10 marks)
4. a) Consider the following partially complete OWL (Web Ontology Language) class (line numbers are for reference only):

```
(1) <owl:Class rdf:about="#FirstYearCourse">
(2)   <rdfs:subClassOf>
(3)     <owl:MISSING1>
(4)       <owl:MISSING2 rdf:resource="#isTaughtBy"/>
(5)       <owl:allValuesFrom rdf:resource="#Professor"/>
(6)     </owl:MISSING1>
(7)   </rdfs:subClassOf>
(8) </owl:Class>
```

What should take the place of MISSING1 in lines 3 and 6, and of MISSING2 in line 4? (2 marks)

Having filled in the missing components, what can be said about the construct between lines 3 and 6? (1 mark)

How would you express the meaning of the FirstYearCourse class in plain English? (1 mark)

b) In an OWL ontology of living things, which I am developing using Protégé 4, I specify primitive classes for Grass, Leafy_plant and Cow, and a defined class for Herbivore, as follows (using Manchester syntax):

i)  Herbivore = Animal AND ii) Grass = Plant
    eats ONLY Plant

iii) Cow = Mammal        iv) Leafy_plant = Plant
    eats SOME Grass
    eats SOME Leafy_plant

Animal and Plant are already specified as primitive subclasses of Living_Thing; Mammal is already specified as a primitive subclass of Animal. I now run a classifier, but find that Cow is not classified as a Herbivore. Explain why this result is obtained, and state what changes would be necessary to allow Cow to be automatically classified as a Herbivore. (3 marks)

c) A principle of ontological engineering is that subclasses should be pairwise disjoint. E.g. The class Cow should not contain any members of the class Dog. Under which circumstances would you break this principle? (2 marks)

(Question 4 continues on the following page)
(Question 4 continues from the previous page)

d) Consider the following constraint:

"Any resource is allowed to be only a class, a data type, a data type property, an object property, an individual, a data value, or part of the built-in vocabulary, and not more than one of these." (Antoniou & van Harmelen, 2004)

Which of the three sub-languages of OWL does the above constraint apply to? Justify your answer. (2 marks)

e) Ontology design patterns are often referred to in discussions of best practice in building ontologies. Explain what such design patterns are, in relation to building OWL ontologies. Give examples of the design patterns you describe and discuss which are likely to be of greatest help to the novice ontology builder. Motivate your ranking. (9 marks)
5. a) The Semantic Web Challenge defines a Semantic Web application as having to meet the following requirements:

1. The meaning of data has to play a central role:
   - Meaning must be represented using formal descriptions
   - Data must be manipulated/processed in interesting ways to derive useful information and
   - This semantic information processing has to play a central role in achieving things that alternative technologies cannot do as well, or at all.

2. The information sources used
   - should have diverse ownerships
   - should be heterogeneous
   - should contain real world data.

3. The application assumes an open world (information is never complete).

Demonstrate, with reference to examples, to what extent applications meeting these criteria exist in convincing form. Discuss technological barriers to more widespread deployment of semantic applications, assess the scale of the individual barriers you identify, and comment on advances needed to overcome these barriers. (10 marks)

b) “For most organizations thus far, semantic technologies have been considered a "future" technology. In other words, it was considered either too cutting edge and risky, or too immature to be of value. To gain faster and more widespread adoption, especially within mainstream corporate development environments, there needs to be a clearer definition of the benefits and opportunities of semantic technology.” (SemTech 2007 Panel Topic)

Imagine you are a participant in the SemTech panel discussion and are called on to give a short position paper on the above topic. Set out your position, with appropriate argumentation and justification. You may agree, disagree, adopt a compromise position, be controversial, etc.: however, the panel chairman has warned you that unsupported opinion will not go down well with your audience. (10 marks)