Two hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

M.Sc. in Advanced Computer Science

Semi-Structured Data and the Web

Date: Tuesday 2\textsuperscript{nd} June 2009
Time: 14:00 – 16:00

Please answer BOTH Questions
This is a CLOSED book examination
The exam will be taken on line.
This is the paper format, which will be available as a backup

The use of electronic calculators is NOT permitted
1. You must answer ALL subquestions. (30 marks total)

a) Describe each of the following terms in one or two sentences.

i) well-formedness of an XML document (1 mark)
ii) validity of an XML document (1 mark)
iii) an XML parser (1 mark)
iv) a validating XML parser (1 mark)
v) the DOM tree or infoset of an XML document (1 mark)

b) Describe two aspects of an XML document that most XML schema languages are designed to describe and constrain. For each aspect, give an example of a constraint with respect to this aspect. (4 marks)

c) We have discussed the concepts of tree grammars, single-typed and local tree grammars. Answer each of the following questions in one or two sentences.

i) What are tree grammars used for in the context of XML? (1 mark)

ii) What is a "single-type" tree grammar? (1 mark)

iii) Why is being "single-type" relevant for tree grammars in the context of XML? (2 marks)

iv) What kind of efficient validator can we build to validate a document against a Relax NG schema, and how much memory will it need (in the size of the document)? (2 marks)
d) Consider the following XML schema declarations. Provide an element UKShoesize and an element Company that validates against these declarations.

```
<x:s:complexType name="shoesizeType">
  <xs:simpleContent>
    <xs:extension base="xs:integer">
      <xs:attribute name="gender" type="xs:string"/>
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>

<x:s:element name="UKShoesize">
  <xs:complexType>
    <xs:simpleContent>
      <xs:restriction base="shoesizeType">
        <xs:minExclusive value="2"/>
        <xs:maxExclusive value="17"/>
      </xs:restriction>
    </xs:simpleContent>
  </xs:complexType>
</xs:element>

<x:s:element name="Company">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="name" type="xs:string"/>
      <xs:element name="holding" type="xs:string" maxOccurs="3" minOccurs="1"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

(2 marks)
(Question 1 continues from the previous page)

e) For the following XQuery, describe the result when answered over the two XML documents articles.xml and nodes.xml given in the box below and explain your answer.

```
<mylist>
  {
    for $d in doc("nodes.xml")//deeper,
      $art in doc("article.xml")//@price > "2"
    where $d/@type = $art/@cat
    return <product id="{$art/@cat}">
      { $art/description,
        <morestuff>
          { $d/T }
        </morestuff>
      }
    </product>
  }
</mylist>
```

articles.xml:
```
<?xml version="1.0" encoding="UTF-8"?>
<articlelist>
  <article cat="A" price="3.00">
    <description>super wheel</description>
    <supplier>t2</supplier>
  </article>
  <article cat="A" price="1.00">
    <description>like a bike</description>
    <supplier>t2</supplier>
  </article>
  <article cat="B" price="3.50">
    <description>great bike</description>
    <supplier>t1</supplier>
  </article>
  <article cat="F" price="3.60">
    <description>nice scooter</description>
    <supplier>t3</supplier>
  </article>
</articlelist>
```
f) Give two XPath expressions that retrieve the 'bolded' element (the one containing "nice scooter") from article.xml above. (2 marks)

g) Discuss an application where using a graph-based formalism such as RDF is more suitable than using a tree-based one like XML. Describe the application and explain advantages and disadvantages of using RDF. (3 marks)

h) Describe and compare inference and validation. (3 marks)
You must answer all subquestions. (30 marks total)

a) Give three uses of a schema for XML documents. (6 marks)

b) State Postel's law. (2 marks)

c) **FILL IN THE BLANKS:** Please fill in the prefix and namespace URIs in such a way as to make this XML document not "super confusing". Use the prefixes: "a" and "b" and the namespace uri "http://ex.org". (Note: this is a bit tricky!) (2 marks)

```xml
<a:name xmlns:a="http://ex.org">
  <_:lastname xmlns:_="_______________"></_:lastname>
</a:name>
```

d) Compare the error handling strategy of XML, CSS (according to the spec), and HTML (according to HTML5). Give examples which exemplifies each strategy. (6 marks)

e) List which of schema languages we studied in the course (DTD, WXS, RELAX NG, and Schematron) can constrain the content of the element <age> to be positive integers, that is, can rule out: <age>twenty-nine</age> or <age>-900</age> while accepting: <age>16</age> (2 marks)

f) Write a DTD that describes the following XML document. Strive for a fairly tight, but sensible DTD. (It should be about 6 lines long): (6 marks)

```xml
<argument>
  <topic id="htk">Hume's theory of knowlege</topic>
  <arguers>
    <person nick="bijan"/>
    <person nick="ruth"/>
  </arguers>
</argument>
```

g) What are the benefits of an XSugar grammar (.xsg file)? (2 marks)

(Question 2 continues on the following page)
(Question 2 continues from the previous page)

h) Explain the basic philosophy of Schematron with respect to validation and error reporting.

i) MULTIPLE CHOICE: Assume these rules appear in the same scope:

<table>
<thead>
<tr>
<th>Rule 1</th>
<th>h1 p.standout {color: red}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 2</td>
<td>h1 p#a1 {color: blue}</td>
</tr>
</tbody>
</table>

Select the answer which gives the right priority and reason for that priority.

1. Rule 1 has priority because it is lexically first.
2. Rule 1 has priority because it has greater specificity in its selector.
3. Rule 2 has priority because it is lexically last.
4. Rule 2 has priority because it has greater specificity in its selector.
5. They have equal priority.

END OF EXAMINATION