



UNIVERSITY of LIMERICK

O L L S C O I L L U I M N I G H

COLLEGE of INFORMATICS and ELECTRONICS

Department of Computer Science
and Information Systems

End-of-Semester Assessment Paper

Academic Year:	2006/07	Semester:	1
Module Title:	Document Architectures	Module Code:	CS4146
Duration of Exam:	2½ Hours	Percent of Total Marks:	80
Lecturer(s):	Richard F. E. Sutcliffe	Paper marked out of :	100

Instructions to Candidates:

- Answer any FIVE questions
- Do not answer more than FIVE questions
- All questions carry equal marks

Q1. a) Korean is normally written using Hangul. How does Hangul relate to Jamo?

2 Marks

b) Are Hangul a form of Chinese Hanzi or not? Explain.

2 Marks

c) The Japanese writing system involves four different scripts: Kanji, Hiragana, Katakana and Roma-ji. Explain what these are and what purpose they serve in the language.

6 Marks

d) A Japanese Kanji or Chinese Hanzi is made up of one or more radicals each comprised of one or more strokes. Explain this statement using an ideograph of your own invention.

Note: The ideograph does not need to exist in Japanese or Chinese; Anything is acceptable as long as it illustrates your points clearly.

6 Marks

e) Two methods of entering Kanji or Hanzi are by stroke count and by pronunciation. Explain how these work.

4 Marks

Q2. a) What is meant by the term Document Type Definition (DTD)?

2 Marks

```
01 <?xml version="1.0" encoding="UTF-8" ?>
02
03 <!ELEMENT corpus (q+)>
04
05 <!ELEMENT q (language+)>
06 <!--ATTLIST q
07         cnt CDATA #REQUIRED
08         category (D|F) "D"
09         answer_type (LOCATION|ORGANISATION|PERSON|TIME)
10         #IMPLIED>
11
12 <!--ELEMENT language (question,answer+)>
13 <!--ATTLIST language
14         val (BG|DE|EN|ES|FR|IT|NL|PT) #REQUIRED
15         original (FALSE|TRUE) #REQUIRED>
16
17 <!--ELEMENT question (#PCDATA)>
18 <!--ATTLIST question
19         group (BTB|DFKI|ELDA|IRST|LING|UNED|UoA) #REQUIRED>
20
21 <!--ELEMENT answer (#PCDATA)>
22 <!--ATTLIST answer
23         n CDATA #REQUIRED
24         docid CDATA #IMPLIED>
25
26
27 <?xml version="1.0" encoding="UTF-8" ?>
28
29 <!DOCTYPE corpus SYSTEM "qa_corpus.dtd">
30 <corpus>
31 <q cnt="1" category="G" answer_type="LOCATION">
32 <language val="EN" original="TRUE">
33 <question group="IRST">The question</question>
34 <answer n="1" docid="1234">The answer</answer>
35 <reason>The reason</reason>
36 </language>
37 </q>
38 </corpus>
```

b) The figure above shows an example Document Type Definition (DTD) followed by a document instance. Note that the numbers on the far left are to identify lines in this question and are not part of the document itself. What is the purpose of line 1?

2 Marks

c) Explain exactly what the ELEMENT definitions of lines 5, 12 and 17 mean.

6 Marks

d) Explain exactly what the ATTLIST definition starting on line 6 means.

2 Marks

e) Look at the example document in lines 27-38. Does it conform to the DTD or not? Answer by going through the document explaining carefully how each tag, attribute or entity arises from the definitions in the DTD.

8 Marks

Q3. a) What is the purpose of an XML Schema?

2 Marks

b)

```
01 <?xml version="1.0"?>
02 <xs:schema
03   xmlns:xs="http://www.w3.org/2001/XMLSchema"
04   elementFormDefault="qualified">
05   <xs:element name="message_seq">
06     <xs:complexType>
07       <xs:sequence>
08         <xs:element name="message" maxOccurs="unbounded">
09           <xs:complexType>
10             <xs:sequence>
11               <xs:element name="contents" type="xs:string"/>
12               <xs:element name="route" type="xs:int"/>
13             </xs:sequence>
14             <xs:attribute name="id" type="xs:positiveInteger"/>
15             <xs:attribute name="date" type="xs:date"/>
16           </xs:complexType>
17         </xs:element>
18       </xs:sequence>
19     </xs:complexType>
20   </xs:element>
21 </xs:schema>
22
23 <?xml version="1.0" encoding="UTF-8"?>
24 <message_seq
25   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
26   xsi:noNamespaceSchemaLocation="message_seq.xsd">
27   <message id="10" date="2006-09-20">
28     <contents>The contents</contents>
29     <route>10a</route>
30   </message>
31   <message id="11" date="2006-09-21">
32     <contents>The contents</contents>
33   </message>
34 </message_seq>
```

In the above, lines 1-21 show a schema while lines 23-34 show a sample document referring to it.

What does line 3 state and what is the exact significance of `xmlns:xs`?

4 Marks

c) Is there a processing instruction in lines 1-21 and if so what does it indicate?

2 Marks

d) An element called `message_seq` is being defined. Describe the required structure and content of this element including any attributes.

4 Marks

e) Look at the document on lines 23-34. Does it conform to the schema or not? Answer by going through it line-by-line explaining how it conforms to the schema. **Note:** You are not required to say anything about the attributes of element `message_seq`.

8 Marks

Q4. a)

```

01 <?xml version="1.0"?>
02 <xsl:stylesheet
03   version="1.0"
04   xmlns:xsl=
05   "http://www.w3.org/1999/XSL/Transform">
06
07 <xsl:template match="doc">
08   <html>
09     <head>
10       <title>Transformed Doc1</title>
11     </head>
12     <body>
13       <xsl:apply-templates select="title"/>
14     </body>
15   </html>
16 </xsl:template>
17
18 <xsl:template match="title">
19   <h1><xsl:apply-templates select="/doc/title/text()"/></h1>
20   <h2><xsl:apply-templates select="../sec1/title/text()"/></h2>
21   <h2><xsl:apply-templates select="../sec2/title/text()"/></h2>
22 </xsl:template>
23
24 </xsl:stylesheet>
25
26 <?xml version="1.0"?>
27 <doc>
28   <title>
29     The Document Title
30   </title>
31   <sec1 id="1">
32     <title>Section 1</title>
33     <text>Text of Section 1</text>
34   </sec1>
35   <sec2 id="2">
36     <title>Section 2</title>
37     <text>Text of Section 2</text>
38   </sec2>
39 </doc>

```

Lines 1-24 show an XSL stylesheet while lines 26-39 show a document. What is the effect of lines 4 and 5?

2 Marks

b) State the effect of the template shown on lines 7-16.

2 Marks

c) Draw a document tree for the example document on lines 26-39.

4 Marks

d) Explain what nodes will be matched by the XPATH expressions `/doc/title/text()` and `../sec1/title/text()` assuming you are at node `title` which is the child of `doc`.

4 Marks

e) Using your tree, explain exactly how the stylesheet will be applied to the document and write down the output which will be produced.

6 Marks

f) If the `apply-templates` on line 13 was removed, what effect would this have on the output produced?

2 Marks

Q5. a) In the context of XML what is a White Space character? Name the characters involved. 2 Marks

b) What is the purpose of Line End Normalisation? 2 Marks

c)

```
<doc>
  <writer>
    <given>Niamh</given>
    <family>Flynn</family>
  </writer>
  </contact>john@xyz.com</contact>
...
```

The above fragment is formatted to enable its structure to be readily understood. However, is the formatting between `<writer>` and `<given>` a part of the document text or not? Write down a fragment of a Document Type Definition (DTD) which would show the XML parser that the space was not significant and explain carefully what it shows.

6 Marks

d) Now write another fragment which would show that the space *was* in fact significant. Explain your reasoning.

6 Marks

e) What is a Mixed Content Model? What restrictions are placed on such models within a DTD?

4 Marks

Q6. a) What are Cascading Style Sheets (CSS) and in exactly what sense do they control the style of a document?

4 Marks

b)

```
01  H1 { font-size: 1.5em }
02  BODY { background: url(parchment.jpg) red; color: green }
03  P EM { background: yellow; color: black }
04  .note { margin-left: 5em; margin-right: 5em }
05  .firstwords { font-variant: small-caps }
06  A:link { color: red }
07  A:visited { color: green; font-size: 60% }
```

In the above example what is the precise effect of line 2?

2 Marks

c) Which line in the above is a Contextual/Descendant Selector? Such selectors apply even if containment is not direct. Explain the effect of the selector including an XML fragment in your answer.

4 Marks

d) The rule on line 5 uses a Class Selector. What material in a document will match such a rule? How could a portion of text be demarcated in order for line 5 to apply to it?

4 Marks

e) Explain the effect of a Pseudo Element Selector. Which lines in the above example use such selectors and what is their effect?

4 Marks

f) Name one strength and one weakness of CSS.

2 Marks

- Q7. a) What is the purpose of XPath in the XSL language? 4 Marks
- b) An XPath expression consists of an Axis Specifier, a Node Test and zero or more Predicates. Explain briefly what these are. 3 Marks
- c) How can the structure of a document be represented as a tree? Answer by drawing trees corresponding to the following descriptions:
- A manual consists of a version_spec, three chapters and an index.
 - A manual consists of a version_spec, an introduction, two chapters and an index. Each chapter comprises an outline followed by two sections.
- 4 Marks
- d) In the context of document trees, what is the difference between a relative path and an absolute path? Answer by showing any two absolute paths and any two relative paths using the second tree you have drawn. 4 Marks
- e) Explain the meaning of the following XPath expressions:
- ancestor::system/child::action
- child::one
- system/*/*/content
- contents//footnote
- author[@firstname="Conor"]
- 5 Marks
- Q8. a) What is Information Retrieval and how does it differ from Question Answering? 4 Marks
- b) Explain in outline the method of retrieval by inverted indexing. 6 Marks
- c) Explain the idea of Term Weighting within Information Retrieval. In your answer, include a description of the TF*IDF weighting scheme. 10 Marks