THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF COMPUTING

EXAMINATION

Course : MSc Scheme - 61030

Subject : COMP5323 Web Database Technologies & Applications

Group : 201, 202, 205, 2888

Session : 2009 / 2010 Semester II

Date : 11 May 2010 Time : 18:30-20:30

Time Allowed: 2 Hours Subject Lecturer: Vincent Ng

This question paper has 6 pages (cover included).

Instructions to Candidates:

1. This is an open book examination.
2. Students should attempt ALL questions.
3. Marks for each question are shown next to the question. Total marks = 100.

Do not turn this page until you are told to do so!

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
    xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
        http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd">

<url>
 <loc>http://www.comp.polyu.edu.hk/</loc>
 <priority>1.00</priority>
 <lastmod>2009-01-19T04:22:20+00:00</lastmod>
 <changefreq>daily</changefreq>
</url>
<url>
 <loc>http://www.comp.polyu.edu.hk/Data/userimages/profDocument.pdf</loc>
 <priority>0.51</priority>
 <changefreq>daily</changefreq>
</url>
<url>

<loc>http://www.comp.polyu.edu.hk/contentpage.php?id=20070823112439&amp;lang=e
</loc>
 <priority>0.64</priority>
 <changefreq>daily</changefreq>
</url>
<url>

<loc>http://www.comp.polyu.edu.hk/contentpage.php?id=20070319145752&amp;lang=e
</loc>
 <priority>0.64</priority>
 <changefreq>daily</changefreq>
</url>
</urlset>
```

Figure 1.

Question 1 ***[Total = 25 marks]***

(a) For the XML sitemap shown in Figure 1, write XQuery expressions to satisfy the queries below. If a query cannot be answered, provide justification. The date format is DD/MM/YY. (20 marks)

i. Find the priorities of webpages which contain "prof" as part of their path locations.
ii. How many webpages are there having priorities more than 0.55?
iii. For the webpages with the last modification date before 2010 and priority less than 0.5, list out their locations.
iv. For webpages related to conferences, they have the "conference" word contained in the location tag, <loc>. List out the locations of conferences which have only one webpage.
v. The file type of a location is indicated at the suffix of the last part in <loc>. List out the locations referencing MS word files (with .doc suffix) for those conferences have more than one webpage.

(b) Provide a DTD for the XML file in Figure 1. (5 marks)

Question 2 ***[Total = 25 marks]***

```
1.  <xs:element name="Customer">
2.     <xs:complexType>
3.        <xs:sequence>
4.        <xs:element name="Dob" type="xs:date" />
5.           <xs:element name="Address">
6.              <xs:complexType>
7.              <xs:sequence>
8.                 <xs:element name="Line1" type="xs:string" />
9.                 <xs:element name="Line2" type="xs:string" />
10.                <xs:element name="phone" type="xs:integer" />
11.             </xs:sequence>
12.             </xs:complexType>
13.          </xs:element>
14.          </xs:sequence>
15.       </xs:complexType>
16. </xs:element>

17. <xs:element name="Vendor">
18. <xs:complexType>
19. <xs:sequence>
20. <xs:element name="type" type="xs:string" />
21. <xs:element name="Address">
22. <xs:complexType>
23. <xs:sequence>
24. <xs:element name="Line1" type="xs:string" />
25. <xs:element name="Line2" type="xs:string" />
26. <xs:element name="phone" type="xs:integer" />
27. </xs:sequence>
28. </xs:complexType>
29. </xs:element>
30. </xs:sequence>
31. </xs:complexType>
32. </xs:element>
```

Figure 2.

(a) Re-write the above XML schema so that the element Address has a type and can be re-used in defining the Customer and Vendor elements. Indicate the locations of changes by using line numbers. (6 marks)

(b) There is a need to support addresses in 2 different places, HK and US. In addition to existing information, a HK address would include district and estate, while an US address would have state and zipcode information. Modify the answer in (a) to further support the extensions of the Address element. (10 marks)

(c) Revise the type definition of the phone element so that element values are restricted to a correct telephone number format in Hong Kong. (4 marks)

(d) There is a many-to-many relationship in between customers and vendors. Revise the above schema to establish this relationship. (5 marks)

Question 3 ***[Total = 25 marks]***

(a) Draw the corresponding DOM tree in Figure 1. You can skip the contents of the elements. (4 marks)

(b) Label the nodes of the tree in (a) with the indexing scheme of iNode. The maximum fanout is 5 for this case. Attributes of an element are considered as child nodes of it. (4 marks)

(c) Provide a SQL expression corresponding to the XPath expression '/urlset/url/priority' when iNode is used to store the data in Figure 1. (5 marks)

(d) Use the DOM tree in (a) to illustrate the potential structural conflict of between 2 transactions in concurrency control. (8 marks)

(e) Why node numbering scheme is important to support lock management in concurrency control? (4 marks)

Question 4 ***[Total = 25 marks]***

```
<!ELEMENT CallCenter (reception*)>
<!ATTLIST  CallCenter      location CDATA #REQUIRED
                           type CDATA #REQUIRED>
<!ELEMENT reception (customer*)>
<!ATTLIST  reception staffname CDATA #REQUIRED>
<!ELEMENT customer (name, message*)>
<!ATTLIST  customer        id CDATA #REQUIRED
                           package CDATA #REQUIRED>
<!ELEMENT name (#PCDATA)>
<!ELEMENT message (#PCDATA)>
```

Figure 3.

(a) For the DTD shown in Figure 3, it represents a call center that has many receptions receiving messages from different customers. Each customer may have talked to more than one reception. Provide a sample XML file for this DTD. (4 marks)

(b) Show the output structure after using XMill to compress the XML file in (a). (7 marks)

(c) Determine if the DTD shown in Figure 3 is XNF or not. If not, revise the DTD to conform to the XNF requirements; otherwise, explain why it is in XNF. (8 marks)

(d) A company would like to access the records of customer messages in the call center but does not know the DTD in Figure 3. The company issues a query in an XPath expression '/CallCenter/customer/message'. Which is a better approach, query re-writing or data re-writing, to handle the request? Justify your answer. (6 marks)

***** end*****