© The Hong Kong Polytechnic University

THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF COMPUTING

EXAMINATION

Course : MSc Scheme - 61030

Subject : COMP5323 Web Database Technologies and Applications

Group : 201, 202, 2011

Session : 2008 / 2009 Semester II

Date : 08 May 2009 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!

© The Hong Kong Polytechnic University

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema>
<xs:element name="Passenger">
        <xs:complexType> <xs:sequence>
        <xs:element name="name" type="xs:string"/>
        <xs:element name="retCard" type="xs:string"/>
        <xs:element name="homeTown" type="xs:string"/>
        </xs:sequence> </xs:complexType>
</xs:element>

<xs:element name="Flight">
        <xs:complexType><xs:sequence>
        <xs:element name="seats" type= "xs:int"/>
        <xs:element name= date" type="xs:date"/>
        <xs:element name="source" type="xs:string"/>
        <xs:element name="destination" type="xs:string"/>
        </xs:sequence>
        <xs:attribute name="flightId">
                <xs:simpleType> <xs:restriction base="xs:string">
                        <xs:minLength value="0"/><xs:maxLength value="5"/>
                </xs:restriction></xs:simpleType>
        </xs:attribute>
</xs:complexType> </xs:element>

<xs:element name="Airport">
        <xs:complexType><xs:sequence>
                <xs:element name="name" type="xs:string"/>
                <xs:element name="fee" type="xs:float"/>
        </xs:sequence>
        <xs:attribute name="airId" type="xs:string"/>
</xs:complexType> </xs:element>

<xs:element name="Reservation">
        <xs:complexType><xs:sequence>
                <xs:element name="date" type="xs:date"/>
                <xs:element name="flightRef" type="xs:string"/>
                <xs:element name="passRef" type="xs:string"/>
        </xs:sequence> </xs:complexType>
</xs:element>

<xs:element name="records">
        <xs:complexType> <xs:sequence>
                <xs:element ref="Airport" minOccurs="2" maxOccurs="unbounded" />
                <xs:element ref="Flight" minOccurs="1" maxOccurs="unbounded" />
                <xs:element ref="Passenger" minOccurs="0" maxOccurs="unbounded" />
                <xs:element ref="Reservation" minOccurs="0" maxOccurs="unbounded" />
        </xs:sequence> </xs:complexType>
</xs:element>
</xs:schema>
```

Figure 1.

© The Hong Kong Polytechnic University

Question 1 *[Total = 25 marks]*

Consider the XML Schema file in Figure 1.

(a) Provide a valid XML file with the fewest elements. (6 marks)

(b) Suggest a modification of the given schema file so as to define a key for the reservation element with keyref. The key should be a compound key of flight ID and date. (5 marks)

(c) Suggest a modification of the given schema file so as to re-define the number of seats be within the range from 100 to 180. (6 marks)

(d) Suggest a modification of the given schema file so that a passenger can have at most one reservation. (4 marks)

(e) Provide 2 constraints in the given XML schema file which DTD cannot represent. (4 marks)

© The Hong Kong Polytechnic University

Question 2 *[Total = 25 marks]*

(a) Develop a DTD for the description below. (8 marks)

The HallLife Monthly is a student newsletter and has a number of reporters. Each reporter is responsible to provide articles of a given section. The articles can be written by the reporters or other students. Articles are reviewed by an editorial board to decide whether to be published or not. The editorial board is elected among the student reporters. The information of each article includes its title, one or more author names, submission date, reviewer name, review date and a possible publication date.

(b) For the XML schema 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.

i. Provide the list of flights on April 20, 2009 which have Beijing as the destination. (3 marks)
ii. Find out the names and retCard information of all passengers originating from Hong Kong. (3 marks)
iii. Find the names of airports which appear the maximum number of times among all flights on April 20, 2009. (5 marks)
iv. Suppose there can be connecting flights with one intermediate stop when travelling on the same day. For example, a passenger from Hong Kong goes to Shanghai, then connecting to Beijing. Retrieve the pairs of flights from Beijing to Xian with an intermediate stop on April 20, 2009. (6 marks)

© The Hong Kong Polytechnic University

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

```
<?xml version="1.0" ?>
<courses>
<course name="WWW Databases" code="5323">
<lecturer fname="Alvin" sname="Ng">
        <affiliation>Business Computing</affiliation>
</lecturer>
<material>
          <required>
          <book><title>Query</title>
          <author>A. John</author> <author>S. Chan</author>
          <publ>Hallway</publ> </book>
          </required>
          <paper> <title>Query</title>
          <author>J. Peter</author>
          <journal>ACM SIGMOD Record</journal>
         </paper>
</material>
<room>PQ305</room>
<room>PQ604B</room>
</course>
<course name="Datamining" code="2115">
<lecturer fname="Vincent" sname="Chan"/>
<material>
          <required>
          <book><title>Query</title>
          <author>A. Peter</author>
          </required>
</material><room>PQ303</room>
</course>
</courses>
```

Figure 2.

(a) Provide the node numbers of the XML elements in Figure 2 using the ORDPATH method. You can ignore all the attributes. (4 marks)

(b) What are the possible node numbers of two lecturers after adding to the course 2115 before Vincent Chan with the answer of (a)? (4 marks)

(c) What are the context resemble measures with the partial matching approach for the Xpath expression /courses//required/book/title[./='Query']? (3 marks)

(d) Why keyword proximity is important in ranking elements during querying? (3 marks)

(e) Can the ORDPATH numbering method be used to support the DIL querying process? Would there be any change in the sorting step? (5 marks)

(f) Why node numbering scheme is important to support lock management in concurrency control? How the use of the CX lock in taDOM2 can improve the level of concurrency? (6 marks)

© The Hong Kong Polytechnic University

Question 4 *[Total = 25 marks]*

(a) Show the output structure after using XGRIND to compress the XML file shown in Figure 2. (8 marks)

(b) Describe how XGRIND can be applied to support an exact match query when an XPath expression is given. (4 marks)

(c) If XML files of the same XML schema are to be compressed, can XGRIND be modified to have a better compression performance?(4 marks)

(d) For query-aware compressions, why disobeying homomorphism is not always a drawback? What is the tradeoff? (5 marks)

(e) Discuss whether XQueC can be integrated with the content synopsis and position filtering indexing method to speed up query performance. (4 marks)

***** end*****