CSE 4/560 Project 3: XML and XQuery

Due 23:59 12/03/2018 EST

November 26, 2018

1 Submission

Failure to comply with the submission specifications will incur penalties for EACH violation

1.1 What to submit

A zip file has to be submitted through the 'submit_cse460'/'submit_cse560' submit script by 12/03/2018 11:59PM EST. **ONLY** zip extension will be accepted.

1.2 Zip file naming convention

Write your *ubit_proj3* (**NO SPACE!**) for the filename, e.g. *jsmith_proj3.zip*, where *jsmith* is the ubit. The project is an **INDIVIDUAL** project, copied solutions will be considered violations of academic integrity.

1.3 Sub-structure of zip file

- On unzipping the zip file, there should be a folder named with *ubit_proj3*.
- Under the folder *ubit_proj3*, there should be a .txt file named *ubit_proj3*.txt, this file contains your answers for all the problems.

2 Problem 1 (3 pts) XML Database Design

You are asked to design an XML database for TinyExpedia, which is a travel booking website. TinyExpedia has three types of data to manage: customer data, airline information and flight ticket booking data. Specifically, you are given the following list of requirements:

- TinyExpedia creates a record for each **customer**, and each record consists of:
  - A passport number, the passport number is also the key for the customer records.

  - A legal name.
  - A date of birth (DOB).
  - At least one address, each address consists of street, city, state, and zipcode.
  - A list of flight ticket orders that belong to the customer. (The requirements of order data will be given soon).

- TinyExpedia creates a record for each **airline**, and each record consists of:
  - An airline code, which is also a key.
  - An airline name.

- TinyExpedia also creates a record for each **flight ticket order**, and each record consists of:
  - An order number, the order number is also the key for the flight ticket order records.
  - At least one airline code, airline codes indicate which airline companies the order associates to.
  - A payment type, which indicates if the order is paid by a credit card, a check, or a debit card, note only these three types are valid types for payment type.

Write a DTD to model the above requirements. Use any xml validation tool to validate your DTD with your test data, e.g., an XML Validation tool can be seen at `https://www.xmlvalidation.com/`. For this problem, your solution should include your DTD and your test data that is valid against your DTD (note: test data shouldn't be empty, and your test data should be able to show all the modeling components of your DTD).

3 Problem 2 (7 pts) XQuery

Given the following DTD describes the information of the buildings of a university: *department* indicates the departments that are using the buildings, *name* is the name of the building, *type* indicates how the building is used, e.g., 'teaching' means that the building is used for teaching purpose, *year* is the year the building was built. Assume you have a xml database called *buildings.xml* that is valid against the given DTD, write the following queries in your solution file, use XQuery comments to separate your answers, e.g., (: answer for 2.1 :). Use eXistDB to test and verify your answers, the file path of buildings.xml should be **"/db/buildings.xml"**. **Note**: Your solution should include **only** your queries, please do **not** include the given DTD or any test data.

```
<!DOCTYPE buildings[
 <!ELEMENT buildings(building*)>
 <!ELEMENT building (department+, name, type+, year)>
 <!ATTLIST building id ID #REQUIRED>
 <!ELEMENT department (#PCDATA)>
 <!ELEMENT name (#PCDATA)>
 <!ELEMENT type (#PCDATA)>
 <!ELEMENT year (#PCDATA)>
 ]>
```

2.1 (2 pts) Find all the buildings which are used only as a library, i.e., the type is "library".

2.2 (2 pts) Find all the buildings which are used by both CSE department and EE department.

2.3 (3 pts) For every department, find the department name and the total number of buildings built in 1900 the department is using, you can ignore those departments that are not using any building built in 1900, the result should be sorted in lexicographical order by the department name.