CSCI 585- Database Systems

Spring 2015

Homework Assignment 4

Deadline: April 29th 11:59PM (Pacific Time zone)

Deadline: April 29th 11:59PM (Pacific Time zone)
**The goal of this assignment is to design and query an xml database using
oXygen XML Developer (download link is provided below).**

Project specification:
Consider a web service MyMdb "my movie database" which stores movies, reviews and reviewers.

For each movie, MyMdb stores the following information.
- Movie_ID, Title, Director, Writer, Genre, Release_Date, Runtime

Constraints:
- Release_Date is in the format(YYYY-MM-DD)
- Movie_ID, Title, Director, Genre and Release_Date are mandatory and occur only once
- Writer is mandatory and can occur more than once
- Runtime is optional

For each review, MyMdb stores the following information
- Reviewer_ID, Movie_ID, Rating, Review_Date

Constraints:
- Review_Date format is YYYY-MM-DD
- Reviewer_ID, Movie_ID, Rating, Review_Date are mandatory & can occur only once
- Rating values are in the range 1-10
- Order of elements for review is:
  - Reviewer_ID
  - Movie_ID
  - Rating
  - Review_Date

For each reviewer, MyMdb stores the following information
- Reviewer_ID, Reviewer_Name, Reviewer_DOB

Constraints:
- Reviewer_DOB format is YYYY-MM-DD
- Reviewer_ID, Reviewer_Name, Reviewer_DOB are mandatory & can occur only once

Data for movies, reviews and reviewers are in the mymdb.xlsx file.
- All dates are in the format: YYYY-MM-DD
- Writer column is comma separated list of writers
- Genre column is space separated list of genres

Assignment:
Part1: Database setup
A. XML schema (30 points)
   Create 3 xml schemas (movies.xsd, reviews.xsd, reviewers.xsd), which satisfy the above project description.
B. XML document (10 points)
   Create 3 xml files (movies.xml, reviews.xml, reviewers.xml) using the data from "mymdb.xls". Ensure that these xmls are valid according to their corresponding xml schemas.

Part2: XPATH & XQUERY

1. Find movie/s with the largest runtime.

2. Sort movies written and directed by "Christopher Nolan" in ascending order using Release_Date.

3. Show the total number of movies released in each year.

4. Group reviewers for movies "Goodfellas" and "Inception" using movie Title.
   Note: Show the movie Title and reviewer's name.

5. Find the movie/s with the highest average rating.

6. Group movies by Genre where Genre is either "Crime" or "Action" or "Sci-Fi".
   Note: Show the Movie_ID's for each Genre

Note: filter duplicate results, if any.

Submission Instructions:
You need to submit the following files (the filenames should be EXACTLY the same as mentioned below):

1. 3 xml schema files: movies.xsd, reviews.xsd, reviewers.xsd
2. 3 xml files: movies.xml, reviews.xml, reviewers.xml
3. Six xquery files: q1.xquery, q2.xquery, q3.xquery, q4.xquery, q5.xquery, q6.xquery

Useful Links:

- **oXygen XML Developer** Version 16.1
  http://www.oxygenxml.com/download_oxygenxml_developer.html
  Note: Use the trial license to complete this assignment.
- XML Schema
  http://www.w3schools.com/schema/default.asp
- XQuery
  http://www.w3schools.com/xquery/default.asp
- XPATH
  http://www.w3schools.com/xpath/default.asp