

MSDS 7330

File Organization and Database Management

Homework XML

Name:

This is a homework assignment for MSDS 7330, File Organization and Database Management. For this assignment, turn in a single pdf file containing all of your answers. The file should be named ;yourLastName;HW-XML.pdf. For example, the file name for my homework assignment would be 'EngelsHW-XML.pdf'. Insert your answer pages into this file with the answer for Question 1 inserted immediately after Question 1 and before Question 2, the answer for Question 2 inserted immediately after Question 2, etc. You may insert a front page containing your name and date if you do not wish to or cannot electronically add that information to the first page of this homework sheet.

Collaboration is expected and encouraged; however, each student must hand in their own homework assignment. To the greatest extent possible, answers should not be copied but, instead, should be written in your own words. Copying answers from anywhere is plagiarism, this includes copying text directly from the textbook. Do not copy answers. Always use your own words and your own code. Directly under each question list all persons with whom you collaborated and list all resources used in arriving at your answer. Resources include but are not limited to the textbook used for this course, papers read on the topic, and Google search results. Don't forget to place your name on the first page of the pdf document.

XML Database

Question 1 : The file `baseball_salaries_2003.txt` contains salary information for certain professional baseball players from the year 2003. Define an XML schema for this file. Write a Python script that processes this file and stores it in a single XML file: `baseball_salaries_2003.xml`.

Turn in the Python script, XML schema definition, and resulting XML file.

Collaborators:

Resources:

Question 2 : The file `baseball_salaries_2003.xml` contains salary information for certain professional baseball players from the year 2003. Write a Python script that processes the XML file from Question 1 to determine, for each position, the average salary of the players in that position. Note that the seven player positions that can occur in the input file are “Catcher”, “First Baseman”, “Outfielder”, “Pitcher”, “Second Baseman”, “Shortstop” and “Third Baseman”. The output should appear sorted in descending order of average salary.

Capture the resulting output in a file. Turn in a pdf of your script and the results.

Collaborators:

Resources:

Question 3 : Create an XML database called “University” for the University database in the textbook. Note that you will need to define the XML schema. Load the textbook relations into your *University* database and save the database in an XML file.

Using your *University* XML file, obtain answers for the following queries (write your queries in Python):

- 1) Produce a list of all the students in the student relation, including their ID, name and department name, sorted into ascending order by their name.
- 2) Produce a list of the names and salaries of professors in the Comp. Sci. and Elec. Eng. departments ordered by decreasing salary.

Capture the output in an XML file. Turn in your XML schema, University XML database, Python script and output XML file.

Collaborators:

Resources:

Question 4 : Compare and contrast your use of XML versus your use of MySQL for these data sets. Under what conditions does XML provide a better database approach? Under what conditions does MySQL provide a better database approach?

Collaborators:

Resources: