CSE 560: Project #2 (due 12/08/08)

Problem 1 (70 pts)

Suppose you have to represent the information about *scientific ancestry*. For every scientist you need to store:

- the name,
- the university awarding the PhD degree,
- the year of the PhD degree,
- the list of the scientists who are the graduated PhD students.

You can assume that each scientist has exactly one PhD degree and that the names of the scientists are unique.

For every university you need to store:

- the name,
- the country,
- the year when it was established.

You can assume that university names are unique.

- 1. Define the schema of the XML database containing the ancestry information using DTDs.
- 2. Give an example of a document instance which is valid under the DTD.
- 3. Write the following queries in XQuery:
 - (a) Q1: Find all Tarski's students.
 - (b) Q2: Find all Tarski's descendants who graduated from a US university.
 - (c) Q3: Who are the common ancestors of Bonner and Revesz?
 - (d) Q4: For every scientist, compute the number of descendants.

Problem 2 (30 pts)

Show how to represent the above database in the relational model using the interval representation. Using this representation, write the query Q1 in SQL.

Extra credit (20 pts)

Write queries Q2 and Q4 in SQL, using the representation from Problem 2.

Submission

Submit everything in electronic form using submit_cse560. There will be no deadline extensions.