WebDB Assignment 1 (Spring 2012)

There are 4 questions in this assignment.

Question 1 – Schema Design with IBM DB2 (30 marks)

The purpose of this question is to gain experience of querying XML data using IBM DB2. In PolyU, we encourage students to serve the community and there are funding available to support different initiatives. One good example is the CSLP grant. You can find the CSLP grant application form of SAO at http://www.polyu.edu.hk/sao/cslp/download.htm. Develop an XML Schema for the form.

Populate the XML schema with some data and answer each of the following queries with XQuery statements in DB2. Also, for each query, discuss and provide the expression if XPath alone is sufficient.

(a) List the names of the group leaders who are full time and COMP students.
(b) How many group leaders are also members of other groups?
(c) List out the project names which have activities held on June 20, 2012.
(d) List out the pairs of project names which have activities held on June 20, 2012 and at the same place.
(e) Provide the project name and the number of members whose project total expenditure is more than HKD 30,000 and number of activities planned is less than 3.
(f) Group the projects according to their number of service users. There are 3 groups: below 100, 101-1000 and more than 1000. Find the average amount spent for each service user for projects in the first group.

Question 2 - DTD (15 marks)

A *Directed acyclic graph* (*DAG)* has a unique name and consists of a set of nodes and a set of edges. Each node has a label that uniquely identifies it. A node may be a part of more than one graph. Each edge connects two distinct nodes and has an edge weight. An edge is directed from a "head" node to a "tail" node, i.e., an edge connecting head node *u* to tail node *v* is different from an edge connecting head node *v* to tail node *u*.

In a single graph, given a head node and a tail node, there is at most one edge from the head node to the tail node. A single graph may contain an edge from *u* to *v* and an edge from *v* to *u*, since these two edges are distinct. In addition, an important property of a DAG is that it does not have any cycles. A sequence of nodes forms a *cycle* if an edge connects each node in the sequence to the next node in the sequence and an edge also connects the last node in the sequence to the first.

(a) Design a DTD to represent the data model for the DAGs
(b) Is it possible to model the fact that a DAG does not contain a cycle? Justify your answer.

Question 3 Schema Design (20 marks)

This question is to provide a data model based on the television show “Star Chefs with Ma.” Do not worry if you have not watched this show! The show features a set of four “Star Chefs" in each episode and an expert would be in precisely one of the following cuisines: Chinese, Indo, Thai, and Japanese. Star Chefs are initially located for the first episode. Their name serves to identify them uniquely.

Each episode, identified by an episode number, features a competition between a challenger and one of the Star Chefs. Each challenger is also quite famous; his/her name and restaurant are enough to identify the challenger. The challenger selects the Star Chef he/she wants to compete with. Each competition features a secret ingredient unveiled at the beginning of the episode. The secret ingredient never repeats, i.e., two different competitions do not have the same secret ingredient. The challenger and the selected Star Chef each have one hour to prepare a dish that articulates the theme ingredient. At the end of the hour, a panel of four judges tastes these dishes. Each judge awards each competitor scores of two categories: taste and appearance; the maximum score of each category is 20. The competitor with the total maximum number of scores wins. If there are ties, the head judge, one of the four judges, would determine the winner. The winner would become one of the Star Chefs afterwards. A person can be a judge in multiple competitions.

a. Using XMLSpy or similar tool to develop a data model in XML Schema to support an XML database application.
b. Provide the XQuery statement that can determine which Star Chefs staying the longest of the show.
c. Discuss the possible ways to model M:N relationships with XML Schema, and their pros and cons together.

Question 4 – Oracle and another XML DB (15 marks)

This question is to compare the XML support of IBM DB2 with another native XML database, Senda (http://www.modis.ispras.ru/sedna/). The comparison should include the following aspects:

- Storage method
- Schema/DTD design support
- Query support and indexing
- Transaction management
- Interfacing with relational database systems, like MS SQL Server