Distribution: Friday, November 16, 2007 (Revised November 21, 2007)
Due date: 1:00 pm, Friday, December 7, 2007

Instructions

1. This assignment is worth 150 marks and counts for 15% of your final grade in this course.

2. Please read instructions carefully before you proceed with the assignment. Failure to follow instructions may affect your grade.

3. You must submit your assignment both in paper format in the drop-box in Bahen room BA2220, and electronically via CDF by the due date.

4. Late assignments will not be marked.

5. To submit your assignment through CDF

    submit -c csc343h -a A3 cover.txt Assignment3(.pdf or .ps depending on file type)

    For more information on submit, use **man submit**.

6. You may work in pairs for this assignment. Please submit only one copy of your solutions per team.

7. All questions regarding this assignment must be asked through the newsgroup, or in person during the instructor's office hours. Questions sent to the instructor via email may not be answered. You must *not* post any part of your solution to the newsgroup — this is considered cheating, and you may be penalized for doing this.

8. Assignments received at 1:01 pm on the due date or later, will *not* be marked.

Question 1. Entity-Relationship Model [30 marks]

ReadBooks is a running shoes manufacturer. To support its operations, *ReadBooks* maintains a complete database of all the running shoe types that it offers for sale. The database stores sale figures for each country where the company sells shoes. Shoes are targeted for specific cultural markets so that a style of shoe sold in one market may remain approximately the same but have external stylistic additions or a different sole added to fit a different market. For example, the United States has a large basketball shoe market, but the turns and twists on the basketball court are the same as those found on a volleyball court. Taiwan has a huge volleyball shoe market. Thus, the same shoe is sold in both places but with a different set of markings and names.

Some shoes are also different gradations of the same class of shoes, i.e., the stitching may be a little less well done to save money or a cheaper form of foam cushioning or reinforcement plastic may be built into the cheaper shoe. The foreign exchange rate and inflation in some markets dictate that only lower priced shoes are sold in these countries. Markets also differ by climate. For example, hot, humid countries need mesh inserts to remove moisture from the shoes and polypropylene linings to absorb moisture from the foot.

Shoe models may create a new line of show products, or may substitute an older model that is taken off the market. For new models you need to keep information on the designers (name, address, phone number) and the date it was introduced. For models that substitute an older model you need to store the relationship between the new and the substituted model, also the date it was introduced. When a model is an adaption of another one, you need to capture the relationship between the two, the date when the adaption was introduced and the country for which the adaption is intended.

A database expert (you!) has been called in to design a new database for *ReadBook*. The database needs to accommodate for each shoe model, its name, the older model it has replaced (if any), manufactured cost, retail price and shipping costs in each country it is sold, design type, soles type, cushion support type and the major sports it is intended for. In cases where a model is adapted for a particular country, the database needs to store a model name, the name of the adapted model, the designated country and climate type for which the model is designed.

In addition, the database needs to store yearly sales by country, including number of pairs sold, price, and net profit, as well as climate type for each country.

Design and draw an ER schema that captures the information given above. Your schema should model explicitly entities and relationships in the domain of shoe sales around the world, also their attributes, generalization relationships, keys and cardinality constraints.

Here is a sample of the kinds of data your database is supposed to handle. Note that your ER schema may have very different structure from what is shown below.

Shoe Models

Model Name	Subst Model	Mfg Cost	Retail Price	Type Design	Type Soles	Cushion Support	Major Sport
Actif	Pasif	$12	$89	RX45	Neo	Neo	Bsktbl
HiJump	Hitop	$23	$95	K389	Poly	Poly	Runng
RB Xtra	Actif	$28	$113	RX45	Neo	Neo	Bsktbl
RB Tops	Pasif	$13	$93	RX45	Poly	Poly	Bsktbl
Air 1000	Actif	$24	$109	K389	Prion	Prion	Runng
Air 2000	Air 1000	$43	$189	K389	Prion	Prion	Runng

Yearly Sales by Country

Country Name	Model Name	No. Sold	Net Profit
Mali	Actif	23,000	$1.2M
Korea	Air 1000	45,000	$4.3M
Japan	Air 2000	109,000	$8.9M
Ireland	Hitop	12,000	$0.9M

Model Adaption

Model Name	Lower Priced Model Name	Designated Country	Climate Type
Air 2000	Air 500	Ethiopia	Dry
RB Tops	RB Bottoms	Ethiopia	Dry
RB Xtra	RB Less	Ethiopia	Dry
Air 1000	Air Less	Panama	Humid
Actif	R Sets	Panama	Humid
HiJump	FlatFeet	Bhutan	Mountainous
Pasif	Flyers	Bhutan	Mountainous

Question 2. Relational Schema Design [30 marks]

Assume the following workload for the database:

- **Operation 1**: Insert a new shoe model [10 times/year]

- **Operation 2**: Insert an adapted version of a shoe model [50 times/year]

- **Operation 3**: Insert a country [2 times/year]

- **Operation 4**: Add a country sales report to the database [100 times/year]

- **Operation 5**: Report annual sales in all countries for a particular show model [20 times/year]

- **Operation 6**: Report all sales for all shoe models [10 times/year]

- **Operation 7**: Report all adaptions for a particular shoe model, and their sales in all countries [10 times/year]

You may assume that on average there are 500 shoe models, and 80 countries in the database. On average, there are 20 shoe models sold in each country. You may make additional assumptions about the workload as you need them. Please be sure to state clearly any such assumptions you make.

Design a relational database from the ER schema you created for question 1. Show explicitly the input and output of each step of your design process.

Question 3. Functional Dependencies [40 marks]

Consider the following functional dependencies over the attribute set `ABCDEFGH`:

$$A \rightarrow E, \; BE \rightarrow D, \; AD \rightarrow BE, \; BDH \rightarrow E, \; AC \rightarrow E,$$

$$F \rightarrow A, \; E \rightarrow B, \; D \rightarrow H, \; BG \rightarrow F, \; CD \rightarrow A$$

(a) [10 marks] Find a minimal cover for this set of functional dependencies.

(b) [15 marks] Decompose the relation `ABCDEFGH` into a lossless 3NF schema.

(c) [15 marks] Check whether your answer to (b) is in BCNF. If not, decompose it into a lossless BCNF schema.

Question 4. XML DTDs and Queries [50 marks]

Consider the following student database. We are interested in placing this information in document format.

Student Information:

STID	Name	Phone	St. No	Street name	City	Prov.	Courses taken	Student Status	Degree
995435245	Ali	416-4245979	406	Main	Toronto	ON	1 3 4 5	Degree Student	BSc
995267842	Bob	613-5345660	12	Charles	Ottawa	ON	3 4 5	Special Student	
997458623	Carlos	905-2348638	5	King	Oshawa	ON		Special Student	
998112455	Fernando		101	Avenue	Montreal	QC	2 4 8 10	Degree Student	BEng
993457622	Jason	204-4562983	32	Main	Winnipeg	MB	1 2 3	Degree Student	BA
996112321	Jun	204-7893242	160	Pembina	Winnipeg	MB	5 7 10	Degree Student	BEng
995987345	Lee	647-9982342	35	Charles	Toronto	ON	3 4 7 9	Degree Student	BSc
997821345	Lueng		11	Yong	Toronto	ON		Degree Student	BSc
996453222	Mark	613-4561190	30	University	Ottawa	ON	1 5 6 7	Special Student	
997424563	Maria		11	Queen	Ottawa	ON	5	Special Student	
997345632	Nicolas	613-8932456	13	Cumberland	Ottawa	ON	9 10	Special Student	

Courses information:

ID	Course Name	Department	Allowable Student Status
1	Introduction to Databases	CS	Degree Student, Special Student
2	Numerical Methods	CS	Degree Student, Special Student
3	Operating Systems	CS	Degree Student
4	Computer Graphics	CS	Degree Student
5	Calculus Sci I	MAT	Special Student
6	Complex Variables	MAT	Degree Student, Special Student
7	Groups and Symmetry	MAT	Degree Student
8	Introduction to Economics	Economics	Degree Student, Special Student
9	Microeconomic Theory	Economics	Degree Student
10	Energy & Resources	Economics	Special Student

(a) [10 marks] Define a DTD, *students.dtd*, for documents that list all courses offered, and then the courses taken by each student. The DTD should allow each registered student to take 0 or more courses. Phone numbers are optional for. Also, only degree students have a degree attribute. A course may be taken by students whose status is degree, special or both. A student may be taking a course for which she can't get credit (e.g., a special student taking a course for which only degree students get credit).

(b) [10 marks] Provide an XML document, *students.xml*, that captures the information given in the tables above and is consistent with *students.dtd*.

(c) [30 marks] Represent the following queries in XQuery:

I. Display student name and status for all students. Sort the result by student name in ascending order.

II. Display the number of students for each city mentioned in the database; sort the result by number of students in a descending order. You need to output city names and the number of students for each city.

III. Display the number of students who can get a credit for each course. This includes all students whose status is allowed by any given course. Display course ID, course name and the number of students who are allowed for each course.

IV. Display the name of students who have no credits for CS department courses. Include in your answer students who sit in a CS course without having allowable student status. Sort the result by student name in an ascending order.

V. Display the names of students who take at least two courses they are allowed to get credit for from the same department. Sort the result by student name in descending order.