CSE 5330: Database Systems I

"Development of a Public Health Information System"

Deadline: November 23, 2014, via Blackboard

Goal: Design and develop a database as a backbone for a public health surveillance system and, based on this, a web-based public health surveillance application. This database is intended to be used by the personnel who works in public health. The application has two functions: it will facilitate the input of information during the surveys and it will allow the public health professionals to navigate through public health indicators and time-series, based on the previously collected information.

The overall procedure

Public health officials collect information in population surveys: this category visits houses (sampling the population) every given period of time (i.e. yearly/every semester) and surveys citizens about various aspects of their health. Public health officials who want to view various health indicators. This process can be part of what we call Health Surveillance. Those health indicators will basically be the results of a series of queries on the data collected by the interviewers. Each citizen who is interviewed, will be asked about their demographics: gender, date of birth, ethnicity, and marital status. We will need to know about the occupation of the citizens: a citizen might work at two different types of jobs. There are the following types of jobs: office employee, farmer, and industrial labor worker.

 Some citizens might have chronic diseases, and we need to capture this information. In specific we need to know the name of the disease (diabetes, hypertension, COPD, myocardial infraction) about the year they were diagnosed with. We need to consider that each citizen might have more than one chronic diseases. Since nutrition is important for our health and well-being, the interviewer will ask about the consumption of fruit servings weekly and the fast food consumption weekly. In addition, the habits will be part of the survey; we need to ask the citizens to report any tobacco and alcohol consumption.

We assume that, depending in the location of each survey, which, by the way, is tracked down, there is some preloaded contextual information about the environmental parameters.

You will need to capture the carbon monoxide measurements, the pollen concentration and traffic accidents of a specific location. Each of those data is produced individually by different agencies. Finally, the database has information about healthcare resources (location of hospitals

and rehabilitation centers) of each location. Each citizen is interviewed at a specific location, and we need to know this location.

Important notice: we are not interested about the individual health of each citizen, but we care about the health status of the population as a whole. Therefore, our main focus is on the survey information and not the patient information. Think carefully about the above consideration and how it is going to affect your schema.

Task 1 *(11 points)*

Provide the appropriate Entity-Relational Diagram which conceptualized the above process. Please use all the appropriate notations on your ERD.

Task 2 *(11 points)*

Apply the ERD to relational algorithm in order to transform the ERD to a relational schema and present the relational schema (still at this point your task is DBMS independent). Describe, in detail, step by step your decisions.

Task 3 *(12.5 points)*

Write appropriate CREATE SQL code to design your database schema. The schema must be based on what your colleague designed in Task 2. In your code, define all the appropriate constraints, including ON DELETE and ON UPDATE constraints. Please, deliver, alongside with your code, the schema, with the use of the reverse engineering functionality.

Task 3.1 *(5 points)*

Insert Synthetic Data:

36 surveys: 6 for the year 2012, 6 for the year 2013 and 6 for the year 2014.Add the following locations (small towns): Town A, Town B and Town C. Please add 4 surveys per year for Town A, 5 for Town B and 3 for Town C. Demographics, nutrition and habits data can be of your own choice. For each one of the surveys insert chronic diseases form the following disease list: diabetes, hypertension, COPD, myocardial infraction. In each survey assign an occupation of the surveyed person among the following: office employee, farmer, and industrial labor worker

For each town add the following information:

4 traffic accidents for Town A, 3 for Town B and 2 for Town C (some can be car, motorcycle, and pedestrian). Try to assign accidents in each year. Add 18 CO_2 measurements, two for each town yearly. Finally, add 2 healthcare infrastructures for each town. Those health infrastructures can either be hospitals or rehabilitation centers.

Task 4 *(17.5 points)*

Develop appropriate queries for the following

1. The surveyed female population for each town during the year 2012
2. Most common accident type for Town A.
3. Prevalence of diabetes in Town B for the year 2014
4. Town with the lowest diabetes prevalence for the year 2012
5. The town(s) where no accidents have been reported
6. The availability of health resources for Town C for the whole observation period. This is the ratio of citizens per health infrastructure.
7. The change in the smoking habit between in 2014 compared to 2012, for all towns

Prevalence of a disease is defined as the proportion of a population found to have a condition. It is calculated by comparing the number of people found to have the condition with the total number of people studied, and is usually expressed as a fraction, as a percentage or as the number of cases per 10,000 or 100,000 people

Task 5 *(32 points)*

Develop a web application (using java preferably) based on your database, and queries with the following functionalities

    (i)   A web page with a form where the user can insert the information during the interviews with the citizens.

    (ii)   Exploration of the information coming from the queries of Task 4.

    (iii)   The user should be able to select a town from a list and view aggregated information about the population of this town.

12 points: evaluation of 15 minute in-class presentation (3 points per student)

Technical considerations

Your DBMS should be MySQL Server. Your web application can be built with any programming language you prefer. Java can be a good option. Do not use tools like i.e. PowerBuilder.

Deliverables (via Blackboard)

1. **A word document with the name Team_X_Database (X is the number of your team) which is going to include the following:**
    a. The ERD of the process
    b. Description of the transformation of the ERD to relational algorithm
    c. The SQL Create code and the relational diagram (via reverse engineering)
    d. The INSERT code
    e. The SQL code for your queries and screenshots with the output of each query.
2. **A word document (1.5 – 2 pages) with the name Team_X_Web_Application (X is the number of your team) which describes the technical considerations for your web application. Refer to the following:**
    a. Language you used for the web development
    b. Methodology for the connection of your database with the web application
    c. Instructions on how to setup the application on a new machine.
3. **A folder with the name Team_X_Source (X is the number of your team) including:**
    a. SQL script files for your schema and data insertion.
    b. The files with the source code of your web application

Compress the above 1, 2 and 3 into a zip file with the name Team_X_Project2_Submission (X is the number of your team) and upload this zip file on the Blackboard before the deadline.

Grading Policy for Project 2

Each one of the four members of the team has a very distinct role
Student 1: tasks 1 and 2 (22 + 3 points for the presentation =25 maximum)
Student 2: task 3, task 4 and queries 6-7 (22 + 3 points for the presentation =25 maximum)
Student 3: queries 1-5 and collaboration with student 4 (approximately one third of the effort) for the task 5 (22 + 3 points for the presentation =25 maximum)
Student 4: task 5, approximately two thirds of the effort required for the web app development (22 + 3 points for the presentation =25 maximum)

Good luck!