Lab 2: Database requirements analysis and design

Due date: Wednesday, January 18, end of lab period/midnight.

Note: This lab involves a number of deliverables that are due before January 18. The due date above is for the final set of the deliverables.

This is a **team lab**. Each team submits one set of deliverables. Each team member receives the same grade for the assignment.

Course Project Information

This quarter, CPE 366 comes with a group course project. General information about the project is provided below.

Customer. Our customer for this quarter is Dr. Dan Peterson from the Animal Science program at Cal Poly.

Project. You are asked to design a database for storing information about the regulatory sequences of proteins found in the promoter regions of the genes responsible for the synthesis of these proteins. You are also asked to implement a database application that can analyze the similarities and differences between the collections of regulatory sequences for different proteins.

Lab Overview

In the course of this lab, you will

- receive an overview of the customer requirements via direct communication with the customer and a tour of the Department of Biology laboratory facilities in which DNA analysis is conducted (January 9);

- prepare an initial design of the database for the project (January 11 - 18);
- elicit additional requirements for the database via direct communication with the customer (January 17);

The lab assignment spans three lab periods (technically, four), and has three distinct steps, each of which serves a different goal and yields a set of deliverables. This document specifies your assignment for each of the periods as well as the final deliverable.

The overview of the lab structure is below.

January 9, 2012 *(This has already happened. I am keeping this part in, as it is all part of the process).* Teams were announced in class. Following that, we had a meeting with the customer and a presentation of the project by the customer. A hardcopy of initial project documentation has been released.

January 11, 2012 Teams work on understanding the nature of the project, and on preliminary conceptual design.

January 17, 2012 Q&A session with the customer during our lab period. Each team will have an allocated portion of time for questions. Questions/answers must be recorded and posted later.

January 18, 2012 teams work on conceptual design, discuss it with me.

Assignment

January 9: Group Creation and Initial Presentation

We have formed eight (8) teams. Two teams have five members, the remaining teams have four members. I made an effort to balance the teams for overall expertise in all areas except web programming.

The project can be executed in a number of different ways. The customer does not have a preference as to the eventual architecture of the software solution. There are two possible types of solutions to explore:

- **Standalone desktop application.** This requires solely the skills you acquired in CSC 365. A standalone application is best implemented in Java using either Oracle DBMS or MySQL DBMS as the database back end. The deliverable will be a desktop application that installs on the customer's computer. (No Oracle installation from scratch is expected/required).
- **Web-based application.** This requires skills not taught in CSC 365 and therefore, each team that wants to pursue a web-based application will be on its own as far as software development and database administration support is concerned. A web-based application will hosted on a CSL Virtual Machine (VM), which will be provided for each team upon request. The front-end of the application will be a collection of dynamic web pages that present the database to the user and allow for interaction as specified in the list of use cases you will receive. The back end of the application will be MySQL DBMS installed on the VM.

In the next two weeks each team needs to determine which path it wants to take as far as application development is concerned. I will facilitate the setup process for both the teams that choose to implement desktop applications and the teams that choose web-based applications.

Deliverables. Each team shall come up with a name. Further, you shall create a team wiki page on the csc366 wiki (see below). The page, at the very least shall contain team name and names and emails (Cal Poly userIds are fine) for all team members.

January 17: Question and Answer Session

Our next interaction with the customer is a Q&A session on January 17. During this session each team gets a chance to ask questions about the project.

Preparation. Before *January 17*, each team shall study the materials provided by the customer. The goal of each team is to identify any aspects of the application domain, information about objects to be stored in the database, customer needs (w.r.t. the database), *that require clarification.* Each team shall prepare a list of questions that the team wants to ask the customer during the Q&A session. Note, that in addition to the customer, you may direct your questions at the instructor.

Please note: initial list of questions **must be prepared** by each team independently of other teams. However, once lists are prepared, I allow for teams to compare them. I expect that many questions will coincide or be similar (it is ok if multiple groups have similar/same questions). It is also ok for one group to come up with a question, no other group has asked — there is no need for other groups to add it to their lists, since each question only needs to be asked once, and all teams benefit from the answer.

You have a week to prepare for the Q& A session, so, please, be thorough in your preparation. Any aspect of the project that you do not understand needs to be clarified.

Each team shall post its list of questions to the csc366 wiki (inside the team's workspace).

Q&A session. Each team will be given around 10 minutes of time to ask their questions and receive answers from the customer and the instructor. We will set the order of questioning by a simple lottery at the beginning of the class. I strongly suggest that each team sits and works together during the Q&A. Please note the following:

- Listen carefully to the questions other groups are asking. It is very likely, that some of your team's questions will be posed by teams who get to ask questions earlier.
- If your group runs out of questions yield your time to the next group.
- If your group has unanswered questions left after its Q&A period, wait until all groups ask their questions. If the question has still not been asked, you will have a chance to ask it at the end of the lab.
- You are certainly allowed to ask questions that are not on your list - some questions might occur to you as a result of things said during the Q&A session.

Each team is responsible for recording the answers to all questions its members have asked (whether those questions were prepared or were asked on the spot).

After each team had a chance to ask questions, if any additional questions remain, a free-form Q&A period will begin and will last until the end of the class. Any questions that went unasked, or unanswered during the Q&A session should be submitted in writing (via email) to the instructor. The questions will be forwarded to the customer, who, in turn, will provide answers in writing as soon as he can.

Deliverables. After the Q&A session, all teams, in collaboration, shall create the project knowledge base as part of the csc366 wiki. The structure of the knowledge base is left up to the teams (generally speaking, you want to break the entire knowledge base by topic: e.g., group information about one type of data in one place, and information about another type of data — in another). At the "bottom" of the knowledge base hierarchy shall lie wiki pages that contain questions that were asked and answers that were received from the customer (plus any additional information that teams think may be useful). A single question can appear in multiple places in the knowledge base (as it may be applicable to more than one "leaf" topic).

The knowledge base is one of the two **main deliverables** for this lab. The initial state of the knowledge base shall be completed by the main Lab 2 due date. Since the knowledge base is part of the csc366 wiki, feel free to maintain and update it throughout the project. The intent is to have a useful and accessible resource for your further development.

Note: You can view the knoweldge base created by students from the previous CSC 366 classes here:

https://wiki.csc.calpoly.edu/cs366-2008/wiki/Q&A_Information
https://wiki.csc.calpoly.edu/csc366-2009/wiki/KnowledgeBase
https://wiki.csc.calpoly.edu/csc366-2010/wiki/KnowledgeBase

January 11 — January 18: Database Design

Assignment. The main task of this lab is for each team to prepare an initial full database design document. By the end of the *January 18* lab session, each team shall have the initial design document ready to be submitted by the end of the day.

Our *January 11* lab period will be devoted, in part, to the preliminary analysis of the customer documentation. We will attempt to figure out the overall structure of the data model. (Additionally, teams can spend time putting together the lists of questions).

During the *January 18* lab session, each team will will work with the instructor on the full database model. Depending on the level of progress, we may conduct one or two structured activities at the beginning of the lab period, or, the entire lab period may be given to the teams for team meeting activities, with instructor participating in each team's meeting in turn.

Deliverable. Each team is given time until **Wednesday, January 18, midnight** to finalize and submit the database model document. Lab 3 assignment will be handed out on January 18, and will require teams to proceed with the next step of the project starting January 19 (i.e., prior to our next lab period). I will grade the database design documents over the weekend and will hand them back to teams on January 23 during the lab period. I will discuss the design briefly with each team.

The design document prepared and submitted by each team shall, at a minimum, contain the following information:

- List of entity sets for the proposed database.
- List of attributes for each proposed entity set.
- Identification of primary keys for each proposed entity set.
- Identification of all weak entity sets in the proposed database, and of their discriminating attributes.
- Identification of any other entity set constraints for the proposed entity sets.
- List of relationship sets for the proposed database. For each relationship set, the following must be indicated:
  - All participating enitity sets.
  - All clarifying relationship set attributes.
  - Type of the relationship set (one-to-one, one-to-many, many-to-many).
  - If the relationship set is an identifying one for a weak entity set.
- List of other relationship set constraints.

- List of any class hierarchies and/or aggregates in the database model.
- List of relationship sets (with all the information as above) associated with the aggregates.
- E-R diagram of the proposed database design.

The design document must be typeset (handwritten submissions will not be accepted). See comments above about the software to use to draw the E-R diagram. The E-R diagram shall contain all entity sets and relationship sets. It shall also contain all key attributes for all entity sets and all relationship set attributes. Other attributes can be omitted from the diagram for clarity. All constraints that can be shown on the diagram, shall be shown.

The design document shall begin with the name of the group, and the list of group members. Note, that in general, **all documents** submitted by each team during the course of the project must contain the team name and the list of students.

Course Wiki

The course wiki is

`https://wiki.csc.calpoly.edu/366-Winter2012`

The wiki will be set up and prepared for your use on Tuesday, January 10.

Deliverables Overview and Submission Instructions

Overall, the lab has the following formal deliverables:

No.	Deliverable	Due:	Softcopy	Hardcopy
1	Stub of team wiki page	January 17, 4:10pm	Yes	No
2	List of questions for Q&A session	January 17, 4:10pm	Yes	No
3	Project knowledge base	January 18, midnight	Yes	No
4.	Database Design document	January 18, midnight	Yes	Yes

Submission. Softcopy submissions must appear on your team's wiki page by the due date/time. Your team's contribution to the project knowledge base shall be up on the wiki by the due date/time.

Hardcopy submissions: usually are required. Due to printing restrictions in CSL, I am waiving this requirement for the course. I will print your softcopy submission.

Please, keep the soft copies of all submitted documents. You will be working with them in the labs that follow.