

Lab 2: Database requirements analysis and design

Due date: Tuesday, April 12, midnight.

Note: This lab involves a number of deliverables that are due before April 14. 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 customers for this quarter are Dr. Chris Kitts and Dr. Michael Black from the Department of Biology at Cal Poly.

Project. The customers ask us to design a database for storing information about forensic signatures of *E.coli* bacteria collected from different animals and from the environment. *E.coli* bacteria have a large number of different *strains*. The database will support a variety of *microbial source tracking* research at Cal Poly. The forensic signatures, otherwise known as *pyroprints* or *pyrograms* are a result of using DNA sequencing technology called *pyrosequencing* to obtain quantitative information about the structure of a number of short DNA fragments. The customers ask us to develop a prototype web-based application capable of performing a number of basic database manipulations (storage of data, browsing of data) as well as the key forensic analysis operation: comparison between forensic signatures (pyrograms) from the database.

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 (April 5);
- elicit additional requirements for the database via direct communication with the customer (April 7);
- prepare an initial design of the database for the project (April 12);

The lab assignment spans three lab periods, 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.

April 5, 2011: 8:10 - 11:00am: Groups are announced in class. Following this, we embark on a field trip to the Department of Biology. During the first part of the field trip, you will meet one of the customers, Dr. Michael Black, who will introduce the problem for you, will discuss the topics of microbial source tracing, digital forensics, and will present the need for the database and the software suite you will be building.

April 7, 2011: 8:10 - 9:30am: Q&A session with the customers, Drs. Black and Kitts. Each team will have an allocated portion of time for questions. Questions/answers must be recorded and posted later.

April 7 - April 12, 2011: teams work on the full database design.

April 12, 2011: 9:40 - 11:00am: teams discuss their designs with me.

Assignment

April 5: Group Creation and Initial Presentation

We will spend the first few minutes of the April 5 class on group creation. Group lineup will be announced by the instructor at the beginning of the class. We will ensure that the group lineups are feasible (e.g., that we are not missing three people on a single 4-person team). At the moment, there are 18 people in the class. We are looking to form four teams: two teams with five people and two teams with four people in them.

Following the team formation, the class will move to the Department of Biology seminar room on the first floor of the Fisher Science building, where Dr. Michael Black will join us for the presentation of the project. The main purpose of the presentation is to give you a good understanding of what the project entails. After the presentation, a tour of the Department of Biology laboratory facilities will follow. You will get to see the process of creating the forensic DNA footprints (pyrograms) that will need to be stored in the database. At the end of the tour we will release customer-supplied documentation:

- Project and database description/requirements.
- Use Cases.

These two documents, combined with the data, which will be released upon your completion of Lab 2 will serve as the basis for your project.

Assignment. Field trip attendance is required for everyone. During the presentation, you are welcome to take notes - those will come in handy later. Upon receiving customer documentation, study it in preparation for the April 7 Q&A session.

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

April 7: Question and Answer Session

The process of designing your database will come in two steps. The first step is a Q&A session with the customers, during which each team gets a chance to ask questions about the project. The second step is the actual database design process that will take place after the Q&A session.

Preparation. Before *April 7*, 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 customers 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.

Also, note that we are on a rather tight schedule: each team has about 48 hours to come up with a list of questions.

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 15-20 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 customers, who, in turn, will provide answers prior to the Tuesday, April 12 lab.

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 the teams asked and answers 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>

April 17- April 12: 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 *April 12* lab session, each team shall have the completed design, to be submitted by the end of the day.

During the *April 12* lab session, each 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 **Tuesday, April 12, midnight** to finalize and submit the database model document. I will try to return graded designs back to each team during the April 14 lab period which will be devoted to analysis of the submitted designs and preparation for Lab 3: the relational model of the database.

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 entity 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.

Deliverables Overview and Submission Instructions

Overall, the lab has the following formal deliverables:

No.	Deliverable	Due:	Softcopy	Hardcopy
1	Stub of team wiki page	April 7, 8:10am	Yes	No
2	List of questions for Q&A session	April 7, 8:10am	Yes	No
3	Project knowledge base	April 12, midnight	Yes	No
4.	Database Design document	April 12, 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: Leave a hard copy of your team's design document with me during my Wednesday, April 13 office hours (9:00am – 12:00pm) (or earlier). If I am not in the office, leave the document in the yellow envelope pinned to the corkboard outside my office (14-215).

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