Lab 3: Database design

Due date: January 30, in-class.

This is a **team lab**. Each group submits one set of deliverables.

Lab Overview

This lab includes the following tasks:

- Revisit your Lab 2 design submission and enhance/update it to take instructor comments (once you receive them!), and unsatisfied customer requirements into account.
- Determine database constraints that are not expressible outright in the E-R model.
- Prepare the relational database schema for your design.

Assignment

Database Redesign

You should receive your initial design documents with instructor's comments on Monday, January 23 during the lab period.

Each team shall prepare a revised design document. The document shall contain a revised version of your Entity-Relationship model. The initial model was built based on incomplete knowledge, and thus, needs to be be extended (as well as corrected as needed).

The redesign shall essentially follow the same structure as the initial design and shall contain the Entity-Relationship Diagram of the database, accompanied by the lists of entity sets (with all attributes) and relationship sets (with any attributes, multiplicity specification and full lists of participating entity sets). All weak entity sets shall be identified together with their identifying relationship sets and owners.

The revision shall be prepared as a single, text-processed file, and submitted to the course Wiki preferably in PDF format. All figures, tables and diagrams shall be included in the design document. The document **shall contain the team name and the names of all team members**. You can combine the contents of the revision with the new material you have to submit into a single document.

The design document must be typeset (handwritten submissions will not be accepted). The E-R diagram shall be designed using drawing software. (if you are using Windows, you can use MS Powerpoint; if you are using Linux, CS labs have xfig, which allows exporting designed diagrams into .eps (Encapsulated Postscript) format).

Change Log

To simplify grading, and to simplify tracking the changes in your design, each team shall compile a changelog document. This document shall include the following:

- List of specific changes in the design of the database.
- List of responses to any instructor's comments on your Lab 3 submission, which were not addressed by the redesign (e.g., you believe that your original design already does what was needed).

The changelog shall be maintained on the wiki. A hard copy version of the changelog is one of the deliverables for this lab.

Database Constraints

The Regulatory Sequences database may have constraints on the contents (i.e., on the data that can reside in the database at any moment of time) that cannot be expressed in E-R model, but that, nevertheless need to be articulated. For each constraint discovered by each team, the team will need to make an eventual decision on how this constraints will be dealt with. (The standard options are either to monitor the database, and resolve inconsistencies as they come, or prevent the constraint violation in the software by running appropriate checks).

An example of a constraint that exists in the database is:

- A regulatory sequence string must match the substring of the promoter sequence for the same gene that starts at the position indicated, and goes in the direction specified by the sense attribute.

You shall compile a list of the database constraints. You will maintain the list on the wiki. The hard copy of the list will also be one of the deliverables.

Note: At this stage, you need not make any specific decisions on **how** you will be dealing with each constraint. You are simply asked to enumerate them.

Logical Database Design

Based on your E-R model redesign, each group shall prepare the initial logical database design.

The logical database design is the relational database schema obtained from your E-R diagram. Each group shall prepare the following:

- Database description, which consists of the following information:
    1. List of relational tables in the database.
    2. For each relational table, list of all attributes.
    3. Identification of primary keys for **all** tables.
    4. Identification of any foreign keys (this can be done on separate lines, in the form, `"Attributes X, Y, Z are a foreign key referencing table R"`.
    5. Specification of any constraints on the database that cannot be preserved in the database schema, but must be kept track of by the software.
- SQL DDL commands creating the database.

Submission Instructions

Each group shall document their design activities fully on the wiki. Each deliverable shall exist in soft copy as part of the team's wiki space. (you can either attach documents, or create wiki pages, or both; the specific decisions are left to you). Note, that you can submit all deliverables as a single PDF file to the wiki (as long as you maintain previous submissions intact).

The following deliverables shall be submitted in hardcopy on the due date:

1. Your **original** Lab 2 submission — **the one with all the comments**. I will be using my comments to your Lab 2 submission to judge your Lab 3 submission.
2. Hardcopy of the redesign document.
3. Hardcopy of the change log.
4. Hardcopy of the list of constraints.
5. Hardcopy of the logical database design document, which includes DDL commands.

6. Electronic copy of the DDL commands. Name the file `DB-setup.sql`. This file must be available on the wiki prior to the submission deadline. (Note: I will run your DDL commands, they must work, the (empty) database must be created!)

7. Electronic copy of the DDL commands deleting all the tables in the database. Name the file `DB-cleanup.sql`. (The script must properly drop all tables. It will be tested in conjunction with `DB-setup.sql`.)