BF768 Homework 0

Problem 1 is due on Thursday, 1/27/22 and will be discussed in class. Do not turn in. Problems 2 and 3 are due on Tuesday 2/1/22 by 3:30 pm, and should be turned in through blackboard.

General Policy on Homework Collaboration:
Except as otherwise noted, all problem sets/homeworks are to represent individual effort, and are to be written up and turned in individually. This does not preclude talking about a problem set with other class members; in fact, working together is encouraged, since it is one of the skills of modern science. However you are not allowed to copy each other's answers. If you work on a problem set with other people, please state that (include their names) on your write-up.

1. Design and draw an ER diagram that captures the information below.

- Patients are recorded by an id, social security number (SSN), name, address, and date of birth (DOB).
- Doctors are identified by an id, name, specialty, and start date of service.
- Each pharmaceutical company is identified by a name.
- For each drug, the trade name and formula are recorded.
- Each drug is manufactured by a pharmaceutical company and the trade name identifies the drug uniquely from among the products of that company.
- Every pharmacy has a name, address, and phone number.
- Every patient has a primary physician. Every doctor has at least one patient.
- Each pharmacy sells several drugs and has a price for each. A drug could be sold at several pharmacies, and the price could vary from pharmacy to pharmacy.
- Doctors prescribe drugs for patients. Usually, a doctor writes prescriptions for more than one patient and a patient could receive a prescription from more than one doctor. Each prescription has a date and a quantity.
- Pharmaceutical companies have long-term contracts with pharmacies. A pharmaceutical company will contract with more than one pharmacy, and a pharmacy will contract with more than one pharmaceutical company. For each contract, there is a start date, end date, and the text of the contract.

2. Design and draw an ER diagram that captures the information below.

- A gene is described by a chromosome number, an organism, a name, a start position in the chromosome, and an end position.
- A gene is composed of one or more exons and zero or more introns.
- An exon has a start position, a stop position, and a number in the gene sequence (1 means first exon, etc.).
- An intron has a start position, a stop position, and a number in the gene sequence.
- Each intron and each exon belongs to exactly one gene.
- Each gene has zero or more associated transcription promoter elements.
- A promoter element has a name and a type (examples: TATA box, transcription factor binding site).
- Each promoter stored will be associated with at least one gene.
- Experiments are run to test gene activity.
- Each experiment involves one or more genes and is run by one student using one of several protocols. The lab, date, and gene activity are recorded with each experiment.
- Each student has an id number and a name.
- Each protocol has a set of experimental conditions (this could be a text file listing the conditions). Not all protocols are necessarily used and some will be used repeatedly.
- Students can run more than one experiment. Some students will not run any experiments.
- No two experiments are the same on the same day.

3. Design and draw an ER diagram that captures the information below.

- A protein is described by a name, a sequence, and a structure (this could be a file storing PDB structure data).
- Each protein has one or more functions.
- A function has a name, a class, and a subclass (for example, a class could be DNA binding protein, a subclass could be zinc finger, which is a particular protein structure that binds DNA).
- A protein's function is either predicted or experimentally confirmed.
- If confirmed, there is at least one journal reference.
- If predicted, there is a gene annotation program that was used for the prediction.
- Each gene annotation program has a unique journal references and no single reference discusses more than one gene annotation program.
- Some proteins are known to function in one or more cellular pathways. The protein's function can be different in different pathways and a protein can have more than one function in the same pathway.