Database Systems, CSCI 4380-01
Homework # 3
Due Wednesday February 7, 2018 at 11:00:00 PM

Homework Statement. This homework is worth 3% of your total grade. If you choose to skip it, Midterm #1 will be worth 3% more. Remember, practice is extremely important to do well in this class. I recommend that not only you solve this homework, but also work on homeworks from past semesters. Link to those is provided in the Piazza resources page.

This homework targets decompositions and E-R diagrams.

SUBMISSION INSTRUCTIONS. Submit a PDF document for this homework using Gradescope. No other format and no hand written homeworks please. No late submissions will be allowed. Please put answer for each question on a separate page for easiest grading!

Question 1 (20 points). You are given the following:

$\mathcal{F} = \{AB \to CD, AE \to G, GD \to H, HB \to ED, D \to B\}$ for relation $R(A, B, C, D, E, F, G, H)$

Use 3NF decomposition algorithm to find a set of relations that satisfy the third normal form.

For each resulting relation, list all functional dependencies projected over the relation. Also show whether the relation is in BCNF or not.

Question 2 (30 points). You are given the following:

$\mathcal{F} = \{AB \to C, BD \to E, E \to F, AF \to G\}$ for relation $R(A, B, C, D, E, F, G)$

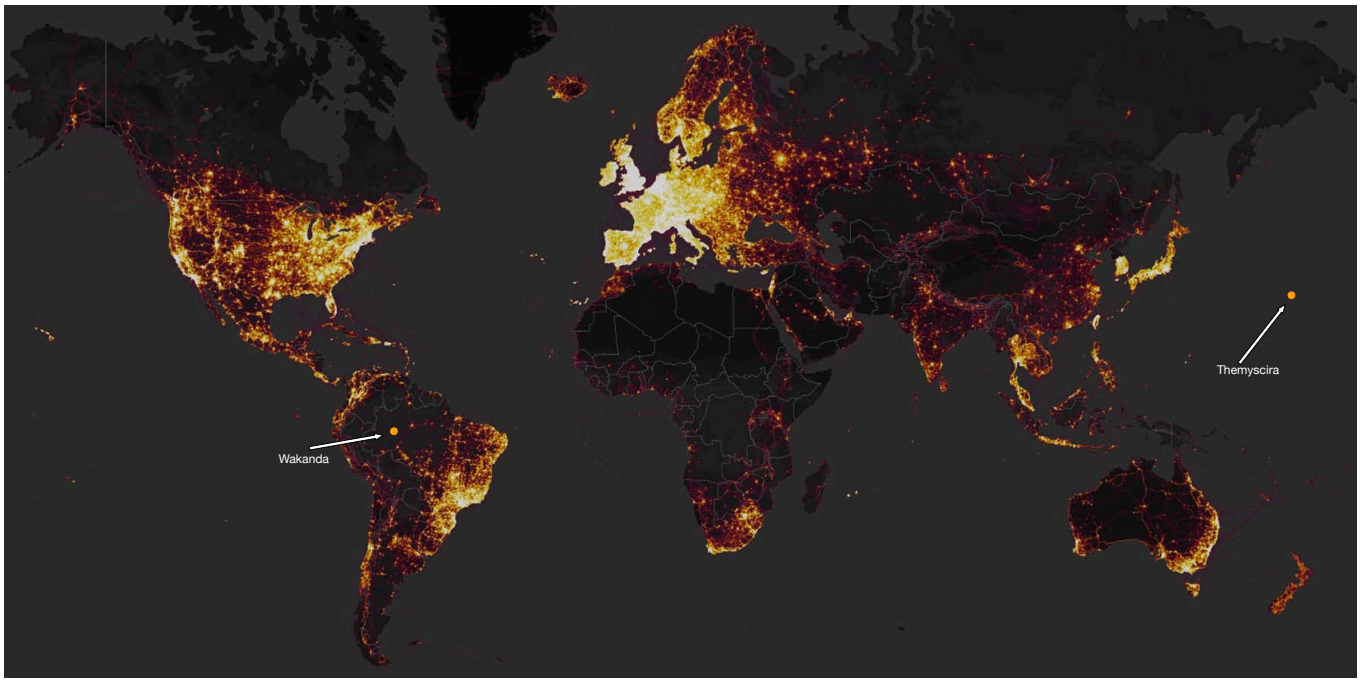
Use BCNF decomposition to find a set of relations that satisfy the Boyce-Codd normal form.

For each resulting relation, list all functional dependencies projected over the relation. Also show whether the resulting decomposition is functional dependency preserving or not.

To show that a decomposition is functional dependency preserving, you must show that the original set of functional dependencies is equivalent to the union of all projected functional dependences.

To show the decomposition is not functional dependency preserving, show a single functional dependency in the original set that is not implied by the union of all projected functional dependencies.

Question 3 (50 points). In this question, you are asked to create an E-R diagram that represents all the requirements below precisely. Note that while there are multiple potential correct models, not all answers are right. Be extra careful about choosing entities and correctly implementing the participation requirements.
A note on drawing ER diagrams. You will find that the diagrams can get very large if entities have a lot of attributes. To avoid this, you can simply list the attributes for the entities inside the box for that entity. Remember to underline the key attributes.



In this database, we will store data similar to what is stored in Strava. What is Strava you say? Well, it records performance activities for runners and cyclists. It also supports people to friend, monitor and compete with other users. There was a recent news article on how Strava put a heatmap of where people are running the most, which apparently revealed some secret sites. Obviously, our main worry that it will also reveal the locations of Wakanda and Themyscira, putting them in peril. So, let's see if this is possible.

In this database, there are users who store their username, email, password, address. Furthermore, the database holds information about the different devices that they use, make, year, OS while exercising (phone, fitbit, watch, etc.). A person can have many such personal devices.

Users may also friend other users. Friendship is mutual, i.e. both sides agree. Friends are first requested by one side (only requests at this point) and if approved by the other side, it becomes stored friendships. Both requests and friendships must be stored.

The database holds a (very large) number of road segments, each segment has a starting and end point, both are latitude and longitude values. Plus, each segment may have an associated geographic landmark. Each landmark is given by a name, type, state, city, zip and country attached. Landmarks may be associated with multiple segments.

Each user stores multiple events. Each event has an id, type (running or cycling), day, start time and end time, recording device and starting segment. The event also has a sequence of data points. Each data point has an associated segment, associated event, the exact times the start and end points of the segment were visited, and the sequence number in the event (an integer, for example 3 means this was the third data point in the segment). Note that the sequence number is needed because people can be running in a circle, hitting the same segment many times.

The events also have comments, made by the friends of the user. Each comment has the author, associated event, date, time and text.

We stand ready to analyze the database to see if it reveals these important secret locations!