Database Systems, CSCI 4380-01
Homework # 2
Due Thursday February 18, 2016 at 11:59:59 PM

Homework Statement. This homework is on data modeling, normalization and ER diagrams.

Question 1 (10 points each). For the following relations: first find the keys, then discuss whether it is in 3NF and/or BCNF with a short explanation of why or why not.

(a) $R(A, B, C, D, E, F)$, $\mathcal{F} = \{ABC \to DEF, CE \to A\}$

(b) $R(A, B, C, D, E, F)$, $\mathcal{F} = \{AB \to CDE, ABE \to F\}$

(c) $R(A, B, C, D, E, F)$, $\mathcal{F} = \{AB \to DE, CD \to A\}$

(d) $R(A, B, C, D, E, F)$, $\mathcal{F} = \{BCD \to AEF, AB \to C\}$

Question 2 (10 points). You are given the following set $\mathcal{F}$ of functional dependencies. Convert this to a minimal basis. Show your work at each step by providing sufficient detail. You do not have to show all unsuccessful steps (what you tried to remove and decided cannot be removed), but you must show what needs to be removed and why.

$R(A, B, C, D, E, F, G)$ with $\mathcal{F} = \{BC \to ADE, ABD \to BEF, E \to A, ABC \to G\}$

Question 3 (10 points each). You are given the following relation:

$R(A, B, C, D, E, F)$, $\mathcal{F} = \{A \to C, CE \to D, CD \to A, DF \to A\}$

Find whether the following decompositions of this relation are lossless or not using the Chase algorithm:

(a) Decomposition 1: $R1(A, C, E, D)$, $R2(B, E, F)$, $R3(A, E, F)$

(b) Decomposition 2: $R1(A, B, C, D)$, $R2(A, D, E)$, $R3(B, C, E, F)$

Question 4 (20 points). You are given the following relation:

$R(A, B, C, D, E, F)$, $\mathcal{F} = \{A \to BC, C \to D, D \to E\}$

Convert it to BCNF using the BCNF decomposition, starting with the functional dependency: $C \to D$.

In each step, find the projection of the functional dependencies for the decomposed relations and check whether the resulting relations are in BCNF or not. Continue until all relations are in BCNF.

When finished, find whether the resulting decomposition is dependency preseving or not (take the union of all projected functional dependencies and check if it is equivalent to the original set above).

Question 5 (10 points). Convert the relation from Question 4 to 3NF using the 3NF decomposition. Show your work.

Question 6 (30 points). You are given the following relation. Create an ER diagram that represents all the requirements below precisely.

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.

Voter databases have become extremely important for election campaigns to find individuals who might vote for a candidate based on their past voting records and other issues they care about. Often campaign target specific neighborhooods and even individuals to contact, even customize their messages towards these individuals. Also, an amazing amount of personal information is collected and sold in these databases. A recently discovered security flaw exposed all sorts of extremely personal information about 191 million voters to the world. This shows you the scale of these databases and you can expect there are much larger ones out there.

Ok, you are hired by the campaign manager of BirdPerson (see on the left) to create one of these scary databases, with the hope that BirdPerson can win the next big elections.

Your database will store information about individual voters. First, you will store information for each person that is entered on voter registration as shown here:

http://www.elections.ny.gov/NYSBOE/download/voting/voteform.pdf

The information for a voter includes last and first name, middle initial, gender, birth date, phone, email, address for where the voter lives (street address, apt. number, city, state, zip) and address for where the voter receives mail, has the person voted before (Y/N) and which year. In addition, the voter is asked her DMV number or last 4 digits of SSN, or that the voter can indicate that she does not have either of these. Each voter can enroll in one party or none at all. You also store the occupation of the person, such as teacher or business owner, and the employer of the person.

An important part of each voter is their voting record. Basically, the database stores all the past elections that they voted in. Elections take place in a specific year and can be presidential, primaries, senate/congress or state level. Primaries are for a specific party. Other elections simply have a type. You do not know who the person voted for, but that they have voted.

Also, you will store political donations that they have made to a specific candidate for a specific election. Many donations can be made for a specific candidate and a person can donate to multiple candidates. For each donation, you need to store the dollar amount. You can see that donation data is public, for example at opensecrets.org.

You also want to store charitable donations made by the voter, to different organizations. You need to store the amount and date of each donation. Of course, it is possible to make multiple donations in a specific year.

Other information stored for voters include the names of magazines the person subscribes to and has subscribed to in the past.

You realize that you need to relate some of this information to each other for deeper analysis. So, you will store a number of issues that voters likely care for (immigration, education, gun control, tax cuts), for each issue you will store the name and description. Then, you will store whether a specific charitable organization stands for an issue and whether the standing is pro or con. You also want to know which candidates in a specific elections stood for a specific issue, again whether the candidate's position was pro or con on this issue.

Finally, to show off your extra skills (because you are from RPI after all), you will also store the various social media usernames for each voter, if they are known. You will mine as much information as you can from each social media outlet and store the following for each outlet: overall number of friends the voter has, number of messages per week, how frequently (percent of messages) the voter talks about the issues in the database and the date the analysis was performed.

BirdPerson thanks you for your contributions to his campaign!

SUBMISSION INSTRUCTIONS. Submit a single PDF for this homework using Homework Submission Server. No other format, especially no WORD documents or hardcopy please. The penalty for late submissions in effect, make yourself familiar with the course policies.

The course website for homework submissions will become available sometime next week. We will announce it on Piazza.