

CS 564 Problem Set #2

February 25, 2018

DELIVERABLES

Submit your answers using the `submission_template.txt` file that is posted on the class website. Follow the instructions on the file! Upload the file at Canvas (under PS2).

INSTRUCTIONS / NOTES

Read these carefully

- You **may** create new IPython notebook cells to use for e.g. testing, debugging, exploring, etc.- this is encouraged in fact!- **just make sure that your final answer for each question is in its own cell and clearly indicated**
- When you see In [*]: to the left of the cell you are executing, this means that the code / query is *running*.
 - **If the cell is hanging- i.e. running for too long: To restart the SQL connection, you must restart the entire python kernel**
 - To restart kernel using the menu bar: "Kernel >> Restart >> Clear all outputs & restart"), then re-execute the sql connection cell at top
 - You will also need to restart the connection if you want to load a different version of the database file
- Remember:
 - `%sql [SQL]` is for *single line* SQL queries
 - `%%sql [SQL]` is for *multi line* SQL queries
- *Have fun!*

1 Problem 1: Verifying Functional Dependencies [24 points]

For this part, you will need to provide a *single* SQL query which will check whether a certain condition holds on the **hospital** table in the provided database:

```
In [ ]: %sql select * from hospital limit 2;
```

You need to evaluate any requested conditions in the following way: **your query should return an empty result if and only if the condition holds on the instance**. If the condition *doesn't hold*, your query should return something non-empty, but it doesn't matter what this is.

Note our language here: the conditions that we specify cannot be proved to hold **in general** without knowing the externally-defined functional dependencies; so what we mean is, *check whether they **are not violated** for the provided instance.*

You may assume that there are no NULL values in the tables.

1.1 Part (a) [14 points]

Is $\{provider\}$ a **superkey** for relation *Hospital*?

1.2 Part (b) [10 points]

Does $\{Zip\} \rightarrow \{City, State\}$ hold for relation *Hospital*?

2 Problem 2: Superkeys & Decompositions [40 points]

Consider a relation $S(A, B, C, D, E, F)$ with the following functional dependencies:

- $\{A\} \rightarrow \{D\}$
- $\{A\} \rightarrow \{E\}$
- $\{D\} \rightarrow \{C\}$
- $\{D\} \rightarrow \{F\}$

In each part of this problem, we will examine different properties the provided schema.

To answer **yes**, provide python code that assigns the variable `answer` to `True` and assigns `explanation` to be a python string which contains a (short!) explanation of why. For example:

```
answer = True
explanation = "All keys are superkeys."
```

To answer **no**, provide python code that assigns the variable `answer` to `False` and assigns `explanation` to be a python string which contains a (short!) explanation of why. For example:

```
answer = False
explanation = "D is not a superkey because its closure is {D,C,F}."
```

2.1 Part (a) [8 points]

Is it correct that A, B is a superkey?

2.2 Part (b) [8 points]

Is it correct that the decomposition ABC, CDE, EFA is lossless-join?

2.3 Part (c) [8 points]

Is it correct that the decomposition ABC, CDE, EFA is dependency preserving?

2.4 Part (d) [8 points]

Is the functional dependency $\{A\} \rightarrow \{C, F\}$ logically implied by FDs present in the relation?

2.5 Part (e) [8 points]

Is it correct that relation S is in BCNF?

3 Problem 3: Relational Algebra [36 points]

Consider the following relational schema:

- `JournalArticle(articleID, title, journal, year, month)`
- `ConferenceArticle(articleID, title, conference, year, location)`
- `Person(name, affiliation)`
- `Author(name, articleID)`

Express the following queries in the extended Relational Algebra (you can also use the aggregation operator if necessary). To write the RA expression, use the LaTeX mode that ipython notebook provides. For example:

$$\pi_{name}(\sigma_{affiliation="UW-Madison"}(Person))$$

3.1 Part (a) [12 points]

Output the names of everyone who is affiliated with UW-Madison and who has submitted during 2016 a journal article but not a conference article:

3.2 Part (b) [12 points]

Output the names of the people who coauthored a conference *or* journal article with John Doe for. Be careful: a person cannot be coauthor with herself!

3.3 Part (c) [12 points]

Count how many journal *or* conference articles were published in 2016 by John Doe.

4 Bonus Problem [+10 points]

Suppose we are given a relation R that has exactly one key. Prove that R is in BCNF if and only if it is in 3NF.