Assignment 5: Dependency Theory

Jelle Hellings

3DB3: Databases – Fall 2021

Department of Computing and Software
McMaster University

Deadline: November 19, 2021

Cheating and plagiarism. This assignment is an *individual* assignment: do not submit work of others. All parts of your submission *must* be your own work and be based on your own ideas and conclusions. If you *submit* work, then you are certifying that you have completed the work for that assignment by yourself. By submitting work, you agree to automated and manual plagiarism checking of the submitted work.

Cheating and plagiarism are serious academic offenses. All cases of academic dishonesty will be handled in accordance with the Academic Integrity Policy via the Office of Academic Integrity.

Late submission policy. There is a late penalty of 20% on the score per day after the deadline. Submissions five days (or later) after the deadline are not accepted. Do not wait until the deadline to ask questions or raise problems.

Description

The Professor S. Marty Pants, a recent faculty hire of the University, is convinced that they are the smartest database person of all times. To impress people, the Professor often states problems that the Professor then claims are *almost impossible* to solve and then shows how to solve them. Next, we will take a look at a few of these problems from three categories:

D1 Reasoning with dependencies

1. Prove that $\{AB \longrightarrow C, A \longrightarrow D, CD \longrightarrow EF\} \models AB \longrightarrow F$ holds using only the Armstrong Axioms.
2. Prove the soundness of the following inference rule directly from the definition of functional dependencies (without using any inference rules):

$$\text{if } X \longrightarrow Y \text{ and } YW \longrightarrow Z\text{, then } XW \longrightarrow Z.$$

3. Prove the soundness of the following inference rule for inclusion dependencies:

$$\text{if } R[X] \subseteq S[Y] \text{ and } S[Y] \subseteq T[Z]\text{, then } R[X] \subseteq T[Z].$$

4. Prove the soundness of the following inference rule

$$\text{if } X \twoheadrightarrow Y \text{ and } XY \longrightarrow Z\text{, then } X \longrightarrow Z \setminus (X \cup Y).$$

5. Prove that the following inference rule *is not sound*:

$$\text{if } XW \longrightarrow Y \text{ and } XY \longrightarrow Z\text{, then } X \longrightarrow Z.$$

Hint: Look for a counterexample by constructing a table in which $XW \longrightarrow Y$ and $XY \longrightarrow Z$ hold, but $X \longrightarrow Z$ does not hold.

D2 A completeness proof for Armstrong

Consider the following inference rules.

R1. *Reflexivity.* If $Y \subseteq X$, then $X \longrightarrow Y$.

R2. *Augmentation.* If $X \longrightarrow Y$ then $XZ \longrightarrow YZ$ for any Z.

R3. *Transitivity.* If $X \longrightarrow Y$ and $Y \longrightarrow Z$, then $X \longrightarrow Z$.

6. Prove that the inference rules R1, R2, and R3 are *complete*: prove that if $\mathfrak{S} \models X \longrightarrow Y$ holds for some set of functional dependencies $\mathfrak{S}$, then we can derive $X \longrightarrow Y$ from $\mathfrak{S}$ using only the inference rules R1, R2, and R3.

   Hint: Use the fact that the Closure algorithm is complete. Can you prove that any *sound* derivation made by the Closure algorithm can also be derived using the inference rules R1–R3? You may use the Union rule and the Decomposition rule (as we have derived them using the inference rules R1–R3).

D3 The human computer

Consider the relational schema $\mathbf{r}(A, B, C, D, E)$ and the following set of functional dependencies:

$$\mathfrak{S} = \{AB \longrightarrow D, AC \longrightarrow E, BC \longrightarrow D, C \longrightarrow A, D \longrightarrow A, E \longrightarrow B\}.$$

Answer the following questions:

7. Provide the attribute closure of set of attributes C (hence, C^+) and of set of attributes EA (hence, $(EA)^+$) with respect to $\mathfrak{S}$. Explain your steps.

8. Compute the closure $\mathfrak{S}^+$. Explain your steps. Based on the closure, indicate which attributes are *superkeys* and which attributes are *keys*.

   Hint: A *superkey* is a set of attributes that determines *all* attributes from $\mathbf{r}$. A *key* k is a superkey of minimal size (if we remove any attribute from k, it is no longer a key).

9. Provide a minimal cover for $\mathfrak{S}$. Explain your steps.

Assignment

The goal of the assignment is to show that Professor S. Marty Pants is just an average genius by showing that many people can solve its "problems". To do so, you have to write a report in which you solve each of the above problems. Your submission:

1. must be a PDF file;

2. must have clearly labeled solutions to each of the stated problems;

3. must include a correct proof and all proof details for problems 1–6;

4. must include explanation of the steps taken for problems 7–9;

5. must be clearly presented;

6. must *not* be hand-written: prepare your document in Microsoft Word or another word processor (printed or exported to PDF) or in LaTeX.

Submissions that do not follow the above requirements will get a grade of zero.

Grading

The presented solutions for the problems in Section D1 will account for 50% of the maximum grade (equally divided over all problems in the section); the presented solution for Section D2 will account for 20% of the maximum grade; and the presented solutions for the problems in Section D3 will account for 30% of the maximum grade (equally divided over all problems in the section).