Information and Database Management Systems I

(CIS 4301 UF Online)

Fall 2024

Instructor: Alexander Webber

TA: Kyuseo Park

Homework 4

Printed Name:	
UFID:	
Email Address:	

Instructions: Please provide your answers to the questions of the following pages in Word or handwritten on separate sheets of paper. Mark clearly to which question each answer belongs. Then convert or scan your work into PDF (the latter by using either a scanner or a suitable scanner app on your smartphone). Note that *only the PDF format* is allowed and that your submission must be a *single PDF file*. Finally, upload your PDF file into *Gradescope* and follow the instructions there. In order to enable the graders to fast find the solutions to your questions, it is important that you correctly specify the location of your answer for each question in Gradescope, as it is described there. Otherwise, 0.25 points will be deducted for each answer.

Note: All homework assignments are designed for a period of two, three, or even four weeks (see course deadline sheet). This means they cannot be solved in two or three hours but require a considerable amount of time and effort. Therefore, the first recommendation is to start with them as soon as they are posted. The second recommendation is to distribute the work on a homework assignment over the entire available period. The third recommendation is to submit the homework solutions *on time before the deadline*.

Pledge (Must be signed[1] according to the UF Honor Code):

On my honor, I have neither given nor received unauthorized aid in doing this assignment.

____

Student signature

[1] Each student is obliged to print out this page, fill in the requested information in a handwritten and readable manner, make the *handwritten* signature, scan this page into PDF, and put this page as the first page of the PDF submission.

Question 1 (Functional Dependencies – Armstrong's Axioms) [19 points]

(a) [9 points] Let $R(A,B,C,D,E)$ be a relation schema, and let $F = \{A \to BC, B \to D, D \to E, C \to A\}$ be a set of functional dependencies (FDs). Show for each of the following FDs whether they can be logically implied from F by using Armstrong's axioms and their derived inference rules. Show each step.

(1) [3 points] $BC \to AE$
(2) [3 points] $A \to ED$
(3) [3 points] $BD \to CE$

(b) [5 points] Let $R(A,B,C,D,E,F)$ be a relation schema, and let $F = \{A \to BC, CD \to E, E \to F, B \to D\}$ be a set of functional dependencies (FDs). Which of the following attribute sets is a candidate key? Explain your answer by using Armstrong's axioms and their derived inference rules.

- AC
- AD
- A

(c) [5 points] Let $R(A,B,C,D,E,F,G)$ be a relation schema, and let $F = \{A \to B, AB \to C, CD \to E, E \to FG\}$ be a set of functional dependencies (FDs). Infer at least five new FDs by using Armstrong's axioms and their derived inference rules. Each applied rule in your answer should be different.

Question 2 (Functional Dependencies – Equivalence and Closure) [29 points]

(a) [14 points] Consider the relation schema $R(W,X,Y,Z)$ with the functional dependencies $F = \{X \rightarrow Y, YZ \rightarrow W, W \rightarrow Z\}$ and $G = \{X \rightarrow YZ, Y \rightarrow W, W \rightarrow Z, Z \rightarrow Y\}$.

(1) [7 points] Explain why the two sets F and G are not equivalent by calculating the attribute closures with respect to each set. Explain your answer in detail.

(2) [7 points] Adjust or add necessary FDs to F or G so that they become equivalent. Explain your answer in detail.

(b) [9 points] Consider the relation schema $R(V,W,X,Y,Z)$ with the functional dependencies $F = \{W \rightarrow V, X \rightarrow Y, XW \rightarrow Z, YZ \rightarrow XW\}$.

(1) [5 points] Which of the following sets of attributes functionally determine VWX? Show each step.

- YZ
- XY
- W

(2) [5 points] Which of the three attribute sets of (1) is a candidate key? Give reasons.

(c) [6 points] Consider the relation schema $R(A,B,C,D,E,F)$ with the functional dependencies $F = \{AB \rightarrow C, CD \rightarrow E, E \rightarrow F, B \rightarrow D\}$. By using the algorithm for calculating attribute closures provided in the lecture slides, calculate the closure of the attribute set CD. Show each step.

Question 3 (Functional Dependencies – Minimal Cover) [12 points]

Find a minimal cover for the relation schema $R(A,B,C,D,E,F,G)$ with the set $F = \{AB \rightarrow C, C \rightarrow A, BC \rightarrow D, ACD \rightarrow B, D \rightarrow EG, BE \rightarrow C, CG \rightarrow BD, CE \rightarrow G\}$ of functional dependencies. Show each step.

Question 4 (Functional Dependencies − Candidate Keys) [40 points]

This question aims at getting a deeper insight into the algorithmic aspects of the determination of all candidate keys of a given relation schema R and a set F of functional dependencies that hold true on R. In our course, we learned the "7-step method" for this purpose.

(a) [3 point] Look at the slides that introduce the 7-step method. For reasons of clarity and uniformity, in steps 1 to 6 of the method, the attributes of an attribute set are ordered lexicographically (that is, like in a dictionary). An attribute $X \in R$ can be regarded as a pair (A,D) where A is the *name*[1] of X and D is the *data type* of X. We assume that an attribute name (such as "salary" or "A") is of type *string*. We further assume two functions *name* and *dom* that return the name and the type of an attribute respectively. That is, given an attribute $X = (A,D) \in R$, we obtain $name(X) = A$ and $dom(X) = D$. The latter function was already introduced earlier as part of the discussion of relational schemas. What remains at this point is to define a lexicographical order $<_A$ as well as the equality predicate $=_A$ on the attributes of R. For this purpose, replace the question marks in the two lines below by corresponding formal definitions. Explain your definitions.

(1) $\forall X,Y \in R : X <_A Y \Leftrightarrow ?$

(2) $\forall X,Y \in R : X =_A Y \Leftrightarrow ?$

(b) [6 points] The next step, as a continuation of (a), is to "normalize" all possible attribute sets of a relation schema $R(A_1,A_2,\ldots,A_n)$ in the sense that the attributes of an attribute set are lexicographically ordered. [Example: If H and B are attributes of R, we only allow to write BH but not HB, although they are semantically equivalent since $\{H,B\} = \{B,H\}$.] Formally complete the line below to enforce this normalization.

$\forall X = \{B_1,B_2,\ldots B_m\} \subseteq R : ?$

(c) [7 points] Based on our considerations in (b), we implicitly assume that all attribute subsets of R are normalized. Step 7 of the 7-step method orders attribute sets in a certain order in a list. For this purpose, complete the line below to formally define the order relation $<_{AS}$ on two attribute subsets $X,Y \subseteq R$. Explain your definition. [Note the $\subseteq$ symbol instead of the $\in$ symbol in (a).]

$\forall X = \{B_1,\ldots,B_k\}, Y = \{C_1,\ldots,C_l\} \subseteq R, 1 \leq k,l \leq n : X <_{AS} Y \Leftrightarrow ?$

(d) [11 points] The 7-step method determines all candidate keys of a relation schema $R(A_1,A_2,\ldots,A_n)$ and a set F of functional dependencies that hold true on R. This method classifies the attributes of R with respect to their appearance or location in the functional dependencies of F. The task here is to design a simpler method for the candidate key determination problem that, compared to the 7-step method, does not make use of an attribute classification, is straightforward ("brute-force"), is not as efficient, shares some similar features, and mainly rests on the defining properties of uniqueness and minimality of candidate keys. Write down your method in pseudo-code[2], and make use of the order relation $<_{AS}$ from (c).

---

[1] Usually, we use the name A to denote X. This means we equate the name A with the attribute object X.

[2] In Wikipedia, the term *pseudo-code* is characterized as follows: "In computer science, pseudocode is a description of the steps in an algorithm using a mix of conventions of programming languages (like assignment operator, conditional operator, loop) with informal, usually self-explanatory, notation of actions and conditions. Although pseudocode shares features with regular programming languages, it is intended for human reading rather than machine control. Pseudocode typically omits details that are essential for machine implementation of the algorithm. The programming language is augmented with natural language description details, where convenient, or with compact mathematical notation. The purpose of using pseudocode is that it is easier for people to understand than conventional programming language code, and that it is an efficient and environment-independent description of the key principles of an algorithm. It is commonly used in textbooks and scientific publications to document algorithms and in planning of software and other algorithms.".

(e) [13 points] Based on the insight gotten from (d) and the lecture slides, the task is now to formulate an algorithm for the 7-step method in pseudo-code. We assume again a relation schema $R(A_1, A_2, \ldots, A_n)$ and a set F of functional dependencies on R. You can reuse parts of the pseudo-code from (d) if appropriate. Further, we predefine the functions *isolatedAttr*, *leftSideOnlyAttr*, *rightSideOnlyAttr*, and *bothSidesAttr* that, applied to a relation schema R and a set F of FDs, yield the set of isolated attributes (that is, those attributes that do not appear in any FD of F), the set of attributes that only appear on left-hand sides of FDs in F, the set of attributes that only appear on right-hand sides of FDs in F, and the set of attributes that appear on both sides of FDs in F.

An algorithmic step that should be performed at the beginning of the 7-step method and that was not mentioned in the lecture is the "cleaning" of F by computing a minimal cover F_c of F. F_c should then be taken as a basis for the 7-step method. Include this step as the first step in your pseudo-code.