

Assignment 2

Due: Wednesday, March 1 at 11:59pm

General Instructions

- Feel free to talk to other members of the class in doing the homework. You should, however, write down your solutions yourself. *List the names of everyone you worked with at the top of your submission.*
- Keep your solutions brief and clear.
- Please use Piazza if you have questions about the homework but do not post answers. Feel free to use private posts or come to the office hours.

Homework Submission

- We DO NOT accept late homework submissions.
- We will be using Compass for collecting the homework assignments. Please submit your answers via [Compass](#). Hard copies are not accepted.
- Contact the TAs if you are having technical difficulties in submitting the assignment; attempt to submit well in advance of the due date/time.
- The homework must be submitted in **pdf** format. Scanned handwritten and/or hand-drawn pictures in your documents won't be accepted.
- Please do not zip the answer document (PDF) so that the graders can read it directly on Compass. You need to submit one answer document, named as **hw2_netid.pdf**.
- Please see the [assignments](#) page for more details. In particular, we will be announcing errata, if any, on this page.

Question 1. Short Questions (15 points)

1. Consider two relations given below. State if it is possible to perform the set union of these two relations i.e. $R1 \cup R2$. If union is possible provide the resulting schema after union, if not, state your reasoning.

R1: (FirstName(char), LastName(char), DatefBirth(date))

R2: (FName(char), LName(char), PhoneNumber(number))

2. Use relational algebra to show 5 different ways to express: $\sigma_{c>20} (R \bowtie S)$. You can use basic and derived relational algebra operations covered in class.
3. The following formulae are valid under set semantics. For each formula, mention if it is valid under bag semantics as well. If not, provide a counter example that demonstrates that it is not valid under bag semantics.

(a) $R \cup S = S \cup R$

(b) $(R \cup S) - T = (R - T) \cup (S - T)$

(c) $(R \cap S) - T = R \cap (S - T)$

4. Consider a relation $R(A,B,C,D,E)$ with dependencies $ABC \rightarrow DE$, $E \rightarrow BCD$. Then the relation R is in 3NF. State true or false with reasoning.

Question 2. Relational algebra to English (20 points)

Consider a relation $\text{Speed}(\text{pitcher}, \text{fast}, \text{slow})$ recording the fastest (maximum) and slowest (minimum) pitch speeds for baseball pitchers. A key for the relation is (pitcher). Consider the following relational algebra expression, written in linear notation.

$P1(\text{pitcher}, f) := \pi_{\text{pitcher}, \text{fast}}(\text{Speed})$
 $P2(\text{pitcher}, s) := \pi_{\text{pitcher}, \text{slow}}(\text{Speed})$
 $P3(\text{pitcher}) := \pi_{P1.\text{pitcher}}(P1 \bowtie_{f < \text{fast}} \text{Speed})$
 $P4(\text{pitcher}) := \pi_{P2.\text{pitcher}}(P2 \bowtie_{s > \text{slow}} \text{Speed})$
 $P5(\text{pitcher}) := \pi_{\text{pitcher}}(\text{Speed}) - P3$
 $P6(\text{pitcher}) := \pi_{\text{pitcher}}(\text{Speed}) - P4$
 $\text{Answer}(p) := P5 \cup P6$

State in English what is computed as the final Answer briefly. Long-winded answers will be deducted points.

Question 3. English to Relational algebra (20 points)

Consider a relation $\text{Taking}(\text{studentId}, \text{courseId})$ with no duplicates. Write a relational algebra expression to find all courses with at least five students. Your solution should be a single expression (i.e., not a tree or linear notation), and it should be as compact as possible while giving the correct answer on all databases.

Question 4. Data to functional dependency (20 points)

Which of the following rules for functional dependencies are correct (i.e., the rule holds over all databases) and which are incorrect (i.e., the rule does not hold over some database)? For incorrect rules, give the simplest example relation instance you can come up with where the rule does not hold. For correct rules, prove them using Armstrong axioms.

1. If $A \rightarrow B$ and $BC \rightarrow D$, then $AC \rightarrow D$
2. If $AB \rightarrow C$ then $A \rightarrow C$
3. If $A \rightarrow B_1, \dots, B_n$ and $C_1, \dots, C_m \rightarrow D$ and C_1, \dots, C_m is a subset of B_1, \dots, B_n , then $A \rightarrow D$
4. If $A \rightarrow C$ and $B \rightarrow C$ and $ABC \rightarrow D$, then $A \rightarrow D$

Question 5. Normalization (25 points)

Consider the following relational schema:

UnivInfo(studID, studName, course, profID, profOffice)

Each tuple in relation UnivInfo encodes the fact that the student with the given ID and name took the given course from the professor with the given ID and office. Assume that students have unique IDs but not necessarily unique names, and professors have unique IDs but not necessarily unique offices. Each student has one name; each professor has one office.

- (a) Specify a set of completely nontrivial functional dependencies for relation UnivInfo that encodes the assumptions described above and no additional assumptions.
- (b) Based on your functional dependencies in part (a), specify all minimal keys for relation UnivInfo.
- (c) Is UnivInfo in Boyce-Codd Normal Form (BCNF) according to your answers to (a) and (b)? If not, give a decomposition of UnivInfo into BCNF.
- (d) Now add the following two assumptions: (1) No student takes two different courses from the same professor; (2) No course is taught by more than one professor (but a professor may teach more than one course). Specify additional functional dependencies to take these new assumptions into account.
- (e) Based on your functional dependencies for parts (a) and (d) together, specify all minimal keys for relation UnivInfo.
- (f) Is UnivInfo in BCNF according to your answers to (d) and (e)? If not, give a decomposition of UnivInfo into BCNF.