CS145 Midterm Examination

- Please read all instructions (including these) carefully.
- The exam is open book and open notes; any written materials may be used.
- There are four parts on the exam, with a varying number of points for a total of 75 points. There are also two extra-credit questions with a total of 10 points. You should look through the entire exam before getting started, in order to plan your strategy. You have 75 minutes to complete the exam.
- There is no penalty for guessing multiple-choice questions. For short-answer questions, *simplicity and clarity of solutions will count.* You may get as few as 0 point for a problem if your solution is far more complicated than necessary.
- Unless otherwise specified, assume that all relations are duplicate-free in questions not related to SQL. All questions about SQL refer to the SQL2 or SQL3 standard, *not* necessarily to the Oracle implementation of SQL used in programming assignments.

NAME: ____________________

In accordance with both the letter and spirit of the Honor Code, I have neither given nor received assistance on this examination.

SIGNATURE: ____________________

Part	Max points	Points
I	20	
II	24 (+4)	
III	16	
IV	15 (+6)	
TOTAL	75 (+10)	

Part I. E/R, ODL, and Basic Relational Design (20 points)

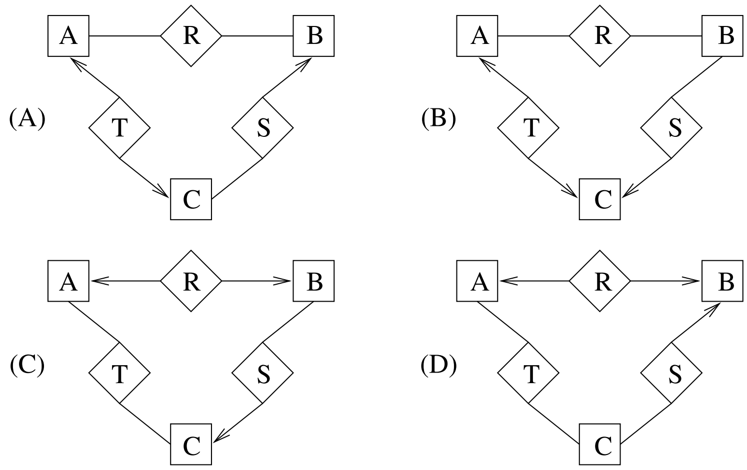
Question 1: (4 points) Below is an ODL specification:

```
interface A {
  relationship Set<B> R
    inverse B::R;
  relationship C T
    inverse C::T;
};
```

```
interface B {
  relationship Set<A> R
    inverse A::R;
  relationship C S
    inverse C::S;
};
```

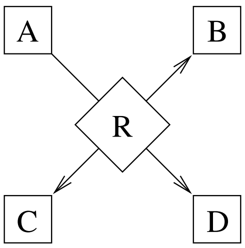
```
interface C {
  relationship Set<B> S
    inverse B::S;
  relationship A T
    inverse A::T;
};
```

For simplicity, all attributes are omitted. Which one of the E/R diagrams below best captures the intent of the ODL specification above?



Answer: $\square$

Question 2: (4 points) Consider the following E/R diagram.

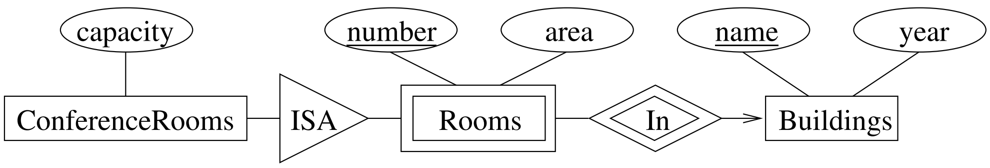


Suppose the key of entity set A is attribute A, the key of B is B, the key of C is C, and the key of D is D. If we translate relationship set R into a relation $R(A, B, C, D)$, what are all the keys of R?

(A) $\{A\}$
(B) $\{B\}$, $\{C\}$, and $\{D\}$
(C) $\{B, C, D\}$
(D) $\{A, B, C\}$, $\{A, B, D\}$, and $\{A, C, D\}$

Answer: $\square$

Questions 3–5 below refer to the following E/R design for a database that keeps track of buildings, rooms, and in particular, conference rooms.



Question 3: (4 points) Suppose that we convert the above E/R diagram into relations using the E/R-style translation for subclasses. What will we get for the ConferenceRooms entity set?

(A) `ConferenceRoom(capacity)`
(B) `ConferenceRoom(`<u>`room_number`</u>`, capacity)`
(C) `ConferenceRoom(`<u>`building_name`</u>`, `<u>`room_number`</u>`, capacity)`
(D) `ConferenceRoom(`<u>`building_name`</u>`, `<u>`room_number`</u>`, area, capacity)`

Answer: ☐

Question 4: (4 points) Suppose that we have specified an ODL schema that captures the design of the above E/R diagram. We then convert the ODL schema into three relations $\texttt{Building}_{\texttt{ODL}}$, $\texttt{Room}_{\texttt{ODL}}$, and $\texttt{ConferenceRoom}_{\texttt{ODL}}$, using the ODL-style translation for subclasses. Suppose there are m rooms in the database, and among these, n are conference rooms. How many tuples are in relations $\texttt{Room}_{\texttt{ODL}}$ and $\texttt{ConferenceRoom}_{\texttt{ODL}}$, respectively?

(A) m and n (B) m and $n - m$ (C) $m - n$ and n (D) 0 and m

Answer: ☐

Question 5: (4 points) Which of the following statements are true according to the constraints encoded by the E/R diagram above? Do not make any assumptions other than those encoded by the E/R diagram.

I. The number of entities in the Rooms entity set must be greater than or equal to the number of entities in the ConferenceRooms entity set
II. The number of entities in the Rooms entity set must be greater than or equal to the number of entities in the Buildings entity set

(A) I only (B) II only (C) Both I and II (D) Neither I nor II

Answer: ☐

Part II. FD's and BCNF (24 (+4) points)

Question 6: (4 points) In the instance of the relation $R(A, B, C, D, E)$ shown below, which of the following functional dependencies (FD's) hold?

A	B	C	D	E
1	2	3	4	5
1	4	3	4	5
1	2	4	4	1

I. $AB \to C$ II. $B \to D$ III. $DE \to A$

(A) I only   (B) II only   (C) I and III only   (D) II and III only

Answer: $\boxed{}$

Questions 7–9 below refer to a relation $S(A, B, C, D, E)$ with the FD's:

$$AB \to C, \quad B \to D, \quad DE \to A$$

Question 7: (4 points) What are all the keys of S?

(A) $\{A, B, D, E\}$
(B) $\{A, B\}$ and $\{D, E\}$
(C) $\{B, E\}$
(D) $\{A, B, E\}$ and $\{B, D, E\}$

Answer: $\boxed{}$

Question 8: (4 points) $B \to D$ is a BCNF violation for S. Suppose we decide to decompose S into $S_1(B, D)$ and $S_2(A, B, C, E)$. Which of the following statements are true?

I. $\{AB \to C\}$ is a minimal basis for the FD's that hold in S_2
II. $AB \to C$ is a BCNF violation for S_2
III. S_2 should be decomposed further into $S_3(A, B, C)$ and $S_4(C, E)$

(A) I only   (B) II only   (C) I and II only   (D) II and III only   (E) I, II, and III

Answer: $\boxed{}$

Question 9: (4 points) Which of the following statements are true?

I. Instead of decomposing S using $B \to D$, we could decompose S using $DE \to A$ first
II. It does not matter whether we start with $B \to D$ first or $DE \to A$ first: at the end of the BNCF decomposition algorithm we will get the same set of relations

(A) I only   (B) II only   (C) Both I and II   (D) Neither I nor II

Answer: $\boxed{}$

Questions 10–12 below refer to a relation $T(A, B, C, D, E)$ with the following FD's:

$$A \rightarrow BC, \quad CD \rightarrow E, \quad \clubsuit \rightarrow D$$

Unfortunately we don't know what $\clubsuit$ is—it could be any nonempty subset of T's attributes. (In particular, $\clubsuit$ might even contain D itself, which would make $\clubsuit \rightarrow D$ a trivial dependency.)

Question 10: (4 points) Which of the following must be true regardless of what is inside $\clubsuit$?

(A) Every key of T contains A
(B) No key of T contains A
(C) Some key of T contains A while some other key does not
(D) None of the above

Answer: $\boxed{}$

Question 11: (4 points) Which of the following must be true regardless of what is inside $\clubsuit$?

(A) Every key of T contains C
(B) No key of T contains C
(C) Some key of T contains C while some other key does not
(D) None of the above

Answer: $\boxed{}$

Question 12: (4 points)[**extra credit**] Which of the following must be true regardless of what is inside $\clubsuit$?

I. If some key of T contains E, then some other key must contain D
II. If some key of T contains D, then some other key must contain E

(A) I only (B) II only (C) Both I and II (D) Neither I nor II

Answer: $\boxed{}$

Part III. Relational Algebra and SQL: Multiple-Choice Questions (16 points)

Question 13: (4 points) Suppose that two relations $R(A, B)$ and $S(A, B)$ have exactly the same schema. Which of the following equalities hold in relational algebra?

I. $R \cap S = R - (R - S)$
II. $R \cap S = S - (S - R)$
III. $R \cap S = R \bowtie S$

(A) I only (B) I and II only (C) I, II, and III (D) None of the above

Answer: $\boxed{}$

Question 14: (4 points) Suppose we have two relations $R(\underline{A}, B)$ and $S(\underline{A}, B)$ with the same schema. The only key of R is $\{A\}$; the only key of S is $\{A\}$ as well. Let relation $T(A, B)$ be the set union of R and S, i.e., $T = R \cup S$. What are the keys of T?

(A) $\{A\}$ (B) $\{B\}$ (C) $\{A\}$ and $\{B\}$ (D) $\{A, B\}$

Answer: $\boxed{}$

Questions 15 and 16 below refer to the following database schema:

```
Person(SSN, employerSymbol, salary)
Holding(SSN, symbol, numShares)
```

A person is uniquely identified by a social security number (SSN). A company is uniquely identified by its stock ticker symbol. Each person is employed by exactly one company, but may hold any number of different stocks.

Question 15: (4 points) Suppose we wish to find the SSN's of the persons who do not own stocks of their employers. Which of the following queries will return the correct set of SSN's?

I. $\pi_{\texttt{SSN}}\big(\sigma_{\texttt{employerSymbol}\neq\texttt{symbol}}(\texttt{Person} \bowtie \texttt{Holding})\big)$

II. $\pi_{\texttt{SSN}}\big(\pi_{\texttt{SSN,sym}}(\rho_{\texttt{P(SSN,sym,sal)}}(\texttt{Person})) - \pi_{\texttt{SSN,sym}}(\rho_{\texttt{H(SSN,sym,num)}}(\texttt{Holding}))\big)$

III.
```
SELECT SSN
FROM   Person
WHERE  employerSymbol <> ALL
       (SELECT symbol FROM Holding WHERE Person.SSN = Holding.SSN);
```

(A) II only (B) I and II only (C) I and III only (D) II and III only

Answer:

Question 16: (4 points) Suppose we wish to find the average salary of the persons who own more than 100 shares of Microsoft (MSFT) or more than 100 shares of Yahoo! (YHOO). Which of the following queries will correctly compute the desired average?

I.
```
SELECT AVG(salary)
FROM   Person
WHERE  SSN IN (SELECT SSN FROM Holding
               WHERE  (symbol = 'MSFT' OR symbol = 'YHOO')
               AND    numShares > 100);
```

II.
```
SELECT AVG(salary)
FROM   Person, Holding
WHERE  Person.SSN = Holding.SSN
AND    ((symbol = 'MSFT' AND numShares > 100) OR
        (symbol = 'YHOO' AND numShares > 100));
```

(A) I only (B) II only (C) Both I and II (D) Neither I nor II

Answer:

Part IV. Relational Algebra and SQL: Short-Answer Questions (15 (+6) points)

Questions 17–20 below refer to the following database schema:

```
Person(SSN, employerSymbol, salary)
Holding(SSN, symbol, numShares)
StockPrice(symbol, date, price)
```

`Person` and `Holding` relations are identical to the ones used by Questions 15 and 16. We have added a third relation `StockPrice`, which tracks the closing price (per share) of each stock on each trading day.

Question 17: (5 points) Write a **relational algebra** query to find the SSN's of all Informix (IFMX) employees who own more than 50 shares of Oracle (ORCL) stock.

Question 18: (5 points) Write a **SQL** query to find the total number of shares of Oracle (ORCL) stock owned by Informix (IFMX) employees.

Question 19: (5 points) Write a **relational algebra** query to find the ticker symbols of all "superstocks". A superstock is a stock whose closing price always rises on every trading day. You may compare `date` values using $=$, $>$, etc. (*Hint:* you do not need arithmetics on `date` values.)

Question 20: (6 points)[**extra credit**] Let us define a "widely-held" stock to be one that is owned by more than 40% of the investors in our database. Write a **SQL** query to find the latest closing price for each widely-held stock. Note that some quotes may be delayed: for example, the latest closing price of Microsoft stored in our database might be one day old, while the latest closing price of Macrohard might be two days old.