WEEK 8

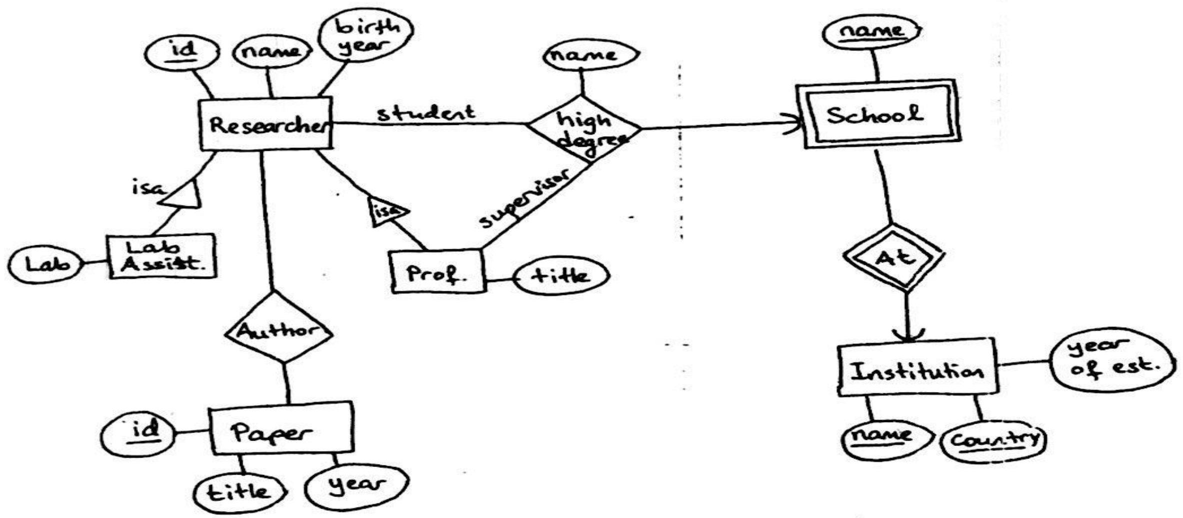
SAMPLE MIDTERM WITH SOLUTIONS

QUESTION 1 AND 2 MAIN TEXT

- Questions 1-2 of this exam is concerned with a database design that contains information on researchers, academic institutions, research papers and collaborations among researchers. The following (and any necessary info such as ID’s) should be stored in the database:
- * For each researcher, his/her name, year of birth. A researcher can be employed as either a professor or a lab assistant. For a professor we need his/her title: Assistant, associate, and full professor. For a lab assistant store the name of the lab he/she works at.
- * For each institution, its name, country, and year of establishment. For each institution, the names of its schools (*e.g., School of law, School of Business, School of Computer Science, etc.)*. A school belongs to exactly one institution.
- * For each research paper the title and the year of publishing. Information about co-authorship, i.e., which researchers have co-authored a research paper.
- * For each researcher, information on his/her highest degree (BSc, MSc or PhD), including who was the main supervisor, and at what school. (Assume that the highest degree is unique)

Question 1

- **(15 p)** Draw the E/R diagram for the above database system, indicate keys, many-one relationships, weak entity sets, and other features of E/R diagrams.



Question 2

- **(10 p)** Determine the relations (tables) together with their key attributes. Use NULLS approach for converting subclass structures to relations.
- **Answer:**
- Researcher (ID, name, birthyear, title, lab, degreeName, supervisorID, degreeSchool, degreeInstitution, degreeCountry) (6 p)
- Paper (ID, title, year) (1 p)
- Author (researcherID, paperID) (1 p)
- Institution (name, country, year-of-establishment) (1 p)
- School (name, institution-name, country) (1 p)

Question 3

- Given the relation schema R (A, B, C, D, E) and a set of FD's: AB → C ; DE →C ; B → D ;
- **a)** Check if R is in BCNF or not. **(5 pts.)**
- **Solution:**
- **Keys:** Any key contains ABE. ABE^+ = ABCDE. **ABE is the only key.**
- AB→C is a BCNF violation. Not BCNF.
-
- **b)** Decompose R into collections of relations, as necessary, that are in BCNF. **(7 pts.)**
- **Solution:**
- AB^+ = ABCD. Decompose as R1(A, B, C, D) and R2(A, B, E). R1 has AB → C and B → D. Key is AB.
- B → D is a BCNF violation. B^+ =BD. Decompose R1 as R11(BD) and R12(ABC) both of which are BCNF. Since R2 has no projected dependency, it is in BCNF.
-
- **c)** Check if the functional dependencies of the original table are preserved in BCNF decomposition of part b). (**5 pts.)**
- **Solution:** The projected dependencies are B → D and AB → C from which we cannot enforce DE →C. Therefore, the decomposition does not preserve the original dependencies.

Question 3 (cont'd)

- Given the relation schema R (A, B, C, D, E) and a set of FD's: AB → C ; DE →C ; B → D ;
- **d)** Check if R is 3NF. **(5 pts.)**
- **Solution:**
- C is not a prime attribute. Hence, AB→C is a 3NF violation. Therefore, R is also not 3NF.
-
- **e)** Decompose R into collections of relations, as necessary, that are in 3NF. **(8 pts.)**
- **Solution:**
- Minimal set of FDs for R are: AB → C ; DE →C ; B → D. Therefore, ABC, CDE and BD are decomposed tables. Since none of these contains a key, ABE should also be added to the list of decomposed tables.
-
- **f)** Check if the functional dependencies of the original table are preserved in 3NF decomposition. (**5 pts.)**
- **Solution:**
- The projected dependencies are B → D, AB → C and DE →C. All of the dependencies are preserved.

Question 4

- Given the relation schema R (A, B, C, D, E) and a set of FD's: AB → C ; C →B ; A → D. Suppose that we have decomposed R as R1 (ABC), R2 (AD) and R3 (ABE). By using chase test, verify that this decomposition has a lossless join.
- **Solution:**
- Start with $abcd_1e_1$, $ab_2c_2de_2$ and abc_3d_3e and try to obtain abcde by using the given dependencies. From AB → C, $abcd_1e_1$ and abc_3d_3e, we conclude that c_3=c leading to $abcd_3e$. From A → D, $ab_2c_2de_2$ and $abcd_3e$ we conclude that d_3 = d. Therefore we have shown that the decomposition has a lossless join.

Question 5

- Consider the following database schema for a BOOKSTORE database:
- Books (bookID, title, authorID, year)
- Authors(authorID, name, address)
- Customers (customerid, name, email)
- Purchases (customerid, bookID, year, price)
- The Books relation stores information about books sold by the bookstore. An example tuple is as follows: (105, 'JAVA PROGRAMMING', '22', 2001). The Authors relation stores information about the authors of the books sold by the bookstore. An example is as follows: (22, ‘JOHN DOE’, ‘HIDDEN VALLEY NO 13 CALIFORNIA’). The Customers relation stores information about the customers of the bookstore. An example tuple is as follows: (210, 'JOHN SMITH', 'SMITH@YAHOO.COM'). The Purchases relation stores information about the customer purchases of books. An example tuple is as follows: (210, 105, 2002, 49.95), indicating 'JOHN SMITH' with customerid 210 purchased the 'JAVA PROGRAMMING' book with bookID 105 in the year 2002 at 49.95$. Given the above schema**, write SQL statements** for the following:

Question 5 (cont'd)

- Consider the following database schema for a BOOKSTORE database:
- Books (bookID, title, authorID, year)
- Authors(authorID, name, address)
- Customers (customerid, name, email)
- Purchases (customerid, bookID, year, price)
- **a) (5 pts.)** Delete purchases either made before 2000 or had a price less than 10 $.
- **Solution:**
- DELETE FROM Purchases
- WHERE year < 2000 OR
- price < 10;
-
- **b) (7 pts.)** Give the names of those authors who have published books after the year 2000.
- **Solution:**
- SELECT name
- FROM Authors
- WHERE authorID IN
- (SELECT authorID
- FROM Books
- WHERE year > 2000);

Question 5 (cont'd)

- Consider the following database schema for a BOOKSTORE database:
- Books (bookID, title, authorID, year)
- Authors(authorID, name, address)
- Customers (customerid, name, email)
- Purchases (customerid, bookID, year, price)
- **c) (7 pts.)** Find the title(s) and price(s) of the most expensive book(s) purchased in year 2002.
- **Solution:**
- SELECT DISTINCT title, price
- FROM Purchases P, Books B
- WHERE price >= ALL(SELECT price
- FROM Purchases
- WHERE year=2002)
- AND P.bookID = B. bookID
- AND P.year = 2002;

Question 5 (cont’d)

- Consider the following database schema for a BOOKSTORE database:
- Books (bookID, title, authorID, year)
- Authors(authorID, name, address)
- Customers (customerid, name, email)
- Purchases (customerid, bookID, year, price)
- **d) (5 pts.)** Find the average prices of the book purchases through years.
- **Solution:**
- SELECT year, AVG(price)
- FROM Purchases
- GROUP BY year;
- **e) (6 pts.)** Find the number of customers without an e-mail addres (i.e., e-mails are NULLs).
-
- **Solution:**
- SELECT COUNT(*) - COUNT(email)
- FROM Customers;
-