Dr. Stefan Brass
School of Information Sciences
University of Pittsburgh

January 5, 2000

INFSCI 2710 "Database Management" — Example for Midterm Exam —

Instructions

- This is the midterm exam from the Spring 1999 edition of the course.
- There were 2 hours, 50 minutes time for solving the exercises.
- Note that the midterm exam in this course will include some new types of exercises. I will give some hints during the last lecture before the exam and on the web page. There may be some multiple choice exercises.
- Examples of new types of exercises are, e.g.:
  - Find errors in relational algebra queries.
  - Given pairs of relational algebra queries, which ones are equivalent?
  - Given pairs of ER-schemas, which ones are equivalent.
  - Given functional dependencies, determine a key.
- However, it will certainly be a good preparation to do these example exercises under exam-like conditions. Do it relatively early so that you still have time to practice more or ask questions if you find this exam difficult.
- In the real midterm exam, space for the solutions will be provided.
- You may use books and notes in the exam. But don't forget that the time is limited. Books will only help if you know exactly where to search (i.e. just must have read them before). It might be helpful to make your own summary notes when you repeat the material for the exam.
- If you do not understand the exercises, please ask (also during the real exam).
- The midterm exam will be on February 23. Please check the web page for any changes.

Exercise 1 (ER Design) 10 Points

Please design an ER-Schema for a very simplified database of a physician (doctor). The physician wants to store data about:

- Patients: Social security number, first name, last name.
- Some, but not all patients belong to a health management organisation (HMO). In this case, also the organisation name and membership number has to be stored.
- Meetings with patients (appointments): Which patient came on which day to the doctors office. You can assume that there cannot be two meetings with the same patient on the same day (or else this would somehow be treated as a single meeting). The doctor wants to store the diagnosis or some other text about each such meeting. For simplicity, you do not have to consider scheduled meetings (lying still in the future).
- Prescriptions: Which drug was prescribed to a patient on a specific day (as an outcome of a meeting). It is possible that several drugs are prescribed to a patient in a single meeting.
- Drugs: The name of the drug and some information about the drug collected in a single text-valued attribute (so you do not have to worry about any internal structure in this information).

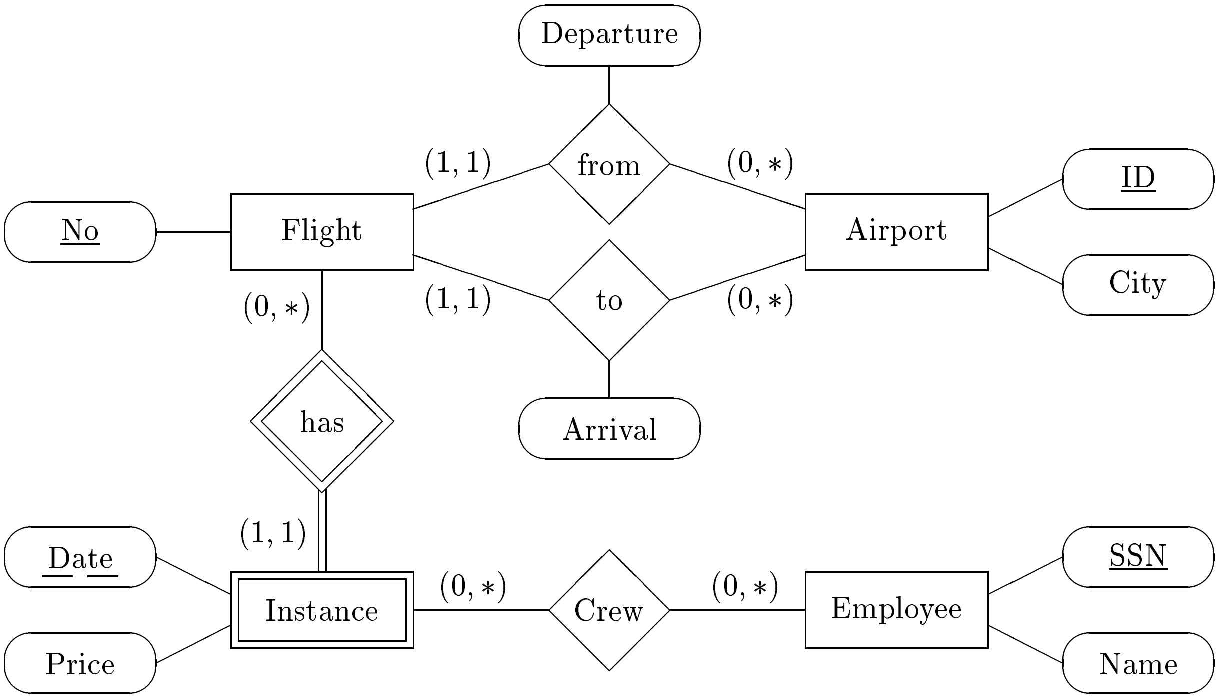
Note that the doctor wants to keep historical information (data about past meetings and prescriptions).

Please define also keys and cardinalities. If other constraints should be needed, it suffices to sketch them in natural language.

Exercise 2 (ER to Relational Mapping) 10 Points

You are given the following ER-Schema for a simplified airline database. It contains information about:

- Airports: The airport id (e.g. PIT, JFK) and the name of the city.
- Flight Routes: The number of the flight (e.g. “US 781”), the airport from which the flight starts and the goal airport (e.g. from Frankfurt to Pittsburgh), and the planned departure and arrival times (e.g. 11:35 and 14:45). For simplicity, intermediate stops are not represented.
- Flight Instances: A flight scheduled for a particular date, e.g. flight “US 781” on February 25. Also the enconomy class price for this flight instance is stored (it depends on the date).
- Employees: Their social security number and name.
- Crew Assignments: Which employee is scheduled to work on which flight instance? For simplicity, different types of employees (e.g. pilots and stewards) are not represented.



Translate this schema into the relational model. Specify keys, foreign keys, and not null constraints. If further constraints are needed to make the relational schema equivalent to the ER-schema, please explain them in natural language.

Exercise 3 (Constraints) 2 Points

Please explain one additional constraint for the flights database which excludes impossible database states. You do not have to specify the constraint formally. Only explain a situation which the ER-schema allows, but which cannot happen in the real world.

Example Database for Exercises 4 to 7

The following relational database contains information about multivitamin/multimineral tablets (nutrition supplements). It consists of three relations (tables):

- Supplement(Name: string, Producer: string not null)

Supplement	
Name	Producer
Centrum	Lederle
One A Day	Bayer
⋮	⋮

- Vitamin(Substance: string, Daily Value: number not null, Unit: String not null)
  /* Despite the name, this relation also contains minerals etc. */

Vitamin		
Substance	Daily Value	Unit
Vitamin A	5000	IU
Vitamin C	60	mg
Biotin	300	mcg
Iron	18	mg
⋮	⋮	⋮

- Contents(Name → Supplement, Substance → Vitamin, Quantity: number not null)
  /* Quantity is the contents in one tablet, same units as in table “Vitamin” */

Contents		
Name	Substance	Quantity
Centrum	Vitamin A	5000
One A Day	Vitamin A	5000
Centrum	Vitamin C	60
One A Day	Vitamin C	60
Centrum	Biotin	30
Centrum	Iron	18
One A Day	Iron	27
⋮	⋮	⋮

All attributes are not null. Since the “not null” constraint is implied by the primary key constraint, it is explicitly listed only for the non-key attributes.

Of course, the database state is only an example. Your queries must work with any database state.

Exercise 4 (Relational Algebra) — 9 Points

Formulate the following queries in relational algebra:

a) Which nutrition supplements contain "Iron"? List the name of the supplement and the quantity of iron.

b) List all substances (Vitamins and Minerals) which are contained in the nutrition supplement "Centrum". Print the name of the substance, the quantity, and the unit (IU, mg, mcg, etc.).

c) Which nutrition supplements contain more than the recommended daily value of any of their substances? List name, producer and the substance.

Exercise 5 (Relational Algebra, Advanced Queries) — 6 Points

Please formulate also the following queries to the multivitamin tables database in relational algebra:

a) Which nutrition supplements (the name suffices) do not contain "Biotin".

b) Which nutrition supplements contain "Vitamin C", but no other Substances? So you have to select only the pure Vitamin C tablets. Print name and producer. You can assume that all tablets contain at least one substance (there are no dummies). So it suffices to check that the supplement contains no other substance (then it must automatically contain Vitamin C).

Exercise 6 (SQL) — 9 Points

Formulate the queries of Exercise 4 also in SQL:

a) Which nutrition supplements contain "Iron"? List the name of the supplement and the quantity of iron.

b) List all substances (Vitamins and Minerals) which are contained in the nutrition supplement "Centrum". Print the name of the substance, the quantity, and the unit (IU, mg, mcg, etc.).

c) Which nutrition supplements contain more than the recommended daily value of any of their substances? List name, producer and the substance.

Exercise 7 (Reverse Engineering) 5 Points

Define an ER-schema which corresponds to the above "Nutrition Supplement" database. So when your ER-schema is translated into the relational model, the result should be exactly the given relational schema.

Exercise 8 (BCNF) 2 Points

Suppose you have a relation

Homework_Result(Student_ID, Exercise_No, Points, Max_Points)

and the following two functional dependencies:

- Student_ID, Exercise_No $\rightarrow$ Points
- Exercise_No $\rightarrow$ Max_Points

Is the relation in BCNF? Please give a short explanation of your answer.

Exercise 9 (ER Diagrams) 4 Points

The following ER-schema contains (at least) 4 errors. Please identify them.