

Exercises Week 1: ER and Data Model

Exercise 1.1 Explain the following terms briefly: attribute, domain, entity, relationship, entity set, relationship set, one-to-many relationship, many-to-many relationship, participation constraint, overlap constraint, covering constraint, weak entity set, aggregation.

Exercise 1.2 Consider the following information about a university database:

- Professors have an SSN, a name, an age, a rank, and a research specialty.
- Projects have a project number, a sponsor name (e.g., NSF), a starting date, an ending date, and a budget.
- Graduate students have an SSN, a name, an age, and a degree program (e.g., M.S. or Ph.D.).
- Each project is managed by one professor (known as the project's principal investigator).
- Each project is worked on by one or more professors (known as the project's co-investigators).
- Professors can manage and/or work on multiple projects.
- Each project is worked on by one or more graduate students (known as the project's research assistants).
- When graduate students work on a project, a professor must supervise their work on the project. Graduate students can work on multiple projects, in which case they will have a (potentially different) supervisor for each one.
- Departments have a department number, a department name, and a main office.
- Departments have a professor (known as the chairman) who runs the department.
- Professors work in one or more departments, and for each department that they work in, a time percentage is associated with their job.
- Graduate students have one major department in which they are working on their degree.
- Each graduate student has another, more senior graduate student (known as a student advisor) who advises him or her on what courses to take.

Design and draw an ER diagram that captures the information about the university. Use only the basic ER model here; that is, entities, relationships, and attributes. Be sure to indicate any key and participation constraints.

Exercise 1.3 The Prescriptions-R-X chain of pharmacies has offered to give you a free lifetime supply of medicine if you design its database. Given the rising cost of health care, you agree. Here's the information that you gather:

- Patients are identified by an SSN, and their names, addresses, and ages must be recorded.
- Doctors are identified by an SSN. For each doctor, the name, specialty, and years of experience must be recorded.
- Each pharmaceutical company is identified by name and has a phone number.
- Each pharmacy has a name, address, and phone number.
- For each drug, the trade name and formula must be recorded. Each drug is sold by a given pharmaceutical company, and the trade name identifies a drug uniquely from

among the products of that company. If a pharmaceutical company is deleted, you need not keep track of its products any longer.

- Every patient has a primary doctor. Every doctor has at least one patient.
 - Doctors prescribe drugs for patients. A doctor could prescribe one or more drugs for several patients, and a patient could obtain prescriptions from several doctors. Each prescription has a date and a quantity associated with it. You can assume that, if a doctor prescribes the same drug for the same patient more than once, only the last such prescription needs to be stored.
 - Each pharmacy sells several drugs and has a price for each. A drug could be sold at several pharmacies, and the price could vary from one pharmacy to another.
 - Pharmaceutical companies have long-term contracts with pharmacies. A pharmaceutical company can contract with several pharmacies, and a pharmacy can contract with several pharmaceutical companies. For each contract, you have to store a start date, an end date, and the text of the contract.
 - Pharmacies appoint a supervisor for each contract. There must always be a supervisor for each contract, but the contract supervisor can change over the lifetime of the contract.
1. Draw an ER diagram that captures the preceding information. Identify any constraints not captured by the ER diagram.
 2. How would your design change if each drug must be sold at a fixed price by all pharmacies?
 3. How would your design change if the design requirements change as follows: If a doctor prescribes the same drug for the same patient more than once, several such prescriptions may have to be stored.

Exercise 1.4 Define the following terms: *relation schema*, *relational database schema*, *domain*, *attribute*, *attribute domain*, *relation instance*, *relation cardinality*, and *relation degree*.

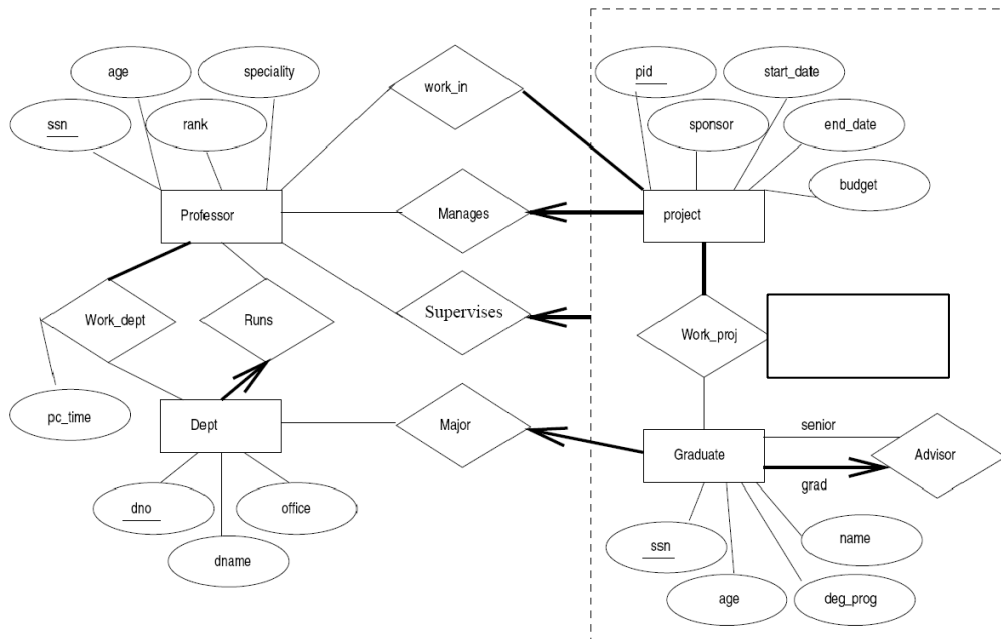
Solutions

Answer 1.1 Term explanations:

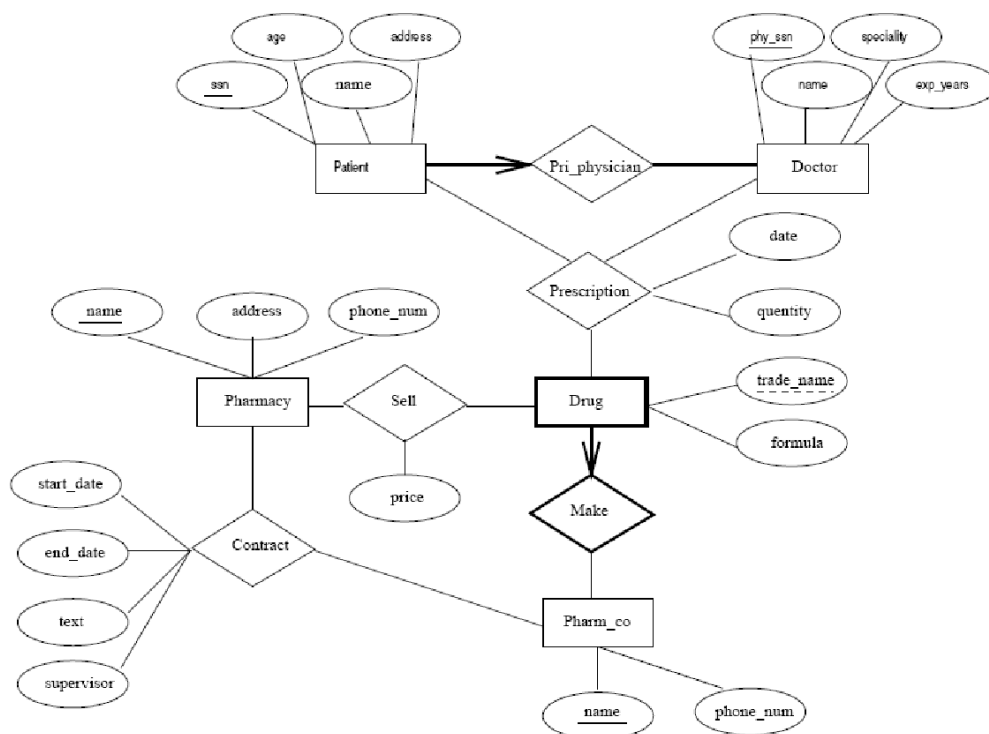
- *Attribute* - a property or description of an entity. A toy department employee entity could have attributes describing the employee's name, salary, and years of service.
- *Domain* - a set of possible values for an attribute.

- **Entity** - an object in the real world that is distinguishable from other objects such as the green dragon toy.
- **Relationship** - an association among two or more entities.
- **Entity set** - a collection of similar entities such as all of the toys in the toy department.
- **Relationship set** - a collection of similar relationships
- **One-to-many relationship** - a key constraint that indicates that one entity can be associated with many of another entity. An example of a one-to-many relationship is when an employee can work for only one department, and a department can have many employees.
- **Many-to-many relationship** - a key constraint that indicates that many of one entity can be associated with many of another entity. An example of a many-to-many relationship is employees and their hobbies: a person can have many different hobbies, and many people can have the same hobby.
- **Participation constraint** - a participation constraint determines whether relationships must involve certain entities. An example is if every department entity has a manager entity. Participation constraints can either be total or partial. A total participation constraint says that every department has a manager. A partial participation constraint says that every employee does not have to be a manager.
- **Overlap constraint** - within an ISA hierarchy, an overlap constraint determines whether or not two subclasses can contain the same entity.
- **Covering constraint** - within an ISA hierarchy, a covering constraint determines where the entities in the subclasses collectively include all entities in the superclass. For example, with an Employees entity set with subclasses HourlyEmployee and SalaryEmployee, does every Employee entity necessarily have to be within either HourlyEmployee or SalaryEmployee?
- **Weak entity set** - an entity that cannot be identified uniquely without considering some primary key attributes of another identifying owner entity. An example is including Dependent information for employees for insurance purposes.
- **Aggregation** - a feature of the entity relationship model that allows a relationship set to participate in another relationship set. This is indicated on an ER diagram by drawing a dashed box around the aggregation.

Answer 1.2 The ER diagram is shown in the figure.



Answer 1.3 The ER diagram is shown in the figure.



Note: We allow a pharmacy to exist in the database even if it does not participate in the 'sell' relationship, since in the very first moments of its existence we might not have yet any data about what drugs it sells.

2. If the drug is to be sold at a fixed price we can add the price attribute to the Drug entity set and eliminate the price from the Sell relationship set.

3. The date information can no longer be modeled as an attribute of Prescription. We have to create a new entity set called Prescription date and make Prescription a 4-way relationship set that involves this additional entity set.

Answer 1.4 A *relation schema* can be thought of as the basic information describing a table or *relation*. This includes a set of column names, the data types associated with each column, and the name associated with the entire table. For example, a relation schema for the relation called Students could be expressed using the following representation:

cStudents(*sid*: string, *name*: string, *login*: string, *age*: integer, *gpa*: real)

There are five fields or columns, with names and types as shown above.

- A *relational database schema* is a collection of relation schemas, describing one or more relations.
- *Domain* is synonymous with *data type*. *Attributes* can be thought of as columns in a table. Therefore, an *attribute domain* refers to the data type associated with a column.
- A *relation instance* is a set of tuples (also known as *rows* or *records*) that each conform to the schema of the relation.
- The *relation cardinality* is the number of tuples in the relation.
- The *relation degree* is the number of fields (or columns) in the relation.