3. Conceptual Database Design

Introduction

- most known conceptual data model on a high abstraction level, easy to understand, independent of aspects of data organization and data management
- E-R model (besides **UML**) has great importance in practice
- two-phase procedure for DB design
  - phase 1: requirements analysis and design of an E-R model
  - phase 2: transformation of the E-R model into a concrete logical model
- goal: modeling of an interesting part of the “real world” by abstraction so that questions about it can be answered with the aid of the model

**E-R model describes the “real world” by**

- **entities** (objects)
- **attributes** (properties)
- **relationships** between entities
Entity

- **entities** are distinguishable, independent, self-contained, physically or intellectually existing concepts of the mini-world to be modeled

- similar entities are collected in an **entity set**, e.g., the set of all books, the set of all cars

- an entity is described by a set of pertaining properties (attributes), e.g., each book has an ISBN number, an author, a publisher, ...

- The values of an attribute are from domains like *integer*, *real*, *string*, ... e.g., the name of an author is of type *string*

- a minimal set of attributes whose values uniquely characterize the associated entity among all entities of its type is called **key**, e.g., ISBN number identifies a book, an article number an article
Relationship

- A relationship describes a connection between several entities, e.g., student Smith attends lecture COP 4720, teaching assistant Benson works for professor Meyer.

- A homogeneous set of relationships is collected in a relationship set, e.g., relationship sets attends_lecture or works_for.

- Formal: Relationship set $R$ between the entity sets $E_1, E_2, ..., E_n$ as a relation, i.e.,
  $$R \subseteq E_1 \times E_2 \times ... \times E_n,$$
  where $n$ is the degree of relationship set $R$.
  
  For example:
  - $\text{attends}_\text{lecture} \subseteq \text{students} \times \text{lectures}$
  - $\text{works}_\text{for} \subseteq \text{TAs} \times \text{professors}$

- Attributes may characterize relationships, e.g., frequency as an attribute for attends_lecture.

- An entity set can occur more than once in a relationship set.

- If there is only one entity set $E$ participating in a binary relationship $R(E, E)$, each of these entity sets can be assigned roles.
  For example, $\text{is}_\text{precondition}_\text{of} \subseteq \text{lectures} \times \text{lectures}$.
  First lecture / second lecture has the role of a predecessor / successor.
Constraints of binary relationship sets

- **1:1-relationship** (one-to-one relationship)
  if for a binary relationship set $R(E_1, E_2)$ each entity in $E_1$ is associated with at most one entity in $E_2$, and vice versa

- **1:m-relationship** (one-to-many relationship)
  if for a binary relationship set $R(E_1, E_2)$ each entity in $E_1$ is associated with any number (zero or more) of entities in $E_2$, and each entity in $E_2$ is associated with at most one entity in $E_1$

- **m:1-relationship** (many-to-one relationship)
  analogous to the 1:m-relationship

- **m:n-relationship** (many-to-many relationship)
  if for a binary relationship set $R(E_1, E_2)$ each entity in $E_1$ is associated with any number (zero or more) of entities in $E_2$, and vice versa

- constraints considered as partial functions, e.g.
  for 1:1-relationship: has_husband: women $\rightarrow$ husbands, has_wife: men $\rightarrow$ wives
  for m:1-relationships: employed_by: persons $\rightarrow$ companies
E-R diagrams

- graphical representation of entity sets, relationship sets, and their attributes by means of a graph

Notations

- rectangles represent entity sets: $E$

- ellipses represent attributes: $A$
  - they are connected with their entity set by undirected edges
  - key attributes are underlined

- relationship sets are represented by diamonds: $R$
  - relationship sets are connected with their pertaining entity sets by edges
  - edges carry information about cardinality according to imposed constraints

- a role of a relationship set is attached to the corresponding edge