Semantic Data Modeling

- What is Conceptual Data Modeling
- Entity-Relationship (E-R) Modeling
- Limitations of E-R Modeling
- Object-oriented Modeling: Another semantic model (discussed later in this course)
What Is Conceptual Data Modeling?

A process that represents the entities, relationships, and activities of an enterprise in terms of a set of abstract concepts of a chosen data model for specific purposes.

*Enterprise Modeling, Business Modeling*
Bridge the Gap

STUDENT( ID, Name, Age, Address, GPA )
INSTRUCTOR ( Emp#, Name, Rank, Dept )
COURSE ( Course#, Credits, Title )
CLASS ( Emp#, ID, Course#, Time, Room )
Data Semantics

Static Information

- Data -- Entities
- Associations -- Relationships among entities

Dynamic Information

- Activities -- Operations/transactions
- Integrity constraints -- Business rules/regulations and data meanings
A conceptual data model consists of:

- A collection of formal concepts
- A set of usage rules

Different model has different modeling capability

- Conventional (Logical data modeling)
  -- Hierarchical
  -- Network
  -- Relational

- Semantic data modeling
  -- E-R
  -- EER
  -- etc.

- Object-Oriented data modeling
E-R Modeling

Introduced by Peter Chen in 1976

Basic modeling concepts:
- Entities, entity types, and attributes
- Relationships

Diagram:
- Instructor
  - Emp#
  - FName
  - MInit
  - LName
  - Language
- Department
  - Date
- Assigned
  - Date
- Office
  - Time
  - Location
- Teaches
  - Course
  - Time
  - Location
- Works_for
  - 1
  - N
  - 1
- Course
  - M
E-R Notation

- Entity
- Relationship
- Attribute
- Primary Key
Entities

- An entity is a conceptual object
- Physically exists
- Usually a noun in requirement specification
Entity Types

- A collection of similar entities
- An abstraction of "physical" entities
  - A noun in requirement specifications
  - Having "independent" meaning

Diagram:
- Department
  - CIS
  - Acct
- Student
  - Jose
  - Steve
  - Alice
- Course
  - CIS 2010
  - CIS 3730
Weak Entity Types

Can't exist in DB independently

Must be identified by its owner

- Owner entity type
- Identify relationship
- Partial key
- Total participation
Attributes

Properties or characteristics of entities and entity types

- Attribute values -- Properties of entities
- Value set - All acceptable attribute values
- Attributes (definitions) -- Properties of entity types

- A noun or an adjective in requirement specifications
- No "independent" meaning

Diagram:

- Student
  - ID
    - "123-45-6789"
  - Age
    - 25
- Jose
Key Attributes

One or a group of attributes that can uniquely identify individual entities of an entity type

- A key refers to one or a group of attributes as a whole
- A key attribute is a component attribute of a key
- Key changes with data semantics

An entity type may have several qualified keys

- Primary key -- One of the candidate keys
- Alternate key - Candidate keys not used as the primary key
- Secondary key -- An identifier of records with similar properties of interest

- The primary key attribute(s) is(are) underlined
More Attributes

- **Simple attribute**
  - Contains atomic values only

- **Composite attribute**
  - Has component attributes

- **Single-valued attribute**
  - Has exactly one value per entity

- **Multi-valued attribute**
  - Contains repeating values per entity

- **Derived attribute**
  - Attribute values computed by means of other attributes
Relationships

Associations among entities

- Relationships -- Associations among entities
- Usually a verb in requirement specification
Relationship Degrees

The number of entity types associated with that relationship. Each entity in a relationship is known as a participant.
N-ary Relationships

- A link must associate with all participants
- Cardinality is with respect to individual relationships
  - A N-ary relationship is not equivalent to N binary relationships
Relationship Attribute

- Describes the association
  - A adverb or noun in requirement specification

![ER diagram for Instructor teaches Course relationship with attributes sect#, time, room#, and date]
How entities are connected through a relationship

- **One-to-One** -- An entity of $E_1$ is connected to at most one entity of $E_2$ and vice versa.

- **One-to-Many** -- An entity of $E_1$ may be connected to one or more entities of $E_2$, but an entity of $E_2$ can only be mapped to at most one entity of $E_1$.

- **Many-to-Many** -- An entity of $E_1$ may be linked to one or more entities of $E_2$, and vice versa.
• Recursive Relationship
  – Relationship type where *same* entity type participates more than once in *different roles*.

• Relationships may be given role names to indicate purpose that each participating entity type plays in a relationship.
Recursive Relationship called \textit{Supervises} with Role Names

‘Staff (Supervisor) supervises staff (Supervisee)’
Entities associated through two distinct Relationships with Role Names

- Manager manages branch office
- Manager manages branch office
- Branch office has member of staff
- Branch office has member of staff
Participation Constraints

Let $R$ be a relationship type involves entity type $E$ as a participant. If every instance of $E$ participates in at least one instance of $R$, then the participation is said to be total; otherwise it is said to be partial [Date, 2004]

\[\text{Partial (optional) participation}\]

An entity of $E$ does not have to be mapped to another entity through the relationship.

\[\text{Total (mandatory) participation}\]

Every entity of $E$ must be connected through the relationship to other entity (or entities).
Structural Constraints

- Main type of constraint on relationships is called multiplicity.

- Multiplicity - number (or range) of possible occurrences of an entity type that may relate to a single occurrence of an associated entity type through a particular relationship.

- Represents policies (called business rules) established by user or company.
Multiplicity
Other E-R Models

1. Entity 1 \rightarrow Relationship \rightarrow Entity 2

2. attr. 1 \rightarrow Relationship \rightarrow attr. 2

3. Entity 1 \rightarrow Relationship \rightarrow Entity 2

4. Entity 1  (1:m) Relationship  (0:5) Entity 2
Pros and Cons of E-R

Advantages

- Simple and easy to understand.
- Very popular.
- Semantic richer than classical data models.

Disadvantages:

- Not a formally defined data model.
- Deals with some integrity constraints.
- Difficult to distinguish entities from relationships.
- Has redundant modeling information.
Is-A Relationship

Generalization and specialization hierarchy

- Supertypes -- Hide the differences of subtypes
- Subtypes -- Reveal specific properties

**Generalization**

- Person
  - Student
    - UnderGrad
    - Grad
  - Staff
    - Faculty
    - Secretary

**Specialization**
Inheritance

- A supertype contains the common properties of all its subtype entities.
- Subtype inherits properties of its supertype and may have its own properties.

```
<table>
<thead>
<tr>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID, Name, Address</td>
</tr>
</tbody>
</table>

- Employee
  
  | Project Manager |
  | Salary, Skill   |
  | Ratio-of-success, ... |

- Dependent
  
  | Project Member |
  | Birth-date    |
  | Languages, ... |
```
A database is being constructed to keep track of the teams and games of a baseball league. A team has a number of players, not all of whom participate in each game. It is desired to keep track of the players participating in each game for each team, the positions they played in that game, and the result of the game. Design an ER schema diagram for this application.

Assumptions:
Each game in the schedule is identified by a unique Game#, and a game is also identified uniquely by the combination of Date, starting Time, and Field where it is played. A performance attribute is used to store information on the individual batting performance of each player in a game.
American Airlines Company

The American Airlines Company publishes a monthly flight log report that tracks which type of aircraft and the number of hours that were flown by an individual pilot. A separate report is prepared for each pilot and is used to monitor pilot flight proficiency for the two types of aircraft (fixed-wing and rotorcraft) which a pilot may be qualified to fly.

The following business rules apply to this report. Pilots may be assigned to fly different aircraft each day by the flight scheduling manager. Each aircraft has a single crew chief permanently assigned to perform maintenance on the aircraft, although a crew chief may crew more than one aircraft. Each aircraft is identified by an aircraft number. There are several aircraft types in the fleet. An aircraft number is unique within an aircraft type.

Identify the entities in this situation and draw an E-R diagram of the entities to include their relationships and any attributes identified in this example. Give examples of additional attributes that might be associated with each entity.