ER Modeling

Problem statement

Suppose someone asks you to model data for their business or organization.

Problems:

- where do you even begin?
- who has the information? who are the stakeholders?
- should you write a relational schema right away?
- how to avoid need for big changes in future?
- how to share your model with others?

What is a good first model?

Some things we might want in a good first model:

- uses simple, everyday concepts
- general – can map onto various detailed models
- is easy for non-experts to understand
- can be presented as a picture
- ties closely to elements of human language
- provides a conceptual framework for later work

Would Java be a good first modeling language?

Would English be a good first modeling language?

Phases of the database design process

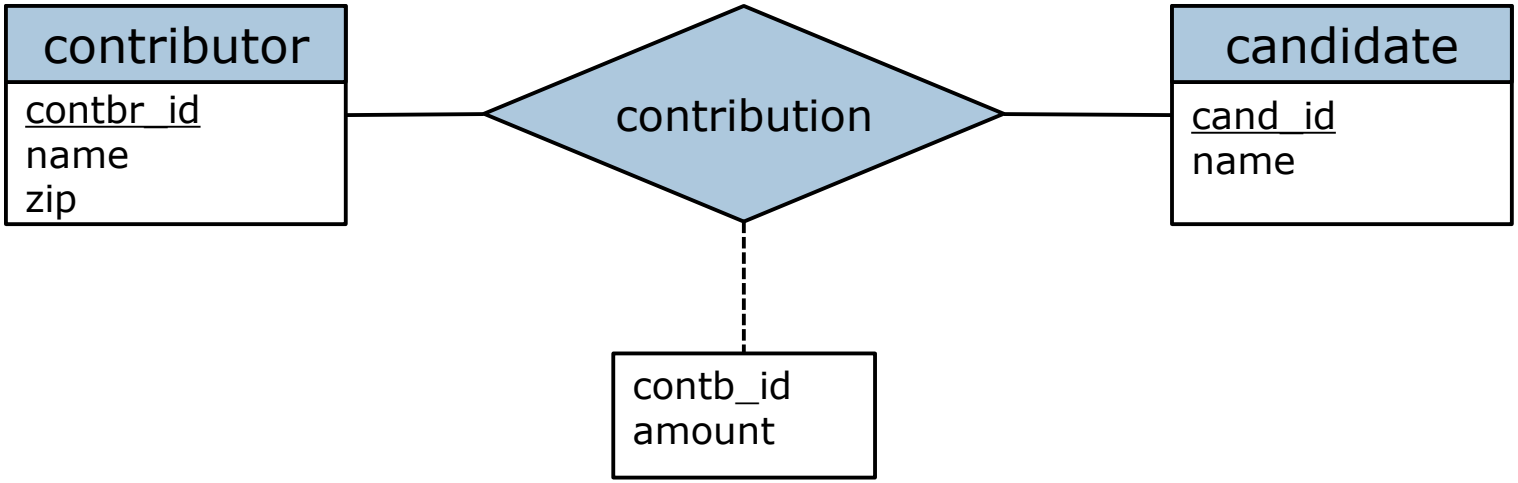
1. Understand data needs of future users
   - talk to users and domain experts; write things down
2. Create a conceptual design
   - identify objects, their relationships, etc.
   - ER model
3. Identify functional requirements
   - what will users do with the data?
4. Logical and physical design
   - map conceptual design to data model
   - specify physical features of DB, like file organization

Entity-Relationship (ER) Models

- Developed by Peter Chen of MIT
- Used for conceptual design
- Building blocks of an ER model:
  - entities
  - relationships
  - attributes

We'll use a digital book library as our running example

Example ER Model



Entities and entity sets

Entity

- a real-world thing that is distinguishable from others
- it has a set of properties, or attributes
- can be physical (book) or abstract (course section)

Entity set

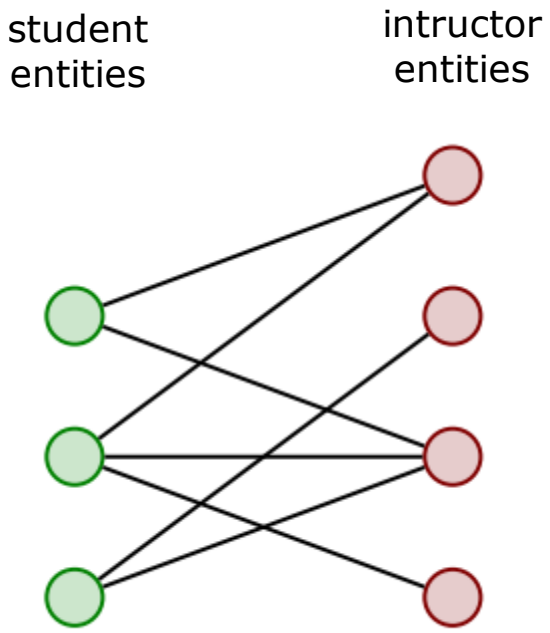
- a collection of similar entities
- example: “instructor” is the collection of people at a university that share some properties

Attributes

- Similar to object-oriented prog. languages
- Each entity has a value for each of its attributes
- ‘Instructor’ attributes could include ID, name, salary, etc.
- Example: instructor has value “Wu” for name

Relationships and relationship sets

- Relationship
  - an association between entities
  - for example, student Shankar is advised by instructor Gold
  - a relationship can involves 2 or more entities
- Relationship set
  - a collection of relationships of the same type
  - for example, ‘teaches’ and ‘takes’ are relationship sets



(figure from slides by Peter Wood, University of London)

More on relationship sets

- Participation
  - 'instructor' participates in the 'teaches' relationship set
  - 'student' participates in the 'takes' relationship set
- Roles
  - this is useful in relationship sets like 'prerequisite', where the same entity set appears more than once
  - example: "parent-of" relationship between people
- Attributes of relationships
  - 'advisor' relationship has a date when advising began
  - 'takes' relationship has a grade attribute

More on attributes

- Every attribute has a domain of permitted values
  - like a type in a programming language
- Attributes can be simple or composite
  - simple: age, salary
  - composite: date, address
- Attributes can be multi-valued
  - example: my contact phone numbers
- Attributes can take value 'null'