Modeling: Entity-Relationship Diagrams

1

Scenario

- http://www.imdb.com wants to store information about movies and has chosen **you** to help them
- Four steps:
  - Requirements Analysis: Discover what information needs to be stored, how the stored information will be used, etc. Taught in course on system analysis and design
  - Conceptual Database Design: High level description of data to be stored (ER model)
  - Logical Database Design: Translation of ER diagram to a relational database schema (description of tables)
  - Physical Database Design: Done by the DB system

2

Requirements

- For actors and directors, we want to store their name, a unique identification number, address and birthday (why not age?)
- For actors, we also want to store a photograph
- For films, we want to know the title, year of production and type (thriller, comedy, etc.)
- We want to know who directed and who acted in each film. Every film has one director. We store the salary of each actor for each film
- Etc…

3

ER-Diagrams: General Information

- ER-diagrams are a formalism to model real-world scenarios
- There are many versions of ER-diagrams that differ both in their appearance and in their meaning
  - We will use the version appearing in the book **Database Systems: The Complete Book**
- ER-diagrams have a formal semantics (meaning) that must be thoroughly understood, in order to create correct diagrams
- **Goal** of modeling is to translate informal requirements to a precise diagram. This diagram can then be translated into to the desired **data model**, to allow data to be stored in a database

4

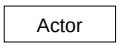
Basic Concepts: Entities, Attributes, Relationships

5

Entities, Entity Sets

- **Entity** (ישות): An object in the world that can be distinguished from other objects
  - Examples of entities:
  - Examples of things that are not entities:
- **Entity set** (קבוצת ישויות): A set of similar entities
  - Examples of entity sets:

Entity sets are drawn as rectangles



6

Attributes

- **Attributes** (תכונות): Used to describe entities
  - All entities in the set have the same attributes
  - A minimal set of attributes that uniquely identify an entity is called a **key**
  - An attribute contains a single piece of information (and not a list of data)

7

Attributes (2)

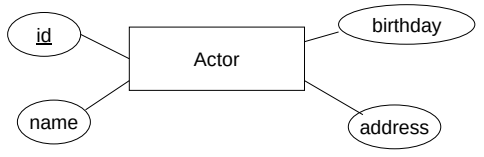
- Examples of attributes:
- Examples of things that cannot be attributes:

✎ Attributes are drawn using ovals

✎ The names of the attributes which make up a key are underlined

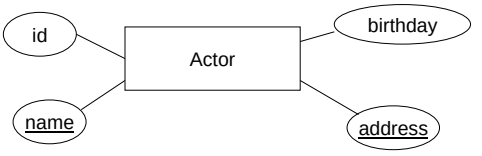
8

Example



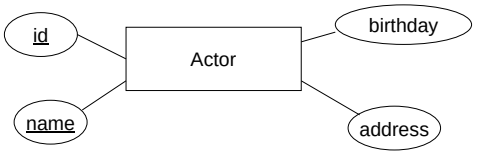
9

Another Option for a Key?



10

Another Option for a Key?



11

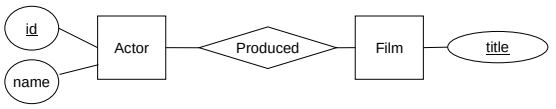
Relationships, Relationship Sets

- **Relationship** (קשר): Association among two or more entities
  - Examples of Relationships:
- **Relationship Set** (קבוצת קשרים): Set of similar relationships
  - Examples of Relationship sets:

✎ Relationship sets are drawn using diamonds

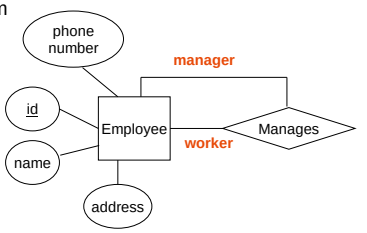
12

Example



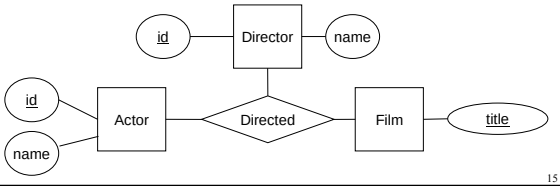
Recursive Relationships

- An entity set can participate more than once in a relationship
- In this case, we add a description of the role to the ER-diagram



n-ary Relationship

- An *n*-ary relationship set *R* involves exactly *n* entity sets: E_1, …, E_n.
- Each relationship in *R* involves exactly *n* entities: e_1 in E_1, …, e_n in E_n
- Formally, $R \subseteq E_1 x \ldots x E_n$

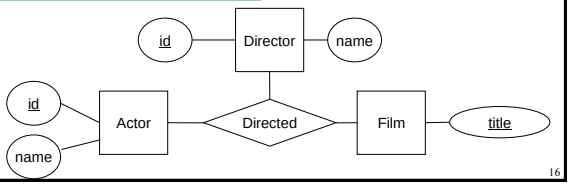


Example

- Suppose that there are:
  - 2 Actors
  - 3 Directors
  - 4 Film

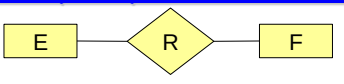
How many pairs can be in the relationship set "Directed"?

How many triples can be in the relationship set "Directed"?



Multiplicity of Relationships

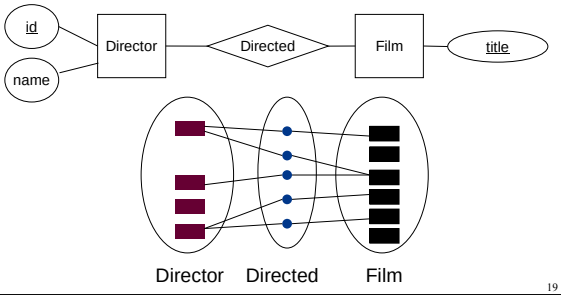
Multiplicity of Relationships



- A member of E may be connected by R to any number of members from F, and vice versa
- This is called a **many-many relationship**

Many-to-Many

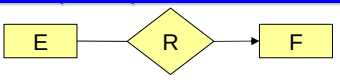
- ✓ A film is directed by any number of directors
- ✓ A director can direct any number of films



Director Directed Film

19

Multiplicity of Relationships

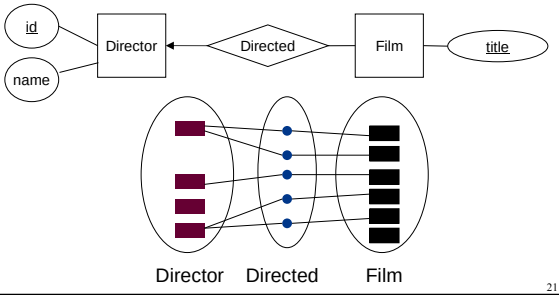


- By adding arrows to the diagram, we can indicate constraints on the relationship
- An arrow towards F indicates that:
  - A member of E may be connected by R to at most one members from F
  - (Still, a member of F may be connected by R to any number of members from E)
- This is called a **many-one relationship**

20

One-to-Many

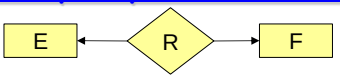
- ✓ A film is directed by at most one director
- ✓ A director can direct any number of films



Director Directed Film

21

Multiplicity of Relationships

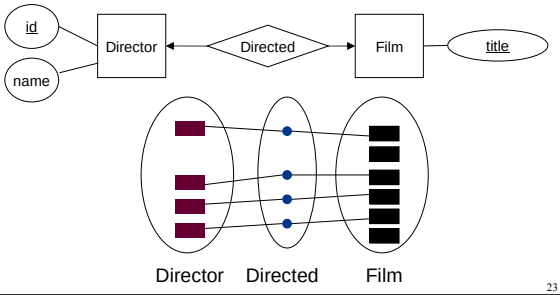


- An arrow towards F and towards E indicates that:
  - A member of E may be connected by R to at most one member from F
  - A member of F may be connected by R to at most one member from E
- This is called a **one-one relationship**

22

One-to-One

- ✓ A film is directed by at most one director
- ✓ A director can direct at most one film

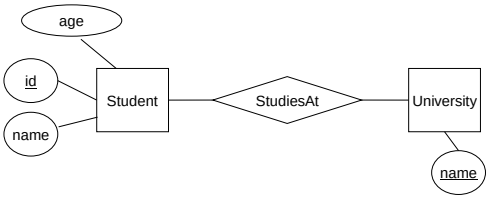


Director Directed Film

23

Example

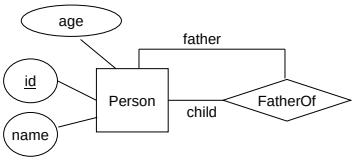
Where would you put the arrow?



24

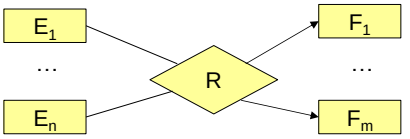
Another Example

Where would you put the arrow?



25

Multiplicities in Multiway Relationships

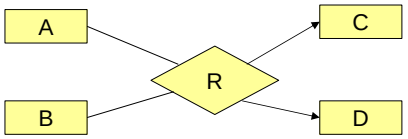


For any $1<=i<=m$

For any tuple of entities $e_1,\ldots,e_n,f_1,\ldots f_{i-1},f_{i+1},\ldots,f_m$ there is at most one f_i, such that $e_1,\ldots,e_n,f_1,\ldots,f_m$ are connected by R

26

Example



Suppose that there are

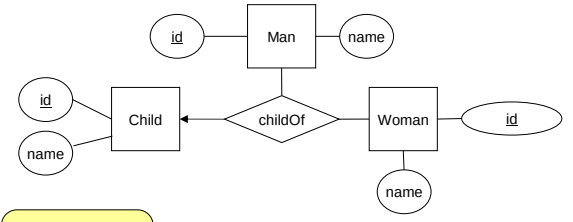
- a entities in A
- b entities in B
- c entities in C
- d entities in D

What is the maximum number of 4-tuples in R?

What is the minimum number of 4-tuples in R?

27

Example

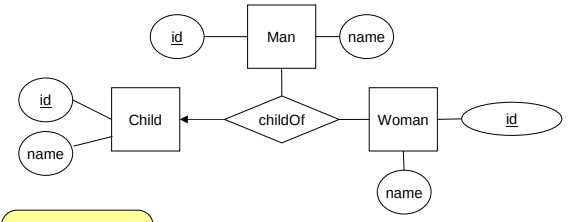


Note that for many reasons, this is a *bad* modeling

What does this mean?

28

Multiplicities in Multiway Relationships



Note that for many reasons, this is a *bad* modeling

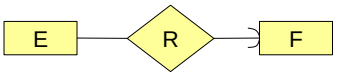
Each pair of a man and a woman can have at most one child

29

Referential Integrity and Degree Constraints

30

Referential Integrity



- So far, we can say that an entity participates *at most one time*, but cannot require it to participate *at least one time*
- The rounded arrow above indicates that each entity in E must participate **exactly one time** in an R-relationship with an entity in F

31

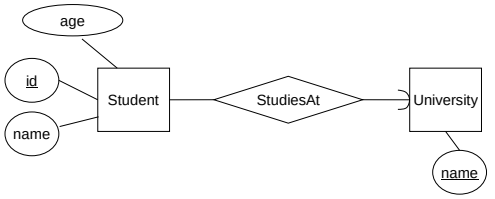
Degree Constraints

- We can attach a bounding number to edges to indicate limits on the number of entities that can be connected to a single entity via a relationship set
- In the example above, a move has at most 10 stars
- Note: a regular arrow is the constraint <=1
- Note: a rounded arrow is the constraint =1

32

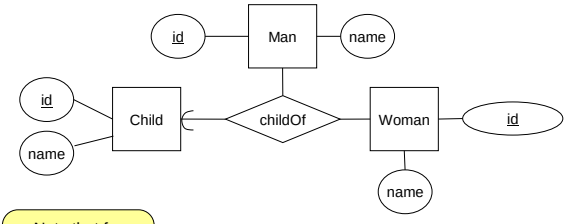
Example (1)

What does this mean?



33

Example (2)



Note that for many reasons, this is a *terrible* modeling

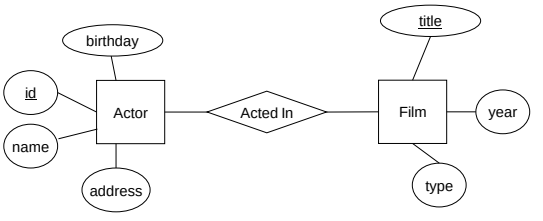
What does this mean?

34

Relationship Sets with Attributes

35

Relationship Sets Can have Attributes



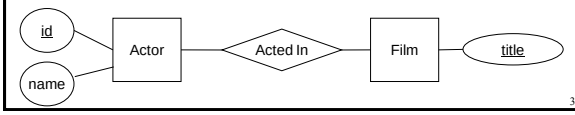
Where does the salary attribute belong?

salary

36

Important Note

- The entities in a relationship set must identify the relationship
- Attributes of the relationship set cannot be used for identification!
- Suppose we wanted to store the role of an actor in a film.
  - How should we store the role of the actor?
  - How would we store information about a person who acted in one film in several roles?



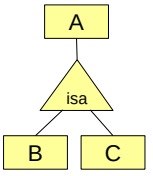
Subclasses

ISA Hierarchies

ISA Relationships: Defines a hierarchy between entity sets

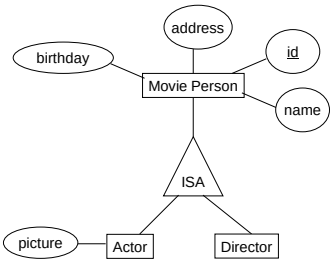
- ISA is similar to inheritance
- ISA relationships are drawn as a triangle with the word ISA inside it. The "super entity-set" is above the triangle and the "sub entity-sets" are below

Implications of an ISA Relationship



- Every entity in B or in C belongs to A
- There may be entities in A that do not belong to B or to C
- There may be entities that belong to both B and C

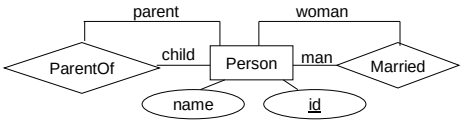
Example



What are the keys of:

1. Movie Person
2. Actor
3. Director

Example



Is this good method of modeling data for the רבנות database on marriage?

How can you fix it?

Weak Entity Sets

43

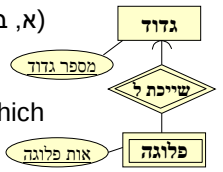
Intuition

- Sometimes, entities cannot be identified by their own attributes.
- To identify such an entity, we need information about a "supporting relationship"
  - Example: Given a bank account number, you cannot identify the actual bank account. For identification, you also have to know the name of the bank.

44

Example

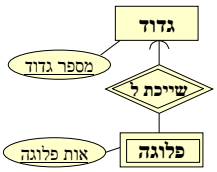
- A פלוגה has a letter (א, ב, ...)
- To uniquely identify the פלוגה, one must know which גדוד it belongs to
- פלוגה is a **weak entity set**
- The relationship set שייכת ל is the **supporting relationship** for פלוגה



45

Notation

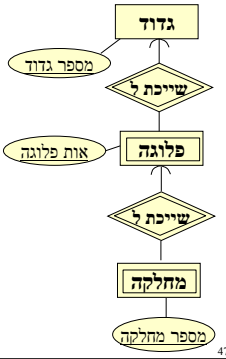
- Weak entity set has a double line
- Supporting relationships have double lines
- Rounded arrow pointing into the identifying entity sets



46

Example (1)

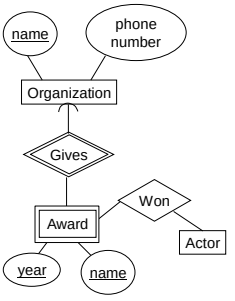
- Keys:
- גדוד: מספר גדוד
- פלוגה: אות פלוגה, מספר גדוד
- מחלקה: מספר מחלקה, אות פלוגה, מספר גדוד



47

Example (2)

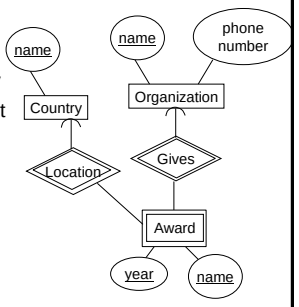
- Same award can be given by several organizations ("Academy award for Best Actor 2007"),
- A year, award name and organization name uniquely define an award
- Weak entity set can participate in additional (non-supporting) relationships



48

Example (3)

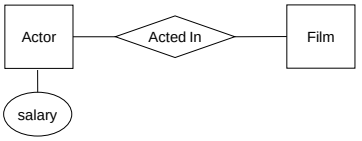
- Awards are now identified by organization and country
- Same award can be given by same organization in different countries ("Academy award for Best Actor Israeli 2007")
- Weak identity set has 2 supporting relationships
- What is identifying key for award?



Design Principles

Faithfulness

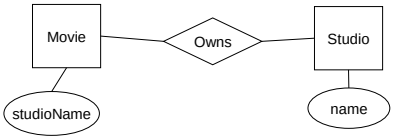
- The design should be accurate to the specifications



- This is ok ***only*** if each actor has a set salary, regardless of all movies

Avoiding Redundancy

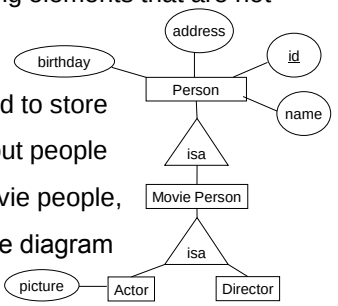
- The design should not model the same information in multiple ways



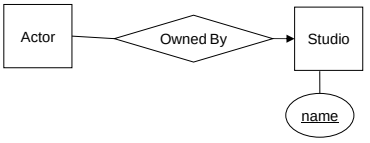
- Leads to fact repetition
- Leads to inconsistencies

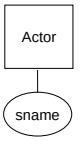
Simplicity Counts

- Avoid introducing elements that are not needed
- If we never need to store information about people that are not movie people, don't put it in the diagram



Picking the Right Kind of Element





- The bottom diagram is sufficient if Studio has no attributes other than name

Summary

- Given a set of requirements, to translate the requirements into a diagram:
  - Identify the entity sets
  - Determine if there are hierarchies (ISA or weak relationships) among entity sets
  - Identify the relationship sets
  - Identify the attributes
  - Determine constraints on relationship participation

55

The Relational Model

56

Data Models

- A **data model** is a notation for describing data
  - Conceptual structure of the data
  - Operations on the data
  - Constraints on the data
- In this course we focus on the **relational data model**

57

Conceptual Structure of the Data

- The basic element of the relational model is a **relation** (which is similar to a table)
- A relation has a **schema**, consisting of a
  - **Name**
  - **List of attributes, possibly with domains**
- A relation may also have an **instance**, which is a set of **tuples** (i.e., rows) in the relation

58

Example

Movies

Title	Year	Length	Genre
Follow…	1985	90	children
Who …	1987	90	mystery

- Schema: Movies(title, year, length, genre)
- Relation name: Movies
- Attributes: title, year, length, genre
- Possible tuple instance
  - ("Follow that Bird", 1985, 90, children)
- Scheme with domains:

  Movies(title: string, year: number, length: number, genre: string)

59

Operations on the Data

- Relational algebra
  - Selection, projection, union, minus, join
- Stay tuned… Discussed in detail next week

60

Constraints on the Data

- We discuss complex constraints later on in the course
- For now, we introduce **key constraints**
- A set of attributes forms a **key** for a relation if there cannot be 2 different tuples with the same values for all attributes of the key
  - Noted with underline
- Examples:
  - Movies(title, year, length, genre)
  - Actor(teudatZehut, name, address)

61

A Step Closer

- Once we have a set of relational schemas in the relational model, we are a step closer to storing data in a DBMS
  - A DBMS has a **data definition language (DDL)**, used to define **tables** in the database
  - Once we have decided on the relational schemas, these can be directly translated into the database using the DDL

62

ER Diagrams to Relational Schemas

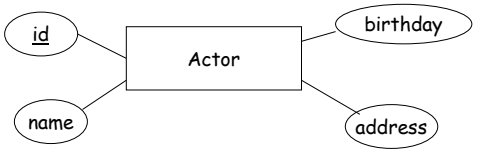
63

General Principles

- When converting ER diagrams to Relations, we should:
  - Reduce duplicate information
  - Constrain as tightly as possible
- Notes:
  - Some scenarios can be represented in different ways.
  - Sometimes we will not be able to fully represent constraints, or will be forced to represent information more than once.

64

Entity Set Translation



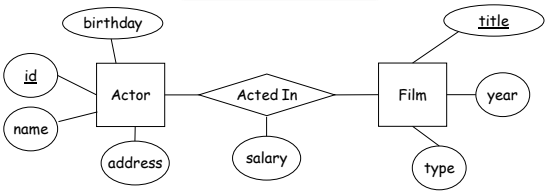
General Rule:

Actor (id, name, birthday, address)

- Create a relation with the name of the Entity.
- There is a column for each attribute
- The key in the diagram is the primary key of the relation

65

Relationship Sets (without constraints)

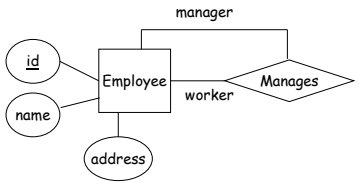


General Rule:

- Create a table with the name of the relationship set
- Relationship table attributes: its own attributes (salary) + all keys of the relating entities (title, id)
- What is the primary key of the table?
- Note: Do not define two attributes with the same name – instead rename one of them in the schema

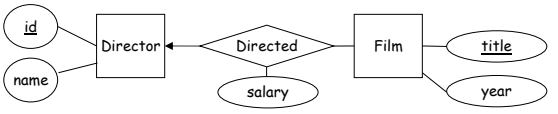
66

Translating Recursive Relationship Sets (without constraints)



What are all the relations created for this diagram?

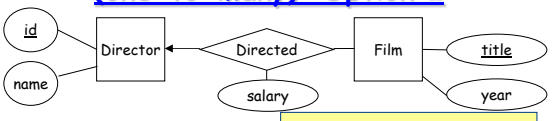
Translating relationships (one-to-many): Option 1



Option 1:

Same as without key constraints (3 tables), except that the primary key of Directed is now title (why?)

Translating relationships (one-to-many): Option 2

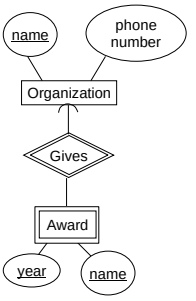


Film (title, year, salary, id)
Director(id, name)

Option 2:

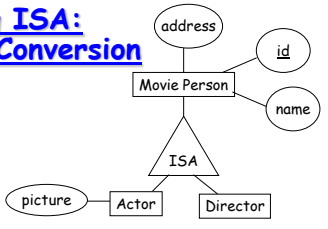
- Do not create a table for the relationship
- Add information columns that would have been in the relationship's relation to the relation of the **entity which does not have the incoming arrow**

Translating Weak Entity Sets



- Key of relation for weak entity set includes its own keys and the keys of its supporting entity sets
- No relation is created for the supporting relationship sets
- **Example:** Translate diagram on the left to relations

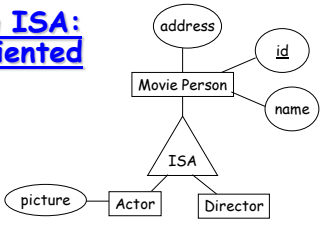
Translating ISA: E/R Style Conversion



- A relation for each entity set. An entity may appear in more than one relation

MoviePerson(id, address, name)
Actor(id, picture)
Director(id)

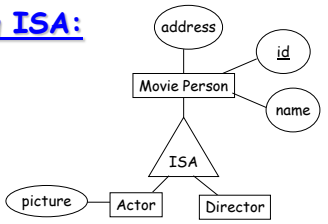
Translating ISA: Object-Oriented Approach



- A relation for each possible combinations of how entities may appear in the entity sets.

MoviePerson(id, address, name)
MoviePersonActor(id, address, name, picture)
MoviePersonDirector(id, address, name)
MoviePersonActorDirector(id, address, name, picture)

Translating ISA: Null Value Approach



- A single relation for containing all values
- Possible only if NULL values (i.e., missing values) are allowed – not in the pure relational model

MoviePerson(id, address, name, picture)

73