Discrete Structures for Computer Science

William Garrison

bill@cs.pitt.edu

6311 Sennott Square

Lecture #28: N-ary relations

Based on materials developed by Dr. Adam Lee

Binary relations establish a relationship between elements of two sets

Definition: Let A and B be two sets. A binary relation from A to B is a subset of $A \times B$.

In other words, a binary relation R is a set of ordered pairs (a_i, b_i) where $a_i \in A$ and $b_i \in B$.

Notation: We say that

- $a\ R\ b$ if $(a,b) \in R$
- $a\ \not R\ b$ if $(a,b) \notin R$

Example: Course Enrollments

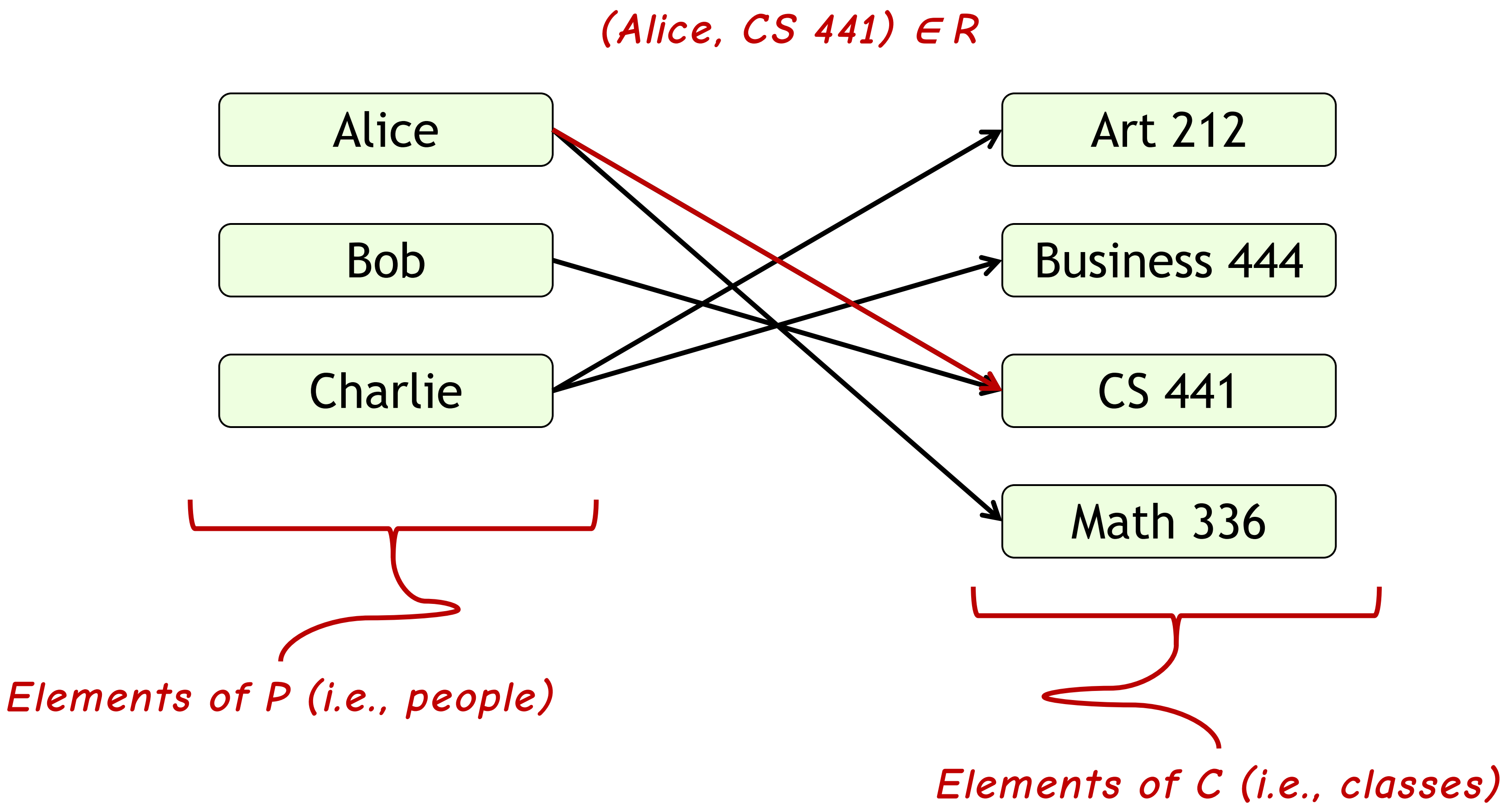
Let's say that Alice and Bob are taking CS 441. Alice is also taking Math 336. Furthermore, Charlie is taking Art 212 and Business 444. Define a relation R that represents the relationship between people and classes.

Solution:

- Let the set P denote people, so P = {Alice, Bob, Charlie}
- Let the set C denote classes, so C = {CS 441, Math 336, Art 212, Business 444}
- By definition $R \subseteq P \times C$
- From the above statement, we know that
  - (Alice, CS 441) ∈ R
  - (Bob, CS 441) ∈ R
  - (Alice, Math 336) ∈ R
  - (Charlie, Art 212) ∈ R
  - (Charlie, Business 444) ∈ R
- So, R = {(Alice, CS 441), (Bob, CS 441), (Alice, Math 336), (Charlie, Art 212), (Charlie, Business 444)}

A relation can also be represented as a graph

Let's say that Alice and Bob are taking CS 441. Alice is also taking Math 336. Furthermore, Charlie is taking Art 212 and Business 444. Define a relation R that represents the relationship between people and classes.



A relation can also be represented as a table

Let's say that Alice and Bob are taking CS 441. Alice is also taking Math 336. Furthermore, Charlie is taking Art 212 and Business 444. Define a relation R that represents the relationship between people and classes.

Name of the relation

(Bob, CS 441) ∈ R

Elements of C (i.e., courses)

R	Art 212	Business 444	CS 441	Math 336
Alice			X	X
Bob			X	
Charlie	X	X		

Elements of P (i.e., people)

We can also “relate” elements of more than two sets

Definition: Let $A_1, A_2, \ldots, A_n$ be sets. An n-ary relation on these sets is a subset of $A_1 \times A_2 \times \ldots \times A_n$. The sets $A_1, A_2, \ldots, A_n$ are called the domains of the relation, and n is its degree.

Example: Let R be the relation on $\mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}^+$ consisting of triples (a, b, m) where $a \equiv b \pmod{m}$.

- What is the degree of this relation?
- What are the domains of this relation?
- Are the following tuples in this relation?
  - (8,2,3)
  - (-1,9,5)
  - (11,0,6)

N-ary relations are the basis of relational database management systems

Data is stored in relations (a.k.a., tables)

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

Columns of a table represent the attributes of a relation

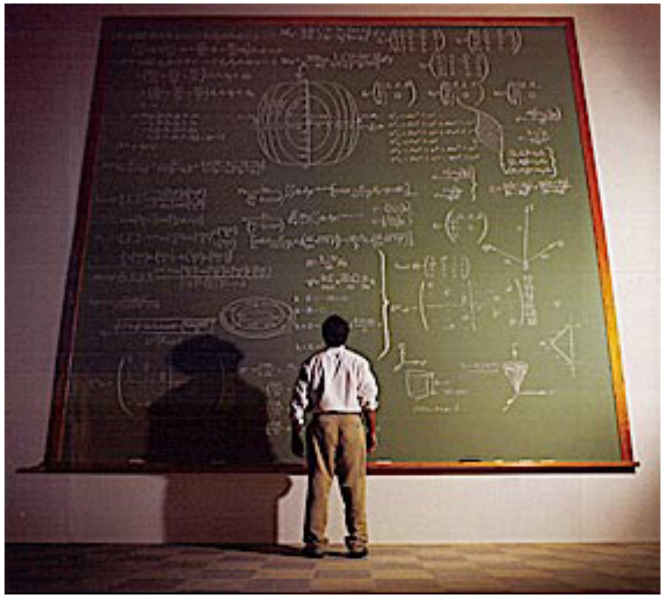
Rows, or records, contain the actual data defining the relation

Operations on an RDBMS are formally defined in terms of a relational algebra

Relational algebra gives a formal semantics to the operations performed on a database by rigorously defining these operations in terms of manipulations on sets of tuples (i.e., records)

Operators in relational algebra include:

- Selection ☆
- Projection ☆
- Rename
- Join
  - Equijoin ☆
  - Left outer join
  - Right outer join
  - …
- Aggregation



The selection operator allows us to filter the rows in a table

Definition: Let R be an n-ary relation and let C be a condition that elements in R must satisfy. The selection s_C maps the n-ary relation R to the n-ary relation of all n-tuples from R that satisfy the condition C.

Example: Consider the Students relation from earlier in lecture. Let the condition C1 be Major="CS" and let C2 be GPA > 2.5. What is the result of $s_{C1 \wedge C2}$(Students)?

Answer:

- (Alice, 334322, CS, 3.45)
- (Charlie, 045628, CS, 2.75)

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

The projection operator allows us to consider only a subset of the columns of a table

Definition: The projection $P_{i1,\ldots,im}$ maps the n-tuple $(a_1, a_2, \ldots, a_n)$ to the m-tuple $(a_{i1}, \ldots, a_{im})$ where $m \leq n$

Example: What is the result of applying the projection $P_{1,3}$ to the Students table?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Name	Major
Alice	CS
Bob	Math
Charlie	CS
Denise	Art

The equijoin operator allows us to create a new table based on data from two or more related tables

Definition: Let R be a relation of degree m and S be a relation of degree n. The equijoin $J_{i1=j1,\ldots,ik=jk}$, where $k \le m$ and $k \le n$, creates a new relation of degree m+n-k containing the subset of $S \times R$ in which $s_{i1} = r_{j1}, \ldots, s_{ik}=r_{jk}$ and duplicate columns are removed (via projection).

Example: What is the result of the equijoin $J_{2=1}$ on the Students and Enrollment tables?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

What is the result of the equijoin $J_{2=1}$ on the Students and Enrollment tables?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

Name	ID	Major	GPA	Course
Alice	334322	CS	3.45	CS 441
Alice	334322	CS	3.45	Math 336
Bob	546346	Math	3.23	Math 422
Denise	964389	Art	4.0	Art 707

SQL queries correspond to statements in relational algebra

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

$\Rightarrow$

Name	ID
Alice	334322
Charlie	045628

SELECT Name, ID FROM Students WHERE Major = "CS" AND GPA > 2.5

SELECT is actually a projection (in this case, $P_{1,2}$)

The WHERE clause lets us filter (i.e., $S_{major=\text{"CS"} \wedge GPA>2.5}$)

SQL: An Equijoin Example

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

```
SELECT Name, ID, Major, GPA, Course FROM Students, Enrollment
WHERE ID = Stud_ID
```

Name	ID	Major	GPA	Course
Alice	334322	CS	3.45	CS 441
Alice	334322	CS	3.45	Math 336
Bob	546346	Math	3.23	Math 422
Denise	964389	Art	4.0	Art 707

Why are n-ary relations and relational algebra interesting?

Reason 1: Formal representation of DB state

Reason 2: Efficient way to process SQL queries

- Parse/tokenize SQL query
- Compile into a tree of relational operators
- Optimize tree for efficient execution
- Execute plan and return results

Example

Assume table T has 1,000,000 tuples with two attributes, a and b. Ten of these tuples have a = 5, while the other 999,990 have a ≠ 5.

Query: `SELECT b FROM T WHERE a = 5`

- Parse 1: $S_{a=5}\ P_2\ T$
- Parse 2: $P_2\ S_{a=5}\ T$

In-class exercises

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment	
Stud_ID	Course
334322	CS 441
334322	Math 336
546346	Math 422
964389	Art 707

Problem 1: What is $P_{1,4}$(Students)?

Problem 2: What relational operators would you use to generate a table containing only the names of Math and CS majors with a GPA > 3.0?

Problem 3: Write an SQL statement corresponding to the solution to problem 2.

Final Thoughts

Relations allow us to represent and reason about the relationships between sets in a more general way than functions did

Without n-ary relations, DBMS systems would not exist!

Do you find this stuff interesting?

- CS 1555: Introduction to Database Systems
- CS 1571: Introduction to Artificial Intelligence

Next: Review and wrap up!