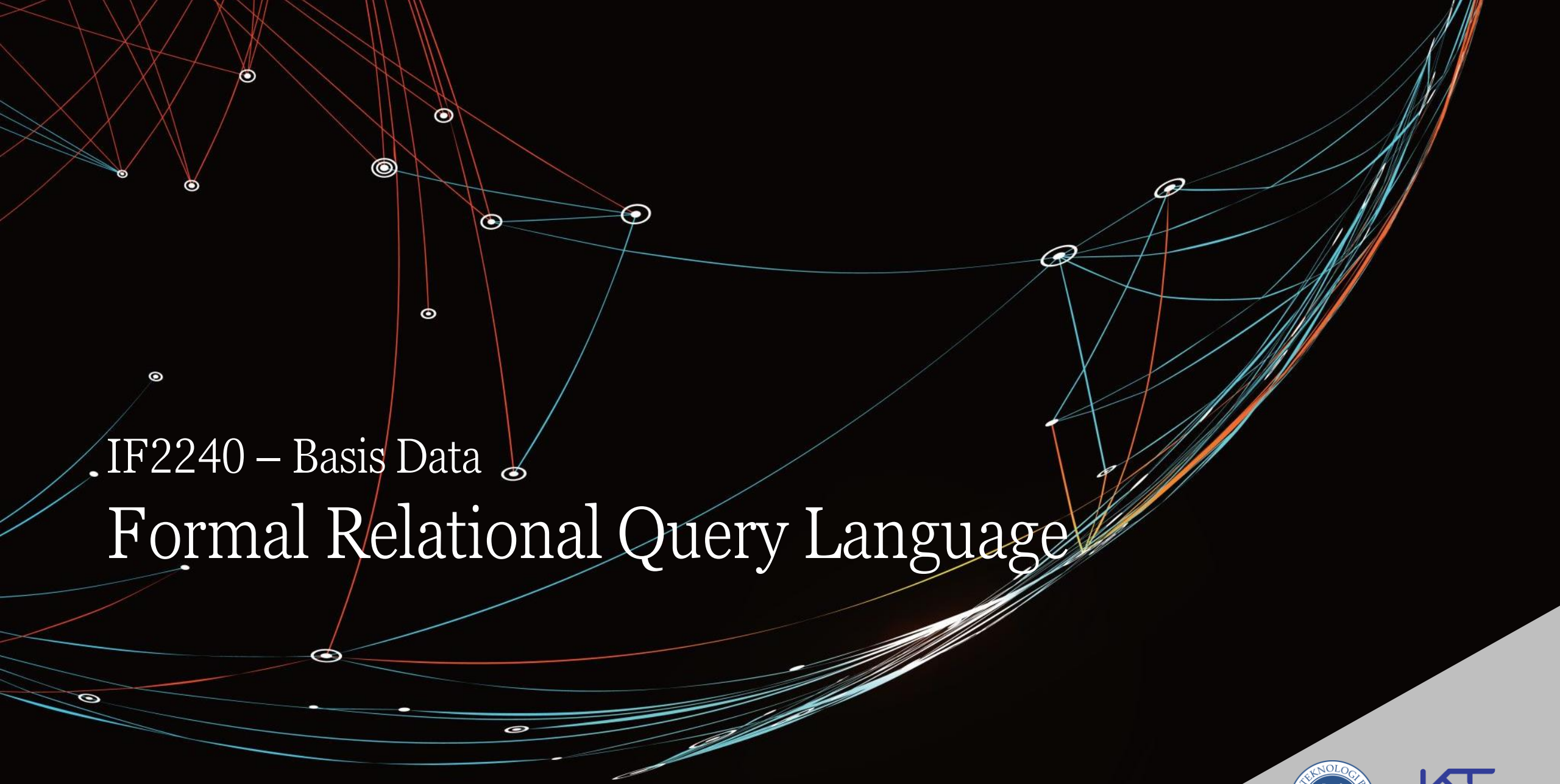
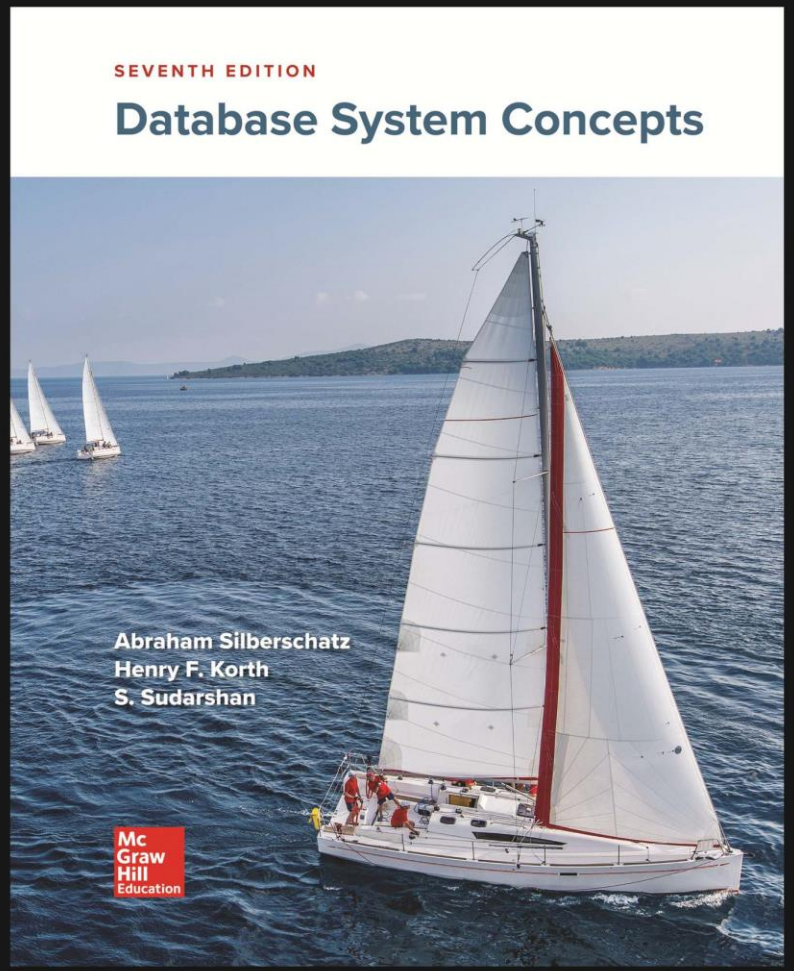


IF2240 – Basis Data

Formal Relational Query Language

KNOWLEDGE & SOFTWARE ENGINEERING



Sumber

Silberschatz, Korth, Sudarshan: "Database System Concepts", 7th Edition

- Chapter 2: Relational Model
  - Section 2.5. Relational Query Languages
  - Section 2.6. The Relational Algebra
  - Page 47 – 58
- Chapter 27: Formal-Relational Query Languages (*online chapter*)
  - Section 27.1. The Tuple Relational Calculus
  - Section 27.2. The Domain Relational Calculus

Objective

Demonstrate use of the relational algebra operations from mathematical set theory (union, intersection, difference, and Cartesian product) and the relational algebra operations developed specifically for relational databases (select (restrict), project, join, and division)

Demonstrate queries in the relational algebra

Demonstrate queries in the domain relational calculus

Demonstrate queries in the tuple relational calculus

Query Languages

Categories of languages

1

Functional/Procedural

- **Relational algebra**

HOW?

2

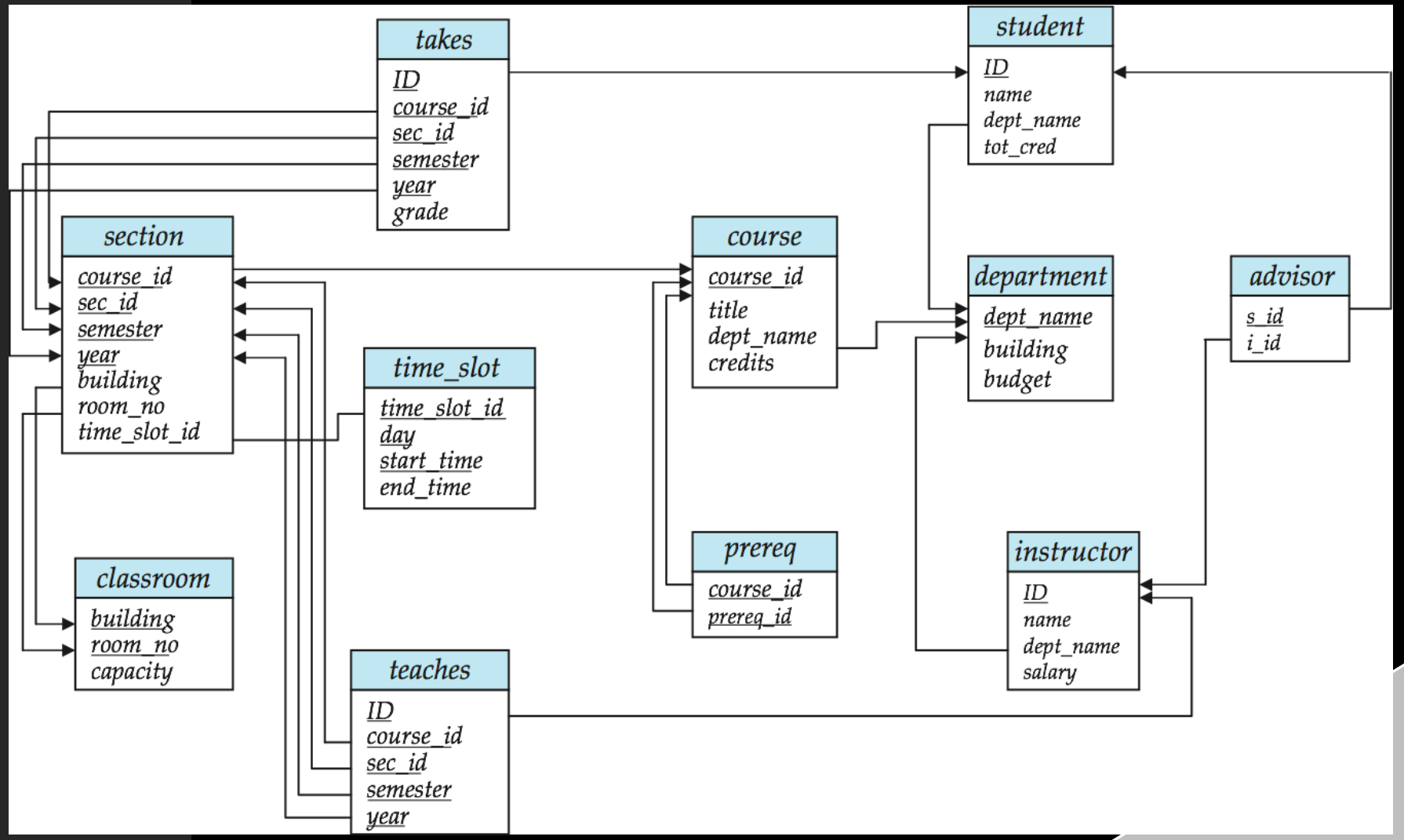
Non-procedural, or declarative

- Tuple relational calculus
- Domain relational calculus

WHAT?

Pure languages form underlying basis of query languages that people use

Schema Diagram for University Database



KNOWLEDGE & SOFTWARE ENGINEERING

RELATIONAL ALGEBRA

KNOWLEDGE & SOFTWARE ENGINEERING

Relational Algebra

Procedural language

Six basic operators

- select: σ
- project: Π
- union: $\cup$
- set difference: $-$
- cartessian product: x
- rename: ρ

The operators take one or two relations as inputs and produce a new relation as a result.

- Additional Operators
  - intersection
  - natural join
  - assignment
  - outer join
  - division
- Extended Operators
  - generalized projection
  - aggregation

Basic Operators

Select Operation

Notation

$$\sigma_p(r)$$

p = selection predicate

Defined as

$$\sigma_p(r) = \{t \mid t \in r \text{ and } p(t)\}$$

- Where p is a formula in propositional calculus: terms connected by $\wedge$ (**and**), $\vee$ (**or**), $\neg$ (**not**)
- Term = <attribute> *op* <attribute> OR <constant>

Select Operation – Example 1

INSTRUCTOR (ID, NAME, DEPT_NAME, SALARY)

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

$\sigma_{DEPT\_NAME= \text{“PHYSICS”}}(INSTRUCTOR)$

ID	name	dept_name	salary
22222	Einstein	Physics	95000
33456	Gold	Physics	87000

Select Operation – Example 2

RELATION R

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

$\sigma_{A = B \wedge D > 5}(R)$

A	B	C	D
α	α	1	7
β	β	23	10

Project Operation

Notation

$$\Pi_{A1, A2, \ldots, Ak}\ (r)$$

where A_1, A_2, … are attribute names and r is a relation name.

Defined as

- The relation of k columns obtained by erasing the columns that are not listed
- Duplicate rows removed from result, since relations are sets

Project Operation – Example 1

INSTRUCTOR (ID, NAME, DEPT_NAME, SALARY)

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

$\Pi_{ID,\ NAME,\ DEPT\_NAME}\ (INSTRUCTOR)$

ID	name	dept_name
22222	Einstein	Physics
12121	Wu	Finance
32343	El Said	History
45565	Katz	Comp. Sci.
98345	Kim	Elec. Eng.
76766	Crick	Biology
10101	Srinivasan	Comp. Sci.
58583	Califieri	History
83821	Brandt	Comp. Sci.
15151	Mozart	Music
33456	Gold	Physics
76543	Singh	Finance

Project Operation – Example 2

RELATION *R*:

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

$\Pi_{A,C}(R)$

A	C
α	1
α	1
β	1
β	2

A	C
α	1
β	1
β	2

Composition of Relational Operations

The result of a relational-algebra operation is a relation and therefore some of relational-algebra operations can be composed together into a **relational-algebra expression**.

Consider the query -- Find the names of all instructors in the Physics department.

$$\Pi_{name}(\sigma_{dept\_name = \text{“Physics”}}(instructor))$$

Union Operation

- Notation

$$r \cup s$$

- Defined as

$$r \cup s = \{t \mid t \in r \text{ or } t \in s\}$$

- Requirement
  1. r, s must have the *same* ***arity***
  2. The attribute domains must be **compatible**

Union Operation – Example 1

RELATIONS

R

A	B
α	1
α	2
β	1

S

A	B
α	2
β	3

R ∪ S:

A	B
α	1
α	2
β	1
β	3

Union Operation – Example 2

section
course_id
sec_id
semester
year
building
room_no
time_slot_id

Find all courses taught in the Fall 2017 semester, or in the Spring 2018 semester, or in both

$$\Pi_{course\_id}\ (\sigma_{semester=\text{“Fall”}\ \wedge\ year=2017}\ (section))\ \cup$$
$$\Pi_{course\_id}\ (\sigma_{semester=\text{“Spring”}\ \wedge\ year=2018}\ (section))$$

Set Difference Operation

- Notation

$$r - s$$

- Defined as

$$r - s = \{t \mid t \in r \text{ and } t \notin s\}$$

- Requirement

  Set differences must be taken between *compatible* relations

Set Difference Operation – Example 1

RELATIONS

R

A	B
α	1
α	2
β	1

S

A	B
α	2
β	3

R - S:

A	B
α	1
β	1

Set Difference Operation – Example 2

section
course_id
sec_id
semester
year
building
room_no
time_slot_id

Find all courses taught in the Fall 2017 semester, but not in the Spring 2018 semester

$$\Pi_{course\_id}\,(\sigma_{semester=\text{“Fall”}\ \wedge\ year=2017}\,(section)) - \Pi_{course\_id}\,(\sigma_{semester=\text{“Spring”}\ \wedge\ year=2018}\,(section))$$

Cartesian-Product Operation

Notation

$$r \times s$$

Defined as

$$r \times s = \{t\ q \mid t \in r \text{ and } q \in s\}$$

Requirement

- Attributes of r(R) and s(S) are disjoint
- If attributes of *r(R)* and *s(S)* are not disjoint, then:
  renaming OR attach its source relation

instructor.ID

Cartesian-Product Operation – Example 1

RELATIONS

R

A	B
α	1
β	2

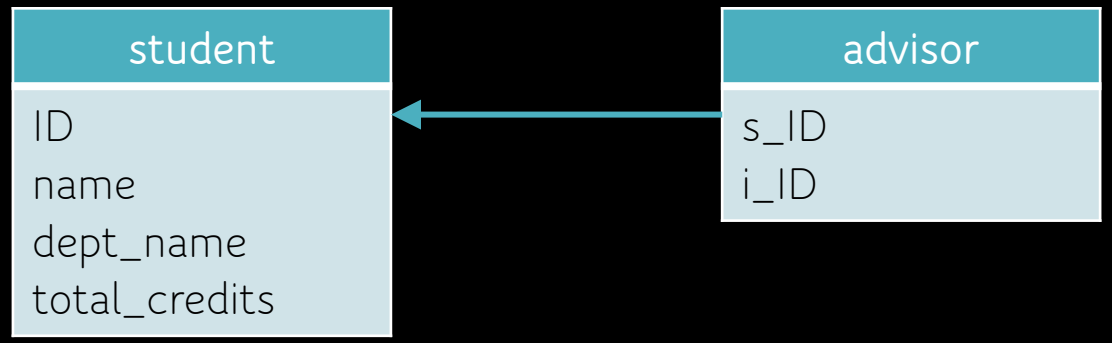
S

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

R X S:

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

Cartesian-Product Operation – Example 2



Find the names of all students whose advisor's id is 22222

$\Pi_{name}\ (\sigma_{i\_ID=22222}\ (\sigma_{ID=s\_ID}\ (student\ x\ advisor)))$

Rename Operation

- Usage
  - To name the results of relational-algebra expressions.
  - To refer to a relation by more than one name
- Notation (relation)

$$\rho_{X}(E)$$

returns the result of expression E under the name X

- Notation (attributes)

$$\rho_{X(A1, A2, \ldots, An)}(E)$$

returns the result of expression E under the name X, and with the attributes renamed to $A1, A2, \ldots, An$.

Rename Operation - Example

instructor
ID
name
dept_name
salary

Find the ID and name of those instructors who earn more than the instructor whose ID is 12121

$$\Pi_{z.ID,\ z.name} \left(\sigma_{z.salary > w.salary} \left(\rho_z (instructor) \times \left(\sigma_{ID=12121} (\rho_w (instructor)) \right) \right) \right)$$

Formal Definition

A basic expression in the relational algebra consists of either one of the following:

- A relation in the database
- A constant relation

Let E_1 and E_2 be relational-algebra expressions; the following are all relational-algebra expressions:

- $E_1 \cup E_2$
- $E_1 - E_2$
- $E_1 \times E_2$
- $\sigma_P (E_1)$, P is a predicate on attributes in E_1
- $\prod_S (E_1)$, S is a list consisting of some of the attributes in E_1
- $\rho_x (E_1)$, x is the new name for the result of E_1

Example Queries

Find the names of all instructors in the Physics department, along with the *course_id* of all courses they have taught

- Query 1

$$\Pi_{instructor.name,course\_id}\ (\sigma_{dept\_name=\text{“Physics”}}\ (\sigma_{instructor.ID=teaches.ID}\ (instructor \times teaches)))$$

- Query 2

$$\Pi_{instructor.name,course\_id}\ (\sigma_{instructor.ID=teaches.ID}\ (\sigma_{dept\_name=\text{“Physics”}}\ (instructor) \times teaches))$$

Example Queries

Find all instructors with salary over $80000

Find the names of all instructors in Comp. Sci. dept with salary > 80000

Find the names of all person in the university

schema

Example Queries

Find the title of all courses offered in Fall 2017 semester.

Find the title of all courses offered in Fall 2017 semester that do not have any prerequsites.

schema

Example Queries

Find the names of all instructors who are located at Archway Bldg.

schema

Example Query

Find the largest salary in the university

- Step 1: find instructor salaries that are less than some other instructor salary (i.e. not maximum)
  - using a copy of *instructor* under a new name *d*

- Step 2: Find the largest salary

schema