RELATIONAL ALGEBRA

CS-A1153 - Databases (Summer 2020)

LUKAS AHRENBERG

Aalto University
School of Science

WHYYYYYYYYYYYY?

- To *reason* about relations
- To construct new relations from given ones
- To analyze given relations
- And so, to *plan*, *analyze*, and *reason* about efficient database systems (for fun/science/profit)

[U&W (New Int. Ed.) - 2:2 & 2:4]

ALGEBRA?

- In school we learned about *arithmetic* algebra
  - Operators: $+$, $-$, $\times$, $/$
  - Operands: variables (e.g. x), and constants (e.g. 14)
- *An algebra* in a broader sense, is a mathematical structure of operators and operands
  - We will be working with relations instead of numbers
  - Operations such as *union*, *selection* and *projection*

RELATIONS - REPETITION

- A relation can be thought of as a table where
  - The **columns** are **attributes**
  - The **rows** are data **tuples**
- So, the structure is a **set of tuples**
- Described by a **schema** giving the name and the column order

For example: `Booker(author,title,year)`

author	title	year
Kazuo Ishiguro	The Remains of the Day	1989
Ben Okri	The Famished Road	1991
Margaret Atwood	The Blind Assassin	2000

SET OPERATIONS - REVIEW

$$A = \{1, 2, 5\}, B = \{3, 4, 5\}$$

Union

$$A \cup B = \{1, 2, 3, 4, 5\}$$

Intersection

$$A \cap B = \{5\}$$

Difference

$$A - B = \{1, 2\}$$

Note: $A - B \neq B - A$

SET OPERATIONS ON RELATIONS

Union, Intersection, and Difference works the same way for Relations, if

- R, S has schemas with identical sets of attributes and domains
- The attributes **must** have the same order

UNION EXAMPLE

R

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd, Brentwood	M	8/8/88

S

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99
Harrison Ford	789 Palm Dr, Beverly Hills	M	7/7/77

R ∪ S

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd, Brentwood	M	8/8/88
Harrison Ford	789 Palm Dr, Beverly Hills	M	7/7/77

INTERSECTION EXAMPLE

R

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd, Brentwood	M	8/8/88

S

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99
Harrison Ford	789 Palm Dr, Beverly Hills	M	7/7/77

R ∩ S

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99

SET DIFFERENCE EXAMPLE

R

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd, Brentwood	M	8/8/88

S

name	address	gender	birthdate
Carrie Fisher	123 Maple St, Hollywood	F	9/9/99
Harrison Ford	789 Palm Dr, Beverly Hills	M	7/7/77

R − S

name	address	gender	birthdate
Mark Hamill	456 Oak Rd, Brentwood	M	8/8/88

SELECTION – σ

- Produces a new relation by selecting a subset of the tuples (rows) from an existing relation
- $\sigma_C\ (\mathtt{R})$, where C is the condition for which tuples of $\mathtt{R}$ to select
- C is the conditional expression (think `if C` in many programming languages) for which tuples should be included in the new relation

SELECTION – σ : EXAMPLE 1

Movies

title	year	length	genre	studioName
Star Wars	1977	124	sciFi	Fox
Galaxy Quest	1999	104	comedy	DreamWorks
Wayne's World	1992	95	comedy	Paramount

$\sigma_{length \geq 100}$ (**Movies**)

title	year	length	genre	studioName
Star Wars	1977	124	sciFi	Fox
Galaxy Quest	1999	104	comedy	DreamWorks

(Movies with a running time of at least 100 minutes.)

SELECTION – σ : EXAMPLE 2

Movies

title	year	length	genre	studioName
Star Wars	1977	124	sciFi	Fox
Galaxy Quest	1999	104	comedy	DreamWorks
Wayne's World	1992	95	comedy	Paramount

$$\sigma_{length \geq 100 \ AND \ studioName = 'Fox'} (\text{Movies})$$

title	year	length	genre	studioName
Star Wars	1977	124	sciFi	Fox

(Movies with a running time of at least 100 minutes and made by Fox.)

PROJECTION – π

- Creates a new relation by using a subset of an existing relation's attributes.
- $\pi_{A_1, A_2, \ldots, A_n}\ (\mathbf{R})$, where $\mathbf{R}$ is a relation, and $A_1, \ldots, A_n$ are some of its attributes.

PROJECTIONS – π : EXAMPLE 1

Movies

title	year	length	genre	studioName
Star Wars	1977	124	sciFi	Fox
Galaxy Quest	1999	104	comedy	DreamWorks
Wayne's World	1992	95	comedy	Paramount

$\pi_{title,year,genre}$ (**Movies**)

title	year	genre
Star Wars	1977	sciFi
Galaxy Quest	1999	comedy
Wayne's World	1992	comedy

CARTESIAN PRODUCT – $\times$

- Creates a new relation as all ordered pairings of tuples from two relations.
- $\mathtt{R} \times \mathtt{S}$

CARTESIAN PRODUCT – $\times$: EXAMPLE 1

`Titles`

author	title
Ben Okri	The Famished Road
Margaret Atwood	The Blind Assassin

`Stock`

title	product	copies
The Famished Road	hardback	2
The Blind Assassin	hardback	1
The Blind Assassin	pocket	7

`Titles` $\times$ `Stock`

author	Titles.title	Stock.title	product	copies
Ben Okri	The Famished Road	The Famished Road	hardback	2
Ben Okri	The Famished Road	The Blind Assassin	hardback	1
Ben Okri	The Famished Road	The Blind Assassin	pocket	7
Margaret Atwood	The Blind Assassin	The Famished Road	hardback	2
Margaret Atwood	The Blind Assassin	The Blind Assassin	hardback	1
Margaret Atwood	The Blind Assassin	The Blind Assassin	pocket	7

(Combine **all** rows of `Titles` with **all** rows `Stock`.)

CARTESIAN PRODUCT – $\times$: EXAMPLE 2

$$\texttt{Titles} \times \texttt{Stock}$$

author	Titles.title	Stock.title	product	copies
Ben Okri	The Famished Road	The Famished Road	hardback	2
Ben Okri	The Famished Road	The Blind Assassin	hardback	1
Ben Okri	The Famished Road	The Blind Assassin	pocket	7
Margaret Atwood	The Blind Assassin	The Famished Road	hardback	2
Margaret Atwood	The Blind Assassin	The Blind Assassin	hardback	1
Margaret Atwood	The Blind Assassin	The Blind Assassin	pocket	7

$$\sigma_{Titles.title=Stock.title}\ (\texttt{Titles} \times \texttt{Stock})$$

author	Titles.title	Stock.title	product	copies
Ben Okri	The Famished Road	The Famished Road	hardback	2
Margaret Atwood	The Blind Assassin	The Blind Assassin	hardback	1
Margaret Atwood	The Blind Assassin	The Blind Assassin	pocket	7

(Combine **all** rows of `Titles` with **all** rows `Stock` and then select those where the attribute **title** agrees.)

CARTESIAN PRODUCT – $\times$: EXAMPLE 3

$$\texttt{Titles} \times \texttt{Stock}$$

author	Titles.title	Stock.title	product	copies
Ben Okri	The Famished Road	The Famished Road	hardback	2
Ben Okri	The Famished Road	The Blind Assassin	hardback	1
Ben Okri	The Famished Road	The Blind Assassin	pocket	7
Margaret Atwood	The Blind Assassin	The Famished Road	hardback	2
Margaret Atwood	The Blind Assassin	The Blind Assassin	hardback	1
Margaret Atwood	The Blind Assassin	The Blind Assassin	pocket	7

$$\pi_{author,Titles.title,product,copies}\left(\sigma_{Titles.title=Stock.title}\left(\texttt{Titles} \times \texttt{Stock}\right)\right)$$

author	Titles.title	product	copies
Ben Okri	The Famished Road	hardback	2
Margaret Atwood	The Blind Assassin	hardback	1
Margaret Atwood	The Blind Assassin	pocket	7

(Combine **all** rows of `Titles` with **all** rows `Stock`; select those where the attribute **title** agrees; finally project to get rid of one of the **title** columns.)

NATURAL JOIN – ⋈

- Creates a new relation by joining two relations on their common attributes, leaving out anything else.
- $\mathtt{R} \bowtie \mathtt{S}$, where $\mathtt{R}$,$\mathtt{S}$ are relations sharing at least one attribute.

NATURAL JOIN – ⋈ : EXAMPLE 1

Booker

author	title	year
Kazuo Ishiguro	The Remains of the Day	1989
Ben Okri	The Famished Road	1991
Margaret Atwood	The Blind Assassin	2000

Nationality

author	nationality
Ben Okri	Nigerian
Kazuo Ishiguro	British
Margaret Atwood	Canadian
Ursula K. Le Guin	U.S.

Booker ⋈ Nationality

author	title	year	nationality
Kazuo Ishiguro	The Remains of the Day	1989	British
Ben Okri	The Famished Road	1991	Nigerian
Margaret Atwood	The Blind Assassin	2000	Canadian

NATURAL JOIN – ⋈ : GOTCHA

Booker

author	title	year
Kazuo Ishiguro	The Remains of the Day	1989
Ben Okri	The Famished Road	1991
Margaret Atwood	The Blind Assassin	2000

Nobel

author	year	nationality
Nadine Gordimer	1991	South African
Günter Grass	1999	German
Kazuo Ishiguro	2017	British

Booker ⋈ Nobel

$$\emptyset$$

*A natural join works over **all** common attributes.*

THETA-JOIN – $\bowtie_C$

- Creates a new relation by joining two other relations under some condition
- $\mathbf{R} \bowtie_C \mathbf{S}$, where $\mathbf{R}$,$\mathbf{S}$ are relations, and C is the conditions under which tuples are joined
- C is a conditional expression of attributes in the two relations
- Product + Filtering

THETA-JOIN – $\bowtie_C$: EXAMPLE 1

Booker

author	title	year
Kazuo Ishiguro	The Remains of the Day	1989
Ben Okri	The Famished Road	1991
Margaret Atwood	The Blind Assassin	2000

Nobel

author	year	nationality
Nadine Gordimer	1991	South African
Günter Grass	1999	German
Kazuo Ishiguro	2017	British

Booker $\bowtie_{Booker.author=Nobel.author}$ Nobel

Booker.author	title	Booker.year	Nobel.author	Nobel.year	nationality
Kazuo Ishiguro	The Remains of the Day	1989	Kazuro Ishiguro	2017	British

(Get table of authors with both a Booker and Nobel prizes by a theta-join on the author name.)

THETA-JOIN – $\bowtie_c$: EXAMPLE 2

St

title	product	copies
The Famished Road	hardback	2
The Blind Assassin	hardback	1
The Blind Assassin	pocket	7

Wh

title	product	copies	shelf
The Famished Road	hardback	10	I11
The Blind Assassin	hardback	0	A13
The Blind Assassin	pocket	22	A13

St $\bowtie_{St.title=Wh.title\ AND\ St.product=Wh.product\ AND\ St.copies<5\ AND\ Wh.copies>0}$ **Wh**

St.title	St.product	St.copies	Wh.title	Wh.product	Wh.copies	Wh.shelf
The Famished Road	hardback	2	The Famished Road	hardback	10	I11

(Joining Store and Warehouse on title and product, where store copies are less than five, and warehouse copies more than zero.)

RENAME – ρ

- Names or renames a relation or its attributes
- ρ_{sc}, where sc is a new naming E.g. $\rho_{Films}\ (\texttt{Movies})$
- Useful
  - Many other operations pair attributes by name
  - When combining a relation with itself

RENAME – ρ : EXAMPLE 1

Movies

title	year	length	genre	studioName
Star Wars	1977	124	sciFi	Fox
Galaxy Quest	1999	104	comedy	DreamWorks
Wayne's World	1992	95	comedy	Paramount

$\rho_{Films(flick,AD,minutes,kind,st)}$ (**Movies**)

flick	AD	minutes	kind	st
Star Wars	1977	124	sciFi	Fox
Galaxy Quest	1999	104	comedy	DreamWorks
Wayne's World	1992	95	comedy	Paramount

(Rename **Movies** and all of its attributes.)

RENAME – ρ : EXAMPLE 2

Credit

name	credit	production
Kip Thorne	producer	Interstellar
Michael Caine	actor	Inception
Alice Munroe	writer	Julieta

Nobels

laureate	category	year
Alice Munroe	Literature	2013
Bengt R. Holmström	Economics	2016
Kip Thorne	Physics	2017

$$\texttt{Credit} \bowtie \rho_{N(name,category,year)} (\texttt{Nobels})$$

name	credit	production	category	year
Alice Munroe	writer	Julieta	Literature	2013
Kip Thorne	producer	Interstellar	Physics	2017

(Get people with both movie credits and Nobel prizes by first renaming **laureate** to **name** in **Nobels** and performing a natural join with **Credit**.)

RENAME – ρ : EXAMPLE 3 - RENAME WHEN JOINING A TABLE WITH ITSELF

Booker

author	title	year
Kazuo Ishiguro	The Remains of the Day	1989
Ben Okri	The Famished Road	1991
Margaret Atwood	The Blind Assassin	2000

$$(\rho_{B1(author,title,year)}\ (\texttt{Booker})) \bowtie_{B1.year<B2.year}\ (\rho_{B2(author,title,year)}\ (\texttt{Booker}))$$

B1.author	B1.title	B1.year	B2.author	B2.title	B2.year
Kazuo Ishiguro	The Remains of the Day	1989	Ben Okri	The Famished Road	1991
Kazuo Ishiguro	The Remains of the Day	1989	Margaret Atwood	The Blind Assassin	2000
Ben Okri	The Famished Road	1991	Margaret Atwood	The Blind Assassin	2000

(Pair authors with those who got the price earlier; Achieved by theta-join after renaming the relation so to be able to refer to the attribute **year** in either.)

LINEAR NOTATION AND THE ASSIGNMENT SYMBOL :=

- Nested expressions of relational algebra can get quite hairy
- In 'normal' algebra we are used to do things in multiple steps
- := can be used to assign a calculation on the right hand side to some temporary relation on the left

LINEAR NOTATION EXAMPLE 1

Booker

author	title	year
Kazuo Ishiguro	The Remains of the Day	1989
Ben Okri	The Famished Road	1991
Margaret Atwood	The Blind Assassin	2000

Nobel

author	year	nationality
Nadine Gordimer	1991	South African
Günter Grass	1999	German
Kazuo Ishiguro	2017	British

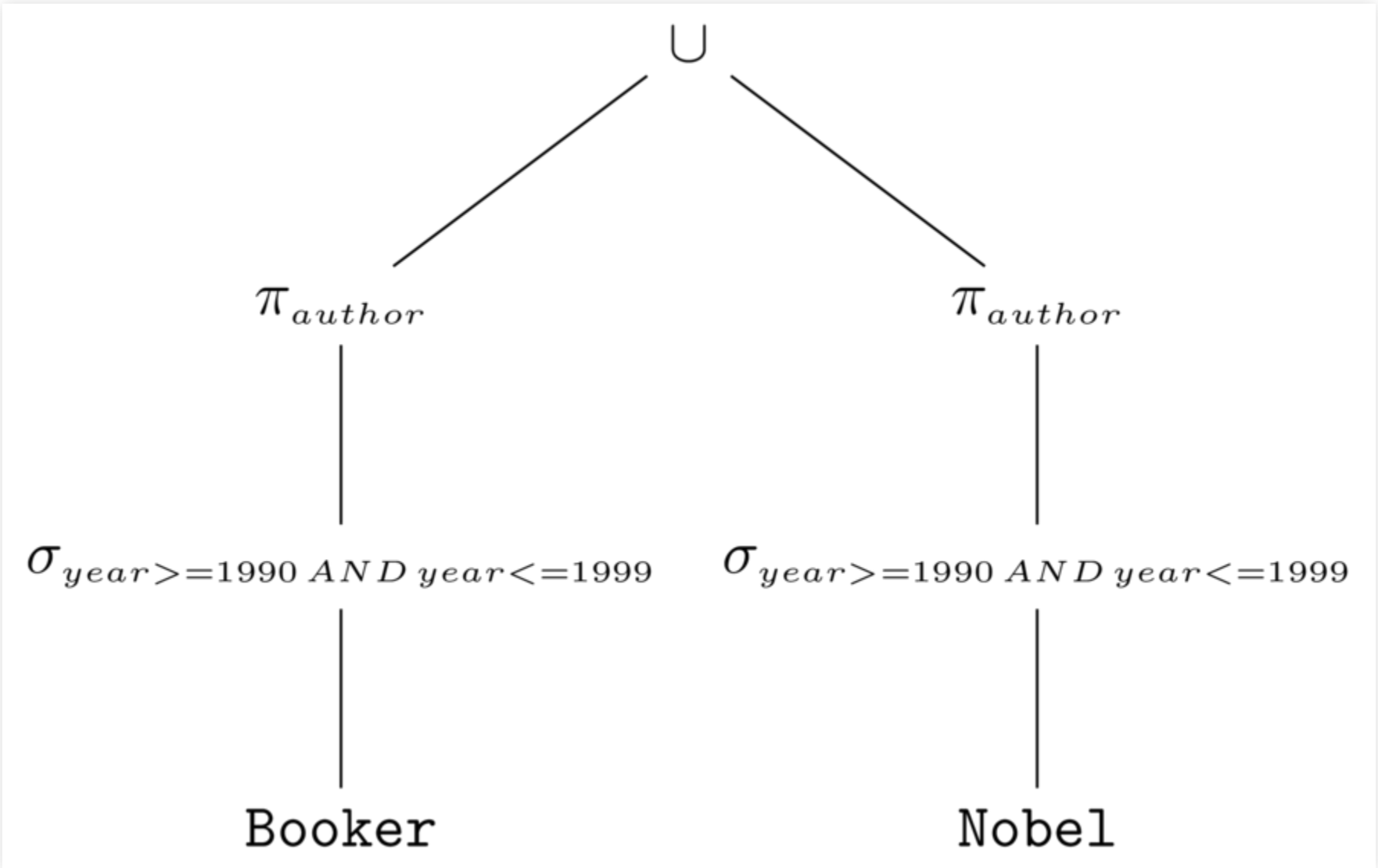
$$\pi_{author}\left(\sigma_{year>=1990\ AND\ year<=1999}(\texttt{Booker})\right) \cup \pi_{author}\left(\sigma_{year>=1990\ AND\ year<=1999}(\texttt{Nobel})\right)$$

Or on linear form:

$$\texttt{R(a,t,y)} := \sigma_{year>=1990\ AND\ year<=1999}(\texttt{Booker})$$
$$\texttt{S(a,y,n)} := \sigma_{year>=1990\ AND\ year<=1999}(\texttt{Nobel})$$
$$\texttt{U(a)} := \pi_{a}\left(\texttt{R}\right)$$
$$\texttt{V(a)} := \pi_{a}\left(\texttt{S}\right)$$
$$\texttt{Answer(author)} := \texttt{U} \cup \texttt{V}$$

EXPRESSION TREE NOTATION

$$\pi_{author}\ (\sigma_{year>=1990\ AND\ year<=1999}(\texttt{Booker})) \cup \pi_{author}\ (\sigma_{year>=1990\ AND\ year<=1999}(\texttt{Nobel}))$$



COURSE SPECIFICS

- [U&W (New Int. Ed.) - 2:2 & 2:4]
- Go to A+ Exercise 1.1 where you can practice the relational algebra notation!
  - Read the instructions
  - There's also an instruction video on panopto
- Use the direct notation on A+, it doesn't support $:=$ and thus the linear notation
- On A+, if using the same relation twice in an expression (for instance when joining it with itself) you may get an error if you rename (ρ) only one of the sides. Solution: rename both.

A+ RENAME EXAMPLE

$\texttt{Students(ID,name,program,year)}$

- Encoding this will yield an error:

$$\texttt{Students} \bowtie_{year<S2.year} (\rho_{S2(ID,name,program,year)} (Students))$$

- Rename on both sides instead:

$$\rho_{S1(ID,name,program,year)} (\texttt{Students}) \bowtie_{S1.year<S2.year} (\rho_{S2(ID,name,program,year)} (\texttt{Students}))$$

(Join every student table with itself, matching up entries with those of students those who started studying before.)