



Algebra relacji

Relacyjne języki zapytań

- » Model relacyjny – zbiór tabel do reprezentacji danych i zależności między nimi
 - opisuje dane na poziomie logicznym i widoku.
- » Języki zapytań
 - w jaki sposób użytkownicy określają żądania pobierania i aktualizowania danych
- » Proceduralne vs nieprocedurale, lub deklaratywne
- » “Czyste” języki:
 - Algebra relacji
 - Wnioskowane na krotkach (tuple relational calculus)
 - Wnioskowanie dziedzinowe (domain relational calculus)

Język proceduralny składający się z zestawu operacji, które przyjmują jedną (unarne) lub dwie (binarne) relacje jako dane wejściowe, a ich wynikiem jest nowa relacja.

- Sześć podstawowych operatorów:
 1. **selekcja** (*select*): σ
 2. **projekcja** (*project*): Π
 3. **unia** (*union*): \cup
 4. **różnica zbiorów** (*set difference*): $-$
 5. **produkt kartezjański** (*Cartesian product*): \times
 6. **przemianowanie** (*assignment*): ρ
- Dodatkowe operatory:
 - **przecięcie zbiorów** (*set intersection*): \cap
 - **złączenie naturalne** (*natural join*): \bowtie
 - **przypisanie** (*assignment*): \leftarrow

Operacja selekcji

- » Operacja **selekcji** wybiera krotki spełniające dany predykat.
- » Notacja: $\sigma_p(r)$
- » p zwane **predykatem selekcji**
- » Np : wybierz te krotki relacji *instructor* gdzie instruktor pracuje na wydziale "Physics".
 - Zapytanie

$\sigma_{dept_name="Physics"}(instructor)$

– Wynik

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>salary</i>
22222	Einstein	Physics	95000
33456	Gold	Physics	87000

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



AGH

Operacja selekcji c.d.

Dozwolone korzystanie z porównań

$=, \neq, >, \geq, <, \leq$

w predykanie selekcji.

- » Można łączyć kilka predykatów w większy predykat za pomocą łączników :

\wedge (**and**), \vee (**or**), \neg (**not**)

- » Np : Znajdź instruktorów z Physics zarabiających więcej niż \$90,000:

$\sigma_{dept_name="Physics" \wedge salary > 90,000}(instructor)$

- » Predykat selekcji może zawierać porównania między dwoma atrybutami.
 - Np znajdź wszystkie wydziały, których nazwa jest taka sama jak nazwa ich budynku:
 - $\sigma_{dept_name=building}(department)$

Operacja projekcji/rzutowania

- » Operator jednoargumentowy, który zwraca relację będącą argumentem pomijając pewne atrybuty.
- » Notacja:

$$\Pi_{A_1, A_2, A_3 \dots A_k}(r)$$

gdzie A_1, A_2 – nazwy atrybutów i r - nazwa relacji.

- » Wynik: relacja z k kolumnami uzyskana w wyniku usunięcia tych niewymienionych
- » Zduplikowane wiersze usunięte z wynikowej relacji, bo relacje są zbiorami

Operacja projekcji c.d.

- » Np : eliminacja atrybutu *dept_name* z *instructor*
- » Zapytanie:

$$\Pi_{ID, name, salary} (instructor)$$

- » Wynik:

<i>ID</i>	<i>name</i>	<i>salary</i>
10101	Srinivasan	65000
12121	Wu	90000
15151	Mozart	40000
22222	Einstein	95000
32343	El Said	60000
33456	Gold	87000
45565	Katz	75000
58583	Califieri	62000
76543	Singh	80000
76766	Crick	72000
83821	Brandt	92000
98345	Kim	80000

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

Złożenie operacji relacyjnych

- » Wynikiem operacji algebry relacji jest relacja i dlatego operacje algebry relacji można łączyć w **wyrażenie algebry relacji**.
- » Np. zapytanie – Znajdź nazwiska wszystkich instruktorów na wydziale fizyki.

$$\Pi_{name}(\sigma_{dept_name = "Physics"} (instructor))$$

- » Zamiast podawać nazwę relacji jako argument operacji projekcji, można podać wyrażenie, które zwraca relację.

Operacja unii

- » Operator unii pozwala łączyć dwie relacje
- » Notacja: $r \cup s$
- » Aby $r \cup s$ było poprawne.
 1. r, s muszą mieć ten sam **stopień** (tę samą liczbę atrybutów)
 2. Dziedziny atrybutów muszą być **kompatybilne** (np: druga kolumna r ma ten sam typ wartości co druga kolumna s)
- » Np.: znajdź wszystkie kursy prowadzone w semestrze jesiennym 2017 lub wiosennym 2018 lub w obydwu

$\Pi_{course_id} (\sigma_{semester="Fall" \wedge year=2017}(section)) \cup$

$\Pi_{course_id} (\sigma_{semester="Spring" \wedge year=2018}(section))$

Operacja unii c.d.

» Wynik:

$$\Pi_{course_id} (\sigma_{semester="Fall" \wedge year=2017}(section)) \cup$$

$$\Pi_{course_id} (\sigma_{semester="Spring" \wedge year=2018}(section))$$

<i>course_id</i>
CS-101
CS-315
CS-319
CS-347
FIN-201
HIS-351
MU-199
PHY-101

<i>course_id</i>	<i>sec_id</i>	<i>semester</i>	<i>year</i>	<i>building</i>	<i>room_number</i>	<i>time_slot_id</i>
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

Operacja różnicy zbiorów

- » Pozwala znaleźć krotki, które występują jednej relacji ale nie w drugiej.
- » Notacja $r - s$
- » Różnica zbiorów musi zachodzić między **kompatybilnymi** relacjami.
 - r i s muszą mieć ten sam stopień
 - dziedziny atrybutów r i s muszą być kompatybilne
- » Np : aby znaleźć wszystkie kursy prowadzone w semestrze jesiennym 2017 ale nie w wiosennym 2018

$$\Pi_{course_id}(\sigma_{semester="Fall" \wedge year=2017}(section)) - \Pi_{course_id}(\sigma_{semester="Spring" \wedge year=2018}(section))$$

<i>course_id</i>
CS-347
PHY-101

Operacja iloczynu kartezjańskiego

- » Operator iloczynu kartezjańskiego (oznaczony przez X) pozwala łączyć informację z dowolnych dwóch relacji.
- » Np: Iloczyn kartezjański relacji *instructor* i *teaches*:
$$\textit{instructor} \times \textit{teaches}$$
- » Krotka wynikowa tworzona z każdej możliwej pary krotek: jedna z relacji *instructor* i jedna z relacji *teaches*
- » Ponieważ *ID* instruktora występuje w obydwu relacjach, rozróżnia się te atrybuty dołączając do atrybutu nazwę relacji, z której pierwotnie pochodził atrybut.
 - *instructor.ID*
 - *teaches.ID*

Iloczyn kartezjański

instructor

teaches

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009

Inst.ID	name	dept_name	salary	teaches.ID	course_id	sec_id	semester	year
10101	Srinivasan	Comp. Sci.	65000	10101	CS-101	1	Fall	2009
10101	Srinivasan	Comp. Sci.	65000	10101	CS-315	1	Spring	2010
10101	Srinivasan	Comp. Sci.	65000	10101	CS-347	1	Fall	2009
10101	Srinivasan	Comp. Sci.	65000	12121	FIN-201	1	Spring	2010
10101	Srinivasan	Comp. Sci.	65000	15151	MU-199	1	Spring	2010
10101	Srinivasan	Comp. Sci.	65000	22222	PHY-101	1	Fall	2009
...
...
12121	Wu	Finance	90000	10101	CS-101	1	Fall	2009
12121	Wu	Finance	90000	10101	CS-315	1	Spring	2010
12121	Wu	Pinance	90000	10101	CS-347	1	Fall	2009
12121	Wu	Pinance	90000	12121	FIN-201	1	Spring	2010
12121	Wu	Finance	90000	15151	MU-199	1	Spring	2010
12121	Wu	Pinance	90000	22222	PHY-101	1	Fall	2009
...
...

Operacja przemianowania

- » Wyniki wyrażeń algebry relacji nie mają nazwy, do której można by się było odwoływać. Do tego służy operator przemianowania ρ
- » Wyrażenie:

$$\rho_x(E)$$

zwraca wynik wyrażenia E pod nazwą x

- » Jeżeli wyrażenie algebry relacji E ma licznosc n , to wtedy :

$$\rho_{x(A_1, A_2, \dots, A_n)}(E)$$

zwraca wynik wyrażenia E pod nazwą X z atrybutami przemianowanymi na A_1, A_2, \dots, A_n .

Operacja przemianowania - przykład

» Znajdź najwyższą pensję

(1) Oblicz relację tymczasową składającą się z tych wynagrodzeń, które nie są największe:

$\Pi_{instructor.salary}(\sigma_{instructor.salary < d.salary}(instructor \times \rho_d(instructor)))$

(2) Znajdź różnicę między $\Pi_{salary}(instructor)$ a tą relacją tymczasową

$\Pi_{salary}(instructor) - \Pi_{instructor.salary}(\sigma_{instructor.salary < d.salary}(instructor \times \rho_d(instructor)))$

salary
95000

salary
65000
90000
40000
60000
87000
75000
62000
72000
80000
92000

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

Formalna definicja wyrażenia algebry relacji

- » Podstawowe wyrażenie w algebrze relacji składa się z:
 - relacji w bazie danych albo
 - stałej relacja (np. $\{(22222, \text{Einstein}, \text{Physics}, 95000), (76543, \text{Singh}, \text{Finance}, 80000)\}$)
- » Niech E_1 i E_2 będą wyrażeniami algebry relacji; to poniższe są również wyrażeniami algebry relacji:
 - $E_1 \cup E_2$
 - $E_1 - E_2$
 - $E_1 \times E_2$
 - $\sigma_P(E_1)$, P jest predykatem na atrybutach w E_1
 - $\Pi_S(E_1)$, S jest listą zawierającą wybrane atrybuty z E_1
 - $\rho_x(E_1)$, x jest nową nazwą dla wyniku E_1



AGH

Operacja przecięcia zbiorów

Pozwala znaleźć krotki, które są w obydwu relacjach wejściowych.

- » Notacja: $r \cap s$
- » Założenie:
 - r, s mają ten *sam stopień*
 - atrybuty relacji r i s są kompatybilne
- » Np : Znajdź wszystkie kursy prowadzone zarówno w semestrze jesiennym 2017 semester jak i w wiosennym 2018.

$$\Pi_{course_id} (\sigma_{semester="Fall" \wedge year=2017}(section)) \cap \Pi_{course_id} (\sigma_{semester="Spring" \wedge year=2018}(section))$$

<i>course_id</i>
CS-101

Relacja *section*

<i>course_id</i>	<i>sec_id</i>	<i>semester</i>	<i>year</i>	<i>building</i>	<i>room_number</i>	<i>time_slot_id</i>
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

Operacja złączenia

» Iloczyn kartezjański

instructor X teaches

kojarzy każdą krotkę z *instructor* z każdą krotką z *teaches*.

- Większość powstałych wierszy zawiera informacje o instruktorach, którzy NIE nauczali konkretnego kursu.

» Aby uzyskać tylko te krotki z “*instructor X teaches*” które odnoszą się do instruktorów i kursów, które nauczali:

$\sigma_{instructor.id = teaches.id} (instructor \times teaches)$

- dostajemy tylko te krotki “*instructor X teaches*” które dotyczą instruktorów i kursów, których nauczali.

Wspólny atrybut ID

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

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

Złączenie naturalne

ID	name	dept_name	salary	ID	course_id	sec_id	semester	year
10101	Srinivasan	Comp. Sci.	65000	10101	CS-101	1	Fall	2009
12121	Wu	Finance	90000	10101	CS-315	1	Spring	2010
15151	Mozart	Music	40000	10101	CS-347	1	Fall	2009
22222	Einstein	Physics	95000	12121	FIN-201	1	Spring	2010
32343	El Said	History	60000	15151	MU-199	1	Spring	2010
3345				22222	PHY-101	1	Fall	2009
4556				32343	HIS-351	1	Spring	2010
4556	10101	Srinivasan	Comp. Sci.	65000	CS-101	1	Fall	2009
5858	10101	Srinivasan	Comp. Sci.	65000	CS-315	1	Spring	2010
7654	10101	Srinivasan	Comp. Sci.	65000	CS-347	1	Fall	2009
7676	12121	Wu	Finance	90000	FIN-201	1	Spring	2010
8382	15151	Mozart	Music	40000	MU-199	1	Spring	2010
9834	22222	Einstein	Physics	95000	PHY-101	1	Fall	2009
	32343	El Said	History	60000	HIS-351	1	Spring	2010
	45565	Katz	Comp. Sci.	75000	CS-101	1	Spring	2010
	45565	Katz	Comp. Sci.	75000	CS-319	1	Spring	2010
	76766	Crick	Biology	72000	BIO-101	1	Summer	2009
	76766	Crick	Biology	72000	BIO-301	1	Summer	2010
	83821	Brandt	Comp. Sci.	92000	CS-190	1	Spring	2009
	83821	Brandt	Comp. Sci.	92000	CS-190	2	Spring	2009
	83821	Brandt	Comp. Sci.	92000	CS-319	2	Spring	2010
	98345	Kim	Elec. Eng.	80000	EE-181	1	Spring	2009

Złączenie naturalne

- » Znajdź nazwiska wszystkich instruktorów wraz z „course_id” wszystkich prowadzonych przez nich kursów”

$\Pi_{name, course_id} (instructor \bowtie teaches)$

<i>name</i>	<i>course_id</i>
Srinivasan	CS-101
Srinivasan	CS-315
Srinivasan	CS-347
Wu	FIN-201
Mozart	MU-199
Einstein	PHY-101
El Said	HIS-351
Katz	CS-101
Katz	CS-319
Crick	BIO-101
Crick	BIO-301
Brandt	CS-190
Brandt	CS-319
Kim	EE-181

Złączenie naturalne

- » Rozważmy dwie relacje $r(R)$ i $s(S)$.
- » Złączenie naturalne między r i s oznaczane $r \bowtie s$ jest relacją o schemacie $R \cup S$ zdefiniowaną jako:

$$r \bowtie s = \Pi_{R \cup S} (\sigma_{r.A1 = s.A1 \wedge r.A2 = s.A2 \wedge \dots \wedge r.An = s.An} (r \times s))$$

gdzie $R \cap S = \{A_1, A_2, \dots, A_n\}$

Złączenie naturalne

- » Znajdź imiona i nazwiska wszystkich instruktorów z wydziału Comp. Sci. wraz z nazwami wszystkich kursów przez nich prowadzonych

$\Pi_{name, title}(\sigma_{dept_name="Comp.Sci"}(instructor \bowtie teaches \bowtie course))$

<i>name</i>	<i>title</i>
Brandt	Game Design
Brandt	Image Processing
Katz	Image Processing
Katz	Intro. to Computer Science
Srinivasan	Intro. to Computer Science
Srinivasan	Robotics
Srinivasan	Database System Concepts

$instructor \bowtie (teaches \bowtie course)$

$(instructor \bowtie teaches) \bowtie course$

Złączenie naturalne - problemy

- Znajdź nazwiska wszystkich wykładowców z wydziału informatyki wraz z nazwami prowadzonych przez nich kursów
 - $\Pi_{name, title} (\sigma_{dept_name="Comp. Sci."} (instructor \bowtie teaches \bowtie course))$
- teaches(ID,name,dept_name,salary,course_id,sec_id)
course(course_id,title,dept_name,credits)
instructor(ID,name,dept_name,salary)
- Niepoprawne dopasowanie: `course.dept_name=instructor.dept_name`

Złączenie theta

- Operator **złączenia** pozwala łączyć operację selekcji i iloczynu kartezjańskiego w pojedynczą operację.
- Rozważmy relacje $r (R)$ i $s (S)$
- Niech "theta" bądź predykatem atrybutów w schemacie R „union” S. Operator złączenia $r \bowtie_{\theta} s$ jest zdefiniowany następująco:

$$r \bowtie_{\theta} s = \sigma_{\theta}(r \times s)$$

- Tak więc

$$\sigma_{instructor.id = teaches.id}(instructor \times teaches)$$

- Można równoważnie zapisać:

$$instructor \bowtie_{Instructor.id = teaches.id} teaches.$$

Operacja przypisania

- Czasami wygodnie jest napisać wyrażenie algebry relacji poprzez przypisanie jego części do zmiennych relacji tymczasowych.
- Operator przypisania oznaczany \leftarrow działa jak przypisanie w języku programowania.
- Np.. : Znajdź instruktorów z wydziału "Physics" i "Music".

$Physics \leftarrow \sigma_{dept_name="Physics"}(instructor)$

$Music \leftarrow \sigma_{dept_name="Music"}(instructor)$

$Physics \cup Music$

- Za pomocą operatora przypisania zapytanie można zapisać jako program sekwencyjny składający się z serii przypisań, po których następuje wyrażenie, którego wartość jest pokazywana jako wynik zapytania.

Złączenie zewnętrzne (*outer join*)

- Rozszerzenie operatora złączenia, który zapobiega utracie danych.
- Wylicza złączenie, a następnie dodaje do wyniku krotki z jednej z relacji, które nie pasują do krotek w drugiej relacji.
- Używa wartości *null*:
 - *null* oznacza, że wartość jest nieznana lub nie istnieje
 - wszystkie porównania z *null* są (generalnie) z definicji **false**.

Przykład – złączenie zewnętrzne

- Relacja *instructor*

<i>ID</i>	<i>name</i>	<i>dept_name</i>
10101	Srinivasan	Comp. Sci.
12121	Wu	Finance
15151	Mozart	Music

- Relacja *teaches*

<i>ID</i>	<i>course_id</i>
10101	CS-101
12121	FIN-201
76766	BIO-101

Przykład – złączenie zewnętrzne

- Złączenie

instructor ⋈ *teaches*

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>course_id</i>
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201

- Left Outer Join

instructor ⋈_□ *teaches*

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>course_id</i>
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201
15151	Mozart	Music	<i>null</i>

Przykład – złączenie zewnętrzne

- Right Outer Join

instructor ⋈ *teaches*

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>course_id</i>
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201
76766	null	null	BIO-101

- Full Outer Join

instructor ⋈ *teaches*

<i>ID</i>	<i>name</i>	<i>dept_name</i>	<i>course_id</i>
10101	Srinivasan	Comp. Sci.	CS-101
12121	Wu	Finance	FIN-201
15151	Mozart	Music	<i>null</i>
76766	null	null	BIO-101

Równoważne zapytania

- » Najczęściej istnieje więcej niż jeden sposób napisania zapytania w algebrze relacji.
- » Np.: Znajdź informacje o kursach prowadzonych przez instruktorów na wydziale fizyki, których wynagrodzenie przekracza 90 000

- » Zapytanie 1

$$\sigma_{dept_name="Physics" \wedge salary > 90,000} (instructor)$$

- » Zapytanie 2

$$\sigma_{dept_name="Physics"} (\sigma_{salary > 90.000} (instructor))$$

- » Dwa zapytania nie są identyczne; są jednak równoważne - dają taki sam wynik w dowolnej bazie danych.

Równoważne zapytania c.d.

- » Np.: Znajdź informacje o kursach prowadzonych przez instruktorów na wydziale fizyki

- » Zapytanie 1

$\sigma_{dept_name="Physics"} (instructor \bowtie_{instructor.ID = teaches.ID} teaches)$

- » Zapytanie 2

$(\sigma_{dept_name="Physics"} (instructor)) \bowtie_{instructor.ID = teaches.ID} teaches$

- » Dwa zapytania nie są identyczne; są jednak równoważne - dają taki sam wynik w dowolnej bazie danych.

Dostarczają możliwości pisania zapytań, które nie mogą być wyrażone przy pomocy podstawowych operatorów algebry relacji

- Uogólniona projekcja
- Funkcje agregujące

Uogólniona projekcja

- Rozszerza operację projekcji zezwalając aby operatory arytmetyczne były użyte w liście projekcji.

$$\Pi_{F_1, F_2, \dots, F_n}(E)$$

- E jest dowolnym wyrażeniem algebry relacji
- Każdy F_1, F_2, \dots, F_n jest wyrażeniem arytmetycznym ze stałymi i atrybutami ze schematu E .
- Dana relacja *instructor*(*ID*, *name*, *dept_name*, *salary*) gdzie *salary* jest roczną pensją, pokaż te same dane, ale z pensją miesięczną

$$\Pi_{ID, name, dept_name, salary/12}(instructor)$$

Funkcje i operacje agregujące

- Relacja r :

A	B	C
α	α	7
α	β	7
β	β	3
β	β	10

$$\square \mathcal{G}_{\text{sum}(c)}(r)$$

sum(c)
27

Funkcja agregująca bierze zbiór wartości i zwraca w wyniku pojedynczą wartość:

avg, min, max, sum, count

$$\mathcal{G}_{\text{sum}(\text{salary})}(\text{instructor})$$

(Kaligraficzne G)

Operacje agregujące - przykład

- Znajdź średnią pensję na każdym wydziale

dept_name **G** **avg**(salary) (*instructor*)

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

dept_name	avg_salary
Biology	72000
Comp. Sci.	77333
Elec. Eng.	80000
Finance	85000
History	61000
Music	40000
Physics	91000

- **Operacja agregacji** w algebrze relacji

$$G_1, G_2, \dots, G_n \mathcal{G}_{F_1(A_1), F_2(A_2), \dots, F_n(A_n)}(E)$$

E jest dowolnym wyrażeniem algebry relacji

- G_1, G_2, \dots, G_n lista atrybutów po której ma być grupowanie (może być pusta)
 - Każde F_i jest funkcją agregującą
 - Każde A_i jest nazwą atrybutu
- Uwaga: Czasem używane jest γ zamiast \mathcal{G}

Modyfikacje bazy danych

- Zawartość bazy danych może być zmieniana przy użyciu następujących operatorów :
 - Usuwanie
 - $r \leftarrow r - E$ (r - relacja, E – wyrażenie algebry relacji)
 - Wstawianie
 - $r \leftarrow r \cup E$
 - Modyfikacja
 - $r \leftarrow \Pi_{F_1, F_2, \dots, F_l}(r)$
- Wszystkie te operacje mogą być wyrażone przy pomocy operatora przypisania

Algebra relacji wielozbiorów

- „Czysta” algebra relacji usuwa wszystkie duplikaty
 - np. po projekcji
- Algebra relacji wielozbiorów pozostawia duplikaty aby dopasować się do semantyki SQL
 - zostawianie duplikatów w SQL było początkowo zrobione w celach efektywnościowych, ale teraz jest jego cechą
- Algebra wielozbiorów definiowana następująco
 - *selekcja*: ma tyle duplikatów krotki co na wejściu, jeżeli krotka spełnia warunek selekcji
 - *projekcja*: jedna krotka na krotkę wejściową, nawet jeżeli jest duplikatem
 - *iloczyn kartezyjski*: jeżeli jest m kopii $t1$ w r i n kopii $t2$ w s , to jest $m \times n$ kopii $t1.t2$ w $r \times s$
 - pozostałe operacje podobnie definiowane
 - np. unia: $m + n$ kopii, przecięcie: $\min(m, n)$ kopii
 - różnica: $\max(0, m - n)$ kopii

- **select** $A1, A2, \dots, An$
from $r1, r2, \dots, rm$
where P

jest równoważne następującemu wyrażeniu w algebrze relacji wielozbiorów

$$\Pi_{A1, \dots, An} (\sigma_P (r1 \times r2 \times \dots \times rm))$$

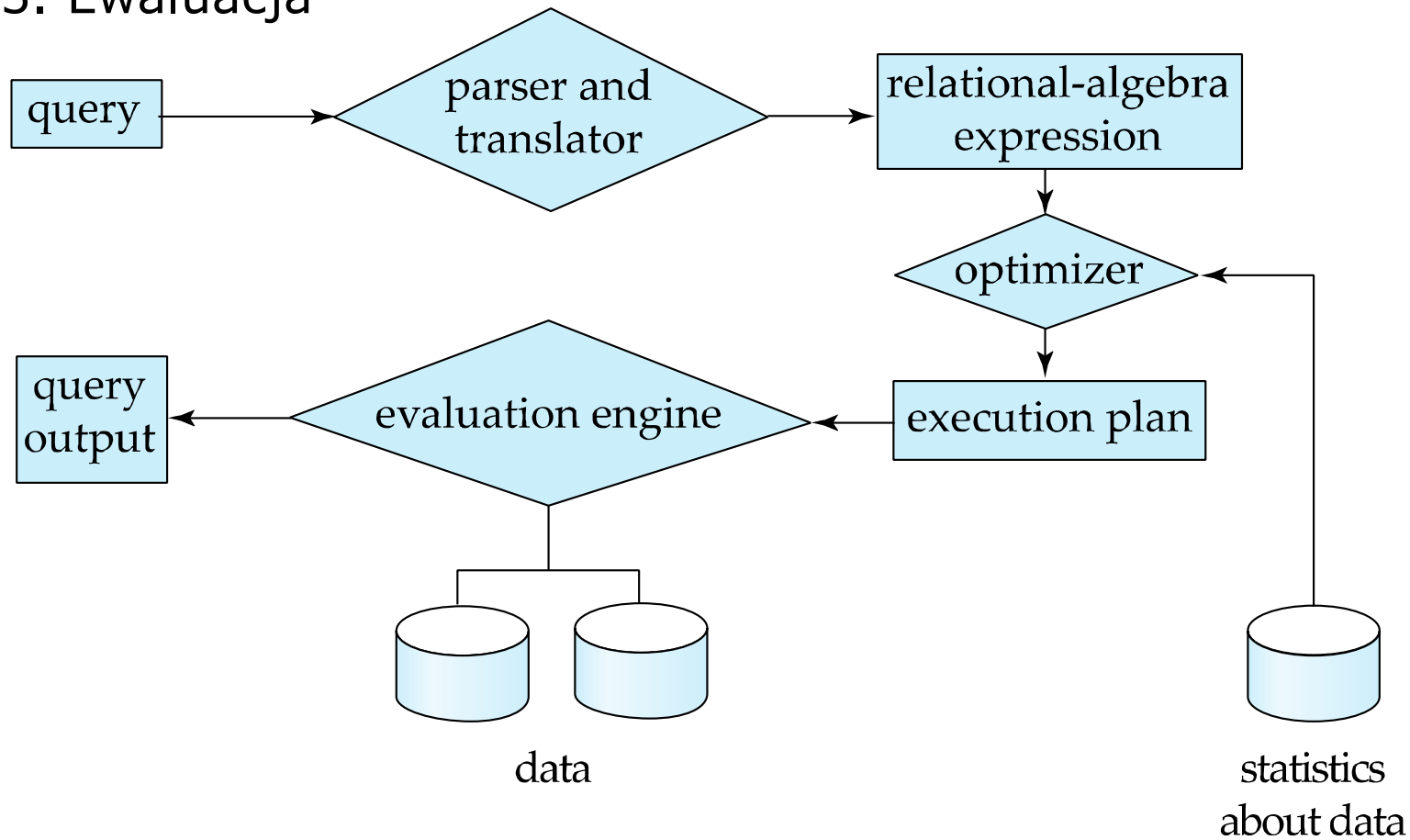
- **select** $A1, A2, \mathbf{sum}(A3)$
from $r1, r2, \dots, rm$
where P
group by $A1, A2$

jest równoważne następującemu wyrażeniu w algebrze relacji wielozbiorów

$$A1, A2 \mathcal{G} \mathbf{sum}(A3) (\Pi_{A1, \dots, An} (\sigma_P (r1 \times r2 \times \dots \times rm)))$$

Podstawowe kroki w przetwarzaniu zapytań

1. Parsowanie i translacja
2. Optymalizacja
3. Ewaluacja



Zapytanie i plan jego wykonania

select salary
from instructor
where salary < 75000

może być przetłumaczone na
jedno z wyrażeń algebry relacji:

$\sigma_{salary < 75000} (\Pi_{salary}(instructor))$
 $\Pi_{salary}(\sigma_{salary < 75000}(instructor))$

π_{salary}

$\sigma_{salary < 75000; \text{use index 1}}$

instructor

Optymalizacja planu

