

ALGEBRA RELACJI – LISTA 1

ZAD 4

Z4. (2 pkt.) Rozważmy bazę danych oznaczoną w kalkulatorze relacji jako Silberschatz - UniversityDB (lista relacji z atrybutami poniżej).

Potrenuj samodzielnie pisząc w kalkulatorze algebra relacji wyrażenia algebra relacji dla poniższych pytań. Na zajęciach trzeba będzie zapisać w algebrze relacji podobne, ale nie takie same zapytania.

- Dla każdego filmu sprzed 1960 roku wypisz jego tytuł, nazwisko reżysera i gatunek (genre).
- Wypisz imiona i nazwiska aktorów, którzy nie zagrały w żadnym filmie *Quentin Tarantino*.
- Wypisz imiona i nazwiska aktorów, którzy dokładnie raz zagrały w jakimś filmie.
- Wypisz tytuły filmów, które są jednocześnie gatunku *Drama* oraz *Sci-Fi*.
- Wypisz pełne dane filmów z najwyższą wartością atrybutu *rank*.
- Wypisz nazwiska aktorów, którzy zagrały taką samą rolę w co najmniej dwóch, różnych filmach.
- Wypisz nazwiska reżyserów, którzy nigdy nie nakręcili horroru.
- Wypisz nazwiska reżyserów, którzy nakręcili film, w którym wśród aktorów nie było kobiet.

<https://dbis-uibk.github.io/relax/calc/gist/4f7866c17624ca9dfa85ed2482078be8/relax-silberschatz-english.txt/0>

A)

- Dla każdego filmu sprzed 1960 roku wypisz jego tytuł, nazwisko reżysera i gatunek (genre).

$$\pi_{name, last_name, genre}((\rho_{movie_id \leftarrow id}(\sigma_{year < 1960}(movies))) \bowtie movies_directors \bowtie (\rho_{director_id \leftarrow id}(directors)) \bowtie movies_genres)$$

$\pi name, last_name, genre ((\rho movie_id \leftarrow id (\sigma year < 1960 (movies))) \bowtie movies_directors \bowtie (\rho director_id \leftarrow id (directors)) \bowtie movies_genres)$

B)

- Wypisz imiona i nazwiska aktorów, którzy nie zagrały w żadnym filmie *Quentin Tarantino*.

$$\pi_{first_name, last_name, id}(actors) \setminus (\pi_{actors.first_name, actors.last_name, actor.id}(\rho_{actor.id \leftarrow id}(actors)) \bowtie roles \bowtie movies_directors \bowtie \sigma_{last.name_d='Tarantino'}(\rho_{director.id \leftarrow id, first.name_d \leftarrow first.name, last.name_d \leftarrow last.name}(directors)))$$

$\pi actors.first_name, actors.last_name, actors.id (actors) - \pi actors.first_name, actors.last_name, actor.id (\rho actor.id \leftarrow id (actors)) \bowtie roles \bowtie movies_directors \bowtie \sigma_{last.name_d='Tarantino'} (\rho director.id \leftarrow id, first.name_d \leftarrow first.name, last.name_d \leftarrow last.name (directors))$

druga wersja (nie do końca działa)

$\pi_{\text{first_name}, \text{last_name}}(\text{actors} - \pi_{\text{actor_id}}(\text{actors}), \text{actors.first_name}, \text{actors.last_name}, \text{actors.gender}) \rho_{\text{actor_id} \leftarrow \text{id}(\text{actors})} \bowtie \text{roles} \bowtie \text{movies_directors} \bowtie \sigma_{\text{last_name_d} = \text{'Tarantino'}} (\rho_{\text{director_id} \leftarrow \text{id}(\text{directors})}, \text{first_name_d} \leftarrow \text{first_name}(\text{directors}), \text{last_name_d} \leftarrow \text{last_name}(\text{directors})))$

C)

- (c) Wypisz imiona i nazwiska aktorów, którzy dokładnie raz zagraли w jakimś filmie.

$\pi_{\text{actors.first_name}, \text{actors.last_name}}(\text{actors}) \setminus (\pi_{\text{actors.first_name}, \text{actors.last_name}}((\sigma_{\text{actor_id} = \text{actor_id}2 \wedge \text{movie_id} \neq \text{movie_id}2} (\text{roles} \times \rho_{\text{actor_id}2 \leftarrow \text{actor_id}}, \text{movie_id}2 \leftarrow \text{movie_id}, \text{role2} \leftarrow \text{role}(\text{roles}))) \bowtie \rho_{\text{actor_id} \leftarrow \text{id}}(\text{actors})))$

$\pi_{\text{actors.first_name}, \text{actors.last_name}}(\text{actors}) - \pi_{\text{actors.first_name}, \text{actors.last_name}}((\sigma_{\text{actor_id} = \text{actor_id}2 \wedge \text{movie_id} \neq \text{movie_id}2} (\text{roles} \times \rho_{\text{actor_id}2 \leftarrow \text{actor_id}}, \text{movie_id}2 \leftarrow \text{movie_id}, \text{role2} \leftarrow \text{role}(\text{roles}))) \bowtie \rho_{\text{actor_id} \leftarrow \text{id}}(\text{actors}))$

D

- (d) Wypisz tytuły filmów, które są jednocześnie gatunku *Drama* oraz *Sci-Fi*.

$\pi_{\text{movies.name}}((\sigma_{\text{movie_id} = \text{movie_id}2 \wedge \text{genre} = \text{'Sci-Fi'} \wedge \text{genre2} = \text{'Drama'}} (\text{movies_genres} \times \rho_{\text{movie_id}2 \leftarrow \text{movie_id}, \text{genre2} \leftarrow \text{genre}}(\text{movies_genres}))) \bowtie \rho_{\text{movie_id} \leftarrow \text{id}}(\text{movies}))$

$\pi_{\text{movies.name}}((\sigma_{\text{movie_id} = \text{movie_id}2 \wedge \text{genre} = \text{'Sci-Fi'} \wedge \text{genre2} = \text{'Drama'}} (\text{movies_genres} \times \rho_{\text{movie_id}2 \leftarrow \text{movie_id}, \text{genre2} \leftarrow \text{genre}}(\text{movies_genres}))) \bowtie \rho_{\text{movie_id} \leftarrow \text{id}}(\text{movies}))$

E

MAX / MIN

- (e) Wypisz pełne dane filmów z najwyższą wartością atrybutu *rank*.

$\text{movies} \setminus \pi_{\text{movies.id}, \text{movies.name}, \text{movies.year}, \text{movies.rank}} (\sigma_{\text{rank} < \text{rank}2} (\text{movies} \times \rho_{\text{id}2 \leftarrow \text{id}, \text{name2} \leftarrow \text{name}, \text{year2} \leftarrow \text{year}, \text{rank2} \leftarrow \text{rank}}(\text{movies})))$

$\text{movies} - \pi_{\text{movies.id}, \text{movies.name}, \text{movies.year}, \text{movies.rank}} (\sigma_{\text{rank} < \text{rank}2} (\text{movies} \times \rho_{\text{id}2 \leftarrow \text{id}, \text{name2} \leftarrow \text{name}, \text{year2} \leftarrow \text{year}, \text{rank2} \leftarrow \text{rank}}(\text{movies})))$

F

- (f) Wypisz nazwiska aktorów, którzy zagrały taką samą rolę w co najmniej dwóch, różnych filmach.

$\pi_{\text{actors.last_name}} (\sigma_{\text{role} = \text{role}2 \wedge \text{movie_id} \neq \text{movie_id}2 \wedge \text{actor_id} = \text{actor_id}2} (\rho_{\text{actor_id}2 \leftarrow \text{actor_id}, \text{movie_id}2 \leftarrow \text{movie_id}, \text{role2} \leftarrow \text{role}(\text{roles})} \bowtie \text{roles} \bowtie \rho_{\text{actor_id} \leftarrow \text{id}}(\text{actors})))$

$$\pi_{actors.last_name}(\sigma_{role=role2 \wedge movie_id \neq movie_id2 \wedge actor_id=actor_id2}(\rho_{actor_id2 \leftarrow actor_id, movie_id2 \leftarrow movie_id, role2 \leftarrow role}(roles) \times roles)) \bowtie \rho_{actor_id \leftarrow id}(actors)$$
$$\pi_{actors.last_name}(\sigma_{role=role2 \wedge movie_id \neq movie_id2 \wedge actor_id=actor_id2}(\rho_{actor_id2 \leftarrow actor_id, movie_id2 \leftarrow movie_id, role2 \leftarrow role}(roles) \times roles) \bowtie \rho_{actor_id \leftarrow id}(actors))$$

G

(g) Wypisz nazwiska reżyserów, którzy nigdy nie nakręcili horroru.

$$\pi_{last_name}(directors) \setminus \pi_{directors.last_name}(\sigma_{genre='Horror'}(\rho_{director_id \leftarrow id}(directors) \bowtie directors_genres))$$
$$\pi_{last_name}(directors) - \pi_{directors.last_name}(\sigma_{genre='Horror'}(\rho_{director_id \leftarrow id}(directors) \bowtie directors_genres))$$

H

(h) Wypisz nazwiska reżyserów, którzy nakręcili film, w którym wśród aktorów nie było kobiet.

$$\pi_{directors.last_name}((\pi_{movie_id}(roles) \setminus \pi_{movie_id}(\sigma_{gender='F'}(roles \bowtie \rho_{actor_id \leftarrow id}(actors)))) \bowtie movies_directors \bowtie \rho_{director_id \leftarrow id}(directors))$$
$$\pi_{directors.last_name}((\pi_{movie_id}(roles) - \pi_{movie_id}(\sigma_{gender='F'}(roles \bowtie \rho_{actor_id \leftarrow id}(actors)))) \bowtie movies_directors \bowtie \rho_{director_id \leftarrow id}(directors))$$

RELACJĘ JNY KROTEK I RACHUNEK DLA EDU - USTAWA

RRK - wybiera krótkie mojące dane właściwość

RRD - wybiera wektor wartości zmiennych mających dane właściwość

ZAD 6

6. (2 pkt. - po 0.5 pkt. za podpunkt) Baza danych składa się z relacji:

- **F(idf,tytul,rezyser,rokProd,czas)** — idf jest kluczem; tytuł i inne atrybuty nie muszą być unikalne; czas oznacza czas trwania filmu i jest podany w minutach;
- **S(idf,sala,data,godz)** — w podanej sali i terminie jest projekcja filmu o podanym identyfikatorze;
- **A(pesudo,imie,nazwisko,narodowosc,rokUr)** — informacje o aktorach; pseudonim jest unikalny;
- **P(pesudo,idf,nastep_gara)** — informacja, że aktor o podanym pseudonimie i idf jest kolejnym gospodarzem.

- ~~$r.pseudo.idf$~~ , postać, gaza) — informacja, że aktor o podanym pseudonimie grał w filmie daną postać i otrzymał za to daną gażę.
- M(~~$pseudo$~~ , rok, minGaza) — informacja, że aktor o podanym pseudonimie w danym roku na podanym poziomie ustalił minimalną gażę za grę w filmie.

Zapisz poniższe zapytania w rrd lub rrk.

1

- Podaj dane aktorów (pseudonim, imię, nazwisko, rok urodzenia, narodowość), którzy pojawiły się w filmach produkowanych tylko w jednym roku (powiedzmy, że są to gwiazdy jednego sezonu).

rrd

$$\{(ps, im, naz, nar, rU) | A(ps, im, naz, nar, rU) \wedge (\exists idf, post, gaza)(R(ps, idf, post, gaza)) \wedge \\ \wedge (\forall idf', post', gaza')(R(ps, idf', post', gaza') \implies (\exists tyt, rez, rok, czas))(F(idf, tyt, rez, rok, czas)) \\ \wedge (\exists tyt', rez', rok', czas')(F(idf', tyt', rez', rok', czas') \wedge rok == rok')\}$$

rrk

$$\{a | a \in A \wedge (\exists r)(r \in R \wedge r.pseudo = a.pseudo \wedge (\exists f)(f \in F \wedge f.idf = r.idf) \wedge \\ (\forall r')(r' \in R \wedge r'.pseudo = r.pseudo \implies (\exists f')(f' \in F \wedge f'.idf = r'.rokProd < f'.rokProd))\}$$

2

- Podaj pełne krotki filmów, które są najnowszymi filmami reżyserów.

rrk

$$\{f | f \in F \wedge \neg(\exists f')(f' \in F \wedge f.idf \neq f'.idf \wedge f.rezyser = f'.rezyser \wedge f.rokProd < f'.rokProd)\}$$

3

- Dla każdego filmu znajdź aktora, który dostał najwyższą gażę w tym filmie (został najlepiej opłacony z obsady filmu). W relacji wynikowej podaj pseudonim aktora, idf oraz gażę.

rrk

$$\{Z^{[pseudo,idf,gaza]} | (\exists r)(r \in R \wedge r.pseudo = Z.pseudo \wedge r.idf = Z.idf \wedge r.gaza = Z.gaza) \\ \wedge \neg(\exists r')(r' \in R \wedge r.pseudo \neq r'.pseudo \wedge r.idf = r'.idf \wedge r.gaza < r'.gaza)\}$$

4

- Podaj pełne krotki aktorów, którzy nigdy nie obniżyli swojej minimalnej gaży (w późniejszych latach mogła ona najwyżej rosnąć). Na wynik nie wpływają lata, w których aktor nie podał minimalnej gaży.

rrk

$$\{a | a \in A \wedge (\exists m)(m \in M \wedge a.pseudo = m.pseudo) \wedge \neg(\exists m')(m' \in M \\ \wedge a.pseudo = m'.pseudo \wedge m.rok < m'.rok \wedge m.gaza > m'.gaza)\}$$

**SQL – USTY NIEOBOWIĄZKOWE
Skos**

WARTOŚĆ NULL i UNKNOWN

TRUE/
FALSE \wedge NULL = NULL

TRUE/
FALSE \vee NULL = TRUE/FALSE

NULL \wedge/\vee NULL = NULL

tok samo
sytuacja
wygląda dla
UNKNOWN

• COUNT(*) zlicza NULL

• COUNT(kol) nie zlicza NULL

• SUM(kol) ignoriuje NULL, sumuje reszte

jeśli będą same NULL w kol to zwrócić NULL

↑ podstanie inne f. agregujace (oprócz COUNT - on zwraca 0)

B-DRzewa

Są drzewkowozłożonymi drzewami wyszukiwania, które dostosły do projektowane w taki sposób, żeby dobrze dostosować się dyskach magnetycznych lub innych pamięciach zewnętrznych o dostępie bezpośrednim.

INDEKSOWANIE

Indeksy to pomocnicze struktury danych.

PRAWIŚPIEZA: wyszukiwanie, dżascenie, kontrolę unikalności i innych warunków

SPÓWALNIA: wstawianie, usuwanie i modyfikacie indeksowanych krotek

PRZYKŁAD:

- jeśli często będziemy wykonywać zapytanie:

SELECT * FROM users

WHERE displayname = 'Isaac'

[CREATE INDEX i-users-displayname
ON users (displayname)]

- jeśli indeks nie zawiera wszystkich potrzebnych kolumn

SELECT y FROM tab

WHERE x = 'key'

[CREATE INDEX tab-x-y ON tab(x,y)]

lub

[CREATE INDEX tab-x-y ON tab(x)]

INCLUDE y

JAK SPROWADZIĆ W JAKI
JEST PLAN WYKONANIA
ZAPYTANIA I W JAKIM
Czasie się ono wykona?

EXPLAIN ANHME (ZAPYTANIE)

METODY DOSTĘPU:

- sequential scan - czyta wszystko
- index scan - sprawdza tylko posiadające krotki (dobry gdy małe)
- bitmap scan - zaznacza strony x posiadającymi krotkami i je przegąda
- index only scan - potrafi tylko do indeksów

KLUCZE I KLUCZE OBCE

KLUCZ OBCY - kolumna lub grupa kolumn w jednej tabeli stanowiących klucz główny w innej tabeli.

Klucz obcy służy do zapewnienia integralności danych. Wykorzystuje się go do tworzenia relacji między tabelami.

Co prawda najczęściej **klucz obcy w SQL** jest to odwołanie do klucza głównego w innej tabeli, jednak są inne możliwości.

- klucz obcy może odwoływać się do kolumny w innej tabeli która nie jest oznaczona jako PRIMARY KEY, kolumna ta jednak musi być oznaczona jako UNIQUE (zapewnia unikatowe wartości w kolumnie bądź grupie kolumn)
- klucz obcy może odwoływać się do kolumny z tej samej tabeli

Klucz obcy może odwoływać się jedynie do tabel w tej samej bazie danych. Integralność danych z innymi bazami danych trzeba wtedy zaimplementować przy pomocy wyzwalaczy (Trigger).

KLUCZ RELACJI

Kluczem relacji R jest taki podzbior K

zakresem całego relacji i jej
jej atrybutów, który

- $K \rightarrow R$ (wyznacza funkcjonalnie wszystkie
atrybuty R)

- jest minimalnym zbiorem o tej
własności

NADKUJCZ - dowolny zbiór atrybutów
zawierający klucz relacji

KLUCZ GŁÓWNY - jeden z kluczy
relacji

KLUCZ ALTERNATYWNY - klucz relacji
inny niż klucz
główny

ATRIBUT GŁÓWNY - atrybut (dowolnego)
klucza relacji

ZALEŻNOŚĆ FUNKCJONA

Dla relacji $R = A_1 A_2 \dots A_K$ oraz
zbiorów jej atrybutów $\alpha, \beta \subseteq \{A_1 \dots A_K\}$
zachodzi zależność funkcjonalna
 $\alpha \rightarrow \beta$, jeżeli dla każdego r
relacji R zachodzi:

$$(\forall t_1, t_2 \in r)((t_1 \cdot \alpha = t_2 \cdot \alpha) \Rightarrow (t_1 \cdot \beta = t_2 \cdot \beta))$$

(24L1)

jeżeli dla każdych dwóch krotek
 t_1, t_2 rekord α jest taki sam,
to rekord β też jest taki sam.

wtedy mówimy o
zależności funkcyjnej d $\rightarrow \beta$

PRZYKŁADY ZALEŻNOŚCI:

- Tabele PR(id, tytuł, nazwisko, adres)
 $\text{id} \rightarrow \text{nazwisko, adres, tytuł}$
- ST(id, indeks, nazwisko, adres)
 $\text{id} \rightarrow \text{indeks, nazwisko, adres}$
 $\text{indeks} \rightarrow \text{id, nazwisko, adres}$
- TS(id-grupy, termin, sala)
 $\text{termin, sala} \rightarrow \text{id-grupy}$

POSTACIE NORMALNE

1NF

atrbuty są otomane, tabele
miają klucz główny

kobieta	Asia, Ole, Kasia...	X
mężczyzna	Tomek, Olek	

kobieta	Asia
kobieta	Ole
kobieta	Kasia
	:

2NF

jest w 1NF oraz
kolumny, które nie są w kluce,
mogą być funkcyjnie zależne
od części klucza

a więc



3NF

jest w 2NF oraz dla nietrywialnej zależności

$\alpha \rightarrow \beta$

- zbior α jest domknięciem
- ma po prawej stronie atrybut główny (β jest częścią jekiego's klucza)

PRZYKŁAD:

- Gong → Miasto
- Miasto, Proceder → Gong
- Gong → Siet

Rozkład do 3NF:

Nefde(Miasto, Gong, Proceder)
Sætawje(Gong, Siet)

BCNF

dla każdej nietrywialnej zależności $\alpha \rightarrow \beta$ ($\alpha \cap \beta = \emptyset$)
zbior α jest domknięciem

KAŻDA RELACJA MA ODWRACALNY ROZKŁAD NA SKŁADOWE W BCNF!

ISTNIEJĄ RELACJE, KTÓRE NIE MAJĄ ODWRACALNEGO
I ZACHOWUJĄCEGO ZALEŻNOŚCI ROZKŁADU NA SKŁADOWE
W BCNF!

M | G | P

→ G | M G | P

Gong → Miejsce

Miejsce, Proceder →
→ Gong

Miejsce (Miejsce, Gong)

Procedery (Gong, Proceder)

REDUNDACJA - nadmierowość w stosunku do tego co konieczne

NORMALIZACJA - proces doprowadzający bazy danych do postaci bez nadmierzności.

ANOMALIE USUWANIA - usuwanie części informacji powoduje utratę innej informacji, której nie określibyśmy utracić. (co po usunięciu ostatniej grupy powstanie?)

ANOMALIA MODYFIKACJI - informacje zostawia zmodyfikowane w pewnych krotkach, a w innych nie. Które informacje jest właściwo prowadzić?

ANOMALIA WSTAWIENIA - wprowadzenie pewnej informacji jest możliwe tylko wtedy, gdy jednocześnie wprowadzony jest inna informacja, która może być obecnie niedostępna. (jak dodać nowego prawnego dotyczącego bez grupy)

INNER JOIN

- zwraca wynikowe te wiersze dla których kolumny użyte do łączenia mają te same wartości

OUTER JOIN

LEFT OUTER JOIN zwraca:

- wiersze dla których warunek łączenia jest spełniony
- wiersze z lewej tabeli dla których nie ma odpowiedników w prawej

RIGHT OUTER JOIN

jak wyżej tylko odpowiednie dla prawej tabeli

FULL OUTER JOIN zwraca:

- wiersze dla których warunek łączenia jest spełniony
- wiersze z lewej tabeli dla których nie ma odpowiedników w prawej
- tak samo z prawej do lewej

