



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

Projekt współfinansowany ze
środków Unii Europejskiej w ramach
Europejskiego Funduszu
Społecznego

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



7 Ćwiczenia: Trudne zapytania w algebrze relacji

7.1 Przykładowa baza danych

- OSOBY(ido,nazwisko,imię) — atrybut ido jest jednoznaczny identyfikatorem osoby;
- BARY(idb,nazwa,adres) — atrybut idb jest jednoznaczny identyfikatorem baru;
- PODAJĄ(idb,sok,cena) — w barze o identyfikatorze idb jest podawany sok o nazwie sok w cenie cena;
- BYWA(ido,idb) — osoba o identyfikatorze ido bywa w barze o identyfikatorze idb
- LUBI(ido,sok) — osoba o identyfikatorze ido lubi sok o nazwie sok

7.2 Przykład z wykluczeniem

PYTANIE: Osoby, które nie chodzą do ŻADNEGO baru, w którym jest podawany sok kiwi

Zupełnie niezadawalającą odpowiedzią jest wybranie osób (ich identyfikatorów), które chodzą do jakiegoś baru, w którym jest podawany sok inny niż kiwi:

$$\pi_{ido}(\text{BYWA} \bowtie \sigma_{\text{sok} \neq \text{'kiwi'}}(\text{PODAJĄ}))$$

W takim barze może być podawany także sok kiwi, a poza tym, osoba ta może chodzić też do innego baru, gdzie sok kiwi jest podawany.

Podejrzmy więc do problemu metodą wykluczania: weźmy wszystkie osoby i wyrzucmy te, które chodzą do co najmniej jednego baru, w którym jest podawany sok kiwi.

- Identyfikatory osób bywających w CO NAJMNIEJ JEDNYM barze, w którym jest podawany sok kiwi:

$$\pi_{ido}(\text{BYWA} \bowtie \sigma_{\text{sok} = \text{'kiwi'}}(\text{PODAJĄ}))$$

- Identyfikatory osób niebywających w ŻADNYM barze, w którym jest podawany sok kiwi:

$$\underbrace{\pi_{ido}(\text{OSOBY}) \setminus \pi_{ido}(\text{BYWA} \bowtie \sigma_{\text{sok} = \text{'kiwi'}}(\text{PODAJĄ}))}_{\text{osoby niebywające w żadnym barze, w którym jest podawany sok kiwi}}$$

wszystkie osoby osoby bywające w co najmniej jednym barze, gdzie jest podawany sok kiwi

- Powiedzmy, że zależało nam na pełnych danych osób, a nie tylko ich identyfikatorach. Pozostaje więc połączyć wyselekcjonowane identyfikatory z relacją OSOBY:

$$\text{OSOBY} \bowtie \underbrace{\left(\pi_{ido}(\text{OSOBY}) \setminus \underbrace{\pi_{ido}(\text{BYWA} \bowtie \sigma_{\text{sok}='kiwi'}(\text{PODAJA}))}_{\text{identyfikatory osób niebywających w żadnym barze, w którym jest podawany sok kiwi}} \right)}_{\text{pełne krotki osób niebywających w żadnym barze, w którym jest podawany sok kiwi}}$$

7.3 Przykład z dokładnym wyborem

PYTANIE: Osoby, które chodzą TYLKO do barów, w których jest podawany sok kiwi

W algebrze relacji nie mamy operatora TYLKO. Łatwo możemy wybrać osoby (ich identyfikatory) chodzące do CO NAJMNIEJ jednego baru, gdzie jest podawany sok kiwi:

$$\underbrace{\pi_{ido}(\text{BYWA} \bowtie \underbrace{\sigma_{\text{sok}='kiwi'}(\text{PODAJA}))}_{\text{gdzie jest podawany sok kiwi}})}_{\text{bywanie w barach, w których jest podawany sok kiwi}} \\ \text{osoby, bywające w co najmniej jednym barze, gdzie jest podawany sok kiwi}$$

Nie jest to jednak poprawna odpowiedź na pierwotne pytanie, bo taka osoba może TAKŻE chodzić do barów, gdzie soku kiwi nie dostanie.

Do pytania, które nas interesuje, musimy podejść z drugiej strony — wykluczyć osoby chodzące do barów, gdzie nie jest podawany sok kiwi, czyli od **wszystkich osób**, odjąć te, które chodzą do barów, gdzie **nie jest podawany sok kiwi**. Wybranie barów, w których nie jest podawany sok kiwi wymaga powtórzenia tego samego triku: od **wszystkich barów** trzeba odjąć te, w których **jest podawany sok kiwi**.

- Identyfikatory barów, w których nie jest podawany sok kiwi:

$$\underbrace{\pi_{idb}(\text{BARY})}_{\text{wszystkie bary}} \setminus \underbrace{\pi_{idb}(\sigma_{\text{sok}='kiwi'}(\text{PODAJA}))}_{\text{bary, w których jest podawany sok kiwi}} \\ \text{bary, w których nie jest podawany sok kiwi}$$

- Identyfikatory osoby bywających w barach, w których nie jest podawany sok kiwi:

$$\pi_{ido}(\text{BYWA} \bowtie \underbrace{\left(\pi_{idb}(\text{BARY}) \setminus \pi_{idb}(\sigma_{\text{sok}='kiwi'}(\text{PODAJA})) \right)}_{\text{bary, w których nie jest podawany sok kiwi}}) \\ \underbrace{\hspace{10em}}_{\text{bywanie w barze, w którym nie jest podawany sok kiwi}} \\ \text{identyfikatory osoby bywających w barach, w których nie jest podawany sok kiwi}$$

- Identyfikatory osoby bywających TYLKO w barach, w których jest podawany sok kiwi:

$$\underbrace{\pi_{ido}(\text{OSOBY})}_{\text{wszystkie osoby}} \setminus \underbrace{\pi_{ido}(\text{BYWA} \bowtie \left(\pi_{idb}(\text{BARY}) \setminus \pi_{idb}(\sigma_{\text{sok}='kiwi'}(\text{PODAJA})) \right))}_{\text{identyfikatory osoby bywających w barach, w których nie jest podawany sok kiwi}} \\ \text{identyfikatory osoby bywających TYLKO w barach, w których jest podawany sok kiwi}$$

- Łączymy wyselekcjonowane identyfikatory z krotkami z relacji OSOBY:

$$\underbrace{\text{OSOBY}}_{\text{pełne krotki osób}} \times \underbrace{(\pi_{ido}(\text{OSOBY}) \setminus \pi_{ido}(\text{BYWA} \times (\pi_{idb}(\text{BARY}) \setminus \pi_{idb}(\sigma_{\text{sok}='kiwi'}(\text{PODAJA}))))))}_{\text{identyfikatory osoby bywających TYLKO w barach, w których jest podawany sok kiwi}}$$

pełne krotki osoby bywających TYLKO w barach, w których jest podawany sok kiwi

7.4 Przykład z kwantyfikatorem uniwersalnym

PYTANIE: Osoby, które chodzą do WSZYSTKICH barów, w których jest podawany sok kiwi

W algebrze relacji nie ma operatora wszystko. Proste zapytanie zademonstrowane poniżej wyszukuje osoby (tylko ich identyfikatory), które chodzą do CO NAJMNIEJ JEDNEGO baru, w którym jest podawany sok kiwi (a więc realizuje kwantyfikator egzystencjalny):

$$\pi_{ido}(\text{BYWA} \times \sigma_{\text{sok}='kiwi'}(\text{PODAJA}))$$

Choć wydawać by się mogło, że to krok w dobrym kierunku, niestety tak nie jest. Nie mamy narzędzi, by z powyższego grona wyselekcjonować tych, którzy chadzają do wszystkich barów serwujących sok kiwi.

Sformułujmy więc pytanie inaczej, zastępując słowo "wszystkich" słowami "nie" i "co najmniej jednego". Od **wszystkich osób** odejmijmy te **osoby, które nie chodzą do przynajmniej jednego baru serwującego sok kiwi**. Aby znaleźć osoby, które nie chodzą do przynajmniej jednego baru serwującego sok kiwi, możemy wygenerować **wszystkie pary: (osoba, bar serwujący sok kiwi)** i odjąć od nich **wszystkie pary: (osoba, bar)** z relacji **BYWA**.

- Od wszystkich par: (identyfikator osoby, identyfikator baru serwującego sok kiwi) odejmujemy te, które występują w relacji BYWA.

$$\underbrace{(\underbrace{\pi_{ido}(\text{OSOBY})}_{\text{identyfikatory wszystkich osób}} \times \underbrace{\pi_{idb}(\sigma_{\text{sok}='kiwi'}(\text{PODAJA}))}_{\text{identyfikatory barów, w których serwują sok kiwi}}) \setminus \text{BYWA}}_{\text{wszystkie pary: (ido, idb baru serwującego sok kiwi)}}$$

wszystkie pary: (ido, idb baru serwującego sok kiwi), takie że osoba nie bywa w tym barze

- Ze zbioru wszystkich osób usuniemy osoby, które nie bywają w co najmniej jednym barze serwującym sok kiwi:

$$\underbrace{\pi_{ido}(\text{OSOBY})}_{\text{identyfikatory wszystkich osób}} \setminus \underbrace{\pi_{ido}(\underbrace{(\pi_{ido}(\text{OSOBY}) \times \pi_{idb}(\sigma_{\text{sok}='kiwi'}(\text{PODAJA}))) \setminus \text{BYWA}}_{\text{wszystkie pary: (ido, idb baru serwującego sok kiwi), takie że osoba nie bywa w tym barze}}))}_{\text{identyfikatory osób, które nie bywają w jakimś barze serwującym sok kiwi}}$$

identyfikatory osób, które bywają we wszystkich barach serwujących sok kiwi

- Jeśli zależy nam na pełnych danych (krotkach) osób, to otrzymane wyrażenie musimy złączyć naturalnie z relacją OSOBY, co spowoduje wybranie z tej relacji krotek osób chodzących do wszystkich barów, w których jest podawany sok kiwi:

$$\text{OSOBY} \times \underbrace{(\pi_{ido}(\text{OSOBY}) \setminus \pi_{ido}(\underbrace{(\pi_{ido}(\text{OSOBY}) \times \pi_{idb}(\sigma_{\text{sok}='kiwi'}(\text{PODAJA}))) \setminus \text{BYWA}}_{\text{wszystkie pary: (ido, idb baru serwującego sok kiwi), takie że osoba nie bywa w tym barze}}))}_{\text{identyfikatory osób, które bywają we wszystkich barach serwujących sok kiwi}}}$$

pełne krotki osób, które bywają we wszystkich barach serwujących sok kiwi

7.5 Przykład z zawieraniem zbioru

PYTANIE: Osoby, które chodzą do KAŻDEGO baru, do którego chodzi Puk

Możemy znaleźć osoby, które chodzą do CO NAJMNIEJ JEDNEGO baru, do którego chodzi Puk:

- To są identyfikatory barów, do których chodzi Puk:

$$\pi_{idb}(\sigma_{nazwisko='Puk'}(OSOBA) \bowtie BYWA)$$

- To są identyfikatory osoby, które chodzą do barów, do których chodzi Puk:

$$\pi_{ido}(BYWA \bowtie \pi_{idb}(\sigma_{nazwisko='Puk'}(OSOBA) \bowtie BYWA))$$

Aby znaleźć osoby, które chodzą do KAŻDEGO baru, do którego chodzi Puk, musimy ze zbioru **wszystkich osób** usunąć osoby, które **nie chodzą do co najmniej jednego baru, do którego chodzi Puk**. Ostatnie pytanie też nie jest banalne — nie da się go zadać wprost. Ale możemy wygenerować **wszystkie pary: (identyfikator osoby, identyfikator baru odwiedzanego przez Puka)** i od nich odjąć **pary z relacji BYWA** — w ten sposób pozostaną pary: (ido, idb baru odwiedzanego przez Puka) takie, że osoba o identyfikatorze ido nie bywa w barze o identyfikatorze idb (do którego chodzi Puk).

- Generujemy wszystkie pary: (ido, idb baru odwiedzanego przez Puka)

$$\underbrace{\underbrace{\pi_{ido}(OSOBA)}_{\text{identyfikatory wszystkich osób}} \times \underbrace{\pi_{idb}(\sigma_{nazwisko='Puk'}(OSOBA) \bowtie BYWA)}_{\text{idb barów odwiedzanych przez Puka}}}_{\text{wszystkie pary: ido, idb baru odwiedzanego przez Puka}}$$

- Od wygenerowanych par odejmujemy pary: (ido, idb baru odwiedzanego przez tę osobę), czyli krotki relacji BYWA; wynik tej operacji rzutujemy na atrybut ido, by pozostawić tylko identyfikatory osób i pozbyć się informacji o barze:

$$\underbrace{\pi_{ido}\left(\underbrace{\pi_{ido}(OSOBA) \times \pi_{idb}(\sigma_{nazwisko='Puk'}(OSOBA) \bowtie BYWA)}_{\text{wszystkie pary: ido, idb baru odwiedzanego przez Puka}} \setminus BYWA\right)}_{\text{wszystkie pary: ido, idb baru odwiedzanego przez Puka i nieodwiedzanego przez osobę o identyfikatorze ido}} \\ \underbrace{\hspace{15em}}_{\text{identyfikatory wszystkich osób, które nie chodzą do jakiegos baru odwiedzanego przez Puka}}$$

- Pozostaje od identyfikatorów wszystkich osób odjąć wybrane w poprzednim zapytaniu i otrzymujemy zbiór identyfikatorów wszystkich osób, które nie omijają żadnego baru odwiedzanego przez Puka, czyli chodzą do wszystkich barów odwiedzanych przez Puka:

$$\underbrace{\pi_{ido}(OSOBA)}_{\text{identyfikatory wszystkich osób}} \setminus \underbrace{\pi_{ido}(\pi_{ido}(OSOBA) \times \pi_{idb}(\sigma_{nazwisko='Puk'}(OSOBA) \bowtie BYWA) \setminus BYWA)}_{\text{identyfikatory osób, które nie chodzą do jakiegos baru odwiedzanego przez Puka}} \\ \underbrace{\hspace{15em}}_{\text{identyfikatory osób, które chodzą do każdego baru odwiedzanego przez Puka}}$$

- W ostatnim kroku łączymy zbiór wybranych identyfikatorów z relacją OSOBY otrzymując pełne dane wybranych osób:

OSOBY \bowtie

$$\underbrace{(\pi_{ido}(\text{OSOBA}) \setminus \pi_{ido}(\pi_{ido}(\text{OSOBA}) \times \pi_{idb}(\sigma_{nazwisko='Puka'}(\text{OSOBA}) \bowtie \text{BYWA}) \setminus \text{BYWA}))}_{\text{identyfikatory osób, które chodzą do każdego baru odwiedzanego przez Puka}}$$

7.6 Przykład z wyborem ekstremum

PYTANIE: Bary (tylko ich identyfikatory), w których sok mango jest najdroższy

W algebrze relacji nie ma operatora wyboru maksimum. Wartości największej możemy szukać porównując ją z innymi — powinna być od nich większa (a dokładniej, niemniejsza). Zapytanie zaprezentowane poniżej zwróci nam jednak tylko takie bary, dla których istnieje bar, w którym sok mango jest tańszy, a więc zwróci nam wszystkie bary, w których sok mango ma cenę powyżej minimalnej:

$$\underbrace{\pi_{idb}(\sigma_{c1 < cena}(\underbrace{\pi_{b1,c1}(\sigma_{s1='mango'}(\rho_{P1(b1,s1,c1)}(\text{PODAJA}))}_{\text{pary bar (b1) i cena (c1) soku mango w tym barze}}) \times \underbrace{\pi_{idb,cena}(\sigma_{sok='mango'}(\text{PODAJA}))}_{\text{pary: idb i cena soku mango w tym barze}}))}_{\text{czwórki: b1, c1, idb i cena, w których cena soku mango w b1 (czyli c1) jest niższa niż cena w barze idb (czyli 'cena')}}}_{\text{identyfikatory barów, w których sok mango nie jest najtańszy}}$$

Jak widać, powyższe wyrażenie nie daje odpowiedzi na pytanie o bary, gdzie sok mango jest NAJDROŻSZY, tylko o takie, gdzie NIE JEST NAJTAŃSZY. Choć nie jest to odpowiedź na pytanie, to daje wskazówkę, jak ją uzyskać. Od **wszystkich barów, gdzie jest serwowany sok mango** odejmijmy te bary, w **których sok mango nie jest najdroższy**.

- Identyfikatory barów, w których jest serwowany sok mango:

$$\pi_{idb}(\sigma_{sok='mango'}(\text{PODAJA}))$$

- Identyfikatory barów, w których sok mango nie jest najdroższy (czyli istnieje bar, gdzie ten sok jest droższy):

$$\underbrace{\pi_{idb}(\sigma_{c1 > cena}(\underbrace{\pi_{b1,c1}(\sigma_{s1='mango'}(\rho_{P1(b1,s1,c1)}(\text{PODAJA}))}_{\text{pary identyfikator baru (b1) i cena (c1) soku mango w tym barze}}) \times \underbrace{\pi_{idb,cena}(\sigma_{sok='mango'}(\text{PODAJA}))}_{\text{pary: identyfikator baru i cena soku mango w tym barze}}))}_{\text{czwórki: b1, c1, idb i cena, w których cena soku mango w b1 (czyli c1) jest wyższa niż cena w barze idb (czyli 'cena')}}}_{\text{identyfikatory barów, w których sok mango nie jest najdroższy}}$$

- Odejmując od identyfikatorów wszystkich barów serwujących sok mango identyfikatory tych, w których nie jest on nadroższy, otrzymujemy poszukiwane bary:

$$\underbrace{\pi_{idb}(\sigma_{sok='mango'}(\text{PODAJA}))}_{\text{identyfikatory barów serwujących sok mango}}$$

$$\setminus \underbrace{\pi_{idb}(\sigma_{c1 > cena}(\pi_{b1, c1}(\sigma_{s1 = 'mango'}(\rho_{P1(b1, s1, c1)}(PODAJA)))) \times \pi_{idb, cena}(\sigma_{sok = 'mango'}(PODAJA)))}_{\text{identyfikatory barów, w których sok mango nie jest najdroższy}}$$

$$\underbrace{\hspace{10em}}_{\text{identyfikatory barów, w których sok mango jest najdroższy}}$$