

Relációs algebra 1.rész

Tankönyv: Ullman-Widom:
Adatbázisrendszerek Alapvetés
Második, átdolgozott kiadás,
Panem, 2009



Lekérdezések a relációs modellben

2.4. Egy algebrai lekérdező nyelv

- 01B_RelAlg1alap: alapműveletek folyt.köv.:
- 02A_RelAlg2peldak: lekérdező nyelv itt: Tk. 2.4.1.Termék-feladatai a)-k)

Milyen típusú feladatokat fogalmazhatunk meg?

Például (lásd az 1.előadáson kiadott lapon a konkrét táblákat)

Legyen adott az alábbi **relációs sémák** feletti relációk:

Termék (gyártó, modell, típus)

PC (modell, sebesség, memória, merevlemez, ár)

Laptop (modell, sebesség, memória, merevlemez, képernyő, ár)

Nyomtató (modell, színes, típus, ár)

Feladatok Tk.2.4.1.feladat (ezeket a)-k) kérdéseket konkrét táblák alapján természetes módon meg lehet válaszolni, majd felírjuk relációs algebrában)

a) Melyek azok a PC modellek, amelyek sebessége legalább 3.00?

b) Mely gyártók készítenek legalább egy gigabájt méretű merevlemezzel rendelkező laptopot?

c) Adjuk meg a B gyártó által gyártott összes termék modellszámát és árát típustól függetlenül!

!! i) Melyik gyártó gyártja a leggyorsabb számítógépet (laptopot vagy PC-t)?

!! k) Melyek azok a gyártók, akik pontosan három típusú PC-t forgalmaznak?

Mit nevezünk algebrának?

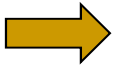
- Nyelv: a kérdés szintaktikai alakja és a kérdés kiértékelése (algoritmus) kiértékelési szemantika
- Algebra **műveleteket** és **atomi operandusokat** tartalmaz.
- **Relációs algebra**: az atomi operandusokon és az algebrai kifejezéseken végzett műveletek alkalmazásával kapott relációkon műveleteket adunk meg, kifejezéseket építünk (a kifejezés felel meg a kérdés szintaktikai alakjának).
- Fontos tehát, hogy **minden művelet végeredménye reláció**, amelyen további műveletek adhatók meg.
- A relációs algebra atomi operandusai a következők:
 - a relációkhoz tartozó **változók**,
 - **konstansok**, amelyek véges relációt fejeznek ki.

Halmazműveletek (jelölése a szokásos)

- Reláció előfordulás véges sok sorból álló halmaz. Így értelmezhetők a szokásos halmazműveletek: az **unió** (az eredmény halmaz, csak egyszer szerepel egy sor) értelmezhető a **metszet** és a **különbség**. Milyen művelet van még halmazokon? Értelmezhető-e relációkon?
- R, S és azonos típusú, $R \cup S$ és $R - S$ típusa ugyanez
 $R \cup S := \{t \mid t \in R \vee t \in S\}$, $R - S := \{t \mid t \in R \wedge t \notin S\}$
- Az alpműveletekhez az **unió** és **különbség** tartozik, **metszet** műveletet származtatjuk $R \cap S = R - (R - S)$

- | A | B | C |
|---|---|---|
| a | b | c |
| c | d | e |
| g | a | d |

A	B	C
a	b	c
c	d	e
g	d	f

Példa: különbségre
 $R - S$ 

A	B	C
g	a	d

Vetítés (project, jelölése π)

- **Vetítés** (projekció). Adott relációt vetít le az alsó indexben szereplő attribútumokra (attribútumok számát csökkentik)
- $\Pi_{\text{lista}}(R)$ ahol lista: $\{A_{i_1}, \dots, A_{i_k}\}$ R-sémájában levő attribútumok egy részhalmazának felsorolása
eredmény típusa $\langle A_{i_1} : \text{értéktípus}_{i_1}, \dots, A_{i_k} : \text{értéktípus}_{i_k} \rangle$
 $\Pi_{\text{lista}}(R) := \{ t.A_{i_1}, t.A_{i_2}, \dots, t.A_{i_k} \mid t \in R \} = \{ t[\text{lista}] \mid t \in R \}$
- Reláció soraiból kiválasztja az attribútumoknak megfelelő $A_{i_1}, \dots, A_{i_k}$ -n előforduló értékeket, ha többször előfordul akkor a duplikátumokat kiszűrjük (hogy halmazt kapjunk)

■ **Példa:**

A	B	C
a	b	c
c	d	e
c	d	d

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

➔

A	B
a	b
c	d

Kiválasztás (select, jelölése: kis-szigma)

- **Kiválasztás** (szűrés). Kiválasztja az argumentumban szereplő reláció azon sorait, amelyek eleget tesznek az alsó indexben szereplő feltételnek.
- $\sigma_{\text{Feltétel}}(R)$ és R sémája megegyezik
- $\sigma_{\text{Feltétel}}(\mathbf{R}) := \{ t \mid t \in R \text{ és } t \text{ kielégíti az } F \text{ feltételt} \}$
- $R(A_1, \dots, A_n)$ séma feletti reláció esetén a σ_F kiválasztás F feltétele a következőképpen épül fel:
 - **elemi feltétel**: $A_i \theta A_j$, $A_i \theta c$, ahol c konstans, θ pedig $=, \neq, <, >, \leq, \geq$
 - **összetett feltétel**: ha B_1, B_2 feltételek, akkor $\neg B_1$, $B_1 \wedge B_2$, $B_1 \vee B_2$ és zárójelezésekkel is feltételek

- **Példa:**

A	B	C
a	b	c
c	d	e
g	a	d

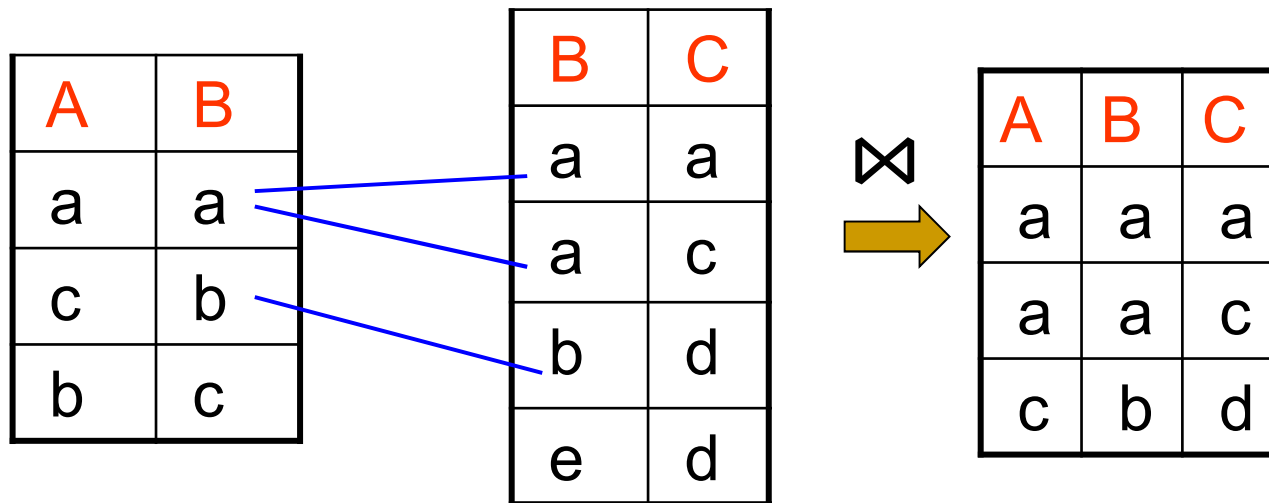
$\sigma_{A=a \vee C=d}(R)$



A	B	C
a	b	c
g	a	d

Természetes összekapcsolás ---1

- **Szorzás jellegű műveletek** (attribútumok számát növeli) többféle lehetőség, amelyekből csak egyik alpművelet:
- Angolul: Natural Join (jelölése: „csokornyakkendő”)
- **Természetes összekapcsolás**: közös attribútum-nevekre épül. $R \bowtie S$ azon sorpárokat tartalmazza R-ből illetve S-ből, amelyek R és S azonos attribútumain megegyeznek.



Természetes összekapcsolás ---2

- **Természetes összekapcsolás:**
- Legyen $R(A_1, \dots, A_k, B_1, \dots, B_n)$, illetve $S(B_1, \dots, B_n, C_1, \dots, C_m)$
- $R \bowtie S$ típusa $(A_1, \dots, A_k, B_1, \dots, B_n, C_1, \dots, C_m)$ vagyis a két attribútum-halmaz uniója
- $R \bowtie S = \{ \langle A_1: t(A_1), \dots, A_k: t(A_k), B_1: t(B_1), \dots, B_n: t(B_n), C_1: s(C_1), \dots, C_m: s(C_m) \rangle \mid t \in R, s \in S, t(B_i) = s(B_i) \ i=1, \dots, n \}$
- $R \bowtie S$ elemei $v \in R \bowtie S$
$$R \bowtie S = \{ v \mid \exists t \in R, \exists s \in S: t[B_1, \dots, B_n] = s[B_1, \dots, B_n] \wedge v[A_1, \dots, A_k] = t[A_1, \dots, A_k] \wedge v[B_1, \dots, B_n] = t[B_1, \dots, B_n] \wedge v[C_1, \dots, C_m] = s[C_1, \dots, C_m] \}$$

Természetes összekapcsolás ---3

- **Példákban:** két azonos nevű attribútumot úgy tekintünk, hogy ugyanazt jelenti és a közös érték alapján fűzzük össze a sorokat.
- **Milyen problémák lehetnek?**
- Filmek adatbázisban ugyanarra a tulajdonságra más névvel hivatkozunk: Filmek.év és SzerepelBenne.filmÉv, illetve FilmSzínész.név és SzerepelBenne.színészNév
- Termékek adatbázisban pedig ugyanaz az azonosító mást jelent: Termék.típus más, mint Nyomtató.típus
- Emiatt a Filmek és a Termékek adatbázisokban ahhoz, hogy jól működjön az összekapcsolás **szükségünk van** egy technikai műveletre, és ez: **az átnevezés (rename)**

Átnevezés (rename, jelölése: ρ)

- Miért van erre szükség? Nem tudjuk a reláció saját magával való szorzatát kifejezni, $R \bowtie R = R$ lesz.
- Láttuk, hogy egyes esetekben szükség lehet relációnak vagy a reláció attribútumainak **átnevezésére**:

$$\rho_{T(B_1, \dots, B_k)}(R(A_1, \dots, A_k))$$

- Ha az attribútumokat nem szeretnénk átnevezni, csak a relációt, ezt $\rho_T(R)$ -rel jelöljük. Ha ugyanazt a táblát használjuk többször, akkor a táblának adunk másik hivatkozási (alias) nevet.
- Az attribútumok átnevezése helyett alternatíva: $R.A$ (vagyis relációnév.attribútumnév hivatkozás) amivel meg tudjuk különböztetni a különböző táblákból származó azonos nevű attribútumokat.

A relációs algebra hat alapművelete

- $R \cup S$ unió
- $R - S$ különbség
- $\Pi_{\text{lista}}(R)$ vetítés
- $\sigma_{\text{Feltétel}}(R)$ kiválasztás
- $R \bowtie S$ természetes összekapcsolás
- $\rho_{T(B_1, \dots, B_k)}(R(A_1, \dots, A_k))$ átnevezés

Minimális készlet, vagyis bármelyiket elhagyva az a többivel nem fejezhető ki. A halmazműveletekből miért nem az unió és metszetet, ehelyett az unió és különbséget választottuk ki?

A különbség nem monoton, ha új sorokat veszünk be valamelyik táblába, ettől az eredmény nem biztos, hogy növekedni fog. Az összes többi alapművelet monoton (plusz még a metszet is).

Szorzás jellegű műveletek ---1

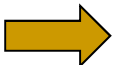
- Szorzás jellegű műveletek többféle lehetőség, az egyik: **Természetes összekapcsolás**: közös attribútumnevekre épül. $R \bowtie S$ azon sorpárokat tartalmazza R -ből illetve S -ből, amelyek R és S azonos attribútumain megegyeznek.
- Egy másik lehetőség: **direkt-szorzat (Descartes-szorzat)** Ezt is tekintjük alpműveletnek (és bizonyos esetekben egyszerűbb ezt venni alpműveletnek) az ennél sokkal gyakrabban használt **természetes összekapcsolás** helyett.
- **$R \times S$** : az R és S minden sora párban összefűződik, az első tábla minden sorához hozzáfűzzük a második tábla minden sorát

$$R \times S := \{ t \mid t[R] \in R \text{ és } t[S] \in S \}$$

Szorzás jellegű műveletek ---2

- A **direkt-szorzat** (vagy szorzat, **Descartes-szorzat**) esetén természetesen nem fontos az attribútumok egyenlősége. A két vagy több reláció azonos nevű attribútumait azonban meg kell különböztetni egymástól. Hivatkozás séma: oszlopok átnevezése illetve azonos nevű oszlop esetén: $R.A_1, \dots, R.A_k, S.A_1, \dots, S.A_k$

- **Példa:**

A	B	C		B	D						
a	b	c		b	r		a	b	c	q	s
c	d	e		q	s		c	d	e	b	r
g	a	d					c	d	e	q	s
							g	a	d	b	r
							g	a	d	q	s

Szorzás jellegű műveletek ---3

- Ha R, S **sémái megegyeznek**, akkor $R \bowtie S = R \cap S$.
- Ha R, S sémáiban **nincs közös attribútum**, akkor $R \bowtie S = R \times S$.
- Később nézünk még további szorzás jellegű műveletet:
Théta összekapcsolás $\bowtie_{\theta}$, félig összekapcsolás $\ltimes$, és a rel.algebra kiterjesztésénél külső összekapcsolásokat.
- Hogyan fejezhető ki az $R \times S$ **direkt szorzat** relációs algebrában? (ha a **természetes összekapcsolást** tekintjük alpműveletnek, ebből és az átnevezés segítségével felírható a direkt szorzat).
- Hogyan fejezhető ki a **természetes összekapcsolás**, ha a **direkt szorzatot** soroljuk az alpműveletek közé? Lásd köv.lapon (kiválasztás és a vetítés segítségével)

Műveletek közötti kapcsolatok

Felszolgál

kocsmá	sör
Makk 7-es	Dreher
Lórúgás	Kozel
Lórúgás	Gösser



Látogat

név	kocsmá
Péter	Makk 7-es
Feri	Lórúgás



kocsmá	sör	név
Makk 7-es	Dreher	Péter
Lórúgás	Kozel	Feri
Lórúgás	Gösser	Feri

A természetes

összekapcsolás

kifejezhető a direkt

szorzat, kiválasztás,

vetítés műveletekkel:

$$R \bowtie S \equiv \pi_L(\sigma_C(R \times S)),$$

itt: **C** a közös

attribútumok

egyenlőségét írja elő,

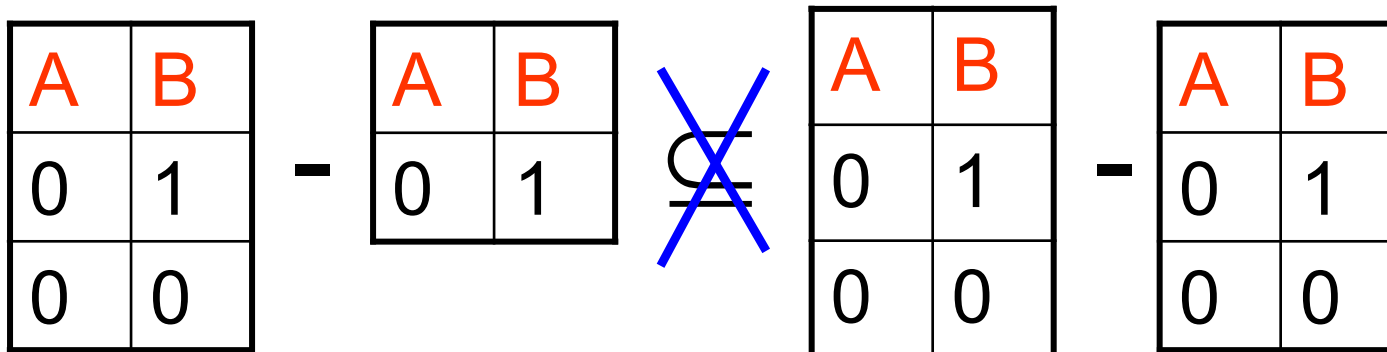
L pedig csak egyszer

veszi fel a közös

attribútumokat.

Monotonitás ---1

- **Monoton nem csökkenő** (röviden **monoton**) kifejezés: bővebb relációra alkalmazva az eredmény is bővebb:
Ha $R_i \subseteq S_i$, $i=1, \dots, n$, akkor $E(R_1, \dots, R_n) \subseteq E(S_1, \dots, S_n)$.
- A kivonás kivétel az alpműveletek monoton műveletek (ha az alpműveletet közül elhagyjuk a különbséget az monoton relációs algebra, lásd később a Datalognál is).



Monotonitás ---2

- **DE:** Monoton kifejezésben is szerepelhet kivonás:
 $R \cap S = R - (R - S)$ monoton, uis a metszet monoton.
- Ha E, E_1, E_k monoton kifejezések, és $E(E_1(\dots), \dots, E_k(\dots))$ helyes kifejezés, akkor monoton is.
- Következmény: **A kivonás nem fejezhető ki a többi alpművelettel** (független a többi művelettől).
- A relációs algebra hat alpművelete minimális készlet, vagyis bármelyiket elhagyva az a többivel nem fejezhető ki. Hogyan tudná ezt a többi műveletre is belátni? Vagyis milyen sajátos tulajdonsága van az egyes műveleteknek, ami miatt azt elhagyva a többivel nem állítható elő?

Relációs algebra hátrányai

- Az így bevezetett természetes összekapcsolásnak vannak azért hátrányai...

Például:

- Személyek (Szülő, Gyermeke) adatbázis felhasználásával szeretnénk elkészíteni (Nagyszülő, Unoka) párok listáját.
- Hogyan végezzük el? (átnevezés után vesszük a reláció önmagával vett direkt szorzatát, majd kiválasztjuk a megfelelő sorokat és vetítjük a megfelelő oszlopokra) (lásd a következő héten nézzünk ilyen jellegű példákat)
- Mi van ha az összes (Ős, Utód) párokat kellene megkapni? Ezt már **nem tudjuk relációs algebrában** kifejezni (később lesz Datalog és PL/SQL programmal)