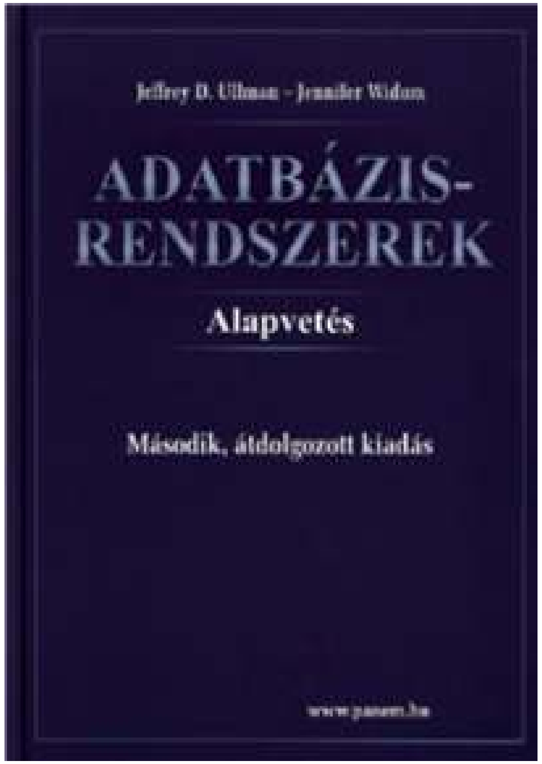
# Relációs algebra 1.rész alapok

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




Lekérdezések a relációs modellben
2.4. Egy algebrai lekérdező nyelv,
relációs algebrai műveletek

Jön: Relációs algebra 2.részben:
a kezdetekhez való példák: Tk.2.4.1.Termék lekérdezések,
4.részben haladó példák: pl.Tk.2.4.8. és Tk.2.4.10.feladatai

# Milyen típusú feladatokat fogalmazhatunk meg?

Legyen adott az alábbi relációs sémák feletti relációk:

Termék (gyártó, <u>modell</u>, típus)

PC (<u>modell</u>, sebesség, memória, merevlemez, ár)

Laptop (<u>modell</u>, sebesség, memória, merevlemez, képernyő, ár)

Nyomtató (<u>modell</u>, színes, típus, ár)

Feladatok Tk.2.4.1.feladat (ezeket a 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 száz gigabájt méretű merevlemezzel rendelkező laptopot?

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

stb…

!! 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.

# Relációs algebrai lekérdező nyelv ---1

Relációs algebrai kifejezés, mint lekérdező nyelv

Lekérdező nyelv: L -nyelv

Adott az adatbázis sémája: $\mathbb{R} = \{R_1, \ldots, R_k\}$

$q \in L \qquad q: R_1, \ldots, R_k \rightarrow V$ (eredmény-reláció)

E - relációs algebrai kifejezés: $E(R_1, \ldots, R_k) = V$ (output)

Relációs algebrai kifejezések formális felépítése

- Elemi kifejezések (alapkifejezések)

  (i) $R_i \in \mathbb{R}$ (az adatbázis-sémában levő relációnevek)

  $R_i$ kiértékelése: az aktuális előfordulása

  (ii) konstans reláció (véges sok, konstansból álló sor)
- Összetett kifejezések (folyt. köv.oldalon)

# Relációs algebrai lekérdező nyelv ---2

(folyt.) Relációs algebrai kifejezések felépítése

- Összetett kifejezések
- Ha $E_1$, $E_2$ kifejezések, akkor a következő E is kifejezés
  - $E := E_1 \cup E_2$ unió, ha azonos típusúak (és ez a típusa)
  - $E := E_1 - E_2$ különbség, ha $E_1$, $E_2$ azonos típusúak (típus)
  - $E := \Pi_{lista}( E_1 )$ vetítés (típus a lista szerint)
  - $E := \sigma_{Feltétel}( E_1 )$ kiválasztás (típus nem változik)
  - $E := E_1 \bowtie E_2$ term. összekapcsolás (típus attr-ok uniója)
  - $E := \rho_{S(B1, \ldots, Bk)}(E_1 (A_1, \ldots A_k))$ átnevezés (típ.új attr.nevek)
  - $E := ( E_1 )$ kifejezést zárójelezve is kifejezést kapunk
- Ezek és csak ezek a kifejezések, amit így meg tudunk adni

# 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 alapműveletekhez 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 pí: $\Pi$)

- Vetítés (projekció). Adott relációt vetít le az alsó indexben szereplő attribútumokra (attribútumok számát csökkentik)
- $\Pi_{lista}(R)$ ahol lista: $\{A_{i_1}, \ldots, A_{i_k}\}$ R-sémájában levő attribútumok egy részhalmazának felsorolása

  eredmény típusa $<A_{i_1}$: értéktípus$_{i_1}, \ldots, A_{i_k}$:értéktípus$_{i_k}>$

  $\Pi_{lista}(R) := \{ t.A_{i_1}, t.A_{i_2}, \ldots, t.A_{i_k} \mid t \in R\} = \{ t[lista] \mid t \in R\}$
- Reláció soraiból kiválasztja az attribútumoknak megfelelő $A_{i_1}, \ldots, A_{i_k}$ -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 |

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

| A | B |
|---|---|
| a | b |
| c | d |

# Kiválasztás (select, jelölése szigma: σ)

- Kiválasztás (szűrés). Kiválasztja az argumentumban szereplő reláció azon sorait, amelyek eleget tesznek az alsó indexben szereplő feltételnek.
- $\sigma_{Feltétel}(R)$ és R sémája megegyezik
- $\sigma_{Feltétel}(R) := \{ t \mid t \in R$ és t kielégíti az F feltételt$\}$
- $R(A_1, \ldots, A_n)$ séma feletti reláció esetén a $\sigma_F$ 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 alapművelet:
- 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.

| A | B |
|---|---|
| a | a |
| c | b |
| b | c |

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

$\bowtie$

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

# Természetes összekapcsolás ---2

- Természetes összekapcsolás:
- Legyen $R(A_1,\ldots,A_k,B_1,\ldots,B_n)$, illetve $S(B_1,\ldots,B_n,C_1,\ldots,C_m)$
- $R \bowtie S$ típusa $(A_1,\ldots,A_k,B_1,\ldots,B_n,C_1,\ldots,C_m)$ vagyis a két attribútum-halmaz uniója
- $R \bowtie S = \{ <A_1: t(A_1),\ldots, A_k: t(A_k), B_1: t(B_1),\ldots, B_n: t(B_n),$
  $C_1: s(C_1),\ldots, C_m: s(C_m)> \mid t \in R, s \in S,$
  $t(B_i) = s(B_i)$ i=1, …, n $\}$
- $R \bowtie S$ elemei $v \in R \bowtie S$

  $R \bowtie S = \{ v \mid \exists\, t \in R, \exists\, s \in S: t[B_1,\ldots,B_n] = s[B_1,\ldots,B_n] \wedge$
  $\wedge\ v[A_1,\ldots,A_k] = t[A_1,\ldots,A_k] \wedge v[B_1,\ldots,B_n] = t[B_1,\ldots,B_n] \wedge$
  $\wedge\ v[C_1,\ldots,C_m] = s[C_1,\ldots,C_m] \}$

# 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 ró: $\varrho$)

- 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:

  $\rho_{T(B_1, \ldots, B_k)}(R(A_1, \ldots A_k))$
- 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.

# Szorzás jellegű műveletek ---1

- Szorzás jellegű műveletek többféle lehetősége közül csak az egyiket vesszük alapműveletnek: join vagy természetes összekapcsolást, amely közös attribútumnevekre épül. R ⋈ S azon sorpárokat tartalmazza R-ből illetve S-ből, amelyek R és S azonos attribútumain megegyeznek.
- Egy másik lehetőség: direkt-szorzat (Descartes-szorzat) Ez is tekinthető alapműveletnek (és bizonyos esetekben egyszerűbb ezt venni alapműveletnek) 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: $R.A_1, \ldots, R.A_k, S.A_1, \ldots, S.A_k$
- **Példa:**

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

| B | D |
|---|---|
| b | r |
| q | s |

$R \times S$

| A | R.B | C | S.B | D |
|---|---|---|---|---|
| a | b | c | b | r |
| a | b | c | q | s |
| c | d | e | b | r |
| 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 x S direkt szorzat relációs algebrában? (ha a természetes összekapcsolást tekintjük alapműveletnek, ebből és az átnevezés segítségével felírható a direkt szorzat).
- Hogyan fejezhető ki a természetes összekapcsolás, ha a direkt szorzatot soroljuk az alapműveletek közé?

# Szorzás jellegű műveletek ---4

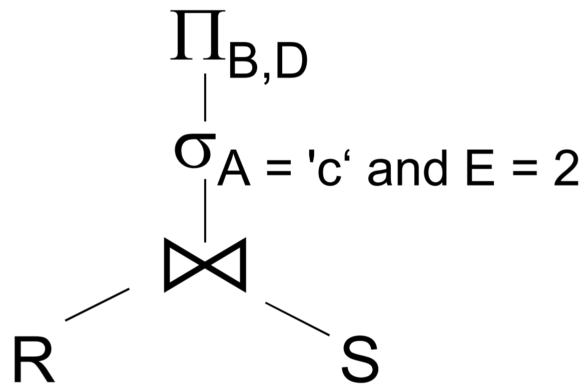
- További szorzás jellegű műveletek:
- Théta összekapcsolás $\bowtie_{\theta}$
- Félig összekapcsolás ⋉ („lógó sorok” kifejezése)
- **Tankönyv !2.4.8. feladata:** Az *R* és *S* relációk félig-összekapcsolása, *R* ⋉ *S* az *R* azon sorainak halmaza, amely sorok megegyeznek az *S* legalább egy sorával az *R* és *S* összes közös attribútumán. Adjunk meg olyan három különböző relációs algebrai kifejezést, amelyek ekvivalensek az *R* ⋉ *S* kifejezéssel.

# Lekérdezések kifejezése algebrában ---1

- Kifejezés kiértékelése: összetett kifejezést kívülről befelé haladva átírjuk kiértékelő fává, levelek: elemi kifejezések.
- A relációs algebra procedurális nyelv, vagyis nemcsak azt adjuk meg, hogy mit csináljunk, hanem azt is hogyan.
- Legyen R, S az R(A, B, C), S(C, D, E) séma feletti reláció

$\Pi_{B,D}\ \sigma_{A = 'c' \text{ and } E = 2}\ (R \bowtie S)$

- Ehhez a kiértékelő fa: (kiértékelése alulról felfelé történik)



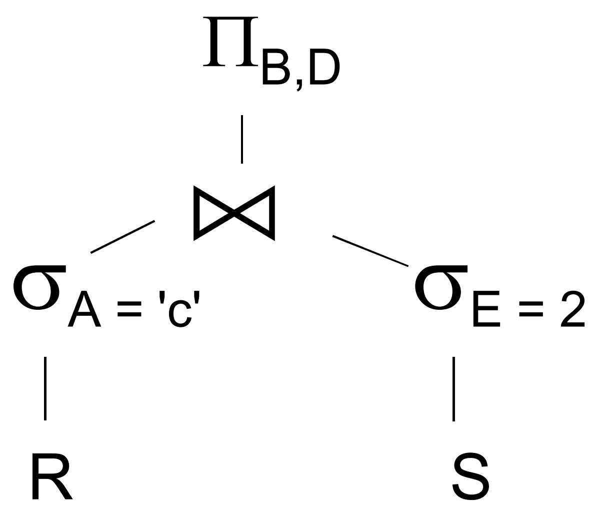

- Tudunk-e ennél jobb, hatékonyabb megoldást találni?

# Lekérdezések kifejezése algebrában ---2

- Ekvivalens átalakítási lehetőségekkel, relációs algebrai azonosságokkal át tudjuk alakítani a fentivel ekvivalens másik relációs algebrai kifejezésre. Hatékonyabb-e?

$$\Pi_{B,D}\left(\sigma_{A='c'}(R) \bowtie \sigma_{E=2}(S)\right)$$

- Ehhez is felrajzolva a kiértékelő fát:



# Lekérdezések kifejezése algebrában ---3

- Ekvivalens átalakítás: oly módon alakítjuk át a kifejezést, hogy az adatbázis minden lehetséges előfordulására (vagyis ugyanazokra a táblákra) minden esetben ugyanazt az eredményt (ugyanazt a táblát) adja a két kiértékelő fa.
- Adatbázisok-2 tárgyból lesznek az **ekvivalens átalakítási szabályok**, a szabály alapú optimalizálás első szabálya például, hogy a kiválasztási műveletet minél előbb kell végrehajtani (közbülső táblák lehetőleg kicsik legyenek)
- Egy-egy részkifejezést, ha gyakran használjuk, akkor új változóval láthatjuk el, segédváltozót vezethetünk be:

  $T(C_1, \ldots C_n) := E(A_1, \ldots A_n)$, de a legvégén a bevezetett változók helyére be kell másolni a részkifejezést.