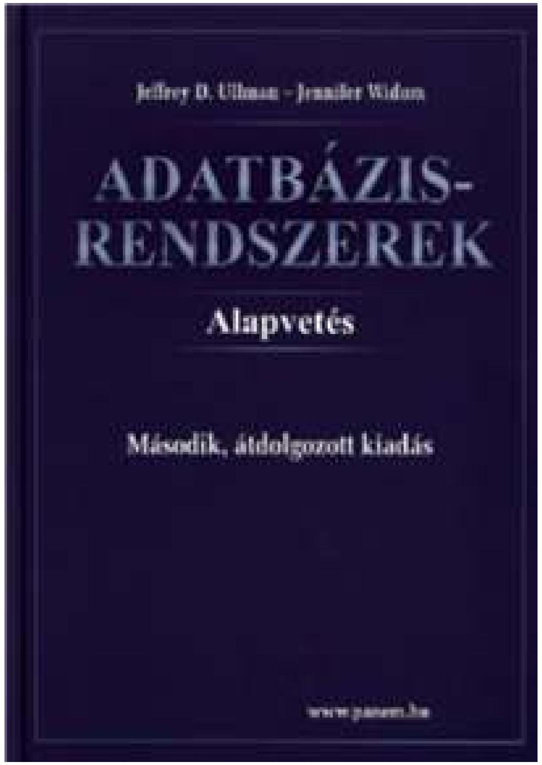
Relációs algebra

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



2.4. Relációs algebra
Példák: Tk.Termékek feladatai

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
  - $E := \Pi_{lista}(E_1)$ vetítés (típus a lista szerint)
  - $E := \sigma_{Feltétel}(E_1)$ kiválasztás (típus nem változik)
  - $E := E_1 \cup E_2$ unió, ha azonos típusúak (és ez a típusa)
  - $E := E_1 - E_2$ különbség, ha E_1, E_2 azonos típusúak (típus)
  - $E := E_1 \bowtie E_2$ 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

Vetítés (project, jelölése pí: Π)

- 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

$\Pi_{A, B}(R)$ ⇒

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.
- $\sigma_{\text{Feltétel}}(R)$ és R sémája megegyezik
- $\sigma_{\text{Feltétel}}(R) := \{\, t \mid t \in R$ és t kielégíti az F feltételt$\}$
- $R(A_1, \ldots, A_n)$ 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

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

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 ⋈ 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: ϱ)

- Miért van erre szükség? Nem tudjuk a reláció saját magával való szorzatát kifejezni, R ⋈ 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. $R \bowtie S$ 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

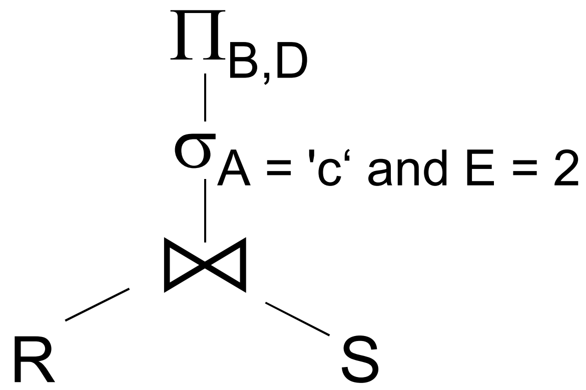
- Ha R, S sémái megegyeznek, akkor $R \bowtie S = R \cap S$.
- Ha R, S sémáiban nincs közös attribútum, akkor
  $R \bowtie S = R \times S$.
- Később nézünk még további szorzás jellegű műveletet:
  Théta összekapcsolás $\bowtie_\theta$, félig összekapcsolás $\ltimes$, és a rel.algebra kiterjesztésénél 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?

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)



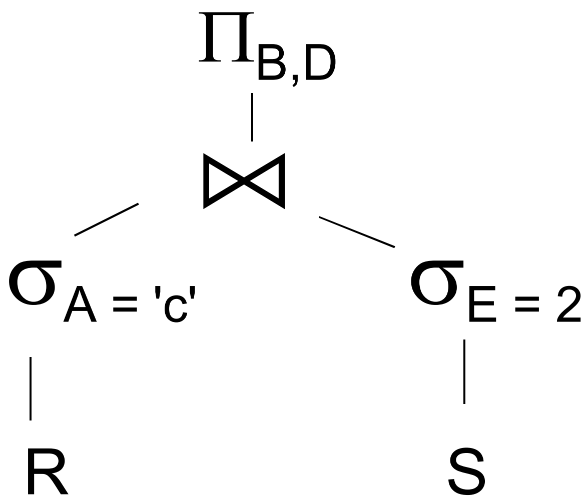
- Tudunk-e ennél jobb, hatékonyabb megoldást találni?

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 bármilyen is a táblák tartalma) minden esetben ugyanazt az eredményt (vagyis ugyanazt az output táblát) adja az eredeti és az átalakított 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)
- Ha 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.

Példák rel.algebrai kifejezések átírása

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)

- Tk.2.4.14. (54-57.o.) 2.4.1.feladata Termékek feladatai először relációs algebrában táblákkal gondolkodva felírva kifejezőfákkal, majd átírva SQL lekérdezésekre többféle megoldási lehetőséget vizsgáljunk meg, vessünk össze

Gépes környezetben való kipróbáláshoz create table:
http://people.inf.elte.hu/sila/eduAB/create_termekek.txt

Példák rel.algebrai kifejezések átírása

- Relációs algebra kifejezések ilyen bevezetése valóban használható a lekérdezések megadására?
- Tk.2.4.1.feladat
- **Példa:** Adottak az alábbi relációs sémák feletti relációk

Termék (gyártó, <u>modell</u>, típus)

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

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

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

- Jelölje: T(gy, m, t)
  PC(m, s, me, ml, ár)
  L(m, s, me, ml, k, ár)
  Ny(m, sz, t, ár)

Megj.: a két típus attr.név nem ugyanazt fejezi ki és így T $\bowtie$ Ny természetes összekapcsolásnál „zűr”

Példák rel.algebrai kif. átírása (a.)

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

Relációs algebrai kifejezéssel:

$\Pi_m(\sigma_{s \geq 3.00}(PC))$

Kifejezésfával:

$$\begin{array}{l} \Pi_m \\ | \\ \sigma_{s >= 3.00} \\ | \\ PC \end{array}$$

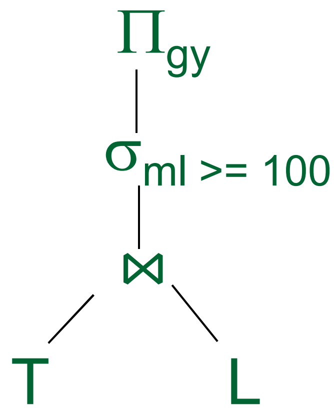
SQL SELECT utasításával:

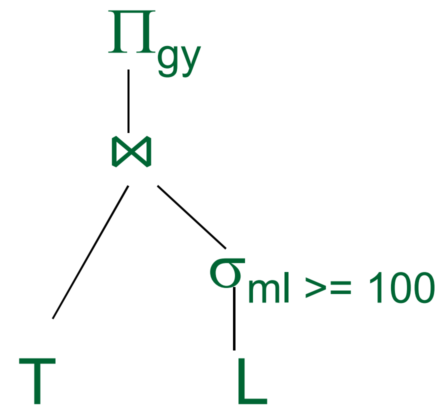
```
SELECT modell
  FROM PC
  WHERE sebesség>=3.00;
```

Példák rel.algebrai kif. átírása (b.)

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

Π_{gy} ($\sigma_{ml\geq 100}$ (T ⋈ L)) vagy ekv. Π_{gy}(T ⋈ ($\sigma_{ml\geq 100}$(L))






```
SELECT gyarto
    FROM Termek T, Laptop L
  WHERE T.modell=L.modell
      AND merevlemez>=100
```

Példák rel.algebrai kif. átírása (c.)

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

(Nyomtató táblánál típus attr. mást jelent, kezelni kell $\bowtie$)
-- segédváltozót vezetek be, legyen **BT :=** $\prod_{m} \sigma_{gy='B'}$**(T)**

$\prod_{m,\, ár}(BT \bowtie PC) \cup \prod_{m,\, ár}(BT \bowtie Laptop) \cup$

$\cup \prod_{m,\, ár}(BT \bowtie Ny)$

```
with BT as
(select modell from termek where gyarto='B')
 select modell, ar from pc natural join BT
        union
 select modell, ar from laptop natural join BT
        union
 select modell, ar from nyomtato natural join BT;
```

Példák rel.algebrai kif. átírása (d.)

d.) Adjuk meg valamennyi színes lézernyomtató modellszámát: $\prod_m(\sigma_{sz='i'}(Ny)) \cap \prod_m(\sigma_{t='lézer'}(Ny))$

-- elvégezhető más módon is: $\prod_m(\sigma_{sz='i' \wedge t='lézer'}(Ny)) =$

$= \prod_m(\sigma_{sz='i'}\, \sigma_{t='lézer'}(Ny)) = \prod_m(\sigma_{t='lézer'}\, \sigma_{sz='i'}(Ny))$

Példák rel.algebrai kif. átírása (e.)

e) Melyek azok a gyártók, amelyek laptopot árulnak, PC-t viszont nem? (ha laptop gyártó több pc-t gyárt, akkor az eredménytábla csökken, nem monoton művelet: R - S)

$$\Pi_{gy}(T \bowtie L) - \Pi_{gy}(T \bowtie PC)$$

Példák rel.algebrai kif. átírása (f.)

! f) Melyek azok a merevlemezméretek, amelyek legalább két PC-ben megtalálhatók? (táblát önmagával szorozzuk)

-- segédváltozót vezetek be, legyen **PC_1 := PC**

$\Pi_{PC.ml}(\sigma_{PC_1.m \neq PC.m \wedge PC_1.ml = PC.ml}(PC_1 \times PC))$

Példák rel.algebrai kif. átírása (g.)

! g) Adjuk meg azokat a PC-modell párokat, amelyek ugyanolyan gyorsak és a memóriájuk is ugyanakkora. Egy pár csak egyszer jelenjen meg, azaz ha már szerepel az (i, j), akkor a (j, i) ne jelenjen meg.

$$\Pi_{PC_1.m,\ PC.m}(\sigma_{PC_1.m<PC.m\ \wedge\ PC_1.s=PC.s\ \wedge\ PC_1.me=PC.me}(PC_1 \times PC))$$

Példák rel.algebrai kif. átírása (h.)

!! h) Melyek azok a gyártók, amelyek gyártanak legalább két, egymástól különböző, legalább 2.80 gigahertzen működő számítógépet (PC-t vagy laptopot)

-- segédváltozó: **Gyors** $:= \prod_{m}(\sigma_{s\geq 2.8}(PC)) \cup \prod_{m}(\sigma_{s\geq 2.8}(L))$

-- és ezzel legyen: $T_1 := T \bowtie Gyors$ és $T_2 := T \bowtie Gyors$

$$\prod_{T_1.gy}(\sigma_{T_1.gy = T2.gy \wedge T_1.m \neq T_2.m}(T_1 \times T_2))$$

Példák rel.algebrai kif. átírása (i.)

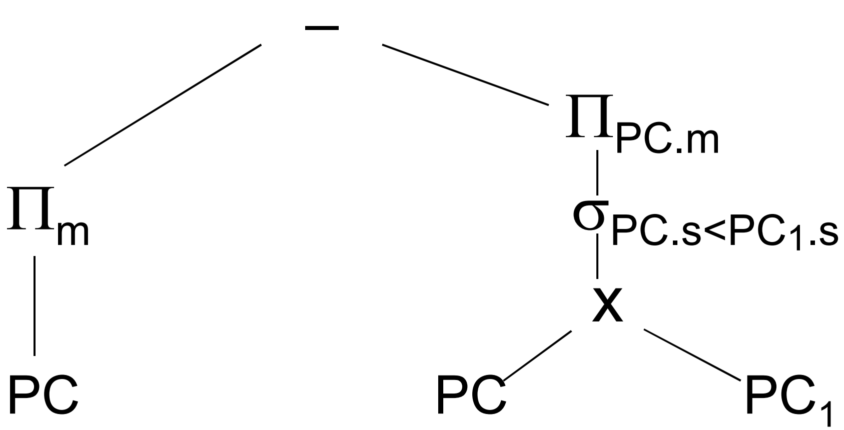
!! i) Melyik gyártó gyártja a leggyorsabb számítógépet (PC-t vagy laptopot)? Lásd még az „elhagyás” típusú lekérdezéseket (köv.oldalon pl. maximum kifejezése)

Kiválasztjuk azokat a PC-ket, amelyiknél van gyorsabb, ha ezt kivonjuk a PC-ékből megkapjuk a leggyorsabbat:

EnnélVanNagyobb = $\Pi_{PC.m}(\sigma_{PC.s<PC_1.s}(PC \times PC1))$

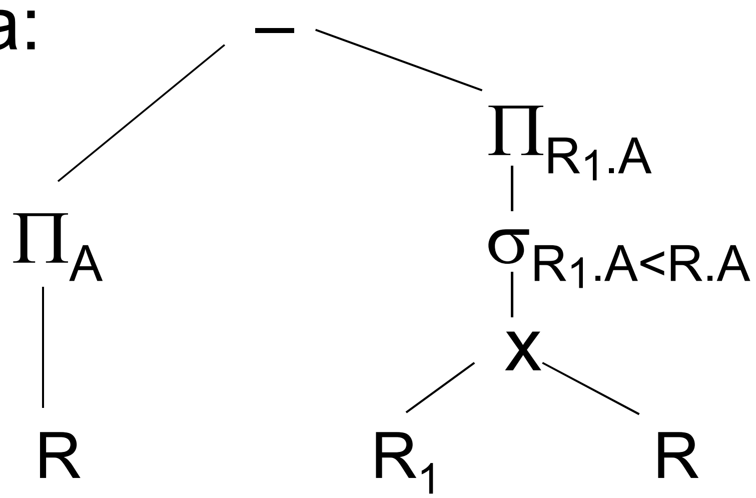
Leggyorsabb: $\Pi_m(PC)$ – **EnnélVanNagyobb**

Ehhez rajzoljuk fel a kiértékelő fát is: (folyt.: PC helyett számítógép kell és a válaszban is a gyártó kell…)



Példa: MAX előállítása rel.algebrában

- Nézzük meg a maximum előállításának a kérdését! Legyen R(A,B). Feladat: Adjuk meg MAX(A) értékét! (Ez majd átvezet az új témára, aggregáló függvényekre, illetve csoportosításra).
- $\pi_A(R) - \pi_{R1.A}(\sigma_{R1.A<R.A}(\rho_{R1}(R) \times R))$
- Kiértékelő fa:



Példa: Rel.alg. kifejezés átírása SQL

- Előző oldal folyt.max előállítás átírása SQL-re:
- Kiértékelő fa szerinti átírás SQL-be:

```
(SELECT A FROM R)
      EXCEPT
(SELECT R1.A AS A
      FROM R R1, R R2
      WHERE R1.A<R2.A);
```

- Nézzük meg korrelált (függő) alkérdéssel is:

```
SELECT A FROM R MAXA
      WHERE NOT EXISTS
            (SELECT A FROM R
              WHERE A > MAXA.A);
```

Példák rel.algebrai kif. átírása (j.)

!! j) Melyik gyártó gyárt legalább három, különböző sebességű PC-t? mint a legalább kettő, csak ott 2x, itt 3x kell a táblát önmagával szorozni. Legyenek **S, S_1, S_2 := T ⋈ $\prod_{m,s}$(PC)**

$$\prod_{S.gy}(\sigma_{S_1.gy=S.gy \wedge S_2.gy=S.gy \wedge S_1.s\neq S.s \wedge S_2.s\neq S.s \wedge S_1.s\neq S_2.s}(S \times S_1 \times S_2))$$

!! k) Melyek azok a gyártók, amelyek pontosan három típusú PC-t forgalmaznak? legalább 3-ból - legalább 4-t kivonni

Példák rel.algebrai kifejezések átírása

- Mire érdemes felhívni a figyelmet?

  Mi a leggyakrabban előforduló típus, amiből építkezek?

  $\Pi_{\textbf{lista}}(\sigma_{\textbf{feltétel}}(\textbf{táblák szorzata}))$

  Ezt a komponenst támogatja legerősebbben majd az SQL:

  SELECT s-lista FROM f-lista WHERE feltétel;

Kérdés/Válasz

- Köszönöm a figyelmet! Kérdés/Válasz?

- Tk.2.4.14. (54-57.o.) 2.4.1.feladata Termékek feladatai először relációs algebrában táblákkal gondolkodva felírva kifejezőfákkal, majd átírva SQL lekérdezésekre többféle megoldási lehetőséget vizsgáljunk meg, vessünk össze create table: http://people.inf.elte.hu/sila/eduAB/create_termekek.txt