Adatbázisok elmélete 7. előadás

Csima Judit
Budapesti Műszaki és Gazdaságtudományi Egyetem
Számítástudományi Tsz.
I. B. 136/b
csima@cs.bme.hu

2003. Március 5.

Származtatott műveletek

Bal (jobb) félillesztés

- $R(A_1,\ldots,A_k,B_1,\ldots,B_r), S(A_1,\ldots,A_k,C_1,\ldots,C_s)$ relációk
  $\Longrightarrow R \rtimes S =$ R azon sorai, amelyhez vannak passzoló sorok S-ben
  $R \rtimes S = \pi_R(R \bowtie S)$
- $R \rtimes S \subseteq R$
- $R \ltimes S =$ ugyanez jobbról
- Ha R és S ugyanaz, mint múltkor:

$R \rtimes S$	A	B	C
	a	b	2
	a	c	3

$R \ltimes S$	D	C
	a	2
	b	3
	x	2

Származtatott műveletek

θ-illesztés

- R, S relációk
  $\Longrightarrow R \underset{R.A_i\theta S.B_j}{\bowtie} S = R \times S$ azon sorai, amelyben az adott oszlopok θ relációban vannak
  $R \underset{R.A_i\theta S.B_j}{\bowtie} S = \sigma_{R.A_i\theta S.B_j}(R \times S)$
- Példa:

R	A	B	C
	a	b	2
	a	c	3
	b	a	4

S	D	E
	a	2
	b	3
	x	2

$R \underset{C\leq E}{\bowtie} S$	A	B	C	D	E
	a	b	2	a	2
	a	b	2	b	3
	a	b	2	x	2
	a	c	3	b	3

Példák relációs algebra alkalmazására

ÁRU(ÁRUKÓD, ÁRUNÉV, EGYSÉGÁR)
MENNYISÉG(DÁTUM, ÁRUKÓD, DB)
BEVÉTEL(DÁTUM, ÖSSZEG)
BEFIZ(ÖSSZEG, BEFIZ) BEFIZ=ÖSSZEG-4000

A 2002. jan. 1. utáni napok bevételei a dátummal együtt:

$$\sigma_{\text{DÁTUM>'2002-01-01'}}\left(\text{BEVÉTEL}\right)$$

A 2002. jan. 15-i befizetett összeg és bevétel:

$$\pi_{\text{ÖSSZEG, BEFIZ}}\left(\sigma_{\text{DÁTUM='2002-01-15'}}\left(\text{BEVÉTEL}\bowtie\text{BEFIZ}\right)\right)$$

$$\pi_{\text{ÖSSZEG, BEFIZ}}\left(\sigma_{\text{DÁTUM='2002-01-15'}}\left(\text{BEVÉTEL}\right)\bowtie\text{BEFIZ}\right)$$

Hány darabot adtak el 2002. jan. 15-én az A123 kódú áruból, mi a neve és az ára?

$$\pi_{\text{DB, ÁRUNÉV, EGYSÉGÁR}}\left(\sigma_{\text{ÁRUKÓD='A123'} \wedge \text{DÁTUM='2002-01-15'}}\left(\text{MENNYISÉG}\bowtie\text{ÁRU}\right)\right)$$

$$\pi_{\text{DB, ÁRUNÉV, EGYSÉGÁR}}\left(\sigma_{\text{ÁRUKÓD='A123'} \wedge \text{DÁTUM='2002-01-15'}}\left(\text{MENNYISÉG}\right) \bowtie \text{ÁRU}\right)$$

Mely nevű áruk azok, amelyekkel van azonos egységárú másik áru?

Az ÁRU reláció két sorát kell összevetni.

Átnevezés

- Technikai segítség, ha pl. két relációban ugyanolyan attribútum név van, és direkt szorzatot akarunk. Nem változtatja meg a reláció sorait, csak az attribútumok és a reláció nevét, ezért nem igazi művelet.
- $R(A_1,\ldots,A_n)$ egy reláció
  $\Longrightarrow$ $\rho_{S(B_1,\ldots,B_n)}(R) =$ sorai megegyeznek R soraival, a reláció neve S, attribútumai rendre $B_1,\ldots,B_n$.
- Ha csak a relációt akarjuk átnevezni: $\rho_S(R)$

Megoldás:

$$\text{ÁRU1} = \rho_{\text{ÁRU1(ÁRUKÓD1, ÁRUNÉV1, EGYSÉGÁR1)}}(\text{ÁRU})$$
$$\text{ÁRU2} = \rho_{\text{ÁRU2(ÁRUKÓD2, ÁRUNÉV2, EGYSÉGÁR2)}}(\text{ÁRU})$$
$$\text{ÁRU3} = \text{ÁRU1} \underset{\text{EGYSÉGÁR1 = EGYSÉGÁR2} \wedge \text{ÁRUKÓD1} \neq \text{ÁRUKÓD2}}{\bowtie} \text{ÁRU2}$$
$$\text{ÁRU4} = \pi_{\text{ÁRUNÉV1}}\left(\text{ÁRU3}\right)$$

További példák

TERMÉK(GYÁRTÓ, MODELL, TÍPUS)
PC(MODELL, SEBESSÉG, MEMÓRIA, MEREVLEMEZ, CD, ÁR)
LAPTOP(MODELL, SEBESSÉG, MEMÓRIA, MEREVLEMEZ, KÉPERNYŐ, ÁR)
NYOMTATÓ(MODELL, SZÍNES, TÍPUS, ÁR)

A relációk jelentése:
TERMÉK: az adott nevű gyártó gyártja az adott modellszámú és adott típusú (PC, Laptop vagy nyomtató) terméket
PC: modellszám, sebesség megaHz-ben, memória megabájtban, merevlemez gigabájtban, a CD sebessége (pl. 4x), az ár
Laptop: mint PC-nél, plusz a képernyő mérete hüvelykben
Nyomtató: modellszám, színes-e (i/n), típusa (tintasugaras, lézer, mátrix), ára

A modellszámokról feltesszük, hogy egyediek.

- Melyek azok a PC modellek, amelynek sebessége legalább 150?

$$\pi_{\text{MODELL}}\left(\sigma_{\text{SEBESSÉG} >= 150}(\text{PC})\right)$$

- Mely gyártók készítenek legalább egy gigás merevlemezű laptopot?

$$\pi_{\text{GYÁRTÓ}}\left(\text{TERMÉK} \bowtie \sigma_{\text{MEREVLEMEZ}>=1}(\text{LAPTOP})\right)$$

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

$$\pi_{\text{MODELL, ÁR}}\left(\sigma_{\text{GYÁRTÓ='B'} \wedge \text{TÍPUS = 'PC'}}\left(\text{TERMÉK}\right) \bowtie \text{PC}\right) \cup$$
$$\pi_{\text{MODELL, ÁR}}\left(\sigma_{\text{GYÁRTÓ='B'} \wedge \text{TÍPUS = 'LAPTOP'}}\left(\text{TERMÉK}\right) \bowtie \text{LAPTOP}\right) \cup$$
$$\pi_{\text{MODELL, ÁR}}\left(\sigma_{\text{GYÁRTÓ='B'} \wedge \text{TÍPUS = 'NYOMTATÓ'}}\left(\text{TERMÉK}\right) \bowtie \text{NYOMTATÓ}\right)$$

- Melyek azok a gyártók, akik laptopot gyártanak, de PC-t nem?

$$\text{TERMÉK1} = \rho_{\text{TERMÉK1}}(\pi_{\text{GYÁRTÓ, TÍPUS}}\left(\text{TERMÉK}\right))$$
$$\pi_{\text{GYÁRTÓ}}\left(\sigma_{\text{TÍPUS='LAPTOP'}}\left(\text{TERMÉK1}\right)\right) \setminus \pi_{\text{GYÁRTÓ}}\left(\sigma_{\text{TÍPUS='PC'}}\left(\text{TERMÉK1}\right)\right)$$

- Melyek azok a gyártók, amelyek gyártanak legalább két, egymástól különböző, legalább 133 Mhz-en működő PC-t vagy Laptopot? (Nincs két azonos modellszám!)

$$R1 = \pi_{\text{MODELL, SEBESSÉG}}(\text{PC}) \cup \pi_{\text{MODELL, SEBESSÉG}}(\text{LAPTOP})$$
$$R2 = \pi_{\text{GYÁRTÓ, MODELL}}\left(\sigma_{\text{SEBESSÉG}>=133}(\text{R1}) \Join \text{TERMÉK}\right)$$
$$R3 = \rho_{R3(\text{GYÁRTÓ2, MODELL2})}(R2)$$
$$R4 = R2 \underset{\text{GYÁRTÓ = GYÁRTÓ2} \wedge \text{MODELL <> MODELL2}}{\Join} R3$$
$$R5 = \pi_{\text{GYÁRTÓ}}(\text{R4})$$

Megjegyzés: kifejezésfával is meg lehet adni a relációs algebrai kifejezéseket, de ezeket most nem rajzolom ide le....

További műveletek

Ezek nincsenek benne a relációs algebrában, de fontosak, nem túl műveletigényesek.

aggregátumok: MIN, MAX, AVG, SUM, CNT (szám)
Pl. leggyorsabb gép, átlagár, hányféle printer
eredmény mindig egy szám

aggregátum csoportosítva: Bizonyos feltételek szerinti partíciókban aggregátumok.
Pl. átlagos ár tintasugaras nyomtatók között, egy gyártónak hány terméke van
$\Longrightarrow$ eredmény egy reláció pl. (gyártó, szám) párokból.

rekurzív lezárás: (hagyományos adatkezelésben ritka, intelligensebb rendszerekben inkább)
Pl. reláció: ki főnöke kinek $\Longrightarrow$ lezárás: ki felettese kinek
reláció: melyik városból melyikbe van repülő járat $\Longrightarrow$ lezárás: átszállással el lehet-e jutni
Ezt a relációs algebra nem tudja, csak fix mélységre: pl. max 4 átszállás, (SQL nem igazán támogatja, de azért kicsit igen: SQL3, majd lesz)

A NULL érték

Lehet, hogy vannak kitöltetlen mezők, ezt meg akarjuk engedni. Pl. ha csak a termelő anyjának neve hiányzik, attól még a termék ára kell. $\Longrightarrow$ NULL érték

2 alapvető értelmezés (majd SQL-nél lesz, hogy hogyan kell megmondani, hogy melyik van éppen, illetve, hogy lehet-e egyáltalán NULL valahol):

- $\nexists$
- $\exists$, de nem ismerjük.

Attól függően, hogy hogyan értelmezzük a NULL-t:
Mi legyen egy ilyen kérdéssel?:
Pl. $\pi_{\text{CÍM='BP'}}\left(\text{TERMELŐ}\right)$
Ilyenkor belevegyük-e ha a cím NULL?

Külső illesztés (outer join)

Definíció. *R, S relációk $\Longrightarrow$ $R ⟕ S$* ***bal külső illesztés****: $R \Join S$-hez azokat az R-beli sorokat is hozzávesszük, amihez nem illeszkedik S-beli. Hiányzó helyekre NULL kerül.*

Pl. SZEMÉLY(NÉV, KÓD), CÍM(KÓD, CÍM)
SZEMÉLY $\Join$ CÍM $\Longrightarrow$ akinek nincs címe nem lesz rajta
SZEMÉLY $⟕$ CÍM $\Longrightarrow$ kiderül, kinek nincs meg a címe

SQL-ben van, relációs algebrával elvileg nem fejezhető ki (NULL miatt), de elkerülhető. Lényegében: $(R \Join S) \cup (R \setminus (R \ltimes S))$

Van **jobb külső illesztés** is: $R ⟖ S$
Teljes külső illesztés: $R ⟗ S := (R ⟕ S) \cup (R ⟖ S)$

Példa:

R	A	B	C
	a	b	2
	a	c	3
	b	a	4

S	D	C
	a	2
	b	3
	x	2
	y	1

R ⟕ S	A	B	C	D
	a	b	2	a
	a	b	2	x
	a	c	3	b
	b	**a**	**4**	NULL

R ⟖ S	A	B	C	D
	a	b	2	a
	a	b	2	x
	a	c	3	b
	NULL	NULL	**1**	**y**

R ⟗ S	A	B	C	D
	a	b	2	a
	a	b	2	x
	a	c	3	b
	b	**a**	**4**	NULL
	NULL	NULL	**1**	**y**

Külső unió

Részben kompatibilis relációk egyesítésére:
DIÁK(NÉV, TÉMAVEZ, TSZK)
TANÁR(NÉV, TSZK, BEOSZT)

DIÁK $\cup_k$ TANÁR

	NÉV	TSZK	TÉMAVEZ	BEOSZT
diák				NULL
tanár			NULL	

Multihalmazos szemantika

A relációs algebrában ugyan minden reláció halmaz, ezért nincsenek többszörös sorok, de pl. SQL-nél lesznek. A multihalmazokkal kicsit máshogy vannak a halmazműveletek:

Ha a t sor $m_R(t)$ példányban van meg R-ben és $m_S(t)$ példányban van meg S-ben, akkor

- $m_{(R\cup S)}(t) := m_R(t) + m_S(t)$ példányban lesz meg R és S uniójában
- $m_{(R\cap S)}(t) := \min\{m_R(t), m_S(t)\}$ példányban lesz meg R és S metszetében
- $m_{(R\setminus S)}(t) := \max\{m_R(t) - m_S(t), 0\})$ példányban lesz meg $R\setminus S$-ben

ODL séma átírása relációsémává

Legegyszerűbb eset $\Longrightarrow$ az osztályoknak csak attribútumai vannak, amik atomi típusúak:

```
interface Film (keys (cím, év)){
    attribute cím;
    attribute év;
    attribute hossz;
    attribute szalagFajta;
};
```

$\Longrightarrow$ `Film(cím, év, hossz, szalagFajta)`

A kulcs az ODL-es kulcs lesz, ha egy van csak belőle; ezt aláhúzással jelöljük. Ha több kulcs van: egyiket jelöljük, a többit írásban (de persze azok is kulcsok a relációsémában is).

Összetett típusú attribútumok: pl. rekordszerkezet OK $\Longrightarrow$

```
interface Színész (key név){
    attribute String név;
    attribute Struct Cím
        { string város, string utca } lakcím;
};
```

⟹ Színész(<u>név</u>, város, utca)

Kulcs: ugyanúgy, mint a nem összetett attribútumnál

De pl. halmaz csak rosszul oldható meg: halmaz minden eleméhez új sor.

```
interface Színész {
    attribute String név;
    attribute Set <
        Struct Cím{ string város, string utca }
        > lakcím;
};
```

⟹ Színész(név, város, utca)

név	város	utca
Gálvölgyi J.	Budapest	Nyereg u. 2.
Gálvölgyi J.	Budapest	Kantár u. 3.
⋮		

Kulcs: elromlik az ODL-es kulcs, lehet, hogy ami ott kulcs volt, itt már nem lesz az ⟹ baj