Adatbázisok elmélete 5. előadás

Katona Gyula Y.
Budapesti Műszaki és Gazdaságtudományi Egyetem
Számítástudományi Tsz.
I. B. 137/b
kiskat@cs.bme.hu
http://www.cs.bme.hu/˜kiskat

2005

A relációs algebra alapműveletei (emlékeztető)

- Halmazműveletek (bármilyen halmazra mennének)
  - ⋆ unió: $\cup$
  - ⋆ különbség: $\setminus$
  - ⋆ szorzat: $\times$
- Relációs műveletek (ezek már kihasználják, hogy itt relációkról van szó)
  - ⋆ vetítés, projekció: π
  - ⋆ kiválasztás, szelekció: σ

Ezek mind tiszta műveletek: reláció → reláció
$\Longrightarrow$ gond nélkül egymásba ágyazhatók

Kiválasztás, szelekció

- R egy reláció $\Longrightarrow$ $\sigma_F(R)$ = a reláció azon sorai, amelyekre az F formula teljesül.
- Teljesülni fog, hogy $\sigma_F(R) \subseteq R$
- Nincs megszorítás, csak hogy F értelmes legyen.
- Az eredmény örökli R típusait és attribútum neveit
- Példa:

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

$\sigma_{A \neq B \wedge C>2}(R)$	A	B	C
	a	c	3
	b	c	4

Az F formula:

Atomok: $A\ \theta\ B$, $A\ \theta\ c$, $c\ \theta\ B$,
ahol A, B attribútumok, c érték (konstans), $\theta \in \{<, >, =, \boxed{\leq, \geq, \neq}\}$

Építkezés: $\wedge, \vee, \neg, (,)$ Kvantorok, nincsenek!

- Példa:

DOLGOZÓ(NÉV,CÍM,FIZETÉS)

$$\sigma_{\text{CÍM='BP., Várna u.'} \wedge \text{FIZETÉS>'50000'}}(\text{DOLGOZÓ})$$

Relációs algebra

Definíció. *Alapreláció:* *A bevezetés, tervezés során definiált tábla, ami meg van adva.*
A relációs algebra relációi: amik kifejezhetők az alaprelációkból $\cup, \setminus, \times, \pi, \sigma$ *segítségével.*
Származtatott reláció: nem alapreláció, de kifejezhető.
Definíció. *Egy lekérdező nyelv (igazi vagy modell) relációsan teljes, ha benne megvalósíthatók a relációs algebra alapműveletei:* $\cup, \setminus, \times, \pi, \sigma$

Ez fontos követelmény, általában tudja is mindegyik.
Inkább az a baj, hogy néha túl sokat tudnak, de nincs hatékony implementáció.

Származtatott műveletek

Hasznosak, de mivel kifejezhetők az 5 alapművelettel, ezért lényegében csak rövidítések.

Metszet

- R, S relációk $\Longrightarrow R \cap S = R \setminus (R \setminus S)$ azok a sorok, amelyek mindkettőben benne vannak.
- nincs kompatibilitási követelmény $\Longleftarrow \setminus$ tulajdonságából
- Az eredmény örökli R típusait és attribútum neveit $\Longleftarrow \setminus$ tulajdonságából
- Példa:

R	A	B
	a	a
	a	c
	b	a

S	A	C
	a	a
	a	d
	a	c
	b	b

$R \cap S$	A	$B \cap C$
	a	a
	a	c

Származtatott műveletek

Természetes illesztés (natural join)

- $R(A_1, \ldots, A_k, B_1, \ldots, B_r), S(A_1, \ldots, A_k, C_1, \ldots, C_s)$ relációk $\Longrightarrow R \bowtie S =$
  - Vegyük $R \times S$-t
  - Vesszük azokat a sorokat, ahol $R.A_1 = S.A_1, \ldots, R.A_k = S.A_k$, a többit kidobjuk.
  - $\forall A_i$-ból az egyik példányt eldobjuk, azaz vetítünk $R.A_1, \ldots, R.A_k, R.B_1, \ldots, R.B_r, S.C_1, \ldots, S.C_s$-re
  - Azonos sorokból csak egyet tartunk meg, a többit kidobjuk.
- $$R \bowtie S = \pi_{R.A_1,\ldots}(\sigma_{R.A_1=S.A_1,\ldots}(R \times S))$$

  $R \bowtie S$-nek $k + r + s$ oszlopa lesz.
  Ha nincs közös attribútum. $\Longrightarrow R \bowtie S = R \times S$.
- nincs kompatibilitási követelmény
- Az eredmény örökli R és S típusait és attribútum neveit
- Gyakorlatban ennél hatékonyabban számítjuk ki.
- Az oszlopok sorrendje nem definiált, de általában: R oszlopai, aztán S saját oszlopai.

- Példa:

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

S	D	C
	a	2
	b	3
	x	2

$R \bowtie S$	A	B	C	D
	a	b	2	a
	a	b	2	x
	a	c	3	b

Miért „természetes"?

Példa: TERMELŐ(TNÉV,TERMÉK,ÁR,CÍM)

- TNÉV → CÍM
- TNÉV, TERMÉK → ÁR

Gond: TERMELŐ címét minden terméknél tároljuk
⟹ redundancia + veszélyek : cím mindig kell, minden módosításhoz; könnyen sérülhet a fent megadott függés, ha elírom a címet; akkor is kell a cím, ha csak új árut akarok felvenni)

Megoldás: Inkább tároljuk két táblában:

$R = \pi_{\text{TNÉV, CÍM}}(\text{TERMELŐ})$ és
$S = \pi_{\text{TNÉV, TERMÉK, ÁR}}(\text{TERMELŐ})$

⟹ TERMELŐ=$R \bowtie S$ (ha kell egyben a tábla, vissza lehet állítani)

Jó-e bármilyen felbontás?

$R' = \pi_{\text{TNÉV, CÍM, ÁR}}(\text{TERMELŐ})$ és
$S' = \pi_{\text{TNÉV, TERMÉK}}(\text{TERMELŐ})$

⟹ minden terméknek ugyanannyi lesz az ára (sok ára lesz)
⟹ TERMELŐ$\subsetneq R' \bowtie S'$

Az lesz majd a kérdés, hogy mik lesznek a jó felbontások?

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
  $\Longrightarrow R \ltimes S = R$ azon sorai, amelyhez vannak passzoló sorok S-ben
  $R \ltimes S = \pi_R(R \bowtie S)$
- $R \ltimes S \subseteq R$
- $R \rtimes S =$ ugyanez jobbról
-

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

S	D	C
	a	2
	b	3
	x	2

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

$R \rtimes S$	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

$R \underset{C\le E}{\bowtie} S$	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 2004. jan. 1. utáni napok bevételei a dátummal együtt:

$$\sigma_{\text{DÁTUM}>\text{'2004-01-01'}}\left(\text{BEVÉTEL}\right)$$

A 2004. jan. 15-i befizetett összeg és bevétel:

$$\pi_{\text{ÖSSZEG, BEFIZ}}\left(\sigma_{\text{DÁTUM='2004-01-15'}}\left(\text{BEVÉTEL}\bowtie\text{BEFIZ}\right)\right)$$

$$\pi_{\text{ÖSSZEG, BEFIZ}}\left(\sigma_{\text{DÁTUM='2004-01-15'}}\left(\text{BEVÉTEL}\right)\bowtie\text{BEFIZ}\right)=\sigma_{\text{DÁTUM='2004-01-15'}}\left(\text{BEVÉTEL}\right)\bowtie\text{BEFIZ}$$

Hány darabot adtak el 2004. jan. 15-én az A123 kódú áruból, mi a neve és az ára?

$$\pi_{\text{DB, ÁRUNÉV, EGYSÉGÁR}}\left(\sigma_{\text{ÁRUKÓD='A123'}\wedge\text{DÁTUM='2004-01-15'}}\left(\text{MENNYISÉG}\bowtie\text{ÁRU}\right)\right)$$

$$\pi_{\text{DB, ÁRUNÉV, EGYSÉGÁR}}\left(\sigma_{\text{ÁRUKÓD='A123'}\wedge\text{DÁTUM='2004-01-15'}}\left(\text{MENNYISÉG}\right)\bowtie\text{ÁRU}\right)$$

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:

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

Sorkalkulus (Tuple calculus)

Lekérdezőnyelvek típusai:

- relációs algebra (**LEAP**, letölthető, SIGMOD-ról link), **ISBL** *nehezen emészthetőbb, algebrai alapú; ez volt: láttuk, hogy relációs algebrával jól meg lehet adni relációkat*
- sorkalkulus (**SQL egy kicsit, QUEL**, INGRES nyelve), *könnyebben emészthető, logikai alapú*
- oszlopkalkulus (**QBE, SQL**) nagyon hasonló

Sorkalkulus (Tuple calculus)

Formális modell, de már hasonlít az igazihoz.

Elsőrendű nyelv relációk kifejezésére.

Változók: t, r, s sorváltozók, a reláció sorainak felel meg
$t^{(k)}$**:** k oszlopos reláció sorainak felel meg
$t^{(k)}[i]$**:** A t sorváltozó i-edik komponense.

Pl. egy sor $\Longrightarrow$ (R. M., Budapest, hamburger, 180), akkor
$t^{(4)}[3] =$'hamburger' és $t^{(4)}[\text{ÁR}] = 180$

Sorkalkulus

Cél: sorkalkulussal relációkat megadni, úgy, mint relációs algebrával

A sorkalkulus által kifejezett reláció:

$$\{t^{(k)} \mid \phi(t)\}$$

$\Longrightarrow$ a kifejezett reláció azon t-kből áll, amikre $\phi(t)$ igaz, ahol ϕ egy megengedett formula + valami még.

Megengedett formulák (amik a $\phi(t)$ helyén állhatnak):

atomok :

- $R^{(k)}(t^{(k)})$: *(ahol R alapreláció), akkor igaz, ha $t \in R$, azaz a sor benne van a relációban.*
- ★ $t^{(k)}[i]\ \theta\ s^{(l)}[j]$
  ★ $t^{(k)}[i]\ \theta\ c$
  ★ $c\ \theta\ t^{(k)}[i]$
  ahol $\theta \in \{<, >, =, \neq, \leq, \geq\}$, t, s sorváltozók, c konstans érték.
  Világos, mikor igaz.

építkezési szabályok :

- ϕ, ψ formulák, akkor $\phi \vee \psi, \phi \wedge \psi, \neg\phi$ is formulák.
  Világos, hogy mikor igaz.
- ϕ formula, s sorváltozó, akkor $\forall s\phi, \exists s\phi$ is formula.
  Világos, hogy mikor igaz.
  Kötött változó: ha vonatkozik rá kvantor,
  Szabad változó: ha nem,

Sorkalkulus által kifejezett reláció (pontosan):

$$\{t^{(k)} \mid \phi(t)\}$$

$\Longrightarrow$ a kifejezett reláció azon k hosszú t vektorokból áll, amikre $\phi(t)$ igaz, ahol ϕ egy megengedett formula és ϕ*-ben* t *az egyetlen szabad* változó.

Példák sorkalkulus alkalmazására

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

Az 2004. jan. 1. utáni napok bevételei a dátummal együtt:

$$\sigma_{\text{DÁTUM>'2004-01-01'}}(\text{BEVÉTEL})$$
$$\{t^{(2)} \mid \text{BEVÉTEL}(t) \wedge t[1] \geq \text{2004-01-01}\}$$

Az 2004. jan. 15-i bevétel és a befizetett összeg:

$$\pi_{\text{ÖSSZEG, BEFIZ}}(\sigma_{\text{DÁTUM='2004-01-15'}}(\text{BEVÉTEL}) \bowtie \text{BEFIZ})$$
$$\{u^{(2)} \mid \text{BEFIZ}(u) \wedge \exists v(\text{BEVÉTEL}(v) \wedge v[1] = \text{2004-01-15} \wedge v[2] = u[1])\}$$

Hány darabot adtak el 2004. jan. 15-én az A123 kódú áruból, mi a neve és az ára?

$$\pi_{\text{DB, ÁRUNÉV, EGYSÉGÁR}}(\sigma_{\text{ÁRUKÓD='A123'} \wedge \text{DÁTUM='2004-01-15'}}(\text{MENNYISÉG} \bowtie \text{ÁRU}))$$

$$\{s^{(3)} \mid \exists u \exists v(\text{MENNYISÉG}(u) \wedge \text{ÁRU}(v) \wedge u[1] = \text{2004-01-15} \wedge u[2] = {}'A123' \wedge$$
$$v[1] = {}'A123' \wedge s[1] = u[3] \wedge s[2] = v[2] \wedge s[3] = v[3])\}$$

Példák sorkalkulus alkalmazására

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

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

$$\{s^{(1)} \mid \exists u \exists v(\text{ÁRU}(v) \wedge \text{ÁRU}(u) \wedge s[1] = v[2] \wedge v[3] = u[3] \wedge \neg(v[1] = u[1]))\}$$