Adatbázisok elmélete 7. 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

2004

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:

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

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

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

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

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

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

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?

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

Sorkalkulus ereje

Mivel lehet több mindent kifejezni?
Relációs algebrával, vagy sorkalkulussal?

Tétel. *A sorkalkulus relációsan teljes.*

Bizonyítás: Be kell látni, hogy minden reláció, ami relációs algebrával megadható, megadható sorkalkulussal is. Ehhez azt elég megmutatni, hogy
1. az alaprelációk megadhatók
2. a relációs algebrai alapműveletek (unió, különbség, szorzat, vetítés, szelekció) alaprelációkra alkalmazva megvalósíthatók
3. ha R és S nem alapreláció és ezekre alkalmazunk valami relációs alapműveletet, akkor az eredmény kifejezhető sorkalkulussal

Bizonyítás

- **alapreláció:** Tegyük fel, hogy R k oszlopos alapreláció
  $R = \left\{t^{(k)} \mid R(t)\right\}$
- **Unió:** *Tfh. S is k oszlopos*
  $R \cup S = \left\{t^{(k)} \mid R(t) \vee S(t)\right\}$
- **különbség:**
  $R \setminus S = \left\{t^{(k)} \mid R(t) \wedge \neg S(t)\right\}$
- *metszet:*
  $R \cap S = \left\{t^{(k)} \mid R(t) \wedge S(t)\right\}$
- **szorzat:** R legyen k oszlopos, S pedig l oszlopos

$$R \times S = \Big\{t^{(k+l)} \mid \exists r^{(k)} \exists s^{(l)} \,(R(r) \wedge S(s) \wedge r[1] = t[1] \wedge \ldots$$
$$\wedge r[k] = t[k] \wedge s[1] = t[k+1] \wedge \ldots \wedge s[l] = t[k+l])\}$$

- **vetület:** Legyen $R(A_1, \ldots, A_d, A_{d+1}, \ldots, A_k)$ reláció, vetítsük az első d-re
  $\pi_{A_1,\ldots,A_d}(R) = \left\{t^{(d)} \mid \exists r^{(k)} \,(R(r) \wedge r[1] = t[1] \wedge \ldots \wedge r[d] = t[d])\right\}$

- **kiválasztás:**
  $\sigma_F(R) = \{t^{(k)} \mid R(t) \wedge F'\}$, ahol F' átfordítása sorkalkulusra $\Longrightarrow$ az i-edik attribútum helyett $t^{(n)}[i]$-t írunk.
  Pl. (evidenciával történő meggyőzés)
  $\sigma_{\text{ÁR}>\text{'150'} \wedge \text{TERMÉK}=\text{'hamburger'}}(\text{TERMEL}) =$
  $\{t^{(4)} \mid \text{TERMEL}(t) \wedge t[4] > \text{'150'} \wedge t[3] = \text{'hamburger'}\}$
- Nem lényeges, hogy R, S alaprelációk.
  Ha $R = \{t^{(k)} \mid \phi(t)\}$ és $S = \{t^{(k)} \mid \psi(t)\}$, azaz R és S már valahogy ki van fejezve sorkalkulussal
  $\Longrightarrow R \cup S = \{t^{(k)} \mid \phi(t) \vee \psi(t)\}$
  többinél ugyanígy ($R(t)$ és $S(t)$ helyett $\phi(t)$-t és $\psi(t)$-t írunk). √

Megfordítás

Ki lehet-e fejezni mindet relációs algebrával, amit sorkalkulussal lehet?

[trans='Dissolve']

Nem!

[trans='Replace']

Pl. Ha R egy k változós alapreláció $\Longrightarrow \{t^{(k)} \mid \neg R(t)\}$ nem fejezhető ki algebrával.

Bizonyítás: Relációs algebrában minden reláció véges, *ha az alaprelációk végesek.* Ez viszont lehet végtelen, ha az egyik értékkészlet végtelen. √

Alkalmazásokban, bár elvileg ∀ véges, gyakorlatban azért nagy-nagy véges sem jó. Így ilyen baj tényleg előfordulhat.
Sőt részeredményekben sem lehet ilyen $\Longrightarrow$ túl sok munka.

Megoldás: Nem használunk ilyesmit $\Longrightarrow$ csak **biztonságos formulákat** (safe expression): kiértékelhető úgy, hogy ne kelljen túl nagy halmazt végignézni, csak annyi infó kell hozzá, amit valaki már egyszer korábban beírt $\Longrightarrow$
leszűkítjük a szóba jövő esetek halmazát
Definíció. $Dom(\phi)$= {*ϕ-beli alaprelációk ∀ attribútumának, ∀ értéke*} ∪ {*ϕ-beli konstansok*}

Pl. SZEMÉLY(NÉV, CÍM) alapreláció
$\phi(t) = \text{SZEMÉLY}(t) \wedge t[2] = \text{'Tokyo'}$
$\Longrightarrow Dom(\phi) = \pi_{\text{NÉV}}(\text{SZEMÉLY}) \cup \pi_{\text{CÍM}}(\text{SZEMÉLY}) \cup \{\text{'Tokyo'}\}$

Definíció. *Egy $R = \{t^{(k)} \mid \phi(t)\}$ **reláció biztonságos**, ha*

i) *Minden $\phi(t)$-t kielégítő t minden komponense $\in Dom(\phi)$*
(Ha t kielégíti $\phi(t)$-t, akkor minden komponense $Dom(\phi)$-beli)
(Ez korlátozza keresést! A végeredménybe csak $Dom(\phi)$-ből lehet kerülni.)

ii) *ϕ minden $\exists u \psi(u)$ alakú részformulájára igaz, hogy ha u kielégíti ψ-t a ψ-beli szabad változók valamely értékeire, akkor u minden komponense $\in Dom(\psi)$*
(részformula is biztonságos: $\exists u \psi(u)$ igazsága eldönthető $Dom(\psi)$ végignézésével)

Megj.: *A $\forall u \psi(u)$ alakúakra nem kell, mert ez ugyanaz, mint $\neg\exists u(\neg\psi(u))$ és így elég (ii)-t ellenőrizni a $\exists u(\neg\psi(u))$ részformulára.*

Nem az a kérdés, hogy pontosan mik a biztonságos formulák, hanem:
Hogyan tudunk biztonságos formulákat, kifejezéseket írni?
Csak ilyeneket szeretnénk használni.

Biztonságos formulák

Tipikus technikák a biztonságosság elérésére:

- $\{t \mid R(t) \wedge \phi(t)\}$ biztonságos, ha ϕ biztonságos (mert szűkítés R-re, így (i) teljesül)
- *(Ha nem lehet egy relációra korlátozni pl:)*
  $\{t \mid (R(t) \vee S(t)) \wedge \phi(t)\}$ biztonságos, ha ϕ biztonságos
- $\{t^{(2)} \mid \exists u_1 \exists u_2 (R(u_1) \wedge R(u_2) \wedge t[1] = u_2[1] \wedge t[2] = u_1[1] \wedge \phi(t, u_1, u_2))\}$
- $\exists u(R(u) \wedge \dots)$ ha itt a hátrább levő részek biztonságosak (az R-re való szűkítés miatt (ii) teljesül)
- $\forall u(\neg R(u) \vee \dots)$ ha itt a hátrább levő részek biztonságosak (mert ez átírva $\neg\exists u(R(u) \wedge \dots)$

Tétel. *Biztonságos kifejezésű sorkalkulus reláció kifejezhető relációs algebrával is.*

Bizonyítás: *Ötlet:* Ha $R = \{t^{(k)} \mid \phi(t)\}$ megengedett reláció sorkalkulusban, akkor $R \cap Dom(\phi)^k$ kifejezhető rel. algebrával. Biz. ϕ felépítése szerinti indukcióval.
Ha R biztonságos, akkor $R = R \cap Dom(\phi)^k$ √

Tétel. *A relációs algebra és a biztonságos sorkalkulus ekvivalensek.*

Bizonyítás: Volt: relációs algebrai kifejezésből lehet sorkalkulust csinálni, illetve biztonságos sorkalkulusból relációs algebrát.
Kell még: a relációs alg. átírása sorkalkulusra biztonságos.
De az meg az volt, ahogy csináltuk. √

Példák biztonságos és nem biztonságos kifejezésekre

- Melyik napokon volt a legnagyobb a bevétel?

  Nem bizt.:

  $$\left\{t^{(1)} \mid \exists u^{(2)} \left(\text{BEVÉTEL}(u) \wedge t[1] = u[2] \wedge \right.\right.$$
  $$\left.\left. \forall v^{(2)} \left(\neg\text{BEVÉTEL}(v) \vee v[2] \leq u[2]\right)\right) \wedge \exists w^{(1)}(w[1] \neq t[1])\right\}$$

  Ez azt fejezi ki, csak nem biztonságos, mert (ii) nem oké, bár (i) az. Az a baj, hogy a végén van egy felesleges dolog.

  Bizt.:
  $$\left\{t^{(1)} \mid \exists u^{(2)} \left(\text{BEVÉTEL}(u) \wedge t[1] = u[1] \wedge \forall v^{(2)} \left(\neg\text{BEVÉTEL}(v) \vee v[2] \leq u[2]\right)\right)\right\}$$

  Másik bizt.:
  $$\left\{t^{(1)} \mid \exists u^{(2)} \left(\text{BEVÉTEL}(u) \wedge t[1] = u[1] \wedge \neg\exists v^{(2)} \left(\text{BEVÉTEL}(v) \wedge v[2] > u[2]\right)\right)\right\}$$

- Mely árukból adtak el 100-nál többet egy napon?

  Nem bizt.:
  $$\left\{t^{(3)} \mid \text{ÁRU}(t) \wedge \exists v^{(3)}\left((v[2] = t[1] \wedge v[3] > 100) \boxed{\vee} \text{MENNYISÉG}(v)\right)\right\}$$

  Elírtuk: nem is azt adja, amit kell, meg nem is biztonságos, mert (ii) sérül

  Bizt.:
  $$\left\{t^{(3)} \mid \text{ÁRU}(t) \wedge \exists v^{(3)}\left((v[2] = t[1] \wedge v[3] > 100) \boxed{\wedge} \text{MENNYISÉG}(v)\right)\right\}$$

Oszlopkalkulus (Domain calculus)

Lényegileg ugyanaz, mint a sorkalkulus, csak másféle változókat használ.

Oszlopváltozó: egy koordinátája a sorváltozónak, mint vektornak
Az oszlopváltozó értékkészlete: egy adott attribútum értékkészlete.

A oszlopkalkulus által kifejezett reláció:
$\{u_1, \ldots, u_k \mid \phi(u_1, \ldots, u_k)\}$ $\Longrightarrow$ azon $u = (u_1, \ldots, u_k)$ vektorok, amikre $\phi(u_1, \ldots, u_k)$ igaz, ahol ϕ egy megengedett formula és csak $u_1, \ldots, u_k$ szabad változók benne.

Megengedett formulák:

atomok :

- $R^{(k)}(u_1, \ldots, u_k)$: *akkor igaz, ha $(u_1, \ldots, u_k) \in R$, azaz a sor benne van a relációban.*
- ⋆ $u_i\ \theta\ u_j$
  ⋆ $u_i\ \theta\ c$
  ⋆ $c\ \theta\ u_i$
  ahol $\theta \in \{<, >, =, \neq, \leq, \geq\}$, u_i oszlopváltozók, c konstans érték.
  Világos, mikor igaz.

Oszlopkalkulus (Domain calculus)

építkezési szabályok :

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

Példák oszlopkalkulus 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, legyen előbb a dátum:

$$\left\{t^{(2)} \mid \text{BEVÉTEL}(t) \wedge t[1] \geq \text{2004-01-01}\right\}$$

$$\left\{y, x \mid \text{BEVÉTEL}(y, x) \wedge y \geq \text{2004-01-01}\right\}$$ *(sajtóhiba a Gajdos könyvben)*

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

$$\left\{u^{(2)} \mid \text{BEFIZ}(u) \wedge \exists v(\text{BEVÉTEL}(v) \wedge v[1] = \text{2004-01-15} \wedge v[2] = u[1])\right\}$$

$$\left\{x, y \mid \text{BEFIZ}(x, y) \wedge \exists z(\text{BEVÉTEL}(z, x) \wedge z = \text{2004-01-15})\right\}$$