

# Adatbáziskezelés

## Sor és oszlopkalkulus

Csima Judit

BME, VIK,  
Számítástudományi és Információelméleti Tanszék

2017. szeptember 29.

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. ha a TERMEL(termelőnév, város, termék, ár) relációban egy sor (R. M., Budapest, hamburger, 180), akkor  $t[3] = \text{'hamburger'}$  és  $t[\text{ár}] = 180$

# Relációk megadása sorkalkulussal

**A sorkalkulus által kifejezett reláció:**

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

a kifejezett reláció azon  $t$ -kből áll, amikre  $\phi(t)$  igaz, ahol  $\phi$  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.*

# Megengedett formulák

## Építkezési szabályok:

- $\phi, \psi$  formulák, akkor  $\phi \vee \psi, \phi \wedge \psi, \neg\phi$  is formulák.  
*Világos, hogy mikor igaz.*
- $\phi$  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

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

a kifejezett reláció azon  $k$  hosszú  $t$  vektorokból áll, amikre  $\phi(t)$  igaz, ahol  $\phi$  egy megengedett formula és  $\phi$ -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 2002. jan. 1. utáni napok bevételei a dátummal együtt:

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

$\{t^{(2)} \mid \text{BEVÉTEL}(t) \wedge t[1] \geq 2002-01-01\}$

## Példák sorkalkulus alkalmazására

ÁRU(ÁRUKÓD, ÁRUNÉV, EGYSÉGÁR)

MENNYISÉG(DÁTUM, ÁRUKÓD, DB)

BEVÉTEL(DÁTUM, ÖSSZEG)

BEFIZ(ÖSSZEG, BEFIZ)       $BEFIZ = ÖSSZEG - 4000$

Az 2002. jan. 15-i bevétel és a befizetett összeg:

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

$\{u^{(2)} \mid BEFIZ(u) \wedge \exists v (BEVÉTEL(v) \wedge v[1] = 2002-01-15 \wedge v[2] = u[1])\}$



## Példák sorkalkulus alkalmazására

ÁRU(ÁRUKÓD, ÁRUNÉV, EGYSÉGÁR)

MENNYISÉG(DÁTUM, ÁRUKÓD, DB)

BEVÉTEL(DÁTUM, ÖSSZEG)

BEFIZ(ÖSSZEG, BEFIZ)      BEFIZ=ÖSSZEG-4000

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

$$\pi_{DB, \text{ÁNÉV}, \text{EÁR}} \left( \sigma_{\text{ÁKÓD}='A123' \wedge \text{D}='2002-01-15'} \left( \text{MENNYISÉG} \bowtie \text{ÁRU} \right) \right)$$
$$\{s^{(3)} \mid \exists u \exists v \left( \text{MENNYISÉG}(u) \wedge \text{ÁRU}(v) \wedge u[1] = 2002-01-15 \wedge \right.$$
$$\left. \wedge u[2] = 'A123' \wedge v[1] = 'A123' \wedge s[1] = u[3] \wedge s[2] = v[2] \wedge s[3] = v[3] \right) \}$$

## 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?

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

# Sorkalkulus vs. relációs algebra

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 alpműveletek (unió, különbség, szorzat, vetítés, szelekció) alaprelációkra alkalmazva megvalósíthatók
3. ha  $R$  és  $S$  nem alapreláció és ezekre alkalmazunk valami relációs alpműveletet, akkor az eredmény kifejezhető sorkalkulussal

# A sorkalkulus relációsan teljes

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

## A sorkalkulus relációsan teljes

- **szorzat:**  $R$  legyen  $k$  oszlopos,  $S$  pedig  $l$  oszlopos

$$R \times S = \left\{ t^{(k+l)} \mid \exists r^{(k)} \exists s^{(l)} \left( R(r) \wedge S(s) \wedge \right. \right. \\ \left. \left. \wedge r[1] = t[1] \dots \wedge r[k] = t[k] \wedge s[1] = t[k+1] \wedge \dots \wedge s[l] = t[k+l] \right) \right\}$$

- **vetítés:** Legyen  $R(A_1, \dots, A_d, A_{d+1}, \dots, A_k)$  reláció, vetítsük az első  $d$  attribútumra

$$\pi_{A_1, \dots, A_d}(R) = \left\{ t^{(d)} \mid \exists r^{(k)} \left( R(r) \wedge r[1] = t[1] \wedge \dots \wedge r[d] = t[d] \right) \right\}$$

## A sorkalkulus relációsan teljes

- **kiválasztás:**

$$\sigma_F(R) = \{t^{(k)} \mid R(t) \wedge F'\},$$

ahol  $F'$  átfordítása sorkalkulusra: az  $i$ -edik attribútum helyett  $t^{(n)}[i]$ -t írunk.

$$\text{Pl. } \sigma_{\text{ÁR} > '150' \wedge \text{TERMÉK} = \text{'hamburger'}}(\text{TERMEL}) = \\ \{t^{(4)} \mid \text{TERMEL}(t) \wedge t[4] > '150' \wedge t[3] = \text{'hamburger'}\}$$

# A sorkalkulus relációsan teljes

- **ha  $R$  és  $S$  nem alaprelációk:**

Nem lényeges, hogy  $R, S$  alaprelációk-e. Ha  $R = \{t^{(k)} \mid \phi(t)\}$  és  $S = \{t^{(k)} \mid \psi(t)\}$ , azaz  $R$  és  $S$  már valahogy ki van fejezve sorkalkulussal, akkor

$$R \cup S = \{t^{(k)} \mid \phi(t) \vee \psi(t)\}$$

többinél ugyanígy ( $R(t)$  és  $S(t)$  helyett  $\phi(t)$ -t és  $\psi(t)$ -t írunk.)

# Megfordítás

**Ki lehet-e fejezni mindent relációs algebrával, amit sorkalkulussal lehet?**

**Nem!**

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

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

Alkalmazásokban elvileg  $\forall$  véges, gyakorlatban azért nagy-nagy véges sem jó. Így ilyen baj tényleg előfordulhat. *Sőt részeredményekben sem lehet ilyen, mert túl sok munka lenne.*



## Hogyan érjük el, hogy ne legyenek ilyen csúfságok?

Csak **biztonságos formulákat** (safe expression) használunk: kiértékelhető úgy, hogy ne kelljen túl nagy halmazt végignézni, csak annyi infó kell hozzá, amit valaki már egyszer korábban beírt azaz leszűkítjük a szóba jövő esetek halmazát

**Definíció:**  $Dom(\phi) =$   
 $\{\phi\text{-beli alaprelációk } \forall \text{ attribútumának, } \forall \text{ értéke}\} \cup \{\phi\text{-beli konstansok}\}$

PI. SZEMÉLY(NÉV, CÍM) alapreláció és

$\phi(t) = \text{SZEMÉLY}(t) \wedge t[2] = \text{'Tokyo'}$  formula

$$\implies Dom(\phi) = \pi_{\text{NÉV}}(\text{SZEMÉLY}) \cup \pi_{\text{CÍM}}(\text{SZEMÉLY}) \cup \{\text{'Tokyo'}\}$$

## Biztonságos formula

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

- i) Minden,  $\phi(t)$ -t kielégítő  $t$  minden komponense  $\in \text{Dom}(\phi)$  (Ha  $t$  kielégíti  $\phi(t)$ -t, akkor minden komponense  $\text{Dom}(\phi)$ -beli.)  
(Ez korlátozza keresést, a végeredménybe csak  $\text{Dom}(\phi)$ -ből lehet kerülni.)
- ii)  $\phi$  minden  $\exists u\psi(u)$  alakú részformulájára igaz, hogy ha  $u$  kielégíti  $\psi$ -t a  $\psi$ -beli szabad változók valamely értékeire, akkor  $u$  minden komponense  $\in \text{Dom}(\psi)$   
(részformula is biztonságos:  $\exists u\psi(u)$  igazsága eldönthető  $\text{Dom}(\psi)$  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.

## Biztonságos formulák

*Nem az a kérdés, hogy pontosan mik a biztonságos formulák, hanem: Hogyan tudunk biztonságos formulákat, kifejezéseket írni?*

Tipikus technikák a biztonságosság elérésére:

- $\{t \mid R(t) \wedge \phi(t)\}$  biztonságos, ha  $\phi$  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  $\phi$  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)

## Biztonságos sorkalkulus vs. relációs algebra

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

**Bizonyítás:**

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

Ha pedig  $R$  biztonságos, akkor  $R = R \cap \text{Dom}(\phi)^k$ , vagyis  $R$ -t ki tudtuk fejezni.

# Biztonságos sorkalkulus = relációs algebra

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

**Bizonyítás:** Láttuk, hogy relációs algebrai kifejezésből lehet sorkalkulust csinálni, illetve biztonságos sorkalkulusból relációs algebrát.

Kell még: a relációs alg. megbeszélte átírása sorkalkulusra biztonságos.

Tényleg az, ha megnézzük azt a bizonyítást, akkor látszik. :)

## Példák biztonságos és nem biztonságos kifejezésekre

Melyik napokon volt a legnagyobb a bevétel?

- $\{t^{(1)} \mid \exists u^{(2)}(\text{BEVÉTEL}(u) \wedge \wedge t[1] = u[1] \wedge \forall v^{(2)}(\neg \text{BEVÉTEL}(v) \vee v[2] \leq u[2])))\}$
- $\{t^{(1)} \mid \exists u^{(2)}(\text{BEVÉTEL}(u) \wedge \wedge t[1] = u[1] \wedge \neg \exists v^{(2)}(\text{BEVÉTEL}(v) \wedge v[2] > u[2])))\}$

# Biztonságos vagy nem biztonságos?

Mely árukból adtak el 100-nál többet egy napon?

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

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

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

# Oszlopkalkulus (Domain calculus)

- Lényegileg ugyanaz, mint a sorkalkulus, csak másféle változókat használ.
- Oszlopváltozó: egy koordinátája a sorváltozónak, mint vektornak
- Az oszlopváltozó értékészlete: egy adott attribútum értékészlete.
- A oszlopkalkulus által kifejezett reláció:  
 $\{u_1, \dots, u_k \mid \phi(u_1, \dots, u_k)\} \implies$  azon  $u = (u_1, \dots, u_k)$  vektorok, amikre  $\phi(u_1, \dots, u_k)$  igaz, ahol  $\phi$  egy megengedett formula és csak  $u_1, \dots, u_k$  szabad változók benne.



# Oszlopkalkulus (Domain calculus)

Megengedett formulák:

- atomok:
  - $R^{(k)}(u_1, \dots, u_k)$ : akkor igaz, ha  $(u_1, \dots, u_k) \in R$ , 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.*
- építkezési szabályok:
  - $\phi, \psi$  formulák, akkor  $\phi \vee \psi, \phi \wedge \psi, \neg\phi$  is formulák.  
*Világos, hogy mikor igaz.*
  - $\phi$  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 2002. jan. 1. utáni napok bevételei a dátummal együtt, legyen előbb a dátum:

$\{t^{(2)} \mid \text{BEVÉTEL}(t) \wedge t[1] \geq 2002-01-01\}$

$\{x, y \mid \text{BEVÉTEL}(y, x) \wedge y \geq 2002-01-01\}$

## Példák oszlopkalkulus alkalmazására

ÁRU(ÁRUKÓD, ÁRUNÉV, EGYSÉGÁR)

MENNYISÉG(DÁTUM, ÁRUKÓD, DB)

BEVÉTEL(DÁTUM, ÖSSZEG)

BEFIZ(ÖSSZEG, BEFIZ)      BEFIZ=ÖSSZEG-4000

Az 2002. jan. 15-i bevétel és a befizetett összeg:

$\{u^{(2)} \mid \text{BEFIZ}(u) \wedge \exists v(\text{BEVÉTEL}(v) \wedge v[1] = 2002-01-15 \wedge v[2] = u[1])\}$

$\{x, y \mid \text{BEFIZ}(x, y) \wedge \exists z(\text{BEVÉTEL}(z, x) \wedge z = 2002-01-15)\}$

## Példák oszlopkalkulus alkalmazására

ÁRU(ÁRUKÓD, ÁRUNÉV, EGYSÉGÁR)

MENNYISÉG(DÁTUM, ÁRUKÓD, DB)

BEVÉTEL(DÁTUM, ÖSSZEG)

BEFIZ(ÖSSZEG, BEFIZ)       $BEFIZ = ÖSSZEG - 4000$

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

$$\{s^{(3)} \mid \exists u \exists v (MENNYISÉG(u) \wedge \text{ÁRU}(v) \wedge u[1] = 2002-01-15 \wedge \\ \wedge u[2] = 'A123' \wedge v[1] = 'A123' \wedge s[1] = u[3] \wedge s[2] = v[2] \wedge s[3] = v[3])\}$$
$$\{x, y, z \mid \exists u \exists v (MENNYISÉG(u, v, z) \wedge \text{ÁRU}(u, x, y) \wedge v = \\ 2002-01-15 \wedge u = 'A123')\}$$

## Példák oszlopkalkulus 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?

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

# Biztonságos kifejezés oszlopkalkulusban

$\text{Dom}(\psi)$  és a biztonságos formula és kifejezés ugyanaz, mint sorkalkulusnál.

Ugyanolyan technikák vannak a biztonságosság elérésére, mint sorkalkulusnál.

## Sor és oszlopkalkulus

**Tétel:** A sorkalkulus és az oszlopkalkulus ekvivalensek. A biztonságos sorkalkulus és a biztonságos oszlopkalkulus ekvivalensek.

*Bizonyítás vázlat:*

$$\begin{aligned}\{t^{(k)} \mid \phi(t^{(k)})\} &\longleftrightarrow \{u_1, \dots, u_k \mid \psi(u_1, \dots, u_k)\} \\ t^{(k)} &\longleftrightarrow u_1, \dots, u_k \\ R(t^{(k)}) &\longleftrightarrow R(u_1, \dots, u_k) \\ t^{(k)}[j] &\longleftrightarrow u_j \\ \exists t^{(k)} \phi(t^{(k)}) &\longleftrightarrow \exists u_1 \dots \exists u_k \psi(u_1, \dots, u_k) \\ \text{biztonságos} &\longleftrightarrow \text{biztonságos}\end{aligned}$$

# Lekérdezőnyelvek típusai, általános jellemzőik

- Lehetnek algebrai alapúak: relációs algebrán alapuló lekérdezés, procedurális leírás  
Előny: lekérdezésoptimalizálásra jobb
- Lehetnek logikai alapúak: sor vagy oszlopkalkulusra épülő lekérdezés, deklaratív leírás  
Előny: könnyebben átlátható



# Lekérdezőnyelvek

Általában a konkrét lekérdezőnyelvek eltérnek a modelltől (algebrai és logikai esetben is), van amiben többet tudnak, van amiben kevesebbet, vagy csak máshogy

Lehetséges eltérések:

- logikai alapúakban eleve csak biztonságos kifejezéseknek megfelelő kérdéseket lehet írni (nincsenek kvantorok)
- aritmetika
- aggregátumok
- többszörös sorok kezelése/megengedése
- attribútumok sorrendje kötött

De az igaz mindre, hogy relációsan teljeseek (esetleg bizonyos műveletek nehezebben mennek) és általában van a lekérdező funkció mellett DML és DDL is.