

Databases 1

Relációs algebra és nemrekurzív Datalog

Emlékeztető: Biztonságos szabályok

- ▶ A rule is *safe* if:
 1. Each distinguished variable,
 2. Each variable in an arithmetic subgoal, and
 3. Each variable in a negated subgoal,also appears in a nonnegated, relational subgoal, amivel az x korlátozott:
 - ▶ $\text{pred}(x, y, \dots)$ argumentuma (értéke a táblából)
 - ▶ vagy $x=c$ (konstans)
 - ▶ vagy $x=y$ (ahol y korlátozott)
- ▶ Safe rules prevent infinite results.

Példa: Nem biztonságos szabályokra

- ▶ Each of the following is unsafe and not allowed:
 1. $S(x) \leftarrow R(y)$
 2. $S(x) \leftarrow R(y) \text{ AND } x < y$
 3. $S(x) \leftarrow R(y) \text{ AND NOT } R(x)$
- ▶ In each case, an infinity of x 's can satisfy the rule, even if R is a finite relation.

Datalog kifejező ereje

- ▶ Without recursion, Datalog can express all and only the queries of core relational algebra.
 - ▶ The same as SQL select-from-where, without aggregation and grouping.
- ▶ But with recursion, Datalog can express more than these languages.

Relációs algebra és Datalog ---(1)

Rel.algebrai műveletek hogyan néznek ki Datalogban?

Halmazműveletek: T.f.h $R(x_1, \dots, x_n)$, $S(x_1, \dots, x_n)$

predikátumokhoz tartozó reláció $R(A_1, \dots, A_n)$, $S(A_1, \dots, A_n)$

▶ $R \cap S$ metszetnek megfelelő szabály:

Válasz(x_1, \dots, x_n) $\leftarrow R(x_1, \dots, x_n) \text{ AND } S(x_1, \dots, x_n)$

▶ $R - S$ különbségnek megfelelő szabály:

Válasz(x_1, \dots, x_n) $\leftarrow R(x_1, \dots, x_n) \text{ AND NOT } S(x_1, \dots, x_n)$

▶ $R \cup S$ unió műveletet egyetlen szabállyal nem tudom felírni, mert a törzsben csak AND lehet, OR nem. Ehhez több szabályból álló Datalog program kell:

Válasz(x_1, \dots, x_n) $\leftarrow R(x_1, \dots, x_n)$

Válasz(x_1, \dots, x_n) $\leftarrow S(x_1, \dots, x_n)$

Relációs algebra és Datalog ---(2)

Kiválasztás:

- ▶ $\sigma_{x_i \theta x_j}(R)$ kifejezésnek megfelelő szabály :
Válasz(x_1, \dots, x_n) $\leftarrow R(x_1, \dots, x_n)$ AND $x_i \theta x_j$
- ▶ $\sigma_{x_i \theta c}(E1)$ kifejezésnek megfelelő szabály:
Válasz(x_1, \dots, x_n) $\leftarrow R(x_1, \dots, x_n)$ AND $x_i \theta c$

Vetítés:

- ▶ $\Pi_{A_{i_1}, \dots, A_{i_k}}(R)$ kifejezésnek megfelelő szabály:
Válasz(x_{i_1}, \dots, x_{i_k}) $\leftarrow R(x_1, \dots, x_n)$

Megjegyzés: név nélküli anonymus változók, amelyek csak egyszer szerepelnek és mindegy a nevük azt aláhúzás helyettesítheti. Például:

HosszúFilm(c,é) \leftarrow Film(c,é,h,_,_,_) AND $h \geq 100$

Relációs algebra és Datalog ---(3)

Természetes összekapcsolás: Tegyük fel, hogy $R(A_1, \dots, A_n, C_1, \dots, C_k)$ és $S(B_1, \dots, B_m, C_1, \dots, C_k)$

- ▶ $R \bowtie S$ kifejezésnek megfelelő szabály:

Válasz($x_1, \dots, x_n, y_1, \dots, y_m, z_1, \dots, z_k$) \leftarrow

$\leftarrow R(x_1, \dots, x_n, z_1, \dots, z_k) \text{ AND } S(y_1, \dots, y_m, z_1, \dots, z_k)$

- ▶ A felírt szabályok biztonságosak.
- ▶ Minden Q relációs algebrai kifejezéshez van nem rekurzív, biztonságos, negációt is tartalmazó Datalog program, amelyben egy kitüntetett IDB predikátumhoz tartozó kifejezés ekvivalens a Q lekérdezéssel.
- ▶ A nem rekurzív, biztonságos, negációt is tartalmazó Datalog kifejezőerő tekintetében EKVIVALENS a relációs algebrával.

Példák relációs algebrai lekérdezésekre 1/5

- ▶ A korábbi feladatokat relációs algebra kifejezésekre fejezzük ki relációs algebrában kifejezőfával majd ezt alakítsuk át biztonságos Datalog szabályokká!
- ▶ **Példa:** Adottak az alábbi relációs sémák feletti relációk

Termék (gyártó, modell, típus)

PC (modell, sebesség, memória, merevlemez, cd, ár)

Laptop (modell, sebesség, memória, merevlemez, képernyő, ár)

Nyomtató (modell, 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 relációs algebrai lekérdezésekre 2/5

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

$$\Pi_{\text{mod}}(\sigma_{s \geq 3.00}(\mathbf{PC}))$$

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

$$\Pi_{\text{gy}}(\sigma_{m \geq 100}(\mathbf{T} \bowtie \mathbf{L})) \text{ vagy ekv. } \Pi_{\text{gy}}(\mathbf{T} \bowtie (\sigma_{m \geq 100}(\mathbf{L})))$$

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

három részből áll (Nyomtató táblánál vigyázni, uis term.összekapcsolásnál a típus attr. itt mást jelent!)

$$\begin{aligned} & \Pi_{\text{m, ár}}(\sigma_{\text{gy}='B'}(\mathbf{T}) \bowtie \mathbf{PC}) \cup \Pi_{\text{m, ár}}(\sigma_{\text{gy}='B'}(\mathbf{T}) \bowtie \mathbf{L}) \cup \\ & \cup \Pi_{\text{m, ár}}(\sigma_{\text{gy}='B'}(\mathbf{T}) \bowtie \Pi_{\text{m, ár}}(\mathbf{Ny})) \end{aligned}$$

Példák relációs algebrai lekérdezésekre 3/5

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

elvégezhető más módon is

$$\Pi_m(\sigma_{sz='i' \wedge t='lézer'}(\mathbf{Ny})) =$$

$$\Pi_m(\sigma_{sz='i'} \sigma_{t='lézer'}(\mathbf{Ny})) = \Pi_m(\sigma_{t='lézer'} \sigma_{sz='i'}(\mathbf{Ny}))$$

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**)

$$\Pi_{gy}(\sigma_{t='laptop'}(\mathbf{T})) - \Pi_{gy}(\sigma_{t='pc'}(\mathbf{T}))$$

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

$$\Pi_{ml}(\sigma_{ml=ml1 \wedge m \neq m1}(\mathbf{PC} \bowtie \rho_{PC_1}(m1, s1, me1, ml1, ár1) \mathbf{PC}))$$

Példák relációs algebrai lekérdezésekre 4/5

! 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. **Jelölje ezt PC1 =**

$$\Pi_{m, m1}(\sigma_{s=s1 \wedge me=me1 \wedge m < m1}(\mathbf{PC} \bowtie \rho_{\mathbf{PC1}}(m1, s1, me1, m1, ár1) \mathbf{PC}))$$

!! h) Melyek azok a gyártók, amelyek gyártanak legalább két, egymástól különböző, legalább 2.80 gigahertzen működő számítógépet (PC-t vagy laptopot)? **Vezessünk be munkatáblát (segédváltozót) a rész kifejezéshez:**

$$\mathbf{Gyors} = \Pi_{gy, m}(\mathbf{T} \bowtie (\sigma_{s \geq 2.8}(\mathbf{PC}))) \cup \Pi_{gy, m}(\mathbf{T} \bowtie (\sigma_{seb \geq 2.8}(\mathbf{L})))$$

és ezzel:

$$\Pi_{gy}(\sigma_{gy=gy1 \wedge m \neq m1}(\mathbf{Gyors} \times \rho_{\mathbf{Gy1}}(gy1, m1) \mathbf{Gyors}))$$

Példák relációs algebrai lekérdezésekre 5/5

!! i) Melyik gyártó gyártja a leggyorsabb számítógépet (PC-t vagy laptopot)? **Lásd „elhagyás” típusú lekérdezések...**

Kiválasztjuk azokat a SZ-eket, amelyeknél van gyorsabb, ha ezt kivonjuk a SZ-ékből megkapjuk a leggyorsabbat:

Lassú = $\Pi_m(\sigma_{s < s_1}(\mathbf{SZ} \bowtie \mathbf{SZ}_1))$ ahol SZ: pc vagy laptop
Leggyorsabb: $\Pi_m(\mathbf{SZ}) - \mathbf{Lassú}$ végül behelyettesítem

!! j) Melyik gyártó gyárt legalább három, különböző sebességű PC-t? **mint legalább kettő, ott 2x itt 3x mellé**

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

Mire érdemes felhívni a figyelmet?

- ▶ Relációs algebrai kifejezések átírása nemrekurzív biztonságos Datalogba.
- ▶ Mi a leggyakrabban előforduló típus, amiből építkeznek? $\prod_{Lista}(\sigma_{Felt}(R \bowtie S \bowtie \dots))$

Ezt a komponenst támogatja legerősebben az SQL is: **SELECT lista**

FROM táblák összekapcsolása

WHERE felt

Ez felel meg egy Datalog szabálynak,
halmazműveletek megadása több szabállyal

Következik

- ▶ Rekurzió Datalogban és
- ▶ az SQL-1999 szabvány WITH RECURSIVE