

# Relačný model

## Základné pojmy:

- množina, charakteristická funkcia množiny
- multimnožina, rozplizlá (fuzzy množina)
- prienik, zjednotenie, rozdiel
- relácie a tabuľky
- nadkľúč a kľúč

# Základné definície:

Unárna relácia:  $\langle D, R \subseteq D \rangle \quad P(x) = \mathcal{X}_R(x)$

$\mathcal{X}_R: D \rightarrow \text{Boolean}, \mathcal{X}_R(x) = \text{true}$ , práve vtedy keď  $x \in R$ .

Zovšeobecnenie n-árna relácia:

$\langle D_1, \dots, D_n, R \subseteq D_1 \times \dots \times D_n \rangle$

$\mathcal{X}_R(x_1, \dots, x_n) = \text{true}$ , práve vtedy keď  $\langle x_1, \dots, x_n \rangle \in R$ .

Z teoretického hľadiska rozlišovanie oborov definície nemá zmysel (je nepodstatnou variáciou). Prakticky n-tica oborov definície určuje typ relácie.

Skratka:  $\mathbf{x} = \langle x_1, \dots, x_n \rangle;$

# Relačné operácie

Nech  $R_1$  a  $R_2$  sú relácie rovnakého typu. Potom:

- Prienik  $R_1 \cap R_2 = \{\mathbf{x}: P_1(\mathbf{x}) \wedge P_2(\mathbf{x})\}$
- Zjednotenie  $R_1 \cup R_2 = \{\mathbf{x}: P_1(\mathbf{x}) \vee P_2(\mathbf{x})\}$
- Rozdiel  $R_1 - R_2 = \{\mathbf{x}: P_1(\mathbf{x}) \wedge \neg P_2(\mathbf{x})\}$

Nech  $R$  je typu  $\mathbf{x} \cup \mathbf{y}$  (Premenným priradíme typ podľa oboru, z ktorého môžu nadobúdať hodnoty.)

- Projekcia  $\Pi_{\mathbf{x}} R = \{\mathbf{x}: (\exists \mathbf{y}) P(\mathbf{x}\mathbf{y})\}$

# Relačné operácie (pokračovanie)

Nech  $R_1$  je typu  $\mathbf{x}$  a  $R_2$  je typu  $\mathbf{y}$  a typy  $\mathbf{x}$  a  $\mathbf{y}$  sú disjunktné

- Kartézsky súčin  $R_1 \times R_2 = \{\mathbf{xy}: P_1(\mathbf{x}) \wedge P_2(\mathbf{y})\}$

Nech  $R_1$  je typu  $\mathbf{xy}$  a  $R_2$  je typu  $\mathbf{y}$  a typy  $\mathbf{x}$  a  $\mathbf{y}$  sú disjunktné

- Podiel

$$R_1 : R_2 = \{\mathbf{x}: (\exists \mathbf{y}) P_1(\mathbf{xy}) \wedge (\forall \mathbf{y})(P_1(\mathbf{xy}) \Rightarrow P_2(\mathbf{y}))\}$$

**Veta:**  $R_1 : R_2 = \Pi_{\mathbf{x}} R_1 - \Pi_{\mathbf{x}} ((\Pi_{\mathbf{x}} R_1) \times R_2 - R_1)$

# Tabuľky

Tabuľka ako reprezentácia relácie

hlavička

|       | $x_1$ | . . . |  |  | $x_n$ |
|-------|-------|-------|--|--|-------|
| $r_1$ |       |       |  |  |       |
| $r_2$ |       |       |  |  |       |
| .     |       |       |  |  |       |
| .     |       |       |  |  |       |
| .     |       |       |  |  |       |

# Formalizácia pojmu tabuľka

## Definícia:

Dvojicu  $T = \langle H, R \rangle$ ,

kde  $H$  je množina dvojíc  $\{\langle \textit{meno}_i, \textit{doména}_i \rangle\}_{i=1}^n$

a  $R$  je podmnožina kartézského súčinu  $\times_{i=1}^n \textit{doména}_i$  nazývane tabuľkou. Množinu  $H$  nazývame hlavičkou tabuľky a prvky kartézského súčinu  $R$  riadkami tabuľky. Na riadok  $r$  sa môžeme pozerat' aj ako na funkciu z množiny mien (atribútov) do množiny hodnôt (jednotlivých domén).

## Konvencia:

Z hľadiska významu považujeme tabuľky s rovnakými hlavičkami a rovnakými množinami riadkov za ekvivalentné

# Operácie s tabuľkami

- Množinové operácie
- Projekcia
- Premenovanie (zmena hlavičky)
- Prirodzené spojenie (natural join)

Tabuľky sú skôr multimnožiny ako množiny riadkov.

Operácie sa niekedy vykonávajú s multimnožinami.

Na interpretáciu používame „množinovú ekvivalenciu“.

# „Množinové operácie“

Podmienkou pre nasledujúce operácie sú rovnaké hlavičky tabuliek a výsledku.

- Zjednotenie - zachovanie hlavičky a zlúčenie množín viet (riadkov).
- Rozdiel -  $R_1 - R_2$  z tabuľky  $R_1$  sa vynechajú všetky výskyty viet nachádzajúcich sa v tabuľke  $R_2$ .

Významove sú tieto operácie ekvivalentné rovnomenným relačným operáciám



# Premenovanie a projekcia

(unárne operácie)

- **Premenovanie** - Nech  $T = \langle H_1, R \rangle$ , kde  $H_1 = \{ \langle M_i, D_i \rangle : 1 \leq i \leq n \}$ , nech  $H_2 = \{ \langle N_i, \Delta_i \rangle : 1 \leq i \leq n \}$  a pre každé  $i$ ,  $D_i \subseteq \Delta_i$ . Potom tabuľku  $T' = \langle H_2, R \rangle$ , nazývame premenovaním tabuľky  $T$ .

Premenovanie je technická operácia umožňujúca premenovanie premenných.

- **Projekcia** - Odstránenie niektorých stĺpcov z hlavičky aj multimnožiny viet. Presnejšie povedané projekcia je obmedzenie tabuľky na podhlavičku.

Význam tejto operácie je rovnaký ako v relačných operáciách.

# Prirodzené spojenie (join)

Nech  $T_1 = \langle H_1, R_1 \rangle$ , kde  $H_1 = \{ \langle M_i, D_i \rangle : 1 \leq i \leq m \}$  a  $T_2 = \langle H_2, R_2 \rangle$ , kde  $H_2 = \{ \langle N_i, D_i \rangle : 1 \leq i \leq n \}$  a  $M_k = N_l \Rightarrow D_k = D_l$ .

- **Prirodzeným spojením**  $T = T_1 \bowtie T_2$  rozumieme tabuľku  $T = \langle H, R \rangle$  takú, že  $H = H_1 \cup H_2$  a  $R$  je množina všetkých takých riadkov  $r$ , že projekcia (restrikcia)  $r$  na  $H_1$  patrí do  $R_1$  a projekcia  $r$  na  $H_2$  patrí do  $R_2$ .

# Význam prirodzeného spojenia

Prirodzené spojenie realizuje logickú operáciu *and* v najvšeobecnejšom význame. Ak hlavičky relácií sú rovnaké prirodzené spojenie je prienikom, ak hlavičky relácií sú disjunktné prirodzené spojenie je kartézskym súčinom.

Pojem hlavičky spresňuje, čo je to typ relácie a premennej. Až na toto spresnenie sú tabuľky a relácie to isté.

# Zákony relačnej algebry

- Prirodzené spojenie a zjednotenie sú operácie komutatívne, asociatívne a idempotentné
- Platia distributívne zákony
  - $R \bowtie (S \cup T) = (R \bowtie S) \cup (R \bowtie T)$
  - $R \bowtie (S - T) = (R \bowtie S) - (R \bowtie T)$
- Ak  $\mathbf{y} \subseteq \mathbf{x}$ . Potom  $\Pi_{\mathbf{y}} \Pi_{\mathbf{x}} R = \Pi_{\mathbf{y}} R$
- Ak  $\mathbf{x}$  nepatrí medzi spoločné  $R$  a  $S$ . Potom  $\mathbf{x} = \mathbf{y} \cup \mathbf{z}$ , kde  $\mathbf{y}$  sú atribúty  $R$  a  $\mathbf{z}$  sú atribúty  $S$  a platí  $\Pi_{\mathbf{x}} (R \bowtie S) = (\Pi_{\mathbf{y}} R) \bowtie (\Pi_{\mathbf{z}} S)$

# Typy relácií (tabuliek)

- Extenzionálne – tie, čo sú uložené v báze dát. Vždy konečné.
- Zabudované – porovnávacie reťazcové a aritmetické predikáty. Obvykle nekonečné.
- Intenzionálne – tie, čo sú výsledkami dotazov, databázových programov.

# Efektívnosť relácií

## konečné a nekonečné relácie

Hoci domény môžu byť nekonečné spočítateľné množiny. Databázové relácie sú vždy konečné - tabuľky s konečným počtom riadkov.

Zaujíma nás či dokážeme tabuľku vypísať.

Príklady nekonečných tabuliek:

- $=(x,y)$ ,  $<(x,y)$ ,  $\arcsin(x, \sin x)$
- $\neq(x,y)$
- $\text{perverse}(a,b,c,n) \Leftrightarrow a^n + b^n = c^n$

Vstupné množiny, generátor, rozpoznávač

# Selekcia

Nech  $T = \langle H_1, R_1 \rangle$  je databázová relácia a nech

$F = \langle H_2, R_2 \rangle$  je nekonečná relácia a  $H_2 \subseteq H_1$ .

Potom namiesto  $R \bowtie F$  píšeme  $\sigma_F R$ .

- Selekcia  $\sigma_F R = \{\mathbf{x} : (\mathbf{x} \in R) \wedge \mathbf{x} \in F(\mathbf{x})\}$

**Veta:** Nech  $T = T_1 \bowtie T_2$  a nech  $T_1 = \langle H_1, R_1 \rangle$ , kde

$H_1 = \{\langle M_i, D_i \rangle : 1 \leq i \leq m\}$  a  $T_2 = \langle H_2, R_2 \rangle$ , kde

$H_2 = \{\langle N_i, D_i \rangle : 1 \leq i \leq n\}$ . Označme  $F = \bigwedge M_k = N_l$

pre  $M_k \in H_1 \cap H_2$ . Nech  $T = \langle H, R \rangle$  a  $\mathbf{z} = H_1 \cup H_2$ ,

potom  $R = \Pi_{\mathbf{z}}(\sigma_F(R_1 \times R_2))$ .

# Databáza

Databáza je množina domén a konečných relácií nad týmito doménami.

Dotazy sú výrazy relačnej algebry.

Odpoveď na dotaz získame výpočtom príslušného výrazu.

Je zjavný súvis medzi formulami predikátového počtu a výrazmi relačnej algebry.



# Relačný kalkúl

Formule predikátového kalkulu:

- Negácia sa používa len v pozitívnom kontexte t.j.  $E_1(\mathbf{x}) \wedge \neg E_2(\mathbf{x})$  (to platí rekurzívne aj pre podformuly).
- Kontext univerzálneho kvantifikátoru je relativizovaný na nejaký pozitívny kontext  $(\exists \mathbf{y})E_1(\mathbf{xy}) \wedge (\forall \mathbf{y})(E_1(\mathbf{xy}) \Rightarrow E_2(\mathbf{y}))$  (znovu sa to používa rekurzívne).

Nededuktívny charakter -

dotazy sa vzťahujú na konkrétny model.

# Teória dotazov

Databáza je štruktúra pre predikátový kalkul

Dotaz je formula  $\varphi(\mathbf{x})$  s množinou volných premenných  $\mathbf{x}$

Odpoved' na dotaz  $\varphi(\mathbf{x})$  je  $\{\mathbf{x}: DB \models \varphi(\mathbf{x})\}$

$\phi = false; \quad \{\phi\} = true$

# Funkčné závislosti, kľúče

Hovoríme, že v relácií  $R(\mathbf{x}, \mathbf{y}, \mathbf{z})$  platí funkčná závislosť  $\mathbf{x} \rightarrow \mathbf{y}$  práve vtedy, keď

$$(\forall \mathbf{x} \mathbf{y}_1 \mathbf{y}_2 \mathbf{z}_1 \mathbf{z}_2) R(\mathbf{x}, \mathbf{y}_1, \mathbf{z}_1) \wedge R(\mathbf{x}, \mathbf{y}_2, \mathbf{z}_2) \Rightarrow \mathbf{y}_1 = \mathbf{y}_2$$

Nech  $R(\mathbf{x})$  je relácia a  $\mathbf{k} \subseteq \mathbf{x}$  a platí  $\mathbf{k} \rightarrow \mathbf{x}$ .  
Potom  $\mathbf{k}$  sa nazýva nadkľúč.

Minimálny (v zmysle množinovej inklúzie) nadkľúč sa nazýva kľúč.

# Vstupné množiny, pseudokľúče

Def: Nech  $R(\mathbf{x}, \mathbf{y})$  je relácia, hovoríme, že množina atribútov  $\mathbf{x}$  je vstupná množina pre reláciu  $R$ , ak množina  $\{\mathbf{y}: R(\mathbf{x}, \mathbf{y})\}$  je konečná.

Def: Minimálnu vstupnú množinu v zmysle množinovej inklúzie nazývame ***pseudokľúč***.

Pre konečné relácie pseudokľúč je  $\emptyset$ .

# Vlastnosti a použitie pseudokľúčov

- Pseudokľúč je použiteľný ak databáza má k dispozícii metódu (algoritmus) ako pre danú hodnotu  $\mathbf{x}$  určiť  $\{\mathbf{y}: R(\mathbf{x}, \mathbf{y})\}$ .
- $\sigma_F R = R \bowtie F$ .
- Každý kľúč je vstupná množina a nadmnožina nejakého pseudokľúča.

# Agregácia a agregáčné funkcie

Motivácia: obvyklé „štatistické“ otázky.

Def.: Nech  $R(\mathbf{x}, y, \mathbf{z})$  je konečná relácia a  $f$  je niektorá s funkcií *count*, *sum*, *max*, *min*.

Definujeme  $\Gamma_{\mathbf{x}} R$  ako tabuľku  $R'(\mathbf{x}, y)$ ,  
 $y \leftarrow f(y)$

ktorú dostaneme z tabuľky  $R$  utriedením a vypočítaním funkcie  $f$  pre skupiny s rovnakou hodnotou  $\mathbf{x}$ .

Agregačná funkcia je funkcia 2. rádu prirad'uje množinám hodnoty.

# Agregačný kvantifikátor

Nech  $P(\mathbf{x}, y, \mathbf{z})$  je predikát zodpovedajúci relácii  $R(\mathbf{x}, y, \mathbf{z})$  definujeme:

$$(\bigcup \mathbf{z}, y = f(y)) P(\mathbf{x}, y, \mathbf{z})$$

ako predikát zodpovedajúci relácii

$$\begin{array}{c} \Gamma \\ \mathbf{x} \\ y \leftarrow f(y) \end{array} R(\mathbf{x}, y, \mathbf{z})$$

$$= \{xf(y): (\bigcup \mathbf{z}, y = f(y)) P(\mathbf{x}, y, \mathbf{z})\}$$

# Relačná algebra - sumarizácia

- Nosič: *Množina relácií*
- Konštanty: Konkrétne relácie podľa potreby
- Unárne operácie:
  - projekcia, premenovanie atribútov  $\Pi$
  - agregácia  $\Gamma$
- Binárne operácie
  - kartézsky súčin  $\times$ , prienik  $\cap$ , prirodzené spojenie  $\bowtie$ , selekcia  $\sigma$
  - rozdiel  $-$
  - zjednotenie  $\cup$



# Riešenie rovníc v relačnej algebre

System rovníc  $\bar{R} = \bar{E}(\bar{R}, \bar{R}_0)$

Ak výrazy  $E$  neobsahujú rozdiel a agregáciu

Riešime naivnou iteráciou:

- Začneme  $\bar{R} = \emptyset$
- Opakujeme  $\bar{R}^{i+1} = \bar{E}(\bar{R}^i, \bar{R}_0)$
- Pokiaľ niečo pribúda

# Tarského veta o pevnom bode

Úplný звяз:  $S = \langle D, \sqsubseteq, \sqcup, \sqcap, \perp, \top \rangle$

$\sqsubseteq$  čiastočné usporiadanie

$\sqcup$  l.u.b., sup, join                       $\perp$  dolník

$\sqcap$  g.u.b., inf, meet                       $\top$  horník

Každá neprázdna množina má l.u.b. (suprémum).

**Veta:** Nech  $\mathbf{F}$  je zobrazenie z úplného zväzu  $S$  do  $S$  také, že  
 $x \sqsubseteq y \Rightarrow \mathbf{F}(x) \sqsubseteq \mathbf{F}(y)$ . Potom  $\mathbf{F}$  má aspoň jeden pevný bod.

**Dôkaz:** Nech  $U = \{x : x \sqsubseteq \mathbf{F}(x)\}$ . Množina  $U$  je neprázdna,  $\perp \in U$ .

Označme:  $x_0 = \bigsqcup_{x \in U} x$ . Pre každé  $x \in U$  platí  $x \sqsubseteq \mathbf{F}(x) \sqsubseteq x_0$ .

Preto aj  $x_0 \sqsubseteq \mathbf{F}(x_0)$  t.j.  $x_0 \in U$ . Preto aj  $\mathbf{F}(x_0) \sqsubseteq \mathbf{F}(\mathbf{F}(x_0))$  a  
 $\mathbf{F}(x_0) \in U$ . Z toho ale plynie  $\mathbf{F}(x_0) \sqsubseteq x_0$ . Teda  $x_0 = \mathbf{F}(x_0)$ .