Relačný model

Základné pojmy:

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

Základné definície:

Unárna relácia: $<D, R \subseteq D>$ $P(x)=X_R(x)$

$X_R: D \rightarrow Boolean$, $X_R(x)=true$, práve vtedy keď $x \in R$.

Zovšeobecnenie n-árna relácia:

$<D_1, \ldots, D_n, R \subseteq D_1 \times \ldots \times D_n>$

$X_R(x_1, \ldots, x_n)=true$, práve vtedy keď $< x_1, \ldots, x_n > \in R$.

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

Skratka: $\mathbf{x} = < x_1, \ldots, x_n >$;

Relačné operácie

Nech R_1 a R_2 sú relácie rovnakého typu. Potom:

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

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

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

Relačné operácie (pokračovanie)

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

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

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

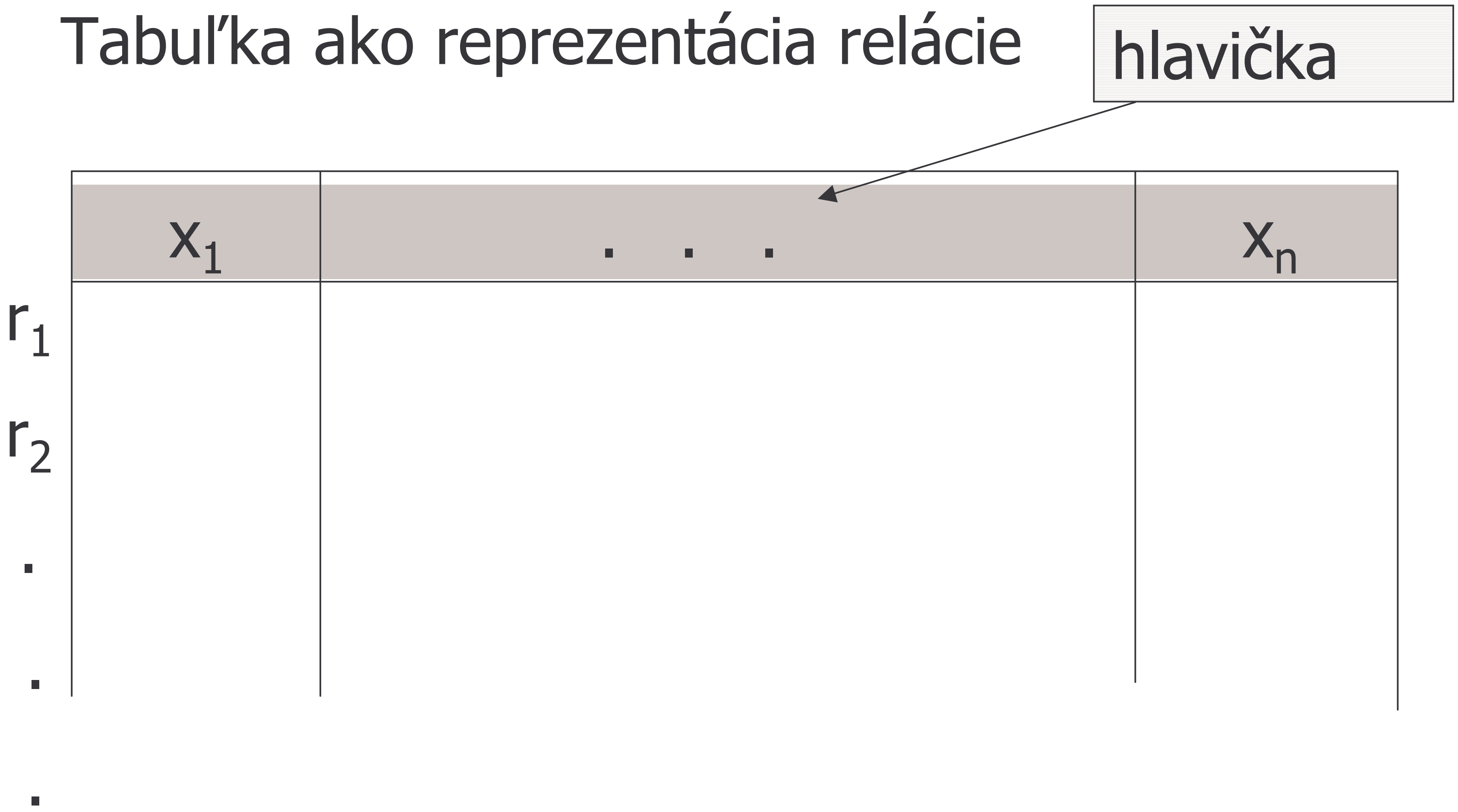
- Podiel

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

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

Tabuľky

Tabuľka ako reprezentácia relácie



Formalizácia pojmu tabuľka

Definícia:

Dvojicu $T = <H, R>$,

kde H je množina dvojíc $\{<meno_i, doména_i>\}_{i=1}^{n}$

a R je podmultimnožina kartézskeho súčinu $\times_{i=1}^{n}$ doména$_i$

nazývane <u>tabuľkou</u>. Množinu H nazývame <u>hlavičkou tabuľky</u> a prvky kartézského súčinu R <u>riadkami tabuľky</u>. Na riadok r sa môžeme pozerať aj ako na funkciu z množiny mien (atribútov) do množiny hodnôt (jednotlivých domén).

Konvencia:

Z hľadiska významu považujeme tabuľky s rovnakými hlavičkami a rovnakými <u>množinami riadkov</u> 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 multimnožín viet (riadkov).
- Rozdiel - $R_1 - R_2$ z tabuľky R_1 sa vynechajú všetky výskyty viet nachádzajúcich sa v tabuľke R_1.

Významove sú tieto operácie ekvivalentné rovnomenným relačným operáciam

Premenovanie a projekcia
(unárne operácie)

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

  Premenovanie je technická operácia umožňujúca zníženie obmedzení na dáta a 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 $T_1 = <H_1, R_1>$, kde $H_1 = \{<M_i, D_i>: 1 \leqslant i \leqslant m\}$ a $T_2 = <H_2, R_2>$, kde $H_2 = \{<N_i, Đ_i>: 1 \leqslant i \leqslant n\}$ a $M_k = N_l \Rightarrow D_k = Đ_l$.

- **Prirodzeným spojením** $T = T_1 \bowtie T_2$ rozumieme tabuľku $T = <H, R>$ 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 $\boldsymbol{y} \subseteq \boldsymbol{x}$. Potom $\Pi_{\boldsymbol{y}} \Pi_{\boldsymbol{x}} R = \Pi_{\boldsymbol{y}} R$
- Ak $\boldsymbol{x}$ nepatrí medzi spoločné R a S. Potom $\boldsymbol{x} = \boldsymbol{y} \cup \boldsymbol{z}$, kde $\boldsymbol{y}$ sú atribúty R a $\boldsymbol{z}$ sú atribúty S a platí $\Pi_{\boldsymbol{x}} (R \bowtie S) = (\Pi_{\boldsymbol{y}} R) \bowtie (\Pi_{\boldsymbol{z}} R)$

Efektívnosť relácií konečné a nekonečné relácie

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

Zaujíma nás či dokážeme tabuľku vypísať.

Príklady nekonečných tabuliek:

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

Vstupné množiny, generátor, rozpoznávač

Selekcia

Nech $T = <H_1, R_1>$ je databázová relácia a nech F $= <H_2, R_2>$ je nekonečná relácia a $H_2 \subseteq H_1$. Potom namiesto $R \bowtie F$ píšeme $\sigma_F R$.

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

Veta: Nech $T = T_1 \bowtie T_2$ a nech $T_1 = <H_1, R_1>,$ kde $H_1 = \{<M_i, D_i>: 1 \leqslant i \leqslant m\}$ a $T_2 = <H_2, R_2>,$ kde $H_2 = \{<N_i, Đ_i>: 1 \leqslant i \leqslant n\}.$ Označme F = $\omega M_k = N_l$ pre $M_k \in H_1 \wedge H_2$. Nech $T = <H, R>$ a $\mathbf{z} = H_1 \wedge H_2,$ 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 kalkul

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 konkretný model.

Teória dotazov

Databáza je štruktúra pre predikátový kalkul

Dotaz je formula $\varphi(\boldsymbol{x})$ s množinou voľných premenných $\boldsymbol{x}$

Odpoveď na dotaz $\varphi(\boldsymbol{x})$ je $\{\boldsymbol{x}: DB \models \varphi(\boldsymbol{x})\}$

ϕ = *false*; $\{\phi\}$ = *true*

Funkčné závislosti, kľúče

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

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

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

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