

Úvod do databázových systémů

9. cvičení

Ing. Petr Lukáš

petr.lukas@nativa.cz

Ostrava, 2012

Teorie zpracování dat

- Intuitivní návrh schématu databáze je pro praktické účely obvykle dostačující
- Chceme-li důkladněji zkoumat vztahy mezi daty, je potřeba tyto vztahy formalizovat
- Existuje teorie, která se tímto problémem zabývá

Pojmy

Univerzální relační schéma

Funkční závislost

Armstrongovy axiomy

Uzávěry množiny atributů

Klíč schématu

Minimální neredundantní pokrytí

Univerzální relační schéma

Předpis pro relaci se všemi možnými atributy.

RU (idZam, jméno, příjmení, idKatedry, název, předmět, kredity, učebna, hodina)

idZam	jméno	příjmení	idKatedry	název	předmět	kredity	učebna	hodina
1	Petr	Lukáš	1	ABC	UDBS	6	A1033	16:00
2	Pavel	Mladý	2	DEF	ML	6	A1034	17:30
3	Jan	Starý	2	DEF	TIS	4	A1033	09:00
1	Petr	Lukáš	1	ABC	DAIS	4	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
4	Karel	Velký	3	XYZ	ML	6	E320	09:00
5	Josef	Dlouhý	3	XYZ	UDBS	6	E320	10:45
7	Martin	Zelený	1	ABC	TIS	4	E322	12:30
2	Pavel	Mladý	2	DEF	DAIS	4	A1033	12:30

Funkční závislost

$$X \rightarrow Y$$
$$X, Y \subset RU$$

Množina atributů Y je funkčně závislá na množině atributů X pokud při stejných hodnotách X jsou stejné hodnoty Y .

Funkční závislost

{idZam} → {jméno, příjmení}

Tam kde bude stejná hodnota *idZam*, budou stejné i hodnoty *jméno* a *příjmení*. Jinými slovy – jestliže znám *idZam*, znám i *jméno* a *příjmení*.

{hodina, učebna} → {předmět}

Pokud znám kombinaci **hodina** a **učebna**, pak znám i **předmět**. Neboli ve všech záznamech, kde je stejná kombinace **hodina** vs. **učebna** bude stejná hodnota **předmět**.

Funkční závislost

RU (idZam, jméno, příjmení, předmět, kredity, učebna, hodina)

Které z následujících funkčních závislostí jsou platné?

{idZam} → {jméno}

{idZam} → {příjmení}

{jméno} → {příjmení}

{učebna} → {hodina}

{předmět} → {kredity}

{idZam, předmět} → {učebna}

{idZam, hodina} → {předmět}

{idZam} → {idZam}

{idZam, jméno} → {příjmení}

{předmět} → {jméno, kredity}

Funkční závislost

RU (idZam, jméno, příjmení, předmět, kredity, učebna, hodina)

Které z následujících funkčních závislostí jsou platné?

{idZam} → {jméno}

{idZam} → {příjmení}

{idZam, hodina} → {předmět}

{idZam} → {idZam}

{idZam, jméno} → {příjmení}

{předmět} → {kredity}

Armstrongovy axiomy

Sada pravidel pro odvozování dalších funkčních závislostí z dané množiny funkčních závislostí (dále jen FZ).

1. Dekompozice
2. Sjednocení
3. Tranzitivita
4. Rozšíření

Armstrongovy axiomy

Dekompozice

$$A \rightarrow b, c, d$$

$$\Downarrow$$

$$A \rightarrow b$$

$$A \rightarrow c$$

$$A \rightarrow d$$

Jestliže si z množiny hodnot atributů **A** umím odvodit kombinaci hodnot atributů **b, c, d**, určitě si umím odvodit i hodnoty těchto atributů zvlášť.

Armstrongovy axiomy

Sjednocení

$$A \rightarrow b$$

$$A \rightarrow c$$

$$A \rightarrow d$$

\Downarrow

$$A \rightarrow b, c, d$$

Pokud si z množiny atributů **A** umím samostatně odvodit hodnoty atributů **b, c, d**, pak si přirozeně umím odvodit kombinaci hodnot **b, c, d**.

Armstrongovy axiomy

Rozšíření

$$\begin{array}{c} A \rightarrow B \\ \Downarrow \\ AZ \rightarrow BZ \end{array}$$

Pokud vím, že z množiny hodnot atributů **A** si umím odvodit hodnoty **B**, pak upřesněním levé strany o množinu hodnot atributů **Z** jednak nepřijdu o hodnoty **B** a jednak budu triviálně znát i hodnoty **Z**.

Armstrongovy axiomy

Tranzitivita

$$A \rightarrow B$$

$$B \rightarrow C$$

$$\Downarrow$$

$$A \rightarrow C$$

Pokud vím, že z množiny hodnot atributů **A** si umím odvodit hodnoty **B** a z množiny hodnot **B** odvodit hodnoty **C**, umím si z **A** odvodit **C**.

Uzávěr množiny atributů

Uzávěrem množiny atributů X vzhledem k množině funkčních závislostí F je množina všech funkčně závislých atributů na X . Tento uzávěr označujeme jako X^+ .

Jinými slovy, pokud mám množinu atributů X , které všechny atributy si můžu odvodit?

Uzávěr množiny atributů

$\{\text{idZam}\} \rightarrow \{\text{jméno, příjmení, idKatedry}\}$

$\{\text{idKatedry}\} \rightarrow \{\text{název, adresa}\}$

$\{\text{předmět}\} \rightarrow \{\text{kredity}\}$

$\{\text{idZam, hodina}\} \rightarrow \{\text{předmět, učebna}\}$

$\{\text{učebna, hodina}\} \rightarrow \{\text{idZam, předmět}\}$

tzv. triviální funkční závislost

$\{\text{idZam}\}^+ = \{\text{idZam}\}$

Uzávěr množiny atributů

$\{\text{idZam}\} \rightarrow \{\text{jméno, příjmení, idKatedry}\}$

$\{\text{idKatedry}\} \rightarrow \{\text{název, adresa}\}$

$\{\text{předmět}\} \rightarrow \{\text{kredity}\}$

$\{\text{idZam, hodina}\} \rightarrow \{\text{předmět, učebna}\}$

$\{\text{učebna, hodina}\} \rightarrow \{\text{idZam, předmět}\}$

jednoduše použiju FZ

$\{\text{idZam}\}^+ = \{\text{idZam, jméno, příjmení, idKatedry}\}$

Uzávěr množiny atributů

$\{\text{idZam}\} \rightarrow \{\text{jméno, příjmení, idKatedry}\}$

$\{\text{idKatedry}\} \rightarrow \{\text{název, adresa}\}$

$\{\text{předmět}\} \rightarrow \{\text{kredity}\}$

$\{\text{idZam, hodina}\} \rightarrow \{\text{předmět, učebna}\}$

$\{\text{učebna, hodina}\} \rightarrow \{\text{idZam, předmět}\}$

z idKatedry se dostanu na její název a adresu

$\{\text{idZam}\}^+ = \{\text{idZam, jméno, příjmení, idKatedry, název, adresa}\}$

Uzávěr množiny atributů

$\{\text{idZam}\} \rightarrow \{\text{jméno, příjmení, idKatedry}\}$

$\{\text{idKatedry}\} \rightarrow \{\text{název, adresa}\}$

$\{\text{předmět}\} \rightarrow \{\text{kredity}\}$

$\{\text{idZam, hodina}\} \rightarrow \{\text{předmět, učebna}\}$

$\{\text{učebna, hodina}\} \rightarrow \{\text{idZam, předmět}\}$

Žádné další funkční závislosti mi už množinu nerošíří

$\{\text{idZam}\}^+ = \{\text{idZam, jméno, příjmení, idKatedry, název, adresa}\}$

Uzávěr množiny atributů

Algoritmus

Vytvářím uzávěr množiny atributů X vzhledem k dané množině funkčních závislostí F .

1. Do X^+ umístím všechny atributy z X
2. V F vyhledám takovou FZ , jejíž atributy na levé straně jsou obsaženy ve vznikající X^+ .
3. Do vznikajícího výsledku, tzn. X^+ vložím atributy z pravé strany nalezené FZ .
4. Pokračuji bodem 2 dokud se mi výsledek X^+ zvětšuje.

Klíč schématu relace

Množina atributů K je klíčem R , pokud jsou všechny atributy z R funkčně závislé na K .

Jinými slovy uzávěr množiny K obsahuje všechny atributy z R .

Klíčů může být více, obvykle nás zajímá pouze ten nejkratší.

Klíč schématu relace

$\{\text{idZam}\} \rightarrow \{\text{jméno, příjmení, idKatedry}\}$

$\{\text{idKatedry}\} \rightarrow \{\text{název, adresa}\}$

$\{\text{předmět}\} \rightarrow \{\text{kredity}\}$

$\{\text{idZam, hodina}\} \rightarrow \{\text{předmět, učebna}\}$

$\{\text{učebna, hodina}\} \rightarrow \{\text{idZam, předmět}\}$

$\{\text{idZam, hodina}\}_+ = \{\text{idZam, hodina, jméno, příjmení, idKatedry, název, adresa, učebna, předmět, kredity}\}$

Tzn. znám-li idZam a hodinu, pak už vím úplně všechno!

Minimální neredundantní pokrytí

Máme dānu ňĕjakou množinu funkčních závislostí F . Nĕkteré FZ v této množinĕ mohou být nadbytečné. Tak samo ňĕkteré atributy na levé stranĕ ňĕkterých FZ mohou být nadbytečné.

1. Rozložíme dané FZ na tzv. elementární FZ.
2. Předpočítáme si uzávĕry pro levé strany všech FZ.
3. Nalezneme a odstraníme redundantní atributy na levých stranách FZ.
4. Nalezneme a odstraníme celé redundantní FZ.

Minimální neredundantní pokrytí

Rozklad na elementární FZ

Mějme schéma $R(X, Y, Z)$ a množinu FZ $\{X \rightarrow YZ, Y \rightarrow XZ\}$

Po rozkladu na elementární pravidla bude množina FZ vypadat:

- $X \rightarrow Y$
- $X \rightarrow Z$
- $Y \rightarrow X$
- $Y \rightarrow Z$

Minimální neredundantní pokrytí

Odstranění nadbytečných atributů (hledání minimálního pokrytí)

Mějme schéma $R(A, B, C, D, E)$ a množinu FZ
 $\{ABC \rightarrow D, E \rightarrow C, AB \rightarrow E, C \rightarrow E\}$

- i) Testujeme samozřejmě jen tam, kde je na levé straně více atributů.
Tzn. začneme závislostí $ABC \rightarrow D$.
 $\{A, B, C\}^+ = \{A, B, C, D, E\}$
- ii) Zkusíme postupně odstranit atributy a zjistit, zda se mi nějak změní uzávěr. Tzn. zkusím odstranit atribut C – mám tedy $AB\cancel{C} \rightarrow D$.
- iii) Zkusím znova vytvořit uzávěr tentokrát $\{A, B\}^+$ (s upravenou FZ).
 $\{A, B\}^+ = \{A, B, C, D, E\}$
- iv) Vidím, že uzávěr se nezměnil, **atribut je nadbytečný**, takže jej z dalšího zpracování můžu skutečně vyloučit.

Minimální neredundantní pokrytí

Odstranění redundantních pravidel

Mějme schéma $R(A, B, C)$ a množinu FZ $\{A \rightarrow B, A \rightarrow C, B \rightarrow A, B \rightarrow C\}$.

- i) Pokusíme se postupně odstraňovat jednotlivé FZ a zjišťovat, zda se mi změní uzávěr jejich levých stran.
- ii) Podívejme se např. na závislost $A \rightarrow C$.
 $\{A\}^+ = \{A, B, C\}$
- iii) Zkusme celou tuto FZ odstranit za znova zkonstruovat uzávěr.
 $\{A\}^+ = \{A, B, C\}$
- iv) Vidíme, že uzávěr se nezměnil, takže FZ $A \rightarrow C$ je redundantní (nadbytečná) a z dalšího zpracování ji vypustíme.

Shrnutí

Univerzální relační schéma – „široká“ nepřehledná tabulka, která obsahuje všechny atributy (např. modelovaného systému)

Funkční závislost – ze znalosti hodnot nějaké množiny hodnot atributů (a samozřejmě obsahu databáze) znám i množinu hodnot jiných atributů.

Armstrongovy axiomy – odvozovací pravidla pro funkční závislosti.

Uzávěry množiny atributů – hodnoty kterých všech atributů jsem schopný získat na základě určité dané množiny atributů

Klíč schématu – atributy, kterými jednoznačně identifikuji celý záznam, tzn. pokud znám hodnoty těchto atributů, umím v univerzálním schématu dohledat obsah celého záznamu

Minimální neredundantní pokrytí – rozložím FZ na elementární FZ, z levých stran odstraním nadbytečné atributy a nakonec celé nadbytečné FZ.

Úvod do databázových systémů

www.dbedu.cs.vsb.cz

jméno: **student**

heslo: **tuodbedu**

Vpravo sloupec -> České kurzy -> UDBS -> Cvičení – Relaçní algebra