

Úvod do databázových systémů

11. cvičení

Ing. Petr Lukáš
petr.lukas@nativa.cz
Ostrava, 2013

- Intuitivní návrh schématu databáze je pro praktické účely při troše zkušeností obvykle dostačující
- **Jak ale poznáme, že databáze je navržena správně? Existuje nějaká „míra“ správnosti návrhu databáze?**
- Existuje teorie, která se těmito problémy zabývá a která navíc definuje postupy, jak automaticky dojít ke správnému databázovému schématu (tzn. ke správně navrženým tabulkám)

Teorie zpracování dat

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
- Sestavení univerzálního relačního schématu je vždy prvním krokem automatického návrhu databáze

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Funkční závislost

$$X \rightarrow Y$$

$$X, Y \subset RU$$

Nechť $R(\{A_1, A_2, \dots, A_n\})$ je relační schéma a X, Y jsou podmnožiny atributů $\{A_1, A_2, \dots, A_n\}$.

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je $\{\text{idZam}\} \rightarrow \{\text{jméno, příjmení}\}$ platná FZ ?

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je $\{\text{idZam}\} \rightarrow \{\text{jméno, příjmení}\}$ platná FZ ?

ANO, všude, kde jsou stejné hodnoty atributu **idZam**, jsou i stejné hodnoty atributů **jméno** a **příjmení**.

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je $\{\text{předmět}\} \rightarrow \{\text{učebna}\}$ platná FZ ?

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je {**předmět**} \rightarrow {**učebna**} platná FZ ?

Určitě **NE**, jednou máme TIS na A1033, podruhé na G317b.
Podobný problém je v případě DAIS.

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je $\{\text{předmět, hodina}\} \rightarrow \{\text{učebna}\}$ platná FZ ?

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je $\{\text{předmět, hodina}\} \rightarrow \{\text{učebna}\}$ platná FZ ?

ANO, všude, kde je stejná kombinace hodnot atributů **předmět** a **hodina**, jsou i stejné hodnoty atributu **učebna**.

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je $\{\text{jméno}\} \rightarrow \{\text{příjmení}\}$ platná FZ ?

Funkční závislost

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro **každou možnou aktuální** relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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	6	G317b	10:45
2	Pavel	Mladý	2	DEF	TIS	4	G317a	09:00
2	Pavel	Mladý	2	DEF	DAIS	6	A1033	12:30

Je $\{\text{jméno}\} \rightarrow \{\text{příjmení}\}$ platná FZ ?

Na první pohled to sice vypadá, že ano, ale pozor! V definici se mluví o **každé možné aktuální relaci**. Určitě je možné, že bude existovat relace, kde to platit nebude. Takže odpověď je **NE**.

Funkční závislost

$$X \rightarrow Y$$

$$X, Y \subset RU$$

- Intuitivně platí představa, že znám-li hodnoty atributů X , pak umím jednoznačně dohledat („odvodit“) hodnoty atributů Y .
- Pozor, tato představa sice je většinou správná, ale přesná definice říká něco jiného.

Množina atributů Y je funkčně závislá na množině atributů X , píšeme $X \rightarrow Y$, když pro každou možnou aktuální relaci R platí, že mají-li dva prvky relace R stejné hodnoty atributů X , pak mají i stejné hodnoty atributů Y .

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**

Dekompozice

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

$$\Downarrow$$

$$A \rightarrow b$$

$$A \rightarrow c$$

$$A \rightarrow d$$

Jestliže si podle množiny hodnot atributů **A** umím dohledat kombinaci hodnot atributů **b, c, d**, jistě si umím dohledat i hodnoty těchto atributů samostatně.

Sjednocení

$$A \rightarrow b$$

$$A \rightarrow c$$

$$A \rightarrow d$$

\Downarrow

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

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

Rozšíření

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

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

Tranzitivita

$$A \rightarrow B$$

$$B \rightarrow C$$



$$A \rightarrow C$$

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

Uzávěr množiny atributů

Nechť X je množina atributů a F množina funkčních závislostí.

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 budu znát hodnoty atributů X , které všechny další hodnoty (atributy) umím podle funkčních závislostí dohledat?

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

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

Algoritmus

Vytvářím uzávěr X^+ 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^+ mění.

Klíč schématu relace

Nechť $R(\{A_1, A_2, \dots, A_n\})$ je relační schéma.

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

Jak zjistím, že nějaká množina atributů K je skutečně klíč? Jednoduše. Sestrojím si uzávěr K^+ a zjistím, jestli obsahuje všechny atributy.

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

Klíč schématu relace

{idZam} → {jméno, příjmení, idKatedry}

{idKatedry} → {název, adresa}

{předmět} → {kredity}

{idZam, hodina} → {předmět, učebna}

{učebna, hodina} → {idZam, předmět}

Je atribut **idZam** klíčem univerzálního schématu?

Klíč schématu relace

{idZam} → {jméno, příjmení, idKatedry}

{idKatedry} → {název, adresa}

{předmět} → {kredity}

{idZam, hodina} → {předmět, učebna}

{učebna, hodina} → {idZam, předmět}

Je atribut **idZam** klíčem univerzálního schématu?

{idZam}₊ = {idZam, jméno, příjmení, idKatedry, název, adresa}

NE, uzávěr neobsahuje všechny atributy.

Klíč schématu relace

{idZam} → {jméno, příjmení, idKatedry}

{idKatedry} → {název, adresa}

{předmět} → {kredity}

{idZam, hodina} → {předmět, učebna}

{učebna, hodina} → {idZam, předmět}

Jsou atributy **idZam** a **hodina** dohromady klíčem univerzálního schématu?

Klíč schématu relace

{idZam} → {jméno, příjmení, idKatedry}

{idKatedry} → {název, adresa}

{předmět} → {kredity}

{idZam, hodina} → {předmět, učebna}

{učebna, hodina} → {idZam, předmět}

Jsou atributy **idZam** a **hodina** dohromady klíčem univerzálního schématu?

{idZam, hodina}⁺ = {idZam, hodina, jméno, příjmení, idKatedry, název, adresa, učebna, předmět, kredity}

ANO, pokud znám **idZam** a **hodinu**, pak už umím najít hodnoty všech ostatních atributů.

Minimální neredundantní pokrytí

Máme dānu ňĕjakou množinu funkčnĕch zāvislostĕ F . Nĕkterĕ FZ v tĕto množinĕ mohou bĕt nadbytečnĕ. Tak samo ňĕkterĕ atributy na levĕ stranĕ ňĕkterĕch FZ mohou bĕt nadbytečnĕ.

- 1. Rozložíme dané FZ na tzv. elementární FZ**
- 2. Předpočítáme si uzávĕry pro levĕ strany vĕsch FZ**
- 3. Nalezneme a odstranĕme redundantnĕ atributy na levĕch stranĕch FZ**
- 4. Nalezneme a odstranĕme celĕ redundantnĕ FZ**

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.

www.dbedu.cs.vsb.cz

- Přihlášení přes **jednotný login a heslo**
- Vpravo sloupec -> *České kurzy* -> *UDBS*