

Databázové systémy

Úvod do teorie normalizace

Vilém Vychodil

KMI/DATA1, Přednáška 12

Databázové systémy

Přednáška 12: Přehled

- 1 Boyce-Coddova normální forma:
 - formulace normální formy,
 - dekompozice na základě funkčních závislostí,
 - normalizace pomocí dekompozice.

Opakování: Funkční závislosti (PŘEDNÁŠKA 11)

Definice (funkční závislost, angl.: *functional dependency*)

Nechť R je relační schéma. Pak **funkční závislost** nad schématem R je formule ve tvaru $A \Rightarrow B$, kde $A, B \subseteq R$.

Definice (pravdivost funkční závislosti v datech)

Nechť R je relační schéma a \mathcal{D} je relace nad schématem R . Pak funkční závislost $A \Rightarrow B$ nad schématem R je **pravdivá** v \mathcal{D} , což označujeme $\mathcal{D} \models A \Rightarrow B$, pokud pro každé n -tice $r_1, r_2 \in \mathcal{D}$ platí:

$$\text{pokud } r_1(A) = r_2(A), \text{ pak } r_1(B) = r_2(B).$$

V opačném případě říkáme, že $A \Rightarrow B$ neplatí v \mathcal{D} a píšeme $\mathcal{D} \not\models A \Rightarrow B$. Funkční závislost se nazývá **triviální**, pokud je pravdivá v každé \mathcal{D} .

pozorování: $A \Rightarrow B$ je triviální p. k. $B \subseteq A$

Příklad (Funkční závislosti, které jsou/nejsou pravdivé v \mathcal{D})

mějme relaci $\mathcal{D} = \{r_1, r_2, r_3, r_4\}$ nad schématem $R = \{\text{FOO}, \text{BAR}, \text{BAZ}, \text{QUX}\}$:

FOO	BAR	BAZ	QUX
10	22	a	222
10	33	b	333
10	22	a	444
20	33	a	555

$$r_1 = \{\langle \text{FOO}, 10 \rangle, \langle \text{BAR}, 22 \rangle, \langle \text{BAZ}, a \rangle, \langle \text{QUX}, 222 \rangle\}$$

$$r_2 = \{\langle \text{FOO}, 10 \rangle, \langle \text{BAR}, 33 \rangle, \langle \text{BAZ}, b \rangle, \langle \text{QUX}, 333 \rangle\}$$

$$r_3 = \{\langle \text{FOO}, 10 \rangle, \langle \text{BAR}, 22 \rangle, \langle \text{BAZ}, a \rangle, \langle \text{QUX}, 444 \rangle\}$$

$$r_4 = \{\langle \text{FOO}, 20 \rangle, \langle \text{BAR}, 33 \rangle, \langle \text{BAZ}, a \rangle, \langle \text{QUX}, 555 \rangle\}$$

$$\mathcal{D} \not\models \{\text{BAZ}\} \Rightarrow \{\text{BAR}\} \quad (\text{kvůli } r_1 \text{ a } r_4)$$

$$\mathcal{D} \not\models \{\text{FOO}, \text{BAZ}\} \Rightarrow \{\text{QUX}\} \quad (\text{kvůli } r_1 \text{ a } r_3)$$

$$\mathcal{D} \not\models \{\text{BAR}\} \Rightarrow \{\text{BAZ}\} \quad (\text{kvůli } r_2 \text{ a } r_4)$$

$$\mathcal{D} \models \{\text{FOO}, \text{BAR}\} \Rightarrow \{\text{BAZ}\}$$

$$\mathcal{D} \models \{\text{QUX}\} \Rightarrow S \quad (\text{pro jakékoliv } S \text{ je triviálně splněná})$$

$$\mathcal{D} \models \{\text{FOO}, \text{BAR}\} \Rightarrow \{\text{FOO}\}$$

Motivace pro normalizaci

motivace:

Na základě znalostí (funkčních) závislostí v datech je potřeba navrhnout relační schémata v databázi tak, aby se minimalizovala redundance dat a nedocházelo k patologickým situacím souvisejícím s modifikací dat.

příklad relace nad nevhodným schématem:

SCHOOL	DEAN	DEPT	HEAD	ID	COURSE	YEAR
SCI	Blangis	AF	Durcet	7	QOPT1	2012
SCI	Blangis	AF	Durcet	8	LASR1	2012
:	:	:	:	:	:	:
SCI	Blangis	CS	Curval	6	PAPR1	2012

problémy:

- 1 *redundance dat* (zbytečná duplikace hodnot)
- 2 *anomálie spojená s výmazem dat* (výmaz kurzů katedry „odstraní vedoucího“)
- 3 *anomálie spojená s aktualizací hodnot* (změna vedoucího katedry na víc místech)

Boyce-Coddova normální forma

zdroj anomálií v předchozím příkladě: některá množina atributů je funkčně závislá na jiné množině atributů, která není nadklíč; zavádíme proto:

Definice (Boyce-Coddova normální forma, BCNF)

Mějme relační schéma R a teorii Γ . Pak R je v BCNF vzhledem k Γ pokud pro každou netriviální $A \Rightarrow B \in \Gamma$ platí, že $\Gamma \models A \Rightarrow R$.

normalizace pomocí dekompozice: pokud není R v BCNF vzhledem k Γ , pak:

- 1 vezmeme netriviální $A \Rightarrow B \in \Gamma$ takovou, že $\Gamma \not\models A \Rightarrow R$
- 2 položíme $R_1 = A \cup B$ a $\Gamma_1 = \{C \cap R_1 \Rightarrow D \cap R_1 \mid C \cap R_1 \Rightarrow D \in \Gamma\}$
- 3 položíme $R_2 = A \cup (R \setminus B)$ a $\Gamma_2 = \{C \cap R_2 \Rightarrow D \cap R_2 \mid C \cap R_2 \Rightarrow D \in \Gamma\}$
- 4 proces se pokusíme opakovat pro dvojici R_1 a Γ_1 pokud R_1 není v BCNF vzhledem k Γ_1 a analogicky pro R_2 a Γ_2

poznámka: BCNF nemusí být dosažitelná, více kurs *Databázové systémy 2* (!!)

Příklad (Schéma, které není v BCNF)

Uvažujme relační schéma

$$R = \{\text{SCHOOL, DEAN, DEPT, HEAD, ID, COURSE, YEAR}\}$$

a teorii Γ popisující závislosti mezi atributy:

$$\begin{aligned}\Gamma = \{ & \{\text{DEPT}\} \Rightarrow \{\text{HEAD, SCHOOL, DEAN}\}, \\ & \{\text{SCHOOL}\} \Rightarrow \{\text{DEAN}\}, \\ & \{\text{COURSE, YEAR}\} \Rightarrow \{\text{ID}\}, \\ & \{\text{ID, YEAR}\} \Rightarrow \{\text{DEPT}\}\}.\end{aligned}$$

Schéma R není v BCNF vzhledem k Γ , protože (například):

- $\{\text{DEPT}\} \Rightarrow \{\text{HEAD, SCHOOL, DEAN}\} \in \Gamma$, ale $[\{\text{DEPT}\}]_{\Gamma} = \{\text{SCHOOL, DEAN, DEPT, HEAD}\} \neq R$, to jest $\Gamma \not\models \{\text{DEPT}\} \Rightarrow R$, nebo:
- $\{\text{SCHOOL}\} \Rightarrow \{\text{DEAN}\} \in \Gamma$, ale $[\{\text{SCHOOL}\}]_{\Gamma} = \{\text{SCHOOL, DEAN}\} \neq R$, to jest $\Gamma \not\models \{\text{SCHOOL}\} \Rightarrow R$.

Příklad (Normalizace schématu pomocí dekompozice)

$R = \{\text{SCHOOL, DEAN, DEPT, HEAD, ID, COURSE, YEAR}\}$

$\Gamma = \{\{\text{DEPT}\} \Rightarrow \{\text{HEAD, SCHOOL, DEAN}\}, \dots\}$ (viz předchozí příklad)

- $R_1 = \{\text{SCHOOL, DEAN, DEPT, HEAD}\}$

$\Gamma_1 = \{\{\text{DEPT}\} \Rightarrow \{\text{HEAD, SCHOOL, DEAN}\}, \{\text{SCHOOL}\} \Rightarrow \{\text{DEAN}\}\}$

- $R_{11} = \{\text{SCHOOL, DEAN}\}$

$\Gamma_{11} = \{\{\text{SCHOOL}\} \Rightarrow \{\text{DEAN}\}\}$

- $R_{12} = \{\text{SCHOOL, DEPT, HEAD}\}$

$\Gamma_{12} = \{\{\text{DEPT}\} \Rightarrow \{\text{HEAD, SCHOOL}\}, \{\text{SCHOOL}\} \Rightarrow \{\}\}$

- $R_2 = \{\text{DEPT, ID, COURSE, YEAR}\}$

$\Gamma_2 = \{\{\text{DEPT}\} \Rightarrow \{\}, \{\text{COURSE, YEAR}\} \Rightarrow \{\text{ID}\}, \{\text{ID, YEAR}\} \Rightarrow \{\text{DEPT}\}\}$

- $R_{21} = \{\text{DEPT, ID, YEAR}\}$

$\Gamma_{21} = \{\{\text{DEPT}\} \Rightarrow \{\}, \{\text{ID, YEAR}\} \Rightarrow \{\text{DEPT}\}\}$

- $R_{22} = \{\text{ID, COURSE, YEAR}\}$

$\Gamma_{22} = \{\{\text{COURSE, YEAR}\} \Rightarrow \{\text{ID}\}, \{\text{ID, YEAR}\} \Rightarrow \{\}\}$

Příklad (Reprezentace výchozích dat v normalizované databázi)

SCHOOL	DEAN	DEPT	HEAD	ID	COURSE	YEAR
SCI	Blangis	AF	Durcet	7	QOPT1	2012
SCI	Blangis	AF	Durcet	8	LASR1	2012
SCI	Blangis	AF	Durcet	8	LASR1	2013
SCI	Blangis	CS	Curval	3	ALMA1	2012
SCI	Blangis	CS	Curval	3	ALMA1	2013
SCI	Blangis	CS	Curval	6	DATA1	2012
SCI	Blangis	CS	Curval	6	DATA1	2013
SCI	Blangis	CS	Curval	6	PAPR1	2012

=

SCHOOL	DEAN
SCI	Blangis

⋈

SCHOOL	DEPT	HEAD
SCI	AF	Durcet
SCI	CS	Curval

⋈

DEPT	ID	YEAR
AF	7	2012
AF	8	2012
AF	8	2013
CS	3	2012
CS	3	2013
CS	6	2012
CS	6	2013

⋈




ID	COURSE	YEAR
3	ALMA1	2012
3	ALMA1	2013
6	DATA1	2012
6	DATA1	2013
6	PAPR1	2012
7	QOPT1	2012
8	LASR1	2012
8	LASR1	2013

Přednáška 8: Závěr

pojmy k zapamatování:

- anomálie, Boyce-Coddova normální forma, normalizace pomocí dekompozice

použité zdroje:

-  Date C. J.: *Database in Depth: Relational Theory for Practitioners*
O'Reilly Media 2005, ISBN 978-0596100124
-  Maier D: *Theory of Relational Databases*
Computer Science Press 1983, ISBN 978-0914894421
-  Simovici D.: Tenney R.: *Relational Database Systems*
Academic Press 1995, ISBN 978-0126443752