Databázové systémy

Úvod do funkčních závislostí a normalizace

Vilém Vychodil

KMI/DATA1, Přednáška 8

Databázové systémy

Přednáška 8: Přehled

1. Funkční závislosti:
   - pravdivost funkčních závislostí v systémech relací,
   - teorie, modely,
   - sémantické vyplývání,
   - kanonické relace, kanonické modely,
   - testování sémantického vyplývání,
   - sémantický uzávěr množiny atributů,
   - algoritmus pro výpočet uzávěru,
   - charakterizace vyplývání pomocí sémantického uzávěru,
   - vztah k pojmu klíč.
2. Boyce-Coddova normální forma:
   - formulace normální formy,
   - dekompozice na základě funkčních závislostí,
   - normalizace pomocí dekompozice.

Opakování: Funkční závislosti (Přednáška 7)

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 závloslost 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 $R = \{\texttt{FOO}, \texttt{BAR}, \texttt{BAZ}, \texttt{QUX}\}$:

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

$r_1 = \{\langle \texttt{FOO}, 10\rangle, \langle \texttt{BAR}, 22\rangle, \langle \texttt{BAZ}, \mathsf{a}\rangle, \langle \texttt{QUX}, 222\rangle\}$
$r_2 = \{\langle \texttt{FOO}, 10\rangle, \langle \texttt{BAR}, 33\rangle, \langle \texttt{BAZ}, \mathsf{b}\rangle, \langle \texttt{QUX}, 333\rangle\}$
$r_3 = \{\langle \texttt{FOO}, 10\rangle, \langle \texttt{BAR}, 22\rangle, \langle \texttt{BAZ}, \mathsf{a}\rangle, \langle \texttt{QUX}, 444\rangle\}$
$r_4 = \{\langle \texttt{FOO}, 20\rangle, \langle \texttt{BAR}, 33\rangle, \langle \texttt{BAZ}, \mathsf{a}\rangle, \langle \texttt{QUX}, 555\rangle\}$

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

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

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

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

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

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

Věta (O pravdivosti funkčních závislostí)

Pro každé $A, B, C \subseteq R$ a libovolné relace $\mathcal{D}, \mathcal{D}_1, \mathcal{D}_2$ nad R platí:

1. *pokud $\mathcal{D} \models A \Rightarrow B$, pak $\mathcal{D} \models A \Rightarrow B \cap C$;*
2. *pokud $\mathcal{D} \models A \Rightarrow B$, pak $\mathcal{D} \models A \cup C \Rightarrow B$;*
3. *$\mathcal{D} \models A \Rightarrow B$ právě tehdy, když $\mathcal{D} \models A \Rightarrow B \setminus A$;*
4. *pokud $\mathcal{D}_1 \subseteq \mathcal{D}_2$ a $\mathcal{D}_2 \models A \Rightarrow B$, pak $\mathcal{D}_1 \models A \Rightarrow B$.*

Důkaz.

Předpokládejme, že $\mathcal{D} \models A \Rightarrow B$ a vezměme libovolné $r_1, r_2 \in \mathcal{D}$ takové, že $r_1(A) = r_2(A)$. Dle předpokladu pak $r_1(B) = r_2(B)$ a tím spíš $r_1(B \cap C) = r_2(B \cap C)$, protože $B \cap C \subseteq B$, to dokazuje 1 a jednu stranu tvrzení 3. Analogicky 2. Pokud $\mathcal{D} \models A \Rightarrow B \setminus A$ a platí $r_1(A) = r_2(A)$, pak z $r_1(B \setminus A) = r_2(B \setminus A)$ a $r_1(A) = r_2(A)$ dostaneme $r_1(A \cup (B \setminus A)) = r_2(A \cup (B \setminus A))$, to jest platí 3, protože $B \subseteq A \cup (B \setminus A)$. Bod 4 je přímým důsledkem toho, že pro každé $r_1, r_2 \in \mathcal{D}_1$ platí $r_1, r_2 \in \mathcal{D}_2$. □

Věta (Dekompozice na základě funkčních závislostí)

Mějme $\mathcal{D}$ na R a $A, B \subseteq R$. Pokud $\mathcal{D} \models A \Rightarrow B$, pak $\mathcal{D}$ má bezeztrátovou dekompozici vzhledem k $A \cup B$ a $A \cup (R \setminus B)$.

Důkaz.

Máme ukázat, že $\pi_{A\cup B}(\mathcal{D}) \bowtie \pi_{A\cup(R\setminus B)}(\mathcal{D}) \subseteq \mathcal{D}$ (obrácená inkluze platí vždy). Vezměme n-tice $r_1 \in \pi_{A\cup B}(\mathcal{D})$ a $r_2 \in \pi_{A\cup(R\setminus B)}(\mathcal{D})$, které jsou spojitelné. Jelikož

$$(A \cup B) \cap (A \cup (R \setminus B)) = A \cup (B \cap (R \setminus B)) = A \cup \emptyset = A,$$

platí, že $r_1(A) = r_2(A)$. Vezměme $s \in \mathcal{D}$ takovou, že $r_2 = s(A \cup (R \setminus B))$. Jelikož je B funkčně závisá na A, pak zřejmě z $r_1(A) = r_2(A) = s(A)$ dostáváme $r_1(B) = s(B)$, to jest $r_1 = s(A \cup B)$. To znamená, že $r_1 r_2 = s \in \mathcal{D}$. $\square$

triviální použití:

- pokud je $A \Rightarrow B$ triviální, pak je $B \subseteq A$ a potom se $\mathcal{D}$ na R dekomponuje na relace na schématech $A \cup B = A$ a $A \cup (R \setminus B) \supseteq A \cup (R \setminus A) = R$ (nezajímavé)

Příklad (Dekompozice použitím $\{\text{DEPT}\} \Rightarrow \{\text{HEAD}, \text{SCHOOL}, \text{DEAN}\}$)

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	DEPT	HEAD
SCI	Blangis	AF	Durcet
SCI	Blangis	CS	Curval

$\bowtie$

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

Pravdivost závislostí v systémech relací

Definice (pravdivost funkční závislosti v datech)

Nechť R je relační schéma a $\mathcal{M}$ je množina relací nad schématem R. Položíme $\mathcal{M} \models A \Rightarrow B$ pokud $\mathcal{D} \models A \Rightarrow B$ pro každou $\mathcal{D} \in \mathcal{M}$.

Věta (pravdivost funkčních závislostí v systému dvouprvkových relací)

Pro každou $\mathcal{D}$ nad R existuje konečný systém $\mathcal{M}$ nejvýš dvouprvkových relací nad R tak, že pro každé $A, B \subseteq R$ platí: $\mathcal{D} \models A \Rightarrow B$ p. k. $\mathcal{M} \models A \Rightarrow B$.

Důkaz.

Položme $\mathcal{M} = \big\{\{r_1, r_2\} \,|\, r_1, r_2 \in \mathcal{D}\big\}$, to jest, $\mathcal{M}$ je systém dvouprvkových tabulek složených ze všech n-tic z relace $\mathcal{D}$. Zřejmě platí, že $\mathcal{D} \models A \Rightarrow B$ p. k. pro libovolné $r_1, r_2 \in \mathcal{D}$ takové, že $r_1(A) = r_2(A)$, platí $r_1(B) = r_2(B)$, což je p. k. pro libovolné $r_1, r_2 \in \mathcal{D}$ platí, že $\{r_1, r_2\} \models A \Rightarrow B$ p. k. $\mathcal{M} \models A \Rightarrow B$. □

Příklad (Dvouprvkové relace odpovídající $\mathcal{D}$)

FOO	BAR	BAZ	QUX
10	22	a	222
20	33	a	555

FOO	BAR	BAZ	QUX
10	22	a	222
10	22	a	444

FOO	BAR	BAZ	QUX
10	22	a	444
10	33	b	333

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

FOO	BAR	BAZ	QUX
10	22	a	222
10	33	b	333

FOO	BAR	BAZ	QUX
10	22	a	444
20	33	a	555

FOO	BAR	BAZ	QUX
10	33	b	333
20	33	a	555

Teorie a modely

motivace:

Chceme se zabývat tím, které funkční závislosti vyplývají z jiných závislostí. Primárně se zajímáme o sémantické vyplývání, pro jeho zavedení potřebujeme pojmy teorie a model.

Definice (teorie, angl.: *theory*)

Množinu funkčních závislostí (nad schématem R) nazveme **teorie** (nad R). Pokud je Γ teorie a $A \Rightarrow B \in \Gamma$, pak říkáme, že $A \Rightarrow B$ je **předpokladem** z Γ.

Definice (model, angl.: *model*)

Mějme teorii Γ. Relace $\mathcal{D}$ je **model** Γ pokud pro každou $A \Rightarrow B \in \Gamma$ platí, že $\mathcal{D} \models A \Rightarrow B$. Množinu všech modelů Γ označujeme $\mathrm{Mod}(\Gamma)$, to jest:

$$\mathrm{Mod}(\Gamma) = \{\mathcal{D} \,|\, \text{pro každou } A \Rightarrow B \in \Gamma \text{ platí } \mathcal{D} \models A \Rightarrow B\}.$$

Příklad (Teorie a modely)

uvažujme následující teorii:

$$\begin{aligned}\Gamma = \{&\{\mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAR}, \mathtt{BAZ}\},\\ &\{\mathtt{BAR}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{FOO}\},\\ &\{\mathtt{FOO}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{BAR}\},\\ &\{\mathtt{FOO}, \mathtt{BAR}\} \Rightarrow \{\mathtt{BAZ}\}\}\end{aligned}$$

potom pro danou Γ například:

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

je model

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

není model

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

není model

Sémantické vyplývání

Definice (sémantické vyplývání, angl.: *semantic entailment*)

Funkční závislost $A \Rightarrow B$ **sémanticky plyne** z teorie Γ pokud je $A \Rightarrow B$ pravdivá v každém modelu Γ, to znamená pokud $\mathrm{Mod}(\Gamma) \models A \Rightarrow B$. Fakt, že $A \Rightarrow B$ sémanticky plyne z Γ značíme $\Gamma \models A \Rightarrow B$.

slovně:

Funkční závislost $A \Rightarrow B$ plyne z teorie Γ pokud je $A \Rightarrow B$ pravdivá v každé relaci, ve které jsou pravdivé všechny formule z teorie Γ.

speciální případ:

- pro $\Gamma = \emptyset$ píšeme $\models A \Rightarrow B$ místo $\emptyset \models A \Rightarrow B$
- význam: $\models A \Rightarrow B$ p. k. $A \Rightarrow B$ je pravdivá v každé relaci
- z předchozího víme: $\models A \Rightarrow B$ p. k. $A \Rightarrow B$ je triviální, to jest $B \subseteq A$

Věta (Vlastnosti teorií, modelů a sémantického vyplývání)

Pro libovolné teorie $\Gamma, \Gamma_1, \Gamma_2$ *nad* R *platí:*

1. *pokud* $\Gamma_1 \subseteq \Gamma_2$, *pak* $\mathrm{Mod}(\Gamma_2) \subseteq \mathrm{Mod}(\Gamma_1)$;
2. $\mathrm{Mod}(\emptyset)$ *je množina všech relací nad* R;
3. $\mathrm{Mod}(\Gamma) \neq \emptyset$ *pro každou teorii* Γ;
4. *pokud* $\Gamma_1 \subseteq \Gamma_2$ *a* $\Gamma_1 \models A \Rightarrow B$, *pak* $\Gamma_2 \models A \Rightarrow B$;
5. $\Gamma \models A \Rightarrow B$ *pro každou* $A \Rightarrow B \in \Gamma$.

Důkaz.

Pokud $\mathcal{D} \in \mathrm{Mod}(\Gamma_2)$, pak pro každou $A \Rightarrow B \in \Gamma_2$ platí, že $\mathcal{D} \models A \Rightarrow B$. Jelikož ale $\Gamma_1 \subseteq \Gamma_2$, tím spíš platí $\mathcal{D} \models A \Rightarrow B$ pro každou $A \Rightarrow B \in \Gamma_1$, což ukazuje 1; 2 plyne přímo z definice; 3 platí, protože $\mathcal{D} \in \mathrm{Mod}(\Gamma)$ pro každou $\mathcal{D}$ splňující $|\mathcal{D}| < 2$; 4 je důsledkem 1; 5 platí triviálně, protože $\mathcal{D} \models A \Rightarrow B$ pro každé $A \Rightarrow B \in \Gamma$ a $\mathcal{D} \in \mathrm{Mod}(\Gamma)$. □

Příklad (Sémantické důsledky teorie)

teorie z předchozího příkladu:

$$\Gamma = \{\{\mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAR}, \mathtt{BAZ}\}, \{\mathtt{BAR}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{FOO}\},$$
$$\{\mathtt{FOO}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{BAR}\}, \{\mathtt{FOO}, \mathtt{BAR}\} \Rightarrow \{\mathtt{BAZ}\}\}$$

netriviální funkční závislosti $A \Rightarrow B$, kde $A \cap B = \emptyset$ a $\Gamma \models A \Rightarrow B$:

$\{\mathtt{BAR}, \mathtt{BAZ}, \mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}\}$, $\{\mathtt{BAR}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{FOO}\}$, $\{\mathtt{BAR}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAZ}\}$,
$\{\mathtt{BAR}, \mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAZ}\}$, $\{\mathtt{BAR}, \mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}\}$, $\{\mathtt{BAZ}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAR}\}$,
$\{\mathtt{BAZ}, \mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAR}\}$, $\{\mathtt{BAZ}, \mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}\}$, $\{\mathtt{FOO}, \mathtt{BAR}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAZ}\}$,
$\{\mathtt{FOO}, \mathtt{BAR}\} \Rightarrow \{\mathtt{BAZ}\}$, $\{\mathtt{FOO}, \mathtt{BAZ}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAR}\}$, $\{\mathtt{FOO}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{BAR}\}$,
$\{\mathtt{FOO}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAR}, \mathtt{BAZ}\}$, $\{\mathtt{FOO}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAR}\}$, $\{\mathtt{FOO}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAZ}\}$,
$\{\mathtt{QUX}\} \Rightarrow \{\mathtt{BAR}, \mathtt{BAZ}\}$, $\{\mathtt{QUX}\} \Rightarrow \{\mathtt{BAR}\}$, $\{\mathtt{QUX}\} \Rightarrow \{\mathtt{BAZ}\}$,
$\{\mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAR}, \mathtt{BAZ}\}$, $\{\mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAR}\}$, $\{\mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAZ}\}$,
$\{\mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}\}$

Kanonické relace a modely

Definice (kanonická relace, angl.: *canonical relation*)

Pro libovolnou $M \subseteq R$ definujeme relaci $\mathcal{D}_M$ nad schématem R tak, že $\mathcal{D}_M = \{r_1, r_2\}$, přitom $r_1(y) = p$ pro každý $y \in R$ a

$$r_2(y) = \begin{cases} p, & \text{pokud } y \in M, \\ q, & \text{pokud } y \notin M, \end{cases}$$

přitom p, q označují dva různé (fixní) prvky domén atributů z R. Takto zavedenou relaci $\mathcal{D}_M$ nazveme **kanonická relace** nad schématem R.

Definice (kanonický model, angl.: *canonical model*)

Mějme teorii Γ. Kanonická relace $\mathcal{D}_M$, která je modelem Γ, se nazývá **kanonický model** Γ. Množinu všech kanonických modelů Γ značíme $\mathrm{Mod}_\mathrm{C}(\Gamma)$, to jest:

$$\mathrm{Mod}_\mathrm{C}(\Gamma) = \{\mathcal{D}_M \,|\, \mathcal{D}_M \in \mathrm{Mod}(\Gamma)\}.$$

Věta (Pravdivost funkčních závislostí v kanonických relacích)

Pro libovolné $A, B, M \subseteq R$ jsou následující tvrzení ekvivalentní:

1. $\mathcal{D}_M \models A \Rightarrow B$,
2. *pokud* $A \subseteq M$, *pak* $B \subseteq M$.

Důkaz.

Nejprve ukážeme, že pro $\mathcal{D}_M = \{r_1, r_2\}$ a libovolnou $C \subseteq R$ platí, že $r_1(C) = r_2(C)$ p. k. $C \subseteq M$. Dle definice kanonické relace máme, že $r_1(C) = r_2(C)$ znamená $r_2(y) = p$ pro každý $y \in C$. To znamená, $y \in M$ pro každý $y \in C$, to jest $C \subseteq M$.

S použitím tohoto faktu, $\mathcal{D}_M \models A \Rightarrow B$ právě tehdy, když $r_1(A) = r_2(A)$ implikuje $r_1(B)r_2(B)$, to platí právě tehdy, když $A \subseteq M$ implikuje $B \subseteq M$. To jest body 1 a 2 jsou ekvivalentní. □

poznámka:

- pozorování slouží ke zjednodušenému vyjádření $\mathcal{D}_M \models A \Rightarrow B$

Věta (Charakterizace $\models$ pomocí kanonických modelů)

$\Gamma \models A \Rightarrow B$ *právě tehdy, když* $\mathrm{Mod}_{\mathrm{C}}(\Gamma) \models A \Rightarrow B$.

Důkaz.

Pokud $\Gamma \models A \Rightarrow B$, to jest dle definice $\mathrm{Mod}(\Gamma) \models A \Rightarrow B$ a tím spíš $\mathrm{Mod}_{\mathrm{C}}(\Gamma) \models A \Rightarrow B$, protože $\mathrm{Mod}_{\mathrm{C}}(\Gamma) \subseteq \mathrm{Mod}(\Gamma)$.

Obráceně, předpokládejme, že $\mathrm{Mod}_{\mathrm{C}}(\Gamma) \models A \Rightarrow B$ a vezměme libovolný $\mathcal{D} \in \mathrm{Mod}(\Gamma)$. To znamená, že $\mathcal{D} \models C \Rightarrow D$ pro každou $C \Rightarrow D \in \Gamma$. Dle jednoho z předchozích tvrzení existuje množina dvouprvkových relací $\mathcal{M}$ tak, že $\mathcal{M} \models C \Rightarrow D$ pro každou $C \Rightarrow D \in \Gamma$. Ke každé $\mathcal{D}' \in \mathcal{M}$ můžeme vzít množinu atributů, na kterých jsou si obě n-nice z $\mathcal{D}'$ rovny, to jest

$$M = \{y \in R \,|\, r_1(y) = r_2(y)\},$$

kde $\mathcal{D}' = \{r_1, r_2\}$. Potom $\mathcal{D}_M$ je kanonický model Γ. Jelikož $\mathrm{Mod}_{\mathrm{C}}(\Gamma) \models A \Rightarrow B$, pak i $\mathcal{D}_M \models A \Rightarrow B$. Protože jsme vzali $\mathcal{D}' \in \mathcal{M}$ libovolně, máme $\mathcal{D} \models A \Rightarrow B$. $\square$

Příklad (Kanonické relace odpovídající $\mathcal{D}$)

FOO	BAR	BAZ	QUX
10	22	a	222
20	33	a	555

FOO	BAR	BAZ	QUX
10	22	a	222
10	22	a	444

FOO	BAR	BAZ	QUX
10	22	a	444
10	33	b	333

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

FOO	BAR	BAZ	QUX
10	22	a	222
10	33	b	333

FOO	BAR	BAZ	QUX
10	22	a	444
20	33	a	555

FOO	BAR	BAZ	QUX
10	33	b	333
20	33	a	555

Příklad (Kanonické relace odpovídající $\mathcal{D}$)

FOO	BAR	BAZ	QUX
0	0	1	0
1	1	1	1

FOO	BAR	BAZ	QUX
1	1	1	0
1	1	1	1

FOO	BAR	BAZ	QUX
1	0	0	0
1	1	1	1

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

FOO	BAR	BAZ	QUX
1	0	0	0
1	1	1	1

FOO	BAR	BAZ	QUX
0	0	1	0
1	1	1	1

FOO	BAR	BAZ	QUX
0	1	0	0
1	1	1	1

Vztah k sémantickému vyplývání z výrokové logiky

motivace:

Funkční závislosti lze chápat jako výrokové formule. V jakém vztahu je tedy výrokové sémantické vyplývání a $\models$ *tak, jak jsme jej zavedli my?*

funkční závislosti jako výrokové formule:

$$\{y_1, \ldots, y_m\} \Rightarrow \{z_1, \ldots, z_n\} \text{ vs. } (y_1 \curlywedge \cdots \curlywedge y_m) \Rightarrow (z_1 \curlywedge \cdots \curlywedge z_n)$$

Pro $A \Rightarrow B$, Γ a $\mathcal{D}_M$ lze proto brát jejich výrokové protějšky:

- odpovídající výrokovou formuli $\varphi_{A,B}$;
- odpovídající výrokovou teorii $\Gamma_{\mathrm{V}} = \{\varphi_{A,B} \,|\, A \Rightarrow B \in \Gamma\}$;
- odpovídající ohodnocení výrokových proměnných: $e_M(y) = 1$ p. k. $y \in M$;

přitom jako důsledek předchozí věty máme:

$$\Gamma \models A \Rightarrow B \text{ p. k. } e(\varphi_{A,B}) = 1 \text{ pro každý model } e \text{ teorie } \Gamma_{\mathrm{V}}.$$

důsledek: vyplývání funkčních závislostí lze ověřovat *tabulkovou metodou*

Příklad (Test sémantického vyplývání pomocí tabelace)

$M \subseteq R$	φ_1	φ_2	φ_3	φ_4	ψ_1	ψ_2	ψ_3
$\{\}$	1	1	1	1	1	1	1
$\{\mathtt{QUX}\}$	0	1	1	1	1	1	1
$\{\mathtt{BAZ}\}$	1	1	1	1	1	0	1
$\{\mathtt{BAZ}, \mathtt{QUX}\}$	0	1	1	1	1	0	0
$\{\mathtt{BAR}\}$	1	1	1	1	1	1	1
$\{\mathtt{BAR}, \mathtt{QUX}\}$	0	1	1	1	1	1	1
$\{\mathtt{BAR}, \mathtt{BAZ}\}$	1	0	1	1	1	0	1
$\{\mathtt{BAR}, \mathtt{BAZ}, \mathtt{QUX}\}$	0	0	1	1	1	0	1
$\{\mathtt{FOO}\}$	1	1	1	1	1	1	1
$\{\mathtt{FOO}, \mathtt{QUX}\}$	0	1	1	1	1	1	1
$\{\mathtt{FOO}, \mathtt{BAZ}\}$	1	1	0	1	0	0	1
$\{\mathtt{FOO}, \mathtt{BAZ}, \mathtt{QUX}\}$	0	1	0	1	1	0	0
$\{\mathtt{FOO}, \mathtt{BAR}\}$	1	1	1	0	1	1	1
$\{\mathtt{FOO}, \mathtt{BAR}, \mathtt{QUX}\}$	0	1	1	0	1	1	1
$\{\mathtt{FOO}, \mathtt{BAR}, \mathtt{BAZ}\}$	1	1	1	1	0	1	1
$\{\mathtt{FOO}, \mathtt{BAR}, \mathtt{BAZ}, \mathtt{QUX}\}$	1	1	1	1	1	1	1

teorie:

$\varphi_1 : \{\mathtt{QUX}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAR}, \mathtt{BAZ}\}$

$\varphi_2 : \{\mathtt{BAR}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{FOO}\}$,

$\varphi_3 : \{\mathtt{FOO}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{BAR}\}$,

$\varphi_4 : \{\mathtt{FOO}, \mathtt{BAR}\} \Rightarrow \{\mathtt{BAZ}\}$

testované formule:

$\psi_1 : \{\mathtt{FOO}, \mathtt{BAZ}\} \Rightarrow \{\mathtt{QUX}\}$

$\psi_2 : \{\mathtt{BAZ}\} \Rightarrow \{\mathtt{FOO}, \mathtt{BAR}\}$

$\psi_3 : \{\mathtt{BAZ}, \mathtt{QUX}\} \Rightarrow \{\mathtt{BAR}\}$

algoritmus pro výpočet $[M]_\Gamma$

Data: $M \subseteq R$ a teorie Γ nad R
Result: množina $[M]_\Gamma \subseteq R$ (**sémantický uzávěr M vzhledem k Γ**)

```
W := M; /* W je pomocná proměnná */
repeat
    L := W; /* L označuje poslední vypočtnou hodnotu W */
    foreach E ⇒ F ∈ Γ do
        if E ⊆ W then
            W := W ∪ F; /* aktualizuj W */
            Γ := Γ \ {E ⇒ F}; /* E ⇒ F už není potřeba v Γ */
        end
    end
until L = W;
return W;
```

proč zavádíme: pomocí $[M]_T$ ukážeme efektivní test sémantického vyplývání

Příklad (Průběh výpočtu sémantického uzávěru)

W :	Γ :
$\{\mathtt{A}\}$	$\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}$, $\{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}$, $\{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}$, $\{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}$, $\{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C}\}$	

W :	Γ :
$\{\mathtt{A},\mathtt{E}\}$	$\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}$, $\{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}$, $\{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}$, $\{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}$, $\{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E}\}$	$\{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}$, $\{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}$, $\{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}$, $\{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{G}\}$	

W :	Γ :
$\{\mathtt{E},\mathtt{F}\}$	$\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}$, $\{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}$, $\{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}$, $\{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}$, $\{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{E},\mathtt{F}\}$	$\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}$, $\{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}$, $\{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}$, $\{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{F}\}$	$\{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}$, $\{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}$, $\{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{F},\mathtt{G}\}$	

odtud: $[\{\mathtt{A}\}]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C}\}$, $[\{\mathtt{A},\mathtt{E}\}]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{G}\}$, $[\{\mathtt{E},\mathtt{F}\}]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{F},\mathtt{G}\}$

Věta (Základní vlastnost sémantického uzávěru I)

Pro libovolnou Γ *nad* R *a* $A \subseteq R$ *platí, že* $\Gamma \models A \Rightarrow [A]_\Gamma$.

Důkaz.

Nechť $\mathcal{D}$ je libovolný model $\mathcal{D} \in \mathrm{Mod}(\Gamma)$. Vezměme libovolné $r_1, r_2 \in \mathcal{D}$ takové, že $r_1(A) = r_2(A)$. Stačí ukázat, že $r_1([A]_\Gamma) = r_2([A]_\Gamma)$. Toto tvrzení prokážeme indukcí přes počet kroků algoritmu pro výpočet $[A]_\Gamma$.

Pro počáteční hodnotu $W = A$ je tvrzení zřejmé, protože $A \Rightarrow A$ je triviální funkční závislost. Předpokládejme, že pro W platí $r_1(W) = r_2(W)$ a uvažujme $E \Rightarrow F \in \Gamma$ takovou, že $E \subseteq W$. Pak stačí ukázat, že $r_1(W \cup F) = r_2(W \cup F)$. Z faktů, že $\mathcal{D} \in \mathrm{Mod}(\Gamma)$, $r_1, r_2 \in \mathcal{D}$, $r_1(W) = r_2(W)$, $E \subseteq W$ a $E \Rightarrow F \in \Gamma$ ihned dostáváme, že $r_1(F) = r_2(F)$ což dohromady s naším předpokladem $r_1(W) = r_2(W)$ dává, že $r_1(W \cup F) = r_2(W \cup F)$.

Odtud plyne, že pro každou průběžnou hodnotu W v algoritmu je $\mathcal{D} \models A \Rightarrow W$, tedy $\Gamma \models A \Rightarrow W$. Zbyek plyne z toho, že $[A]_\Gamma$ je poslední hodnotou W. $\square$

Věta (Základní vlastnost sémantického uzávěru II)

Pokud $\Gamma \models A \Rightarrow B$, *pak* $B \subseteq [A]_\Gamma$.

Důkaz.

Tvrzení prokážeme obměnou: Za předpokladu, že $B \not\subseteq [A]_\Gamma$ ukážeme, že $\Gamma \not\models A \Rightarrow B$. K tomu stačí ukázat, že za předpokladu $B \not\subseteq [A]_\Gamma$ existuje model $\mathcal{D} \in \mathrm{Mod}(\Gamma)$ tak, že $\mathcal{D} \not\models A \Rightarrow B$.

Předpokládejme tedy, že $B \not\subseteq [A]_\Gamma$. Hledaný model budeme uvažovat jako kanonickou relaci. Konkrétně vezmeme $\mathcal{D}_{[A]_\Gamma}$. Z věty o platnosti $A \Rightarrow B$ v kanonické relaci dostáváme, že $\mathcal{D}_{[A]_\Gamma} \not\models A \Rightarrow B$, protože $A \subseteq [A]_\Gamma$, ale $B \not\subseteq [A]_\Gamma$ (předpoklad). Stačí tedy ukázat, že $\mathcal{D}_{[A]_\Gamma} \in \mathrm{Mod}(\Gamma)$.

Vezměme libovolnou $E \Rightarrow F \in \Gamma$. Z algoritmu pro výpočet $[A]_\Gamma$ je patrné, že pokud $E \subseteq [A]_\Gamma$, pak musí platit i $F \subseteq [A]_\Gamma$, protože $[A]_\Gamma$ je poslední průběžnou hodnotou W z algoritmu. To jest, $E \subseteq [A]_\Gamma$ implikuje $F \subseteq [A]_\Gamma$, což dává $\mathcal{D}_{[A]_\Gamma} \models E \Rightarrow F$ a tedy $\mathcal{D}_{[A]_\Gamma} \in \mathrm{Mod}(\Gamma)$, protože $E \Rightarrow F \in \Gamma$ jsme zvolili libovolně. $\square$

Věta (Charakterizace sémantického vyplývání)

Pro každou teorii Γ *a* $A, B \subseteq R$ *jsou následující tvrzení ekvivalentní:*

1. $\Gamma \models A \Rightarrow B$,
2. $B \subseteq [A]_\Gamma$,
3. $\mathcal{D}_{[A]_\Gamma} \models A \Rightarrow B$.

Důkaz.

Z bodu 1 plyne bod 2 díky předchozí větě. Předpokládejme, že platí 2, pak ihned dostáváme 3 z věty o platnosti funkčních závislostí v kanonických relacích. Předpokládejme, že platí 3. Pak z $A \subseteq [A]_\Gamma$ ihned dostáváme, že $B \subseteq [A]_\Gamma$. Jelikož ale $\Gamma \models A \Rightarrow [A]_\Gamma$ (viz základní vlastnost sémantického uzávěru), tím spíš i $\Gamma \models A \Rightarrow B$, protože $B \subseteq [A]_\Gamma$. □

důsledky:

- $[A]_\Gamma$ je největší prvek množiny $\{B \subseteq R \mid \Gamma \models A \Rightarrow B\}$
- podmínka 2 je *efektivní test sémantického vyplývání*

Příklad (Testování sémantického vyplývání pomocí uzávěrů)

Pro $\Gamma = \{\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}, \{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}, \{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}, \{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}, \{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}\}$ z předchozího příkladu a množinu atributů $M = \{\mathtt{A},\mathtt{E}\}$ máme:

W:	Γ:
$\{\mathtt{A},\mathtt{E}\}$	$\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}, \{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}, \{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}, \{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}, \{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E}\}$	$\{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}, \{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}, \{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}, \{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{G}\}$	

To jest $[M]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{G}\}$ a platí, že $\Gamma \models M \Rightarrow N$ p. k. $N \subseteq \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{G}\}$.

Například tedy platí:

$$\Gamma \models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{G}\} \qquad \Gamma \models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{C}\} \qquad \Gamma \models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{C},\mathtt{G}\}$$

$$\Gamma \models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{B},\mathtt{G}\} \qquad \Gamma \models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{B},\mathtt{C}\} \qquad \Gamma \models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{B},\mathtt{C},\mathtt{G}\}$$

Na druhou stranu například:

$$\Gamma \not\models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{D}\} \qquad \Gamma \not\models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{H}\} \qquad \Gamma \not\models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{D},\mathtt{E}\}$$

$$\Gamma \not\models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{F},\mathtt{G}\} \qquad \Gamma \not\models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{C},\mathtt{D}\} \qquad \Gamma \not\models \{\mathtt{A},\mathtt{E}\} \Rightarrow \{\mathtt{C},\mathtt{G},\mathtt{H}\}$$

Klíče z pohledu funkčních závislostí

Definice (**nadklíč**, **klíč**, angl.: *superkey, key*)

Mějme relační schéma R a teorii Γ nad R. Pak **nadklíč** schématu R vzhledem k Γ je libovolná $K \subseteq R$ taková, že $\Gamma \models K \Rightarrow R$. Pokud je K nadklíč R vzhledem k Γ a žádná $K' \subset K$ není nadklíč R vzhledem k Γ, pak je K **klíč** R vzhledem k Γ.

otázka:

Jak souvisí s pojmem klíče relační proměnné (PŘEDNÁŠKA 2)?

následující je ekvivalentní:

1. $K_1, \ldots, K_n$ je množina klíčů (relační proměnné) typu R;
   to jest $\{K_1, \ldots, K_n\} \neq \emptyset$ a $K_i \not\subseteq K_j$ pro každé $i \neq j$ (PŘEDNÁŠKA 2) p. k.
2. každý K_i je klíč R vzhledem k $\Gamma = \{K_1 \Rightarrow R, \ldots, K_n \Rightarrow R\}$.

pozorování: každý nadklíč (R vzhledem k Γ) obsahuje nějaký klíč (R vzhledem k Γ)

Příklad (Nadklíče a klíče)

Pro $\Gamma = \{\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}, \{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}, \{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}, \{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}, \{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}\}$ z předchozích příkladů jsou klíče schématu $R = \{\mathtt{A},\dots,\mathtt{H}\}$ vzhledem k Γ následující:

$$K_1 = \{\mathtt{A},\mathtt{D}\}, \quad K_2 = \{\mathtt{B},\mathtt{D}\}, \quad K_3 = \{\mathtt{C},\mathtt{D},\mathtt{E}\}, \quad K_4 = \{\mathtt{D},\mathtt{F}\}, \quad K_5 = \{\mathtt{D},\mathtt{G}\},$$

protože pro K_1 platí:

$$[\{\mathtt{A},\mathtt{D}\}]_\Gamma = R, \qquad [\{\mathtt{A}\}]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C}\} \neq R, \qquad [\{\mathtt{D}\}]_\Gamma = \{\mathtt{D}\} \neq R,$$

to znamená: $\Gamma \models K_1 \Rightarrow R$ a $\Gamma \not\models M \Rightarrow R$ pro každou $M \subset K_1$, to jest K_1 je klíč. Analogicky se dá ukázat pro $K_2,\dots,K_5$. Například pro K_3 je $[K_3]_\Gamma = R$ a

$$[\{\mathtt{C},\mathtt{D}\}]_\Gamma = \{\mathtt{C},\mathtt{D}\} \neq R, \quad [\{\mathtt{C},\mathtt{E}\}]_\Gamma = \{\mathtt{C},\mathtt{E},\mathtt{G}\} \neq R, \quad [\{\mathtt{D},\mathtt{E}\}]_\Gamma = \{\mathtt{D},\mathtt{E}\} \neq R.$$

Následující jsou nadklíče, ale nejsou klíče:

$$\{\mathtt{A},\mathtt{D},\mathtt{E}\}, \qquad \{\mathtt{B},\mathtt{D},\mathtt{G}\}, \qquad \{\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{F}\}, \qquad \{\mathtt{D},\mathtt{F},\mathtt{H}\}, \qquad \{\mathtt{A},\dots,\mathtt{H}\}$$

Například $\{\mathtt{A}\}$, $\{\mathtt{A},\mathtt{E}\}$, $\{\mathtt{E},\mathtt{F}\}$ nejsou nadklíče, protože $[\{\mathtt{A}\}]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C}\} \neq R$, $[\{\mathtt{A},\mathtt{E}\}]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{G}\} \neq R$ a $[\{\mathtt{E},\mathtt{F}\}]_\Gamma = \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{F},\mathtt{G}\} \neq R$.

Příklad (Nalezení klíče postupnou redukcí nadklíče)

Pokud je dána Γ nad R, (některý) klíč schématu R vzhledem k Γ lze nalézt tak, že vyjdeme z nadklíče $K = R$ a postupně z něj odebíráme atributy, dokud je množina pořád nadklíč. Pokud už žádný atribut nelze odebrat, výsledná množina K je klíč.

Pro $\Gamma = \{\{\mathtt{A}\}\Rightarrow\{\mathtt{B},\mathtt{C}\}, \{\mathtt{B},\mathtt{D}\}\Rightarrow\{\mathtt{E},\mathtt{F}\}, \{\mathtt{C},\mathtt{E}\}\Rightarrow\{\mathtt{G}\}, \{\mathtt{F}\}\Rightarrow\{\mathtt{A},\mathtt{B}\}, \{\mathtt{D},\mathtt{G}\}\Rightarrow\{\mathtt{A},\mathtt{C},\mathtt{H}\}\}$ z předchozích příkladů můžeme najít klíče K_1, K_2, K_3 takto:

$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{F},\mathtt{G},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{F},\mathtt{G},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{F},\mathtt{G},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{G},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{F},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{G},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{D},\mathtt{E},\mathtt{G},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{D},\mathtt{E},\mathtt{F},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{B},\mathtt{D},\mathtt{E},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{D},\mathtt{F},\mathtt{H}\}$	$\{\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{H}\}$
$\{\mathtt{A},\mathtt{D},\mathtt{E},\mathtt{H}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{D},\mathtt{H}\}$	$\{\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E}\}$
$\{\mathtt{A},\mathtt{D},\mathtt{E}\}$	$\{\mathtt{A},\mathtt{B},\mathtt{D}\}$	$\{\mathtt{C},\mathtt{D},\mathtt{E}\} = K_3$
$\{\mathtt{A},\mathtt{D}\} = K_1$	$\{\mathtt{B},\mathtt{D}\} = K_2$	

Analogicky pro K_4 a K_5.

pozor: algoritmus závisí na výběru prvku z aktuálního nadklíče (**!!**)

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 $\Gamma_1 = \{C \cap R_1 \Rightarrow D \cap R_1 \,|\, C \cap R_1 \Rightarrow D \in \Gamma\}$
3. položíme $R_2 = A \cup (R \setminus B)$ a $\Gamma_2 = \{C \cap R_2 \Rightarrow D \cap R_2 \,|\, C \cap R_2 \Rightarrow D \in \Gamma\}$
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

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

a teorii Γ popisující závislosti mezi atributy:

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

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

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

Příklad (Normalizace schématu pomocí dekompozice)

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

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

- $R_1 = \{\mathtt{SCHOOL}, \mathtt{DEAN}, \mathtt{DEPT}, \mathtt{HEAD}\}$
  $\Gamma_1 = \{\{\mathtt{DEPT}\} \Rightarrow \{\mathtt{HEAD}, \mathtt{SCHOOL}, \mathtt{DEAN}\}, \{\mathtt{SCHOOL}\} \Rightarrow \{\mathtt{DEAN}\}\}$
  - $R_{11} = \{\mathtt{SCHOOL}, \mathtt{DEAN}\}$
    $\Gamma_{11} = \{\{\mathtt{SCHOOL}\} \Rightarrow \{\mathtt{DEAN}\}\}$
  - $R_{12} = \{\mathtt{SCHOOL}, \mathtt{DEPT}, \mathtt{HEAD}\}$
    $\Gamma_{12} = \{\{\mathtt{DEPT}\} \Rightarrow \{\mathtt{HEAD}, \mathtt{SCHOOL}\}, \{\mathtt{SCHOOL}\} \Rightarrow \{\}\}$
- $R_2 = \{\mathtt{DEPT}, \mathtt{ID}, \mathtt{COURSE}, \mathtt{YEAR}\}$
  $\Gamma_2 = \{\{\mathtt{DEPT}\} \Rightarrow \{\}, \{\mathtt{COURSE}, \mathtt{YEAR}\} \Rightarrow \{\mathtt{ID}\}, \{\mathtt{ID}, \mathtt{YEAR}\} \Rightarrow \{\mathtt{DEPT}\}\}$
  - $R_{21} = \{\mathtt{DEPT}, \mathtt{ID}, \mathtt{YEAR}\}$
    $\Gamma_{21} = \{\{\mathtt{DEPT}\} \Rightarrow \{\}, \{\mathtt{ID}, \mathtt{YEAR}\} \Rightarrow \{\mathtt{DEPT}\}\}$
  - $R_{22} = \{\mathtt{ID}, \mathtt{COURSE}, \mathtt{YEAR}\}$
    $\Gamma_{22} = \{\{\mathtt{COURSE}, \mathtt{YEAR}\} \Rightarrow \{\mathtt{ID}\}, \{\mathtt{ID}, \mathtt{YEAR}\} \Rightarrow \{\}\}$

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

$\bowtie$

SCHOOL	DEPT	HEAD
SCI	AF	Durcet
SCI	CS	Curval

$\bowtie$

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

$\bowtie$

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:

- teorie, model, sémantické vyplývání
- kanonická relace, kanonický model
- sémantický uzávěr množiny atributů, charakterizace sémantického vyplývání
- 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