Databázové systémy

Integritní omezení, funkční závislosti a klíče

Petr Krajča

Katedra informatiky
Univerzita Palackého v Olomouci

Integritní omezení

Opakování:

- **Instance databáze** je konečná množina relační proměnných, jejich aktuálních hodnot a integritních omezení.
- **Dotaz** je částečně rekurzivní funkce z množiny všech instancí databáze do množiny všech relací (nad relačními schématy).

Úloha integritních omezení

Zajistit, aby bylo možné provádět pouze ta relační přiřazení (modifikace relačních proměnných), která zaručují, že databáze bude v konzistentním stavu – všechna integritní omezení budou splněna.

Relační algebra

Dosud jsme ukázali:

- formalizaci pojmu relace (a souvisejících pojmů)
- množinu relačních operací (některé odvoditelné z ostatních)
- korespondenci v jazyce SQL (slabší)

výrazy relační algebry

- skalární
- relační

dva aspekty relačních výrazů

- **syntaxe výrazů** $\mathfrak{E}, \mathfrak{F}, \ldots$ (definujeme jejich přípustné tvary)
- **sémantika výrazů** $\mathfrak{E}^{\mathrm{D}}, \mathfrak{F}^{\mathrm{D}}, \ldots$ (definujeme jejich interpretaci v instancích databáze)

Relační model: integritní omezení

Množina integritních omezení relační databáze je konečná množina skalárních výrazů typu ,,pravdivostní hodnota". Instance databáze splňuje danou množinu integritních omezení, pokud jsou všechny dané skalární výrazy *pravdivé*.

integritní omezení vs. efektivita

- integritní omezení = prostředek zaručení konzistence dat
- součást logického modelu dat (již od počátku jedna z komponent rel. modelu)
- neplést s prostředky efektivity: indexy apod. (záležitost fyzické vrstvy)

Jak formulovat?

- obecné pravidlo – čím více, tím lépe
- je možné rozlišit: integritní omezení na straně serveru vs. uživatele (klienta)
- na straně uživatele (klienta) se považuje za nedostatečné

SQL: integritní omezení

```
CREATE TABLE jmeno-tabulky (
definice-atributu CONSTRAINT jmeno-omezeni1 definice-omezeni1,
CONSTRAINT jmeno-omezeniN definice-omezeniN);

ALTER TABLE jmeno-tabulky ADD CONSTRAINT jmeno-omezeni definice-omezeni;
ALTER TABLE jmeno-tabulky DROP CONSTRAINT jmeno-omezeni;

CREATE TABLE tab (
foo NUMERIC NOT NULL PRIMARY KEY,
bar NUMERIC NOT NULL,
baz VARCHAR NOT NULL);

ALTER TABLE tab ADD CONSTRAINT tab_nnega CHECK (foo < 0);
ALTER TABLE tab ADD CONSTRAINT tab_foobar CHECK (foo < bar);

INSERT INTO tab VALUES (10, 20, 'aaa');
INSERT INTO tab VALUES (-1, 20, 'bbb'); /* fail */
INSERT INTO tab VALUES (30, 20, 'ccc'); /* fail */
```

Opakování: Klíč

Množina klíčů (relační) proměnné X je libovolná neprázdná množina $\{K_1, \ldots, K_n\}$, jejíž prvky jsou podmnožiny R a splňují podmínku, že $K_i \nsubseteq K_j$ pro každé $i \neq j$.

Relační přiřazení

Mějme relační proměnnou X typu R a nechť $\{K_1, \ldots, K_n\}$ je množina klíčů proměnné X.

Pak relaci $\mathcal{D}$ typu R lze **přiřadit jako hodnotu** proměnné X, pokud je splněna podmínka, že pro každé $i = 1, \ldots n$ a libovolné $r_1, r_2 \in \mathcal{D}$ platí:

$$\text{Pokud } r_1(y) = r_2(y) \text{ pro každý } y \in K_i, \text{ pak } r_1 = r_2.$$

V opačném případě říkame, že $\mathcal{D}$ porušuje integritní omezení dané některým klíčem X.

Klíč jako integritní omezení

Předpoklad:

- relační proměnná $\mathbb{R}$ typu R
- množina atributů $K \subseteq R$

Integritní omezení vyjadřující, že K je klíč $\mathbb{R}$ (resp. nadklíč) pro $R = \{y_1, \ldots, y_k, \ldots, y_n\}$ a $K = \{y_1, \ldots, y_k\}$, uvažujeme jako výraz:

$$\sigma_{y_1=y'_1 \wedge \cdots \wedge y_k=y'_k \wedge \neg(y_{k+1}=y'_{k+1} \wedge \cdots \wedge y_n=y'_n)}(\mathbb{R} \bowtie \rho_{y'_1 \leftarrow y_1, \ldots, y'_n \leftarrow y_n}(\mathbb{R})) = \emptyset_R$$

Předchozí výraz je pravdivý v instanci databáze D, pokud pro každé dvě n-tice $r, r' \in \mathbb{R}^{\mathtt{D}}$ platí:

$$\text{pokud } r(K) = r'(K), \text{ pak } r = r'.$$

SQL

- klauzule `UNIQUE`
- unikátní indexy

Motivace pro funkční závislosti (1/2)

vztahy mezi atributy:

- uvažujme relaci evidující adresy osob žijící v ČR s atributy: jméno (JMENO), město (MESTO), ulice (ULICE), číslo popisné (CP), číslo orientační (CO), PSČ (PSC).
- platí:
  - $\{\mathtt{PSC}\} \Rightarrow \{\mathtt{MESTO}\}$
  - $\{\mathtt{MESTO}, \mathtt{ULICE}, \mathtt{CO}\} \Rightarrow \{\mathtt{CP}\}$
  - $\{\mathtt{PSC}, \mathtt{ULICE}, \mathtt{CO}\} \Rightarrow \{\mathtt{CP}\}$
  - $\{\mathtt{MESTO}, \mathtt{CP}\} \Rightarrow \{\mathtt{MESTO}, \mathtt{ULICE}, \mathtt{CO}\}$
  - $\{\mathtt{JMENO}\} \Rightarrow \{\mathtt{MESTO}, \mathtt{ULICE}, \mathtt{CP}, \mathtt{CO}, \mathtt{PSC}\}$
  - ...

Otázky:

- Jsou to všechny závislosti v datech?
- Jak to zjistím?
- Proč je to důležité?

Motivace pro funkční závislosti (2/2)

SCHOOL	DEAN	DEPT	HEAD	ID	COURSE	YEAR
SCI	Adams	AF	Black	7	QOPT1	2012
SCI	Adams	AF	Black	8	LASR1	2012
SCI	Adams	AF	Black	8	LASR1	2013
SCI	Adams	CS	Davis	3	ALMA1	2012
SCI	Adams	CS	Davis	3	ALMA1	2013
SCI	Adams	CS	Davis	6	DATA1	2012
SCI	Adams	CS	Davis	6	DATA1	2013
SCI	Adams	CS	Davis	6	PAPR1	2012

Problémy

- **redundance dat** (zbytečná duplikace hodnot)
- **anomálie spojená s výmazem dat** (výmaz kurzů katedry „odstraní vedoucího")
- **anomálie spojená s aktualizací hodnot** (změna vedoucího katedry na více místech)

Bezeztrátová dekompozice relace

- Je možné vyjádřit výchozí relaci pomocí spojení některých jejich projekcí?

Relace $\mathcal{D}$ na schématu $R_1 \cup \cdots \cup R_n$ má bezeztrátovou dekompozici vzhledem k množině $\{R_1, \ldots, R_n\}$, pokud $\mathcal{D} = \bowtie_{i=1}^{n} \pi_{R_i}(\mathcal{D})$.

Poznámka

- Každá relace $\mathcal{D}$ na schématu R má jednoprvkovou bezeztrátovou dekompozici $\{R\}$ (triviální, nezajímavé).
- Duální pojem k úplnému spojení.

Příklad bezeztrátové dekompozice (1/2)

SCHOOL	DEAN	DEPT	HEAD	ID	COURSE	YEAR
SCI	Adams	AF	Black	7	QOPT1	2012
SCI	Adams	AF	Black	8	LASR1	2012
SCI	Adams	AF	Black	8	LASR1	2013
SCI	Adams	CS	Davis	3	ALMA1	2012
SCI	Adams	CS	Davis	3	ALMA1	2013
SCI	Adams	CS	Davis	6	DATA1	2012
SCI	Adams	CS	Davis	6	DATA1	2013
SCI	Adams	CS	Davis	6	PAPR1	2012

=

SCHOOL	DEAN	DEPT	HEAD
SCI	Adams	AF	Black
SCI	Adams	CS	Davis

$\bowtie$

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

Příklad bezeztrátové dekompozice (2/2)

$\mathcal{D}$

FOO	BAR	BAZ	QUX
100	aaa	444	GGG
100	aaa	444	HHH
200	bbb	555	III
300	ccc	555	III

$\pi_{\{\mathtt{FOO},\mathtt{BAR},\mathtt{BAZ}\}}(\mathcal{D})$

FOO	BAR	BAZ
100	aaa	444
200	bbb	555
300	ccc	555

$\pi_{\{\mathtt{BAZ},\mathtt{QUX}\}}(\mathcal{D})$

BAZ	QUX
444	GGG
444	HHH
555	III

$\pi_{\{\mathtt{BAR},\mathtt{QUX}\}}(\mathcal{D})$

BAR	QUX
aaa	GGG
aaa	HHH
bbb	III
ccc	III

$\pi_{\{\mathtt{FOO},\mathtt{BAR}\}}(\mathcal{D})$

FOO	BAR
100	aaa
200	bbb
300	ccc

Například platí

- $\mathcal{D} = \pi_{\{\mathtt{FOO},\mathtt{BAR},\mathtt{BAZ}\}}(\mathcal{D}) \bowtie \pi_{\{\mathtt{BAZ},\mathtt{QUX}\}}(\mathcal{D})$
- $\mathcal{D} = \pi_{\{\mathtt{FOO},\mathtt{BAR},\mathtt{BAZ}\}}(\mathcal{D}) \bowtie \pi_{\{\mathtt{BAR},\mathtt{QUX}\}}(\mathcal{D})$

Například neplatí

- $\mathcal{D} \neq \pi_{\{\mathtt{FOO},\mathtt{BAR}\}}(\mathcal{D}) \bowtie \pi_{\{\mathtt{BAZ},\mathtt{QUX}\}}(\mathcal{D})$

Funkční závislosti

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

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 $\mathcal{D}$. (Např. $A \Rightarrow B$ je triviální, právě když $A \subseteq B$.)

Příklady funkčních závislostí

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

FOO	BAR	BAZ	QUX		
10	22	a	222	$\ldots$	r_1
10	33	b	333	$\ldots$	r_2
10	22	a	444	$\ldots$	r_3
20	33	a	555	$\ldots$	r_4

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

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

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

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

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

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

Vlastnosti 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 $\mathcal{D}_1 \models A \Rightarrow B$

Dekompozice na základě funkčních závislostí

Mějme relaci $\mathcal{D}$ 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 $A \cup (R \setminus B)$.

Dekompozice na základě funkčních závislostí

SCHOOL	DEAN	DEPT	HEAD	ID	COURSE	YEAR
SCI	Adams	AF	Black	7	QOPT1	2012
SCI	Adams	AF	Black	8	LASR1	2012
SCI	Adams	AF	Black	8	LASR1	2013
SCI	Adams	CS	Davis	3	ALMA1	2012
SCI	Adams	CS	Davis	3	ALMA1	2013
SCI	Adams	CS	Davis	6	DATA1	2012
SCI	Adams	CS	Davis	6	DATA1	2013
SCI	Adams	CS	Davis	6	PAPR1	2012

$=$

SCHOOL	DEAN	DEPT	HEAD
SCI	Adams	AF	Black
SCI	Adams	CS	Davis

$\bowtie$

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

Využita závislost: $\{\mathtt{DEPT}\} \Rightarrow \{\mathtt{HEAD}, \mathtt{SCHOOL}, \mathtt{DEAN}\}$

Teorie a model

- 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.

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.

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 $\mathrm{Mod}(\Gamma)$, tj.:

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

Teorie a model (příklad)

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 teorie 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
20	22	a	444
20	33	a	555

není model

Sémantické vyplývání

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 $\Gamma \models A \Rightarrow B$.

Jinými slovy

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 případy

- pro $\Gamma = \emptyset$ píšeme pouze $\models A \Rightarrow B$
- $\models A \Rightarrow B$ právě když $A \Rightarrow B$ je pravdivá v každé relaci (tj. triviální)

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 $A \Rightarrow B \in \Gamma$

Příklad netriviálních funkčních závislostí

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 $\Gamma \models A \Rightarrow B$:

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

Algoritmus pro výpočet sémantického uzávěru

- umožňuje nám zavést efektivní test sémantického vyplývání

Data: $M \subseteq R$ a teorie Γ nad R
Result: Množina $[M]_\Gamma \subseteq R$ (**sémantický uzávěr** M **vzhledem k** Γ)
$W := M$; /* W – pomocná proměnná */
repeat
  $L := W$; /* L – poslední vypočtená hodnota W */
  foreach $E \Rightarrow F \in \Gamma$ **do**
    if $E \subseteq W$ **then**
      $W := W \cup F$; /* aktualizuj W */
      $\Gamma := \Gamma \backslash \{E \Rightarrow F\}$; /* $E \Rightarrow F$ už není potřeba v Γ */
    end
  end
until $L = W$;
return W;

Příklad

$W:$ $\{A\}$, $\{A,B,C\}$

$\Gamma:$ $\{A\}\Rightarrow\{B,C\}, \{B,D\}\Rightarrow\{E,F\}, \{C,E\}\Rightarrow\{G\}, \{F\}\Rightarrow\{A,B\}, \{D,G\}\Rightarrow\{A,C,H\}$

$W:$ $\{A,E\}$, $\{A,B,C,E\}$, $\{A,B,C,E,G\}$

$\Gamma:$
$\{A\}\Rightarrow\{B,C\}, \{B,D\}\Rightarrow\{E,F\}, \{C,E\}\Rightarrow\{G\}, \{F\}\Rightarrow\{A,B\}, \{D,G\}\Rightarrow\{A,C,H\}$
$\{B,D\}\Rightarrow\{E,F\}, \{C,E\}\Rightarrow\{G\}, \{F\}\Rightarrow\{A,B\}, \{D,G\}\Rightarrow\{A,C,H\}$

$W:$ $\{E,F\}$, $\{A,B,E,F\}$, $\{A,B,C,E,F\}$, $\{A,B,C,E,F,G\}$

$\Gamma:$
$\{A\}\Rightarrow\{B,C\}, \{B,D\}\Rightarrow\{E,F\}, \{C,E\}\Rightarrow\{G\}, \{F\}\Rightarrow\{A,B\}, \{D,G\}\Rightarrow\{A,C,H\}$
$\{A\}\Rightarrow\{B,C\}, \{B,D\}\Rightarrow\{E,F\}, \{C,E\}\Rightarrow\{G\}, \{D,G\}\Rightarrow\{A,C,H\}$
$\{B,D\}\Rightarrow\{E,F\}, \{C,E\}\Rightarrow\{G\}, \{D,G\}\Rightarrow\{A,C,H\}$

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

Vlastnosti sémantického uzávěru

1. Pro libovolnou Γ nad R a $A \subseteq R$ platí, že $\Gamma \models A \Rightarrow [A]_\Gamma$.
2. Pokud $\Gamma \models A \Rightarrow B$, pak $B \subseteq [A]_\Gamma$.

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

1. $\Gamma \models A \Rightarrow B$
2. $B \subseteq [A]_\Gamma$

Poznámky

- $[A]_\Gamma$ je největší prvek množiny $\{B \subseteq R \mid \Gamma \models A \Rightarrow B\}$
- druhá podmínka 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 $M = \{A, E\}$ 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 $\Gamma \models M \Rightarrow N$ právě když $N \subseteq \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{E},\mathtt{G}\}$.

Například tedy platí:

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

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

Na druhou stranu například:

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

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

Klíče z pohledu funkčních závislostí

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 $K' \subseteq K$ není nadklíč R vzhledem k Γ, pak je K **klíč** R vzhledem k Γ.

Vztah s pojmem klíče relační proměnné?

Následující je ekvivalentní

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

Poznámka

- každý nadklíč (klíč R vzhledem k Γ) obsahuje nějaký klíč (klíč R vzhledem k Γ)
- někdy užívány pojmy:
  - kandidátní klíče – množina klíčů
  - primární klíč – jeden zvolený kandidátní klíč

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: $R = \{\mathtt{A}, \ldots, \mathtt{H}\}$ vzhledem ke 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:

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

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 dokázat pro $K_2, \ldots K_5$. 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:

$$\{\mathtt{A}, \mathtt{D}, \mathtt{E}\}, \quad \{\mathtt{B}, \mathtt{D}, \mathtt{G}\}, \quad \{\mathtt{C}, \mathtt{D}, \mathtt{E}, \mathtt{F}\}, \quad \{\mathtt{D}, \mathtt{F}, \mathtt{H}\}, \quad \{\mathtt{A}, \ldots, \mathtt{H}\}.$$

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

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 (!!)