

Databázové systémy \diamond poznámky k přednášce

6. Funkční závislosti

verze z 6. listopadu 2024

Je dáno relační schéma S . Pro dvě množiny $X, Y \subseteq S$ se výraz

$$X \rightarrow Y$$

nazývá **funkční závislost** nad S . Například

$$\{\text{movie_title}, \text{movie_year}\} \rightarrow \{\text{movie_length}\}$$

je funkční závislost nad

$$\{\text{movie_title}, \text{movie_year}, \text{movie_length}, \text{actor_name}\}.$$

Místo $\{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_m\}$ můžeme psát $x_1, \dots, x_n \rightarrow y_1, \dots, y_m$. Tedy předchozí funkční závislost můžeme zapsat

$$\text{movie_title}, \text{movie_year} \rightarrow \text{movie_length}.$$

Relace R nad S **splňuje** funkční závislost $X \rightarrow Y$, jestliže pro každé dvě n -tice $t_1, t_2 \in R$ platí, že jestliže se t_1, t_2 shodují v X , pak se shodují i v Y . Pokud R splňuje funkční závislost $X \rightarrow Y$, pak říkáme, že v R je Y **funkčně závislé** na X .
Například v relaci

movie_title	movie_year	movie_length	actor_name
The Aviator	1985	98	Christopher Reeve
The Aviator	2004	170	Leonardo DiCaprio
The Aviator	2004	170	Cate Blanchett
Eternal Sunshine of the Spotless Mind	2004	108	Jim Carrey
My Fair Lady	1964	170	Audrey Hepburn
Breakfast at Tiffany's	1961	115	Audrey Hepburn

danou vlastností „Film `movie_title` dlouhý `movie_length` minut byl vytvořený roku `movie_year` a hraje v něm herec `actor_name`.“ je $\{\text{movie_length}\}$ funkčně závislé na $\{\text{movie_title}, \text{movie_year}\}$, ale $\{\text{movie_title}\}$ není funkčně závislé na $\{\text{actor_name}\}$.

Následující věta objasňuje původ názvu funkční závislost.

Věta 1. V relaci R je Y funkčně závislé na X , právě když binární relace

$$\{\langle t(X), t(Y) \rangle \mid t \in R\}$$

je zobrazení množiny $\pi_X(R)$ do množiny $\pi_Y(R)$.

Důkaz: Relaci $\{\langle t(X), t(Y) \rangle \mid t \in R\}$ si označíme F . Předpokládejme nejprve, že v R je Y funkčně závislé na X . Vezměme $t' \in \pi_X(R)$. Z definice projekce vyplývá, že existuje $t \in R$ tak, že $t(X) = t'$, tedy $\langle t', t(Y) \rangle = \langle t(X), t(Y) \rangle \in F$. Dále vezměme $\langle t', t'_1 \rangle, \langle t', t'_2 \rangle \in F$. Z definice relace F plyne, že existují $t_1, t_2 \in R$ tak, že $t_1(X) = t' = t_2(X)$, $t_1(Y) = t'_1$ a $t_2(Y) = t'_2$. Tedy t_1 a t_2 se shodují v X a protože je Y v R funkčně závislé na X , shodují se i v Y . Tedy $t'_1 = t_1(Y) = t_2(Y) = t'_2$. Ukázali jsme, že F je zobrazení $\pi_X(R)$ do $\pi_Y(R)$.

Obráceně předpokládejme, že F je zobrazení $\pi_X(R)$ do $\pi_Y(R)$ a vezměme n -tice $t_1, t_2 \in R$, které se shodují v X . Z definice zobrazení F plyne, že $F(t_1(X)) = t_1(Y)$ a $F(t_2(X)) = t_2(Y)$. Protože $t_1(X) = t_2(X)$ dostáváme, že $t_1(Y) = F(t_1(X)) = F(t_2(X)) = t_2(Y)$. Tedy t_1, t_2 se shodují v Y . Ukázali jsme, že v R je Y funkčně závislé na X . \square

Podíváme se, jak je možné funkční závislost charakterizovat pomocí relačních operací.

Věta 2. Předpokládejme, že $X = \{x_1, \dots, x_m\}$ a $Y = \{y_1, \dots, y_n\}$ a vezměme schéma Z takové, že $X \cup Y$ je disjunktní se Z , a přejmenování $h: (X \cup Y) \rightarrow Z$. Dále označme θ podmínku

$$(x_1 = h(x_1)) \wedge \dots \wedge (x_m = h(x_m)) \wedge \neg((y_1 = h(y_1)) \wedge \dots \wedge (y_n = h(y_n))).$$

Pak v R je Y funkčně závislé na X , právě když

$$\sigma_\theta(\pi_{X \cup Y}(R) \times \rho_h(\pi_{X \cup Y}(R))) = \emptyset.$$

Důkaz: Označme $X' = \{h(x) \mid x \in X\}$ a $Y' = \{h(y) \mid y \in Y\}$. Zvolme libovolné n -tice $t_1, t_2 \in R$ a položme $t = t_1(X \cup Y) \cup \rho_h(t_2(X \cup Y))$. Jistě $t \in \pi_{X \cup Y}(R) \times \rho_h(\pi_{X \cup Y}(R))$. Dostáváme, že $t \in \sigma_\theta(\pi_{X \cup Y}(R) \times \rho_h(\pi_{X \cup Y}(R)))$, právě když $\rho_h(t(X)) = t(X')$ a $\rho_h(t(Y)) \neq t(Y')$ a $t \in \pi_{X \cup Y}(R) \times \rho_h(\pi_{X \cup Y}(R))$, právě když $\rho_h(t(X)) = t(X')$ a $\rho_h(t(Y)) \neq t(Y')$, právě když $t_1(X) = t_2(X)$ a $t_1(Y) \neq t_2(Y)$. \square

Podle předchozí věty může v SQL rozhodnout, zda funkční závislost platí. Například v hodnotě proměnné:

```
# TABLE movie;
  title      | length
-----+-----
The Matrix   |    136
My Fair Lady |    170
The Aviator  |    170
```

dané vlastností „Film title je dlouhý length minut.“ je length funkčně závislé na title, protože:

```
# SELECT m1.title AS title1,
```

```

        m1.length AS length1,
        m2.title AS title2,
        m2.length AS length2
FROM    movie AS m1, movie AS m2
WHERE   m1.title = m2.title
AND     NOT ( m1.length = m2.length );

title1 | length1 | title2 | length2
-----+-----+-----+-----
(0 rows)

```

je prázdná množina, ale title už není závislé na length, protože:

```

# SELECT m1.title AS title1,
        m1.length AS length1,
        m2.title AS title2,
        m2.length AS length2
FROM    movie AS m1, movie AS m2
WHERE   m1.length = m2.length
AND     NOT ( m1.title = m2.title );

title1 | length1 | title2 | length2
-----+-----+-----+-----
My Fair Lady | 170 | The Aviator | 170
The Aviator | 170 | My Fair Lady | 170
(2 rows)

```

je neprázdná množina. Ve výsledku dostáváme protipříklady platnosti funkční závislosti.

Funkční závislost na R se nazývá **triviální**, pokud platí v každé relaci nad S . Například $\text{movie_title}, \text{movie_year} \rightarrow \text{movie_title}$ je triviální. Triviální funkční závislosti můžeme jednoduše charakterizovat.

Věta 3. *Funkční závislost $X \rightarrow Y$ je triviální, právě když $Y \subseteq X$.*

Důkaz: Předpokládejme, že $X \rightarrow Y$ je triviální a $Y \not\subseteq X$. Z druhého předpokladu plyne, že existuje $y \in Y$, které není v X . Označme $S = X \cup Y$ a vezměme libovolnou relaci $R = \{t_1, t_2\}$ nad S takovou, že $t_1(X) = t_2(X)$, ale $t_1(y) \neq t_2(y)$. Relace R nespĺňuje funkční závislost $X \rightarrow Y$. Dostáváme spor s triviálností $X \rightarrow Y$. Proto je nutně Y podmnožina X .

Předpokládejme naopak, že $Y \subseteq X$, vezměme relaci R nad S a $t_1, t_2 \in R$ shodující se v X . Protože Y je podmnožina X , musí se t_1, t_2 shodovat i v Y . Tedy R splňuje $X \rightarrow Y$. Ukázali jsme, že $X \rightarrow Y$ je splněna v každé relaci nad S a tedy, že je triviální. \square

Funkční závislosti lze použít k charakterizaci nadklíčů relace.

Věta 4. *Množina $K \subseteq S$ je nadklíč relace R nad S , právě když v R je S funkčně závislé na K .*

Důkaz: Vezměme dvě libovolné n -tice $t_1, t_2 \in R$ shodující se v K . Pokud K je nadklíč R , pak $t_1(S) = t_1 = t_2 = t_2(S)$ a tedy t_1, t_2 se shodují v S . Pokud naopak S je v R funkčně závislé na K , pak se t_1 a t_2 shodují v S a tedy se rovnají. \square

Vezměme relační proměnou *relation* nad S . Proměnná *relation* **splňuje** funkční závislost $X \rightarrow Y$, pokud každá možná hodnota proměnné *relation* splňuje $X \rightarrow Y$. Pokud *relation* splňuje funkční závislost $X \rightarrow Y$, pak taky říkáme, že $X \rightarrow Y$ **je** funkční závislost proměnné *relation* nebo také, že v *relation* je Y **funkčně závislé** na X .

Například vezměme relační proměnnou `movie_library` danou vlastností „Film `movie_title` dlouhý `movie_length` minut byl vytvořený roku `movie_year` a hraje v něm herec `actor_name`.“ s aktuální hodnotou

<code>movie_title</code>	<code>movie_year</code>	<code>movie_length</code>	<code>actor_name</code>
The Aviator	1985	98	Christopher Reeve
The Aviator	2004	170	Leonardo DiCaprio
The Aviator	2004	170	Cate Blanchett
Eternal Sunshine of the Spotless Mind	2004	108	Jim Carrey
Breakfast at Tiffany's	1961	115	Audrey Hepburn

která sice splňuje funkční závislost $\text{movie_length} \rightarrow \text{movie_title}$, ale proměnná `movie_library` ji již nespĺňuje – dva různé filmy mohou mít tutěž délku. Jako možnou hodnotu, kde závislost není splněna se podívejte na dříve zobrazenou relaci.

1 Základní tvar funkční závislosti

Řekneme, že funkční závislost $X \rightarrow Y$ relační proměnné *relation* nad S je v **základním tvaru**, jestliže pro libovolné množiny $X' \subseteq X$ a $Y \subseteq Y' \subseteq S$ takové, že proměnná *relation* splňuje $X' \rightarrow Y'$ platí, že $X = X'$ a $Y = Y'$. Jistě pro funkční závislost $X \rightarrow Y$ proměnné v základním tvaru je $X \subseteq Y$. Například uvažujme základní relaci `movie` danou vlastností „Film `movie_title` vydaný roku `movie_year` je dlouhý `movie_length` minut a režíroval jej `director_name`.“ Předpokládejme, že film je určen jménem a rokem vydání a že každý film režíroval právě jeden režisér.

Všechny funkční závislosti proměnné `movie` v základním tvaru:

1. `movie_title, movie_year` \rightarrow `movie_title, movie_year, movie_length, director_name`
2. `movie_title` \rightarrow `movie_title`
3. `movie_year` \rightarrow `movie_year`

4. $\text{movie_length} \rightarrow \text{movie_length}$
5. $\text{director_name} \rightarrow \text{director_name}$

Následující funkční závislosti nejsou v základním tvaru.

1. $\text{movie_title}, \text{movie_year}, \text{movie_length} \rightarrow \text{movie_title}, \text{movie_year}, \text{movie_length}, \text{director_name}$
2. $\text{movie_title}, \text{movie_year} \rightarrow \text{movie_title}, \text{movie_year}, \text{movie_length}$
3. $\text{movie_title}, \text{movie_year} \rightarrow \text{movie_year}, \text{movie_length}, \text{director_name}$

Věta 5. *Funkční závislost $X \rightarrow Y$ v základním tvaru proměnné relation je triviální, právě když $X = Y$.*

Důkaz: Pokud je $X \rightarrow Y$ triviální, pak ze základního tvaru funkční závislosti $X \rightarrow Y$ dostáváme, že $X \subseteq Y$ a z předpokladu $Y \subseteq X$.

Pokud $X = Y$, pak jistě $Y \subseteq X$ a tedy $X \rightarrow Y$ je triviální. □

Triviální funkční závislosti v základním tvaru můžeme při popisu proměnné vynechat. Například výše uvedená proměnná *movie* má jedinou netriviální funkční závislost v základním tvaru:

$\text{movie_title}, \text{movie_year} \rightarrow \text{movie_title}, \text{movie_year}, \text{movie_length}, \text{director_name}$.

Věta 6. *Jestliže proměnná relation nad S splňuje $X \rightarrow Y$, pak pro $X \subseteq X'$ a $Y' \subseteq Y$ proměnná relation splňuje i $X' \rightarrow Y'$.*

Důkaz: Vezměme dvě n -tice t_1, t_2 nad S , které se shodují v X' . Z $X \subseteq X'$ plyne, že t_1, t_2 se shodují také v X . Z platnosti $X \rightarrow Y$ v *relation* plyne, že t_1, t_2 se shodují v Y , ale $Y' \subseteq Y$, tedy se shodují i v Y' . □

Věta 7. *Pro každou funkční závislost $X \rightarrow Y$ proměnné relation nad S existuje funkční závislost $X' \rightarrow Y'$ proměnné relation v základním tvaru taková, že $X' \subseteq X$ a $Y \subseteq Y'$.*

Důkaz: Opakujme následující, dokud $X \rightarrow Y$ není v základním tvaru. Protože závislost není v základním tvaru, existují $X' \subseteq X$ a $Y \subseteq Y' \subseteq R$ tak, že *relation* splňuje $X' \rightarrow Y'$ a $X' \subsetneq X$ nebo $Y \subsetneq Y'$. Místo $X \rightarrow Y$ vezměme $X' \rightarrow Y'$.

Z konečnosti S plyne, že po konečně mnoha krocích převedeme $X \rightarrow Y$ do základního tvaru. □

Poznamenejme, že převod funkční závislosti na základní tvar podle předchozího důkazu není jednoznačný.

Věta 8. *Schéma K je kandidátní klíč proměnné relation nad S , právě když $K \rightarrow S$ je funkční závislost proměnné relation v základním tvaru.*

Důkaz: Pokud K je kandidátní klíč proměnné *relation*, pak *relation* splňuje $K \rightarrow S$ a neexistuje vlastní podmnožina K' množiny K taková, že *relation* splňuje $K' \rightarrow R$.

Pokud je $K \rightarrow S$ funkční závislost proměnné *relation* v základním tvaru, pak K je nadklíč proměnné *relation* a neexistuje vlastní podmnožina K' množiny K taková, že K' je nadklíč *relations*. \square

2 Vynucení funkční závislosti v PostgreSQL

V PostgreSQL můžeme funkční závislost vynutit. Vezměme základní relaci *relation* a její i -tou funkční závislost $\{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_m\}$, pak následující dva příkazy vynutí, že $\{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_m\}$ je vždy splněna v hodnotě *relation*.

```
CREATE OR REPLACE FUNCTION relation_fd_i_trg () RETURNS TRIGGER AS
$$
BEGIN
  IF ( 0 < ( SELECT count(*) FROM relation
            WHERE  $x_1 = \text{NEW}.x_1$ 
            :
            AND  $x_n = \text{NEW}.x_n$ 
            AND NOT (  $y_1 = \text{NEW}.y_1$ 
            :
            AND  $y_m = \text{NEW}.y_m$  ) ) ) THEN
    RAISE EXCEPTION 'FD  $\{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_m\}$  violated!';
  END IF;
  RETURN NEW;
END;
$$ LANGUAGE plpgsql;

CREATE TRIGGER relation_fd_i BEFORE INSERT OR UPDATE ON relation
FOR EACH ROW EXECUTE PROCEDURE relation_fd_i_trg ();
```

Například uvažujme relační proměnnou *movie* danou vlastností „Film *title* je dlouhý *length* minut“. Deklarujeme, že v *movie* je $\{\text{length}\}$ funkčně závislé na $\{\text{title}\}$:

```
CREATE OR REPLACE FUNCTION movie_fd_1_trg () RETURNS TRIGGER AS $$
BEGIN
  IF ( 0 < ( SELECT count(*) FROM movie
            WHERE title = NEW.title
            AND NOT ( length = NEW.length ) ) ) THEN
    RAISE EXCEPTION 'FD  $\{\text{title}\} \rightarrow \{\text{length}\}$  violated!';
  END IF;
END;
```

```

    RETURN NEW;
END;
$$ LANGUAGE plpgsql;

CREATE TRIGGER movie_fd_1 BEFORE INSERT OR UPDATE ON movie
    FOR EACH ROW EXECUTE PROCEDURE movie_fd_1_trg();

```

Hodnota proměnné je:

```

    title      | length
-----+-----
The Matrix    |    136
My Fair Lady  |    170
The Aviator   |    170
(3 rows)

```

Následující příkaz by porušil deklarovanou funkční závislost a proto skončí chybou:

```

# INSERT INTO movie VALUES
    ( 'The Matrix', 140 );
ERROR:  FD {title} -> {length} violated!
CONTEXT:  PL/pgSQL function movie_fd_1_trg() line 6 at RAISE

```

Otázky a úkoly na cvičení

1. Uvažujme relaci

actor	movie	year	role
Alain Delon	Rocco e i suoi fratelli	1960	Rocco Parondi
Renato Salvatori	Rocco e i suoi fratelli	1960	Simone Parondi
Alain Delon	Plein soleil	1960	Tom Ripley

danou vlastností „Herec `actor` hrál ve filmu `movie` z roku `year` postavu `role`.“ Určete:

- (a) Všechny funkční závislosti, které relace splňuje.
 - (b) Všechny netriviální funkční závislosti, které relace splňuje.
 - (c) Všechny netriviální funkční závislosti v základním tvaru, které relace splňuje.
2. Uvažujme proměnnou `movie_cast` danou vlastností „Herec `actor` hrál ve filmu `movie` z roku `year` postavu `role`.“ Předpokládejte, že film je jednoznačně určen názvem. Určete:
- (a) Všechny funkční závislosti, které relace splňuje.
 - (b) Všechny netriviální funkční závislosti, které relace splňuje.

- (c) Všechny netriviální funkční závislosti v základním tvaru, které relace splňuje.

Předpokládejte, že herec je určený jménem a film názvem a rokem vydání.

3. Určete všechny netriviální funkční závislosti v základním tvaru, které splňuje relace

actor	movie
Alain Delon	Rocco e i suoi fratelli
Renato Salvatori	Rocco e i suoi fratelli
Alain Delon	Plein soleil

daná vlastností „Herec **actor** si zahrál ve filmu **movie**.“

4. Vezměte proměnnou z druhého úkolu a podle funkčních závislostí v základním tvaru určete její kandidátní klíče.
5. Deklarujte proměnnou z druhého úkolu v SQL a vynuťte, aby její hodnota vždy splňovala všechny netriviální funkční závislosti v základním tvaru.