



2. Množinové operace a restrikce

verze z 30. září 2024

1 Relační výrazy

Relační výraz SQL je výraz SQL, jehož hodnota je relace. Každému relačnímu výrazu je přiřazený jeho **typ**, kterým je relační schéma. Každý relační výraz má **charakteristickou vlastnost**, pro kterou vždy platí, že je charakteristickou vlastností hodnoty výrazu.

Pro relační proměnnou *relation* je

```
( TABLE relation )
```

relační výraz, jehož hodnotou je hodnota proměnné *relation*. Typ výrazu je též jako typ proměnné. Charakteristická vlastnost proměnné je i charakteristickou vlastností výrazu.

Vezměme si například základní relaci z minulé přednášky definovanou příkazem:

```
CREATE TABLE movie (  
    title text,  
    year integer  
);
```

s charakteristickou vlastností „Film *title* je z roku *year*.“ Pak relační výraz

```
( TABLE movie )
```

má charakteristickou vlastnost „Film *title* je z roku *year*.“, jeho typ je {*title*, *year*} a hodnota výrazu je hodnota proměnné *movie*.

Nejvíce vnější závorky v relačních výrazech budeme vynechávat, proto můžeme předchozí výraz zjednodušit na:

```
TABLE movie
```

Pro relační výraz *expression* je

```
expression ;
```

příkaz, který vrátí hodnotu relačního výrazu *expression*.

Tedy příkaz

```
TABLE movie;
```

vrátí hodnotu proměnné *movie*.

2 Sjednocení

Pro motivaci předpokládejme, že R_1 je relace:

title	year
The Matrix	1999
The Avengers	2012

daná vlastností: „Vlastním film *title* vydaný v roce *year*.“ a R_2 je relace:

title	year
The Matrix	1999
The Avengers	1998

daná vlastností: „Vlastníš film *title* vydaný v roce *year*.“ Všimněme si, že obě relace jsou nad $\{\text{title}, \text{year}\}$. Relace R_3 daná vlastností: „Film *title* vydaný v roce *year* vlastním já nebo ty.“ je též jako relace:

title	year
The Matrix	1999
The Avengers	1998
The Avengers	2012

Cílem je relaci R_3 z relací R_1 a R_2 vypočítat. Uvědomme si, že tabulkami zapsané relace R_1 , R_2 a R_3 jsou jistými množinami. Konkrétněji, že

$$\begin{aligned} R_1 &= \{ \langle \text{title}, \text{'The Matrix'} \rangle, \langle \text{year}, 1999 \rangle \}, \\ &\quad \{ \langle \text{title}, \text{'The Avengers'} \rangle, \langle \text{year}, 2012 \rangle \}, \\ R_2 &= \{ \langle \text{title}, \text{'The Matrix'} \rangle, \langle \text{year}, 1999 \rangle \}, \\ &\quad \{ \langle \text{title}, \text{'The Avengers'} \rangle, \langle \text{year}, 1998 \rangle \}, \\ R_3 &= \{ \langle \text{title}, \text{'The Matrix'} \rangle, \langle \text{year}, 1999 \rangle \}, \\ &\quad \{ \langle \text{title}, \text{'The Avengers'} \rangle, \langle \text{year}, 1998 \rangle \}, \\ &\quad \{ \langle \text{title}, \text{'The Avengers'} \rangle, \langle \text{year}, 2012 \rangle \}. \end{aligned}$$

Nyní by mělo být jasné, že množinu R_3 můžeme dostat sjednocením množin R_1 a R_2 .

Získaný poznatek zobecníme. Pro dvě relace R_1 a R_2 nad relačním schématem S je **sjednocení** R_1 a R_2 množina

$$R_1 \cup R_2 = \{t \mid t \in R_1 \text{ nebo } t \in R_2\}.$$

Věta 1. Pro dvě relace R_1 a R_2 nad relačním schématem S je sjednocení $R_1 \cup R_2$ opět relace nad S .

Důkaz: Konečnost $R_1 \cup R_2$ plyne z konečnosti R_1 a R_2 . Dále musíme ukázat, že každý prvek $t \in R_1 \cup R_2$ je n -tice nad S . Vezměme libovolnou n -tici $t \in R_1 \cup R_2$. Víme, že $t \in R_1$ nebo $t \in R_2$. Předpokládejme, že $t \in R_1$. Protože R_1 je relace nad S musí být $t \in R_1$ n -ticí nad S . Podobně pro $t \in R_2$. \square

Pokud je R_1 určeno vlastností $V_1(t)$ a R_2 určeno vlastností $V_2(t)$, pak je $R_1 \cup R_2$ určeno vlastností „ $V_1(t)$ nebo $V_2(t)$ “.

Pro výše uvedený příklad dostáváme, že $R_3 = R_1 \cup R_2$ a vlastnost: „Film `title` vydaný v roce `year` vlastním já nebo ty.“ je jen úsporněji zapsaná vlastnost: „Vlastním film `title` vydaný v roce `year` nebo vlastníš film `title` vydaný v roce `year`.“

Každý relační výraz v SQL určuje pořadí atributů v jeho typu. Pořadí atributů relačního výrazu (`TABLE relation`) je shodné s pořadím, v jakém byly atributy základní proměnné `relation` deklarovány příkazem `CREATE TABLE`. Tedy například pořadí atributů `TABLE movie` je `title`, `year`. Zavedeme konvenci, že pořadí atributů můžeme určit pořadím, v jakém se vyskytují atributy v charakteristické vlastnosti, pořadím sloupců v tabulce nebo pořadím atributů v množinovém zápise schématu.

Předpokládejme, že `expression1` a `expression2` jsou relační výrazy nad S určující stejné pořadí atributů. Pak

```
( expression1 UNION expression2 )
```

je relační **výraz sjednocení** nad S zachovávající pořadí atributů.

Označme R_1 hodnotu výrazu `expression1` a R_2 hodnotu výrazu `expression2`. Pak hodnota výrazu sjednocení je $R_1 \cup R_2$. Charakteristickou vlastnost výrazu sjednocení můžeme odvodit z charakteristické vlastnosti operace sjednocení. Například pro relační proměnnou `movie1` danou vlastností: „Vlastním film `title` vydaný v roce `year`.“ s hodnotou:

title	year
The Matrix	1999
The Avengers	2012

a pro relační proměnnou `movie2` danou vlastností: „Vlastníš film `title` vydaný v roce `year`.“ s hodnotou:

title	year
The Matrix	1999
The Avengers	1998

je hodnota výrazu:

```
( TABLE movie1 ) UNION ( TABLE movie2 )
```

rovna:

title	year
The Matrix	1999
The Avengers	1998
The Avengers	2012

Charakteristickou vlastnost výrazu již známe: „Film `title` vydaný v roce `year` vlastním já nebo ty.“

3 Průnik a rozdíl

Podobně, jako jsme zavedli operaci a výraz sjednocení, zavedeme operace a výrazy průniku a rozdílu.

Opět si nejprve vezmeme dvě relace R_1 a R_2 nad schématem S , kde R_1 je dána vlastností $V_1(t)$ a R_2 vlastností $V_2(t)$. Pak **průnik** R_1 a R_2 je množina

$$R_1 \cap R_2 = \{t \mid t \in R_1 \text{ a } t \in R_2\}$$

a **rozdíl** R_1 a R_2 je množina

$$R_1 - R_2 = \{t \mid t \in R_1 \text{ a } t \notin R_2\}.$$

Věta 2. *Pokud je R relace nad S , pak libovolná její podmnožina je opět relace nad S .*

Důkaz: Necht $R' \subseteq R$. Podmnožina konečné množiny musí být konečná. Tedy R' je konečná množina. Zvolme libovolné $t \in R'$. Protože $R' \subseteq R$, platí $t \in R$. Tedy t je n -tice nad S . Ukázali jsme, že každý prvek R' je n -tice nad S . \square

Věta 3. *Množina $R_1 \cap R_2$ je relace nad S .*

Důkaz: Tvrzení plyne z Věty 2 a $R_1 \cap R_2 \subseteq R_1$ (nebo také z $R_1 \cap R_2 \subseteq R_2$). \square

Co je výsledkem průniku dvou relací nad různými schématy?

Věta 4. *Množina $R_1 - R_2$ je relace nad S .*

Důkaz: Opět plyne z Věty 2 a $R_1 - R_2 \subseteq R_1$. \square

Věta 5. Pro množiny A, B platí $A - (A - B) = A \cap B$.

Důkaz: Zvolme libovolný x . Máme $x \in A - (A - B)$, právě když $x \in A$ a $x \notin A - B$, právě když $x \in A$ a $(x \notin A$ nebo $x \in B)$, právě když $(x \in A$ a $x \notin A)$ nebo $(x \in A$ a $x \in B)$, právě když $x \in A$ a $x \in B$, právě když $x \in A \cap B$. \square

Vidíme, že průnik relací můžeme vyjádřit pomocí rozdílu.

Relace $R_1 \cap R_2$ je určena vlastností: „ $V_1(t)$ a $V_2(t)$ “ a relace $R_1 - R_2$ je určena vlastností: „ $V_1(t)$ a ne $V_2(t)$ “.

Podobně jako u sjednocení je pro relační výrazy *expression1* a *expression2* nad S , které určují stejné pořadí atributů,

```
( expression1 INTERSECTION expression2 )
```

relační **výraz průniku** a

```
( expression1 EXCEPT expression2 )
```

relační **výraz rozdílu**. Oba výrazy jsou nad S a zachovávají pořadí atributů. Pokud R_1 je hodnota výrazu *expression1* a R_2 je hodnota výrazu *expression2*, pak hodnota výrazu průniku je $R_1 \cap R_2$ a hodnota výrazu rozdílu je $R_1 - R_2$.

Například pro výše uvedené proměnné *movie1* a *movie2* je hodnota výrazu:

```
( TABLE movie1 ) INTERSECT ( TABLE movie2 )
```

relace:

title	year
The Matrix	1999

Charakteristická vlastnost výrazu je: „Vlastním film *title* vydaný v roce *year* a vlastní film *title* vydaný v roce *year*.“, což je zjednodušeně: „Film *title* vydaný v roce *year* vlastníme oba.“

Hodnota výrazu:

```
( TABLE movie1 ) EXCEPT ( TABLE movie2 )
```

je relace:

title	year
The Avengers	2012

Charakteristická vlastnost výrazu je: „Vlastním film `title` vydaný v roce `year` a nevlastní film `title` vydaný v roce `year`.“ Vlastnost můžeme zjednodušit na: „Film `title` vydaný v roce `year` vlastním jen já.“

Operace můžeme samozřejmě skládat. Například relaci danou vlastností: „Film `title` vydaný roku `year` vlastním jen já nebo jen ty.“ obdržíme jako hodnotu následujícího výrazu.

```
# ( ( TABLE movie1 ) EXCEPT ( TABLE movie2 ) )
      UNION
( ( TABLE movie2 ) EXCEPT ( TABLE movie1 ) );
```

title	year
The Avengers	2012
The Avengers	1998

(2 rows)

4 Podmínky

Je dáno schéma S . **Podmínka** θ nad S je definována následovně.

- jsou-li $y \in S$ a $v \in D_y$, pak $(y = v)$ a $(v = y)$ jsou podmínky nad S ,
- jsou-li $y_1, y_2 \in S$ a $D_{y_1} = D_{y_2}$, pak $(y_1 = y_2)$ je podmínka nad S ,
- jsou-li θ_1 a θ_2 podmínky nad S , pak $(\theta_1 \wedge \theta_2)$, $(\theta_1 \vee \theta_2)$ a $\neg\theta_1$ jsou podmínky nad S .

Například $\theta = ((\text{year} = 1999) \wedge \neg(\text{title} = \text{'The Matrix'}))$ je podmínka nad $S = \{\text{title}, \text{year}\}$.

Opět nejvíce vnější závorky kolem podmínek budeme vynechávat.

Řekneme, že n -tice t nad S **splňuje** podmínku θ nad S , jestliže:

- θ je tvaru $(y = v)$ nebo $(v = y)$ a $t(y) = v$,
- θ je tvaru $(y_1 = y_2)$ a $t(y_1) = t(y_2)$,
- θ je tvaru $(\theta_1 \wedge \theta_2)$ a t splňuje θ_1 a θ_2 ,
- θ je tvaru $(\theta_1 \vee \theta_2)$ a t splňuje θ_1 nebo θ_2 ,
- θ je tvaru $\neg\theta_1$ a t nespĺňuje podmínku θ_1 .

Například uvažujme n -tice:

$$\begin{aligned}r_1 &= \{\langle \text{title}, \text{'American Beauty'} \rangle, \langle \text{year}, 1999 \rangle\}, \\r_2 &= \{\langle \text{title}, \text{'The Matrix'} \rangle, \langle \text{year}, 1999 \rangle\}, \\r_3 &= \{\langle \text{title}, \text{'Titanic'} \rangle, \langle \text{year}, 1997 \rangle\}\end{aligned}$$

nad $S = \{\text{title}, \text{year}\}$ a podmínku $\theta = ((\text{year} = 1999) \wedge \neg(\text{title} = \text{'The Matrix'}))$. Pak r_1 splňuje θ , ale r_2 ani r_3 nesplňují θ .

Podmínce θ můžeme **přřadit výrokovou formu** $V(t)$ rovnou: „ t splňuje podmínku θ “. Například výroková forma podmínky $(\text{year} = 1999) \wedge \neg(\text{title} = \text{'The Matrix'})$ je: „Rok year je roven 1999 a název title je různý od The Matrix.“

V SQL zapisujeme podmínky podle následujících pravidel.

- Podmínku $(y = v)$ zapíšeme $(y = v)$,
- podmínku $(v = y)$ zapíšeme $(v = y)$,
- podmínku $(y_1 = y_2)$ zapíšeme $(y_1 = y_2)$,
- podmínku $(\theta_1 \wedge \theta_2)$ zapíšeme $(\theta_1 \text{ AND } \theta_2)$,
- podmínku $(\theta_1 \vee \theta_2)$ zapíšeme $(\theta_1 \text{ OR } \theta_2)$
- a podmínku $\neg\theta_1$ zapíšeme $(\text{NOT } \theta_1)$.

Například podmínku $(\text{year} = 1999) \wedge \neg(\text{title} = \text{'The Matrix'})$ zapíšeme:

```
( year = 1999 ) AND ( NOT title = 'The Matrix' )
```

5 Restrikce

Pro motivaci uvažujme relaci R :

title	year
American Beauty	1999
The Matrix	1999
Titanic	1997

danou: „Film title vyšel roku year .“ Chceme získat relaci danou: „Film title různý od filmu The Matrix vyšel roku year rovnému 1999.“, což je relace:

title	year
American Beauty	1999

Vezměme relaci R a podmínku θ nad S , pak množinu

$$\sigma_{\theta}(R) = \{t \in R \mid t \text{ splňuje podmínku } \theta\}$$

nazveme **restrikcí** R podle θ . Jistě $\sigma_{\theta}(R) \subseteq R$ a proto podle Věty 2 je $\sigma_{\theta}(R)$ relací na S .

Pokud $V_1(t)$ je charakteristická vlastnost relace R a $V_2(t)$ výroková forma podmínky θ , pak charakteristická vlastnost $\sigma_{\theta}(R)$ je „ $V_1(t)$ a $V_2(t)$.“

Například pro motivační příklad, kde

$$\begin{aligned} S &= \{\text{title, year}\}, \\ R &= \{\{\langle \text{title, 'American Beauty'} \rangle, \langle \text{year, 1999} \rangle\}, \\ &\quad \{\langle \text{title, 'The Matrix'} \rangle, \langle \text{year, 1999} \rangle\}, \\ &\quad \{\langle \text{title, 'Titanic'} \rangle, \langle \text{year, 1997} \rangle\}\}, \\ \theta &= (\text{year} = 1999) \wedge \neg(\text{title} = \text{'The Matrix'}) \end{aligned}$$

dostáváme, že

$$\begin{aligned} \sigma_{\theta}(R) &= \{t \in R \mid t \text{ splňuje } \theta\} \\ &= \{\{\langle \text{title, 'American Beauty'} \rangle, \langle \text{year, 1999} \rangle\}\}. \end{aligned}$$

Charakteristická vlastnost $\sigma_{\theta}(R)$ je: „Film `title` různý od filmu `The Matrix` vyšel roku `year` rovnému 1999.“

Výraz restrikce v SQL je relační výraz:

```
( SELECT *
  FROM  expression
 WHERE  $\theta$  )
```

nad S , kde *expression* je relační výraz nad S a θ podmínka nad S . Výraz zachovává pořadí atributů v *expression*. Hodnotou výrazu restrikce je $\sigma_{\theta}(R)$, kde R je hodnota výrazu *expression*.

Například uvažujme relační proměnnou `movie` danou vlastností: „Film `title` vyšel roku `year`.“ s hodnotou:

title	year
American Beauty	1999
The Matrix	1999
Titanic	1997

Relační výraz:


```
SELECT *
FROM ( TABLE movie )
WHERE ( year = 1999 )
AND ( NOT ( title = 'The Matrix' ) )
```

má charakteristickou vlastnost: „Film `title` různý od filmu The Matrix vyšel roku `year` rovnému 1999.“ a hodnotu:

title	year
American Beauty	1999

Následující věta ukazuje vztah složených podmínek a operací.

Věta 6. *Pro relaci R na schématu S a podmínky θ_1 a θ_2 nad S platí:*

1. $\sigma_{\theta_1 \wedge \theta_2}(R) = \sigma_{\theta_1}(R) \cap \sigma_{\theta_2}(R)$,
2. $\sigma_{\theta_1 \vee \theta_2}(R) = \sigma_{\theta_1}(R) \cup \sigma_{\theta_2}(R)$,
3. $\sigma_{\neg \theta_1}(R) = R - \sigma_{\theta_1}(R)$.

Důkaz: Zvolme libovolnou n -tici t na S .

1. $t \in \sigma_{\theta_1 \wedge \theta_2}(R)$, právě když $t \in R$ a t splňuje $\theta_1 \wedge \theta_2$, právě když $t \in R$ a t splňuje θ_1 a t splňuje θ_2 , právě když $t \in \sigma_{\theta_1}(R)$ a $t \in \sigma_{\theta_2}(R)$, právě když $t \in \sigma_{\theta_1}(R) \cap \sigma_{\theta_2}(R)$;
2. podobně $t \in \sigma_{\theta_1 \vee \theta_2}(R)$, právě když $t \in R$ a (t splňuje θ_1 nebo t splňuje θ_2), právě když $t \in \sigma_{\theta_1}(R) \cup \sigma_{\theta_2}(R)$;
3. nakonec $t \in \sigma_{\neg \theta_1}(R)$, právě když $t \in R$ a t nesplňuje θ_1 , právě když $t \in R - \sigma_{\theta_1}(R)$.

□

6 Mazání n -tic

Uvažujme relační proměnnou *relation* a podmínku θ nad S . Příkaz:

```
DELETE FROM relation WHERE  $\theta$ ;
```

nastaví hodnotu proměnné *relation* na $\{t \in R \mid t \text{ nesplňuje podmínku } \theta\}$, kde R je původní hodnota proměnné *relation*. Například vezměme proměnnou `movie` s hodnotou:

title	year
The Avengers	1998
The Avengers	2012
The Matrix	1999
American Beauty	1999

Pak příkaz:

```
DELETE FROM movie WHERE title = 'The Avengers';
```

změní hodnotu proměnné `movie` na:

title	year
The Matrix	1999
American Beauty	1999

Otázky a úkoly na cvičení

1. Uvažujme základní relaci `actor1`:

name	born
Kyle Merritt McLachlan	1959
Ian Murray McKellen	1939
Catherine Elise Blanchett	1969

určenou vlastností „Zajímám se o herce `name`, který se narodil roku `born`.“ a základní relaci `actor2`:

name	born
Catherine Elise Blanchett	1969
Dame Margaret Natalie Smith	1934
Kristen Noël „Kristy“ Swanson	1969

určenou vlastností „Vlastním film, kde hraje herec `name`, který se narodil roku `born`.“ Napište relační výrazy určené následujícími vlastnostmi.

- „Zajímám se o herce `name`, který se narodil roku `born`, jenž hraje v mnou vlastněném filmu.“
- „Herec `name` hraje ve filmu, který vlastním a narodil se roku `born` rovnému 1969.“
- „Herec `name` rovný 'Kyle Merritt McLachlan' se narodil roku `born` a zajímám se o něj.“
- „Vlastním film, kde hraje herec `name` rovný 'Catherine Elise Blanchett' narozený roku `born` rovnému 1970.“

- (e) „Zajímám se o herce **name**, který se narodil roku **born** rovnému 1959 nebo 1939.“
- (f) „Zajímám se o herce **name** narozeného roku **born**, který hraje ve filmu, jenž nevlastním.“
- (g) „Herec **name**, o kterého se zajímám nebo vlastním film, kde hraje, se narodil roku **born** rovnému 1969.“
2. Určete hodnoty relačních výrazů z prvního úkolu.
3. Zapište výpočet hodnoty relačního výrazu pomocí relačních operací u každého výrazu z prvního úkolu. Předpokládejte, že R_1 je hodnota proměnné `actor1` a R_2 je hodnota proměnné `actor2`.
4. Uvažujme databázi našich oblíbených režisérů a vybraných filmů tvořenou základní relací `movie_director`:

name	title
David Lynch	The Elephant Man
David Lynch	Blue Velvet
Federico Fellini	La Dolce Vita

určenou vlastností „Režisér **name** vytvořil film **title**.“ a základní relaci `favorite_movie`:

name	title
David Lynch	La Dolce Vita
Federico Fellini	Rashomon
Federico Fellini	La Dolce Vita

určenou vlastností „Film **title** je oblíbený film režiséra **name**.“ Napište charakteristické vlastnosti následujících relačních výrazů SQL.

(a)

```
( TABLE favorite_movie ) EXCEPT ( TABLE movie_director )
```

(b)

```
SELECT *
FROM ( TABLE movie_director )
WHERE title = 'Blue Velvet'
```

(c)

```
SELECT *
FROM ( TABLE favorite_movie )
WHERE name = 'David Lynch'
AND ( NOT title = 'La Dolce Vita' )
```

(d)

```
( SELECT *
  FROM ( TABLE favorite_movie )
  WHERE name = 'David Lynch' )
UNION
( SELECT *
  FROM ( TABLE movie_director )
  WHERE name = 'David Lynch' )
```

(e)

```
SELECT *
FROM ( ( TABLE favorite_movie )
      UNION
      ( TABLE movie_director ) )
WHERE title = 'La Dolce Vita'
```

5. Určete hodnoty relačních výrazů ze čtvrtého úkolu.
6. Přepište relační výrazy ze čtvrtého úkolu pomocí relačních operací. Hodnoty použitých relačních proměnných si vhodně označte.
7. Musí být množinové sjednocení dvou relací vždy relace?
8. Co vznikne množinovým rozdílem dvou relací nad různými schématy?
9. Je pro relaci R nad S množina $\text{Tupl}(S) - R$ vždy relace nad S ?
10. Jak upravit základní relaci `actor1`, když se přestaneme zajímat o herce Kyle Merritt McLachlan?
11. Co, když zjistíme, že v roce narození herce je chyba. Dovedeme ji opravit?