УДК 519.689

Д.Б. Буй, С.А. Поляков

КОМПОЗИЦІЙНІ СТРУКТУРИ SQL – ПОДІБНИХ МОВ: ФІЛЬТРАЦІЯ ТА ПОВНИЙ ОБРАЗ

Робота розглядає завдання формальної семантики мов типу SQL [1,2] методами композиційного програмування [3] та продовжує [4].

Введемо множину доменів $Dom_1, Dom_2, ..., Dom_n$, елементи яких назвемо атомарними даними і позначимо $d_1, d_2, ...$; об'єднання всіх доменів назвемо універсальним доменом і позначимо DOM.

1. Кортежом називається скінченна послідовність атомарних даних. Кортежі будемо записувати у вигляді $< d_{i_1}, d_{i_2}, ..., d_{i_k} >$, де $d_{i_j} \in Dom_j$. Множину всіх кортежів позначимо через Tp, а окремі кортежі - через tp. Під схемою кортежа $< d_{i_1}, d_{i_2}, ..., d_{i_k} >$ будемо розуміти послідовність $< D_{i_1}, D_{i_2}, ..., D_{i_k} >$, таку, що $d_{i_j} \in D_{i_{i_j}}$. Розглянемо основні операції на кортежах.

2. Конструктор кортежів $< d_1, d_2, ..., d_n >, n \geq 1$: параметризована за n функція типу $DOM \times ... \times DOM \to TP$, яка формує з аргументів кортеж; конкатенація кортежів $tp_1 \| tp_2$: бінарна функція типу $TP \times TP \to TP$.

3. Введемо поняття рядка. Нехай задана множина імен V, елементи якої позначимо символами v_i. Функцією іменування кортежів назвемо бієкцією множини $\{v_1, v_2, ..., v_k\} \subseteq V$ в початковий відрізок натурального ряду. Такі функції домовимось позначати $\mu_{v_1,...,v_k}$; при цьому припускається, що v_1 відображається в 1, v_2 - в 2, і т.п. Рядком назвемо пару вигляду $(< d_1, ..., d_k >, \ \mu_{v_1,...,v_k})$. Множину всіх рядків позначимо як R, а окремі рядки як r_i. Схемою рядка (tp, μ) назвемо пару вигляду $(sch(tp), \mu)$, де $sch(tp)$ - схема кортежа tp. Розглянемо деякі операції на рядках.

4. Іменування $v \leftarrow (d)$: унарна функція типу $DOM \rightarrow R$; її значення буде рядок $(<d>, \mu_v)$. З'єднання рядків $r_1 \|_R r_2$: бінарна часткова функція типу $R \times R \rightarrow R$. Покладемо за означенням $\mu_{V_1', \ldots, V_m'} \circ \mu_{V_1'', \ldots, V_k''} = \mu_{V_1', \ldots, V_m', V_1'', \ldots, V_k''}$, якщо $\{V_1', \ldots, V_m'\} \cap \{V_1'', \ldots, V_k''\} = \varnothing$, і невизначено в протилежному випадку. Тоді $(tp_1, \mu_1) \|_R (tp_2, \mu_2) = (tp_1 \| tp_2, \mu_1 \circ \mu_2)$, якщо значення $\mu_1 \circ \mu_2$ визначене, та невизначене в іншому випадку.

5. Введемо поняття таблиці. Мультимножиною назвемо множину з дублікатами; їх будемо записувати у вигляді $\{a_1^{i_1}, \ldots, a_k^{i_k}\}$, де a_j - елементи мультимножини, i_j - кількість дублікатів даного елемента. Таблицею назвемо скінченну мультимножину рядків, для яких існує одна схема. Множину всіх таблиць позначимо через T, а окремі таблиці- через t . Схемою таблиці назвемо схему, спільну для всіх рядків. Оскільки схема зіставляється неоднозначно, то й схема таблиці також зіставляється неоднозначно.

Введемо наступні позначення. Нехай $d^n \in S$ означає, що мультимножина S містить рівно n дублікатів елемента d, а запис $d^0 \in S$ - що мультимножина не містить жодного дубліката елемента d.

Для таблиць запис $tp^n Et$ означає, що таблиця містить рівно n дублікатів рядка (tp, μ), де μ - функція іменування кортежів таблиці t; запис $tp^0 Et$ означає, що таблиця не містить жодного дубліката вказаного рядка.

Таблиці назвемо сумісними за об'єднанням, якщо існують їхні схеми, перші компоненти яких збігаються (тобто збігаються схеми кортежів, які містяться в рядках). Розглянемо операції об'єднання таблиць.

6. Об'єднання таблиць $t_1 \cup_1 t_2, t_1 \cup_{ALL} t_2$ - бінарні часткові функції типу $T \times T \rightarrow T$. Нехай μ - функція іменування кортежів, яка належить схемі таблиці t_1. Тоді для сумісних за об'єднанням таблиць функції визначаються таким чином:

$$t_1 \cup_1 t_2 = \{r^1 | r = (tp, \mu) \& (tpEt_1 \vee tpEt_2)\},$$

$$t_1 \cup_{ALL} t_2 = \{r^{i+j} | r = (tp, \mu) \& (tp^i Et_1 \vee tp^j Et_2)\}.$$

7. Композиціями виступають спеціальні операції над функціями та предикатами. Аргументи та значення функцій належать множинам DOM, R та T; аргументи предикатів також

належать цим множинам, а значення- множині $\{True, False, \omega\}$, де ω означає невизначене значення. Нехай $S_1, S_2, \ldots \in \{DOM, R, T\}$. Розглянемо наступні композиції.

Композиція фільтрації Fl зіставляє предикату $p^{(n+2)}$ функцію $f^{(n+1)}$, $n \geq 0$; де $p^{(n+2)}$: $S_1 \times \ldots \times S_n \times T \times R \to \{True, False, \omega\}$ та $f^{(n+1)}$: $S_1 \times \ldots \times S_n \times T \to T$; значення нової функції задається рівністю

$$f^{(n+1)}(s_1, \ldots, s_n, t) = \{r^i \mid p^{(n+2)}(s_1, \ldots, s_n, t, r) = True \& r^i \in t\}.$$

Композиція взяття повного образу Im зіставляє функції $f^{(n+2)}$ функцію $g^{(n+1)}$, $n \geq 0$; де $f^{(n+2)}$: $S_1 \times \ldots \times S_n \times T \times R \to R$ та $g^{(n+1)}$: $S_1 \times \ldots \times S_n \times T \to T$; значення нової функції задається рівністю

$$g^{(n+1)}(s_1, \ldots, s_n, t) = \cup_{ALL} \{f^{(n+2)}(s_1, \ldots, s_n, t, r)^i\}.$$

Введені композиції фільтрації, повного образу і стандартна суперпозиція, структури даних та певні природні функції і предикати дають змогу задати семантику досить потужного фрагмента мови SQL; фрагмент містить оператори SELECT, в яких не використовуються операції зовнішнього з'єднання, агрегатні функції, конструкції групування (GROUP BY та HAVING), впорядкування (ORDER BY), а також розгалуження (вирази типу CASE). Функції, що виступають семантиками операторів цього фрагмента, будуються з вихідних функцій вказаними композиціями.

1. Дейт К. Руководство по реляционной СУБД DB2. - М.: Финансы и статистика, 1988. - 320 с. 2. Грабер М. Справочно руководство по SQL. - М.: ЛОРИ, 1997. -292 с. 3. Редько В.Н. Основания композиционного программирования // Программирование. - 1979. -№5. -С.3-13. 4. Поляков С.А. Композиційна семантика SQL-подібних мов. // Вісн. Київ. ун-ту. Сер. фіз.-мат. науки. -1997. -Вип.3. -С.205-211.