УДК 004.655

Буй Д. Б., Глушко І. М.

УЗАГАЛЬНЕНЕ ЧИСЛЕННЯ РЯДКІВ

У статті подано узагальнення класичного результату про еквівалентність реляційної алгебри Кодда та числення рядків (числення кортежів). Класичне числення рядків поповнено довільними предикатними та функціональними сигнатурами на універсальному домені. Доведено, що при цьому числення рядків залишається не менш виразним, ніж таблична алгебра.

Ключові слова: реляційні бази даних, числення кортежів, реляційна алгебра.

В основі більшості реляційних мов запитів лежить реляційне числення, оскільки на відміну від реляційної алгебри, числення визначає лише, яким повинен бути результат і не передбачає визначення того, як його отримати. Реляційне числення ґрунтується на численні предикатів першого порядку. Є дві форми реляційного числення: числення зі змінними-рядками та числення зі змінними на доменах. Ці форми запропоновано Е. Коддом (E. Codd) [1] та М. Лакруа (M. Lacroix) і А. Піротте (A. Pirotte) [2] відповідно. Це питання також розглядали у своїх роботах Д. Мейєр (D. Maier) [3], Дж. Д. Ульман (J. D. Ullman) [4], К. Дж. Дейт (C. J. Date) [5], Т. М. Коноллі (T. M. Connolly) та К. Беґґ (C. E. Begg) [6], В. В. Пасічник та В. А. Резніченко [7]. Ми зосередили увагу на реляційному численні рядків.

Уточнення реляції, а також усі поняття та позначення ми розуміємо в смислі монографії [8]. Розглядаємо дві множини: $\boldsymbol{A}$ – множину атрибутів і $\boldsymbol{D}$ – універсальний домен. Під табличною алгеброю розуміємо алгебру $<T, \Omega_{P,\Xi}>$, де T – множина усіх таблиць, $\Omega_{P,\Xi}$ = $\left\{\cup_R, \cap_R, \backslash_R, \sigma_p, \pi_X, \otimes, \div^{R_1}_{R_2}, Rt_\xi, \sim\right\}^{p\in P, \xi\in\Xi}_{X,R,R_1,R_2\subseteq A}$ – сигнатура, P, Ξ – множини параметрів. Виразом табличної алгебри називається будь-який вираз, побудований з таблиць множини T при використанні операцій з множини $\Omega_{P,\Xi}$.

Вирази числення рядків мають вигляд $\{x\,(R) \mid P(x)\}$, де P – деякий предикат над змінним рядком x. Цей вираз позначає таблицю t схеми R, $t\in T(R)$, що містить рядки, на яких предикат P істинний.

Введемо множину так званих дозволених формул числення рядків при використанні:

- множини атрибутів $\boldsymbol{A}$ і універсального домену $\boldsymbol{D}$;
- множини предметних змінних (змінних рядків) $x_1, x_2, \ldots$;
- множини предметних констант d_1, $d_2, \ldots, d_i, \ldots \in \boldsymbol{D}$;
- множини функціональних символів $f_1, f_2, \ldots$;
- множини предикатних символів $p_1, p_2, \ldots$.

Наступні вирази є термами (індукція за довжиною термів):

a) будь-яка предметна константа є терм;
b) $x(A)$ – терм, де x – предметна змінна, A – атрибут з множини $\boldsymbol{A}$;
c) якщо $t_1, \ldots, t_n$ – терми і f – n-арний функціональний символ, то $f(t_1, \ldots, t_n)$ – терм;
d) вираз є термом тільки тоді, коли це слідує з правил a), b), c).

Сформулюємо правила побудови формул. Атомарні формули (атоми) можуть бути двох типів.

a1. Нехай t – таблиця, а x – змінний рядок. Тоді $t(x)$ – атом, який означає, що $x \in t$.
a2. Нехай $t_1, \ldots, t_n$ – терми, а p – n-арний предикат на універсальному домені $\boldsymbol{D}$. Тоді $p(t_1, \ldots, t_n)$ – атом.

Використаємо логічні зв'язки $\neg$, $\wedge$, $\vee$, квантори $\exists$, $\forall$ та дужки () для побудови формул із атомів.

f1. Кожний атом – це формула.
f2. Якщо P – формула, то $\neg P$ – теж формула.
f3. Якщо P і Q – формули, то $P \wedge Q$, $P \vee Q$ – формули.
f4. Нехай x – змінний рядок, P – формула, $R \subseteq \boldsymbol{A}$ – схема, тоді $\exists x(R)P$ – формула.
f5. Нехай x – змінний рядок, P – формула, $R \subseteq \boldsymbol{A}$ – схема, тоді $\forall x(R)P$ – формула.
f6. Якщо P – формула, то і (P) – формула.
f7. Інших формул немає.

Зазвичай змінні рядки у формулах можуть бути вільні або зв'язані. Смисл цих понять такий самий, як і у численні предикатів: входження змінної x у формулу називається зв'язаним, якщо x є змінною квантора ($\exists$ або $\forall$), який входить у цю формулу або знаходиться в області дії даного квантора; в протилежному випадку входження змінної x у формулу називається вільним. Змінна називається вільною (зв'язаною) у формулі, якщо існують вільні (відповідно зв'язані) її входження в цю формулу (див., наприклад, [9, с. 55–56]).

© *Буй Д. Б., Глушко І. М.*, 2010

Для кожного змінного рядка x визначено схему $scheme(x, P)$ та множину атрибутів $attr(x, P)$, з якими рядок x зустрічається у формулах. Вирази $scheme(x, P)$ та $attr(x, P)$ визначені, якщо рядок x має вільне входження в формулу P, причому має місце включення $attr(x, P) \subseteq scheme(x, P)$ (що випливає з подальших означень, за умови визначення виразів).

Виділимо клас дозволених формул, використовуючи поняття вільних та зв'язаних змінних рядків, схеми та множини атрибутів, з якими змінний рядок зустрічається у формулах. Визначимо вирази *scheme* та *attr* спочатку для термів:

1. якщо $t = d$, $d \in \boldsymbol{D}$, то схема $scheme(x, t)$ не визначена, а $attr(x, t) = \varnothing$;
2. якщо $t = x(A)$, то, схема $scheme(x, t)$ не визначена, а $attr(x, t) = \{A\}$;
3. якщо $t = f(t_1, \ldots, t_n)$, де t_i – терми, то схема $scheme(x, t)$ не визначена, а $attr(x, t) = \bigcup_{i=1}^{n} attr(x, t_i)$.

Нехай формула P – атом, тоді

a1. якщо $P = t(x)$, то (єдине) входження змінного рядка x є вільним у формулі P, $scheme(x, P) = attr(x, P) = R$, де R – схема таблиці t;

a2. якщо $P = p(t_1, \ldots, t_n)$, де t_i – терми, причому $x_1, \ldots, x_m$ – всі змінні цих термів, то входження цих змінних рядків є вільними у формулі P, схема $scheme(x_i, P)$ не визначена, а $attr(x_i, P) = \bigcup_{j=1}^{n} attr(x_i, t_j)$, $i = 1,\ldots,m$.

Атомарні формули завжди дозволені. Для побудови всіх дозволених формул проведемо індукцію за довжиною формул. Припустимо, що G і Q – дозволені формули.

f2. Якщо $P = \neg G$, то P – дозволена формула, а входження змінних у P – вільні або зв'язані, залежно від того, вільні або зв'язані входження цих самих змінних у G. Якщо x входить у формулу G вільно, то $scheme(x, P) \simeq scheme(x, G)$ і $attr(x, P) = attr(x, G)$. Вище $\simeq$ – узагальнена рівність (тобто обидві частини або одночасно не визначені, або одночасно визначені та рівні).

f3. Якщо $P = G \wedge Q$ або $P = G \vee Q$, то входження змінних у P – вільні або зв'язані, залежно від того, вільні або зв'язані входження цих змінних у G або Q. Нехай змінний рядок x входить у підформули G та/або Q вільно. Визначимо схему та множину атрибутів, з якими рядок x зустрічається у формулах, для формули P. Мають місце наступні випадки.

a. Схеми формул $scheme(x, G)$ та $scheme(x, Q)$ визначені. Для того, щоб формула P була дозволена, має виконуватися рівність $scheme(x, G) = scheme(x, Q)$. Покладаємо за означенням $scheme(x, P) = scheme(x, G)$.

b. Схема визначена лише для однієї з підформул. Нехай схема $scheme(x, G)$ визначена, а схема $scheme(x, Q)$ – не визначена. Для того, щоб формула P була дозволена, має виконуватися включення $attr(x, Q) \subseteq scheme(x, G)$. Покладаємо за означенням $scheme(x, P) = scheme(x, G)$.

c. Схема не визначена для обох підформул. У цьому випадку і схема $scheme(x, P)$ не визначена.

Для будь-якого з розглянутих випадків $attr(x, P) = attr(x, G) \cup attr(x, Q)$.

f4. Якщо $P = \exists x(R)G$ і змінна x входить у формулу G вільно, то P – дозволена формула. Крім того, якщо $scheme(x, G)$ визначена, то повинна виконуватися рівність $scheme(x, G) = R$ і включення $attr(x, G) \subseteq R$. Оскільки змінна x не входить вільно у формулу P, то $scheme(x, P)$ і $attr(x, P)$ не визначені. Якщо $y \neq x$, то будь-яке входження змінної y в P вільне або зв'язане, залежно від того, вільне або зв'язане входження y в G. Якщо y входить в P вільно, то $scheme(y, P) \simeq scheme(y, G)$ і $attr(y, P) = attr(y, G)$.

f5. Якщо $P = \forall x(R)\ G$, то всі визначення й обмеження такі ж, як і у випадку f4 для квантора існування.

f6. Якщо $P = (G)$, то P дозволена формула, а вільні та зв'язані входження змінних, схема та множина атрибутів, з якими змінний рядок зустрічається у формулах, залишаються такими, як і для G.

Після введення множини дозволених формул можемо дати остаточне визначення виразу числення рядків. Отже, вираз числення рядків має вигляд $\{x(R) \mid P(x)\}$, де

1. формула P – дозволена;
2. змінна x – єдина змінна, яка входить у формулу P вільно;
3. якщо $scheme(x, P)$ визначена, то $scheme(x, P) = R$, інакше $attr(x, P)\ R$.

Нехай $P(x)$ – дозволена формула, $R \subseteq \boldsymbol{A}$, і якщо $scheme(x, P)$ визначена, то $scheme(x, P) = R$, інакше $attr(x, P) \subseteq R$. У результаті підставлення конкретного рядка s схеми R замість x у формулу P отримаємо формулу $P(s/x)$. Надамо спочатку значення істинності атомам:

a1. нехай рядок x у підформулі $t(x)$ вільний у P. Атом $t(x)$ набуває значення істина при підставленні конкретного рядка s замість x, якщо $s \in t$, інакше атом $t(x)$ набуває значення хиба;

a2. нехай рядок x_i у підформулі $p(t_1,\ldots, t_n)$ вільний у P, тоді при підставленні конкретного рядка s_i замість x_i замінимо $x_i(A_i)$ на $d_i \in \boldsymbol{D}$, де $\langle A_i, d_i \rangle \in s_i$ (d_i – значення атрибута A_i в рядку s_i). Атом $p(t_1,\ldots, t_n)$, де t_i – терми (предметні константи), набуває значення істина, якщо предикат p істинний на відповідних значеннях x, інакше атом набуває значення хиба.

Набір значень істинності всіх атомів формули називають її інтерпретацією [10]. Нехай формула P – дозволена формула без вільних змінних. Інтерпретація формули P визначається наступним чином.

f2. Якщо $P = \neg G$, то в G немає вільних змінних. Формула P істинна, коли G хибна, і хибна, коли формула G істинна.

f3. Якщо $P = G \wedge Q$ або $P = G \vee Q$, то в G і Q немає вільних змінних. Якщо $P = G \wedge Q$, то формула P істинна, тоді коли G та Q водночас істинні, у всіх інших випадках формула P хибна. Якщо $P = G \vee Q$, то формула P хибна, тоді коли G та Q водночас хибні, у всіх інших випадках формула P істинна.

f4. Якщо $P = \exists x(R)G$, то x – єдина змінна, яка входить у формулу G вільно. Формула P істинна, якщо є принаймні один рядок схеми R $s \in S(R)$, такий, що формула $G(s/x)$ – істинна, інакше формула P хибна.

f5. Якщо $P = \forall x(R)\ G$, то x – єдина змінна, яка входить у формулу G вільно. Формула P істинна, якщо для кожного рядка схеми R $s \in S(R)$ формула $G(s/x)$ істинна, інакше формула P хибна.

f6. Якщо $P = (G)$, то формула P істинна, якщо формула G істинна та хибна, якщо формула G хибна.

Нехай $E = \{x\ (R) \mid P(x)\}$ – вираз числення рядків. Значенням виразу E назвемо таблицю схеми R, яка містить усі рядки $s \in S(R)$, такі, що формула $P(s/x)$ істинна.

Теорема. Якщо E – вираз табличної алгебри, то можна ефективно побудувати еквівалентний йому вираз F числення рядків.

Доведення. При доведенні теореми розглядаємо вирази табличної алгебри, які містять тільки операції об'єднання, перетину, різниці, селекції, проекції, з'єднання та перейменування, оскільки операції ділення та активного доповнення можна виразити через дані операції: $E_1 \div^{R_1}_{R_2} E_2 = \pi_{R'}(E_1) \setminus_{R'} \pi_{R'}(\pi_{R'}(E_1) \otimes E_2 \setminus R_1 E_1)$, де $R_2 \subseteq R_1$, $R' = R_1 \setminus R_2$, $\tilde{E}_1 = C(E_1) \setminus E_1$, де $C(E_1) = \pi_{A_1}(E_1) \otimes \ldots \otimes \pi_{A_n}(E_1)$, $R = \{A_1, \ldots A_n\}$ – схема таблиці, що є значенням виразу E_1 (див., наприклад, [3, с. 45; 8, с. 65, с. 83].

Доведення проведемо індукцією за числом операцій в E.

База індукції (немає операцій). Можливі два випадки. $E = t$, де t – таблиця зі схемою R, покладемо $F = \{x(R) \mid t(x)\}$. E – стала таблиця $t = \{s_1, \ldots, s_n\}$ схеми $\{A_1, \ldots A_m\}$, причому $s_i = \{<A_1, d_{i_1}>, \ldots, <A_m, d_{i_m}>\} = \bigcup_{j=1}^{m} \{<A_j, d_{i_j}>\}$. Покладемо $F = \{\ x(\{A_1, \ldots A_n\}) \mid \bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} x(A_j) = d_{i_j}\ \}$. □

Крок індукції. Припустимо, що твердження теореми виконується для будь-якого виразу табличної алгебри, який містить менше k операцій. Нехай вираз E містить k операцій.

Випадок 1 (об'єднання). $E = E_1 \cup_R E_2$. Тоді E_1 і E_2 мають менше k операцій, тому існують вирази числення рядків $\{x(R) \mid P(x)\}$ і $\{x(R) \mid Q(x)\}$, еквівалентні E_1 і E_2 відповідно. Покладемо F рівним $\{x(R) \mid P(x) \vee Q(x)\}$. □

Випадок 2 (різниця). $E = E_1 \setminus_R E_2$. Тоді, як і у випадку 1, існують вирази числення рядків $\{x(R) \mid P(x)\}$ і $\{x(R) \mid Q(x)\}$, еквівалентні E_1 і E_2 відповідно. Покладемо F рівним $\{\ x(R) \mid P(x) \wedge \neg Q(x)\}$. □

Випадок 3 (перетин). $E = E_1 \cap_R E_2$. Тоді існують вирази числення рядків $\{x(R) \mid P(x)\}$ і $\{x(R) \mid Q(x)\}$, еквівалентні E_1 і E_2 відповідно. Покладемо F рівним $\{\ x(R) \mid P(x) \wedge Q(x)\}$. □

Випадок 4 (селекція). $E = \sigma_{\tilde{p}}\ (E_1)$. Нехай $\{x(R) \mid P(x)\}$ – вираз числення рядків, еквівалентний E_1. Покладемо F рівним $\{x(R) \mid P(x) \wedge p(x(A_1), \ldots, x(A_n))\}$, де $R = \{A_1, \ldots A_n\}$ – схема таблиці, що є значенням виразу E_1. Тут припускається, що предикат-параметр селекції заданий так: $\tilde{p}\ (s) = T \Leftrightarrow p(s(A_1), \ldots, s(A_n)) = T$, $s \in S(R)$, де p – n-арний сигнатурний предикатний символ. □

Випадок 5 (проекція). $E = \pi_X\ (E_1)$. Нехай $\{x(R) \mid P(x)\}$ – вираз числення рядків, еквівалентний E_1. Покладемо $F = \{y(X) \mid \exists x(R)\ (P(x) \bigwedge_{A \in X} y(A) = x\ (A))\}$. □

Випадок 6 (з'єднання). $E = E_1 \otimes E_2$. Нехай $\{x(R_1) \mid P(x)\}$ і $\{y(R_2) \mid Q(y)\}$– вирази числення рядків, еквівалентні E_1 і E_2 відповідно. Покладемо F рівним $\{z(R_1 \cup R_2) \mid \exists x(R_1)\ \exists y(R_2)\ (P(x) \wedge Q(y) \wedge \bigwedge_{A \in R_1} z(A) = x(A) \wedge \bigwedge_{A \in R_2} z(A) = y(A))\}$. □

Випадок 7 (перейменування). $E = Rt_{\xi}(E_1)$, де ξ: $\boldsymbol{A} \tilde{\to} \boldsymbol{A}$ – ін'єктивна, що здійснює перейменування атрибутів. Тоді існує вираз числення рядків $\{x(R_1) \mid P(x)\}$, еквівалентний E_1. Покладемо F рівним $\{y(R_2) \mid \exists x(R_1)\ (P(x) \wedge \bigwedge_{C \in R_1 \setminus \mathrm{dom}\ \xi} y(C) = x(C) \wedge \bigwedge_{A_i \in R_1 \cap \mathrm{dom}\ \xi} x(A) = y(\xi(A))\}$, де $R_2 = R_1 \setminus \mathrm{dom}\ \xi$, $\cup\ \xi[R_1]$. □

Висновки

У роботі визначено синтаксис термів, атомів та формул числення рядків; виділено клас дозволених формул, використовуючи поняття вільних та зв'язаних змінних рядків, введено поняття схеми *scheme*(x, P) та множини атрибутів *attr*(x, P), з якими змінний рядок зустрічається у формулах. Класичне числення рядків поповнено довільними предикатними та функціональними сигнатурами на універсальному домені $\boldsymbol{D}$ (тоді як зазвичай розглядають лише бінарні предикати, а функціональна сигнатура взагалі порожня), при цьому числення рядків залишається не менш виразним, ніж таблична алгебра.

1. Codd E. F. Relational Completeness of Data Base Sublanguages / Codd E. F. // Data Base Systems. – New York : Prentice-Hall. – 1972. – P. 65–93.
2. Lacroix M. Domain-oriented relational languages / M. Lacroix, A. Pirotte // Proc. 3rd Int. Conf. on Very Large Data Bases. – Tokyo, October, 1977. – P. 370–378.
3. Мейєр Д. Теория реляционных баз данных / Д. Мейєр; пер. с англ. – М. : Мир, 1987. – 608 с.
4. Ульман Дж. Основы систем баз данных / Дж. Ульман; пер. с англ. – М. : Финансы и статистика, 1983. – 334 с.
5. Дейт К. Дж. Введение в системы баз данных : [8-е изд.: пер. с англ.] / К. Дж Дейт. – М. : Вильямс, 2005. – 1328 с.
6. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика : [3-изд.: пер. с англ.] / Томас Коннолли, Каролин Бегг. – М. : Вильямс, 2003. – 1440 с.
7. Пасічник В. В. Організація баз даних та знань / В. В. Пасічник, В. А. Резніченко. – К. : Видавнича група BHV, 2006. – 384 с.
8. Реляційні бази даних: табличні алгебри та SQL-подібні мови / В. Н. Редько, Ю. Й. Борона, Д. Б. Буй, С. А. Поляков. – К. : Академперіодика, 2001. – 198 с.
9. Мендельсон Э. Введение в математическую логику / Э. Мендельсон; пер. с англ. – М. : Наука, 1971. – 320 с .
10. Нікольський Ю. В. Дискретна математика : підручник : гриф МОН України / Ю. В. Нікольський, В. В. Пасічник, Ю. М. Щербина. – [3-тє вид.]. – Л. : Магнолія, 2008. – 608 с.
11. Цаленко М. Ш. Моделирование семантики в базах данных / М. Ш. Цаленко. – М. : Наука, 1989. – 287 с.

D. Buy, I. Glushko

GENERALIZATION OF TUPLE CALCULUS

In the article generalization of classic result about the equivalence of relational algebra of Codd and tuple calculus is considered. The classic tuple calculus is filled up arbitrary predicate and functional signatures on the universal domain. It is proved that the tuple calculus remains no less expressive than relation algebra.

Keywords: relation databases, tuple calculus, relation algebra.

УДК 004.655

Буй Д. Б., Кахута Н. Д., Сільвейструк Л. М.

ПОВНИЙ ОБРАЗ, ОБМЕЖЕННЯ, ПРОЕКЦІЯ, ВІДНОШЕННЯ СУМІСНОСТІ

Статтю присвячено дослідженню загальних властивостей теоретико-множинних конструкцій, які використовуються, зокрема, в теорії реляційних баз даних при дослідженні табличних алгебр, побудованих на основі класичних реляційних алгебр Кодда. Розглянуто повний образ множини відносно бінарного відношення та розповсюдження унарних (бінарних) часткових операцій на множини за допомогою повного образу, обмеження бінарного відношення за множиною, проекцію бінарного відношення та відношення сумісності бінарних відношень.

Ключові слова повний образ множини, обмеження бінарного відношення за множиною, проекція бінарного відношення, відношення сумісності.

Загальні зауваження

Зафіксуємо універсум D, елементи якого позначимо $x, y, z,\ldots$ Підмножини універсуму позначимо X, Y ..., бінарні відношення на D (тобто множини пар, компоненти яких належать універсуму) – U, V,...

Зазвичай бінарне відношення U називають *функціональним,* якщо для всіх елементів x, y, z виконується імплікація $\langle x, y\rangle \in U \,\&\, \langle x, z\rangle \in U \Rightarrow y = z$.

Функціональні бінарні відношення (за іншою термінологією *часткові функції*) позначимо f, g,... *Область означеності* (рос. – «область определенности» [1, гл. 1, § 1, с. 20] [1]) функції f

[1] Зауважимо, що в другому видання монографії А. І. Мальцева використовується термін «область определения» [2, гл. 1, § 1, с. 19].

© *Буй Д. Б., Кахута Н. Д., Сільвейструк Л. М.*, 2010