УДК 004.655

До питання про створення мов запитів: узагальнена таблична алгебра, узагальнені реляційні числення

І. М. Глушко
Київський національний університет імені Тараса Шевченка, Україна

В статті проведено узагальнення табличної алгебри, побудованої на основі реляційних алгебр Кодда; узагальнення полягає в тому, що замість таблиць розглядаються пари, які складаються з таблиць та схем таблиць. Числення рядків (кортежів) та числення на домені поповнені довільними предикатними та функціональними сигнатурами. Доведено, що узагальнена таблична алгебра, узагальнене числення рядків та узагальнене числення на домені еквівалентні.

Ключові слова: *реляційні (табличні) бази даних, числення рядків, числення на домені, таблична алгебра.*

В статье проведено обобщение табличной алгебры, построенной на основе реляционных алгебр Кодда; обобщение состоит в том, что вместо таблиц рассматриваются пары, состоящие из таблиц и схем таблиц. Исчисление строк (кортежей) и исчисление на домене пополнены произвольными предикатными и функциональными сигнатурами. Доказано, что обобщенная табличная алгебра, обобщенное исчисление строк и обобщенное исчисление на домене эквивалентны.

Ключевые слова: *реляционные (табличные) базы данных, исчисление строк, исчисление на домене, табличная алгебра.*

In the paper generalization of table algebra built on the basis of Codd's relational algebras is conducted. Generalization is that instead of tables we consider pairs consisting of tables and scheme of tables. The classic tuple calculus and domain calculus are filled up arbitrary predicate and functional signatures on the universal domain. It is proved the equivalence between table algebra, tuple calculus and domain calculus.

Key words: *relation databases, tuple calculus, domain calculus, table algebra.*

1. Вступ

Уточнення реляції в термінах іменних множин було здійснене у монографії [1]. У даній роботі отримані результати узагальнено на таблиці, яким приписується певна схема. Класичне числення рядків та числення на домені поповнені довільними предикатними та функціональними сигнатурами на універсальному домені (у той час як, зазвичай, розглядають лише бінарні предикати, а функціональна сигнатура взагалі порожня, див., наприклад, [2]). Доведено, що узагальнена таблична алгебра, узагальнене числення рядків та узагальнене числення на домені еквівалентні.

2. Узагальнення табличної алгебри

Всі невизначені тут поняття та позначення розуміємо в смислі монографії [1]. Розглядаємо дві множини: $\boldsymbol{A}$ – множину атрибутів і $\boldsymbol{D}$ – універсальний домен. Довільну (скінченну) множину атрибутів $R \subseteq \boldsymbol{A}$ назвемо схемою. Рядком схеми R називається іменна множина на парі R, $\boldsymbol{D}$, проекція якої за першою компонентою рівна R, тобто рядок схеми R – це функція вигляду $s : R \to \boldsymbol{D}$.

Під (узагальненою) таблицею розуміємо пару $\langle t, R\rangle$, де $t \in T(R)$ – таблиця фіксованої схеми R. Тоді $\boldsymbol{T}(R) = \{\langle t, R\rangle \mid t \in T(R)\}$ – множина усіх (узагальнених) таблиць схеми R, а $\boldsymbol{T} = \bigcup_R \boldsymbol{T}(R)$ – множина всіх таблиць. Отже, кожній таблиці приписується певна схема на відміну від визначення таблиці, яке подане у [1]. Це впливає тільки на випадок порожньої таблиці $t_\varnothing$, оскільки за непорожньою таблицею схема відновлюється однозначно. Запис $\langle t_\varnothing, R\rangle$ позначає порожню таблицю схеми R.

Під узагальненою табличною алгеброю розуміємо алгебру $\langle \boldsymbol{T}, \Omega_{P,\Xi}\rangle$, де $\boldsymbol{T}$ – множина усіх таблиць, $\Omega_{P,\Xi} = \{\bigcup_R, \bigcap_R, \setminus_R, \pm_{p,R}, ®_{X,R}, \underset{R_1,R_2}{\otimes}, \div_{R_2}^{R_1}, Rt_{¶,R}, \sim_R\}$ – сигнатура, $p \in P, ¶ \in \Xi$, $X, R, R_1, R_2 \subseteq \boldsymbol{A}$, P, Ξ – множини параметрів. Операції сигнатури $\Omega_{P,\Xi}$ задано в [3]. Виразом табличної алгебри називається вираз, побудований з таблиць множини $\boldsymbol{T}$ при використанні операцій з сигнатури $\Omega_{P,\Xi}$.

Лема 1. Будь-який вираз табличної алгебри можна замінити еквівалентним йому виразом, який використовує сталі таблиці з єдиним атрибутом і єдиним рядком, операції селекції, з'єднання, проекції, об'єднання, різниці й перейменування

(доведення див., наприклад, в [2]).

3. Узагальнене числення рядків

Розглянемо множину атрибутів $\boldsymbol{A}$ і універсальний домен $\boldsymbol{D}$; множини предметних змінних (змінних рядків) $x_1, x_2, ...$; множини предметних констант $d_1, d_2, ...$; множини функціональних символів $f_1^{n_1}, f_2^{n_2}, ...$; множини предикатних символів $p_1^{m_1}, p_2^{m_2}, ...$. Областю інтерпретації предметних констант є універсальний домен $\boldsymbol{D}$, предметних змінних – множина всіх рядків.

Введемо синтаксичні змінні як у [4]. Використовуємо $\boldsymbol{x}$ як синтаксичну змінну, областю зміни якої є предметні змінні; $\boldsymbol{f}$ ($\boldsymbol{p}$) як синтаксичну змінну, областю зміни якої є функціональні (відповідно предикатні) символи; $\boldsymbol{d}$ як синтаксичну змінну, областю зміни якої є предметні константи; $\mathcal{A}$ як синтаксичну змінну, областю зміни якої є атрибути.

Наступні вирази є термами (індукція за довжиною термів):

a) $\boldsymbol{d}$ – терм;

b) $\boldsymbol{x}$ ($\mathcal{A}$) – терм;

c) якщо $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ – терми, $\boldsymbol{f}$ – n-арний функціональний символ, то $\boldsymbol{f}(\boldsymbol{u}_1, ..., \boldsymbol{u}_n)$ – терм;

d) вираз є термом, тільки тоді, коли це слідує з правил a), b), c).

Будемо використовувати $\boldsymbol{u}$ як синтаксичну змінну, областю зміни якої є терми. Сформулюємо правила побудови формул. Атомарні формули (атоми) можуть бути двох типів.

a1. Нехай $\langle t, R \rangle$ – таблиця, а $\boldsymbol{x}$ – змінний рядок. Тоді $t(\boldsymbol{x})$ – атом, який (при інтерпретації) означає, що $\boldsymbol{x} \in t$.

a2. Нехай $\boldsymbol{u}_1,...,\boldsymbol{u}_m$ – терми, а $\boldsymbol{p}$ – m-арний предикат на універсальному домені $\boldsymbol{D}$. Тоді $\boldsymbol{p}(\boldsymbol{u}_1,...,\boldsymbol{u}_m)$ – атом.

Використаємо логічні зв'язки $\neg$, $\wedge$, $\vee$, квантори $\exists, \forall$ та дужки (,) для побудови формул із атомів. Будемо використовувати $\boldsymbol{P}$, $\boldsymbol{Q}$ та $\boldsymbol{G}$ як синтаксичні змінні, областю зміни яких є формули, визначені нижче.

f1. Кожний атом – це формула.

f2. Якщо $\boldsymbol{P}$ – формула, то $\neg\boldsymbol{P}$ – теж формула.

f3. Якщо $\boldsymbol{P}$ і $\boldsymbol{Q}$ – формули, то $(\boldsymbol{P} \wedge \boldsymbol{Q})$, $(\boldsymbol{P} \vee \boldsymbol{Q})$ – формули.

f4. Нехай $\boldsymbol{x}$ – змінний рядок, $\boldsymbol{P}$ – формула, $R \subseteq \boldsymbol{A}$ – схема, тоді $\exists \boldsymbol{x}(R)\boldsymbol{P}$ – формула.

f5. Нехай $\boldsymbol{x}$ – змінний рядок, $\boldsymbol{P}$ – формула, $R \subseteq \boldsymbol{A}$ – схема, тоді $\forall \boldsymbol{x}(R)\boldsymbol{P}$ – формула.

f6. Якщо $\boldsymbol{P}$ – формула, то $(\boldsymbol{P})$ – теж формула.

f7. Інших формул немає.

В загальному випадку змінні рядки у формулах можуть бути вільні або зв'язані. Смисл цих понять такий самий як і у численні предикатів (див., наприклад, [4]). Використовуючи поняття вільних та зв'язаних змінних рядків, схеми $scheme(\boldsymbol{x}, \boldsymbol{P})$ (скінченної множини атрибутів) та множини атрибутів $attr(x, \boldsymbol{P})$, з якими рядок $\boldsymbol{x}$ зустрічається у формулах, виділено клас дозволених формул. Точні означення див. в [3].

Вираз числення рядків має вигляд $\{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x})\}$, де

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

Нехай $\boldsymbol{P}(\boldsymbol{x})$ – дозволена формула, $R \subseteq \boldsymbol{A}$. В результаті підстановки конкретного рядка s схеми R замість $\boldsymbol{x}$ у $\boldsymbol{P}$ отримаємо формулу $\boldsymbol{P}(s / \boldsymbol{x})$.

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

4. Реляційне числення зі змінними на домені

Числення на домені відрізняється від числення рядків тим, що замість змінних рядків розглядаються змінні значення із домену, які представляють компоненти рядка. Також числення на домені підтримує умову приналежності (membership condition) (див., наприклад, в [5]).

Розглянемо множину атрибутів $\boldsymbol{A}$ і універсальний домен $\boldsymbol{D}$; множини предметних констант $d_1, d_2, \ldots$; множини предметних змінних $x_1, x_2, \ldots$; множини функціональних символів $f_1^{n_1}, f_2^{n_2}, \ldots$; множини предикатних символів $p_1^{m_1}, p_2^{m_2}, \ldots$. Областю інтерпретації предметних констант та предметних змінних є універсальний домен $\boldsymbol{D}$. Як і раніше, використовуємо синтаксичні змінні $\boldsymbol{x}$, $\boldsymbol{f}$ ($\boldsymbol{p}$), $\boldsymbol{d}$, $\mathcal{A}$, $\boldsymbol{u}$, $\boldsymbol{P}$, $\boldsymbol{Q}$ та $\boldsymbol{G}$, областю зміни яких є предметні змінні, функціональні (відповідно предикатні) символи, константи, атрибути, терми та формули відповідно.

Наступні вирази є термами (індукція за довжиною термів):

a) будь-яка предметна константа або предметна змінна є термом;

b) якщо $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ – терми і $\boldsymbol{f}$ – n-арний функціональний символ, то $\boldsymbol{f}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ – терм;

c) вираз є термом, тільки тоді, коли це слідує з правил a), b).

Атомарні формули (атоми) мають вигляд:

a1. нехай $\langle t, R \rangle$ – таблиця схеми $R = \{A_1, \ldots, A_n\} \subseteq \boldsymbol{A}$. Тоді $t(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ – атом, де $\boldsymbol{a}_i$ – константа або змінна на універсальному домені $\boldsymbol{D}$;

a2. нехай $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ – терми, а $\boldsymbol{p}$ – m-арний предикат на універсальному домені $\boldsymbol{D}$. Тоді $\boldsymbol{p}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$ – атом.

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

f1. Кожний атом – це формула.

f2. Якщо $\boldsymbol{P}$ – формула, то $\neg \boldsymbol{P}$ – теж формула.

f3. Якщо $\boldsymbol{P}$ і $\boldsymbol{Q}$ – формули, то $(\boldsymbol{P} \wedge \boldsymbol{Q})$, $(\boldsymbol{P} \vee \boldsymbol{Q})$ – формули.

f4. Нехай $\boldsymbol{x}$ – змінна, $\boldsymbol{P}$ – формула, $\mathcal{A}$ – атрибут з множини $\boldsymbol{A}$, тоді $\exists \boldsymbol{x} (\mathcal{A}) \boldsymbol{P}$ – формула.

f5. Нехай $\boldsymbol{x}$ – змінна, $\boldsymbol{P}$ – формула, $\mathcal{A}$ – атрибут з множини $\boldsymbol{A}$, тоді $\forall \boldsymbol{x} (\mathcal{A}) \boldsymbol{P}$ – формула.

f6. Якщо $\boldsymbol{P}$ – формула, то і $(\boldsymbol{P})$ – формула.

f7. Інших формул немає.

Поняття вільних і зв'язаних змінних на домені визначаються в численні на домені так само, як і в численні рядків. Схема змінної $\boldsymbol{x}$ у формулі $\boldsymbol{P}$, $scheme(\boldsymbol{x}, \boldsymbol{P})$ – це або універсальний домен $\boldsymbol{D}$, або невизначена величина. Поняття дозволеної формули визначається як і в численні рядків, а саме: формула дозволена, якщо змінна, зв'язана квантором, входить вільно у формулу, яка слідує за квантором.

Вираз числення зі змінними на домені має вигляд $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{P}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\}$, де

1. формула $\boldsymbol{P}$ – дозволена, а $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ – всі вільні змінні, які входять у формулу $\boldsymbol{P}$;
2. $R = \{A_1, \ldots A_n\}, R \subseteq \boldsymbol{A}$ – схема, порядок атрибутів фіксований;
3. $scheme(\boldsymbol{x_i}, \boldsymbol{P}) = \boldsymbol{D}$, $i = 1, \ldots, n$.

Нехай $\boldsymbol{P}(\boldsymbol{x})$ – дозволена формула. У результаті підстановки константи $\boldsymbol{d}$ з універсального домену $\boldsymbol{D}$ замість змінної $\boldsymbol{x}$ у $\boldsymbol{P}$ отримаємо формулу $\boldsymbol{P}(\boldsymbol{d}/\boldsymbol{x})$.

Нехай $E = \{\boldsymbol{x}_1,...,\boldsymbol{x}_n \mid \boldsymbol{P}(\boldsymbol{x}_1,...,\boldsymbol{x}_n)\}$ – вираз числення на домені. Значенням виразу E назвемо таблицю схеми $R = \{A_1,...,A_n\}$, яка містить всі рядки $s \in S(R)$, $s = \{\langle A_1, d_1\rangle,...,\langle A_n, d_n\rangle\}$ такі, що формула $\boldsymbol{P}(d_1/\boldsymbol{x}_1,...,d_n/\boldsymbol{x}_n)$ істинна.

5. Еквівалентність узагальненої табличної алгебри, узагальненого числення рядків та узагальненого числення на домені

Покажемо, що для довільного виразу табличної алгебри можна побудувати еквівалентний вираз числення рядків. Вираз числення рядків E еквівалентний виразу табличної алгебри F, якщо значення виразу E співпадає із значенням виразу F, за умови, коли табличні константи і табличні змінні інтерпретуються однаково в цих двох виразах.

Теорема 1. Якщо F – вираз табличної алгебри, то можна ефективно (і рівномірно) побудувати еквівалентний йому вираз E числення рядків.

Доведення.

Згідно леми 1 розглядаємо вирази табличної алгебри, які містять лише операції об'єднання, різниці, селекції, проекції, з'єднання та перейменування, причому допускаються тільки одноатрибутні сталі таблиці з одним рядком. Доведення проведемо індукцією за числом операцій в F.

База індукції (немає операцій). Можливі два випадки. По-перше, $F = t$, де t – таблиця схеми R, покладемо $E = \{\boldsymbol{x}(R) \mid t(\boldsymbol{x})\}$. По-друге, F – стала таблиця $t = \{\{\langle A, d\rangle\}\}$. Покладемо $E = \{\boldsymbol{x}(\{A\}) \mid x(A) = d\}$. □

Крок індукції. Припустимо, що твердження теореми виконується для будь-якого виразу табличної алгебри, який містить менше $k = 1, 2, 3...$ операцій. Нехай вираз F містить k операцій.

Випадок 1. $F = F_1 \cup_R F_2$. Тоді F_1 і F_2 мають менше k операцій, тому існують вирази числення рядків $\{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x})\}$ і $\{\boldsymbol{x}(R) \mid \boldsymbol{Q}(\boldsymbol{x})\}$, еквівалентні F_1 і F_2 відповідно. Покладемо E рівним $\{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x}) \vee \boldsymbol{Q}(\boldsymbol{x})\}$. □

Випадок 2. $F = F_1 \setminus_R F_2$. Нехай $\{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x})\}$ і $\{\boldsymbol{x}(R) \mid \boldsymbol{Q}(\boldsymbol{x})\}$ – вирази числення рядків, еквівалентні F_1 і F_2 відповідно. Покладемо $E = \{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x}) \wedge \neg\boldsymbol{Q}(\boldsymbol{x})\}$.□

Випадок 3. $F = \pm_{\tilde{p},R}(F_1)$. Нехай $\{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x})\}$ – вираз числення рядків, еквівалентний F_1. Покладемо $E = \{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x}) \wedge \boldsymbol{p}(x(A_1),...,x(A_m))\}$, де $R = \{A_1,...,A_m\}$ схема таблиці, яка є значенням виразу F_1. Тут припускається, що предикат-параметр селекції заданий так $\tilde{p}(s) = true \Leftrightarrow \boldsymbol{p}(s(A_1),...,s(A_m)) = true$, $s \in S(R)$, де $\boldsymbol{p}$ – m-арний сигнатурний предикатний символ. □

Випадок 4. $F = ®_{X,R}(F_1)$. Нехай $\{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x})\}$ – вираз числення рядків, еквівалентний F_1. Покладемо E рівним $\{\boldsymbol{y}(X \cap R) \mid \exists \boldsymbol{x}(R)(\boldsymbol{P}(\boldsymbol{x}) \wedge \underset{A \in X \cap R}{\wedge} \boldsymbol{y}(A) = \boldsymbol{x}(A))\}$.□

Випадок 5. $F = F_1 \underset{R_1,R_2}{\otimes} F_2$. Нехай $\{\boldsymbol{x}(R_1) \mid \boldsymbol{P}(\boldsymbol{x})\}$ і $\{\boldsymbol{y}(R_2) \mid \boldsymbol{Q}(\boldsymbol{y})\}$ – вирази числення рядків, еквівалентні F_1 та F_2 відповідно. Покладемо E рівним $\{\boldsymbol{z}(R_1 \cup R_2) \mid \exists \boldsymbol{x}(R_1)\exists \boldsymbol{y}(R_2)(\boldsymbol{P}(\boldsymbol{x}) \wedge \boldsymbol{Q}(\boldsymbol{y}) \wedge \underset{A \in R_1}{\wedge} \boldsymbol{z}(A) = \boldsymbol{x}(A) \wedge \underset{A \in R_2}{\wedge} \boldsymbol{z}(A) = \boldsymbol{y}(A))\}$.□

Випадок 6. $F = Rt_{¶,R}(F_1)$, де $¶ : \boldsymbol{A} \tilde{\to} \boldsymbol{A}$ – ін'єктивна функція, що здійснює перейменування атрибутів. Тоді існує вираз числення рядків $\{\boldsymbol{x}(R) \mid \boldsymbol{P}(\boldsymbol{x})\}$, еквівалентний F_1. Покладемо E рівним $\{\boldsymbol{y}(R_2) \mid \exists \boldsymbol{x}(R)(\boldsymbol{P}(\boldsymbol{x}) \wedge \underset{C \in R \setminus dom¶}{\wedge} \boldsymbol{y}(C) = \boldsymbol{x}(C) \wedge \underset{A \in R \cap dom¶}{\wedge} \boldsymbol{x}(A) = \boldsymbol{y}(¶(A)))\}$, де $R_2 = R \setminus dom¶ \cup ¶[R]$. □

Проведемо редукцію числення рядків до числення на домені. Задамо відображення $\ddot{} : F \to E$, яке кожному виразу числення рядків ставить у відповідність еквівалентний вираз числення на домені. Нехай $F = \{\boldsymbol{y}(R) \mid \boldsymbol{P}(\boldsymbol{y})\}$ – вираз числення рядків, змінна $\boldsymbol{y}$ – єдина змінна, яка входить у формулу $\boldsymbol{P}$ вільно, а $R = \{A_1,...,A_n\}$. Причому $\boldsymbol{y}$ – синтаксична змінна, областю зміни якої є змінні числення рядків. Застосуємо спочатку відображення $\ddot{}$ до термів:

1. Кожну предметну константу $\boldsymbol{d}$ числення рядків залишимо без змін.
2. Кожен терм $\boldsymbol{x}(A_i)$ числення рядків замінимо змінною $\boldsymbol{x_i} \in \boldsymbol{D}$ числення на домені.
3. Кожен терм $\boldsymbol{f}(\boldsymbol{v}_1,...,\boldsymbol{v}_n)$, де $\boldsymbol{v}_i$ – терми числення рядків, $i = 1,...,n$, замінимо на $\boldsymbol{f}(\boldsymbol{u}_1,...,\boldsymbol{u}_n)$, де $\boldsymbol{u}_j$ – терми числення на домені, $j = 1,...,n$, отримані в результаті попередніх замін.

Застосуємо відображення $\ddot{}$ до атомів:

a1. Кожен атом $t(\boldsymbol{z})$ числення рядків замінимо на $t(\boldsymbol{z}_1,...,\boldsymbol{z}_m)$, де $R = \{A_1,...,A_m\}$ схема таблиці t.

a2. Кожен атом $\boldsymbol{p}(\boldsymbol{v}_1,...,\boldsymbol{v}_m)$, де $\boldsymbol{v}_i$ – терми числення рядків, $i = 1,...,m$, замінимо на $\boldsymbol{p}(\boldsymbol{u}_1,...,\boldsymbol{u}_m)$, де $\boldsymbol{u}_j$ – терми числення на домені, $j = 1,...,m$, отримані в результаті попередніх замін.

Кожну формулу $\boldsymbol{P}$ числення рядків перетворимо на формулу $\boldsymbol{P}'$ числення на домені, де формула $\boldsymbol{P}'$ – це формула $\boldsymbol{P}$, у якій для кожного атома виконані заміни a1-a2, а кожна вільна змінна $\boldsymbol{z}$ схеми $R = \{A_1,...,A_m\}$ числення рядків замінена m змінними $\boldsymbol{z}_1,...,\boldsymbol{z}_m$ числення на домені. Формулу з квантором $\exists \boldsymbol{z}(R_2)\boldsymbol{G}$, $R_2 = \{A_1,...,A_m\}$ перетворимо на $\exists \boldsymbol{z}_1(A_1)...\exists \boldsymbol{z}_m(A_m)\boldsymbol{G}'$. Формулу з квантором $\forall \boldsymbol{z}(R_2)\boldsymbol{G}$, $R_2 = \{A_1,...,A_m\}$ перетворимо на $\forall \boldsymbol{z}_1(A_1)...\forall \boldsymbol{z}_m(A_m)\boldsymbol{G}'$.

Виконавши вказані заміни, отримаємо вираз числення зі змінними на домені $E = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_m \mid \boldsymbol{P}(\boldsymbol{y}_1, ..., \boldsymbol{y}_m)\}$. Очевидно, що кожна змінна $\boldsymbol{z}_i$ числення на домені може приймати точно ті самі значення, що і $\boldsymbol{z}\,(\mathcal{A}_i)$. Отже, значення виразу E співпадає із значенням виразу F. Тому має місце наступна теорема.

Теорема 2. Якщо F – вираз числення рядків, то можна ефективно (і рівномірно) побудувати еквівалентний йому вираз E числення на домені.

Теорема 3. Для кожного виразу числення на домені E можна ефективно (і рівномірно) побудувати еквівалентний йому вираз F табличної алгебри.

Доведення.

Нехай $E = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n \mid \boldsymbol{P}(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)\}$ – вираз числення на домені, значенням якого є таблиця схеми $R = \{A_1, ..., A_n\}$. Побудуємо для кожної підформули $\boldsymbol{G}$ формули $\boldsymbol{P}$ еквівалентний алгебраїчний вираз F_G. Доведемо індукцією по числу операторів у підформулі $\boldsymbol{G}$ з $\boldsymbol{P}$, що якщо $\boldsymbol{G}$ містить вільні змінні на універсальному домені $\boldsymbol{y}_1, ..., \boldsymbol{y}_m$, то для виразу $\{\boldsymbol{y}_1, ..., \boldsymbol{y}_m \mid \boldsymbol{G}(\boldsymbol{y}_1, ..., \boldsymbol{y}_m)\}$ існує еквівалентний вираз табличної алгебри F_G. Тоді, якщо $\boldsymbol{G}$ є сама $\boldsymbol{P}$, отримаємо алгебраїчний вираз для $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n \mid \boldsymbol{P}(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)\}$. Припускається, що $\boldsymbol{y}_1, ..., \boldsymbol{y}_m$ – всі вільні змінні в $\boldsymbol{G}$ і значенням виразу $\{\boldsymbol{y}_1, ..., \boldsymbol{y}_m \mid \boldsymbol{G}(\boldsymbol{y}_1, ..., \boldsymbol{y}_m)\}$ є таблиця схеми $R_G = \{A_1, ..., A_m\}$.

Замінимо змінні у формулі $\boldsymbol{P}$ таким чином, щоб жодна змінна не входила одночасно і вільно, і зв'язано у $\boldsymbol{P}$ та не була зв'язана квантором у двох різних місцях. Тепер кожна змінна асоційована з атрибутом, який або входить у квантор $\forall \boldsymbol{x}\,(\mathcal{A})$ чи $\exists \boldsymbol{x}\,(\mathcal{A})$, якщо змінна зв'язана у формулі $\boldsymbol{P}$, або, якщо змінна вільна в $\boldsymbol{P}$, співставлений змінній $\boldsymbol{x}_i$, що знаходиться у виразі E зліва від вертикальної лінії. Для будь-якого атрибута $\mathcal{A}$ можна виписати алгебраїчний вираз, значенням якого є одноатрибутна таблиця $\langle t, \{A\} \rangle$, що містить всі рядки $s = \{\langle A, d_i \rangle\}$, $d_i \in \boldsymbol{D}$ – всі елементи множини $\boldsymbol{D}$. Позначимо його через $[\boldsymbol{D}]$.

База індукції (немає операторів, тобто логічних зв'язок і кванторів). Підформула $\boldsymbol{G}$ – атом вигляду $\boldsymbol{p}(\boldsymbol{u}_1, ..., \boldsymbol{u}_m)$ або $t(\boldsymbol{a}_1, ..., \boldsymbol{a}_n)$.

Нехай $\boldsymbol{G}$ має вигляд $\boldsymbol{p}(\boldsymbol{u}_1, ..., \boldsymbol{u}_m)$, де $\boldsymbol{u}_i$ – терми числення на домені, причому $\boldsymbol{y}_1, ..., \boldsymbol{y}_k$ – всі змінні цих термів. Покладемо F_G рівним $\pm_{\tilde{p}}([\boldsymbol{D}]_1 \underset{R_1, R_2}{\otimes} ... \underset{R_1 \cup ... \cup R_{k-1}, R_k}{\otimes} [\boldsymbol{D}]_k)$, де R_i – одноатрибутні схеми таблиць, причому ці таблиці є значеннями алгебраїчних виразів $[\boldsymbol{D}]_i$, $i = 1, ..., k$. □

Нехай $\boldsymbol{G}$ має вигляд $t(\boldsymbol{a}_1, ..., \boldsymbol{a}_k)$, де $\boldsymbol{a}_i$ – константа або змінна на універсальному домені $\boldsymbol{D}$. Нехай $R = \{C_1, ..., C_k\}$ – схема таблиці $\langle t, R \rangle$. Покладемо алгебраїчний вираз F_G рівним $®_X(\pm_{\tilde{p}}(t))$, де $\tilde{p}$ – предикат-параметр селекції, який є кон'юнкцією порівнянь $C_i = \boldsymbol{a}_i$ по всіх i, для яких $\boldsymbol{a}_i$ – константа; X – множина атрибутів $\{C_j \mid \boldsymbol{a}_j$ – змінна$\}$.□

Крок індукції.

Припустимо, що підформула $\boldsymbol{G}$ містить принаймні один оператор і що гіпотеза індукції справедлива для всіх підформул формули $\boldsymbol{P}$, що мають менше операторів, ніж $\boldsymbol{G}$.

Випадок 1. $\boldsymbol{G} = \neg\boldsymbol{Q}$. Нехай F_Q – алгебраїчний вираз для $\boldsymbol{Q}$, а значенням виразу F_Q є таблиця схеми R_Q. Покладемо $F_G = \sim_{R_Q} F_Q$.□

Випадок 2. $\boldsymbol{G} = \boldsymbol{Q} \vee \boldsymbol{Q}'$. Нехай вільні змінні з $\boldsymbol{Q}$ – це $z_1,...,z_k,v_1,...,v_p$, а з $\boldsymbol{Q}'$– це $z_1,...,z_k,w_1,...,w_q$, де змінні $v_1,...,v_p$ та $w_1,...,w_q$ попарно не співпадають. Нехай F_Q і $F_{Q'}$ – алгебраїчні вирази для $\boldsymbol{Q}$ і $\boldsymbol{Q}'$ відповідно, причому C_i, $i = 1,..., p$ – атрибути, асоційовані зі змінними $v_1,...,v_p$, а K_j, $j = 1,...,q$ – атрибути асоційовані із змінними $w_1,...,w_q$.

Покладемо рівним $F_1 = F_Q \underset{R_Q,\{K_1\}}{\otimes} [\boldsymbol{D}]_1 \underset{R_Q \cup \{K_1\},\{K_2\}}{\otimes} ... \underset{R_Q \cup \{K_1,...,K_{q-1}\},\{K_q\}}{\otimes} [\boldsymbol{D}]_q$,

а $F_2 = F_{Q'} \underset{R_{Q'},\{C_1\}}{\otimes} [\boldsymbol{D}]_1' \underset{R_{Q'} \cup \{C_1\},\{C_2\}}{\otimes} ... \underset{R_{Q'} \cup \{C_1,...,C_{p-1}\},\{C_p\}}{\otimes} [\boldsymbol{D}]_p'$, де R_Q та $R_{Q'}$ – схеми таблиць, які є значеннями виразів F_Q та $F_{Q'}$ відповідно, а C_i та K_j – одноатрибутні схеми таблиць, які є значеннями алгебраїчних виразів $[\boldsymbol{D}]_i'$ та $[\boldsymbol{D}]_j$ відповідно, $i = 1,..., p$, $j = 1,...,q$. Позначимо через R_{F_1} та R_{F_2} – схеми таблиць, які є значеннями виразів F_1 та F_2 відповідно. Для відповідним чином побудованих виразів F_1 та F_2, їхні схеми R_{F_1} та R_{F_2} рівні. Тому покладемо $F_G = F_1 \cup_{R_{F_1}} F_2$.□

Випадок 3. $\boldsymbol{G} = \boldsymbol{Q} \wedge \boldsymbol{Q}'$. За законом де Моргана $\boldsymbol{G} = \neg(\neg\boldsymbol{Q} \vee \neg\boldsymbol{Q}')$.□

Випадок 4. $\boldsymbol{G} = \exists\boldsymbol{x}(A)\boldsymbol{Q}$. Нехай F_Q – алгебраїчний вираз для $\boldsymbol{Q}$. Покладемо $F_G = ®_{X \setminus A, X}(F_Q)$, де X – схема таблиці, що є значенням виразу F_Q.□

Випадок 5. $\boldsymbol{G} = \forall\boldsymbol{x}(A)\boldsymbol{Q}$. Даний вираз числення на домені можна виразити через операцію заперечення та квантор існування $\forall\boldsymbol{x}(A)\boldsymbol{Q} = \neg(\exists\boldsymbol{x}(A))(\neg\boldsymbol{Q})$. □

Таким чином, встановлено основний результат

Теорема 4. Узагальнена таблична алгебра, (узагальнене) числення рядків і (узагальнене) числення на домені еквівалентні.

6. Висновки

Проведено узагальнення табличної алгебри на таблиці, яким приписується певна схема. Класичне числення рядків та числення на домені поповнено довільними предикатними та функціональними сигнатурами на універсальному домені $\boldsymbol{D}$. Визначено синтаксис термів, атомів та формул числення рядків і числення на домені. Доведено, що узагальнена таблична алгебра, узагальнене числення рядків та узагальнене числення на домені еквівалентні.

ЛИТЕРАТУРА

1. Реляційні бази даних: табличні алгебри та SQL-подібні мови / В.Н. Редько, Ю.Й. Борона, Д.Б. Буй, С.А. Поляков. – Київ: Видавничий дім «Академперіодика», 2001. – 198 с.
2. Мейер Д. Теория реляционных баз данных. – М.: Мир, 1987. – 608 с.
3. Буй Д.Б., Глушко І.М.: Узагальнена таблична алгебра та узагальнене реляційне числення // Теоретичні та прикладні аспекти побудови програмних систем: міжнародна конференція, 4-8 жовтня 2010 року, Київ, Україна: тези доповідей. – Київ. – 2010. – С. 112-120.
4. Шенфилд Дж. Математическая логика. – М.: Наука, 1975. – 528 с.
5. Дейт К. Дж. Введение в системы баз данных: [8-е изд.: пер. с англ.]. – М.: Издательский дом «Вильямс», 2005. – 1328 с.

Надійшла – 28.04.2011.