

ПРО ЗВ'ЯЗОК МІЖ ТАБЛИЧНОЮ АЛГЕБРОЮ НЕСКІНЧЕННИХ ТАБЛИЦЬ ТА МУЛЬТИМНОЖИННОЮ ТАБЛИЧНОЮ АЛГЕБРОЮ

І.М. Глушко

Дана робота є продовженням робіт, присвячених актуальній проблемі розробки теоретичної основи табличних баз даних, в якості якої виступають табличні алгебри. Розглядається питання про зв'язок між табличною алгеброю нескінченних таблиць та мультимножинною табличною алгеброю. Враховуючи той факт, що 1-мультимножини є аналогами звичайних множин, виникає питання, чи є таблична алгебра нескінченних таблиць підалгеброю мультимножинної табличної алгебри. Ця проблема і досліджується у даній роботі. Застосовуючи теоретико-множинні та логіко-алгебраїчні методи, встановлено, що таблична алгебра нескінченних таблиць не утворює підалгебру мультимножинної табличної алгебри, бо не є замкненою відносно деяких операцій сигнатури мультимножинної табличної алгебри. У роботі встановлено, що це за операції.

Ключові слова. Реляційні бази даних, таблична алгебра нескінченних таблиць, мультимножинна таблична алгебра.

Данная работа является продолжением работ, посвященных актуальной проблеме разработки теоретической основы табличных баз данных, в качестве которой выступают табличные алгебры. Рассматривается вопрос о связи между табличной алгеброй бесконечных таблиц и мультимножественной табличной алгеброй. Учитывая тот факт, что 1-мультимножества являются аналогами обычных множеств, возникает вопрос, является ли табличная алгебра бесконечных таблиц подалгеброй мультимножественной табличной алгебры. Эта проблема и исследуется в этой работе. Применяя теоретико-множественные и логико-алгебраические методы, установлено, что табличная алгебра бесконечных таблиц не образует подалгебру мультимножественной табличной алгебры, потому что не является замкнутой относительно некоторых операций сигнатуры мультимножественной табличной алгебры. В работе установлено, что это за операции.

Ключевые слова. Реляционные базы данных, табличная алгебра бесконечных таблиц, мультимножественная табличная алгебра.

This article is a continuation of the works devoted to the actual problem of the development of the theoretical basis of the table databases. The question of the relationship between table algebra of infinite tables and multiset table algebra is considered. Considering the fact that 1-multisets are analogues of ordinary sets, the question arises, is whether table algebra of infinite tables a subalgebra of multiset table algebra. This paper is devoted to this issue. Applying the theorem-plural and logical-algebraic methods found that this is not the case. The table algebra of infinite tables does not form subalgebra of multiset table algebra, since it is not closed in relation to some signature operations of multiset table algebra. These operations are determined.

Key words. Relation databases, table algebra of infinite tables, multiset table algebra.

Вступ

Уточнення реляції в термінах іменних множин було здійснене В.Н. Редьком, Ю.Й. Бронюю, Д.Б. Буєм, С.А. Поляковим [1]. Традиційно під реляцією розуміється скінченна множина рядків, і автори враховують це обмеження, проте, як правило, математичні твердження про властивості уточнень реляційних операцій залишаються вірними і для нескінченних реляцій. Так, детальний аналіз доведень у [1] показує, що властивість скінченності таблиць в багатьох випадках не використовується. Тому в роботі [2] було здійснено узагальнення, яке полягає в тому, що надалі під реляцією розуміється довільна множина односхемних рядків, зокрема, нескінченна. Також введено до розгляду мультимножинну табличну алгебру, яка є мультимножинним аналогом табличної алгебри, тобто поняття реляції уточнюється, використовуючи поняття мультимножини, зокрема нескінченної [2]. Беручи до уваги, що 1-мультимножини є аналогами звичайних множин, постає питання чи є таблична алгебра нескінченних таблиць підалгеброю мультимножинної табличної алгебри. Цьому питанню й присвячена дана робота.

Таблична алгебра нескінченних таблиць

Дамо означення основних понять табличної алгебри нескінченних таблиць на основі монографії [2].

Розглянемо дві множини: A – множину атрибутів і D – універсальний домен. Довільна (скінченна) множина атрибутів $R \subseteq A$ називається схемою. Рядком схеми R називається іменна множина на парі R, D , проєкція якої за першою компонентою рівна R (тобто по суті розглядається функція вигляду $s: R \rightarrow D$).

Таблицею схеми R є пара $\langle t, R \rangle$, де t – множина (зокрема, нескінченна) рядків вказаної схеми R . Надалі через $\langle \langle t, R \rangle \rangle$ будемо позначати першу компоненту пари $\langle t, R \rangle$, тобто множину t .

Множину усіх рядків (таблиць) схеми R позначимо $S(R)$ (відповідно $T(R)$), а множину всіх рядків (таблиць) – S (відповідно T). Таким чином, $S = \bigcup_{R \subseteq A} S(R)$, $T(R) = \{ \langle t, R \rangle \mid t \in P(S(R)) \}$, $T = \bigcup_{R \subseteq A} T(R)$, де $P(X)$ –

булеан множини X .

У роботі [2] проведена конкретизація, яка полягає в тому, що замість таблиць (як множини рядків) розглядаються пари, які складаються з таблиць та їхніх схем, тобто кожній таблиці однозначно приписується певна схема. Це по суті впливає тільки на випадок порожньої таблиці, оскільки за непорожньою таблицею схема відновлюється однозначно. Запис $\langle t_{\emptyset}, R \rangle$ позначає порожню таблицю схеми R .

Під табличною алгеброю нескінченних таблиць розуміємо (часткову параметричну) алгебру $\langle T, \Omega_{P, \Xi} \rangle$,

де T – множина усіх таблиць, $\Omega_{P, \Xi} = \left\{ \bigcup_{R, R_1, R_2 \subseteq A} \cup, \cap, \setminus, \sigma_{p, R}, \pi_{X, R}, \otimes_{R_1, R_2}, \div_{R_1, R_2}^{R_1}, Rt_{\xi, R}, \sim_R \right\}_{p \in P, \xi \in \Xi}$ – сигнатура,

P, Ξ – множини параметрів. Сигнатура містить аналоги теоретико-множинних операцій (об'єднання, перетин, різницю) та спеціальні операції (селекцію, проєкцію, з'єднання, ділення, перейменування та активне доповнення).

Мультимножинна таблична алгебра

Дамо означення основних понять мультимножинної табличної алгебри на основі монографії [2].

Як і раніше, A – множина атрибутів, D – універсальний домен. Довільна скінченна множина атрибутів $R \subseteq A$ – це схема. Рядком схеми R називається іменна множина на парі R, D , проєкція якої за першою компонентою рівна R . Множина всіх рядків схеми R позначається $S(R)$, а множина всіх рядків – S .

Поняття таблиці задається як пара $\langle \psi, R \rangle$, де перша компонента ψ – це довільна мультимножина, зокрема, нескінченна, а друга компонента R – схема таблиці.

Множину всіх таблиць схеми R позначимо $\Psi(R)$, а множину всіх таблиць $\Psi = \bigcup_{R \subseteq A} \Psi(R)$.

Позначимо через $Occ(s, \psi)$ – кількість дублікатів (екземплярів) рядка s у мультимножині ψ .

Домовимося мультимножину ψ записувати як $\{s_1^{n_1}, \dots, s_k^{n_k}, \dots\}$, де $n_i = Occ(s_i, \psi)$, $i = 1, 2, \dots$, а $\Theta(\psi) = \{s_1, \dots, s_k, \dots\}$ – основа мультимножин ψ .

Під мультимножинною табличною алгеброю розуміємо алгебру $\langle \Psi, \Omega_{P, \Xi} \rangle$, де Ψ – множина всіх

таблиць, $\Omega_{P, \Xi} = \left\{ \bigcup_{R, R_1, R_2 \subseteq A} \cup_{All}^{\Psi, R}, \cap_{All}^{\Psi, R}, \setminus_{All}^{\Psi, R}, \sigma_{p, R}, \pi_{X, R}, \otimes_{R_1, R_2}, Rt_{\xi, R}, \sim_R \right\}_{p \in P, \xi \in \Xi}$ – сигнатура, P, Ξ – множини

параметрів.

Наведемо основні поняття мультимножин в термінах робіт [1, 3]. Зафіксуємо деяку множину U . Під мультимножиною α з основою U будемо розуміти відображення вигляду $\alpha: U \rightarrow N$, де $N = \{1, 2, \dots\}$ – множина натуральних чисел.

Нехай D – універсум елементів основ мультимножин, тоді булеан $P(D)$ – універсум основ мультимножин. Під характеристичною функцією мультимножини α розуміємо функцію вигляду $\chi_\alpha: D \rightarrow Z_+$, значення якої задається наступною кусковою схемою:

$$\chi_\alpha(d) = \begin{cases} \alpha(d), & \text{якщо } d \in \text{dom } \alpha, \\ 0, & \text{інакше;} \end{cases}$$

для всіх $d \in D$.

Мультимножина називається порожньою і позначається як \emptyset_m , якщо її основа – порожня множина.

Мультимножини, областю значень яких є порожня множина або одноелементна множина вигляду $\{1\}$, називаються 1-мультимножинами.

У роботі [1] операції над мультимножинами визначено в термінах характеристичних функцій. Автори вводять операції об'єднання \cup_1 , перетину \cap_1 , різниці \setminus_1 мультимножин, що будують 1-мультимножини, основи яких отримуються відповідно теоретико-множинними об'єднанням, перетином та різницею основ мультимножин-аргументів, та операції об'єднання \cup_{All} , перетину \cap_{All} , різниці \setminus_{All} мультимножин, які будують мультимножини загального вигляду. Також задано операцію декартового з'єднання мультимножин \otimes , операцію $Dist(\alpha)$, що будує 1-мультимножину, основа якої збігається з основою вихідної мультимножини та вводиться аналог повного образу для мультимножин.

Про зв'язок між табличною алгеброю нескінченних таблиць та мультимножинною табличною алгеброю

З'ясуємо чи є таблична алгебра нескінченних таблиць підалгеброю мультимножинної табличної алгебри. Для цього припустимо, що таблицею табличної алгебри нескінченних таблиць є пара $\langle t^1, R \rangle$, де t^1 – 1-множина (зокрема, нескінченна) рядків вказаної схеми R . У цьому випадку очевидно, що множина всіх таблиць табличної алгебри нескінченних таблиць T є підмножиною множини всіх таблиць мультимножинної табличної алгебри Ψ , $T \subseteq \Psi$.

Тепер з'ясуємо чи замкнена множина T відносно кожної операції сигнатури мультимножинної табличної алгебри $\Omega_{p,\Xi}$.

Почнемо з теоретико-множинних операцій: об'єднання \cup_{All}^R , перетину \cap_{All}^R і різниці \setminus_{All}^R . Розглянемо кожну операцію окремо.

Позначимо через $\Theta(t_1^1)$ і $\Theta(t_2^1)$ – основи 1-мультимножин t_1^1 та t_2^1 відповідно. Тоді $\langle t_1^1, R \rangle \cup_{All}^R \langle t_2^1, R \rangle = \langle t_1^1 \cup_{All} t_2^1, R \rangle$, де $\langle t_1^1, R \rangle, \langle t_2^1, R \rangle \in T(R)$.

Основа мультимножини $t_1^1 \cup_{All} t_2^1$ дорівнює об'єднанню основ 1-мультимножин таблиць-аргументів:

$$\Theta(t_1^1 \cup_{All} t_2^1) = \Theta(t_1^1) \cup \Theta(t_2^1).$$

Дублікати рядків, які з'явилися після виконання операції, не будуть вилучатися. Кількість дублікатів кожного рядка визначається за формулою:

$$Occ(s, t_1^1 \cup_{All} t_2^1) = \begin{cases} 1, & \text{якщо } s \in \Theta(t_1^1) \setminus \Theta(t_2^1) \text{ або } s \in \Theta(t_2^1) \setminus \Theta(t_1^1), \\ 2, & \text{якщо } s \in \Theta(t_1^1) \cap \Theta(t_2^1); \end{cases}$$

де $s \in \Theta(t_1^1) \cup \Theta(t_2^1)$. В результаті $t_1^1 \cup_{All} t_2^1$ буде мультимножиною загального вигляду, а не 1-мультимножиною.

Отже, множина T не замкнута відносно операції об'єднання \cup_{All}^R .

З'ясуємо, чи замкнена множина T відносно інших операцій сигнатури мультимножинної табличної алгебри $\Omega_{p,\Xi}$.

Множина T є замкненою відносно перетину: $\cap_{All}^R : T(R) \times T(R) \rightarrow T(R)$, причому $\langle t_1^1, R \rangle \cap_{All}^R \langle t_2^1, R \rangle = \langle t_1^1 \cap_{All} t_2^1, R \rangle$, де $\langle t_1^1, R \rangle, \langle t_2^1, R \rangle \in T(R)$.

Основа мультимножини $t_1^1 \cap_{All} t_2^1$ дорівнює перетину основ 1-мультимножин таблиць-аргументів:

$$\Theta(t_1^1 \cap_{All} t_2^1) = \Theta(t_1^1) \cap \Theta(t_2^1),$$

а кількість дублікатів рядка визначається як

$$Occ(s, t_1^1 \cap_{All} t_2^1) = \min(Occ(s, t_1^1), Occ(s, t_2^1)) = 1, \text{ де } s \in \Theta(t_1^1) \cap \Theta(t_2^1).$$

Оскільки кожен рядок, що входить як в таблицю $\langle t_1^1, R \rangle$, так і в таблицю $\langle t_2^1, R \rangle$ має кількість входжень рівну одиниці, то мінімальна кількість дублікатів цього рядка в таблиці $\langle t_1^1 \cap_{All} t_2^1, R \rangle$ теж буде рівна одиниці.

Отже, $t_1^1 \cap_{All} t_2^1 \in 1$ -мультимножиною і $\langle t_1^1 \cap_{All} t_2^1, R \rangle \in T(R)$, тобто в результаті перетину двох таблиць з множини T отримаємо таблицю з цієї ж множини.

Розглянемо операцію різниці \setminus_{All}^R . Нехай $\langle t_1^1, R \rangle, \langle t_2^1, R \rangle \in T(R)$, тоді $\langle t_1^1, R \rangle \setminus_{All}^R \langle t_2^1, R \rangle = \langle t_1^1 \setminus_{All} t_2^1, R \rangle$.

Основа мультимножини $t_1^1 \setminus_{All} t_2^1$ визначається як

$$\Theta(t_1^1 \setminus_{All} t_2^1) = \Theta(t_1^1) \setminus \Theta(t_2^1).$$

Кількість дублікатів знаходиться так:

$$Occ(s, t_1^1 \setminus_{All} t_2^1) = \begin{cases} 1, & \text{якщо } s \in \Theta(t_1^1) \setminus \Theta(t_2^1), \\ 0, & \text{якщо } s \in \Theta(t_2^1) \setminus \Theta(t_1^1) \end{cases} = Occ(s, t_1^1) \div Occ(s, t_2^1),$$

де $s \in \Theta(t_1^1) \setminus \Theta(t_2^1)$.

В результаті отримаємо, що $t_1^1 \setminus_{All} t_2^1 \in 1$ -мультимножиною і $\langle t_1^1 \setminus_{All} t_2^1, R \rangle \in T(R)$. Отже, множина T замкнена відносно операції різниці \setminus_{All}^R .

Нехай $p : S \xrightarrow{\sim} \{true, false\}$ – частковий предикат на множині рядків. З'ясуємо чи замкнена множина T відносно операції селекції за предикатом p . Маємо, що $\sigma_{p,R}(\langle t^1, R \rangle) = \langle t', R \rangle$, де $\langle t^1, R \rangle \in T(R)$.

Основа мультимножини результуючої таблиці визначається як:

$$\Theta(t') = \{s \mid s \in \Theta(t^1) \wedge p(s) \simeq true\},$$

де \simeq – узагальнена рівність (тобто обидві частини або одночасно не визначені або одночасно визначені і рівні [4]).

В залежності від значення предиката на рядку s усі дублікати цього рядка або входять до мультимножини отриманої таблиці, або ні:

$$Occ(s, t') = Occ(s, t^1) = 1,$$

де $s \in \Theta(t')$. Тому $t' \in 1$ -мультимножиною і $\langle t', R \rangle \in T(R)$. Отже, множина T замкнена відносно операції селекції.

Нехай $X \subseteq A$ – скінченна множина атрибутів. З'ясуємо чи замкнена множина T відносно операції проєкції за множиною атрибутів X . Отже, $\pi_{X,R}(\langle t^1, R \rangle) = \langle t', R \cap X \rangle$, де $\langle t^1, R \rangle \in T(R)$.

Основа мультимножини t' визначається як

$$\Theta(t') = \{s \mid X \mid s \in \Theta(t^1)\}.$$

Дублікати рядків, які з'явилися після виконання операції, як і раніше, не вилучаються. Кількість дублікатів кожного рядка визначається за формулою:

$$Occ(s', t') = \sum_{\substack{s \in \Theta(t^1), \\ s \mid X = s'}} Occ(s, t^1),$$

де $s' \in \Theta(t')$. Таким чином, t' не є 1-мультимножиною. Отже, множина T не замкнена відносно операції проєкції.

З'ясуємо чи замкнена множина T відносно операції з'єднання \otimes_{R_1, R_2} . Маємо, $\langle t_1^1, R_1 \rangle \otimes_{R_1, R_2} \langle t_2^1, R_2 \rangle = \langle t', R_1 \cup R_2 \rangle$, де $\langle t_1^1, R_1 \rangle \in T(R_1)$, $\langle t_2^1, R_2 \rangle \in T(R_2)$. Змістовно кажучи, кожний рядок з t_1^1 з'єднується з кожним рядком із t_2^1 .

Основою мультимножини t' є множина рядків

$$\Theta(t') = \{s' \mid \exists s_1 \exists s_2 (s_1 \in \Theta(t_1^1) \wedge s_2 \in \Theta(t_2^1) \wedge s_1 \approx s_2 \wedge s' = s_1 \cup s_2)\}.$$

Кількість дублікатів знаходиться так:

$$Occ(s', t') = Occ(s' \mid R_1, t_1^1) \cdot Occ(s' \mid R_2, t_2^1) = 1, \text{ де } s' \in \Theta(t').$$

Таким чином, $t' \in 1$ -мультимножиною. Отже, множина T замкнена відносно операції з'єднання.

Розглянемо операцію перейменування. Маємо $R\psi_{\xi,R}(\langle t^1, R \rangle) = \langle Rs_{\eta}[t^1], \eta[R] \rangle$, $t^1 \in T_{\xi}(R)$, де $\eta = \xi \cup id_{A \setminus dom \xi}$, $Rs_{\eta}(s) = \{\langle \eta(A), s(A) \rangle \mid A \in R\}$, $Rs_{\eta}[t^1]$ – повний образ 1-мультимножини t^1 з основою $\Theta(t^1)$ відносно функції Rs_{η} .

Основою мультимножини $Rs_{\eta}[t^1]$ є повний образ множини $\Theta(t^1)$ відносно функції $Rs_{\eta,R}$. А кількість дублікатів рядка у результуючій таблиці задається рівністю

$$Occ(s', Rs_\eta[t^1]) = Occ(s, t^1) = 1,$$

де $s \in Rs_\eta^{-1}(s')$, $s' \in \Theta(Rs_\eta[t^1])$.

Оскільки кількість дублікатів при перейменуванні не змінюється з огляду на ін'єктивність функції Rs_η , то $Rs_\eta[t^1] \in 1$ -мультимножиною. Отже, множина T замкнена відносно операції перейменування.

І наостанок розглянемо операцію активного доповнення \sim_R . Маємо $\sim_R(\langle t^1, R \rangle) = C(\langle t^1, R \rangle) \setminus_{All} \langle t^1, R \rangle$, де насичення визначається за формулою $C(\langle t^1, R \rangle) = \pi_{\{A_1\}, R}(\langle t^1, R \rangle) \otimes_{\{A_1\}, \{A_2\}} \dots \otimes_{\{A_1, \dots, A_{n-1}\}, \{A_n\}} \pi_{\{A_n\}, R}(\langle t^1, R \rangle)$.

Оскільки після застосування операції проєкції до 1-мультимножини, ми не обов'язково отримуємо 1-мультимножину, то результат різниці мультимножини та 1-мультимножини дасть мультимножину. Отже, множина T не замкнена відносно операції активного доповнення.

Таким чином, множина всіх таблиць табличної алгебри нескінченних таблиць T замкнена відносно операцій перетину, різниці, селекції, з'єднання та перейменування і не замкнена відносно об'єднання, проєкції та активного доповнення. Отже, таблична алгебра нескінченних таблиць не утворює підалгебру мультимножинної табличної алгебри.

Висновки

Дана стаття є продовженням робіт присвячених актуальній проблемі розвитку теоретичної основи табличних баз даних. Основна увага в роботі зосереджена на питанні чи є таблична алгебра нескінченних таблиць підалгеброю мультимножинної табличної алгебри. Застосовуючи теоретико-множинні та логіко-алгебраїчні методи встановлено, що це не так. Таблична алгебра нескінченних таблиць не утворює підалгебру мультимножинної табличної алгебри, бо не є замкненою відносно об'єднання, проєкції та активного доповнення.

Література

1. Редько В.Н., Брона Ю.Й., Буй Д.Б., Поляков С.А. Реляційні бази даних: табличні алгебри та SQL-подібні мови. Київ: Видавничий дім "Академперіодика". 2001. 198 с.
2. Буй Д.Б., Глушко І.М. Числення на розширення сигнатур табличних алгебр: монографія. Ніжин: НДУ ім. М. Гоголя, 2016. 151 с.
3. Богатирьова Ю.О. Теорія мультимножин та її застосування: дис. ... канд. фіз.-мат. наук. К., 2011. 113с.
4. Катленд Н. Вычислимость. Введение в теорию рекурсивных функций. М.: Мир, 1983. 256 с.

References

1. REDKO, V.N. et al. (2001) Relational Databases: Table Algebras and SQL-like Language. Kyiv: Publishing house Academperiodica.
2. BUY, D.B & GLUSHKO, I.M. (2016) Calculi and extensions of table algebras signature. Nizhyn: NDU im. M. Gogol.
3. BOGATYREVA, J.A. (2011) *Multisets theory and its applications*. A Thesis Submitted of the Requirements of the Kyiv National Taras Shevchenko University for the Degree of Doctor of Philosophy. Location: Kyiv National Taras Shevchenko University.
4. CUTLAND, N. (1983) Computability. An introduction to recursive function theory. Moscow: Mir.

Про автора:

Глушко Ірина Миколаївна,
кандидат фіз.-мат. наук,
доцент кафедри інформаційних технологій та аналізу даних
Ніжинського державного університету імені Миколи Гоголя.
Кількість наукових публікацій в українських виданнях – приблизно 70.
Кількість наукових публікацій в зарубіжних виданнях – 8.
<https://orcid.org/0000-0003-2549-5356>.

Місце роботи автора:

Ніжинський державний університет імені Миколи Гоголя,
16600, Україна, Чернігівська обл.,
м. Ніжин, вул. Графська 2.
E-mail: glushkoim@gmail.com,
iryna.glushko@ndu.edu.ua.