

ВЗАЄМОЗВ'ЯЗОК МІЖ РОЗШИРЕНОЮ ТАБЛИЧНОЮ АЛГЕБРОЮ НЕСКІНЧЕННИХ ТАБЛИЦЬ ТА РОЗШИРЕНОЮ МУЛЬТИМНОЖИННОЮ ТАБЛИЧНОЮ АЛГЕБРОЮ

Лисенко Ірина

Ніжинський державний університет імені Миколи Гоголя

Анотація

Робота зосереджена на деяких теоретичних питаннях табличних баз даних. Розглянуто два математичні формалізми, такі як розширена таблична алгебра нескінченних таблиць та розширена мультимножинна таблична алгебра. Дано основні визначення, що стосуються цих формалізмів. Розглянуто взаємозв'язок між розширеною табличною алгеброю нескінченних таблиць та розширеною мультимножинною табличною алгеброю. Доведено, що розширена таблична алгебра нескінченних таблиць замкнена відносно додаткових операцій розширеної мультимножинної табличної алгебри.

Abstract

The paper is focused on some theoretical questions of the table databases. Two mathematical formalisms such as extended table algebra of infinite tables and extended multiset table algebra are considered. Basic definitions referring to these formalisms are given. This paper also is dedicated to the topic of the relationship between extended table algebra of infinite tables and extended multiset table algebra. It is proved that extended table algebra of infinite tables is closed with respect to the additional operations of extended multiset table algebra.

Вступ

Прогрес, досягнутий в області технологій баз даних, в значній мірі базується на реляційній моделі, запропонованій Е. Коддом у 70-х роках ХХ ст. Протягом наступних років вчені розширювали та вдосконалювали реляційну модель даних додаючи нові операції відповідно до потреб мов запитів, забезпечуючи маніпулювання невизначеними та неповними даними, пропонуючи реляційну модель даних з мультимножинною семантикою.

У монографії [1] введено два математичні формалізми: таблична алгебра нескінченних таблиць та мультимножинна таблична алгебра, сигнатури операцій яких також були поповнені додатковими операціями. З огляду на те, що таблиці можна розглядати як множини рядків, а отже 1-мультимножини, постає питання про замкненість табличної алгебри нескінченних таблиць відносно операцій мультимножинної табличної алгебри. Саме цій проблемі і присвячена дана робота.

Розширена таблична алгебра нескінченних таблиць

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

Під *таблицею* схеми R розуміємо пару $\langle t, R \rangle$, де t – множина (зокрема, нескінченна) рядків вказаної схеми R .

Під *розширеною табличною алгеброю нескінченних таблиць* розуміємо алгебру $\langle T, \Omega_{P, \Xi} \rangle$, де T – множина усіх таблиць, $\Omega_{P, \Xi}$ – сигнатура. Сигнатура $\Omega_{P, \Xi}$ містить теоретико-множинні операції (об'єднання \cup_R , перетин \cap_R , різницю \setminus_R), спеціальні операції (селекцію $\sigma_{p, R}$, проекцію $\pi_{X, R}$, з'єднання \otimes_{R_1, R_2} , ділення $\div_{R_2}^{R_1}$, перейменування

$Rt_{\xi,R}$ та активне доповнення \sim_R), операції внутрішнього з'єднання (декартове з'єднання Cj_{R_1,R_2} , внутрішнє з'єднання за атрибутами $\otimes_{A_1,\dots,A_n,R_1,R_2}$, внутрішнє з'єднання за предикатом \otimes_{p,R_1,R_2}), операції зовнішнього з'єднання (зовнішнє ліве з'єднання φ_l , зовнішнє праве з'єднання φ_r , повне зовнішнє з'єднання φ_f та зовнішнє з'єднання об'єднання φ_U), зовнішні множинні операції (зовнішнє об'єднання $\cup_{R_1,R_2}/$, зовнішню різницю $\setminus_{R_1,R_2}/$, зовнішній перетин $\cap_{R_1,R_2}/$), агрегатні операції (знаходження суми $Sum_{A,R}$, мінімального $Min_{A,R}$ та максимального $Max_{A,R}$ значення, кількості значень $Count_{A,R}$, середнього арифметичного $Avg_{A,R}$ та кількості рядків $Count_{A,R}(*))$ та операцію напівз'єднання \times_{R_1,R_2} , де $p \in P, \xi \in \Xi, X, R, R_1, R_2 \subseteq A, P, \Xi$ – множини параметрів. Формальні означення операцій сигнатури подано у монографії [1]. Для задання зовнішніх операцій введено особливий елемент універсального домену $NULL$, а для задання агрегатних операцій використовується поняття мультимножини.

Розширена мультимножинна таблична алгебра

Під мультимножинною табличною алгеброю розуміємо алгебру $\langle \Psi, \Omega_{P,\Xi} \rangle$, де Ψ – множина всіх таблиць, $\Omega_{P,\Xi}$ – сигнатура. Сигнатура $\Omega_{P,\Xi}$ містить теоретико-множинні операції (об'єднання $\cup_{All}^{\Psi,R}$, перетин $\cap_{All}^{\Psi,R}$, різницю $\setminus_{All}^{\Psi,R}$), спеціальні операції ($\sigma_{p,R}, \pi_{X,R}, \otimes_{R_1,R_2}, Rt_{\xi,R}$ та \sim_R), операції внутрішнього з'єднання ($Cj_{R_1,R_2}, \otimes_{A_1,\dots,A_n,R_1,R_2}, \otimes_{p,R_1,R_2}$), операції зовнішнього з'єднання ($\varphi_l, \varphi_r, \varphi_f, \varphi_U$), агрегатні операції ($Sum_{A,R}, Min_{A,R}, Max_{A,R}, Count_{A,R}, Avg_{A,R}, Count_{A,R}(*))$ та операцію напівз'єднання \times_{R_1,R_2} , де $p \in P, \xi \in \Xi, X, R, R_1, R_2 \subseteq A, P, \Xi$ – множини параметрів. Формальні означення операцій сигнатури подано у монографії [2]. До операцій сигнатури розширеної мультимножинної табличної алгебри не входять зовнішні мультимножинні операції. Оскільки визначаючи дані операції над таблицями розширеної мультимножинної табличної алгебри, використовуючи міркування, застосовані при визначенні відповідних операцій над таблицями розширеної табличної алгебри нескінченних таблиць, прийдемо до протиріччя у випадку коли схеми таблиць-аргументів рівні.

Поняття *таблиці* задається як пари $\langle \psi, R \rangle$, де перша компонента ψ – це мультимножина, основою якої $\Theta(\psi)$ є довільна множина зокрема, нескінченна, рядків схеми R , а друга компонента R – схема таблиці. Через $Osc(s, \psi)$ позначається кількість дублікатів (екземплярів) рядка s у мультимножині ψ .

Замкненість розширеної табличної алгебра відносно операцій розширеної мультимножинної табличної алгебра

З'ясуємо чи є розширена таблична алгебра нескінченних таблиць підалгеброю розширеної мультимножинної табличної алгебри. Припустимо, що таблицею розширеної табличної алгебри нескінченних таблиць є пара $\langle t^1, R \rangle$, де t^1 – 1-множина (зокрема, нескінченна) рядків вказаної схеми R . 1-мультимножинами називаються мультимножини, областю значень яких є порожня множина або одноелементна множина [2]. У цьому випадку очевидно, що множина всіх таблиць розширеної табличної алгебри нескінченних таблиць T є підмножиною множини всіх таблиць розширеної мультимножинної табличної алгебри Ψ , $T \subseteq \Psi$.

Тепер визначимо чи замкнена множина T відносно кожної операції сигнатури розширеної мультимножинної табличної алгебри $\Omega_{p,\Xi}$. У статті [3] з'ясовано, що таблична алгебра нескінченних таблиць замкнена відносно операцій перетину, різниці, селекції, з'єднання та перейменування мультимножинної табличної алгебри та не замкнена відносно операцій об'єднання, проєкції та активного доповнення. Встановимо чи замкнена множина T відносно додаткових операцій розширеної мультимножинної табличної алгебри.

Як було вказано вище, до додаткових операцій розширеної мультимножинної табличної алгебри належать операції внутрішнього та зовнішнього з'єднання, агрегатні операції та операція напівз'єднання. Розглянемо детально операцію внутрішнього з'єднання за атрибутами A_1, \dots, A_n .

Розглянемо дві таблиці розширеної табличної алгебри нескінченних таблиць $\langle t_1^1, R_1 \rangle$ та $\langle t_2^1, R_2 \rangle$. Позначимо через $\Theta(t_1^1)$ і $\Theta(t_2^1)$ – основи 1-мультимножин t_1^1 та t_2^1 відповідно. Операція внутрішнього з'єднання за атрибутами A_1, \dots, A_n , де всі A_1, \dots, A_n попарно різні, $n \geq 1$, таблиць схем R_1 та R_2 , де $R_1 \cap R_2 \supseteq \{A_1, \dots, A_n\}$ (передбачається, що спільні атрибути, які відрізняються від атрибутів A_1, \dots, A_n , перед з'єднанням перейменовуються) визначається як $\langle t_1^1, R_1 \rangle_{A_1, \dots, A_n, R_1, R_2} \otimes \langle t_2^1, R_2 \rangle = \langle t', R_1 \cup R_2 \rangle$, в припущенні $\langle t_1^1, R_1 \rangle \in T(R_1)$, $\langle t_2^1, R_2 \rangle \in T(R_2)$.

Основою мультимножини t' є множина рядків $\Theta(t') = \left\{ s \mid \exists s_1 \exists s_2 \left(s_1 \in \Theta(t_1^1) \wedge s_2 \in \Theta(t_2^1) \wedge \bigwedge_{i=1}^n s_i(A_i) = s_2(A_i) \wedge s = s_1 \cup s_2 \right) \right\}$.

Кількість дублікатів знаходиться так: $Occ(s, t') = Occ(s_1, t_1^1) \cdot Occ(s_2, t_2^1)$, де $s \in \Theta(t')$, $s = s_1 \cup s_2$ (очевидно, що цей розклад рядка s єдиний) і $s_1 = s | R_1$, $s_2 = s | R_2$. Оскільки t_1^1 та t_2^1 1-мультимножини, тобто вони не мають дублікатів рядків, тому $Occ(s, t') = Occ(s_1, t_1^1) \cdot Occ(s_2, t_2^1) = 1 \cdot 1 = 1$.

Отже, множина T замкнена відносно операції внутрішнього з'єднання за атрибутами A_1, \dots, A_n . Аналогічно, можна показати, що множина T замкнена відносно інших додаткових операцій розширеної мультимножинної табличної алгебри.

Висновок

У даній роботі розглядається питання взаємозв'язку розширеної табличної алгебри нескінченних таблиць та розширеної мультимножинної табличної алгебри. Встановлено, що розширена таблична алгебра нескінченних таблиць замкнена відносно додаткових операцій мультимножинної табличної алгебри.

Список використаних джерел

1. Буй Д.Б. Числення та розширення сигнатур табличних алгебр: монографія / Д.Б. Буй, І.М. Глушко. – Ніжин: НДУ ім. М. Гоголя, 2016. – 151 с. – ISBN 978-617-527-145-2.
2. Реляційні бази даних: табличні алгебри та SQL-подібні мови / В.Н. Редько, Ю.Й. Брона, Д.Б. Буй, С.А. Поляков. – Київ: Видавничий дім "Академперіодика". – 2001. – 198 с. – ISBN 966-8002-00-8.
3. Глушко І.М. Про зв'язок між табличною алгеброю нескінченних таблиць та мультимножинною табличною алгеброю [Електронний ресурс] / І.М. Глушко // Proceedings of the 11th International Conference of Programming. – С.159–163. – Режим доступу: <http://ceur-ws.org/Vol-2139/159-163.pdf>.