*D. Bui, A. Puzikova*

**THEORY OF NORMALIZATION IN RELATIONAL DATABASES: CURRENT SITUATION**

*Given paper contains some attempt to characterize the current state of the theory of normalization in relational databases. We discuss the definitions of some classical normal forms and basic non-classical normal forms. It is shown that both definitions of Fagin's Projection-Join Normal Form are not equivalent. On the basis of analysis of the sources and own results logical connections between the definitions of classical and basic non-classical normal forms are established.*

**Keywords:** relational databases, normalization, normal forms.

*Матеріал надійшов 15.05.2015*

**УДК 004.655**

*Сенченко О. С.*

# ПРО ПОХІДНІСТЬ У ТАБЛИЧНИХ АЛГЕБРАХ ОПЕРАЦІЇ ПЕРЕЙМЕНУВАННЯ ПРИ СКІНЧЕННОМУ УНІВЕРСАЛЬНОМУ ДОМЕНІ

*У роботі розглянуто питання похідності операції перейменування атрибутів. Показано, що за умови скінченності універсального домену ця операція є похідною відносно операцій проекції, селекції та з'єднання з використанням однієї константної таблиці.*

**Ключові слова:** бази даних, таблична алгебра, похідність операцій.

## Вступ

У наш час системи керування базами даних широко використовуються в багатьох сферах діяльності людини. Найпоширенішою є реляційна модель даних, математична модель яких уперше була запропонована Е. Коддом [5; 6]. Формально реляційна база даних є скінченною множиною скінченних відношень різної арності між заздалегідь визначеними множинами елементарних даних.

Табличні алгебри, що були введені В. Н. Редьком і Д. Б. Буєм [3], побудовані на основі реляційних алгебр Е. Кодда та суттєво їх уточнюють. Вони становлять теоретичний фундамент мов запитів сучасних табличних баз даних. Елементи носія табличної алгебри уточнюють реляційні структури даних, а сигнатурні операції побудовані на базі основних табличних маніпуляцій у реляційних алгебрах та SQL-подібних мовах.

У табличних алгебрах деякі сигнатурні операції є похідними відносно інших операцій. Похідність операції дозволяє виконувати еквівалентні перетворення виразів, які можуть спрощувати ці вирази; також еквівалентні перетворення виразів використовуються для розв'язання актуальної та важливої задачі оптимізації запитів [1; 2; 7].

У монографії [4] доведено, що операція проекції не є похідною відносно інших операцій та довільних констант, а деякі операції табличних алгебр (об'єднання, з'єднання, селекція, перейменування) не є похідними в загальному випадку нескінченності універсального домену та/або множини атрибутів. Проте в реальних базах даних у більшості випадків універсальний домен та множина атрибутів є скінченними (хоча й дуже великими за потужністю). У роботі показано, що у випадку скінченності універсального домену та множини атрибутів операція перейменування може бути вираженою через інші операції та константну таблицю.

## Основні визначення

Зафіксуємо деяку непорожню множину атрибутів $\Re = \{A_1,\dots,A_n\}$. Довільну скінченну підмножину множини $\Re$ назвемо схемою, причому схема може бути порожньою множиною. Рядком $s$ схеми $R$ називається множина пар $s=\{(A_1',d_1),\dots,(A_k',d_k)\}$, проекція якої за першою компонентою дорівнює $R$, причому атрибути $A_1',\dots,A_k'$ є попарно різними, тобто рядок є функціональним бінарним відношенням. Таблиця $T$ схеми $R$ ($T(R)$) – скінченна множина рядків схеми $R$. Виділяють дві особливі таблиці: таблицю $T_\varepsilon = \{\varepsilon\}$, де $\varepsilon$ – порожній рядок (порожня множина, яку інтерпретують як усюди невизначену функцію), при цьому схема таблиці $T_\varepsilon$ є порожньою множиною, та таблицю $T_\varnothing = \varnothing$ – порожню множину рядків довільної схеми.

На множині таблиць однакової схеми (для визначеності – схеми $R$) введено операції об'єднання ($\bigcup_R$), перетину ($\bigcap_R$) та різниці ($-_R$) таблиць як обмеження однойменних теоретико-множинних операцій.

Проекцією за множиною атрибутів $X \subseteq R$ називається унарна параметрична операція $\pi_X$, значенням якої є таблиця, що складається з обмежень по $X$ усіх рядків вихідної таблиці: $\pi_X(T)=\{s|X \;\; |s \in T\}$ (тут обмеження розуміємо стандартно: $s|X = s \bigcap (X \times pr_2\, s)$, де $pr_i\, s$ – проекція рядка $s$ за $i$-ю компонентою).

Селекцією за предикатом $P: S \to \{true, false\}$, де $S$ – множина всіх рядків, називається унарна параметрична операція $\sigma_P$, яка зіставляє таблиці її підтаблицю, що містить рядки, на яких предикат $P$ набуває значення «істина»: $\sigma_P(T)=\{s\,|s \in T \wedge P(s)=true\}$.

Для введення операції з'єднання визначимо допоміжне поняття. Бінарні відношення $\rho$ та $\tau$ називаються сумісними (позначається $\rho \approx \tau$), якщо $\rho|X = \tau|X$, де $X = pr_1\,\rho \bigcap pr_1\,\tau$.

З'єднанням називається бінарна операція $\otimes$, значенням якої є таблиця, що складається з усіх можливих об'єднань сумісних рядків вихідних таблиць, тобто $T_1 \otimes T_2 = \{s_1 \bigcup s_2 |\; s_1 \in T_1 \wedge s_2 \in T_2 \wedge s_1 \approx s_2\}$. З цього означення випливає, що для будь-якої таблиці $T$ виконується рівність $T_\varnothing \otimes T = T \otimes T_\varnothing = T_\varnothing$, та якщо схеми таблиць $T_1(R_1)$ і $T_2(R_2)$ не перетинаються, то їх з'єднання $T_1 \otimes T_2$ складається з усіх можливих рядків вигляду $s_1 \bigcup s_2$, де $s_1 \in T_1$ та $s_2 \in T_2$.

Перейменування атрибутів – унарна параметрична операція $RT_\xi$, де $\xi$ – ін'єктивне відображення вигляду $\xi : R \to R$, що здійснює перейменування атрибутів таблиці відповідно до $\xi$, застосування цієї операції зводиться до перейменування перших компонентів пар – елементів рядків. З означення випливає, що перейменування не змінює особливу таблицю $T_\varepsilon$.

Також на множині таблиць введено операції ділення, насичення та активного доповнення, ці операції не буде використано в роботі, тому їх означення не наводимо. Табличною алгеброю назвемо часткову алгебру з носієм – множиною всіх таблиць (довільних схем) – та наведеними вище операціями.

Операція називається похідною відносно деяких інших операцій, якщо її можна побудувати підстановкою з цих операцій (із можливим використанням деяких селекторів) та константних таблиць. Непохідність операцій найчастіше доводять через виділення деяких замкнених класів табличних операцій, які містять довільні селектори та є замкненими відносно підстановки: якщо деяка операція не входить до деякого замкненого класу, то вона не є похідною відносно операцій, що входять до цього класу.

## Основні результати

У монографії [4, с. 118, твердження 2.17.6] доведено, що за умови нескінченності універсального домену операція перейменування не є похідною відносно інших сигнатурних операцій табличних алгебр та довільних констант. Покажемо, що при скінченному універсальному домені операція перейменування може бути вираженою через операції проекції, селекції та з'єднання з використанням однієї константної таблиці.

Нехай $\Re = \{A_1,\dots,A_n\}$ – скінченна множина атрибутів та $D_{A_i} = \{d_1^i,\dots,d_{n_i}^i\}$ – скінченний домен атрибута $A_i$ ($i=\overline{1,n}$). Нехай $A=\{R_1,\dots,R_k\} \subset \Re$ і $T$ – таблиця схеми $A$, $R_j \in A$ та $B \in \Re - A$. Через $RT_\xi(R_j, B)$ позначимо операцію перейменування, що відповідає такому відображенню $\xi$: $\xi(\mathrm{R}_j) = B \,\&\, \forall A \in \left(R - \{R_j\}\right) \xi(A) = A$. Змістовно кажучи, ця операція перейменовує рівно один атрибут таблиці схеми *A* в деякий атрибут, що не належить цій схемі, при цьому для коректності введення операції повинно виконуватись включення $D_{R_j} \subseteq D_B$. Нескладно бачити, що будь-яка операція перейменування будь-якої скінченної кількості атрибутів може бути вираженою через скінченну кількість разів використання наведеної вище операції перейменування, тобто скінченною композицією операції перейменування одного атрибута.

Справедливе твердження.

**Лема**.

Нехай *U* – таблиця, що складається з усіляких рядків вигляду $\left\{\left(A_1, d_{p_1}^1\right), \ldots, \left(A_n, d_{p_n}^n\right)\right\}$, де $d_{p_i}^i \in D_{A_i}$, та $p(R_j, B)$ – предикат, що приймає значення «істина» лише на таких рядках схеми $A \cup \{B\}$, на яких виконується рівність $s(R_j) = s(B)$. Тоді виконується рівність:

$$RT_{\xi}(R_j, B) = \pi_{\left(R - \{R_j\}\right) \cup \{B\}} \left[ \sigma_{p(R_j, B)} \left( T \otimes \pi_B(U) \right) \right].$$

*Доведення*. Для більш компактного викладення введемо такі позначення: $RT_{\xi}(R_j, B) = T'$ $T \otimes \pi_B(U) = T_1$; $\sigma_{p(R_j, B)}(T_1) = T_2$ та $\pi_{\left(R - \{R_j\}\right) \cup \{B\}}(T_2) = T_3$. За визначенням операції перейменування, з урахуванням введеної операції $RT_{\xi}(R_j, B)$, схемою таблиці $T'$ є множина $\left(R - \{R_j\}\right) \cup \{B\}$. Оскільки $B \cap R = \varnothing$, то за властивостями операції з'єднання схемою таблиці $T_1$ є множина $R \cup \{B\}$, за визначенням селекції схемою таблиці $T_2$ – також множина $R \cup \{B\}$, за визначенням проекції схемою таблиці $T_3$ є множина $\left(R - \{R_j\}\right) \cup \{B\}$, таким чином, ми показали, що таблиці в лівій та правій частинах рівності мають однакову схему.

Окремо доведемо рівність для випадків $T = T_{\varnothing}$ та $T \neq T_{\varnothing}$.

Нехай $T = T_{\varnothing}$. Тоді за властивостями перейменування $T' = T_{\varnothing}$, а за властивостями з'єднання $T_1 = T_{\varnothing}$, тому за властивостями селекції та проекції $T_2 = T_3 = T_{\varnothing}$, тобто в цьому випадку рівність виконується.

Нехай тепер $T \neq T_{\varnothing}$. Доведемо включення $T' \subseteq T_3$. Нехай $s = \left\{(R_1, d_1), \ldots, (R_k, d_k), (B, d_B)\right\} \in T'$. За визначенням операції перейменування атрибутів у цьому випадку в таблиці *T* існує рядок $s' = \left\{(R_1, d_1), \ldots, (R_k, d_k), (R_j, d_j)\right\}$. Оскільки $d_B \in D_{R_j}$, то з включення $D_{R_j} \subseteq D_B$ випливає, що $d_B \in D_B$, тому за побудовою таблиці $U$ $\exists u \in U \left| (B, d_B) \in u \right.$, тоді $(B, d_B) \in \pi_B(U)$. За визначенням з'єднання з приналежностей $s' \in T$ та $(B, d_B) \in \pi_B(U)$ випливає існування в таблиці $T_1$ рядка $s_1 = \left\{(R_1, d_1), \ldots, (R_k, d_k), (R_j, d_B), (B, d_B)\right\}$. Тоді на рядку $s_1$ предикат $p(R_j, B)$ приймає значення «істина», тому $s_1 \in T_2$ і за визначенням проекції рядок $\left\{(R_1, d_1), \ldots, (R_k, d_k), (B, d_B)\right\} \in T_3$, це доводить включення $T' \subseteq T_3$.

Доведемо обернене включення. Нехай $s = \left\{(R_1, d_1), \ldots, (R_k, d_k), (B, d_B)\right\} \in T_3$. За визначенням проекції з приналежності $s \in T_3$ випливає існування в таблиці $T_2$ рядка $s_2 = \left\{(R_1, d_1), \ldots, (R_k, d_k), (B, d_B), (R_j, d_j)\right\}$. Тоді, оскільки $T_2 = \sigma_{p(R_j, B)}(T_1)$, то виконується рівність $d_j = d_B$, тому в таблиці $T_1$ існує рядок $s_1 = \left\{(R_1, d_1), \ldots, (R_k, d_k), (R_j, d_B), (B, d_B)\right\}$. Оскільки $T_1 = T \otimes \pi_B(U)$, а схеми таблиць $T$ та $\pi_B(U)$ не перетинаються, то за властивостями з'єднання з $s_1 \in T_1$ випливає існування в таблиці *T* рядка $s' = \left\{(R_1, d_1), \ldots, (R_k, d_k), (R_j, d_B)\right\}$. За визначенням введеної операції перейменування атрибутів $RT_{\xi}(R_j, B)$ з $s' \in T$ випливає $\left\{(R_1, d_1), \ldots, (R_k, d_k), (B, d_B)\right\} \in RT_{\xi}(R_j, B)$, тобто $s \in T'$, що доводить рівність $T' = T_3$.

З цієї леми очевидно випливає теорема про похідність перейменування.

**Теорема** (про похідність перейменування).

За умови скінченності множини атрибутів таблиць та універсального домену перейменування є похідним відносно проекції, селекції, з'єднання та однієї константної таблиці.

## Висновки

У роботі показано, що за умови скінченності універсального домену операція перейменування може бути вираженою через операції проекції, селекції та з'єднання з використанням однієї константної таблиці. Результати роботи становлять теоретичний та практичний інтерес, оскільки на основі рівностей можна виконувати еквівалентні перетворення виразів, які є необхідними для їх оптимізації, зокрема й для оптимізації запитів у реляційних базах даних.

***Список літератури***

1. Буй Д. Б. Модели, методы и алгоритмы оптимизации запросов в базах данных (обзор) / Д. Б. Буй, В. Г. Скобелев // Радіоелектронні і комп'ютерні системи. – 2014. – № 2. – С. 43–58.
2. Мендекович Н. А. Обзор развития методов лексической оптимизации запросов / Н. А. Мендекович, С. Д. Кузнецов // Труды ИСП РАН. – 2012. – Т. 23. – С. 195–214.
3. Редько В. Н. К основаниям теории реляционных моделей баз данных / В. Н. Редько, Д. Б. Буй // Кибернетика и системный анализ. – 1996. – № 4. – С. 3–12.
4. Реляційні бази даних: табличні алгебри та SQL-подібні мови / В. Н. Редько, Ю. Й. Брона, Д. Б. Буй, С. А. Поляков. – К. : Академперіодика, 2001. – 198 с.
5. Codd E. F. A Relational Model of Data for Large Shared Data Banks / E. F. Codd // Comm. of ACM. – 1970. – Vol. 13, № 6. – P. 377–387.
6. Codd E. F. The Relational Model for Database Management: version 2 / E. F. Codd. – Addison-Wesley, 1990. – 538 p.
7. Knuth D. E. The Art of Computer Programming, Volume 4, Fascicle 0: Introduction to Combinatorial Algorithms and Boolean Functions / D. E. Knuth. – Addison-Wesley, 2008. – 240 p.

*O. Senchenko*

## ABOUT DERIVATION OF RENAMING IN TABLE ALGEBRAS WHEN UNIVERSAL DOMAIN IS FINITE

*In this paper the derivation of renaming in table algebras is considered. It is proved that when universal domain is finite, this operation is derived with respect to the operations of projection, selection and join, with the use of one constant table.*

**Keywords:** database, table algebra, derivation.

*Матеріал надійшов 30.04.2015*

**УДК 004.655**

*Буй Д. Б., Глушко І. М.*

# РОЗШИРЕННЯ СИГНАТУР РЕЛЯЦІЙНИХ (ТАБЛИЧНИХ) АЛГЕБР КОДДА: СУЧАСНИЙ СТАН

*Статтю присвячено огляду літератури з розширень реляційної (табличної) алгебри. Розглянуто додаткові операції, що розширюють можливості реляційної алгебри: агрегування, групування, сортування, напівз'єднання, зовнішні з'єднання та ін. Проаналізовано використання null-значень при оперуванні невизначеною та неповною інформацією в базах даних. Оскільки існує ряд прикладних задач, особливістю яких є множинність і повторюваність даних, також приділено увагу питанню розширення можливостей баз даних за рахунок використання мультимножин.*

**Ключові слова:** реляційні бази даних, реляційна (таблична) алгебра, розширення реляційної алгебри, мультимножина.

### Вступ

Системи управління базами даних (СУБД) у наш час використовуються практично в усіх сферах людської діяльності, які пов'язані зі збереженням та обробкою інформації. Прогрес, досягнутий у галузі технологій баз даних (БД), значною мірою ґрунтується на реляційній моделі,