УДК 004.658.2

Буй Д. Б., Пузікова А. В.

ТЕОРІЯ НОРМАЛІЗАЦІЇ В РЕЛЯЦІЙНИХ БАЗАХ ДАНИХ: СУЧАСНИЙ СТАН

Здійснено спробу охарактеризувати сучасний стан теорії нормалізації в реляційних базах даних. Обговорено означення деяких класичних нормальних форм та основних некласичних нормальних форм. Показано нееквівалентність двох означень проективно-з'єднувальної нормальної форми, запропонованих Р. Фейгіним (R. Fagin). На основі аналізу першоджерел та власних результатів встановлено логічні зв'язки між означеннями класичних та основних некласичних нормальних форм.

Ключові слова: реляційні бази даних, нормалізація, нормальні форми.

Логіко-історичний огляд розвитку теорії нормалізації

Уперше термін «нормалізація» був застосований у 1970 р. Е. Коддом для назви процедури усунення непростих доменів [24, с. 381]. У наступні роки серед професіоналів розв'язалась полеміка щодо поняття «непростого домену», що впливало на інтерпретацію першої нормальної форми (1НФ) (див., наприклад, [4; 30; 5]). Характерний факт: можливість використовувати як елементи домену групи (groups) та відношення відстоювали автори праць [37; 42].

У 1971 р. Е. Кодд у праці [25] вказує на надлишковість даних та аномалії, які виникають при здійсненні операцій над відношеннями, вперше представляє концепцію функціональної залежності (ФЗ), демонструє можливість її використання для розв'язання проблем проектування баз даних (БД), наводить означення другої нормальної форми (2НФ), транзитивної ФЗ та третьої нормальної форми (3НФ). У науковій літературі можна натрапити на різні їх інтерпретації. Змістовне означення 2НФ використовується до сьогодні: «Немає непервинного атрибута у відношенні, який би функціонально залежав від повної підмножини потенційного ключа». Надалі повна ФЗ виділяється в окреме поняття, яке використовується у визначенні 2НФ [5; 6; 8].

Означення транзитивної ФЗ, сформульоване Коддом у вигляді: «Нехай A, B і C – різні атрибути (або різні множини атрибутів) у відношенні і нехай вони задовольняють такі умови: виконуються ФЗ $A \to B$, $B \to C$ та не виконується ФЗ $B \to A$. Тоді ФЗ $A \to C$ називається транзитивною залежністю», на сьогодні має два уточнення:

1) атрибут $C \notin A \cup B$, де A і B – множини атрибутів [8, с. 111]. Ця умова виключає з розгляду тривіальні ФЗ[1];

2) $C \nsubseteq B$, $B \nsubseteq A$, де A, B і C – множини атрибутів [6, с. 24]. Цим також виключаються тривіальні ФЗ у посилках $A \to B$ і $B \to C$.

Залежно від використовуваних уточнень означення 3НФ, сформульоване Коддом у вигляді: «Кожний непервинний атрибут не є транзитивно залежним від будь-якого потенційного ключа у відношенні», на сьогодні має два варіанти.

1) Відношення перебуває у 3НФ, якщо воно перебуває у 1НФ і не містить транзитивних залежностей (з наведеним вище уточненням 1)) непервинних атрибутів від потенційних ключів [8]. Додатково доводиться теорема, яка стверджує, що з перебування відношення у 3НФ випливає перебування цього відношення у 2НФ [8, с. 111].

2) Відношення перебуває у 3НФ, якщо воно перебуває у 2НФ і не містить транзитивних залежностей (з наведеним вище уточненням 2)) непервинних атрибутів від потенційних ключів (див., наприклад, [5]).

Відкриття аксіом та правил виведення ФЗ із заданої множини ФЗ [29] та побудова аксіоматики Армстронга для ФЗ [16] дали можливість розробити алгоритми обчислення так званого канонічного покриття для заданої множини ФЗ та замикання для множини атрибутів [8; 20; 12; 3].

Одним із способів приведення відношення до 3НФ є декомпозиція, обґрунтуванням якої стала теорема Хеза (Heath) [35]. До підводних каменів декомпозиції належить залежність проекцій (за термінологією Ріссанена (Rissanen) [46]), яка може стати причиною аномалій. Цю проблему

[1] Нагадаємо, що тривіальною ФЗ називається ФЗ вигляду $X \to Y$, де $X \subseteq Y$. Тривіальні залежності виконуються на довільній таблиці.

розв'язано у статті А. Філіпповича [13], який розглянув взаємні ФЗ (ВФЗ) і умовні взаємні ФЗ (УВФЗ) та дослідив їхні властивості, запропонував алгоритм виявлення ВФЗ, поняття взаємнонезалежної нормальної форми (ВННФ), яку можна розглядати як синонім ациклічної БД, та спосіб зведення до неї.

Інший спосіб запропонував Бернштейн (Bernstein), який використав альтернативне означення 3НФ (відношення перебуває у 3НФ, якщо кожний атрибут, який транзитивно залежить від ключа, є первинним атрибутом) для побудови алгоритму синтезу повної схеми БД у 3НФ для заданої множини ФЗ [21].

Специфіка різних підходів до задачі проектування схеми реляційної БД викликана відмінностями у формальних визначеннях еквівалентності та критеріїв якості схеми [19].

Відомими класичними алгоритмами зведення схеми відношення до 3НФ є алгоритми Ульмана [12], Делобеля – Гейсі (Delobel – Gasey) [29], результатами яких не завжди є схема у 3НФ, Бернштейна [21], Іслура (Isloor) [36], Неклюдової – Цаленка [9], який дає кількісно оптимальну схему БД; оригінальний алгоритм зведення схеми відношення до 3НФ через побудову кільцевих покриттів запропонував Мейєр [8]. Переваги та недоліки більшості з вказаних алгоритмів, а також їх відповідність різним визначенням еквівалентності реляційних схем розглянуто в монографії В. Дрібаса [6]. Пошук ефективних алгоритмів розв'язання задачі синтезу оптимальної схеми БД у 3НФ триває і сьогодні (наприклад, [7; 2; 3]).

Недоліки 3НФ було розглянуто і враховано в роботі Хеза [35] при формулюванні означення посиленої 3НФ та пізніше в роботі Кодда [26] (інша назва НФ – нормальна форма Бойса – Кодда (НФБК)). Одним із перших відомих алгоритмів зведення «майже» до НФБК є алгоритм Бернштейна [21], який дозволяє усувати транзитивні залежності первинних атрибутів від ключів, що не містять цих атрибутів. Більш пізні алгоритми наводяться, наприклад, у працях [39; 17].

Із введенням К. Заніоло (Zaniolo) у розгляд багатозначних залежностей (БЗЗ) [49] Р. Фейгін (Fagin) досліджує їхні властивості та визначає нову четверту нормальну форму (4НФ) [32, с. 267]. Строгий та повний набір правил виведення для БЗЗ, а також правила, які пов'язують ФЗ та БЗЗ, представлено у статті [18], незалежність побудованої системи аксіом обговорюється в роботі Мендельзона (Mendelzon) [43], а проблеми її повноти та ненадлишковості – у статті Біскапа (Biskap) [22].

Залежності з'єднання (ЗЗ) та аномалії, які вони викликають, було розглянуто у статті Рісcанена [46] та досліджено в працях [15; 27].

З результату про відсутність повної скінченної множини правил виведення для ЗЗ, який наведено в роботі С. Петрова [10], випливає неможливість побудови скінченної повної та коректної аксіоматики для ЗЗ.

Концепцію проективно-з'єднувальної нормальної форми (Projection-Join Normal Form – PJ/NF) або класичної п'ятої нормальної форми (5НФ) Р. Фейгін розглянув у статті [33], запропонувавши два означення. У першому – використовується алгоритм приналежності (Membership algorithm), який працює таким чином: на вхід подаються множина потенційних ключів $\Delta = \{K_1, K_2, ..., K_m\}$, відношення із множиною атрибутів R реляційної схеми[2] та ЗЗ $*(R_1, R_2, ..., R_n)$. Ініціюється множина підсхем $S = \{R_1, R_2, ..., R_n\}$. Поки для деякого K_i, $1 \le i \le m$, виконується включення $K_i \subseteq R_j \bigcap R_k$, де $1 \le j, k \le n$, в множині S входження множини R_j замінюється на множину $R_j \bigcup R_k$, а входження R_k знищується. Якщо на виході множина S «збігається» зі схемою R, тобто $S = \{R\}$, то алгоритм приналежності видає значення «*True*». Приймається без доведення (на прикладі), що алгоритм видає результат «*True*» тоді й лише тоді, коли ЗЗ $*(R_1, R_2, ..., R_n)$ логічно (семантично) випливає з множини ключів $\Delta = \{K_1, K_2, ..., K_m\}$. Таким чином, згідно з [33], такі означення є еквівалентними:

1) «схема відношення перебуває у PJ/NF, якщо вона перебуває у 1НФ і для множини ключів Δ та кожної ЗЗ $\sigma \in \Sigma$ алгоритм видає результат "*True*"»;

2) «схема відношення перебуває у PJ/NF, якщо вона перебуває у 1НФ і кожна ЗЗ $\sigma \in \Sigma$ є логічним (семантичним) наслідком множини ключів Δ ($\Delta \vDash \sigma$)».

Перше означення, по суті, є синтаксичним, оскільки застосовується до схеми відношення, а друге – семантичним, оскільки в ньому неявно використовується поняття моделі множини ключів Δ, причому на цій моделі повинна виконуватись кожна залежність з'єднання $\sigma \in \Sigma$. Проаналізуємо детальніше, чи справді ці означення є еквівалентними.

Лема 1. Якщо таблиця t перебуває у PJ/NF за означенням 1), то вона перебуває у PJ/NF за означенням 2).

[2] Зауважимо, що в роботі [33] під реляційною схемою мається на увазі пара $\langle R, \Sigma \rangle$, де R – множина атрибутів, Σ – множина обмежень, яка складається з ФЗ, БЗЗ та ЗЗ і є замкненою відносно операції логічного (семантичного) слідування. Також нагадаємо, що в ЗЗ вигляду $*(R_1, R_2, ..., R_n)$ множини атрибутів (підсхеми) називаються компонентами ЗЗ.

Доведення. Нехай таблиця t є моделлю реляційної схеми $\langle R, \Sigma \rangle$, яка перебуває у PJ/NF за означенням 1). Тоді для довільної ЗЗ $\sigma \in \Sigma$, представленої у вигляді $*(R_1, R_2, ..., R_n)$, на деякому кроці виконання алгоритму приналежності для деякого потенційного ключа $K_i \in \Delta$, $1 \le i \le m$, знайдуться такі компоненти R_j і R_k, $1 \le j, k \le n$, що виконується включення $K_i \subseteq R_j \bigcap R_k$. Звідси за п. 2 твердження 2.10.4 монографії [11, с. 74] випливає, що виконується рівність $\pi_{R_j \cup R_k}(t) = \pi_{R_j}(t) \otimes \pi_{R_k}(t)$.

Вище π_X позначає операцію проекції за множиною X таблиці схеми R, $\otimes$ – операцію природного з'єднання (natural join, за термінологією [11, с. 32] – з'єднання) таблиць.

Продовжуючи застосовувати вказаний пункт для кожного наступного включення вигляду $K_i \subseteq R_j \bigcap R_k$, $1 \le i \le m$, в результаті виконання алгоритму приналежності з урахуванням властивостей комутативності та асоціативності з'єднання [11] отримаємо рівність:

$$\pi_{R_1 \cup ... \cup R_n}(t) = \pi_{R_1}(t) \otimes \pi_{R_2}(t) \otimes ... \otimes \pi_{R_n}(t),$$

оскільки $R = R_1 \bigcup ... \bigcup R_n$,

то $\pi_R(t) = t = \pi_{R_1}(t) \otimes \pi_{R_2}(t) \otimes ... \otimes \pi_{R_n}(t)$.

Вище використали властивість проекції $\pi_X(t) = t$ за умови $R \subseteq X$, де R – схема таблиці t (п. 2 твердження 2.4.1 з [11]).

Звідси випливає, що ЗЗ $\sigma \in \Sigma$ виконується на таблиці t, а отже, реляційна схема $\langle R, \Sigma \rangle$ перебуває у PJ/NF за означенням 2).

Покажемо, що обернена імплікація не виконується; тобто, якщо таблиця *t* перебуває у PJ/NF за означенням 2), то вона необов'язково перебуває у PJ/NF за означенням 1). Для цього достатньо розглянути такий контрприклад.

Приклад 1. Нехай над множиною атрибутів $R = \{A, B, C\}$ з потенційним ключем $\{A\}$ задана ЗЗ $*(ABC, C)$. Очевидно, що на довільній таблиці $t(R)$, яка є моделлю множини ключів $\Delta = \{\{A\}\}$, вказана ЗЗ виконується, оскільки вона є тривіальною [3]. Але алгоритм приналежності видасть результат «*False*».

Таким чином, доведено таке твердження.

Твердження 1. Якщо схема відношення (таблиці) перебуває у PJ/NF за означенням 1), то вона перебуває у PJ/NF за означенням 2). Обернена імплікація в загальному випадку не виконується.

Отже, можемо зробити висновок, що означення 1) для PJ/NF є більш сильним, ніж означення 2).

Викладенню результатів з теорії залежностей та теорії нормалізації для реляційних БД присвячено роботи Д. Мейєра [8], В. Дрібаса [6], М. Цаленка [14].

Некласичні нормальні форми

Означення класичних НФ у реляційних БД застосовуються до одного відношення [24; 5], схеми одного відношення [8; 12; 32; 33], таблиці [11]. Алгоритм зведення відношення до певної НФ повинен задовольняти вимоги (принципи) задачі проектування схеми БД. Одним із важливих принципів розробки схеми БД є принцип представлення [19], який включає в себе можливість відновлюваності початкового відношення r_0 зі схемою R_0 (множиною атрибутів) з відношень $r_1, r_2, ..., r_n$ зі схемами $R_1, R_2, ..., R_n$ відповідно, де r_i – проекції відношення r_0 на схеми R_i для $i = \overline{1, n}$; інакше кажучи, вимагається виконання декомпозиції без втрат $r_0 = \otimes_{i=1}^n r_i$.

Передумовою виникнення пропозицій щодо покращення наявних класичних НФ 2-3 порядків та НФБК є те, що ні декомпозиція (див., наприклад, [13]), ні деякі алгоритми синтезу (наприклад, алгоритм Бернштейна [21]) не забезпечували виконання принципу представлення для відношення з довільно заданими схемою і множиною ФЗ.

Так, якщо на вхід алгоритму синтезу Бернштейна подається реляційна схема $R_0\{A, B, C, D\}$ із множиною ФЗ $\{A \to B, C \to D\}$ (приклад із монографії В. Дрібаса [6, с. 44]), результатом виконання алгоритму є схеми $R_1\{A, B\}$ і $R_2\{C, D\}$, кожна з яких перебуває у 3НФ, але порушується вимога виконання декомпозиції без втрат, оскільки з'єднанням відповідних відношень r_1 і r_2 є їх декартове з'єднання.

Намагаючись уникнути подібної проблеми, Дрібас поширює означення НФ 2-3 порядків та НФБК на набір проекцій відношень, наприклад: «3НФ відношення $r(R)$ – це або дане відношення, якщо воно перебуває у 2НФ і не містить транзитивних залежностей непервинних атрибутів від потенційних ключів, або набір яких-небудь його проекцій, кожна з яких задовольняє вказані умови, причому дане відношення повинно відновлюватись через натуральне з'єднання цих проекцій» [6, с. 25].

Іншу спробу поєднати вимогу виконання декомпозиції без втрат з означенням 3НФ запропонували Лінг (Ling), Томпа (Tompa) і Камеда (Kameda), які розробили концепцію покращеної 3НФ (An Improved Third Normal Form) та алгоритм зведення до неї [40].

Відомо, що алгоритм синтезу Бернштейна моделює схеми відношень, які перебувають «майже» в НФБК. Маючи на меті охарактеризувати

[3] Нагадаємо, що ЗЗ є тривіальною, якщо один з її компонентів R_i збігається зі схемою *R*. Тривіальні ЗЗ виконуються на довільній таблиці. Це випливає з рівностей $\pi_R(t) \otimes \pi_X(t) = t \otimes \pi_X(t) = t$, де *R* – схема *t*; доведення випливає з властивостей з'єднання (підрозділ 2.7, с. 49–56 з [11]).

форму цих відношень, Заніоло запропонував НФ, яка визначається елементарним ключем (Elementary Key Normal Form – EKNF) [50]. На зв'язок EKNF з класичними НФ вказують імплікації: НФБК$\Rightarrow$EKNF$\Rightarrow$3НФ [50].

Перед тим як перейти до наступного блоку некласичних НФ, пов'язаних із означеннями PJ/NF або класичної 5НФ, розглянемо детальніше визначення 5НФ, сформульоване Дейтом [5], та попередню версію 5НФ у монографії Мейєра [8, с. 151]. Наведемо обидва визначення:

1) змінна-відношення перебуває в 5НФ (PJ/NF) за означенням, якщо кожен компонент кожної нетривіальної ЗЗ є суперключем (Дейт);

2) схема відношення R перебуває в попередній PJ/NF за означенням, якщо кожна ЗЗ $*(R_1, R_2, ..., R_n)$ ($R_1 \cup R_2 \cup ... \cup R_n = R$), яка виводиться (синтаксично слідує) з множини обмежень Σ, є або тривіальною, або її компоненти є суперключами (Мейєр).

Використання поняття суперключа в цих означеннях обґрунтовується такою лемою.

Лема 2. Якщо схема відношення $\langle R, \Sigma \rangle$ перебуває у PJ/NF за означенням 1), то кожен компонент довільної ЗЗ $\sigma \in \Sigma$ є суперключем.

Доведення проводиться безпосередньо.

Наслідок 1. Якщо схема відношення $\langle R, \Sigma \rangle$ перебуває у PJ/NF за означенням 1), то кожен атрибут (точніше кажучи, відповідний синглтон) з множини R є суперключем [4].

Доведення. Нехай схема відношення $\langle R, \Sigma \rangle$ перебуває у PJ/NF за означенням 1) і нехай ЗЗ $\sigma \in \Sigma$ представлена у вигляді $*(R_1, R_2, ..., R_n)$. Далі враховуємо, що множина ФЗ і ЗЗ Σ є замкненою відносно операції логічного (семантичного) слідування і виконується включення $\{\sigma \mid \Sigma \models \sigma\} \supseteq \{\varphi \mid \Sigma \vdash \varphi\}$, яке випливає з результату про коректність відповідної аксіоматизації для ЗЗ [10]. Вище $\vdash$ – відношення синтаксичного слідування. Звідси випливає, що $\forall \varphi (\Sigma \vdash \varphi \Rightarrow \varphi \in \Sigma)$.

Нехай $A \in R$, тоді ЗЗ $*(R_1, R_2, ..., R_n, \{A\})$, яка отримана за правилом виведення $\dfrac{*(R_1, R_2, ..., R_n)}{*(R_1, R_2, ..., R_n, X)}$, де $\varnothing \neq X \subseteq R$, належить множині Σ, а отже, за лемою 2 одноелементна множина $\{\ \}$ є суперключем.

Наведемо приклад з [8], на якому Мейєр демонструє нееквівалентність означень попередньої PJ/NF і класичної PJ/NF за означенням 1):

«Нехай задані схема відношення $R = \{A, B, C\}$ і множина обмежень, яка містить ФЗ і ЗЗ: $\Sigma = \{A \to BC, C \to AB, *(AB, BC)\}$. Оскільки множини $\{A, B\}$ і $\{B, C\}$ є суперключами, то схема відношення перебуває в попередній PJ/NF, але вона не перебуває в класичній PJ/NF, оскільки перетин $\{A, B\} \cap \{B, C\} = \{B\}$ не є ключем».

Насправді множина $\{B\}$ в цьому прикладі є ключем; це випливає з такої леми, ідею якої неявно викладено в роботі Вінсента (Vincent) [47].

Лема 3. Якщо на таблиці t схеми R для множин атрибутів R_1, R_2, таких, що $R_1 \cup R_2 = R$, виконується нетривіальна ЗЗ $*(R_1, R_2)$ і R_1, R_2 є суперключами, то $R_1 \cap R_2$ є суперключем.

Доведення. Нехай на таблиці t виконується нетривіальна ЗЗ $*(R_1, R_2)$, де R_1, R_2 – суперключі, і нехай $R_1 \cap R_2 = X$. Розглянемо випадки:

– $X = \varnothing$; тоді з'єднанням проекцій $\pi_{R_1}(t)$ і $\pi_{R_2}(t)$ є їх декартове з'єднання, моделями такої ЗЗ можуть бути порожня таблиця ($t_{\varnothing}$) або однорядкова таблиця, для яких множина $X = \varnothing$ є єдиним ключем [1];

– $X \neq \varnothing$; тоді за теоремою 1 Фейгіна [32, с. 266] з виконання ЗЗ $*(R_1, R_2)$ випливає виконання БЗЗ $X \to\to R_1 \setminus X$. Оскільки R_2 є суперключем, то виконується ФЗ $R_2 \to R_1 \setminus X$. Звідси за спільним правилом для ФЗ і БЗЗ $\dfrac{X \to\to Z,\ Y \to Z'}{X \to Z'}$, де $Z' \subseteq Z$, $Y \cap Z = \varnothing$, випливає виконання ФЗ $X \to R_1 \setminus X$. За правилом поповнення для ФЗ маємо виконання ФЗ $X \to R_1$; звідси за правилом транзитивності з урахуванням ФЗ $R_1 \to R$ (бо R_1 – суперключ) випливає, що ФЗ $X \to R$ також виконується.

Отже, $R_1 \cap R_2$ є суперключем.

Лема 4. Якщо відношення перебуває у PJ/NF за означенням 1) Фейгіна, то воно перебуває в попередній PJ/NF за означенням Мейєра та 5НФ за означенням Дейта.

Доведення випливає безпосередньо з леми 2.

Вінсент у роботі [47] доводить лему про те, що відношення перебуває у 5НФ (за означенням Дейта) або попередній PJ/NF (за означенням

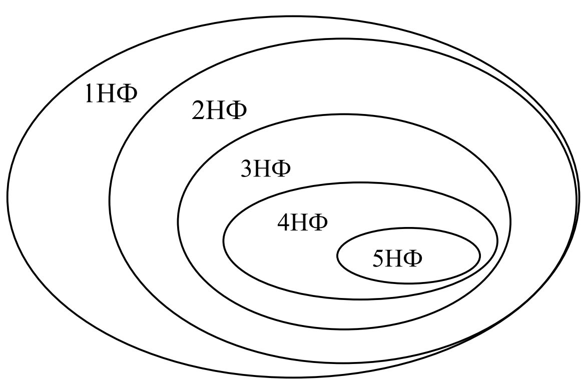

Рис. 1. Співвідношення між НФ

[4] Нагадаємо означення суперключа: «Суперключем є множина атрибутів $X \subseteq R$, яка містить ключ. Кожен ключ є також суперключем» [21, с. 279].

Мейєра) тоді й лише тоді, коли кожен атрибут множини R є ключем.

Вкажемо суттєвий недолік означень Дейта і Мейєра, а також означення 1) для PJ/NF. У формулюванні означення 2) для PJ/NF Фейгін дотримується принципу «включення» НФ вищих порядків до НФ нижчих порядків (див. рис. 1).

Нагадаємо, що в роботі [33] Фейгін сформулював означення класичних НФБК та 4НФ у семантичному вигляді та довів їх еквівалентність з відповідними класичними означеннями НФ, сформульованими в синтаксичній формі. Так, за означенням Фейгіна, «схема відношення перебуває у НФБК (4НФ), якщо вона перебуває у 1НФ і кожна ФЗ (БЗЗ) $\sigma \in \Sigma$ є логічним наслідком множини ключів Δ ($\Delta \models \sigma$)». З того, що ФЗ є окремим випадком БЗЗ, яка своєю чергою є окремим випадком ЗЗ, випливають імплікації: PJ/NF (за означенням 2)) $\Rightarrow$ 4НФ $\Rightarrow$ НФБК. Звідси очевидно, що означення 2) PJ/NF еквівалентно означенню 4НФ за умови, що всі ЗЗ є БЗЗ.

Синтаксичні форми означень Дейта і Мейєра, як і означення 1) для PJ/NF, не є узагальненням 4НФ, тобто не еквівалентні означенню 4НФ, якщо кожна ЗЗ у множині залежностей є БЗЗ [47]. Скористаємось прикладом, наведеним у [47].

Приклад 2. Нехай $R = \{A, B, C\}$ і нехай $\Sigma = \{A \to B, A \to C\}$. Очевидно, що реляційна схема $\langle R, \Sigma \rangle$ перебуває у 4НФ, але не перебуває у 5НФ (за означеннями Дейта або Мейєра), оскільки за правилом виведення $\dfrac{*(R_1, R_2, ..., R_n)}{*(R_1, R_2, ..., R_n, X)}$, де $\varnothing \neq X \subseteq R$, з ЗЗ $*(AC, AB)$ синтаксично слідує ЗЗ $*(AC, AB, C)$, компонент C якої не є ключем.

Враховуючи недоліки наведених означень, Вінсент пропонує зменшити множину ЗЗ, які розглядаються, і запроваджує спрощену 5НФ (reduced-5NF – 5NFR) [47]. «Нехай Σ – множина ФЗ і ЗЗ, причому кожна ЗЗ $*(R_1, R_2, ..., R_p)$ задовольняє умови:

1. $R_1 \cup R_2 \cup ... \cup R_p = R$, де R – схема відношення;

2. $\forall i,\ 1 \le i \le p$, $*(R_1, R_2, ..., R_{i-1}, R_{i+1}, ..., R_p)$ або не належить Σ^+, або $R_1 \cup R_2 \cup ... \cup R_{i-1} \cup R_{i+1} \cup ... \cup R_p \neq R$, де Σ^+ – множина всіх залежностей, які випливають (синтаксично) з множини Σ.

Пара $\langle R, \Sigma \rangle$ перебуває у 5NFR за означенням, якщо ліва частина кожної її нетривіальної ФЗ є суперключем і для кожної ЗЗ $\sigma \in \Sigma$ кожен компонент σ є суперключем».

Альтернативне означення запропонував Норманн (Normann), який розглянув поняття нескоротної ЗЗ [44]: ЗЗ $*(R_1, R_2, ..., R_n)$ є нескоротною, якщо не існує підмножини $\{R_{i_1}, R_{i_2}, ..., R_{i_s}\} \subset \{R_1, R_2, ..., R_n\}$, такої, що ЗЗ $*(R_{i_1}, R_{i_2}, ..., R_{i_s})$ належить Σ^+. Тоді реляційна схема $\langle R, \Sigma \rangle$ перебуває у НФ, яка визначається суперключем (Superkey Normal Form – SKNF), за означенням, якщо кожен компонент кожної нескоротної ЗЗ є суперключем [45].

Лема 5. Якщо схема відношення (таблиці) перебуває в попередній PJ/NF за означенням Мейєра, то вона перебуває у 5NFR за означенням Вінсента та у SKNF за означенням Нормана.

Доведення випливає безпосередньо з означень.

У роботі [48] Вінсент запропонував одразу дві НФ: НФ, яка повністю визначається ключем (ключами) (Key-Complete Normal Form – KCNF), і надлишково-вільну НФ (Redandancy-free normal form – RFNF).

В означенні KCNF використовується поняття залежності з'єднання, яка *повністю визначається ключами* (Key-Complete Join Dependences): «Нехай множина Σ складається з ФЗ і ЗЗ. ЗЗ $*(R_1, R_2, ..., R_p)$ *повністю визначається ключами*, якщо об'єднання тих її компонентів R_i, $1 \le i \le p$, які є суперключами, збігається з R».

За означенням [48] пара $\langle R, \Sigma \rangle$ перебуває у KCNF, якщо «ліва частина кожної нетривіальної ФЗ $\phi \in \Sigma$ є суперключем, а кожна ЗЗ $\sigma \in \Sigma$ повністю визначається ключами».

Положення KCNF і 5NFR по відношенню до інших НФ встановлюють імплікації: PJ/NF (за означенням 1)) $\Rightarrow$ 5NFR $\Rightarrow$ KCNF $\Rightarrow$ 4НФ [48].

В означенні RFNF використовується поняття *надлишковості* для входження значення змінної: «Входження значення змінної $t(A)$ кортежу t відношення r є надлишковим, якщо в результаті заміни $t(A)$ на довільне значення a', таке, що $t(A) \neq a'$, отримане відношення r' не належить множині $SAT(\Sigma)$, де $SAT(\Sigma)$ – множина екземплярів (моделей) множини Σ, яка складається з ФЗ і ЗЗ». Відповідно визначається надлишкове відношення: «Відношення r є надлишковим, якщо існує входження $t(A)$ значення змінної в r, яке є надлишковим». За означенням [48]: «Пара $\langle R, \Sigma \rangle$ перебуває в надлишково-вільній нормальній формі, якщо не існує відношення $r(R) \in SAT(\Sigma)$, яке є надлишковим».

Доводиться теорема про еквівалентність RFNF та 4НФ у випадку, коли множина обмежень Σ складається лише з ФЗ і БЗЗ. Для випадку, коли об'єднання компонентів $R_1, ..., R_p$ кожної ЗЗ $\sigma \in \Sigma$ збігається з R, доводиться теорема про еквівалентність RFNF і KCNF.

Описаний підхід до нормалізації відношення в термінах ненадлишковості кортежів привернув увагу Х'ю Дарвена (Darwen), Кріса Дейта і Рональда Фейгіна. У 2012 р. вони розробили НФ з необхідними кортежами (Essential Tuple Normal Form – ETNF), яка визначається в термінах ФЗ і ЗЗ та займає проміжне положення між 4НФ та PJ/NF (за означенням 1)) [28]. В означенні ETNF використовуються додаткові означення частково надлишкового кортежу та повністю надлишкового кортежу, які є спеціальними випадками поняття надлишковості [48]. На положення ETNF по відношенню до інших НФ вказують імплікації: PJ/NF (за означенням 1)) $\Rightarrow$ SKNF $\Rightarrow$ RFNF $\Rightarrow$ ETNF $\Rightarrow$ 4НФ, причому зворотні імплікації не виконуються [28].

Окремим видом ЗЗ є вкладені залежності (ВЗ), для яких пропонувались інші НФ. Ці НФ у цій роботі не розглядаються. Зауважимо лише, що дослідженню вкладених залежностей, їхніх властивостей та взаємодії з ФЗ присвячено багато робіт. Зокрема, аксіоматику для ВЗ викладено в [23]; відомими є НФ Ніколаса (Nicolas) для вкладених ЗЗ [44], Inclusion Normal Form та алгоритми зведення до неї Лінґа (Ling) і Ґоха (Goh) [41], Inclusion Dependency Normal Form Левена (Levene) [38].

Наостанок розглянемо доменно-ключову НФ (Domain-Key Normal Form – DK/NF або ДКНФ), запропоновану Фейгіним у роботі [31]. Концепція DK/NF базується на поняттях залежності ключа та залежності домену: «Відношення перебуває у DK/NF за означенням, якщо кожне обмеження з його схеми є логічним наслідком з об'єднання множин залежності ключа та залежності домену». Доводиться, що реляційна схема перебуває у DK/NF тоді й лише тоді, коли немає аномалій вставки та знищення (означення яких формулюються в роботі). Показується, що за умови нескінченності потужності всіх доменів DK/NF включається в PJ/NF (за означенням 2)), 4НФ та НФБК. Але під обмеженнями в означенні ДКНФ розуміються не тільки ФЗ, БЗЗ та ЗЗ, означення яких використовуються в попередніх НФ, а й усі, які можна записати у вигляді висловлювання логіки предикатів 1-го порядку. У роботі [34] пропонується форма запису для всіх відомих на той час залежностей:

$$(\forall x_1 \ldots x_m)((A_1 \wedge \ldots \wedge A_n) \Rightarrow \exists y_1 \ldots y_r (B_1 \ldots B_s)),$$

де A_i – атомарна реляційна формула для представлення входження індивідних змінних $z_1 \ldots z_d$ у d -арне відношення P вигляду $P_{z_1 \ldots z_d}$; B_i – або атомарна реляційна формула, або рівність $x = y$, де x і y – індивідні змінні. Зокрема, вбудовані БЗЗ (ВБЗЗ) можна задати у вигляді:

$$(\forall ab_1b_2c_1c_2d_1d_2)((Pab_1c_1d_1 \wedge Pab_2c_2d_2) \Rightarrow \exists d_3 Pab_1c_2d_3.$$

Оскільки задача перевірки імплікації для таких видів обмежень, як ВБЗЗ, є нерозв'язною [14; 34], то неможливо побудувати алгоритм синтезу реляційної схеми у DK/NF.

Логічні зв'язки, встановлені на основі аналізу першоджерел та власних результатів між означеннями класичних НФ та основних некласичних НФ, відображено на діаграмі (див. рис. 2). Окрім розглянутих у цій роботі, на діаграмі

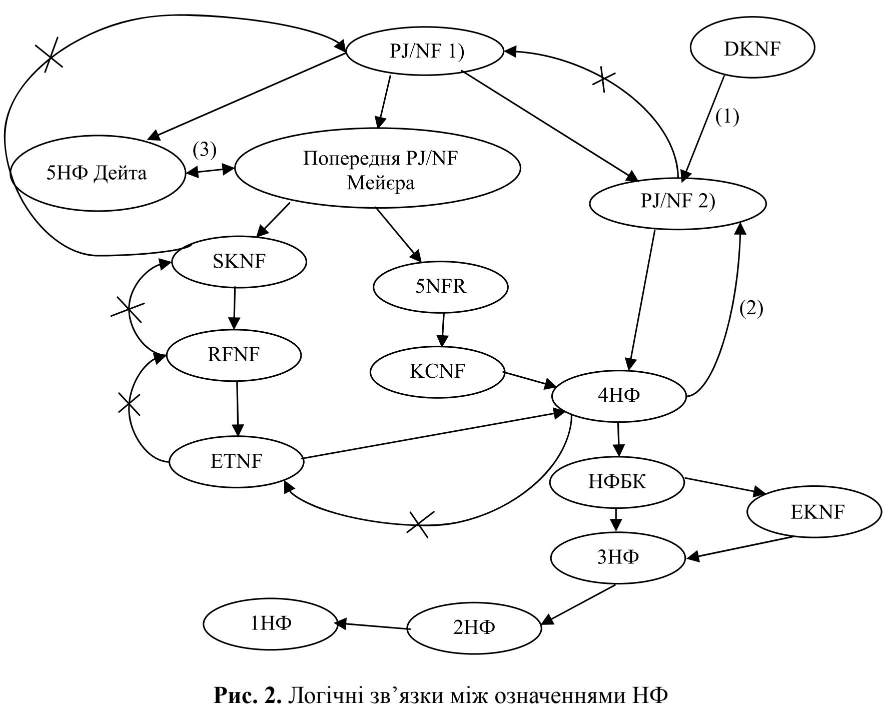

Рис. 2. Логічні зв'язки між означеннями НФ

Умовні позначення:
(1) – за умови нескінченності універсального домену [31];
(2) – за умови, що всі ЗЗ є БЗЗ;
(3) – якщо вважати, що ЗЗ в означенні Дейта синтаксично слідують із множини

вказано логічні зв'язки між класичними НФ, які відображаються імплікаціями: НФБК $\Rightarrow$ 3НФ $\Rightarrow$ 2НФ $\Rightarrow$ 1НФ [1]. Означення розглянутих у роботі НФ зібрано в таблиці.

Висновки

Незважаючи на вагомі результати, теорія нормалізації в реляційних БД має фрагментарний характер і далека ще до задовільного стану. Зауважимо, що нормалізація, основною метою якої є підтримка цілісності даних, вступає в суперечність з ефективністю опрацювання операцій у БД; саме тому нині вже можна говорити про початок створення теорії денормалізації та про природне (точніше кажучи, оптимальне за певним критерієм) поєднання нормалізації та денормалізації.

Таблиця. **Класичні та основні некласичні НФ**

№ з/п	Назва НФ	Автор(и), рік, посилання	Означення
1	1НФ	К. Дж. Дейт, 1975, [4]	Відношення перебуває в 1НФ, якщо кожне значення у відношенні – тобто кожне значення домену в кожному кортежі – є атомарним (таким, що не розкладається) елементом даних, наприклад, число або рядок символів
2	1НФ	Р. Елмасрі (R. Elmasri), С. Р. Наватхе (S. R. Navathe), 2000, [30]	Відношення перебуває в 1НФ, якщо домен кожного атрибута містить тільки атомарні (прості, неподільні) значення і значення кожного атрибута набувають тільки повних значень з цього домену
3	1НФ	К. Дж. Дейт, 2005, [5]	Змінна-відношення (relvar) перебуває у 1НФ тоді й лише тоді, коли в кожному її допустимому значенні кожний кортеж містить тільки одне значення для кожного атрибута
4	2НФ	Д. Мейєр, 1987, [8]	Схема відношення R перебуває у 2НФ відносно множини ФЗ F, якщо вона перебуває у 1НФ і кожен непервинний атрибут повністю залежить від кожного ключа для R
5	3НФ	Д. Мейєр, 1987, [8]	Схема відношення R перебуває у 3НФ відносно множини ФЗ F, якщо вона перебуває у 1НФ і жоден з непервинних атрибутів в R не є транзитивно залежним від ключа для R
6	3НФ	К. Дж. Дейт, 2005, [5]	Змінна-відношення перебуває у 3НФ тоді й лише тоді, коли вона перебуває у 2НФ і кожний неключовий атрибут нетранзитивно залежить від її первинного ключа (в означенні передбачається наявність тільки одного потенційного ключа, який і є первинним ключем відношення)
7	3НФ	В. П. Дрібас, 1982, [6]	3НФ відношення $r(R)$ – це або дане відношення, якщо воно перебуває у 2НФ і не містить транзитивних залежностей непервинних атрибутів від потенційних ключів, або набір яких-небудь його проекцій, кожна з яких задовольняє вказані умови, причому дане відношення повинно відновлюватись через природне з'єднання цих проекцій
8	Elementary Key Normal Form (EKNF)	К. Заніоло (C. Zaniolo), 1982, [50]	Відношення зі схемою R перебуває у EKNF за означенням, якщо для кожної нетривіальної ФЗ $X \to A$: 1) X є суперключем відношення, або 2) атрибут A належить деякому елементарному ключу (ключ X є елементарним, якщо для *деякого* атрибута $A \in R$ ФЗ $X \to A$ є повною)
9	НФБК	К. Заніоло (C. Zaniolo), 1982, [50]	Відношення зі схемою R перебуває у НФБК, якщо для кожної нетривіальної ФЗ $X \to A$ множина атрибутів X є суперключем
10	НФБК	Р. Фейгін (R. Fagin), 1979, [33]	Схема відношення перебуває у НФБК, якщо вона перебуває у 1НФ і кожна ФЗ $\sigma \in \Sigma$ є логічним наслідком множини ключів Δ ($\Delta \models \sigma$)
11	4НФ	Р. Фейгін (R. Fagin), 1977, [32]	Схема відношення перебуває у 4НФ, якщо вона перебуває у 1НФ і кожна БЗЗ $\sigma \in \Sigma$ є логічним наслідком множини ключів Δ
12	PJ/NF (класична 5НФ)	Р. Фейгін (R. Fagin), 1979, [33]	Два означення: 1) схема відношення перебуває у PJ/NF, якщо вона перебуває у 1НФ і для множини ключів Δ та кожної ЗЗ $\sigma \in \Sigma$ алгоритм приналежності видає результат «*True*»; 2) схема відношення перебуває у PJ/NF, якщо вона перебуває у 1НФ і кожна ЗЗ $\sigma \in \Sigma$ є логічним (семантичним) наслідком множини ключів Δ ($\Delta \models \sigma$)
13	5НФ	К. Дж. Дейт, 2005, [5]	Змінна-відношення перебуває в 5НФ (PJ/NF) за означенням, якщо кожен компонент кожної нетривіальної ЗЗ є суперключем

Продовження таблиці

№ з/п	Назва НФ	Автор(и), рік, посилання	Означення
14	Попередня 5НФ	Д. Мейєр, 1987, [8]	Схема відношення R перебуває в попередній PJ/NF за означенням, якщо кожна ЗЗ $*(R_1, R_2, \ldots, R_n)$ $(R_1 \cup R_2 \cup \ldots \cup R_n = R)$, яка виводиться (синтаксично слідує) з множини обмежень Σ, є або тривіальною, або її компоненти є суперключами
15	Reduced-5NF (5NFR)	М. Вінсент (M. Vincent), 1997, [47]	Нехай Σ – множина ФЗ і ЗЗ, причому кожна ЗЗ $*(R_1, R_2, \ldots, R_p)$ задовольняє умови: 1) $R_1 \cup R_2 \cup \ldots \cup R_p = R$, де R – схема відношення; 2) $\forall i,\ 1 \le i \le p,\ *(R_1, R_2, \ldots, R_{i-1}, R_{i+1}, \ldots, R_p)$ або не належить Σ^+, або $R_1 \cup R_2 \cup \ldots \cup R_{i-1} \cup R_{i+1} \cup \ldots \cup R_p \ne R$, де Σ^+ – множина всіх залежностей, які випливають (синтаксично) з множини . Пара $\langle R, \Sigma \rangle$ перебуває у 5NFR за означенням, якщо ліва частина кожної її нетривіальної ФЗ є суперключем і для кожної ЗЗ $\sigma \in \Sigma$ кожен компонент σ є суперключем
16	Superkey Normal Form (SKNF)	Р. Норманн (R. Normann), 1998, [45]	Реляційна схема $\langle R, \Sigma \rangle$ перебуває у НФ, яка визначається суперключем (SKNF), якщо кожен компонент кожної нескоротної ЗЗ є суперключем
17	Key-Complete Normal Form (KCNF)	М. Вінсент (M. Vincent), 1998, [48]	Пара $\langle R, \Sigma \rangle$ перебуває у KCNF, якщо ліва частина кожної нетривіальної ФЗ $\phi \in \Sigma$ є суперключем, а кожна ЗЗ $\sigma \in \Sigma$ повністю визначається ключами
18	Redandancy-free normal form (RFNF)	М. Вінсент (M. Vincent), 1998, [48]	Пара $\langle R, \Sigma \rangle$ перебуває в надлишково-вільній нормальній формі (RFNF), якщо не існує відношення $r(R) \in SAT(\Sigma)$, яке є надлишковим, де $SAT(\Sigma)$ – множина екземплярів (моделей) множини Σ, яка складається з ФЗ і ЗЗ
19	Essential Tuple Normal Form (ETNF)	Х. Дарвен (H. Darwen), К. Дейт (C. Date), Р. Фейгін (R. Fagin), 2012, [28]	Реляційна схема $\langle R, \Sigma \rangle$ перебуває в кортеже-необхідній НФ (ETNF), якщо кожний кортеж кожного екземпляра схеми $\langle R, \Sigma \rangle$ не є частково чи повністю надлишковим
20	Domain-Key Normal Form (DK/NF або ДКНФ)	Р. Фейгін (R. Fagin), 1981, [31]	Відношення перебуває у DK/NF, якщо кожне обмеження з його схеми є логічним наслідком з об'єднання множин залежності ключа та залежності домену

Список літератури

1. Буй Д. Б. Теорія нормалізації табличних баз даних: 2-3НФ, НФБК / Д. Б. Буй, А. В. Пузікова // Теоретичні та прикладні аспекти побудови програмних систем : матеріали міжнар. наук. конф., м. Ялта, 25 травня – 2 червня 2013 р. / редкол. Д. Б. Буй та ін. – Кіровоград : Центр оперативної поліграфії «Авангард», 2013. – С. 34–40.
2. Виноградова М. В. Конструктор баз данных на основе сущностей и их реквизитов с возможностью нормализации [Электронный ресурс] / М. В. Виноградова, Э. Г. Игушев // Электронное научно-техническое издание «Наука и образование». – 2011. – № 10. – Режим доступа: http://technomag.edu.ru/doc/242645.html. – Загл. с экрана.
3. Григорьев Ю. А. Алгоритм синтеза частично оптимальной схемы реляционной базы данных [Электронный ресурс] / Ю. А. Григорьев // Электронное научно-техническое издание «Наука и образование». – 2012. – № 1. – Режим доступа: http://technomag.edu.ru/doc/294486.html. – Загл. с экрана.
4. Дейт К. Дж. Введение в системы баз данных / К. Дж. Дейт. – М. : Наука, 1980. – 464 с.
5. Дейт К. Дж. Введение в системы баз данных : [пер. с англ.] / К. Дж. Дейт. – [8-е изд.]. – М. : Вильямс, 2005. – 1328 с.
6. Дрибас В. П. Реляционные модели баз данных / В. П. Дрибас. – Минск : Изд-во БГУ им. В. И. Ленина, 1982. – 192 с.
7. Зорин И. Теоретико-графовое приведение реляционной базы данных к третьей нормальной форме Э. Кодда [Электронный ресурс] / И. Зорин // Электронное научное издание «Устойчивое инновационное развитие: проектирование и управление». – 2009. – Т. 5. – С. 50–59. – Режим доступа: www.rypravlenie.ru. – Загл. с экрана.
8. Мейер Д. Теория реляционных баз данных : [пер. с англ.] / Д. Мейер. – М. : Мир, 1987. – 608 с.
9. Неклюдова Е. А. Синтез логической схемы реляционной базы данных / Е. А. Неклюдова, М. Ш. Цаленко // Программирование. – 1979. – № 6. – С. 58–68.
10. Петров С. В. Об аксиоматизации зависимостей по соединению / С. В. Петров // Применение методов математической логики : тезисы докл. IV Всес. конф. «Представление знаний и синтез программ». – Таллинн : АН ЭССР, 1986. – С. 151–152.
11. 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.
12. Ульман Дж. Основы систем баз данных / Дж. Ульман. – М. : Финансы и статистика, 1983. – 334 с.
13. Филиппович А. Взаимные функциональные зависимости / А. Филиппович // Системный администратор. – 2002. – № 1. – С. 84–89.
14. Цаленко М. Ш. Моделирование семантики в базах данных / М. Ш. Цаленко. – М. : Наука, 1989. – 287 с.
15. Aho A. V. The theory of joins in relational databases / A. V. Aho, C. Beeri, J. D. Ullman // Proc. 18th Symp. on Foundations of Computer Science, Providence, R. I. – 1977. – P. 107–113.
16. Armstrong W. W. Dependency structures of data base relationships / W. W. Armstrong // Proc. IFIP'74, North Holland Pub. Co., Amsterdam. – 1974. – P. 580–583.

17. Bahmani A. Automatic database normalization and primary key generation / A. Bahmani, M. Naghibzadeh, B. Bahmani // CCECE/CCGEI May 5–7 2008, Niagara Falls, Canada. – 2008. – P. 11–16.
18. Beeri C. A complete axiomatization for functional and multivalued dependencies / C. Beeri, R. Fagin, J. Howard // Proc. ACM-SIGMOD Conf. (Toronto, Canada, Aug. 3–5) ACM. – NY, 1977. – P. 47–61.
19. Beeri C. A sophisticate's introduction to database normalization theory / C. Beeri, P. Bernstein, N. Goodman // Proceedings of 4th International Conference on Very Large Data Bases, West Berlin. – 1978. – P. 113–124.
20. Beeri C. Computational problems related to the design of normal form relation schemas / C. Beeri, P. A. Bernstein // ACM Transactions on Database Systems. – 1979. – Vol. 4, № 1. – P. 30–59.
21. Bernstein P. A. Synthesizing Third Normal Form relations from functional dependencies / P. A. Bernstein // ACM Transactions on Database Systems. – 1976. – Vol. 1, № 4. – P. 277–298.
22. Biskup J. Inferences of multivalued dependencies in fixed and undetermined universes // J. Biskup // Theoretical Computer Science. – 1980. – Vol. 10, № 1. – P. 93–105.
23. Casanova M. A. Inclusion dependencies and their interaction with functional dependencies / M. A. Casanova, R. Fagin, C. H. Papadimitriou // Journal of Computer and System Sciences. – 1984. – № 28. – P. 29–59.
24. 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 (русский перевод: Кодд Е. Ф. Реляционная модель данных для больших совместно используемых банков данных / Е. Ф. Кодд // СУБД. – 1995. – № 1. – С. 145–160).
25. Codd E. F. Further Normalization of the Data Base Relational Model / E. F. Codd // (Presented at Courant Computer Science Symposia Series 6, "Data Base Systems", New York City, May 24–25, 1971.) IBM Research Report RJ909 (August 31, 1971). Republished in Randall J. Rustin (ed.), Data Base Systems: Courant Computer Science Symposia Series 6. Prentice-Hall, 1972.
26. Codd E. F. Recent Investigations into Relational Data Base Systems / E. F. Codd // Proceedings of IFIP Congress 74, Stockholm, Sweden, August 5–10, 1974. – North-Holland, 1974. – P. 1017–1021.
27. Dayal U. The fragmentation problem: lossless decomposition of relations into files / U. Dayal, P. A. Bernstein // Tech. Rep. CCA-78-13, Computer Corporation of America (Cambridge, Mass., November 15, 1978).
28. Darwen H. A Normal Form for Preventing Redundant Tuples in Relational Databases / H. Darwen, C. Date, R. Fagin // Proceedings of the 15th International Conference on Database Theory – ICDT'2012, March 26–30, 2012, Berlin, Germany. – 2012. – P. 114–126.
29. Delobel C. Decomposition of a database and the theory of Boolean switching function / C. Delobel, R. Gasey // IBM Journal of Research and Development. – 1973. – Vol. 17, № 5. – P. 374–386.
30. Elmasri R. Fundamental of Database Systems / R. Elmasri, S. Navathe. – [3rd Edition]. – Addison-Wesley, 2000. – 893 p.
31. Fagin R. A Normal Form for Relational Databases That Is Based on Domains and Keys / R. Fagin // Communications of the ACM. – 1981. – Vol. 6. – P. 387–415.
32. Fagin R. Multivalued Dependencies and a New Normal Form for Relational Databases / R. Fagin // ACM Transactions on Database Systems. – 1977. – Vol. 2, № 1. – P. 262–278.
33. Fagin R. Normal Forms and Relational Database Operators / R. Fagin // Proceedings of the ACM SIGMOD International Conference on Management of Data (Boston, Mass., May 30 – June 1), ACM. – NY, 1979. – P. 153–160.
34. Fagin R. The Theory of Data Dependencies – a Survey / R. Fagin, M. Y. Vardi // Mathematics of Information Processing, Proc. Symposia in Applied Mathematics – vol. 34, American Mathematical Society. – 1986. – P. 19–71.
35. Heath I. J. Unacceptable File Operations in Relational Database / I. J. Heath // ACM SIGFIDET Workshop on Data Description, Access, and Control. – San Diego, Calif, 1971. – P. 19–33.
36. Isloor S. S. An algorithm with logical simplicity for designing third normal form relations data base schema for functional dependencies / S. S. Isloor // Proceedings of International Conference on DBMS (ICMOD 78). – Fast Milano, Italy, 1978. – P. 31–50.
37. Jaeschke G. Remarks on the Algebra of Non First Normal Form Relations / G. Jaeschke, H. J. Schek // Proceedings of the 1st ACM SIGACT-SIGMOD symposium on principles of database systems. – Los Angeles, California, 1982. – P. 124–138.
38. Levene M. Justification for Inclusion Dependency Normal Form / M. Levene, M. W. Vincent // IEEE Transactions on Knowledge and Data Engineering. – 2000. – Vol. 12, № 2. – P. 281–291.
39. Lin W.-Y. Efficient algorithm for BCNF-decomposition / W.-Y. Lin // Information and Software Technology. – 1992. – Vol. 34, № 5. – P. 308–312.
40. Ling T. W. An Improved Third Normal Form for Relational Databases / T. W. Ling, F. W. Tompa, T. Kameda // ACM Transactions on Database Systems. – 1981. – Vol. 6, № 2. – P. 329–346.
41. Ling T. W. Logical Database Design with Inclusion Dependencies / T. W. Ling, C. H. Goh // Proceedings of the Eighth International Conference on Data Engineering, Tempe, Arizona. – 1992. – P. 642–649.
42. Makinouchi A. A Consideration of Normal Form on Not-necessarily Normalized Relations in the Relational Data Model / A. Makinouchi // Proceedings of the Third International Conference on Very Large Data Bases, October 6–8, 1977, Tokyo, Japan. IEEE Computer Society. – 1977. – P. 447–453.
43. Mendelzon A. O. On axiomatizing multivalued dependencies in relational databases / A. O. Mendelzon // ACM Transactions on Database Systems. – 1979. – Vol. 26, № 1. – P. 37–44.
44. Nicolas J. M. Mutual dependencies and same results on indecomposable relations / J. M. Nicolas // Proceedings of the fourth international conference on Very Large Data Bases. – 1978. – Vol. 4. – P. 360-367.
45. Normann R. Minimal lossless decompositions and some normal forms between 4NF and PJ/NF / R. Normann // Information Systems. – 1998. – Vol. 23, № 7. – P. 509–516.
46. Rissanen J. Independent components of relations / J. Rissanen // ACM Transactions on Database Systems. – 1977. – Vol. 2, № 4. – P. 317–325.
47. Vincent M. W. A corrected 5NF definition for relational database design / M. W. Vincent // Theoretical Computer Science (TCS). – 1997. – Vol. 185, № 2. – P. 379–391.
48. Vincent M. W. Redundancy Elimination and a New Normal Form for Relational Database Design / M. W. Vincent // Semantics in Databases / L. Libkin, B. Thalheim, eds. – Vol. 1358 of LNCS. – 1998. – P. 247–264.
49. Zaniolo C. Analysis and design of relational schemata for database systems : Ph. D. dissertation, Tech. Rep. UCLA-Eng-7769 / C. Zaniolo ; Dep. Computer Science, Univ. of California at Los Angeles. – July 1976.
50. Zaniolo C. A New Normal Form for the Design of Relational Database Schemata / C. Zaniolo // ACM Transactions on Database Systems. – 1982. – Vol. 7, № 3. – P. 489–499.

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].