УДК 004.658.2

Буй Д. Б., д.ф.-м.н., проф.,
Пузікова А. В., аспірантка.

Деякі некласичні нормальні форми в реляційних базах даних

Київський національний університет імені Тараса Шевченка, 83000, м. Київ, пр-т. Глушкова 4д,
e-mail: buy@unicyb.kiev.ua
e-mail: anna_inf@mail.ru

D. B. Buy, the doctor of physical and mathematical sciences, the professor,
A. V. Puzikova, postgraduate.

Some nonclassical normal forms in relation databases

Taras Shevchenko National University of Kyiv, 83000, Kyiv, Glushkova st., 4d,
e-mail: buy@unicyb.kiev.ua
e-mail: anna_inf@mail.ru

Обговорюються означення деяких класичних нормальних форм та основних некласичних нормальних форм. Показана нееквівалентність двох означень проективно-з'єднувальної нормальної форми, запропонованих Фейгіним. На основі аналізу першоджерел та власних результатів встановлені логічні зв'язки між означеннями класичних та основних некласичних нормальних форм.

Ключові слова: реляційні бази даних, некласичні нормальні форми.

We discuss the definitions of some classical normal forms, in particular, Fagin's Projection-Join Normal Form, and basic non-classical normal forms, such as: Zaniolo's Elementary Key Normal Form, Date's Fifth Normal Form, Maier's preliminary Projection-Join Normal Form, Vinsent's Reduced Fifth Normal Form, Normann's Superkey Normal Form, Vinsent's Key-Complete Normal Form, Vinsent's Redundancy Free Normal Form, Essential Tuple Normal Form (authors: Darwen, Date and Fagin), Fagin's Domain-Key Normal Form. It is shown that both definitions of Fagin's Projection-Join Normal Form (so-called classical fifth normal form) are not equivalent. In particular, we discuss the features of the relation scheme which is in one of the definitions of Fagin's Projection-Join Normal Form.

On the basis of analysis of the sources and own results are established logical connections between the definitions of classical and basic non-classical normal forms.

Key Words: relational databases, nonclassical normal forms.

Статтю представив д.ф.-м.н., проф. Анісімов А.В.

Означення класичних нормальних форм (НФ) в реляційних базах даних застосовуються до одного відношення [1-2], схеми одного відношення [3-6], таблиці [7]. Алгоритм зведення відношення до певної НФ повинен задовольняти вимогам (принципам) задачі проектування схеми бази даних (БД). Одним з важливих принципів розробки схеми БД є принцип представлення [8], 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 $i = \overline{1,n}$; іншими словами, вимагається виконання декомпозиції без втрат $r_0 = \otimes_{i=1}^{n} r_i$.

Передумовою виникнення пропозицій щодо покращення існуючих класичних НФ 2-3 порядків та НФ Бойса-Кодда (НФБК) є те, що ні декомпозиція (див., наприклад, [9]), ні деякі алгоритми синтезу (наприклад, алгоритм Берштейна (Bernstein) [10]) не забезпечували виконання принципу представлення для відношення з довільно заданими схемою і множиною функціональних залежностей (ФЗ).

Так, якщо на вхід алгоритма синтезу Берштейна подається реляційна схема $R_0\{A, B, C, D\}$ із множиною ФЗ $\{A \to B, C \to D\}$ (приклад з монографії В.П. Дрібаса [11]), результатом виконання алгоритму є схеми $R_1\{A, B\}$ і $R_2\{C, D\}$, кожна з яких знаходиться 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: «Третя нормальна форма (3НФ) відношення $r(R)$ [1] – це або дане відношення, якщо воно знаходиться у другій нормальній формі (2НФ) і не містить транзитивних залежностей непервинних атрибутів від потенційних ключів, або набір яких-небудь його проекцій, кожна з яких задовольняє вказаним умовам, причому дане відношення повинно відновлюватись через натуральне з'єднання (natural join) цих проекцій» [11]; зауважимо, що натуральне з'єднання за термінологією [7] називається з'єднанням та позначається як $\otimes$.

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 [12].

Відомо, що алгоритм синтезу Берштейна моделює схеми відношень, які знаходяться «майже» в НФБК. Маючи на меті охарактеризувати форму цих відношень, Заніоло (Zaniolo) запропонував НФ, яка визначається елементарним ключем (Elementary Key Normal Form – EKNF) [13]. Автор сформулював більш строге означення 3НФ: «Відношення знаходиться у 3НФ за означенням, якщо для кожної нетривіальної ФЗ $X \to A$:

a) множина атрибутів X є суперключем відношення із схемою R, або

b) атрибут A належить деякому потенційному ключу»,

та звернув увагу на різницю між означеннями 3НФ і НФБК – остання не містить умову (b).

В означенні EKNF використовуються додаткові поняття *елементарної ФЗ* та *елементарного ключа*:

– ФЗ $X \to A \in F$ (де F – множина ФЗ) називається *елементарною*, якщо $A \notin X$ і для кожної множини атрибутів $X' \subset X$ ФЗ $X' \to A$ не належить $[F]$ (де $[F]$ – замикання множини ФЗ); зауважимо, що за термінологією [11] така ФЗ $X \to A$ є *повною*;

– ключ X у відношенні схеми R є елементарним, якщо для *деякого* атрибута $A \in R$ ФЗ $X \to A$ є елементарною; наприклад, для схеми відношення $R = \{A, B, C\}$ із заданою множиною ФЗ $\{AB \to C, C \to A\}$ ключ $\{A, B\}$ є елементарним, а ключ $\{B, C\}$ – ні.

В означенні EKNF зберігається умова (a) та посилюється умова (b), вказані вище: «Відношення із схемою R знаходиться у EKNF за означенням, якщо для кожної нетривіальної ФЗ $X \to A$:

a) X є суперключем відношення, або
b) атрибут A належить деякому елементарному ключу».

На зв'язок EKNF з класичними НФ вказують імплікації НФБК $\Rightarrow$ EKNF $\Rightarrow$ 3НФ, які випливають безпосередньо з наведених вище означень.

Як вже згадувалось, схему відношення у EKNF синтезує алгоритм Берштейна; доведення цього факту спирається на теорему про те, що кожен ключ, який отримується в результаті виконання вказаного алгоритму, є елементарним [13].

Наступною НФ, яка викликала неабиякий інтерес, є проективно-з'єднувальна НФ (Projection-Join Normal Form – PJ/NF) або класична 5НФ; для неї Фейгін (Fagin) використовує наступні два означення [6].

В першому означенні PJ/NF використовується алгоритм приналежності (Membership algorirhm), який працює наступним чином: на вхід подаються множина потенційних ключів $\Delta = \{K_1, K_2, ..., K_m\}$ відношення із множиною атрибутів R реляційної схеми[2] та залежність з'єднання $*(R_1, R_2, ..., R_n)$. Ініціюється множина підсхем $S = \{R_1, R_2, ..., R_n\}$. Поки для деякого K_i, $1 \le i \le m$, виконується включення $K_i \subseteq R_j \cap R_k$ де $1 \le j, k \le n$, в множині S входження множини R_j замінюється на множину $R_j \cup R_k$, а входження R_k знищується. Якщо на виході множина S співпадає з схемою R, тобто $S = \{R\}$, то алгоритм видає значення «*True*». Приймається без доведення (на

[1] Запис $r(R)$ означає, що відношення r має схему (множину атрибутів) R.

[2] Зауважимо, що в роботі [6] під реляційною схемою розуміється пара $\langle R, \Sigma \rangle$, де R – множина атрибутів, Σ – множина обмежень, яка складається з ФЗ, багатозначних залежностей (БЗЗ) та залежностей з'єднання (ЗЗ) і є замкненою відносно операції логічного (семантичного) слідування.

прикладі), що алгоритм видає результат «*True*» тоді і тільки тоді, коли залежність з'єднання $*(R_1, R_2, ..., R_n)$ логічно (семантично) слідує з множини ключів $\Delta = \{K_1, K_2, ..., K_m\}$. Таким чином, згідно [6], наступні означення є еквівалентними:

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

2) «схема відношення знаходиться у PJ/NF, якщо вона знаходиться у 1НФ і кожна залежність з'єднання $\sigma \in \Sigma$ є логічним 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 дійсно ці означення є еквівалентними.

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

Доведення. Нехай таблиця t є моделлю реляційної схеми $\langle R, \Sigma \rangle$, яка знаходиться 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 $K_i \subseteq R_j \cap R_k$. Звідси за пунктом 2 твердження 2.10.4 монографії [7] випливає, що виконується рівність $\pi_{R_j \cup R_k}(t) = \pi_{R_j}(t) \otimes \pi_{R_k}(t)$.

Продовжуючи застосовувати вказаний пункт для кожного наступного включення $K_i \subseteq R_j \cap R_k$, $1 \le i \le m$, в результаті виконання алгоритму приналежності з урахуванням властивостей комутативності та асоціативності операції з'єднання [7] 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 $R = R_1 \cup ... \cup R_n$, то

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

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

Покажемо, що обернена імплікація не виконується; тобто, якщо таблиця t знаходиться у PJ/NF за означенням 2), то вона необов'язково знаходиться у PJ/NF за означенням 1). Для цього достатньо розглянути наступний 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. Але алгоритм приналежності видасть результат «*False*».

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

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

Доведення проводиться безпосередньо. □

Наслідок 1. Якщо схема відношення $\langle R, \Sigma \rangle$ знаходиться у PJ/NF за означенням 1), то кожен атрибут з множини R є суперключем[3].

Доведення. Нехай схема відношення $\langle R, \Sigma \rangle$ знаходиться 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 $\{\sigma \mid \Sigma \models \sigma\} \supseteq \{\varphi \mid \Sigma \mid\!- \varphi\}$, яке випливає з результату про відсутність повної скінченної множини правил виведення для ЗЗ [14]. Звідси випливає, що $\forall \varphi \ (\Sigma \mid\!- \varphi \Rightarrow \varphi \in \Sigma)$.

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

Із означеннями PJ/NF або 5НФ пов'язаний наступний блок некласичних НФ:

– НФ, яка визначається суперключем (Superkey Normal Form – SKNF) [15];

– спрощена 5НФ (*reduced*-5NF – 5NFR) [16];

– НФ, яка повністю визначається ключем (Key-Complete Normal Form) [17];

[3] Нагадаймо означення суперключа: «Суперключем є множина атрибутів $X \subseteq R$, яка містить ключ. Кожен ключ є також суперключем» [10].

– надлишково-вільна НФ (Redundancy Free Normal Form) [17].

Розглянемо детальніше кожну з них.

SKNF була запропонована Норманом (Normann) [15], який звернув увагу на «некоректне визначення 5НФ» сформульоване Дейтом [2] та попередню версію 5НФ у монографії Мейера [3]. Наведемо обидва визначення:

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

– схема відношення R знаходиться у PJ/NF за означенням, якщо кожна ЗЗ $*(R_1, R_2, ..., R_n)$ ($R_1 \bigcup R_2 \bigcup ... \bigcup R_n = R$), яка виводиться (синтаксично слідує) з множини обмежень Σ є або тривіальною або її компоненти є суперключами (Мейер).

Наведемо приклад з [3], на якому Мейер 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 знаходиться у підготовчій PJ/NF, але вона не знаходиться у класичній PJ/NF, оскільки $\{A, B\} \bigcap \{B, C\} = \{B\}$ не є ключем».

Насправді, множина $\{B\}$ в даному прикладі є ключем; це випливає з наступної леми, ідея якої неявно викладена в роботі Вінсента (Vincent) [16].

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

Доведення. Нехай на таблиці t виконується нетривіальна ЗЗ $*(R_1, R_2)$, де R_1, R_2 – суперключі, і нехай $R_1 \bigcap R_2 = X$. 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 [18]; ▫

– $X \neq \varnothing$; тоді за теоремою 1 Фейгіна [5] 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:

$$\frac{X \to\to Z,\ Y \to Z',\ Z' \subseteq Z,\ Y \bigcap Z = \varnothing}{X \to Z'}$$

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 $R_1 \to R$ випливає, що ФЗ $X \to R$ також виконується. ▫

Отже, $R_1 \bigcap R_2$ є суперключем. □

Лема 4. Якщо відношення знаходиться у PJ/NF за означенням 1) Фейгіна, то воно знаходиться у попередній PJ/NF за означенням Мейера та 5НФ за означенням Дейта. □

Доведення випливає безпосередньо з леми 2. □

Вінсент у роботі [16] доводить лему про те, що відношення знаходиться у 5НФ (за означенням Дейта) або попередній PJ/NF (за означенням Мейера) тоді і тільки тоді, коли кожен атрибут множини R є ключем.

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

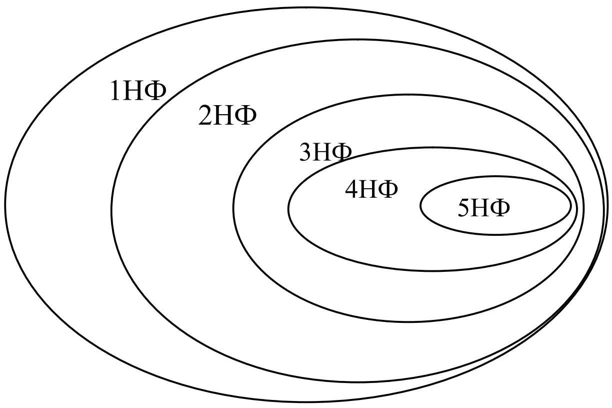

Рис. 1 Співвідношення між НФ

Нагадаймо, що в роботі [6] Фейгін сформулював означення класичних НФБК та 4НФ в семантичному вигляді та довів їх еквівалентність з відповідними класичними означеннями НФ, сформульованими в синтаксичній формі. Так, за означенням Фейгіна, «схема відношення знаходиться у НФБК (4НФ), якщо вона знаходиться у 1НФ і кожна ФЗ (БЗЗ) $\sigma \in \Sigma$ є логічним наслідком множини ключів Δ,

($\Delta \vDash \sigma$)». З того, що ФЗ є окремим випадком БЗЗ, яка в свою чергу 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 усі ЗЗ є БЗЗ.

Синтаксичні форми означень Дейта і Мейера як і означення 1) для PJ/NF не є узагальненням 4НФ, тобто, не еквівалентні означенню 4НФ, якщо кожна ЗЗ у множині залежностей є БЗЗ [16]. Скористаємось прикладом, наведеним у [16].

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

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

1. $R_1 \cup R_2 \cup ... \cup R_p = R$, де R – схема відношення;
2. $\forall R_i, 1 \le i \le p$, $*(R_1, R_2, ..., R_{i-1}, R_{i+1}, ..., R_p)$ 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 усіх залежностей, які слідують (синтаксично) з множини Σ.

Пара $\langle R, \Sigma \rangle$ знаходиться 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 Норман, який розглянув поняття нескоротної ЗЗ. За Норманом 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 $\langle R, \Sigma \rangle$ знаходиться у SKNF за означенням, якщо кожен компонент кожної нескоротної ЗЗ в R є суперключем.

Лема 5. Якщо відношення знаходиться у попередній PJ/NF за означенням Мейера, то воно знаходиться у 5NFR за означенням Вінсента та у SKNF за означенням Нормана. □

Доведення випливає безпосередньо з означень. □

У роботі [17] 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, співпадає з R».

Розглянемо приклад з [17]. Якщо $R = \{A, B, C, D\}$ і $\Sigma = \{A \to BCD,\ B \to ACD,\ *(AB, BC, CD),\ *(AB, BC, ACD)\}$, то множини $\{A\}$ і $\{B\}$ є потенційними ключами. ЗЗ $*(AB, BC, ACD)$ повністю визначається ключами, оскільки всі її компоненти є суперключами, а їх об'єднання $\{AB\} \cup \{BC\} \cup \{ACD\} = R$; тоді як ЗЗ $*(AB, BC, CD)$ не повністю визначається ключами, оскільки суперключами є компоненти $\{AB\}$ і $\{BC\}$, об'єднання яких $\{AB\} \cup \{BC\} \neq R$.

За означенням [17], пара $\langle R, \Sigma \rangle$ знаходиться у KCNF, якщо «ліва частина кожної нетривіальної ФЗ $\varphi \in \Sigma$ 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 [17].

В означенні RFNF використовується поняття *надлишковості* для входження значення змінної: «Входження значення змінної $t(A)$ кортежа 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 [17]: «Пара $\langle R, \Sigma \rangle$ знаходиться у 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 $\sigma \in \Sigma$ співпадає з R, доводиться теорема про еквівалентність RFNF і KCNF.

Описаний підхід до нормалізації відношення в термінах ненадлишковості кортежів привернув увагу Дарвіна (Darwen), Дейта і 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)) [19]. В означенны ETNF використовуються додаткові означення частково надлишкового кортежу, логічного слідування кортежу з множини кортежів та повністю надлишкового кортежу:

– кортеж t екземпляра r схеми R називається частково надлишковим, якщо існує кортеж $t' \in r$, $t' \neq t$, що для нетривіальної ФЗ $X \to A \in \Sigma$ виконується рівність $t[X] = t'[X]$ (тобто, в кортежах t та t' співпадають значення всіх атрибутів з множини X);

– кортеж t логічно слідує з множини кортежів S (відносно схеми $\langle R, \Sigma \rangle$), якщо кожний екземпляр r схеми $\langle R, \Sigma \rangle$, який містить усі кортежі з множини кортежів S, також містить кортеж t;

– кортеж t екземпляра r схеми $\langle R, \Sigma \rangle$ називається повністю надлишковим, якщо існує множина кортежів S в r причому $t \notin S$, така, що з S логічно слідує t (відносно $\langle R, \Sigma \rangle$).

В термінах часткової надлишковості означення НФБК формулюється так: «Схема відношення R знаходиться у НФБК за означенням, якщо не існує її екземпляра, який би містив частково надлишковий кортеж».

Поняття часткової та повної надлишковості кортежу є спеціальними випадками поняття надлишковості [17].

Згідно означення реляційна схема $\langle R, \Sigma \rangle$ знаходиться в кортеже-необхідній НФ (Essential Tuple Normal Form), якщо кожний кортеж кожного екземпляра схеми $\langle R, \Sigma \rangle$ не є частково чи повністю надлишковим.

На положення ETNF по відношенню до інших НФ вказують імплікації: PJ/NF (за означенням 1)) $\Rightarrow$ SKNF $\Rightarrow$ RFNF $\Rightarrow$ ETNF $\Rightarrow$ $\Rightarrow$ 4НФ, причому зворотні імплікації не виконуються.Обговорюються прості умови, за яких схема відношення $\langle R, \Sigma \rangle$ буде знаходитись у ETNF, а саме: 1) схема відношення $\langle R, \Sigma \rangle$ повинна бути зведена до НФБК; 2) деякий ключ повинен складатись з одного атрибута.

Окремим видом ЗЗ є вкладені залежності (ВЗ), для яких пропонувались інші НФ. Ці НФ в даній роботі не розглядаються. Зауважимо лише, що дослідженню вкладених залежностей, їх властивостей та взаємодії з ФЗ присвячено багато робіт. Зокрема, аксіоматика для ВЗ викладена в [20]; відомими є НФ Ніколаса (Nicolas) для вкладених ЗЗ [21], Inclusion Normal Form та алгоритми зведення до неї Лінга (Ling) і Гоха (Goh) [22], Inclusion Dependency Normal Form Левена (Levene) [23].

Наостанок розглянемо доменно-ключову НФ (Domain-Key Normal Form – DK/NF або ДКНФ), запропоновану Фейгіним у роботі [24]. Концепція DK/NF базується на поняттях залежності ключа та залежності домена: «Відношення знаходиться у DK/NF за означенням, якщо кожне обмеження з його схеми є логічним наслідком з об'єднання множин залежності ключа та залежності домена». Доводиться, що реляційна схема знаходиться у DK/NF тоді і тільки тоді, коли немає аномалій вставки та знищення (означення яких формулюються у роботі); проводиться аналогія між класичними та альтернативними означеннями НФБК, 4НФ та PJ/NF (за означенням 2)), розглянутими у роботі [6], пропонуються їх модифікації з урахуванням комбінаторного ефекту розмірів обмежуючого домену та показується, що DK/NF включає в себе ці модифікації; частковим є випадок, коли потужність усіх доменів є нескінченною, тоді DK/NF включає в себе НФБК, 4НФ та PJ/NF (за означенням 2)) у класичному виді. Але під обмеженнями в означенні ДКНФ розуміються не тільки ФЗ, БЗЗ та ЗЗ, означення яких використовуються у попередніх НФ, а й усі, які можна записати у вигляді висловлювання логіки предикатів 1-го порядку. У роботі [25] пропонується форма запису для усіх відомих на той час залежностей:

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

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

$$(\forall ab_1b_2c_1c_2d_1d_2)((Pab_1c_1d_1 \wedge Pab_2c_2d_2) \Rightarrow$$
$$\Rightarrow \exists d_3Pab_1c_2d_3 .$$

Оскільки для ВБЗЗ не існує скінченної дедуктивної системи [26, 27] і задача перевірки імплікації для таких видів обмежень, як ВБЗЗ, є нерозв'язною [25, 28], то неможливо побудувати алгоритм синтезу реляційної схеми у DK/NF.

Висновки. Проаналізовані означення деяких класичних НФ, зокрема PJ/NF, та основних некласичних НФ, а саме: попередньої PJ/NF Мейера, 5НФ Дейта, SKNF Нормана, 5NFR, RFNF та KCNF Вінсента, ETNF (автори: Дарвін, Дейт і Фейгін), EKNF Заніоло, DK/NF Фейгіна. Між означеннями цих НФ встановлені логічні зв'язки, які відображені на діаграмі (див. рис. 2). Окрім розглянутих у даній роботі, на діаграмі вказані логічні зв'язки між класичними НФ, які відображаються імплікаціями: НФБК$\Rightarrow$3НФ$\Rightarrow$ $\Rightarrow$2НФ$\Rightarrow$1НФ [18].

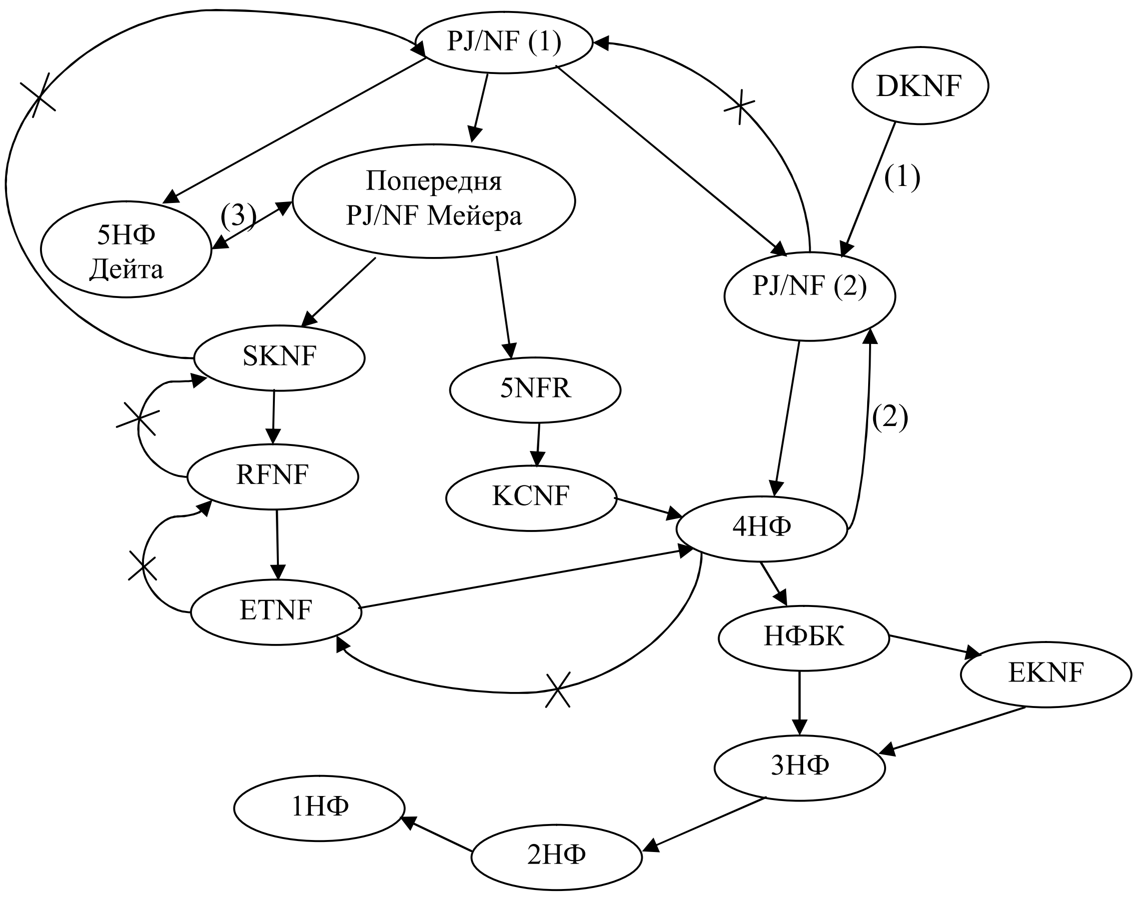

Рисунок 2. Логічні зв'язки між означеннями НФ
Умовні позначення:
(1) – за умови нескінченності універсального домена [24].
(2) – за умови, що усі ЗЗ є БЗЗ.
(3) – якщо вважати, що ЗЗ в означенні Дейта синтаксично слідують з множини Σ.

Список використаних джерел

1. *Codd E. F.* A Relational Model of Data for Large Shared Data Banks / E. F. Codd // Communications of the 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).
2. *Дейт К. Дж.* Введение в системы баз данных, 8-е издание: Пер. с англ. / К. Дж. Дейт. – М.: Вильямс, 2005. – 1328 с.
3. *Мейер Д.* Теория реляционных баз данных / Д. Мейер. – Москва: Мир, 1987. – 608 с.
4. *Ульман Дж.* Основы систем баз данных / Дж. Ульман. – М.: Финансы и статистика, 1983. – 334 с.
5. *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.
6. *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, New York, 1979. – P. 153-160.
7. *Редько В.Н.* 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.
8. *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.
9. *Филиппович, А.* Взаимные функциональные зависимости / А. Филиппович // Системный администратор. – 2002. – № 1. – С. 84-89.
10. *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.
11. *Дрибас В. П.* Реляционные модели баз данных / В. П. Дрибас. – Минск: Изд-во БГУ им. В. И. Ленина, 1982. – 192 с.

References

1. CODD, E. F. (1970) A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM.* 13 (6). p. 377-387.
2. DATE, C. J. (2005) *An Introduction to Database Systems*. 8th Ed. Addison-Wesley.
3. MAIER, D. (1983) *The theory of relational databases*. Computer Science Press.
4. ULLMAN, J. D. (1980) *Principles of Database Systems.* Rockville: Computer Science Press.
5. FAGIN, R. (1977) Multivalued Dependencies and a New Normal Form for Relational Databases. *ACM Transactions on Database Systems*. 2 (1). p. 262-278.
6. FAGIN, R. (1979) Normal Forms and Relational Database Operators. In *International Conference on Management of Data*, Wednesday 30th May to Thursday 1 June 1979. Boston: ACM, New York. – pp. 153-160.
7. REDKO, V. N. at al. (2001) *Reliatsiini bazy danykh: tablychni alhebry ta SQL-podibni movy.* Kyiv: Akademperiodyka.
8. BEERI, C., BERNSTEIN, P. and GOODMAN, N. (1978) A sophisticate’s introduction to database normalization theory. In *4th International Conference on Very Large Data Bases*, West Berlin Germany. ACM, IEEE. pp. 113-124.
9. FILIPPOVICH, A. (2002) Vzaimnye funktsionalnye zavisimisti. *Sistemnyi administrator*. (1). p. 84-89.
10. BERNSTEIN, P. A. (1976) Synthesizing Third Normal Form relations from functional dependencies. *ACM Transactions on Database Systems*. 1 (4). p. 277-298.
11. DRIBAS, V. P. (1982) *Relyacionnye modeli bas dannyh*. Minsk: Izdatelskiy centr Belorusskiy universytet.
12. LING, T. W., TOMPA, F. W. and KAMEDA, T. (1981) An Improved Third Normal Form for Relational Databases. *ACM Transactions on Database Systems*. 6 (2). p. 329-346.
13. ZANIOLO, C. (1982) A New Normal Form for the Design of Relational Database Schemata. *ACM Transactions on Database* Systems. 7 (3). p. 489-499.

12. *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.
13. *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.
14. *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». – Таллин: АН ЭССР, 1986. – С. 151-152.
15. *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.
16. *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.
17. *Vincent M.W.* Redundancy Elimination and a New Normal Form for Relational Database Design / M. W. Vincent // In Semantics in Databases (Libkin, L., Thalheim, B., eds.), vol. 1358 of LNCS. – 1998. – P. 247-264.
18. *Буй Д. Б.* Теорія нормалізації табличних баз даних: 2-3НФ, НФБК / Д. Б. Буй, А. В. Пузікова // Матеріали міжнародної конференції «Теоретичні та прикладні аспекти побудови програмних систем» – TAAPSD’2013 (Ялта, 25 травня – 2 червня 2013 р.). – С. 34-40.
19. *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. – P. 114-126.
20. *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.
21. *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.

14. PETROV, S. V. (1986) Ob aksiomatizacii zavisimostej po soedineniju. Primenenie metodov matematicheskoj logiki. *Tezisy dokladov IV Vsesojuznoj koferencii «Predstavlenie znanij i sintez programm»*. Tallin: AN JeSSR. – pp. 151-152.
15. NORMANN, R. (1998) Minimal lossless decompositions and some normal forms between 4NF and PJ/NF. *Information Systems*. 23 (7). p. 509-516.
16. VINCENT, M. W. (1997) A corrected 5NF definition for relational database design. *Theoretical Computer Science*. 185 (2). p. 379-391.
17. VINCENT, M. W. (1998) Redundancy Elimination and a New Normal Form for Relational Database Design. *In Semantics in Databases*. 1358 of LNCS. p. 247-264.
18. BUI, D. B. and PUZIKOVA, A. V. (2013) Normalization Theory in Table Databases: 2-3NF, BCNF. In *Theoretical and Applied Aspect of Program Systems Development*. Ialta, Saturday 25th May to Sunday 2nd June 2013. Kirovohrad: Avanhard. pp. 34-40.
19. DARWEN, H., DATE, C. and FAGIN, R. (2012) A Normal Form for Preventing Redundant Tuples in Relational Databases. In *Proceedings of the 15th International Conference on Database Theory*. Berlin, Monday 26th to Friday 30th March 2012. pp. 114-126.
20. CASANOVA, M. A., FAGIN, R. and PAPADIMITRIOU, C. H. (1984) Inclusion dependencies and their interaction with functional dependencies. *Journal of Computer and System Sciences*. (28). p. 29-59.
21. NICOLAS, J. M. (1978) Mutual dependencies and same results on indecomposable relations. In *Proceedings of the fourth international conference on Very Large Data Bases*. 4. p. 360-367.
22. LING, T. W. and GOH, C. H. (1992) Logical Database Design with Inclusion Dependencies. In *Proceedings of the Eighth International Conference on Data Engineering*. Tempe, Arizona. pp. 642-649.
23. LEVENE, M. and VINCENT, M. W. (2000) Justification for Inclusion Dependency Normal Form. *IEEE Transactions on Knowledge and Data Engineering*. 12 (2). p. 281-291.

22. *Ling T. W.* Logical Database Design with Inclusion Dependencies / T. W. Ling, C. H. Goh // In Proceedings of the Eighth International Conference on Data Engineering, Tempe, Arizona, 1992. – P. 642-649.
23. *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.
24. *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.
25. *Fagin R.* The theory of data dependencies – a survey / R. Fagin, M. Y. Vardi // In Mathematics of Information Processing, Proceedings of the. Symposia in Applied Mathematics – American Mathematical Society, 1986. – Vol. 34. – P. 19-71.
26. *Цаленко М. Ш.* Моделирование семантики в базах данных / М. Ш. Цаленко. – М.: Наука, 1989. – 287 с.
27. *Sagiv J.* Subset Dependencies and a Completeness Result for a Subset of Embedded Multivalued Dependencies / J. Sagiv, S. F. Walecka // Journal of the ACM. – 1982. – Vol. 29, № 1. – P. 103-117.
28. *Graham M. H.* Notions of dependency satisfaction / M. H. Graham, A. O. Mendelzon, M. Y. Vardi // Journal of the ACM. – 1986. – Vol. 33, № 1. – P. 105-129.

24. FAGIN, R. (1981) A Normal Form for Relational Databases That Is Based on Domains and Keys. *Communications of the ACM*. 6. p. 387-415.
25. FAGIN, R. and VARDI, M. Y. (1986) The theory of data dependencies – a survey. In *Mathematics of Information Processing, Proc. Symposia in Applied Mathematics.* 34. p. 19-71.
26. TSALENKO, M. Sh. (1989) *Modeling semantics in databases*. Moskva: Nauka.
27. SAGIV, J. and WALECKA, S. F. (1982) Subset Dependencies and a Completeness Result for a Subset of Embedded Multivalued Dependencies. *Journal of the ACM*. 29 (1). p. 103-117.
28. GRAHAM, M. H., MENDELZON, A. O. and VARDI, M. Y. (1986) Notions of dependency satisfaction. *Journal of the ACM*. 33 (1). p. 105-129.

Надійшла до редколегії 29.01.15