Міністерство освіти та науки України
Київський національний університет імені Тараса Шевченка

ПУЗІКОВА АННА ВАЛЕНТИНІВНА

УДК 004.655

ТЕОРІЯ НОРМАЛІЗАЦІЇ В ТАБЛИЧНИХ БАЗАХ ДАНИХ

01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем

Автореферат
дисертації на здобуття наукового ступеня
кандидата фізико-математичних наук

Київ – 2016

Дисертацією є рукопис.

Робота виконана на кафедрі теорії та технології програмування факультету кібернетики Київського національного університету імені Тараса Шевченка МОН України.

Науковий керівник – доктор фізико-математичних наук, професор
Буй Дмитро Борисович,
Київський національний університет імені Тараса Шевченка,
професор кафедри теорії та технології програмування

Офіційні опоненти: доктор фізико-математичних наук, професор,
Глибовець Микола Миколайович,
Національний університет "Києво-Могилянська академія",
м. Київ, декан факультету інформатики

кандидат фізико-математичних наук,
Глушко Ірина Миколаївна,
Ніжинський державний університет імені Миколи Гоголя,
м. Ніжин, доцент кафедри прикладної математики,
інформатики та освітніх вимірювань

Захист відбудеться "28" квітня 2016 р. о 14 годині на засіданні спеціалізованої вченої ради Д 26.001.09 Київського національного університету імені Тараса Шевченка за адресою: 03680, м. Київ, проспект академіка Глушкова, 4д, ауд. 40.

З дисертацією можна ознайомитись у науковій бібліотеці ім. М. Максимовича Київського національного університету імені Тараса Шевченка за адресою: 01601, м. Київ, вул. Володимирська, 58.

Автореферат розісланий "26" березня2016 р.

Вчений секретар

спеціалізованої вченої ради **В.П. Шевченко**

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми дослідження. В багатьох сучасних напрямках розвитку технологій баз даних (БД) в тій чи іншій мірі використовується реляційна модель даних, яка була запропонована Е. Коддом у 70-х роках XX ст. Семантичним структурам реляційного підходу в системах керування БД присвячені роботи вітчизняних вчених В.В. Пасічника, В.Н. Редька, А.О. Стогнія, Д.Б. Буя, Ю.Й. Брони та ін.

Однією з ключових вимог до розробників БД є забезпечення їх надійності, яка значною мірою залежить від правильності логічного проектування схеми БД. Проблема звільнення від відомих аномалій (оновлення, вставки, знищення) в реляційних БД, яка виникає на етапі логічного проектування внаслідок наявності різних видів обмежень (наприклад, функціональних і багатозначних залежностей), вимагає здійснення нормалізації – зведення до відповідних нормальних форм. Теорія нормалізації в реляційних БД отримала свій розвиток у роботах вчених Е. Кодда, В. Армстронга, Дж. Ріссанена (Rissanen), Дж. Ульмана, П. Бернштейна, Н. Делобеля (Delobel), Д. Мейера (Mayer), Р. Фейгіна (Fagin), а також Й.І. Брони, В.П. Дрібаса, М.Ш. Цаленко та ін.

Більшість поширених CASE-засобів (Computer-Aided Software Engineering) здійснюють нормалізацію до третьої нормальної форми. Автоматизація процесу нормалізації вимагає залучення формальних, насамперед, математичних методів. Оскільки теорія нормалізації спирається на теорію функціональних і багатозначних залежностей, виникає потреба у розгляді відповідних аксіоматик та тверджень про їх коректність і повноту.

Аналіз наукової і методичної літератури показав, що в ній відсутні доведення згадуваних результатів, які б задовольняли стандартним вимогам до строгості математичного доведення, а також відсутні критерії, за яких аксіоматики функціональних і багатозначних залежностей є повними.

Таким чином, алгоритмічний апарат нормалізації по суті носить евристичний характер, а, значить, вимагає математичного підґрунтя, що і визначає актуальність його розробки.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота є складовою частиною наукових робіт, які велись на кафедрі теорії та технології програмування факультету кібернетики Київського національного університету імені Тараса Шевченка при виконанні фундаментальної теми "Формальні специфікації та методи розробки надійних програмних систем" (№ 0111U007052, 2011-2015 рр.).

Мета і завдання дисертаційного дослідження. *Метою* дисертаційної роботи є математичні результати щодо коректності та повноти аксіоматик функціональних і багатозначних залежностей в табличних (реляційних) БД, а також цілісний несуперечливий фрагмент математичної теорії нормалізації стосовно другої-четвертої нормальних форм.

Із огляду на мету в роботі ставляться такі *задачі*:

– виконати математичне доведення відомого класичного результату в теорії табличних БД про повноту аксіоматики Армстронга для функціональних

залежностей (ФЗ) так, щоб воно задовольняло стандартним вимогам строгості та повноти математичного доведення;
– математично визначити критерій повноти аксіоматики ФЗ Армстронга;
– довести незалежність складових аксіоматики Армстронга;
– формально описати алгебру ФЗ, індуковану аксіоматикою Армстронга;
– виконати математичне доведення відомого класичного результату в теорії табличних БД про повноту аксіоматики багатозначних залежностей (БЗЗ) і ФЗ так, щоб воно задовольняло прийнятій практиці встановлення повноти в аксіоматичних системах;
– математично визначити критерії повноти аксіоматики БЗЗ та аксіоматики БЗЗ і ФЗ;
– математично описати фрагмент теорії нормалізації щодо нормальних форм 2-4 порядків.

Об'єктом дисертаційного дослідження є табличні БД. *Предметом* дослідження є аксіоматики ФЗ і БЗЗ та нормалізація в табличних БД.

У роботі використовуються теоретико-множинні та логіко-алгебраїчні методи.

Наукова новизна одержаних результатів. Доведення повноти аксіоматик ФЗ і БЗЗ доповнено строгим математичним доведенням їх коректності; для цього були введені відношення синтаксичного і семантичного слідування для кожної з цих аксіоматик.

Вперше математично визначені критерії повноти трьох аксіоматик: ФЗ, БЗЗ, а також ФЗ і БЗЗ в термінах потужностей множини атрибутів та універсального домену.

Деякі важливі властивості замикання множини атрибутів в зазначених аксіоматиках та властивості базису в аксіоматиці БЗЗ і ФЗ також розглянуті вперше.

Доведена незалежність складових аксіоматики Армстронга в тому розумінні, що без втрати повноти не можна видалити з аксіоматики аксіому рефлексивності чи жодне з правил виведення.

Побудовано алгебру ФЗ − алгебраїчний аналог аксіоматики Армстронга.

Теоретичне і практичне значення одержаних результатів. Дисертація має теоретико-прикладну спрямованість. Отримані результати можуть бути застосовані при розробці CASE-засобів, які підтримують нормалізацію, та були впроваджені у навчальний процес за спеціальністю "Інформатика" на факультеті кібернетики Київського національного університету імені Тараса Шевченка (нормативний курс "Композиційна семантика SQL-подібних мов").

У Кіровоградському державному педагогічному університеті імені Володимира Винниченка на кафедрі інформатики був прочитаний нормативний курс "Бази даних та СУБД", що включає результати, отримані в дисертації.

Особистий внесок здобувача. Основні результати роботи отримані здобувачем самостійно. Статті [1-11] написані у співавторстві з науковим керівником, якому належить постановка задачі дослідження, вибір методів дослідження та обговорення результатів. Із праць, виконаних зі співавторами, на захист виносяться лише результати, отримані особисто здобувачем.

Апробація результатів дослідження. Основні положення та висновки дисертаційного дослідження обговорювалися на наукових семінарах кафедри теорії та технології програмування Київського національного університету імені Тараса Шевченка.

Результати дисертаційного дослідження оприлюднені у доповідях і повідомленнях на Міжнародних та Всеукраїнських наукових конференціях, семінарах: VII Міжнародній науково-технічній конференції “Dependable Systems, Services and Technologies” – DESSERT’2014 (Київ, Україна, 16–18 травня 2014 р.), XVII Міжнародній конференції “Проблемы теоретической кибернетики” (Казань, РФ, 16–20 червня 2014 р.), Міжнародній науково-технічній конференції “Компьютерное моделирование в наукоемких технологиях” (Харків, 28–31 травня 2014 р.), XI Міжнародній науковій конференції “Теоретичні та прикладні аспекти побудови програмних систем” – TAAPSD’2014 (Київ, Україна, 15–17 грудня 2014 р.), XVII Міжнародній конференції “Дискретные модели в теории управляющих систем” (Москва, РФ, 20–22 травня 2015 р.), Tenth International Conference on Dependability and Complex Systems – DepCoS-RELCOMEX (Brunów, Poland, June 29 – July 3, 2015), Workshop on Foundations of Informatics – FOI-2015 (Chisinau, Republic of Moldova, August 24-29, 2015).

Публікації. Основні результати дисертаційного дослідження опубліковано у 13 працях. Серед них – 7 статей у наукових журналах і збірниках наукових праць [1-7], з них 6 статей опубліковано у фахових виданнях, затверджених ДАК України, 1 стаття опублікована у науковому фаховому іноземному виданні; 6 праць конференцій [8-13].

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел (81 найменування) і додатків. Загальний обсяг дисертації становить 131 сторінку, основний зміст викладено на 107 сторінках. Праця містить 9 таблиць та 5 рисунків.

ОСНОВНИЙ ЗМІСТ

У **вступі** обґрунтовується актуальність теми дисертації, визначається об’єкт, мета та завдання дисертаційної роботи, методи дослідження, наукова новизна одержаних результатів та їх теоретико-практичне значення, відображаються апробації та публікації результатів дисертаційного дослідження.

У **першому розділі "Теорія нормалізації в реляційних базах даних: сучасний стан"** подано огляд існуючої літератури з теорії нормалізації в реляційних (табличних) БД. Розглянуті означення деяких класичних нормальних форм (НФ) та основних некласичних НФ. Показана нееквівалентність двох означень проективно-з’єднувальної (PJ/NF) НФ, запропонованих Фейгіним. На основі аналізу першоджерел та власних результатів встановлені логічні зв’язки між означеннями класичних та основних некласичних НФ.

Аналіз наведеної літератури дозволяє зробити наступні висновки:

1) теорія нормалізації в реляційних БД, незважаючи на вагомі результати, носить фрагментарний характер;

2) накопичених теоретичних досліджень недостатньо для задоволення потреб розробників БД;
3) відсутність ефективного алгоритму зведення до PJ/NF (за другим означенням) пояснюється неможливістю побудови скінченної повної та коректної аксіоматики для залежностей з'єднання (ЗЗ);
4) потреби в автоматизації процесу нормалізації призводять до пошуку НФ, сильніших ніж 4НФ але слабкіших ніж PJ/NF (за другим означенням).

У **другому розділі "Аксіоматика функціональних залежностей Армстронга"** представлено чотири основних результати: перший стосується коректності і повноти аксіоматики ФЗ Армстронга, другий – критерію повноти цієї аксіоматики, третій – незалежності складових вказаної аксіоматики в тому розумінні, що без втрати повноти не можна видалити з аксіоматики аксіому рефлексивності чи жодне з правил виведення, а четвертий – побудови алгебри ФЗ, операції якої визначені відповідно до складових аксіоматики Армстронга.

У **підрозділі 2.1 "Повнота аксіоматики Армстронга"** представлено перший основний результат: побудовано математичне доведення коректності аксіоматики ФЗ табличних БД, а також її повноти через збіжність відношень синтаксичного та семантичного слідувань. Розглянуті властивості замикання множини атрибутів та замикання множини ФЗ. В якості математичного апарата використані властивості теоретико-множинної конструкції обмеження функції за множиною: $f \mid X = f \cap (X \times rang\, f)$, де f – функція, X – множина, $rang\, f$ – область значень функції f, $f \mid X$ – обмеження функції за множиною (функція ототожнюється з її графіком).

Розглядається дві множини: $\boldsymbol{A}$ – *множина атрибутів* і D – *універсальний домен*. Довільну (скінченну) множину атрибутів $R \subseteq \boldsymbol{A}$ назвемо *схемою*.

Рядком схеми R називається іменна множина на парі R, D, проекція якої за першою компонентою рівна R (тобто рядок функція вигляду $s: R \to D$).

Під *таблицею* t схеми R (позначатимемо $t(R)$) розуміємо скінченну множину рядків вказаної схеми R.

Далі X, Y – підмножини схеми R.

На таблиці t виконується ФЗ $X \to Y$, якщо для двох довільних рядків s_1, s_2 таблиці t, які збігаються на множині атрибутів X, має місце їх рівність і на множині атрибутів Y, тобто:

$$(X \to Y)(t) = true \overset{def}{\Leftrightarrow} \forall s_1, s_2 \in t \;\; s_1 \big| X = s_2 \big| X \Rightarrow s_1 \big| Y = s_2 \big| Y \;\; .$$

Отже, з семантичної точки зору ФЗ – це предикат, заданий двома (скінченними) множинами атрибутів.

Скажемо, що таблиця t схеми R є *моделлю* множини ФЗ *F*, якщо кожна ФЗ $X \to Y \in F$ виконується на таблиці t:

$$t \;\; \textit{модель} \;\; F \overset{def}{\Leftrightarrow} \forall (X \to Y)(X \to Y \in F \Rightarrow (X \to Y)(t) = true) .$$

Отже, довільна таблиця є моделлю порожньої множини ФЗ.

Виконуються такі аксіома і правила виведення з аксіоматики Армстронга.

Аксіома рефлексивності Армстронга (точніше кажучи, схема аксіоми рефлексивності): $\forall t (X \to Y)(t) = true$, де $Y \subseteq X$.

Правило поповнення Армстронга (точніше кажучи, схема правила поповнення): $(X \to Y)(t) = true \Rightarrow (X \cup Z \to Y \cup Z)(t) = true$ для $Z \subseteq R$.

Правило транзитивності Армстронга:

$(X \to Y)(t) = true \wedge (Y \to Z)(t) = true \Rightarrow (X \to Z)(t) = true$.

ФЗ вигляду $X \to Y$, де $Y \subseteq X$, називається *тривіальною*.

ФЗ $X \to Y$ *семантично слідує* з множини ФЗ F, якщо на кожній таблиці $t(R)$, яка є моделлю множини ФЗ F, виконується також ФЗ $X \to Y$:

$$F \models X \to Y \overset{def}{\Leftrightarrow} \forall t(R)(t \text{ модель } F \Rightarrow (X \to Y)(t) = true).$$

Таким чином, тривіальні ФЗ виконуються на довільній таблиці; іншими словами, тривіальні ФЗ семантично слідують з порожньої множини ФЗ.

З наведених аксіоми і правил виведення випливають наступні наслідки (властивості відношення семантичного слідування).

1. $\varnothing \models X \to Y$ для $Y \subseteq X$.
2. $F \models X \to Y \Rightarrow F \models X \cup Z \to Y \cup Z$ для $Z \subseteq R$.
3. $F \models X \to Y \wedge F \models Y \to Z \Rightarrow F \models X \to Z$.

Правила поповнення і транзитивності назвемо *правилами виведення*; вони мають синтаксичну природу.

Скажемо, що ФЗ $X \to Y$ *синтаксично слідує* з множини ФЗ F ($F \vdash X \to Y$), якщо існує скінченна послідовність ФЗ $\phi_1, \phi_2, ..., \phi_{m-1}, \phi_m$, така, що $\phi_m = X \to Y$ і $\forall i = \overline{1, m-1}$ кожна ϕ_i є або аксіома рефлексивності, або належить F, або отримана за яким-небудь правилом виведення (поповнення, транзитивності) з попередніх у цій послідовності ФЗ ϕ_j, ϕ_k, $j, k < i$. Послідовність $\phi_1, \phi_2, ..., \phi_{m-1}, \phi_m$ назвемо *доведенням*, наслідуючи традиції математичної логіки[1].

Нехай задана деяка множина ФЗ F. *Замикання* $[F]_{\vdash}$ – це множина усіх ФЗ, які синтаксично слідують з F: $[F]_{\vdash} \overset{def}{=} \{X \to Y \mid F \vdash X \to Y\}$.

Для спрощення позначень далі будемо записувати просто $[F]$ замість $[F]_{\vdash}$.

В наступній лемі викладено природні властивості замикання.

Лема 2.1. Виконуються властивості:

1) $F \subseteq [F]$ (зростання);

2) $[[F]] = [F]$ (ідемпотентність);

3) $F \subseteq G \Rightarrow [F] \subseteq [G]$ (монотонність).□

Таким чином, оператор $F \mapsto [F]$ є оператором замикання в стандартному розумінні.

---

[1] *Линдон Р*. Заметки по логике. – Москва: Мир, 1968. – 128 с.

Нехай задана деяка множина ФЗ F, і нехай $\Im_F$ – сімейство усіх множин ФЗ G, які містять F, таких, що при застосуванні до них аксіом та правил виведення Армстронга не можна отримати жодної ФЗ, яка б не належала G:

$$\Im_F \overset{def}{=} \{G \mid F \subseteq G \wedge [G] \subseteq G\}.$$

Очевидно, що сім'я $\Im_F$ непорожня, вона містить, наприклад, множину всіх ФЗ $\{X \to Y \mid X \subseteq R \wedge Y \subseteq R\}$.

Лема 2.2. Сім'я $\Im_F$ замкнена відносно довільних перетинів. □

Нехай $F^* \overset{def}{=} \operatorname*{I}_{G \in \Im_F} G$ перетин всіх множин ФЗ з сім'ї $\Im_F$. З леми 2.2 випливає, що $F^* \in \Im_F$. Зауважимо, що сім'я ФЗ $\Im_F$ по суті є муровською сім'єю в розумінні абстрактної теорії решіток[2].

Лема 2.3. Виконується рівність: $[F] = F^*$.□

Отже, замикання $[F]$ – це найменша (в розумінні теоретико-множинного включення $\subseteq$) множина ФЗ, що містить F, така, що при застосуванні до неї аксіом та правил виведення Армстронга не можна отримати жодної ФЗ, яка б не належала $[F]$.

Замиканням $[X]_F$ множини атрибутів X (відносно множини ФЗ F) називається об'єднання правих частин всіх ФЗ вигляду $X \to Y$, які слідують (синтаксично) з множини F: $[X]_F \overset{def}{=} \bigcup_{X \to Y \in [F]} Y$.

Для спрощення позначень далі параметр F явно вказувати не будемо.

В наступній лемі викладено властивості замикання множини атрибутів X.

Лема 2.4. Виконуються властивості:

1) $X \subseteq [X]$;

2) $F \text{ ђ } X \to [X]$;

3) $X \to Z \notin [F] \Rightarrow Z \text{ Ъ } [X] \wedge [X] \subset R$. □

Твердження 2.1 (коректність аксіоматики Армстронга)**.** Якщо ФЗ $X \to Y$ синтаксично виводиться з множини ФЗ F, то ФЗ $X \to Y$ виводиться з F семантично: $F \text{ ђ } X \to Y \Rightarrow F \text{ ‘ } X \to Y$. □

Твердження 2.2 (повнота аксіоматики Армстронга). Якщо ФЗ $X \to Y$ семантично виводиться з множини ФЗ F, то ФЗ $X \to Y$ синтаксично виводиться з F: $F \text{ ‘ } X \to Y \Rightarrow F \text{ ђ } X \to Y$. □

Теорема 2.1. Відношення семантичного та синтаксичного слідування збігаються: $F \text{ ‘ } X \to Y \Leftrightarrow F \text{ ђ } X \to Y$.□

Виявляється, що оператор $X \text{ а } [X]$ є теж оператором замикання.

Твердження 2.3. Виконуються властивості:

1) $X \subseteq Y \Rightarrow [X] \subseteq [Y]$ (монотонність);

---

[2] Биркгоф Г. Теория решеток. – Москва: Наука, 1984. – 568 с.

2) $[[X]] = [X]$ (ідемпотентність).□

Окрім того згідно з п.1 леми 2.4 цей оператор є зростаючим оператором.

Детальний аналіз наведеного доведення основного результату про збіжність відношень синтаксичного (ђ) та семантичного (‘) слідування показує, що воно проведено в припущенні: $|D| \geq 2$ та $|R| \geq 2$, тобто універсальний домен містить не менше 2 елементів та схема R має щонайменше 2 атрибути.

У **підрозділі 2.2 "Критерій повноти аксіоматики Армстронга в термінах потужностей множини атрибутів та універсального домену"** досліджено збіжність відношень семантичного і синтаксичного слідувань у випадках, коли $|D| \leq 1$ або $|R| \leq 1$.

Теорема 2.2 (критерій повноти аксіоматики Армстронга). Відношення синтаксичного та семантичного слідувань ђ і ‘ співпадають тоді і тільки тоді, коли в інтерпретації $|D| \geq 2$ або $|R| = 0$.□

У **підрозділі 2.3 "Незалежність аксіоматики Армстронга"** показано, що аксіоматика Армстронга є незалежною в тому розумінні, що без втрати повноти не можна опустити ні єдину аксіому (мова йде про схему аксіоми рефлексивності), ні будь-яке з правил виведення (мова йде про схему правила поповнення та про правило транзитивності).

Лема 2.5. Аксіоми рефлексивності є незалежними від правил виведення аксіоматики Армстронга. □

Лема 2.6. Правила транзитивності є незалежними від аксіом рефлексивності і правил поповнення. □

Лема 2.7. Правила поповнення є незалежними від аксіом рефлексивності і правил транзитивності. □

Враховуючи леми 2.5, 2.6 та 2.7 встановлено третій основний результат.

Теорема 2.3. Аксіоматика Армстронга є незалежною. □

У **підрозділі 2.4 "Алгебра функціональних залежностей в табличних базах даних"** побудована алгебра ФЗ, операції якої визначені відповідно до складових аксіоматики Армстронга. Розглянута підалгебра тривіальних ФЗ та множина, яка її породжує. Побудова цієї алгебри дозволяє формулювати результати щодо властивостей аксіоматики Армстронга на алгебраїчній мові.

Задамо алгебру $A = \langle \Psi, \Omega \rangle$, носієм якої є множина ФЗ $\Psi = \ X \to Y \mid X, Y \in 2^R \ $,

де 2^R – булеан множини R; а сигнатура Ω складається з операцій, які задаються наступним чином:

– нуль-арні операції $X \to Y^{(0)}$, де $X, Y \subseteq R$, причому $Y \subseteq X$, які відповідають аксіомам рефлексивності в аксіоматиці Армстронга;

– унарні параметричні операції $R_Z'^{(1)}$, які відповідають правилам поповнення в аксіоматиці Армстронга:

$$R_Z'^{(1)} : \Psi \to \Psi,\ Z \in 2^R,$$

$$R_Z'^{(1)}(X \to Y) = X \cup Z \to Y \cup Z\,;$$

– бінарна часткова операція $R''^{(2)}$, яка відповідає правилу транзитивності в аксіоматиці Армстронга:

$$R''^{(2)} : \Psi \times \Psi \nrightarrow \Psi,\ dom\ R''^{(2)} = \ \langle X \to Y', Y'' \to Z \rangle \mid Y' = Y''\ ,$$

$$R''^{(2)}(X \to Y, Y \to Z) = X \to Z\,.$$

Таким чином, алгебра ФЗ є частковою і задається як:

$$A = \left\langle\ X \to Y\ \ _{X,Y \in 2^R}, \{R_Z'^{(1)}, R''^{(2)}, X \to Y^{(0)}\}_{Z,X,Y \in 2^R, Y \subseteq X} \right\rangle.$$

Позначимо множину аксіом рефлексивності, які задаються нуль-арними операціями $X \to Y^{(0)}$, через $\Psi_r = \{X \to Y \mid X, Y \in 2^R \wedge Y \subseteq X\}$. Очевидно, що $A' = \left\langle \Psi_r, \{R_Z'^{(1)}, R''^{(2)}\}_{Z \in 2^R} \right\rangle$ є підалгеброю алгебри A, оскільки множина тривіальних ФЗ Ψ_r є замкненою відносно вказаних операцій.

Лема 2.8. Множина $\Psi_{r_0} = \ \ A \to \varnothing \mid A \in R$ породжує множину тривіальних залежностей Ψ_r за допомогою операцій $R_Z'^{(1)}$ і $R''^{(2)}$. □

Лема 2.9. Множина Ψ_{r_0} – найменша множина, яка породжує множину Ψ_r за допомогою основних операцій $R_{\{A\}}'^{(1)}$, $A \in R$ і $R''^{(2)}$. □

Твердження 2.4. Алгебра ФЗ задається як своє збіднення наступного вигляду:

$$A = \left\langle\ X \to Y\ \ _{X,Y \in 2^R}, \{R_{\{A\}}'^{(1)}, R''^{(2)}, \{A\} \to \varnothing^{(0)}\}_{A \in R} \right\rangle. \ \square$$

Зауважимо, що замикання $[F]$ множини ФЗ F, де $F \subseteq\ \ X \to Y\ \ _{X,Y \in 2^R}$ можна розглядати як замикання множини $F \cup \Psi_{r_0}$ операціями $R_{\{A\}}'^{(1)}$, $A \in R$ і $R''^{(2)}$:

$$[F] = [F \cup \Psi_{r_0}]_{\ R_{\{A\}}'^{(1)},\ R''^{(2)}}\ .$$

Третій розділ "Аксіоматика багатозначних і функціональних залежностей" присвячений питанням коректності і повноти аксіоматики БЗЗ і аксіоматики ФЗ та БЗЗ в табличних БД. Встановлено критерії повноти вказаних аксіоматик в термінах потужностей множини атрибутів і універсального домену.

У **підрозділі 3.1 "Аксіоматика багатозначних залежностей"** введені відношення семантичного і синтаксичного слідувань для БЗЗ, розглянуті властивості відношення семантичного слідування, а також властивості замикання множини атрибутів відносно множини БЗЗ та замикання множини БЗЗ.

Як і раніше, зафіксуємо множини R і D.

На таблиці t схеми R виконується БЗЗ $X \to\to Y$, якщо для двох довільних рядків s_1, s_2 таблиці t, які збігаються на множині атрибутів X, існує рядок $s_3 \in t$, який дорівнює об'єднанню обмежень рядків s_1, s_2 на множини атрибутів $X \cup Y$ і $R \setminus (X \cup Y)$ відповідно:

$$(X \to\to Y)(t) = true \overset{def}{\Leftrightarrow} \forall s_1, s_2 \in t(s_1 \mid X = s_2 \mid X \Rightarrow$$
$$\Rightarrow \exists s_3 \in t(s_3 = s_1 \mid (X \cup Y) \cup s_2 \mid R \setminus (X \cup Y)))\,.$$

Отже, з семантичної точки зору БЗЗ – це параметричний предикат на таблицях схеми R, заданий двома (скінченними) множинами-параметрами атрибутів X, Y.

Структура таблиці t, на якій виконується БЗЗ $X \to\to Y$, може бути представлена за допомогою такого відношення. Скажемо, що рядки s_1, s_2 таблиці t знаходяться у (параметричному) відношенні $=_X$, якщо вони збігаються на множині атрибутів X:

$$s_1 =_X s_2 \overset{def}{\Leftrightarrow} s_1 \mid X = s_2 \mid X .$$

Зрозуміло, що відношення $=_X$ є відношенням еквівалентності і тому розбиває множину рядків таблиці t на класи еквівалентності, які мають таке зображення:

$$[s]_{=_X} = \{s \mid X\} \otimes \pi_Y([s]_{=_X}) \otimes \pi_{R \setminus (X \cup Y)}([s]_{=_X}),$$

де s – довільний представник класу, а π_Y, $\otimes$ – операції проекції та з'єднання таблиць відповідно[3].

Скажемо, що таблиця $t(R)$ є *моделлю* множини БЗЗ G, якщо кожна БЗЗ $X \to\to Y \in G$ виконується на таблиці $t(R)$:

$$t(R) \text{ } модель \text{ } G \overset{def}{\Leftrightarrow} \forall (X \to\to Y)(X \to\to Y \in G \Rightarrow (X \to\to Y)(t) = true).$$

Виконуються такі аксіоми і правила виведення:

Аксіома рефлексивності: $\forall t (X \to\to Y)(t) = true$, де $Y \subseteq X$.

Аксіома: $\forall t (X \to\to Y)(t) = true$, де $X \cup Y = R$.

Правило повноти: $(X \to\to Y)(t) = true \Rightarrow (X \to\to R \setminus (X \cup Y))(t) = true$.

Правило поповнення: $(X \to\to Y)(t) = true \wedge Z \subseteq W \Rightarrow (X \cup W \to\to Y \cup Z)(t) = true$.

Правило транзитивності:

$(X \to\to Y)(t) = true \wedge (Y \to\to Z)(t) = true \Rightarrow (X \to\to Z \setminus Y)(t) = true$.

БЗЗ вигляду $X \to\to Y$, де $Y \subseteq X$ або $X \cup Y = R$, називається *тривіальною*.

За умови $Z \cap Y = \varnothing$ із виконання БЗЗ $X \to\to Y$ і $Y \to\to Z$ випливає виконання БЗЗ $X \to\to Z$.

Зауважимо, що за умови $Y \cap X \neq \varnothing$ зображення класів еквівалентності, на які відношення $=_X$ розбиває множину рядків таблиці t, можна уточнити:

$$[s]_{=_X} = \{s \mid X\} \otimes \pi_{Y \setminus X}([s]_{=_X}) \otimes \pi_{R \setminus (X \cup Y)}([s]_{=_X}).$$

Скажемо, що БЗЗ $X \to\to Y$ *семантично слідує* з множини БЗЗ G, якщо на кожній таблиці $t(R)$, яка є моделлю множини БЗЗ G, виконується також БЗЗ $X \to\to Y$:

$$G \models X \to\to Y \overset{def}{\Leftrightarrow} \forall t(R)(t \text{ } модель \text{ } G \Rightarrow (X \to\to Y)(t) = true).$$

Таким чином, тривіальні БЗЗ виконуються на довільній таблиці; іншими словами, тривіальні БЗЗ семантично слідують з порожньої множини БЗЗ.

---

[3] 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.

З наведених аксіом і правил виведення випливають наслідки (властивості відношення семантичного слідування).

1. $\varnothing \models X \to\to Y$ для $Y \subseteq X$.
2. $\varnothing \models X \to\to Y$ для $X \cup Y = R$.
3. $G \models X \to\to Y \Rightarrow G \models X \to\to R \setminus (X \cup Y)$.
4. $G \models X \to\to Y \wedge Z \subseteq W \Rightarrow G \models X \cup W \to\to Y \cup Z$.
5. $G \models X \to\to Y \wedge G \models Y \to\to Z \Rightarrow G \models X \to\to Z \setminus Y$;
6. $G \models X \to\to Y \wedge G \models Y \to\to Z \wedge Z \cap Y = \varnothing \Rightarrow G \models X \to\to Z$. □

Скажемо, що БЗЗ $X \to\to Y$ *синтаксично слідує* з множини БЗЗ G відносно схеми R ($G \vdash_R X \to\to Y$), якщо існує скінченна послідовність БЗЗ $\phi_1, \phi_2, ..., \phi_{m-1}, \phi_m$, така, що $\phi_m = X \to\to Y$ і для $\forall i = \overline{1, m-1}$ кожна ϕ_i є або аксіома рефлексивності, або належить G, або отримана за яким-небудь правилом виведення (повноти, поповнення, транзитивності) для БЗЗ з попередніх у цій послідовності БЗЗ ϕ_j, ϕ_k, $j, k < i$. Послідовність $\phi_1, \phi_2, ..., \phi_{m-1}, \phi_m$ назвемо *доведенням*, наслідуючи традиції математичної логіки.

Нехай задана деяка множина БЗЗ G. *Замиканням* $[G]_R$ називається множина БЗЗ, які синтаксично слідують з G (відносно схеми R):

$$[G]_R \overset{def}{=} \{X \to\to Y \mid G \vdash_R X \to\to Y\}.$$

Для спрощення позначень далі параметр R явно вказувати не будемо.

В наведеній нижче лемі викладено природні властивості замикання.

Лема 3.1. Виконуються властивості:

1) $G \subseteq [G]$ (*зростання*);

2) $[[G]] = [G]$ (*ідемпотентність*);

3) $G \subseteq H \Rightarrow [G] \subseteq [H]$ (*монотонність*).□

Таким чином, оператор $G \mapsto [G]$ є оператором замикання.

З описаних вище аксіоми рефлексивності і правил виведення будуються доведення похідних правил виведення для БЗЗ.

У **підрозділі 3.2 "Аксіоматика багатозначних і функціональних залежностей"** введені відношення семантичного і синтаксичного слідувань для аксіоматики БЗЗ і ФЗ. Розглянуті властивості відношення семантичного слідування для БЗЗ і ФЗ, властивості замикання множини атрибутів відносно об'єднання множин ФЗ і БЗЗ, а також властивості замикання об'єднання множин ФЗ і БЗЗ та властивості базису для об'єднання цих множин.

Нехай задані множини ФЗ F і БЗЗ G. Скажемо, що таблиця $t(R)$ є *моделлю* множини $F \cup G$, якщо кожна залежність $\phi \in F \cup G$ виконується на таблиці t:

$$t(R) \text{ } модель \text{ } F \cup G \overset{def}{\Leftrightarrow} \forall \phi \, (\phi \in F \cup G \Rightarrow \phi(t) = true).$$

Для ФЗ та БЗЗ виконуються спільні правила виведення.

1. Правило *розширення* ФЗ до БЗЗ:

$(X \to Y)(t) = true \Rightarrow (X \to\to Y)(t) = true$.

2. $(X \to\to Z)(t) = true \wedge (Y \to Z')(t) = true \wedge Z' \subseteq Z \wedge Y \cap Z = \varnothing \Rightarrow$
$\Rightarrow (X \to Z')(t) = true$.

Скажемо, що ФЗ або БЗЗ ϕ *семантично слідує* з множини залежностей $F \cup G$ (відносно схеми R), якщо на кожній таблиці $t(R)$, яка є моделлю множини залежностей $F \cup G$, виконується також залежність ϕ:

$$F \cup G \models \phi \overset{def}{\Leftrightarrow} \forall t(R)(t(R) - \textit{модель } F \cup G \Rightarrow \phi(t) = true).$$

З спільних правил виведення для ФЗ і БЗЗ випливають такі властивості відношення семантичного слідування:

1. $F \models X \to Y \Rightarrow F \models X \to\to Y$;
2. $G \models X \to\to Z \wedge F \models Y \to Z' \wedge Z' \subseteq Z \wedge Y \cap Z = \varnothing \Rightarrow F \cup G \models X \to Z'$.

Скажемо, що ФЗ або БЗЗ ϕ *синтаксично слідує* з об'єднання множин ФЗ і БЗЗ $F \cup G$ відносно схеми R ($F \cup G \vdash_R \phi$), якщо існує скінченна послідовність ФЗ або БЗЗ $\phi_1, \phi_2, ..., \phi_{m-1}, \phi_m$, така, що $\phi_m = \phi$ і для $\forall i = \overline{1, m-1}$ кожна ϕ_i є або аксіома рефлексивності (для ФЗ або БЗЗ), або належить $F \cup G$, або отримана за яким-небудь правилом виведення (повноти (для БЗЗ), поповнення (для ФЗ або БЗЗ), транзитивності (для ФЗ або БЗЗ), спільних правил для ФЗ і БЗЗ) з попередніх у цій послідовності ФЗ або БЗЗ ϕ_j, ϕ_k, $j, k < i$.

Як і раніше, послідовність $\phi_1, \phi_2, ..., \phi_{m-1}, \phi_m$ назвемо *доведенням* ϕ з об'єднання множин $F \cup G$.

Нехай задані деякі множини F і G ФЗ та БЗЗ відповідно. *Замикання* $[F \cup G]_R$ – це множина усіх ФЗ і БЗЗ, які синтаксично слідують з $F \cup G$ (відносно схеми R):

$$[F \cup G]_R \overset{def}{=} \{\phi \mid F \cup G \vdash \phi\}.$$

Лема 3.1. Виконуються властивості:

1) $F \cup G \subseteq [F \cup G]$ (*зростання*);
2) $[[F \cup G]] = [F \cup G]$ (*ідемпотентність*);
3) $F' \cup G' \subseteq F \cup G \Rightarrow [F' \cup G'] \subseteq [F \cup G]$ (*монотонність*[4]);
4) $[F] \subseteq [F \cup G]$, $[G] \subseteq [F \cup G]$;
5) $[F] \cup [G] \subseteq [F \cup G]$. □

Замиканням $[X]_{F \cup G, R}$ множини атрибутів X (відносно множини залежностей $F \cup G$ і схеми R) називається сім'я усіх правих частин БЗЗ, які синтаксично слідують з множини $F \cup G$:

$$[X]_{F \cup G, R} \overset{def}{=} \{Y \mid X \to\to Y \in [F \cup G]_R\}.$$

Очевидно, що $[X]_{F \cup G, R} \neq \varnothing$; можна посилити останнє твердження: насправді виконується включення $2^X \subseteq [X]_{F \cup G, R}$, де 2^X – булеан множини атрибутів X.

[4] Зрозуміло, що $F' \subseteq F$ та $G' \subseteq G$.

Нехай $[X]_F$ – замикання множини атрибутів X відносно множини ФЗ F. Зауважимо, що за означенням $[X]_F \subseteq R$.

Лема 3.2. Виконуються властивості:

1) $Y \subseteq [X]_F \Rightarrow Y \in [X]_{F \cup G,R}$;

2) $[X]_{F \cup G,R} = [[X]_F]_{F \cup G,R}$. □

Зауважимо, що оператор X a $[X]_{F \cup G,R}$ не є оператором замикання.

Базисом $[X]_{F \cup G,R}^{bas}$ множини атрибутів X відносно множини залежностей $F \cup G$ і схеми R називається підсім'я замикання $[X]_{F \cup G,R}$, така, що:

1) $\forall W (W \in [X]_{F \cup G,R}^{bas} \Rightarrow W \neq \varnothing)$ (тобто базис містить тільки непорожні множини атрибутів);

2) $\forall W_i W_j (W_i W_j \in [X]_{F \cup G,R}^{bas} \wedge W_i \neq W_j \Rightarrow W_i \cap W_j = \varnothing)$ (тобто різні множини атрибутів базису не перетинаються);

3) $\forall Y (Y \in [X]_{F \cup G,R} \Rightarrow \exists \Im (\Im \subseteq [X]_{F \cup G,R}^{bas} \wedge \Im$ – скінченна $\wedge Y = \bigcup_{W \in \Im} W$) (тобто кожна множина атрибутів з замикання $[X]_{F \cup G,R}$ є скінченним об'єднанням деяких множин атрибутів з базису).

Лема 3.3. Виконуються властивості:

1) $\bigcup_{W \in [X]_{F \cup G,R}^{bas}} W = R$ для $X \subseteq R$ (тобто базис є розбиттям множини атрибутів R);

2) $A \in [X]_F \Rightarrow \{A\} \in [X]_{F \cup G,R}^{bas}$. □

У **підрозділі 3.3 "Коректність та повнота аксіоматики багатозначних і функціональних залежностей"** представлено перший основний результат розділу 3: побудовано математичне доведення коректності аксіоматики ФЗ і БЗЗ та узагальнено і доповнено класичний результат щодо повноти аксіоматики ФЗ і БЗЗ в табличних БД.

Нехай ϕ – ФЗ або БЗЗ.

Твердження 3.1 (коректність аксіоматики БЗЗ і ФЗ). Якщо залежність ϕ синтаксично виводиться з множини залежностей $F \cup G$, то ϕ виводиться з $F \cup G$ семантично:

$$F \cup G \vdash \phi \Rightarrow F \cup G \models \phi. \ \square$$

Твердження 3.2 (повнота аксіоматики БЗЗ і ФЗ). Якщо залежність ϕ семантично виводиться з множини залежностей $F \cup G$, то ϕ виводиться з $F \cup G$ синтаксично за умови $|R| \geq 2$ та $|D| \geq 2$ (при розгляді моделей):

$$F \cup G \models \phi \Rightarrow F \cup G \vdash \phi. \ \square$$

Безпосередньо з тверджень 3.1 і 3.2 випливає доведення теореми 3.1.

Теорема 3.1. Відношення семантичного та синтаксичного слідувань для аксіоматики БЗЗ та ФЗ за умови $|D| \geq 2$ та $|R| \geq 2$ збігаються:

$$F \cup G \models \phi \Leftrightarrow F \cup G \vdash \phi. \ \square$$

Аналогічна теорема має місце і для аксіоматики БЗЗ.

У **підрозділі 3.4 "Критерій повноти аксіоматики багатозначних залежностей"** перевіряється збіжність відношень семантичного ‘ і синтаксичного ђ слідувань для аксіоматики БЗЗ у випадках, коли потужність універсального домену $|D|<2$ або потужність схеми $|R|<2$, і тим самим уточнюються умови повноти аксіоматики БЗЗ.

Основним результатом є наступна теорема.

Теорема 3.2. Відношення семантичного ‘ та синтаксичного ђ слідувань для аксіоматики БЗЗ співпадають тоді і тільки тоді, коли $|R|\leq 1$ або $|R|\geq 2$ і при цьому $|D|\geq 2$. □

У **підрозділі 3.5 "Критерій повноти аксіоматики багатозначних і функціональних залежностей"** також перевіряється збіжність відношень семантичного ‘ і синтаксичного ђ слідувань для аксіоматики БЗЗ і ФЗЗ у випадках, коли $|D|<2$ або $|R|<2$, і тим самим уточнюються умови повноти цієї аксіоматики.

Основним результатом є наступна теорема.

Теорема 3.3. Відношення семантичного ‘ та синтаксичного ђ слідувань для аксіоматики БЗЗ і ФЗ співпадають тоді і тільки тоді, коли $|D|\geq 2$ або $|R|=0$. □

Четвертий розділ "Фрагмент математичної теорії нормалізації: нормальні форми 2-4 порядків" присвячений НФ 2-4 порядків в табличних БД.

У **підрозділі 4.1 "Суперключі, потенційні ключі, первинні та непервинні атрибути, повні та транзитивні ФЗ"** наведені формулювання деяких основних понять з теорії нормалізації, а також, встановлені значення потенційних ключів у спеціальних випадках, коли $|D|\leq 2$ або $|R|\leq 2$.

Як і раніше F – множина ФЗ. Зафіксуємо схему R.

ФЗ $X\to Y$ поставимо у відповідність множину ФЗ:

$$\Phi_{X\to Y}=\{X'\to Y\mid X'\subset X\wedge Y Ъ X'\}.$$

Тобто, множина $\Phi_{X\to Y}$ містить усі нетривіальні ФЗ, такі, що їх права частина співпадає з множиною Y, а ліва частина є власною підмножиною множини X .

Очевидно, що $\Phi_{X\to Y}=\{X'\to Y\mid X'\in 2^X\setminus\{X\}\}$ за умови $X\subseteq Y$; $\Phi_{X\to Y}=\{X'\to Y\mid X'\in\{U\cup V\mid U\in 2^{X\setminus Y}\setminus\{X\}, V\in 2^Y\setminus\{Y\}\}\}$ за умови $Y\subseteq X$ або множини атрибутів X, Y непорівнювані відносно теоретико-множинного включення $\subseteq$. Вище 2^Y як і раніше – булеан множини Y.

Означення 4.1. ФЗ $X\to Y$ називається *повною* (відносно множини ФЗ F), якщо вона виконується на довільній таблиці t, яка є моделлю множини F, а жодна ФЗ з множини $\Phi_{X\to Y}$ не виконується на таблиці t:

$$X\to Y \text{ – } \textit{повна} \text{ (відносно множини ФЗ } F\text{)} \overset{def}{\Leftrightarrow} \forall t(t \text{ – модель } F\Rightarrow (X\to Y)(t)=true\wedge\forall\psi(\psi\in\Phi_{X\to Y}\Rightarrow\psi(t)=false)). \ \square$$

Враховуючи, що відношення синтаксичного і семантичного слідувань для аксіоматики ФЗ збігаються за умови $|D|\geq 2$, для вказаного випадку можна використовувати синтаксичну форму означення 4.1:

$$X \to Y \text{ – } \textit{повна} \text{ (відносно множини ФЗ } F\text{)} \overset{def}{\Leftrightarrow}$$
$$\overset{def}{\Leftrightarrow} X \to Y \in [F] \wedge \forall \psi (\psi \in \Phi_{X \to Y} \Rightarrow \psi \notin [F]),$$

де $[F]$ – синтаксичне замикання множини ФЗ F.

Означення 4.2. Множина атрибутів $W \subseteq R$ називається *суперключем* (відносно множини ФЗ F), якщо ФЗ $W \to R$ виконується на довільній таблиці t, яка є моделлю множини F:

$$W \text{ – } \textit{суперключ} \text{ (відносно множини ФЗ } F\text{)} \overset{def}{\Leftrightarrow} \forall t (t \text{ – модель}$$
$$F \Rightarrow \Rightarrow (W \to R)(t) = true). \ \square$$

Таким чином, наприклад, вся множина атрибутів схеми R є суперключем, оскільки ФЗ $R \to R$ є тривіальною і, значить, виконується на довільній таблиці.

Як і вище для випадку $|D| \geq 2$ можна використовувати синтаксичну форму означення 4.2:

$$W \text{ – } \textit{суперключ} \text{ (відносно множини ФЗ } F\text{)} \overset{def}{\Leftrightarrow} W \to R \in [F].$$

Означення 4.3. Множина атрибутів $K \subseteq R$ називається *потенційним* ключем (відносно множини ФЗ F), якщо K є суперключем і ФЗ $K \to R$ є повною. □

Змістовно кажучи, потенційний ключ – це суперключ, з якого не можна видалити жодного атрибута без втрати властивості суперключа.

Лема 4.1 (про існування потенційного ключа). Для довільної множини ФЗ F, заданої над схемою R, існує не менше одного потенційного ключа. □

Задамо на множині $2^R \times 2^R$ (по суті, на множині всіх ФЗ, бо ФЗ є парою множин атрибутів), наступне бінарне відношення. Скажемо, що ФЗ $X' \to Y'$ і $X \to Y$ знаходяться у відношенні $\leq$, якщо $X' \subseteq X$ та $Y' = Y$:

$$X' \to Y' \leq X \to Y \overset{def}{\Leftrightarrow} X' \subseteq X \wedge Y' = Y.$$

Очевидно, що відношення $\leq$ на множині $2^R \times 2^R$ задає частковий порядок.

Розглянемо частково впорядковану множину $\langle (\Phi_{R \to R} \cup \{R \to R\}) \cap [F], \leq \rangle$, де, як і раніше, $\Phi_{R \to R} = \{X \to R \mid X \subset R\}$. Тоді її кожний мінімальний елемент вигляду $X \to R$ є повною ФЗ і задає потенційний ключ X. Наявність декількох мінімальних елементів визначає декілька потенційних ключів. По суті це ще одне визначення потенційного ключа.

Лема 4.2. Якщо $|R| = 0$ або $|D| \leq 1$, то потенційним ключем буде тільки порожня множина $\varnothing$. □

Лема 4.3. Якщо $|R| > 0$ і $|D| \geq 2$, то ФЗ вигляду $\{A\} \to Y$, де $A \in R$, $\varnothing \neq Y \subseteq R$ є повною за умови $\{A\} \to Y \in [F]$ (тобто виконання ФЗ $\{A\} \to Y$ на довільній моделі множини ФЗ F). □

Означення 4.4. Атрибут $A \in R$ називається *транзитивно залежним* від множини атрибутів $X \subset R$ (відносно множини ФЗ F), якщо існує така множина атрибутів $Y \subset R$, що $A \notin X \cup Y$ і для довільної таблиці t, яка є моделлю множини ФЗ F, виконується $(X \to Y)(t) = true$, $(Y \to X)(t) = false$, $(Y \to \{A\})(t) = true$:

$$\exists Y(Y \subset R \wedge A \in R \setminus (X \cup Y) \wedge \forall t(t - \textit{модель}$$
$$F \Rightarrow (X \to Y)(t) = true \wedge (Y \to X)(t) = false \wedge (Y \to \{A\})(t) = true). \ \square$$

Для випадку $|D| \geq 2$ можна використовувати синтаксичну форму означення 4.4: атрибут A *транзитивно залежний* від множини атрибутів $X \subset R$ (відносно множини ФЗ F) $\overset{def}{\Leftrightarrow} \exists Y(Y \subset R \wedge A \in R \setminus (X \cup Y) \wedge X \to Y \in [F] \wedge Y \to X \notin [F] \wedge \wedge Y \to \{A\} \in [F])$.

Лема 4.4. Якщо K – потенційний ключ (відносно множини ФЗ F) і для деякого атрибута $A \in R$ ФЗ $K \to \{A\}$ не є повною, то атрибут A є транзитивно залежним від множини атрибутів K. □

У **підрозділі 4.2 "Друга нормальна форма"** розглянуті достатні умови для знаходження таблиці $t(R)$ у другій НФ (2НФ). Доведено, що у спеціальних випадках, коли $|R| \leq 2$ або $|D| \leq 1$, таблиця $t(R)$, яка знаходиться у першій НФ (1НФ) і є моделлю множини ФЗ F, знаходиться у 2НФ.

Як і раніше фіксуємо схему R та множину ФЗ F.

Означення 4.5. Скажемо, що таблиця $t(R)$ знаходиться у 2НФ, якщо, по-перше, таблиця $t(R)$ знаходиться у 1НФ[5], по-друге, таблиця t є моделлю множини ФЗ F та, по-третє, кожний непервинний атрибут характеризується повною ФЗ від кожного потенційного ключа, тобто:

$$\forall A \forall K(K \in \Delta \wedge A \in R \setminus \bigcup_{\acute{K} \in \Delta} \acute{K} \Rightarrow K \to \{A\} \text{ – повна ФЗ}). \ \square \quad (1)$$

Вище Δ – множина всіх потенційних ключів.

З означення 4.5 випливає, що, якщо множина непервинних атрибутів порожня, то умова (1) цього означення виконується автоматично (відповідна імплікація тривіально істинна, оскільки умова $A \in R \setminus \bigcup_{\acute{K} \in \Delta} \acute{K}$ хибна).

Твердження 4.1 (достатні умови для знаходження у 2НФ для спеціальних випадків). Таблиця t знаходиться у 2НФ якщо $|R| \leq 2$ або $|D| \leq 1$. □

Змістовна інтерпретація останнього твердження: не більше ніж двоатрибутні таблиці або не більше ніж однорядкові таблиці знаходяться у 2НФ (за умови, що вони є моделями множини ФЗ F).

В наступній лемі Δ – сім'я всіх потенційних ключів як і раніше.

Лема 4.5 (достатні умови для знаходження у 2НФ). Таблиця t, яка є моделлю множини ФЗ F, знаходиться у 2НФ, якщо виконується одна з наступних умов:

1) всі потенційні ключі одноелементні;

2) $\bigcup_{K \in \Delta} K = R$, тобто сім'я потенційних ключів є покриттям схеми R (еквівалентно: множина непервинних атрибутів порожня). □

[5] Оскільки означення 1НФ не підлягає повній формалізації, пропонуємо використовувати один з його варіантів: «таблиця знаходиться у 1НФ тоді і тільки тоді, коли кожний її рядок містить тільки одне значення для кожного атрибуту».

У **підрозділі 4.3 "Третя нормальна форма"** розглянуті достатні умови для знаходження таблиці $t(R)$ у третій НФ (3НФ). Доведено, що в спеціальних випадках, коли $|R| \leq 2$ або $|D| \leq 1$, таблиця $t(R)$, яка знаходиться у 1НФ і є моделлю множини ФЗ F, знаходиться у 3НФ.

Задамо на множині 2^R наступне бінарне відношення. Скажемо, що множини атрибутів $X, Y \subseteq R$ знаходяться у відношенні $\stackrel{\llcorner}{=}$ (відносно множини ФЗ F), якщо F ‘ $X \to Y$ і F ‘ $Y \to X$.

Лема 4.6. Відношення $\stackrel{\llcorner}{=}$ є відношенням еквівалентності. □

Таким чином, відношення $\stackrel{\llcorner}{=}$ розбиває множину 2^R (булеан схеми R) на класи еквівалентності.

Наслідок 4.1. За умов $|D| \leq 1$ або $|R| = 0$ усі елементи множини $2^R = \{X \mid X \subseteq R\}$ належать до одного класу еквівалентності. □

Лема 4.7. Множина суперключів належить до одного класу еквівалентності. □

Позначимо клас еквівалентності суперключів $[R]_{\stackrel{\llcorner}{=}}$, де схема R є представником класу.

Лема 4.8. Виконується імплікація:

$$\forall A, \forall X (\forall Y (A \in R \setminus (X \cup Y) \wedge (Y \to \{A\})(t) = true \Rightarrow X \stackrel{\llcorner}{=} Y) \Rightarrow$$
$$\Rightarrow A \text{ не є транзитивно залежним від } X). \ □$$

Означення 4.6. Скажемо, що таблиця $t(R)$ знаходиться у 3НФ, якщо, по-перше, таблиця $t(R)$ знаходиться у 1НФ, по-друге, таблиця t є моделлю множини ФЗ F та, по-третє, для кожної нетривіальної ФЗ виду $X \to \{A\}$, такої, що F ‘ $X \to \{A\}$, де A – непервинний атрибут, множина атрибутів X є суперключем, тобто:

$$\forall X, \forall A \, (X \subset R \wedge A \in R \setminus (\bigcup_{K \in \Delta} K \cup X) \wedge$$
$$\wedge (X \to \{A\})(t) = true \Rightarrow (X \to R)(t) = true). \ □ \qquad (2)$$

З означення 4.6 випливає, що, якщо множина непервинних атрибутів порожня, то умова (2) цього означення виконується автоматично (відповідна імплікація тривіально істинна). Таким чином, доведена наступна лема.

Лема 4.9 (достатня умова для знаходження у 3НФ). Таблиця t, яка є моделлю множини ФЗ F, знаходиться у 3НФ якщо $\bigcup_{K \in \Delta} K = R$, тобто сім'я потенційних ключів є покриттям схеми R. □

Твердження 4.2 (достатні умови для знаходження у 3НФ для спеціальних випадків). Таблиця знаходиться у 3НФ якщо $|R| \leq 2$ або $|D| \leq 1$. □

Змістовна інтерпретація останнього твердження: не більше ніж двоатрибутні таблиці або не більше ніж однорядкові таблиці знаходяться у 3НФ (за умови, що вони є моделями множини ФЗ F).

Лема 4.10. Якщо таблиця t знаходиться у 3НФ, то вона знаходиться у 2НФ: $3НФ \Rightarrow 2НФ$.□

У **підрозділі 4.4 "Нормальна форма Бойса-Кодда"** розглянуті достатні умови для знаходження таблиці $t(R)$ у нормальній формі Бойса-Кодда (НФБК). Доведено, що в спеціальних випадках, коли $|R| \leq 2$ або $|D| \leq 1$, таблиця $t(R)$, яка знаходиться у 1НФ і є моделлю множини ФЗ F, знаходиться також у НФБК.

Означення 4.7. Скажемо, що таблиця $t(R)$ знаходиться у НФБК, якщо, по-перше, таблиця $t(R)$ знаходиться у 1НФ, по-друге, таблиця t є моделлю множини ФЗ F та, по-третє, для кожної нетривіальної ФЗ виду $X \to \{A\}$ такої, що F ‘ $X \to \{A\}$, множина атрибутів X є суперключем, тобто:

$$\forall X, \forall A\, (X \subset R \wedge A \in R \setminus X \wedge (X \to \{A\})(t) = true \Rightarrow (X \to R)(t) = true). \square \quad (3)$$

Твердження 4.3 (достатні умови для знаходження у НФБК для спеціальних випадків). Таблиця знаходиться у НФБК якщо $|R| \leq 2$ або $|D| \leq 1$. □

Змістовна інтерпретація останнього твердження: не більше ніж двоатрибутні таблиці або не більше ніж однорядкові таблиці знаходяться у НФБК (за умови, що вони є моделями множини ФЗ F).

Доведено, що за умови $|R| \leq 2$ або $|D| \leq 1$ виконуються еквівалентності $НФБК \Leftrightarrow 3НФ \Leftrightarrow 2НФ$.□

Лема 4.11. Якщо таблиця t знаходиться у НФБК, то вона знаходиться у 3НФ: $НФБК \Rightarrow 3НФ$. □

У **підрозділі 4.5 "Четверта нормальна форма"** розглянуті достатні умови для знаходження таблиці $t(R)$ у четвертій НФ (4НФ). Доведено, що в спеціальних випадках, коли $|R| \leq 1$ або $|D| \leq 1$, таблиця $t(R)$, яка знаходиться у 1НФ і є моделлю множини ФЗ F, знаходиться у 4НФ.

Нехай задані множини ФЗ F і БЗЗ G.

Означення 4.8. Скажемо, що таблиця $t(R)$ знаходиться у 4НФ, якщо, по-перше, таблиця $t(R)$ знаходиться у 1НФ, по-друге, таблиця t є моделлю об’єднання множин ФЗ і БЗЗ $F \cup G$ та, по-третє, для кожної нетривіальної БЗЗ $X \to\to Y$, такої, що $F \cup G$ ‘ $X \to\to Y$, множина атрибутів X є суперключем:

$$\forall X, \forall Y (X \subset R \wedge Y \neq R \setminus X \wedge Y \text{Ъ} X \wedge$$
$$\wedge (X \to\to Y)(t) = true \Rightarrow (X \to R)(t) = true). \square \quad (4)$$

Твердження 4.4 (достатні умови для знаходження у 4НФ для спеціальних випадків). Таблиця знаходиться у 4НФ якщо $|R| \leq 1$ або $|D| \leq 1$. □

Зауважимо, що за умови $|R| = 2$ таблиця буде знаходитись у НФБК, але може не знаходитись в 4НФ.

Лема 4.12. Якщо таблиця $t(R)$ знаходиться у 4НФ, то вона знаходиться у НФБК: $4НФ \Rightarrow НФБК$.□

В роботі показано, що сформульовані означення 2-4 НФ задовольняють принципу «включення» НФ вищих порядків в НФ нижчих порядків, тобто виконуються імплікації: $4НФ \Rightarrow НФБК \Rightarrow 3НФ \Rightarrow 2НФ$.

ВИСНОВКИ

Основні результати дисертаційної роботи, які формують математичний фундамент нормалізації – основного механізму підтримки цілісності даних на етапі логічного моделювання (побудови схеми табличних БД).

1. Математично доведені повнота та коректність трьох основних аксіоматик залежностей в реляційних БД: по-перше, аксіоматики ФЗ Армстронга, по-друге, аксіоматики БЗЗ та, по-третє, аксіоматики ФЗ і БЗЗ через збіжність відношень синтаксичного та семантичного слідувань.
2. Математично визначені критерії вказаних трьох аксіоматик в термінах потужностей множини атрибутів та універсального домена.
3. Доведена незалежність складових аксіоматики ФЗ Армстронга (видалення кожної складової зменшує породжуючу силу аксіоматики).
4. Розвинений цілісний та несуперечливий фрагмент математичної теорії нормалізації щодо НФ 2-4 порядків, а також встановлені логічні зв‘язки між основними некласичними НФ.

Вказані результати дозволяють будувати математично обґрунтовані алгоритми нормалізації (а не на евристичних засадах).

СПИСОК ПРАЦЬ ОПУБЛІКОВАНИХ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Буй Д. Б. Повнота аксіоматики Армстронга / Д. Б. Буй, А. В. Пузікова // Вісник Київського національного університету імені Тараса Шевченка. Сер.: фіз.-мат. науки. – 2011. – № 3. – С. 103-108. (БД «ZBMATH – The database Zentralblatt MATH» Springer).
2. Редько В. Н. Аксіоматика багатозначних залежностей табличних баз даних / В. Н. Редько, Д. Б. Буй, А. В. Пузікова // Доповіді НАН України. – 2015. – № 6. – С. 24-29. (ISI (Master Journal List)).
3. Буй Д. Б. Деякі некласичні нормальні форми в реляційних базах даних / Д. Б. Буй, А. В. Пузікова // Вісник Київського національного університету імені Тараса Шевченка. Сер.: фіз.-мат. науки. – 2015. – № 1. – С. 65-74. (БД «ZBMATH – The database Zentralblatt MATH» Springer).
4. Буй Д. Б. Математична теорія нормалізації / Д. Б. Буй, А. В. Пузікова // Вісник Київського національного університету імені Тараса Шевченка. Сер.: фіз.-мат. науки. – 2015. – № 2. – С. 103-112. (БД ВІНІТІ Російської АН, БД «ZBMATH – The database Zentralblatt MATH» Springer).
5. Буй Д. Б. Теорія нормалізації в реляційних базах даних: сучасний стан / Д. Б. Буй, А. В. Пузікова // Наукові записки НаУКМА. Сер.: комп‘ютерні науки. – 2015. – Т. 177. – С. 83-92. (Scopus, Web of Science, Google Scholar).
6. Буй Д. Б. Незалежність аксіоматики Армстронга та алгебра функціональних залежностей / Д. Б. Буй, А. В. Пузікова // Штучний інтелект. – 2015. – № 1-2. – С. 121-126. (Google Scholar).
7. Bui D. Axiomatics for Multivalued Dependencies in Table Databases: Correctness, Completeness, Completeness Criteria / D. Bui, A. Puzikova // Theory and Engineering of Complex Systems and Dependability. Series “Advances in Intelligent Systems and Computing”. Vol. 365. – Springer International Publishing Switzerland, 2015. – P. 45-55. (Indexed by ISI Web of Science (previously ISI Proceedings), DBLP. Ulrich's, EI-Compendex, SCOPUS, Zentralblatt Math, MetaPress, Springerlink).

8. Буй Д. Б. Огляд теорії нормалізації в реляційних базах даних / Д. Б. Буй, А. В. Пузікова // Труды международной научно-технической конференции «Компьютерное моделирование в наукоемких технологиях» (Харьков, 28-31 мая 2014 г.) – С. 60-63.
9. Buy D. B. Theory of Normalization in Relation Databases (Survey) / D. B. Buy, A. V. Puzikova // Presented at 7th International Conference "Dependable Systems, Services and Technologies DESSERT 2014" (Kiev, Ukraine, May 16-18, 2014). Опубліковано в журналі: Радіоелектронні і комп'ютерні системи № 5(69). – Харків "ХАІ", 2014. – С. 45-49.
10. Буй Д. Б. Аксіоматика багатозначних залежностей табличних баз даних: повнота та її критерій / Д. Б. Буй, А. В. Пузікова // Матеріали XI Міжнародної наукової конференції «Теоретичні та прикладні аспекти побудови програмних систем» – TAAPSD'2014 (Київ, 15-17 грудня 2014 року). – С. 35-43.
11. Bui D. Axiomatics for multivalued dependencies in table databases: correctness and completeness / D. Bui, A. Puzikova // Proceedings of the Workshop on Foundations of Informatics – FOI-2015 (August 24-29, 2015, Chisinau, Republic of Moldova). – P. 361-376.
12. Буй Д. Б. Обзор современной теории нормализации в реляционных базах данных / Д. Б. Буй, А. В. Пузикова // Проблемы теоретической кибернетики. Материалы XVII Международной конференции (Казань, 16–20 июня 2014 г.). Под редакцией Ю.И. Журавлева. – Казань: Отечество, 2014. – С. 39-44.
13. Буй Д. Б. Критерии полноты аксиоматик зависимостей в табличных базах данных / Д. Б. Буй, А. В. Пузикова // Материалы XVII Международной конференции «Дискретные модели в теории управляющих систем» (Москва и Подмосковье, 20–22 мая 2015 г.). – М.: МАКС Пресс, 2015. – С. 37-39.

АНОТАЦІЯ

Пузікова А.В. Теорія нормалізації в табличних базах даних. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата фізико-математичних наук за спеціальністю 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем – Київський національний університет імені Тараса Шевченка МОН України. – Київ, 2016.

Робота присвячена аксіоматикам функціональних і багатозначних залежностей в табличних базах даних, а також нормальним формам з другої по четверту.

Побудовано математичні доведення відомих класичних результатів в теорії реляційних баз даних про повноту аксіоматики Армстронга для функціональних залежностей і повноту аксіоматики для функціональних і багатозначних залежностей, які задовольняють стандартним вимогам строгості та повноти математичного доведення. Встановлено критерії повноти для кожної з аксіоматик.

Доведена незалежність складових аксіоматики Армстронга. Побудована алгебра функціональних залежностей, операції якої визначені відповідно до

складових аксіоматики Армстронга. Побудова цієї алгебри дозволяє формулювати результати щодо властивостей аксіоматики Армстронга на алгебраїчній мові.

Побудовано фрагмент математичної теорії нормалізації щодо нормальних форм 2-4 порядків.

На основі аналізу першоджерел та власних результатів встановлені логічні зв'язки між означеннями класичних та основних некласичних нормальних форм (НФ). Показана нееквівалентність двох означень проективно-з'єднувальної (PJ/NF) НФ, запропонованих Фейгіним (Ronald Fagin).

Ключові слова: таблична база даних, теорія нормалізації, функціональна залежність, багатозначна залежність, нормальна форма.

АННОТАЦИЯ

Пузикова А.В. Теория нормализации в табличных базах данных. – Рукопись.

Диссертация на соискание ученой степени кандидата физико-математических наук по специальности 01.05.03 – математическое и программное обеспечение вычислительных машин и систем. – Киевский национальный университет имени Тараса Шевченко. – Киев, 2016.

Целью диссертационной работы являются аксиоматики функциональных и многозначных зависимостей в табличных базах данных (БД), а также фрагмент математической теории нормализации для нормальных форм со второй по четвертую.

Во многих современных направлениях развития технологий БД в той или иной степени используется реляционная модель данных, которая была предложена Е. Коддом в 70-х годах XX в.

Одним из основных требований, которые выдвигаются перед разработчиками БД, является обеспечение их надежности, которая, в свою очередь, в большой степени зависит от правильности логического проектирования схемы БД. Проблема устранения известных аномалий (обновления, вставки, удаления) в реляционных БД, которая возникает на этапе логического проектирования в результате наличия разных видов ограничений (например, функциональных и многозначных зависимостей), требует проведения нормализации – приведения к соответствующим нормальным формам.

В диссертационной работе выполнен обзор существующей литературы по теории нормализации в реляционных (табличных) БД. На основании анализа первоисточников и собственных результатов установлены логические связи между определениями классических и основных неклассических нормальных форм. Показана неэквивалентность двух определений проективно-соединительной нормальной формы, предложенных Фейгиным (Ronald Fagin).

Теория нормализации базируется на теории функциональных и многозначных зависимостей, в основании которой лежат соответствующие аксиоматики и утверждения об их корректности и полноте. Анализ научной и методической литературы показал, что в ней отсутствуют доказательства указанных результатов, которые удовлетворяли бы стандартным требованиям к строгости математического

доказательства, а также отсутствуют критерии, при которых аксиоматики функциональных и многозначных зависимостей являются полными.

В работе даны математические доказательства корректности аксиоматики Армстронга для функциональных зависимостей и аксиоматики для многозначных и функциональных зависимостей, а также обобщены и дополнены известные результаты о полноте этих аксиоматик. Поскольку стандартные доказательства полноты построены в предположении, что универсальный домен содержит не менее двух элементов, а схема – не менее двух атрибутов, то для каждой из указанных аксиоматик установлены критерии полноты, т.е. рассмотрены все возможные случаи для мощностей универсального домена и схемы.

Естественным продолжением, аналогично аксиоматическим системам, стало доказательство независимости компонент аксиоматики Армстронга.

Построена алгебра функциональных зависимостей, операции которой определены в соответствии с составляющими аксиоматики Армстронга. Рассмотрена подалгебра тривиальных функциональных зависимостей и множество, которое ее порождает. Построение этой алгебры позволяет формулировать результаты, касающиеся свойств аксиоматики Армстронга, на алгебраическом языке.

Поскольку различные подходы к понятийному аппарату теории нормализации в табличных БД привели к разнообразию, а часто и к несогласованности в определениях нормальных форм, возникла потребность в построении цельного непротиворечивого фрагмента математической теории нормализации относительно нормальных форм 2-4 порядков, что и было сделано в заключительном разделе работы.

Ключевые слова: табличная база данных, теория нормализации, функциональная зависимость, многозначная зависимость, нормальная форма.

ABSTRACT

Puzikova A.V. Theory of normalization in table databases. – Manuscript.

Candidate's thesis on Physics and Mathematics, speciality 01.05.03 – Theoretical and software of computers and systems. – Taras Shevchenko National University of Kyiv. – Kyiv, 2016.

The purpose of the candidate's thesis is the axiomatics for functional and multivalued dependencies in table databases and a fragment of mathematical normalization theory for 2-4 normal forms.

In the work it is represented a review of the existing literature on the theory of normalization in relational (table) databases. 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. It is shown that two definitions of Fagin's Projection-Join Normal Form (PJ/NF) are not equivalent.

The mathematical proofs of correctness of the axiomatics for functional dependencies and axiomatics for functional and multivalued dependencies are constructed. Known in the literature proofs of the completeness of these axiomatics are reconstructed and complemented (in terms of coincidence of syntactic and semantic consequence relations).

Proofs of completeness of these axiomatics are constructed under the assumption, that universal domain contains at least two elements, a scheme – at least two attributes. To correct it the completeness criteria for each of axiomatic systems in terms of cardinalities of the universal domain and scheme are established.

Also, it is shown that components of Armstrong's axiomatic system for the functional dependences of relational databases are independent.

Algebra of functional dependencies with operations that are induced by components of the Armstrong's axiomatics is built. The construction of this algebra allows us to formulate the results concerning properties of Armstrong's axiomatic system in algebraic language.

Since different approaches to conceptual apparatus of the normalization theory in relational databases have caused a variety, and often, inconsistencies in the definitions of the normal forms, in the work is constructed a consistent fragment of mathematical normalization theory for 2-4 normal forms.

Key words: table database, normalization theory, functional dependency, multivalued dependency, normal form.