

УДК 004.358:681.518

DOI: 10.30837/0135-1710.2025.186.029

*О.В. ЗОЛОТУХІН, М.С. КУДРЯВЦЕВА, В.О. ФІЛАТОВ***ПІДТРИМКА ОБМЕЖЕНЬ ЦІЛІСНОСТІ РЕЛЯЦІЙНОЇ БАЗИ ДАНИХ НА ЕТАПАХ ЕКСПЛУАТАЦІЇ ТА РЕІНЖИНІРИНГУ У ЗАДАЧАХ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ**

Розглянуто вирішення задачі підтримки цілісності реляційної бази даних на етапах експлуатації та реінжинірингу за умов виявлення раніше не відомих функціональних залежностей з множини даних. Завдання виявлення прихованих залежностей є складовою задачі інтелектуального аналізу даних при реінжинірингу і належить до нового класу актуальних завдань користувача. Описаний підхід є варіантом для побудови автоматизованого рішення, безпосередньо орієнтованого на виявлення нових залежностей у даних, що породжуються предметною областю.

1. Вступ

Сучасні інформаційні системи (ІС), побудовані в комплексі із системами управління базами даних (СУБД), покликані виконувати завдання користувача у важких умовах безперервного потоку вимог предметної області (наприклад, удосконалення бізнес-процесів, розширення кола інтересів організації, різні вимоги з боку користувачів). Наслідком цього є адаптація окремих компонентів ІС, зокрема бази даних (БД) під нові вимоги. На сьогоднішній день найбільшого поширення набули реляційні бази даних (РБД), які забезпечують найкраще поєднання простоти, надійності та продуктивності для вирішення широкого класу задач.

Основою всіх сучасних інформаційних, як локальних, так і розподілених систем є БД. Досвід розробки, впровадження та експлуатації таких систем показав, що найефективнішою структурою, яка б задовольняла вимогам як розробника, так і користувача, є реляційна модель даних. Теоретичні основи проектування РБД, запропоновані Коддом, з 1970 року практично не змінилися [1].

Неодноразово робилися спроби ревізії цієї моделі, заміни її на різні варіації з об'єктно-орієнтованим підходом, проте досвід розробників і час показали: за сукупністю виразних засобів та математичної основи реляційна модель даних ще довго буде основною моделлю при проектуванні як локальних, так і розподілених БД.

Дане дослідження присвячене дослідженню методів підтримки цілісності даних у реляційних системах, орієнтованих на методи інтелектуального аналізу та реінжинірингу ІС.

2. Аналіз наукових публікацій і постановка задачі дослідження

Одним з основних понять у технології БД є поняття цілісності. У реляційній моделі даних об'єкти представлені у вигляді сукупності взаємозалежних відношень. Цілісність БД – це правила та засоби, що забезпечують надійну реалізацію встановлених зв'язків між усіма даними, що містяться в БД.

Будь-яка зміна в предметній області на етапах експлуатації, значуща значима для побудованої моделі, повинна відображатися в БД зі зберіганням однозначної інтерпретації інформаційної моделі. Підтримка цілісності в реляційній моделі даних у її класичному розумінні – це підтримка структурної цілісності, яка сприймається як те, що реляційна СУБД повинна допускати роботу лише з однорідними структурами даних типу «реляційне відношення». При цьому поняття «реляційного відношення» має задовольняти всі обмеження, що накладаються на нього в класичній реляційній теорії: відсутність дублікатів кортежів (рядки відношень), відповідно, обов'язкова наявність первинного ключа (один або кілька стовпців (атрибутів), які однозначно ідентифікують кожен запис у таблиці, тобто дозволяють

чітко відрізнити один запис від іншого), відсутність поняття впорядкованості кортежів.

Особливістю проведених досліджень є аналіз змін функціональних залежностей (ФЗ) атрибутів реляційної моделі даних у процесі експлуатації ІС з метою ефективного реінжинірингу схеми реляційної моделі даних з дотриманням умов підтримки цілісності.

Існуючі підходи реінжинірингу РБД ґрунтуються на використанні деякої вихідної множини ФЗ, сформованої на підставі припущення про можливі взаємозв'язки об'єктів предметної області. При цьому велика ймовірність того, що значні зв'язки будуть втрачені з уваги, що негативно вплине (надмірність інформації, непереборні аномалії, протиріччя) на подальше проєктування ІС, що включає цю БД і подальшу експлуатацію.

У той же час у наукових публікаціях [2], [3] велика увага приділяється методам виявлення ФЗ у існуючих екземплярах відношень РБД, що дозволяє виявити приховані взаємозв'язки між даними, які, у свою чергу, можуть бути використані для побудови логічної структури, максимально наближеної до тієї, що обумовлюється предметною областю, що розглядається. Слід зазначити, що методи, що розглядаються, спрямовані, в основному, на використання в системах інтелектуального аналізу даних (Data Mining) і орієнтовані на виявлення наближених ФЗ (Approximate Functional Dependencies — AFD).

У рамках даного дослідження використання таких методів дозволяє отримувати множину строгих ФЗ, тобто справедливих для всього набору вхідних даних на момент проведення обробки.

Найвідоміший та найпоширеніший метод нормалізації, що використовується для проєктування логічної схеми, дозволяє перейти від початкового ставлення до набору відношень у третій нормальній формі (3НФ) шляхом застосування низки правил декомпозиції. Ф. Бернштейн висунув альтернативну теорію: оскільки ФЗ повністю визначають, чи міститься реляційне відношення у 3НФ, то можна побудувати 3НФ, виходячи з набору ФЗ, які задовольняють вихідному універсальному відношенню [4].

У наступній роботі [5] він запропонував послідовність алгоритмів, що дозволяють синтезувати логічну схему РБД у 3НФ. При цьому, при реінжинірингу ІС необхідно провести додатковий аналіз цілей цієї системи і виявити вимоги до неї окремих користувачів.

3. Мета і задачі дослідження

Метою проведених досліджень є аналіз особливостей інформаційних одиниць і структур даних, які впливають на технологію видобування знань, розробка структур зберігання даних, орієнтованих на ефективну спільну роботу з системами аналітичної обробки, за умов підтримки цілісності БД.

Для досягнення цієї мети треба вирішити такі задачі:

- аналіз специфікації реляційної моделі даних з метою формального опису задачі видобування знань;
- виявлення раніше не відомих ФЗ з даних множини цільової БД;
- дослідження багатозначних ФЗ загальної теорії нормалізації відношень за обов'язкової умови – підтримки обмежень цілісності реляційної моделі даних.

4. Матеріали і методи дослідження

Об'єктом дослідження є РБД – БД, заснована на реляційній моделі. Предметом дослідження є методи підтримки цілісності даних у реляційних системах, орієнтованих на інтелектуальний аналіз.

Для проведення подальших досліджень властивостей ФЗ реляційної моделі розглянемо її теоретико-множинне представлення.

Нехай R – кінцева підмножина імен відношень БД; $D = (D_1, \dots, D_i)$ – множина доменів, де всякий домен D_i є іменованою множиною атомарних значень елементів даних; A – кінцева множина імен атрибутів відношення; dom – відображення з A в D , яке визначає,

з якого домену обрані значення атрибутів.

Пару $\langle A_i, \text{dom} A_i \rangle$, де $A_i \in A$, називають атрибутом. Структурну схему S_i відношення R_i ($R_i \in R$) можна представити у вигляді $R_i(A_1, \dots, A_n)$, у якому всі A_i різні. Відношення r_i можна визначити як розширення схеми $S_i: r_i \subseteq \text{dom} A_1 \times \dots \times \text{dom} A_n$. Перестановка атрибутів у схемі не породжує нового розширення та множина $\{A_1, \dots, A_n\}$ атрибутів відношення R_i задає тип відношення. Задля специфікації складу носія використовується вираз $R_i = A_1 \dots A_n$. Структурна схема УРБД – це специфікація вигляду $(R_1, \dots, R_p)$, де $R_i \in R$ і всі R_i різні.

Концептуально реляційна база є інформаційно-логічною моделлю певної предметної області, такою, що кожне розширення відповідає деякому стану даної області у певний дискретний момент поточного часу. Кожен стан моделюється впорядкованою сукупністю значень елементів даних, відповідних значенням властивостей об'єктів предметної області [6].

Зауважимо, що реляційна модель даних передбачає сильну типізацію об'єктів, використання цілком певних категорій, таких як тип об'єкта, атрибут (властивість) об'єкта, домен. Об'єкти мають набір властивостей, що задаються в реляційній моделі схемою відношення.

5. Результати дослідження

5.1. Постановка задачі аналізу даних

Формально задача видобування знань з БД може бути представлена у такому вигляді. Предметна область відображається реляційною моделлю, яка представлена у вигляді універсального відношення R як підмножина кортежів декартового добутку $R = (DX_1, DX_2, \dots, DX_n, DY_1, \dots, DY_m) = \{ \langle x_1, \dots, x_n, y_1, \dots, y_m \rangle \}$, де x_i – значення вхідних атрибутів X_i з домена DX_i ; y_i – значення вихідних атрибутів Y_i з домена DY_i ; $P(x_1, \dots, x_n, y_1, \dots, y_m)$ – предикат як умова відображення конкретної предметної області у кортежі значень атрибутів $\langle x_1, \dots, x_n, y_1, \dots, y_m \rangle$.

5.2. Виявлення нових залежностей у даних внаслідок експлуатації

Одним із сучасних методів інтелектуального аналізу даних може бути метод автоматичного виявлення взаємозв'язків між ними. В основі такого підходу може бути використаний метод аналізу ФЗ у схемі даних. Це зумовлено тим, що ФЗ дозволяють найпростішим чином уявити зв'язки між об'єктами аналізованої предметної області. Іншими типами залежностей, які беруться до уваги, є залежності включення та багатозначні залежності [2].

Слід зазначити, що розглянуті методи спрямовані, переважно, на використання у системах інтелектуального аналізу даних і орієнтовані на виявлення наближених ФЗ. У рамках даного дослідження використання таких методів дозволяє також отримувати множину строгих ФЗ, тобто справедливих для всього набору вхідних даних на момент проведення обробки [7].

5.3. Задача автоматичного виявлення взаємозв'язків між даними

Вихідними даними для вирішення поставленої задачі є: логічна схема РБД $\Sigma = \{\sigma_i, i = \overline{1, n}\}$, де σ_i – схема одного відношення, що входить до БД; n – кількість відношень; схема відношення $\sigma_i = \langle R_i, F_i \rangle$, де R_i – носій відношення (множина атрибутів); F_i – множина ФЗ, що задовольняють даному відношенню, $P = \{\rho_i, i = \overline{1, n}\}$ – множина відношень БД.

Існування ФЗ вигляду $A \rightarrow B$ означає, що для будь-яких двох кортежів u, v деякого відношення ρ_k має місце наслідок: $u(A) = v(A) \Rightarrow u(B) = v(B)$.

Як приклад розглянемо логічну схему $\Sigma = \{\sigma_1, \sigma_2\}$, що складається з двох схем відношень: $\sigma_1 = \langle R_1 = \{A, B, C\}, F_1 \rangle, \sigma_2 = \langle R_2 = \{C, D\}, F_2 \rangle$.

Припустимо, що інформація про F_1, F_2 відсутня або втрачена. Отримати множину ФЗ, що задовольняють даним відношенням, можливо за допомогою методу виявлення ФЗ з екземплярів відношень, що розглядаються, зокрема методу Тапе [8].

В результаті його застосування буде отримано множину мінімальних ФЗ, що задовольняють набору даних на момент проведення обробки даних. Мінімальна ФЗ – це така ФЗ вигляду $X \rightarrow Y, X = \{A_1, \dots, A_n\}, Y = \{B_1, \dots, B_m\}$, в якій не існує множини $Z \subset X$, для якої б дотримувалося $Z \rightarrow Y$.

Тривіальні ФЗ виду $A_i \rightarrow A_i$ ігноруються використанням методом, оскільки є значущими.

Розглянемо як приклад відношення ρ_1, ρ_2 , фрагменти структури яких наведено на рис.1:

ρ_1			ρ_2	
A	B	C	C	D
1	1	1	1	1
2	2	1	2	1
1	2	2	3	1
1	3	3	4	1
3	1	1		
4	2	2		
2	2	4		

а) фрагмент структури відношення ρ_1 б) фрагмент структури відношення ρ_2

Рис.1. Фрагменти структури відношень ρ_1, ρ_2

З використанням методу Тапе отримано такі множини ФЗ для наведених відношень: $F_1 = \{AC \rightarrow B\}, F_2 = \{C \rightarrow D\}$. Беручи до уваги множину ФЗ для F_1 и F_2 , множина ФЗ схеми Σ буде мати вигляд:

$$F_{\Sigma} = \bigcup_{i=1}^2 F_i = \{AC \rightarrow B, C \rightarrow D\}. \quad (1)$$

Носій універсального відношення $R = \{R_1 \cup \dots \cup R_n\}$ для прикладу, що розглядається, виглядає таким чином:

$$R = R_1 \cup R_2 = \{A, B, C\} \cup \{C, D\} = \{A, B, C, D\}. \quad (2)$$

Універсальне відношення може бути отримане через природне поєднання всіх відношень, що входять до схеми. Результат з'єднання наведено на рис. 2.

Із застосуванням методу Тапе для отриманого універсального відношення виявлено такі ФЗ: $\overline{F_{\Sigma}} = \{AC \rightarrow B, C \rightarrow D, A \rightarrow D, B \rightarrow D\}$. Ця множина містить усі мінімальні ФЗ, що належать F_{Σ} , а також додаткові, раніше не відомі ФЗ: $F'_{\Sigma} = \{A \rightarrow D, B \rightarrow D\}$.

Отже, множину ФЗ універсального відношення для логічної схеми Σ можна виразити таким чином:

A	B	C	D
1	1	1	1
2	2	1	1
1	2	2	1
1	3	3	1
3	1	1	1
4	2	2	1
2	2	4	1

Рис.2. Фрагменти структури універсального відношення

$$\overline{F_{\Sigma}} = F_{\Sigma} \cup F'_{\Sigma}, \quad (3)$$

де F_{Σ} – множина ФЗ, що задовольняють вихідним відношенням схеми Σ ; F'_{Σ} – множина додаткових ФЗ.

Необхідно встановити, чи є ФЗ із множини F'_{Σ} виведеними з F_{Σ} , або вони є новою інформацією. Для цього пропонується використовувати метод перевірки належності ФЗ до замикання $(F_{\Sigma})^+$, запропонований Мейєром [9], – вирішення проблеми членства. Принцип полягає в такому: оскільки побудова F^+ пов'язана з перебором всіх підмножин множини атрибутів, що належать F , і має експоненційну складність, то пропонується будувати F – замикання на множині атрибутів. F – замиканням множини X називається така множина атрибутів X^+ , що $X \rightarrow X^+ \in F^+$ і не існує жодного атрибута в R , який би залежав від X та не належав X^+ .

Реалізація методу побудови F – замикання має лінійну складність. Таким чином, метод перевірки належності ФЗ $X \rightarrow Y$ до замикання F^+ полягає у побудові F – замикання X^+ та визначення істинності виразу $Y \subseteq X^+$. Якщо вираз істинний, то $X \rightarrow Y \in F^+$.

Розглянемо приклад: для того, щоб перевірити $A \rightarrow D$ на належність до F_{Σ}^+ , потрібно побудувати A^+ . Відповідно, $A^+ = \{A\}$, оскільки F_{Σ} не містить ФЗ, для яких A була б єдиним атрибутом у лівій частині ФЗ. $D \notin A^+$, отже, $A \rightarrow D \notin F_{\Sigma}^+$. Аналогічним чином можна довести, що $B \rightarrow D \notin F_{\Sigma}^+$. Таким чином, множина F'_{Σ} виявлених залежностей не виводиться і, тим самим, є новою інформацією.

Слід зазначити, що цей підхід не гарантує повну відповідність нових залежностей предметної області, що розглядається. Оскільки він базується на множині даних, що міститься в РБД на момент проведення обробки, і не враховує їхню семантику, велика ймовірність отримання випадкових ФЗ. Випадкова ФЗ – це така ФЗ, яка не є коректною для конкретної предметної області (наприклад, дата народження людини визначає дату народження її дитини) і може бути усунена в будь-який момент шляхом зміни або додавання кортежів з даними, що суперечать виявленій залежності, у процесі функціонування [10].

Таким чином, ставиться ще одне завдання – перевірка нових залежностей щодо коректності для предметної області, яка моделюється аналізованою РБД. У даному дослідженні її рішення не розглядалися і є напрямом для подальших досліджень. Як можливі шляхи рішення можна запропонувати використання експертної оцінки, або ж певного чисельного критерію для кожної окремої ФЗ, що дозволить встановлювати граничне значення, нижче за яку залежність вважатимуться випадковими.

5.4. Обмеження цілісності у реляційній моделі даних

Сучасні СУБД мають низку засобів за для забезпечення підтримки цілісності, як і засобів забезпечення підтримки безпеки [11].

Виділяють три групи правил цілісності: цілісність за сутностями, цілісність за посиланнями, цілісність, що визначається користувачем. Розглянемо опис правил цілісності, загальних для будь-яких РБД.

1. Не допускається, щоб будь-який атрибут, що бере участь у первинному ключі, набував невизначеного значення.

2. Значення зовнішнього ключа має бути рівним значенню первинного ключа або повністю невизначеним.

3. Для будь-якої конкретної БД існує низка додаткових специфічних правил, які належать до неї і визначаються розробником. Найчастіше контролюється унікальність тих чи інших атрибутів, діапазон значень (екзаменаційна оцінка від 60 до 100 балів), приналежність до набору значень (стать «Ч» або «Ж»).

У реляційній моделі об'єкти реального світу представлені у вигляді сукупності взаємозалежних відношень. Будь-яка зміна в предметній галузі, значуща для побудованої моделі, повинна відображатися в БД зі зберіганням однозначної інтерпретації інформаційної моделі в термінах предметної області.

Ми зазначили, що лише суттєві або значущі зміни предметної галузі мають відстежуватись у інформаційній моделі. Дійсно, модель завжди є деяким спрощенням реального об'єкта, в моделі ми відображаємо тільки те, що нам важливо для вирішення конкретного набору завдань.

Розглянемо як приклад фрагмент ІС «Бібліотека» та відповідну реляційну модель БД. Приклад 1.

Як основні сутності предметної області «Бібліотека» можна виділити такі:

- СТУДЕНТ (Номер залікової книжки, Прізвище, Стать, Група, Дата народження);
- ЛІТЕРАТУРА (Код книги, Назва, Автор, Видавництво, Рік, Сторінок, Тип літератури).

Очевидно, як основне логічне (або семантичне) обмеження цілісності буде таке: будь-який студент в один день відвідування бібліотеки не може взяти на руки дві однакові книги або журнали, два однакові довідники. Якими засобами реляційної моделі підтримується це обмеження? Вирішити це завдання досить просто. Відношення «СТУДЕНТ» та відношення «ЛІТЕРАТУРА» засобами реляційної моделі будуть пов'язані між собою проміжним відношенням «СТУДЕНТИ_ЛІТЕРАТУРА».

Підтримку цілісності при цьому буде покладено на СУБД реляційного типу, а як механізм підтримки цілісності виступатимуть властивості складеного ключового атрибуту відношення: СТУДЕНТИ_ЛІТЕРАТУРА (Номер залікової книжки, Код книги, Дата видачі).

СТУДЕНТИ				ЛІТЕРАТУРА		СТУДЕНТИ_ЛІТЕРАТУРА		
Номер	Прізвище	Група	Стать	Код	Автор	Номер	Код	Дата
1	Іванов	ПМ-08-1	Ч	Д.01.001	Мартин Д.	1	Д.01.001	05.02.11
2	Петрова	ПМ-08-1	Ж	Д.01.002	Кодл Е.	1	Д.01.001	12.02.11
3	Сидоров	ПМ-08-2	Ч			1	Д.01.001	15.02.11
						2	Д.01.001	05.02.11
						3	Д.01.002	05.02.11

Рис.3. Фрагменти відношень інформаційної системи «Бібліотека»

Розглянутий вище приклад показує ефективність реляційного підходу: теоретично правильно спроектована модель БД, у разі для ІС «Бібліотека», автоматично захищає себе лише на рівні обмежень цілісності даних протягом усього періоду експлуатації. СУБД реляційного типу ніколи не дозволить порушити задані та реалізовані на рівні моделі обмеження цілісності. Таким чином, можна зробити висновок: основним засобом підтримки логічних обмежень цілісності в реляційній моделі даних є управління ключовими атрибутами відношень. Наведемо ще кілька прикладів на підтвердження викладеного вище затвердження.

Приклад 2.

Обмеження цілісності: студент навчального закладу один предмет може вивчати протягом кількох семестрів. Структуру відношення представлено на рис.4.

СТУДЕНТИ_СЕМЕСТРИ_ПРЕДМЕТЫ_ВИКЛАДАЧИ

Код	Прізвище	Семестр	Код предмета	Предмет	Оцінка	Викладач
1	Іванов	4	221	Фізика	В	Суворов В.А.
1	Іванов	5	221	Фізика	С	Куртин П.А.
2	Петров	4	221	Фізика	В	Сидоров В.А.
2	Петров	5	221	Фізика	В	Куртин П.А.
2	Петров	6	231	Хімія	А	Уваров А.О.

Рис.4. Фрагменти відношення інформаційної системи «Деканат»

Відношення СТУДЕНТИ_СЕМЕСТРИ_ПРЕДМЕТИ_ВИКЛАДАЧИ ключові атрибути – **Код** студента, **Семестр**, **Код** предмета – реалізують технологію підтримки обмежень цілісності.

Приклад 3.

Предметна сфера та обмеження цілісності: в ІС «Склад» при формуванні податкових накладних слід враховувати обмеження цілісності такого вигляду: щодня нумерація накладних починається з першого номера. Таким чином, накладна з номером «один» зустрічатиметься щодня, якщо склад цього дня працював. Структуру відношення, що підтримує цілісність лише на рівні ключових атрибутів, представлено на рис. 5.

НАКЛАДНІ_ПОСТАЧАЛЬНИКИ_СПОЖИВАЧІ

Номер накладної	Дата накладної	Код постачальника	Код споживача	Постачальник	Споживач	Сума накладної
1	05.02.11	1	1	СМІТ	ФОКСМАРТ	3800
2	05.02.11	1	2	СМІТ	АВС	5600
1	06.02.11	1	1	СМІТ	ФОКСМАРТ	12700
2	06.02.11	1	2	СМІТ	АВС	46700
3	06.02.11	1	3	СМІТ	МКС	14540

Рис.5. Фрагмент відношення інформаційної системи «Склад»

У відношенні НАКЛАДНІ_ПОСТАЧАЛЬНИКИ_СПОЖИВАЧІ ключові атрибути – **Номер** накладної, **Дата** накладної – реалізують технологію підтримки цілісності.

Проте існує клас задач, котрим виконати логічні обмеження цілісності з допомогою технології складових ключових атрибутів однозначно не завжди вдається. До таких задач належать завдання складання різноманітних розкладів. Розглянемо докладно загальну постановку такого класу завдань.

Приклад 4.

Нехай у розпорядженні розробника інформаційної системи «Розклад занять» є наведені до ЗНФ реляційні відношення: ВИКЛАДАЧІ, АУДИТОРІЇ, ПАРИ (див. рис. 6).

ВИКЛАДАЧІ		АУДИТОРІЇ		ПАРИ	
Код	Прізвище	Код	Аудиторія	Код	Пара
1	Іванов	1	101	1	перша
2	Петров	2	102	2	друга
3	Мухін	3	103	3	третя

Рис.6. Основні відношення інформаційної системи «Розклад занять»

Обмеження цілісності сформулюємо у такому вигляді:

- викладач може проводити заняття лише один раз на одній парі у будь-якій аудиторії;
- одна аудиторія на одній парі може бути зайнята лише один раз.

На попередньому етапі синтезу загальної моделі даних доцільно об'єднати попарно відношення ВИКЛАДАЧІ та ПАРИ через проміжне відношення ВИКЛАДАЧІ_ПАРИ. У цьому випадку подвійний складовий ключ **Код_викладача** та **Код_пари** дозволить задовольнити обмеження цілісності даних. Аналогічно сформуємо відношення АУДИТОРІЇ_ПАРИ (**Код_аудиторії** та **Код_пари**). І це відношення вирішує завдання підтримки цілісності. Однак залишається відкритим питання – якою буде загальна реляційна модель БД, що дозволить зберігати дані трьох відношень: ВИКЛАДАЧІ, АУДИТОРІЇ, ПАРИ. Сформуємо, згідно з теорією синтезу реляційних моделей, універсальне відношення та всі атрибути визначимо як ключові (див. рис. 7):

ВИКЛАДАЧІ_ПАРИ_АУДИТОРІЇ

<u>Викладач</u>	<u>Пара</u>	<u>Аудиторія</u>
Іванов	перша	101
Іванов	перша	102
Іванов	друга	101
Петров	друга	101

Рис.7. Відношення «ВИКЛАДАЧІ_ПАРИ_АУДИТОРІЇ»

Аналіз наведеного вище відношення дозволяє зробити такий висновок: якщо виділені атрибути об'єднані в подвійний складовий ключ (ВИКЛАДАЧ_ПАРА), тоді виконуються обмеження цілісності – всі кортежі унікальні, і один викладач може бути доданий лише у поєднанні з новою парою занять. Це обмеження одного чіткого правила. Але щойно додається новий атрибут, у нашому прикладі АУДИТОРІЯ, і цей атрибут додається до

складу складеного ключа, попереднє обмеження підтримки цілісності подвійного складеного ключа автоматично порушується.

На рис.7 всі атрибути ключові і про підтримку цілісності лише на рівні ВИКЛАДАЧ_ПАРА мова вже йти не може. Іванов на першій парі може бути як в аудиторії 101, так і в аудиторії 102. З погляду реляційної моделі все вірно – всі кортежі різні, а обмеження цілісності ВИКЛАДАЧ_ПАРА та АУДИТОРІЯ_ПАРА не виконуються.

У реляційній теорії БД розглянутий вище випадок відноситься до дослідження багатозначних ФЗ загальної теорії нормалізації відношень.

5.5. Декомпозиція універсального відношення у випадку багатозначних залежностей

Технологія декомпозиції універсального відношення за багатозначних залежностей атрибутів дозволить автоматично підтримувати обмеження цілісності безпосередньо при введенні даних.

Нехай r – відношення зі схемою R . X, Y, Z – підмножина з R , така що $Z = R - (X, Y)$.

Визначення. Відношення задовольняє множинним ФЗ тоді, коли поділяється без втрат на відношення $R_1 = XY$ і $R_2 = XZ$.

Слідство з визначення. Якщо відношення розділимо на два відношення: $R_1 = XY$ і $R_2 = XZ$ без втрат, то відношення r може бути замінено еквівалентним відношенням r' , яке задовольняє обмеженням цілісності, як для відношення R_1 , так і для відношення R_2 .

Декомпозиція відношення: $R_1(\underline{X}, \underline{Y}) \rightarrow R_{11}(\underline{X}), R_{12}(\underline{Y})$; $R_2(X, \underline{Z}) \rightarrow R_{21}(\underline{Z})$; $r'(X, \underline{Y}, \underline{Z})$.

Перетворення відношення, що містить багатозначні залежності, розглянемо на прикладі (відношення?) «ВИКЛАДАЧІ ПАРИ АУДИТОРІЇ», представленого на рис.7.

Початкове відношення r = «ВИКЛАДАЧІ ПАРИ АУДИТОРІЇ». Відношення, на які без втрат поділяється початкове відношення r : R_1 = «ВИКЛАДАЧ_ПАРА», R_2 = «АУДИТОРІЇ ПАРИ».

Декомпозиція відношень:

R_1 = «ВИКЛАДАЧ_ПАРА» $\rightarrow R_{11}$ = «ВИКЛАДАЧІ»; R_{12} = «ПАРИ»;

R_2 = «АУДИТОРІЇ ПАРИ» $\rightarrow R_{21}$ = «АУДИТОРІЇ»;

r' = «ВИКЛАДАЧІ ПАРИ АУДИТОРІЇ»

Нижнє підкреслення означає ключ відношення.

На рис.8. наведені відношення «ВИКЛАДАЧ_ПАРА», «АУДИТОРІЇ ПАРИ» після декомпозиції відношення «ВИКЛАДАЧІ ПАРИ АУДИТОРІЇ».

ВИКЛАДАЧ_ПАРА		АУДИТОРІЇ ПАРИ	
<u>Викладач</u>	<u>Пара</u>	<u>Пара</u>	<u>Аудиторія</u>
Іванов	перша	перша	101
Іванов	друга	перша	102
Петров	перша	друга	101
Петров	друга	друга	102

Рис.8. Відношення «ВИКЛАДАЧІ ПАРИ АУДИТОРІЇ» після відповідної декомпозиції

Розглянутий під час дослідження формалізований підхід підтримки цілісності даних реляційних моделей даних у разі багатозначних ФЗ на етапах експлуатації та реінжинірингу дозволить підвищити надійність та ефективність ІС для вирішення задач інтелектуального аналізу, заснованих на технології БД [12], [13].

7. Обговорення результатів дослідження

За підсумками аналізу особливостей проєктування реляційної моделі даних розглянуто основні проблеми підтримки цілісності у разі багатозначних ФЗ на етапах експлуатації та реінжинірингу БД. Досліджено багатозначні ФЗ атрибутів та запропоновано метод підтримки цілісності засобами реляційної моделі. Наведено низку прикладів, що пояснюють загальну проблему підтримки цілісності реляційних моделей даних, а також технологію специфічних модельних обмежень та метод вирішення поставленого завдання.

Запропоновано підхід до виявлення раніше невідомих ФЗ, що ґрунтується на аналізі множини даних РБД. Першим кроком є отримання множини ФЗ для кожного відношення. На другому кроці проводиться аналогічна операція для універсального відношення аналізованої РБД. На цьому кроці стає можливим виявити ФЗ між атрибутами різних відношень – взаємозв'язки між даними, що встановилися у процесі функціонування РБД. Запропоновано спосіб визначення їхньої інформаційної новизни, який полягає у перевірці членства ФЗ універсального відношення у замиканні об'єднання множин ФЗ окремих відношень.

Напрямок для подальших досліджень є розробка методів і способів здійснення перевірки отриманих залежностей на предмет коректності для предметної області, яка моделюється, аналізованої РБД.

8. Висновки

Під час дослідження розглядалася задача виявлення інформації про взаємозв'язки між даними, які могли встановитися у процесі функціонування РБД. Взаємозв'язки представляються як залежності різних типів, які потім можна використовувати як вихідні дані для методів повторного проєктування за умов підтримки цілісності реляційної моделі даних.

Одним з основних напрямків розвитку в галузі БД на даний момент є реінжиніринг, що дозволяє проводити перепроєктування існуючих БД, використовуючи максимум корисної інформації, яку можна отримати в результаті аналізу вихідної структури даних. Такий підхід дозволяє суттєво зменшити витрати коштів та часу на проведення перепроєктування.

Отримані результати дозволяють в подальшому вирішити задачу розробки інформаційних систем, орієнтованих на ефективну технологію реінжинірингу РБД з підтримкою обмежень цілісності даних у застосуваннях користувача.

Перелік посилань:

1. Codd E. F. A relational model of data for large shared data banks. Communications of the ACM. 1983. Vol. 26, no. 1. P. 64–69. doi: 10.1145/357980.358007.
2. Wei Q., Chen G. Efficient discovery of functional dependencies with degrees of satisfaction. International Journal of Intelligent Systems, 2004. pp. 1089–1110.
3. Fan W., Geerts F., Li J., Xiong M. Discovering Conditional Functional Dependencies. IEEE Transactions on Knowledge and Data Engineering, 2010.
4. Bernstein P.A., Swenson J.R., Tsichritzis D.C A unified approach to functional dependencies and relations. Proc. ACM 1975 SIGMOD Conf., San Jose, Calif. p. 237-245.
5. Bernstein P.A. Synthesizing Third Normal Form Relations from Functional Dependencies. ACM Transactions on Database Systems (TODS). Vol. 1, Iss. 4. 1976. pp. 277 – 298.
6. Martin J. Computer Database Organization, 2nd Ed. Prentice Hall PTR, 1977. 713 p.
7. Wyss C., Giannella C., Robertson E. FastFDs: A Heuristic-Driven, Depth-First Algorithm for Mining Functional Dependencies from Relation Instances. Data Warehousing and Knowledge Discovery. 2001. pp. 101-110.
8. Huhtala Y. Tane: An Efficient Algorithm For Discovering Functional and Approximate Dependencies. The

Computer Journal 42 (2), 1999. pp. 100-111.

9. Maier D. The theory of relational databases. London: Pitman, 1983. 637 p.

10. Filatov, V., & Doskalenko, S. On the Approach to Searching for Functional Dependences of Data in Relational Systems. Innovative Technologies and Scientific Solutions for Industries, 2018. 1 (3), 54–58. doi:10.30837/2522-9818.2018.3.054.

11. Date C. J. Introduction to database systems. Pearson Education, Limited, 2003. 1024 p.

12. Avrunin, O., Vlasov, O., & Filatov, V. Model of semantic integration of information systems properties in relay database reengineering problems. Innovative Technologies and Scientific Solutions for Industries, 2020. 4 (14), pp. 5–12. doi:10.30837/itssi.2020.14.005

13. Zhou X., Chai C., Li G., Sun J. Database Meets Artificial Intelligence: A Survey. IEEE Transactions on Knowledge and Data Engineering, 2022. Vol. 34, no. 3, pp. 1096-1116. doi: 10.1109/TKDE.2020.2994641.

Надійшла до редколегії 16.07.2025 р.

Золотухін Олег Вікторович – кандидат технічних наук, доцент, завідувач кафедри ШІ ХНУРЕ, м. Харків, Україна; e-mail: oleg.zolotukhin@nure.ua; ORCID: <https://orcid.org/0000-0002-0152-7600>.

Кудрявцева Марина Сергіївна – кандидат технічних наук, доцент, доцент кафедри ШІ ХНУРЕ, Харків, Україна; e-mail: maruna.kudryavtzeva@nure.ua; ORCID: <https://orcid.org/0000-0003-0524-5528>.

Філатов Валентин Олександрович – доктор технічних наук, професор, професор кафедри ШІ, ХНУРЕ, м. Харків, Україна; e-mail: valentin.filatov@nure.ua; ORCID: <https://orcid.org/0000-0002-3718-2077>.