

Міністерство освіти та науки України
Національна академія наук України
Люблінський відділ Польської Академії Наук
Представництво „Польська академія наук” у Києві
Харківський національний університет радіоелектроніки
Харківський національний університет міського господарства
ім. А.М. Бекетова
AGH науково-технологічний університет в Кракові
Миколаївський кораблебудівний університет ім. адмірала Макарова
Одеський національний політехнічний університет
Прикарпатський національний університет ім. В. Стефаника
Українська нафтогазова академія
Українська Федерація Інформатики
Академія Наук Прикладної Радіоелектроніки
Білоруський державний університет інформатики та радіоелектроніки
Білоруський національний технічний університет
Національний університет цивільного захисту України
Запорізький національний технічний університет

«Інформаційні системи та технології» ІСТ-2018

МАТЕРІАЛИ

**7-ї Міжнародної науково-технічної конференції,
присвяченої 55-річчю кафедри Прикладної математики ХНУРЕ,
55-річчю кафедри Програмної інженерії ХНУРЕ
та 40-річчю кафедри Прикладної математики та інформаційних технологій
ХНУМГ імені О.М. Бекетова**

**10-15 вересня 2018
Коблеве-Харків, Україна**

«INFORMATION SYSTEMS AND TECHNOLOGIES» IST-2018

**Proceedings
of the 7-th International Scientific and Technical Conference**

**September 10-15, 2018
Kobleve-Kharkiv, Ukraine**

Харків 2018

Секція 1

Сучасні інформаційні системи та технології: проблеми, методи, моделі.
Управління проектами та програмами.

Section 1

Modern information systems and technologies:
problems, methods, models.
Projects and program managements.

Метод Синтезу Моделі Даних у Задачах Реінжинірингу Реляційних Систем

Валентин Філатов
завідувач кафедри штучного інтелекту
Харківський національний університет
радіоелектроніки
Харків, Україна
valentin.filatov@nure.ua

Валерій Семенець
ректор
Харківський національний університет
радіоелектроніки
Харків, Україна
valery.semenets@nure.ua

Method of Data Model Synthesis for Relational Systems Reengineering

Valentin Filatov
Head of Artificial Intelligence Department
Kharkiv National University
of Radio Electronics
Kharkiv, Ukraine
valentin.filatov@nure.ua

Valerii Semenets
rector
Kharkiv National University
of Radio Electronics
Kharkiv, Ukraine
valery.semenets@nure.ua

Анотація—Запропоновано підхід до пошуку функціональних залежностей, заснований на аналізі множини даних реляційної бази даних. Виділено класи завдань реінжинірингу реляційних баз даних; досліджено етап формування цільової логічної схеми, яка є спільною для задач адаптації та рефакторінга. Розглянуто задачу перевірки відповідності логічної схеми реляційної бази даних третій нормальній формі в рамках даного етапу за допомогою методу синтезу. Показано, що її рішення пов'язане із рядом труднощів, зокрема, необхідністю знаходження множини функціональних залежностей, що виконуються на поточному екземплярі даних реляційної бази даних. Запропоновано підхід для знаходження множини функціональних залежностей примірника даних реляційної структури. Як спрямовуючу силу для подальших досліджень можна виділити реалізацію підтримки порожніх значень на етапі виявлення функціональних залежностей, а також питання перенесення даних без втрат із вихідної структури бази даних у цільову, отриману в результаті застосування методів реінжинірингу.

Abstract—An approach to the search for functional dependencies, based on the analysis of the data set of the relational database is proposed. The classes of tasks of reengineering of relational databases are allocated. The stage of the formation of the target logic scheme, which is common to adaptation and refactoring tasks, is investigated. The task of checking the correspondence of the logical scheme of the relational database to the third normal form within the framework of this stage using the method of synthesis is considered. It is shown that its solution is associated with a number of difficulties, in particular, the need to find the set of functional dependencies performed on the current instance of the relational database. An approach is proposed for finding the set

of functional dependencies of a copy of the relational structure data. As a driving force for further research, the implementation of support for hollow values at the stage of identifying functional dependencies, as well as the issue of transfer of lossless data from the original database structure to the target, obtained as a result of using reengineering methods, can be singled out.

Ключові слова—реінжиніринг, реляційна база даних, функціональна залежність, виявлення залежностей, універсальне відношення, замикання функціональних залежностей.

Keywords—reengineering, relational database, functional dependence, dependency detection, universal relation, closure of functional dependencies.

I. ВСТУП

Розвиток інформаційних технологій призводить до змін усіх етапів життєвого циклу інформаційних систем (ІС), зокрема, етапів підтримки й супроводу. Вимоги до обробки даних предметних областей, що постійно змінюються, негативним чином позначаються як на якості програмного продукту, так і подальших можливостях його розвитку та обслуговування. Нерідко виникають ситуації, коли подальша підтримка окремих компонентів або системи в цілому стає неможливою через значне ускладнення внутрішньої структури та завдань, що постійно змінюються, а проектування нової системи, здатної задовольнити поточні вимоги, є недоцільним із ряду причин. Одним із можливих варіантів вирішення такої проблеми є проведення реінжинірингу, метою якого є поліпшення характеристик вихідної ІС на основі попереднього аналізу поточного стану ІС та її окремих компонентів [1].



II. ОСНОВНІ ПРОБЛЕМИ ПРОЕКТУВАННЯ ІНФОРМАЦІЙНИХ СИСТЕМ

Розробка і реалізація багатьох сучасних великих інформаційних проектів має, як правило, затяжний характер, їх ціна перевищує заплановану, а остаточний продукт виходить ненадійним та складним у супроводі. Це призводить до ситуації, відомої як «криза програмного забезпечення». Перші згадки про кризу були зроблені ще наприкінці 80-х років, але навіть через 30 років її все ще не вдалося подолати. Основні невдачі при створенні програмного забезпечення викликані відсутністю повної специфікації вимог на етапі проектування, прийнятної методології розробки або недостатнім ступенем поділу спільного глобального проекту на окремі компоненти, які піддаються ефективному контролю та управлінню.

У разі часткової реалізації вимог користувачів інформаційної системи або зміні бізнес-процесу таким чином, що система перестає відповідати потребам користувачів, можливі кілька варіантів розвитку: розробка нової системи; модифікація (розвиток) існуючої системи; реінжиніринг існуючої (успадкованої - legacy system) системи [2].

Перший варіант найбільш простий і кращий для розробника, але він найменше задовольняє вимогам користувачів, бо будуть потрібні витрати додаткового часу і фінансових ресурсів, а також існують ризики втрати накопиченої інформації за час існування експлуатованої ІС. Реінжиніринг успадкованих інформаційних систем вимагає залучення експертів в області інформаційних систем і технологій, що відповідно призводить до ускладнення, подорожчання таких робіт. У більшості випадків існує така думка: простіше розробити систему заново, ніж вдатися до її реінжинірингу. Це пов'язано із кваліфікацією фахівців, яких необхідно залучити для проведення робіт. Вони повинні бути досить високого рівня для вирішення комплексу задач проектування і створення модифікованої інформаційної системи [3]. База даних, зазвичай, є найбільш важливою складовою ІС. На сьогоднішній день реляційні бази даних (РБД) займають домінуюче місце і використовуються у переважній більшості існуючих програмних додатків і сервісів. Цим обумовлений вибір РБД як об'єкту дослідження.

Було проведено аналіз існуючих публікацій із дослідженнями даної тематики. Основними напрямками розвитку є: реінжиніринг логічної схеми за допомогою проміжного представлення, такого, як ER-модель або власні метамоделі, і застосування набору визначених правил для трансляції об'єктів моделі в конструкції РБД, реінжиніринг застарілих БД; видобування структури як застарілих, так і реляційних БД і представлення їх у вигляді концептуальної моделі даних, зокрема, ER-моделі.

У даній роботі розглядається задача виявлення інформації про взаємозв'язки між даними, які могли встановитися в процесі функціонування БД. З цією метою розробляються методики відновлення структури даних, обумовленою їх взаємозв'язками, для подальшого аналізу і перенесення даних на сучасну платформу, як правило, реляційну [4].

III. РОЗРОБКА МЕТОДУ ПОШУКУ ФУНКЦІОНАЛЬНО-ЗАЛЕЖНИХ АТРИБУТІВ РЕЛЯЦІЙНОЇ БАЗИ ДАНИХ

Вихідними даними для вирішення поставленої задачі є: логічна схема реляційної БД

$$\Sigma = \{\sigma_i, i = \overline{1, n}\},$$

де σ_i – схема відношення $\sigma_i = \langle R_i, F_i \rangle$, що надходить до БД, R_i – носій відношення (множина атрибутів), а F_i – множина функціональних залежностей (ФЗ), які задовольняють даному відношенню, n – кількість відношень. $P = \{\rho_i, i = \overline{1, n}\}$ – множина відношень розглянутої БД.

Існування функціональної залежності вигляду $A \rightarrow B$ означає [5], що для будь-яких двох кортежів u, v деякого відношення ρ_k справедливий наслідок:

$$u(A) = v(A) \Rightarrow u(B) = v(B).$$

У якості прикладу розглянемо логічну схему $\Sigma = \{\sigma_1, \sigma_2\}$, що складається з двох схем відношень:

$$\sigma_1 = \langle R_1 = \{A, B, C\}, F_1 \rangle, \sigma_2 = \langle R_2 = \{C, D\}, F_2 \rangle.$$

Вдамо, що інформація про F_1, F_2 відсутня або загублена. Отримати множину ФЗ, що задовольняють даному відношенню, можливо за допомогою методу виявлення ФЗ із екземплярів даних відношень, зокрема, методу Тане, принцип и реалізація якого детально викладені у [6].

У результаті його застосування буде отримана множина мінімальних ФЗ, що задовольняють набору даних у відношенні на момент проведення обробки.

Мінімальна ФЗ – це залежність:

$$X \rightarrow Y, X = \{A_1, \dots, A_n\}, Y = \{B_1, \dots, B_m\},$$

у якій не існує така множина $Z \subset X$, для якої б виконувалось $Z \rightarrow Y$. Тривіальні ФЗ вигляду $A_1 \rightarrow A_1$ ігноруються даним методом, оскільки не є суттєвими.

Розглянемо відношення ρ_1, ρ_2 , що наведені у двох наступних таблицях.

ТАБЛИЦЯ I. Відношення ρ_1

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

ТАБЛИЦЯ II. Відношення ρ_2

C	D
1	1
2	1



Застосуємо метод Tane та отримаємо множину ФЗ для наведених відношень:

$$F_1 = \{AC \rightarrow B\}, F_2 = \{C \rightarrow D\}.$$

Приймаючи до уваги множину ФЗ для F_1 и F_2 , множина ФЗ схеми Σ буде мати вигляд:

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

Носій універсального відношення $R = \{R_1 \cup \dots \cup R_n\}$ у цьому випадку виглядає наступним чином: $R = R_1 \cup R_2 = \{A, B, C\} \cup \{C, D\} = \{A, B, C, D\}$.

Універсальне відношення може бути отримано через природне з'єднання всіх відношень, що входять у схему. Результат такого з'єднання наведено у наступній таблиці.

ТАБЛИЦЯ III. УНІВЕРСАЛЬНЕ ВІДНОШЕННЯ

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

Після застосування методу Tane для отриманого універсального відношення, будуть виявлені 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 невідомі ФЗ $F'_\Sigma = \{A \rightarrow D, B \rightarrow D\}$. Отже, множину ФЗ універсального відношення для логічної схеми Σ можна виразити як $\overline{F_\Sigma} = F_\Sigma \cup F'_\Sigma$, де F_Σ – множина ФЗ, що задовольняє вихідним відношенням Σ , F'_Σ – множина додаткових ФЗ.

Далі необхідно встановити, чи дійсно ФЗ із множини F'_Σ виводяться з F_Σ , або ж вони є новою інформацією. Для цього пропонуємо застосувати метод перевірки належності ФЗ до замикання $(F_\Sigma)^+$ – вирішення проблеми членства. Принцип полягає у наступному: у зв'язку з тим, що побудова F^+ пов'язана із перебором усіх підмножин множини атрибутів, які належать F , і мають експоненціальну складність, пропонується будувати F -замикання на множині атрибутів.

F -замиканням множини X назвемо таку множину атрибутів X^+ , що $X \rightarrow X^+ \in F^+$ і не існує жодного атрибуту у R , який б залежав від X і 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^+ і визначенні істинності вислову $Y \subseteq X^+$. У разі, коли вислів істинний, то $X \rightarrow Y \in F^+$.

Щоб перевірити $A \rightarrow D$ на належність до F_Σ^+ , потрібно побудувати A^+ . Відповідно до $A^+ = \{A\}$, F_Σ не містить ФЗ, для яких A була би єдиним атрибутом у лівій частині ФЗ. $D \notin A^+$, відтак, $A \rightarrow D \notin F_\Sigma^+$. Аналогічним чином доводиться, що $B \rightarrow D \notin F_\Sigma^+$. Отже, множина F'_Σ виявлених залежностей є не виведеною і являє собою нову інформацію.

Даний підхід не гарантує повної відповідності нових залежностей розглянутої предметної області. У зв'язку з тим, що він базується на множині даних, яка міститься у РБД на момент проведення обробки і не враховує їх семантику, існує велика ймовірність отримання випадкових ФЗ. Випадкова ФЗ – це така ФЗ, що не є коректною для конкретної предметної області і може бути усунена у будь-коли після зміни або додавання кортежів із даними, що суперечать виявленій залежності, в процесі функціонування РБД.

IV. ВИСНОВКИ

Запропоновано підхід до виявлення раніше невідомих функціональних залежностей, що ґрунтується на аналізі множини даних реляційної БД. Першим кроком пропонується отримання множини ФЗ для кожного відношення. На другому кроці проводиться аналогічна операція для універсального відношення даної РБД. Стає можливим виявити ФЗ між атрибутами різних відношень – взаємозв'язку між даними, що встановилися у процесі функціонування РБД. Запропоновано спосіб визначення їх інформаційної новизни, який полягає у перевірці членства ФЗ універсального відношення у замиканні об'єднання множин ФЗ окремих відносин. Напрямок для подальших досліджень є розробка методів для здійснення перевірки отриманих залежностей на предмет коректності для предметної області РБД, що підлягає реінжинірингу.

ЛІТЕРАТУРА REFERENCES

- [1] В.И. Есин, “Реинжиниринг существующих баз данных” Системы обробки інформації, ХНУ ім. В.Н. Каразіна, Харків 2012, вип. 3 (101), том 2, с. 188-191.
- [1] С.М. Константинов, Ю.Л. Пономаренко, В.О. Філатов, “Часткове відображення моделей даних при інтеграції інформаційних систем” Економіко-математичне моделювання соціально-економічних систем: Зб. наук. пр., вип. 21, Київ, 2016, с. 140-158.
- [2] V. Kosenko, “Principles and structure of the methodology of risk-adaptive management of parameters of information and telecommunication networks of critical application systems” in Innovative technologies and scientific solutions for industries. Kharkiv. 2017. No. 1 (1). pp. 46-52.
- [3] V. Filatov, “Fuzzy models presentation and realization by means of relational systems” in Econtechmod: an international quarterly journal on economics in technology, new technologies and modelling processes. Lublin ; Rzeszow, 2014, Vol.3, № 3, pp. 99-102.
- [4] V. Filatov, V. Radchenko, “Reengineering relational database on analysis functional dependent attribute” Proceedings of the X Intern. Scient. and Techn. Conf. “Computer Science & Information Technologies” (CSIT'2015), 14-17 sept. 2015, Lviv, Ukraine, pp. 85-88.
- [5] V. Filatov, S. Doskalenko, “The Approach to Searching for Functional Dependences of Data in Relational Systems” in Innovative technologies and scientific solutions for industries, Kharkiv, 2018, No. 3 (1), pp. 54-58.

