

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

ЧЕРНЕНКО МИКОЛА ВОЛОДИМИРОВИЧ

УДК 004.658.3

**МЕТОДИ ТА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ІНТЕГРАЦІЇ
БАЗ ДАНИХ КОРПОРАТИВНИХ СИСТЕМ**

05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеню
кандидата технічних наук

Харків – 2013

Дисертацією є рукопис.

Робота виконана в Харківському національному університеті радіоелектроніки Міністерства освіти і науки України.

Науковий керівник – доктор технічних наук, професор
Левикін Віктор Макарович,
Харківський національний університет
радіоелектроніки,
завідувач кафедри інформаційних
управляючих систем.

Офіційні опоненти: доктор технічних наук, професор
Асєєв Георгій Георгійович
Харківська державна академія культури,
завідувач кафедри інформаційних
технологій;

доктор технічних наук, професор
Федорович Олег Євгенович,
Національний аерокосмічний університет
ім. М.Є. Жуковського „ХАІ”,
завідувач кафедри інформаційних
управляючих систем.

Захист відбудеться “_____” червня 2013 р. о _____ годині на засіданні спеціалізованої вченої ради Д 64.052.08 у Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

З дисертацією можна ознайомитись у бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

Автореферат розісланий “_____” травня 2013 р.

Учений секретар
спеціалізованої вченої ради

І.П. Плісс

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Активний розвиток інформаційних технологій і відсутність єдиної централізованої політики автоматизації діяльності організації часто призводять до того, що на одному об'єкті автоматизації можуть функціонувати кілька інформаційних систем. Внаслідок цього успішність реалізації великої долі проєктів інформаційних систем не в повній мірі відповідають потребам користувачів.

Під час функціонування на одному об'єкті управління декількох інформаційних систем (ІС) відбувається накопичення інформації в їх базах даних (БД). З часом при збільшенні обсягу даних у базі даних, ефективність використання ряду інформаційних систем на одному об'єкті управління знижується. Це може бути пов'язано з формуванням некоректних керуючих впливів на підставі неактуальної, а часто, і суперечливої інформації, яка виявляється на перетині предметних областей успадкованих інформаційних систем. Даний факт спричинений тим, що цілісність даних підтримується на рівні системи управління базами даних (СУБД). Складним завданням є підтримка цілісності даних на «стику» систем, особливо по мірі зростання обсягів накопичених даних.

Для вирішення проблеми лоскутної автоматизації організації виникає потреба в інтеграції баз даних інформаційних систем. Особливе місце відводиться базам даних тому, що інформаційні системи спрямовані на задоволення потреб користувачів в інформації, а безпосереднім інструментом збереження і роботи з даними є бази даних та системи управління ними. Аналіз статистики застосування різних моделей даних показує, що в переважній більшості випадків використовуються реляційна модель даних або її розширення - об'єктно-реляційна модель даних.

Завданням інтеграції приділено багато уваги в наукових працях. Починаючи з 70-х років, велися роботи з структурної інтеграції різних моделей даних. Частина праць присвячена питанням інтеграції даних, деякі – питанням інтеграції ІС. Існують також роботи, присвячені інтеграції баз даних, в яких розглядається як інтеграція структурної складової, так і інтеграція даних. Серед закордонних та вітчизняних науковців в області формалізації, проєктування та реорганізації БД можна навести слідуючих: В. Армстронга, К. Делобеля, Е. Кодда, К. Бірі, Е. Тененбаума, Дж. Фрідла, В. Пасічника, Ф. Бернштейна, Д. Мейєра, Т. Лінга, Л. Калиніченко, С. Кузнєцова, К. Чері, А. Карденаса, Дж. Ульмана, Е. Дедікова, М. Бусліка, В. Філатова, С.Таняньського та ін.

Незалежно від пропонованих підходів до інтеграції, всі автори сходяться на думці, що це високовартісний процес, що вимагає участі висококваліфікованих фахівців. Однак у ході аналізу даної проблемної області не було виявлено підходів, інформаційних технологій або рішень, які б повно і детально описували заходи, що пов'язані з переходом від використання ряду локальних БД до інтегрованої БД організації. Також не було виявлено моделей і методів оцінки можливості інтеграції баз даних, які враховували б не тільки моделі або структури даних, але й функціональну забезпеченість СУБД. Це б дозволило на етапах, що передують безпосередньо процесу інтеграції, оцінити доцільність проведення робіт з її реалізації. Разом з тим існує ризик досягнення некоректного результату. З вищевказаного випливає, що задача розробки інформаційної технології, що базується на методах інтеграції не лише

реляційних, але й плоских БД, а також враховує функціональні можливості СУБД за рахунок використання критерію оцінки доцільності інтеграції БД, є актуальною і дозволить знизити можливі ризики в ході інтеграції баз даних успадкованих інформаційних систем.

Зв'язок роботи з науковими програмами, планами, темами. Дослідження, представлені в даній роботі, проводилися автором на кафедрі інформаційних управляючих систем Харківського національного університету радіоелектроніки (ХНУРЕ) в рамках держбюджетних тем: №243 «Методи, моделі та інформаційні технології розбудови соціально-економічної освітньо-наукової мережі з метою інтеграції у Європейський простір» (ДР 0109U002497); №265 «Методи та моделі самоорганізації інфраструктури інтелектуального інформаційного середовища, що базується на використанні принципів хмарних обчислень» (ДР 0112U000207); розділу „Дослідження і розробка моделей інтелектуального аналізу в системах реляційних баз даних” НДР №245 «Еволюційні гібридні системи обчислювального інтелекту зі змінною структурою для інтелектуального аналізу даних» (ДР 0110U000458), де автор був виконавцем, а також науково-технічних послуг за договором № 09-26 «Дослідження та адміністрування апаратного та програмного забезпечення, що входять до комп'ютерної мережі» на замовлення Харківської обласної ради - автор був виконавцем.

Мета і завдання дослідження. Метою даної роботи є розробка методів, інформаційної технології інтеграції реляційних, об'єктно-реляційних, плоских БД і критерію оцінки доцільності застосування запропонованих методів, що дозволять підвищити ефективність процесів інтеграції БД в ході реалізації реінжинірингу інформаційних систем.

Поставлена мета передбачає вирішення таких завдань:

- проведення аналізу реляційної моделі даних, дослідження математичного апарату, що забезпечує високу затребуваність даної моделі, проведення аналізу процесу проектування БД, методів та інструментальних засобів їх інтеграції;
- удосконалення методу побудови узгодженого універсального відношення бази даних за рахунок узагальнення операцій реляційної алгебри: природного з'єднання за відповідними проекціями та зовнішнього природного з'єднання за відповідними проекціями;
- розробка методу синтезу інтегрованої реляційної БД з множини реляційних БД;
- розробка методу перетворення плоских БД до виду реляційних відносин;
- розробка критерію оцінки доцільності інтеграції БД на основі функціональної забезпеченості СУБД;
- розробка інформаційної технології інтеграції БД, формування її формального опису;
- розробка інструментального засобу інтеграції БД, що дозволяє автоматизованим способом реалізувати етапи інформаційної технології.

Об'єкт дослідження: процеси проектування та реінжинірингу інформаційних систем.

Предмет дослідження: методи та інформаційні технології інтеграції баз даних корпоративних систем.

Методи дослідження: для аналізу проблем функціонування успадкованих ІС і пошуку їх вирішення було використано системний аналіз; для опису можливих варіантів перетину предметних областей ІС і введених операцій реляційної моделі даних використана теорія множин; для опису запропонованих алгоритмів використана теорія графів; в ході розробки методів інтеграції БД використовувалася теорія баз даних, технології аналізу і проектування БД.

Наукова новизна отриманих результатів. У ході вирішення поставленого завдання проведені дослідження, які дозволили отримати такі наукові результати:

вперше:

- запропоновано метод синтезу єдиної схеми інтегрованої бази даних, яка описує сумісне представлення розподіленої інформації з урахуванням структурних властивостей даних, що надало можливість реалізації інтеграції різнорідних баз даних в єдину реляційну базу даних із збереженням накопиченої інформації;

- запропоновано інтегральний критерій оцінки доцільності інтеграції баз даних на основі функціональної забезпеченості систем управління базами даних, який базується на використанні показників нормованих трудових витрат, що дозволяє зменшити кількість операцій в процесі побудови схеми інтегрованої бази даних;

дістав подальший розвиток:

- метод побудови узгодженого універсального відношення бази даних за рахунок узагальнення операцій реляційної алгебри: природного з'єднання за відповідними проекціями та зовнішнього природного з'єднання за відповідними проекціями, які, на відміну від існуючих операцій, дозволяють виконувати з'єднання по атрибутам із різними схемами їх іменування з виключенням надлишкових даних, що дозволяє зменшити часові витрати на формування універсального відношення бази даних за рахунок особливостей зазначених операцій;

удосконалено:

- метод перетворення плоских баз даних до виду реляційного відношення, який, на відміну від існуючих, дозволяє не тільки завантажити дані, але і отримати в результаті нормалізоване універсальне відношення за рахунок застосування методів перетворення даних в ході їх завантаження.

Практичне значення отриманих результатів. Запропоновано інформаційну технологію інтеграції БД з використанням розроблених моделей та методів, а також представлено інструментальний засіб інтеграції БД, що реалізовує дану інформаційну технологію і забезпечує інтеграцію реляційних, об'єктно-реляційних і плоских БД (БД середніх і великих OLTP-систем). В ході інтеграції БД знижуються трудові витрати і ймовірність отримання некоректних результатів (ключові операції виконуються автоматизовано). Даний інструментальний засіб дозволяє зберігати історію процесу інтеграції БД, мінімально збільшуючи необхідний простір для її зберігання за рахунок використання представлень (VIEW).

Результати дисертаційної роботи впроваджені у таких установах та організаціях: Харківському національному медичному університеті (акт від 23.05.2011), Прокуратурі Харківської області (акт від 24.01.2011), Головному управлінні МНС в Харківській області (акт від 14.10.2010). Запропоновані моделі, методи, інформаційна технологія та інструментальний засіб, що реалізує запропоновану інформаційну технологію, були застосовані в ході виконання проектних робіт в Центрі інформа-

ційних систем і технологій ХНУРЕ.

Основні результати дисертаційної роботи використовуються у навчальному процесі ХНУРЕ за напрямом 6.050101 «Комп'ютерні науки» за спеціальністю 7.080401 «Інформаційні управляючі системи і технології» у дисциплінах «Проектування інформаційних систем», «Організація баз даних та знань» і «Технології баз даних».

Особистий внесок здобувача. Усі наукові результати, викладені в дисертації, отримані автором особисто. У роботах, які написано у співавторстві, здобувачу належать: метод синтезу інтегрованої реляційної бази даних з кількох реляційних баз даних [1]; критерій оцінки доцільності інтеграції баз даних на основі функціональної забезпеченості СУБД [2]; модель переходу від етапу макропроектування до етапу мікропроектування, як результат проведеного аналізу актуальної задачі формалізації взаємодії замовника та розробника при виконанні робіт по реалізації проекту інформаційної системи або реінжинірингу успадкованих інформаційних систем [4]; удосконалений метод перетворення плоских БД до виду реляційного відношення [5]; критерій оцінки доцільності інтеграції баз даних на основі функціональної забезпеченості СУБД [6]; підхід до управління інформаційними ресурсами з використанням функціональних особливостей СУБД Oracle, аналіз можливості управління інформаційними ресурсами [7]; архітектура системи управління інформаційними ресурсами [8]; технологія формування вихідних документів з використанням окремих функціональних особливостей СУБД [9].

Апробація результатів дисертації. Основні положення дисертаційної роботи доповідалися на таких конференціях і форумах: XI Міжнародний молодіжний форум «Радиоэлектроника и молодежь в XXI веке» (Харків, 2007); Науково-практична конференція «Інформатизація бізнесу очима молодих: прогресивні технології, наука, підприємництво» (Харків, 2007); XII Міжнародний молодіжний форум «Радиоэлектроника и молодежь в XXI веке» (Харків, 2008); Міжнародна науково-практична конференція «Перспективні інформаційні технології в економіці оброблення інформації, моделювання, видавництво» (Харків, 2008); Первая факультетская научно-практическая молодежная школа-семинар «Информационные интеллектуальные системы» (Харків, 2009); VI Міжнародна науково-практична конференція «Наука и социальные проблемы общества: информатизация и информационные технологии» (Харків, 2011).

Публікації. За результатами досліджень опубліковано 11 наукових праць, а саме: 5 статей, з яких 5 у наукових фахових виданнях України з технічних наук, 6 публікацій у працях наукових конференцій та форумів.

Структура дисертації. Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації становить 202 с., що включає 50 рисунків на 14 сторінках, 1 таблицю, що займає окрему сторінку; 9 додатків на 27 сторінках; список використаних джерел з 125 найменувань на 13 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі наведено загальну характеристику роботи: обґрунтовано актуальність

теми дисертаційної роботи, сформульовано мету та задачі дослідження, визначено об'єкт, предмет і методи дослідження, наукову новизну і практичне значення отриманих результатів, наведено дані про публікації, особистий внесок автора в роботи, виконані у співавторстві, та відомості про апробацію результатів дисертації.

У першому розділі проведено аналіз сучасного стану проблеми функціонування декількох різнорідних інформаційних систем (ІС) на одному об'єкті управління. Для вирішення проблем різнорідної автоматизації обрано проведення реінжинірингу успадкованих ІС. В ході цього процесу особливе значення відводиться інформаційному забезпеченню, тому що основна мета використання інформаційних систем – задоволення потреб користувачів в інформації. Саме тому основна увага приділяється методам, підходам та інструментальним засобам проектування баз даних (БД), які можуть бути використані в ході проектування інтегрованої БД рівня організації.

Результати проведеного аналізу дозволяють зробити такі висновки:

- на практиці широко поширена ситуація різнорідної автоматизації в рамках однієї організації в силу різних причин, що призводить до зниження ефективності формування управлінських рішень і, як наслідок, зниження ефективності функціонування організації в цілому;
- реінжиніринг успадкованих ІС – один з ефективних засобів усунення недоліків різнорідної автоматизації в ході розробки інформаційної системи рівня підприємства;
- узгодження та збереження накопиченої інформації успадкованих інформаційних систем в ході реінжинірингу з максимальною підтримкою використовуваних СУБД є одним з ключових факторів успіху реінжинірингу;
- інтеграція БД може проходити із застосуванням одного з існуючих підходів інтеграції даних (консолідація, федералізація, поширення даних);
- аналіз наявності аналогів інструментальних засобів інтеграції показав відсутність таких засобів, які б повною мірою відповідали потребам розв'язуваної задачі.

Наведені висновки відображають перспективність розробки інформаційної технології інтеграції БД, сформульовано мету і задачі дослідження.

У другому розділі запропоновано метод синтезу інтегрованої реляційної БД з множини окремих реляційних БД, удосконалено метод побудови узгодженого універсального відношення бази даних за рахунок узагальнення операцій реляційної алгебри: з'єднання відносин за відповідними проекціями та зовнішнього з'єднання відносин за відповідними проекціями, що використовуються в запропонованому методі.

Оскільки реляційні та об'єктно-реляційні БД є найбільш затребуваними в ході реалізації проектів інформаційних систем (більш 90%), тому основна увага приділена реляційній моделі даних, яка є фундаментальною математичною моделлю в теорії БД на поточному етапі розвитку інформаційних технологій. При описанні складової компоненти представленої моделі даних, особливу увагу приділено операційній специфікації. У даному розділі досліджено недоліки функціонування декількох ІС на одному об'єкті управління. Запропоновані можливі варіанти їх функціонування:

- а) предметні області частково перетинаються (рис.1, а);
- б) предметні області не перетинаються (рис.1, б);
- в) предметні області перетинаються (рис1, в).

Для здійснення інтеграції реляційних БД успадкованих ІС різного ступеню нор-

малізації із збереженням накопичених в них даних в другому розділі розроблено метод синтезу інтегрованої реляційної БД з множини окремих реляційних БД.

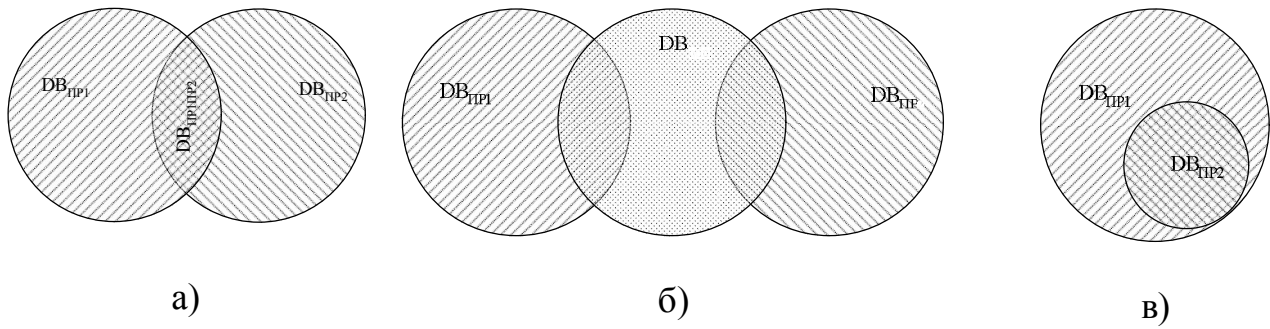


Рисунок 1 – Можливі варіанти функціонування декількох інформаційних систем на одному об'єкті управління

Однак в ході дослідження реляційної моделі даних було виявлено, що існуючі операції не повною мірою дозволяють реалізувати поставлені завдання, і обмежують реалізацію заявленого методу.

Таким чином, постала задача опису операцій, які б усунули недоліки існуючих операцій: виконання умови повного збігу імен атрибутів для операції природного з'єднання і усунення дублювання атрибутів в результаті застосування операції еквіз'єднання, а також запровадження додаткового обмеження – формування додаткової структури (відповідної проекції) згідно якій необхідно проводити з'єднання реляційних відносин. В якості рішення запропонована операція природного з'єднання за відповідними проекціями, модель якої виглядає таким чином:

$$\begin{aligned} &\pi_X(U_k) \leftrightarrow \pi_X(U_p) \\ &U_k \bowtie U_p = \pi_{A_i \cup B_j}(\sigma_{\pi_X(U_k) = \pi_X(U_p)}(U_k \times U_p)), \\ &\pi_X(U_k) = \pi_X(U_p) \end{aligned} \quad (1)$$

де U_k – універсальне відношення БД k-ой успадкованої ІС;

U_p – універсальне відношення БД p-ої успадкованої ІС;

$\pi_X(U_k) \leftrightarrow \pi_X(U_p)$ – відповідні проекції U_k і U_p ;

$A_i \cup B_j$ – правило формування схеми результуючого відношення: схема утворюється шляхом об'єднання схем вихідних універсальних відношень (виключення з неї імен атрибутів, що повторюються);

A_i, B_j – імена атрибутів універсальних відношень U_k, U_p відповідно, що утворюють схему даного універсального відношення.

σ – операція селекції по рівності відповідних проекцій: $\pi_X(U_k) = \pi_X(U_p)$.

Проекція універсального відношення деякої БД успадкованої k-ой ІС ($\pi_X(U_k)$) відповідає проекції універсального відношення БД p-ої успадкованої ІС ($\pi_X(U_p)$) в тому випадку, якщо вони описують один і той же об'єкт реального світу X' , а проекція $\pi_X(U_k)$ утворена атрибутами, що дозволяють ідентифікувати даний

об'єкт поза БД, тобто не є штучно введеними ключами.

Засновником реляційної моделі даних Е.Коддом було введено розширення реляційної моделі даних для кращого відображення семантики, яке отримало свою реалізацію у всіх сучасних СУБД і регламентується стандартами. Тому в ході розробки методу виникла необхідність запропонувати аналог введеної операції природного з'єднання відношень за відповідними проекціями для роботи з невизначеними значеннями типу «властивість не визначено».

Модель розробленої операції зовнішнього природного з'єднання відносин за відповідними проекціями запропоновано представляти у вигляді:

$$U_1[\pi_1 \odot \pi_2] U_2 = \text{comm}[A, \pi_1, C] \cup (U'_1 \times (C: \omega)) \cup ((A: \omega) \times U'_2), \quad (2)$$

де π_1 – проекція універсального відношення U_1 відповідає π_2 проекції універсального відношення U_2 ($\pi_1 \leftrightarrow \pi_2$), π_1 і π_2 повинні бути задані на одному домені;

$\text{comm} = U_1[\pi_1 = \pi_2] U_2$ – множина кортежів, що задовольняють умові рівності відповідних проекцій двох універсальних відношень;

$U'_1 = U_1 - \text{comm}[A, \pi_1]$ – множина кортежів універсального відношення U_1 , які не входять до загальної множини comm , тобто дані, що зберігаються в одній БД про об'єкт реального світу, не зберігаються у другій БД;

$U'_2 = U_2 - \text{comm}[\pi_2, C]$ – множина кортежів універсального відношення U_2 , які не входять до загальної множини comm , тобто дані ситуація аналогічна U'_1 .

Дана операція описана для двох універсальних відношень двох БД. У разі необхідності застосування представленої операції для n баз даних передбачається використовувати її попарно $(n-1)$ раз.

Під завданням інтеграції баз даних декількох успадкованих ІС необхідно розуміти синтез структури єдиної БД для всіх БД успадкованих ІС з збереженням всіх актуальних функціональних залежностей і даних, накопичених в БД кожної окремої успадкованої ІС.

Для її вирішення розроблено метод синтезу інтегрованої реляційної БД з декількох реляційних баз даних у вигляді таких етапів:

Етап 1. Визначення ступеню узгодженості БД в цілому (операція природного з'єднання локальних відносин).

Етап 2. Визначення ступеню перетину предметних областей (ПрО) із використанням інфологічних моделей (ER-діаграми).

Етап 3. Аналіз результатів 2 етапи: якщо ПрО не перетинаються і немає можливості ввести локальні відносини, встановлюють зв'язок між БД успадкованих систем, то перехід до етапу 13 (формування результату «Інтеграція неможлива через загальні відмінності предметних областей»), інакше – перехід на етап 4.

Етап 4. Визначення мов, які застосовуються для:

– опису концептів на рівні БД (назви множин імен локальних відносин (R), атрибутів (A));

– представлення даних (D – множини доменів, $\text{dom}_n : A_n \rightarrow D_n$ – множини відображень атрибутів A_n в домени D_n).

Визначення єдиної мови для представлення описаних елементів структурної компоненти РМД в інтегрованій БД, якщо сутність збережених даних певного атрибута не відповідає назві, то його необхідно перейменувати. Множина доменів $D_1, D_2, \dots, D_n$ визначається відповідно до обраної загальної системи іменування ($D'_1, D'_2, \dots, D'_i, \dots, D'_n$). Потім визначаються множини атрибутів $A_1, A_2, \dots, A_i, \dots, A_n$, з урахуванням проведеного аналізу множин доменів. При необхідності слід виключити надлишкові атрибути в ході формування підсумкової множини $A_s = A_1 \cup A_2 \cup \dots \cup A_i \cup \dots \cup A_n \setminus A_{is}$.

Області припустимих значень атрибутів: $\text{dom}_1 : A_1 \rightarrow D_1, \text{dom}_2 : A_2 \rightarrow D_2, \dots, \text{dom}_i : A_i \rightarrow D_i, \dots, \text{dom}_n : A_n \rightarrow D_n$ визначаються з урахуванням сформованих множин D_s, A_s . В результаті вони утворюють область допустимих значень атрибутів структури цільової бази даних $\text{dom}_s = \text{dom}_1 \cup \text{dom}_2 \cup \dots \cup \text{dom}_i \cup \dots \cup \text{dom}_n$.

Етап 5. Приведення нормалізованих БД до вигляду універсальних відношень (операція природного з'єднання для узгоджених БД і операція зовнішнього природного з'єднання БД, що неузгоджені).

Етап 6. Перехід до 7 етапу якщо ПрО перетинаються (1, 3 варіант перетину), інакше якщо предметні області не перетинаються, але можна ввести додаткові концепти, що дозволяють зв'язати існуючі БД, тоді перехід на крок 12.

Етап 7. Визначення *відповідних проекцій* ($\pi_X(U_k) \leftrightarrow \pi_X(U_p)$), що дозволяють ідентифікувати запис на рівні універсальних відносин без застосування внутрішніх ідентифікаторів кожної окремої БД.

Етап 8. Визначення структури цільової БД, яка дозволяє зберігати накопичені дані успадкованих ІС із збереженням функціональних залежностей, але в той же час усуває надлишкові дані. Для цього необхідно застосовувати введену операцію *природного з'єднання за відповідними проекціями*. Після цього необхідно визначити множини актуальних доменів (D_a) структурної компоненти цільової БД: $D^a = D_1 \cup D_2 \cup \dots \cup D_i \cup \dots \cup D_n \setminus D^* \setminus D_{is}$ (D_{is} – множина сторонніх елементів множин $D_1 \dots D_n$, D^* – множина «застарілих» елементів множин $D_1 \dots D_n$). Результуюча множина доменів (D_s) складається з об'єднання множини фактичних доменів всіх БД і множини доменів, що визначають «історичну» складову (можуть застосовуватися при рефакторингу ІС) $D_s = D^a \cup D^*$.

Етап 9. Перехід до 10 етапу якщо операція природного з'єднання за відповідними проекціями надає набір даних, що дозволяє сформувати повне покриття функціональних залежностей на всьому наборі даних, інакше формування повідомлення про неможливість інтеграції за накопиченими даними і перехід до етапу 14.

Етап 10. Наповнення сформованої структури даними із застосуванням *операції зовнішнього природного з'єднання за відповідними проекціями*. Якщо ступінь перетину ПрО відповідає третьому варіанту, то необхідно привести отримане відношення до виду універсального відношення, тобто виключити всі записи, що дублюються.

Етап 11. Аналіз відповідності структури, наповненої даними тим функціональним залежностям, які були визначені на 7-8 кроках. Якщо є null значення, необхідно або розширити домен значенням «властивість невизначено», або оцінити не-

обхідність існування такого запису. Якщо невизначені значення присутні в ключових атрибутах, то необхідно їх доповнити певними значеннями або виключити кортежі з результуючого універсального відношення.

Етап 12. Нормалізація результуючого універсального відношення методом декомпозиції. Перехід на крок 15 для оформлення результату.

Етап 13. Введення додаткових локальних відношень для сполучення БД успадкованих ІС, предметні області яких не перетинаються, але в рамках організації можуть бути інтегровані в єдину ІС

Етап 14. Застосування методу синтезу БД за виявленими функціональними залежностями. Після формування структури за цим методом проводиться наповнення локальних відносин даними з БД успадкованих ІС з використанням операторів DML та DDL.

Етап 15. Формування результату у вигляді універсального відношення інтегрованої БД або у вигляді повідомлення про неможливість інтеграції із зазначенням причини.

У третьому розділі розглядається розширення області застосування розробленого методу синтезу інтегрованої реляційної БД з декількох реляційних БД за рахунок удосконалення методу перетворення плоских баз даних до виду реляційного відношення, що дозволяє проводити інтеграцію не лише реляційних, але й плоских БД. Розроблено критерій оцінки доцільності інтеграції баз даних на базі функціональної забезпеченості СУБД, який дозволяє оцінити та врахувати складність процесу інтеграції, що пов'язано з використанням функціональних особливостей СУБД, які використовуються в успадкованих інформаційних системах. Запропоновано інформаційну технологію інтеграції БД, що поєднує запропоновані та удосконалені методи, розроблені моделі операцій.

Розроблено дві моделі представлення плоских БД: модель представлення даних плоских БД (FDB) і модель плоских БД, представлених у вигляді таблиці (FDB_t).

У загальному вигляді модель представлення даних плоских БД, що зберігаються в окремому файлі, можна представити таким чином:

$$FDB = \begin{cases} \langle \langle H, del \rangle, cnr \rangle, i = 1 \\ \langle \langle D_i, del \rangle, cnr \rangle, i > 1 \end{cases} \quad (3)$$

де $\langle \langle H, del \rangle, cnr \rangle$ – запис заголовку;

$\langle \langle D_i, del \rangle, cnr \rangle$ – i -й запис даних;

H – множина даних, що формують заголовок;

D_i – множина даних для i -го запису;

del – символ-роздільник даних;

cnr – спеціальний символ переведення рядку.

Таку модель у вигляді таблиці представимо виразом:

$$FDB_t = \text{tag } T(\text{tag } TR_j(\text{tag } TD_{ji}(D'))), \quad (4)$$

де $\text{tag } T$ – парний тег початку і закінчення таблиці;

$\text{tag } TR_j$ – парний тег початку і закінчення j -го рядка таблиці;

$\text{tag } TD_{ji}$ – парний тег початку і закінчення i -ої комірки таблиці в j -ої рядку;

D' – елемент множини даних D (індекс елемента множини даних D відповідає індексу клітинки в рядку).

Даний опис подамо у вигляді орієнтованого дерева (рис.2), висота якого дорівнює 2:

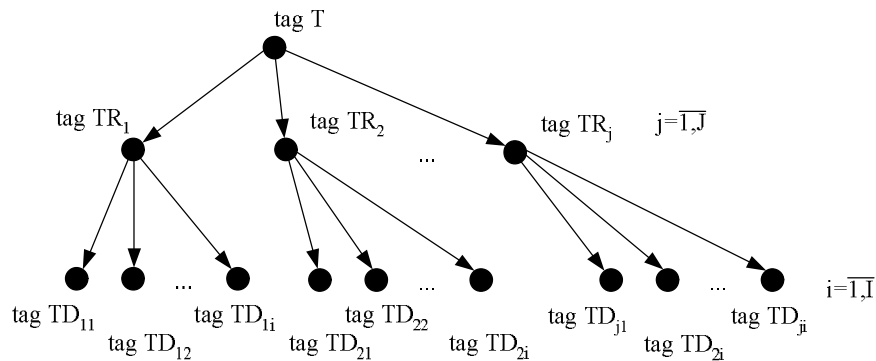


Рисунок 2 – Орієнтоване дерево, що описує структуру FDB_t

У розділі сформульовані умови, необхідні для інтеграції плоских БД з реляційними, удосконалено метод перетворення плоских БД до виду реляційного відношення на базі методів завантаження даних в СУБД, який на відміну від існуючих, дозволяє не тільки завантажити дані, але і сформувати реляційне відношення як кінцевий результат за рахунок вертикальної і горизонтальної фрагментації вихідних даних (при необхідності).

Метод реалізовано такими етапами:

Етап 1. Формування таблиці (T) в цільовій БД зі структурою даних, ідентичною структурі в плоскій БД. Іменування атрибутів проводиться згідно з вимогами реляційної моделі, тобто атрибути повинні мати унікальні імена, відповідні властивостям об'єктів реального світу;

Етап 2. Визначення типів атрибутів згідно типам вихідних даних;

Етап 3. Створення унікального ключа для всього набору атрибутів, що виключає повторюваність записів в вихідних даних шляхом генерації виняткової ситуації на рівні СУБД;

Етап 4. Завантаження даних спеціальними засобами СУБД із застосуванням рішень по фрагментації;

Етап 5. Аналіз файлів журналу спеціальних засобів на наявність у них помилок процесу завантаження даних;

Етап 6. У разі наявності помилок у структурах плоских БД та цільовій таблиці (етап 1) здійснюється коректування структури таблиці або / і структури вихідних даних і перехід до кроку 1. Якщо помилок немає – перехід до кроку 7.

Етап 7. Використання сформованої таблиці (T) з завантаженими даними як універсального відношення локальної предметної області, для автоматизації роботи

якої застосовувалася плоска БД.

Удосконалений метод перетворення плоских БД до виду реляційного відношення на базі методів завантаження даних в СУБД дозволяє застосовувати метод синтезу інтегрованої реляційної БД з декількох реляційних баз даних для інтеграції реляційних і плоских БД.

Розроблено модифікацію методу синтезу інтегрованої реляційної БД з декількох реляційних баз даних, що дозволяє проводити інтеграцію реляційних БД, нормалізованих до 3 НФ, без додаткових трудових витрат. Для опису методу введено визначення відповідних локальних відносин: локальне відношення (R') k -ої реляційної БД в 3 НФ (DB_k) відповідає локальному відношенню (R'') p -ої реляційної БД в 3 НФ (DB_p), якщо вони містять інформацію про одні й ті ж класи об'єктів реально-го світу, позначення відповідних локальних відносин декількох реляційних БД: $R'(DB_k) \leftrightarrow R''(DB_p)$.

Встановлено, що запропонований метод синтезу інтегрованої реляційної БД з декількох реляційних БД і його модифікація, забезпечують інтеграцію БД на рівні моделі даних, однак вони не враховують повною мірою функціональні особливості СУБД. Особливість полягає в тому, що в разі відповідності декількох СУБД єдиному стандарту, їх реалізація буде відрізнятися тому, що розробники, дотримуючись відповідності їх продукту стандартам, наповнюють свої продукти особливими можливостями, які вигідно виділяють його серед конкурентів.

Для мінімізації ризиків інтеграції БД, пов'язаних з такими функціональними особливостями СУБД, в третьому розділі представлений критерій оцінки доцільності інтеграції БД на основі функціональної забезпеченості СУБД і алгоритм його формування.

У загальному вигляді критерій оцінки доцільності інтеграції БД ($K_{\text{оц.ц.}}$) на основі функціональної забезпеченості СУБД представимо у вигляді формули:

$$K_{\text{оц.ц.}} = \max_{k \leq K, i \leq I, j \leq J} \left(\prod_{k=1}^K \text{IND}_{\text{CRk}} * \sum_{i=1}^I (K_{g_i} * \sum_{j=1}^J (K_j * \text{IND}_{ij})) \right), \quad (5)$$

де IND_{CRk} – критичні показники, що можуть приймати такі значення:

$$\text{IND}_{\text{CRk}} = \begin{cases} 0 \\ 1 \end{cases}, \quad k = \overline{1, K};$$

K_{g_i} – ваговий коефіцієнт групи показників, $i = \overline{1, I}$;

K_j – ваговий коефіцієнт показника в групі, $j = \overline{1, J}$;

IND_{ij} – нормоване значення i -го показника в j -й групі;

$K_{\text{оц.ц.}} = [0, 100]$.

Для формування та розрахунку критерію оцінки доцільності інтеграції декількох баз даних з урахуванням функціональної забезпеченості СУБД розроблено алгоритм.

Інтерпретація отриманого результату наступна: чим ближче $K_{\text{оц.ц.}}$ до 100, тим менше вартість реалізації інтеграції БД. На підставі експертних оцінок було запро-

поновано 3 діапазони значень. У разі потрапляння результуючого значення в перший діапазон ($K_{\text{оц.ц}} > 90$) інтеграція БД можлива, для другого діапазону ($K_{\text{оц.ц}} \geq 70$) – можлива, але зі значними витратами, для третього діапазону – не рекомендована (існує ймовірність, що і неможлива).

Якщо $K_{\text{оц.ц}} = 0$ інтеграція декількох БД неможлива. Для випадків інтеграції ідентичних СУБД (включаючи версію) немає необхідності оцінки доцільності інтеграції СУБД. Процес інтеграції знижується на рівень схеми даних.

Для формального опису процесу інтеграції баз даних запропонована прикладна інформаційна технологія інтеграції баз даних (рис. 3).

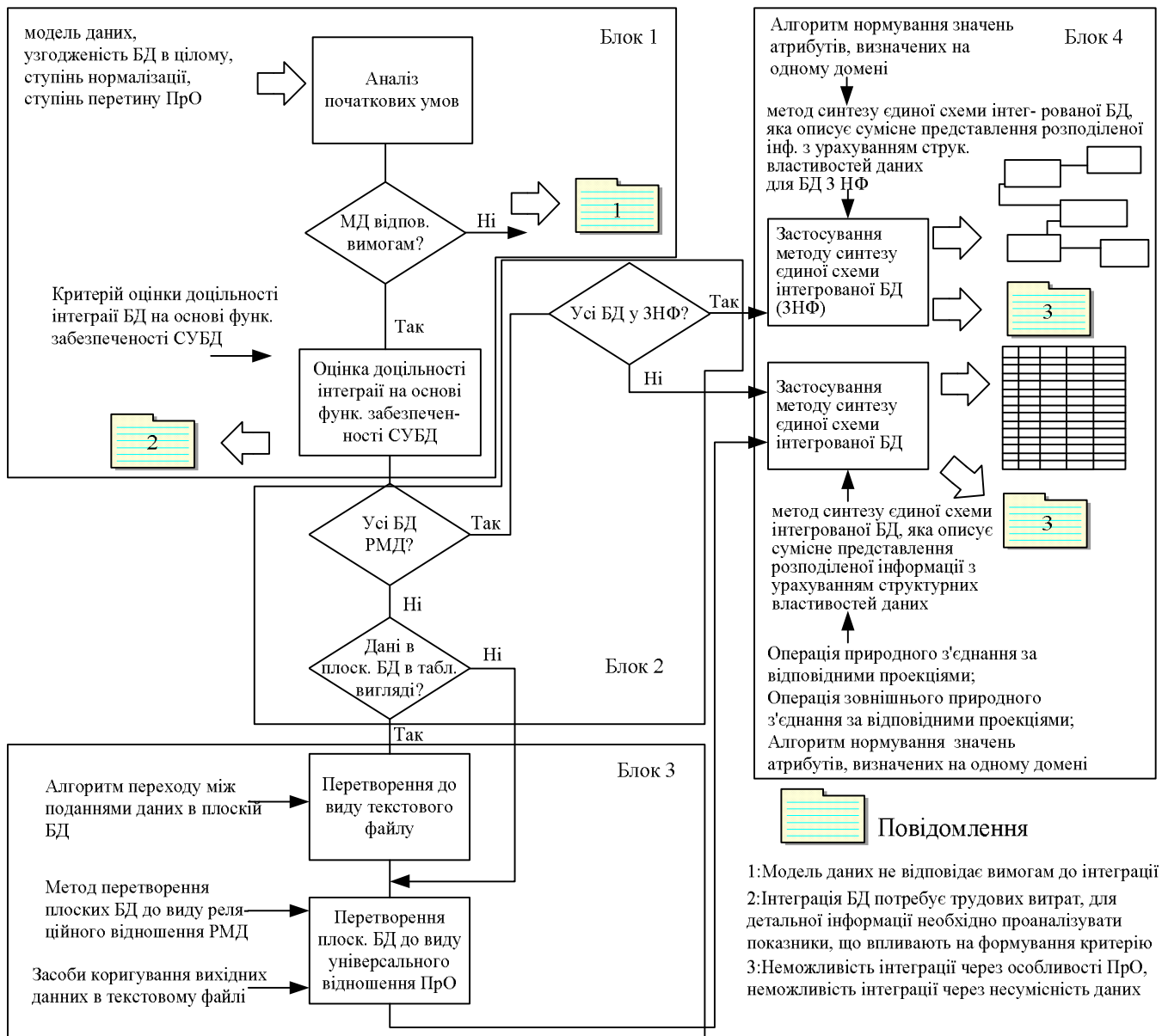


Рисунок 3 – Схема прикладної інформаційної технології інтеграції БД

Запропонована інформаційна технологія дозволяє здійснювати інтеграцію реляційних, об'єктно-реляційних БД і БД, що зберігаються в окремому файлі. Її було детально поетапно описано з усіма необхідними роботами, а також запропоновано

опис технології у вигляді псевдокоду, де максимально формалізовані етапи, які підлягали формалізації.

Розроблена інформаційна технологія заснована на розроблених моделях, запропонованих та удосконалених методах інтеграції БД, дозволяє вирішити задачі інтеграції БД, що досліджуються в дисертаційній роботі.

Четвертий розділ присвячено практичній реалізації розроблених моделей, методів та інформаційної технології інтеграції БД, наведено результати застосування інформаційної технології інтеграції баз даних. З використанням,

В якості практичної реалізації запропонованої інформаційної технології інтеграції БД розроблено інструментальний засіб інтеграції реляційних, об'єктно-реляційних і плоских БД, що зберігаються в окремому файлі. Його розроблено з використанням технології Oracle Application Express, має вигляд двох модулів: «Модуль оцінювання доцільності інтеграції БД» (1 блок IT), «Модуль інтеграції БД» (2-4 блоки IT). Дані модулі можуть використовуватися як спільно, так і окремо. Взаємозв'язок модулів представлений на рис. 4.

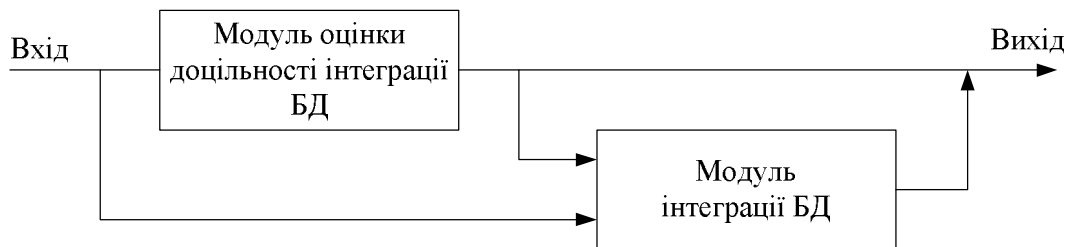


Рисунок 4 – Схема взаємодії модулів інструментального засобу інтеграції БД

Відеограми інструментального засобу приведені на рис.5 та рис.6.

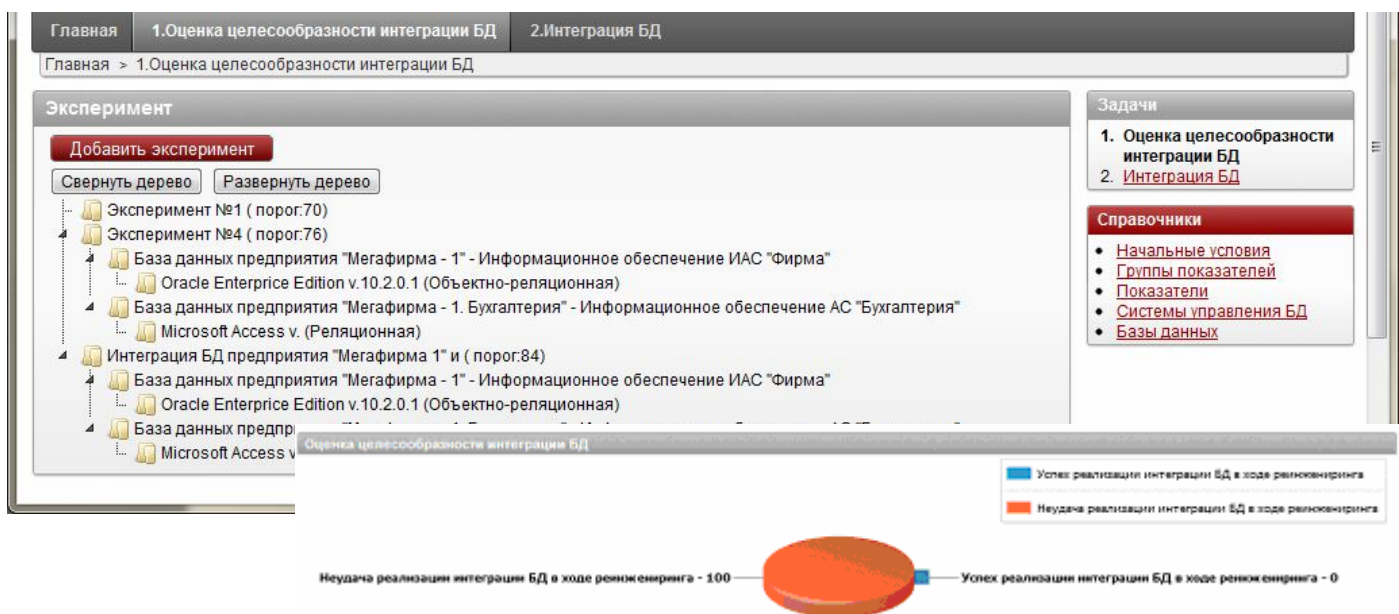


Рисунок 5 – Відеограни інструментального засобу інтеграції баз даних (Модуль оцінки доцільності інтеграції БД)

Главная > 2 Интеграция БД > Множества доменов УО

INT_SCHEMA_115:

EX_CONTRACT END_DATE DATE(7)	EX_CONTRACT IS_MAIN_STAFF CHAR(18)	EX_CONTRACT START_DATE DATE(7)	EX_DEPARTMENT DESCRIPTION VARCHAR2(255)	EX_DEPARTMENT NAME VARCHAR2(255)	EX_EMPLOYEE BIRTHDAY DATE(7)	EX_EMPLOYEE NAME VARCHAR2(20)	EX_EMPLOYEE PATRONYM VARCHAR2(20)	EX_EMPLOYEE SURNAME VARCHAR2(20)	EX_STAFF NAME VARCHAR2(20)
(select distinct END_DATE from EX_CONTRACT)	(select distinct IS_MAIN_STAFF from EX_CONTRACT)	(select distinct START_DATE from EX_CONTRACT)	(select distinct DESCRIPTION from EX_DEPARTMENT)	(select distinct NAME from EX_DEPARTMENT)	(select distinct BIRTHDAY from EX_EMPLOYEE)	(select distinct NAME from EX_EMPLOYEE)	(select distinct PATRONYM from EX_EMPLOYEE)	(select distinct SURNAME from EX_EMPLOYEE)	(select distinct NAME from EX_STAFF)
29-JAN-11, 12-MAR-12,	1, 0,	17-MAR-02, 12-DEC-11, 12-DEC-12, 29-DEC-11, 13-DEC-00, 16-MAR-12, 05-DEC-10,	Отдел, занимающийся реализацией товаров, Реализация маркетинговой стратегии организации,	Отдел продаж, Отдел маркетинга,	13-MAR-84, 31-MAR-54, 13-DEC-91, 20-DEC-73, 09-AUG-83,	Алексей, Алина, Иван,	Александровна, Иванович, Никифорович, Евгеньевич,	Иванов, Петров, Афанасьев, Онопrienko,	главный маркетолог, Реализатор, Руководитель отдела,
.29	.29	1	.29	.29	.57	.43	.57	.57	.43

INT_SCHEMA_116:

FL_MAT_TRANSLATE OTDEL VARCHAR2(255)	FL_MAT_TRANSLATE SOTRUDNIK_1 VARCHAR2(500)	FL_MAT_TRANSLATE SOTRUDNIK_2 VARCHAR2(500)	FL_MAT_TRANSLATE SOTRUDNIK_3 VARCHAR2(500)	FL_MAT_TRANSLATE CENNOSTI VARCHAR2(2000)
(select distinct OTDEL from FL_MAT_TRANSLATE)	(select distinct SOTRUDNIK_1 from FL_MAT_TRANSLATE)	(select distinct SOTRUDNIK_2 from FL_MAT_TRANSLATE)	(select distinct SOTRUDNIK_3 from FL_MAT_TRANSLATE)	(select distinct CENNOSTI from FL_MAT_TRANSLATE)
Отдел продаж, Отдел маркетинга,	Афанасьев, Иванов ,	Алексей, Иван,	Иванович, Никифорович,	ПК инв.№ 1232, ПК инв.№ 12332, ПК инв.№ 8595, принтер инв.№485, принтер инв.№484,
.4	.4	.4	.4	1

Сформировать соответствующие проекции

Рисунок 6 – Відеограма оцінки ступеню вибіркової атрибутів універсальних відносин БД, що інтегруються

В загальному вигляді структура інструментального засобу представлена на рис.7.

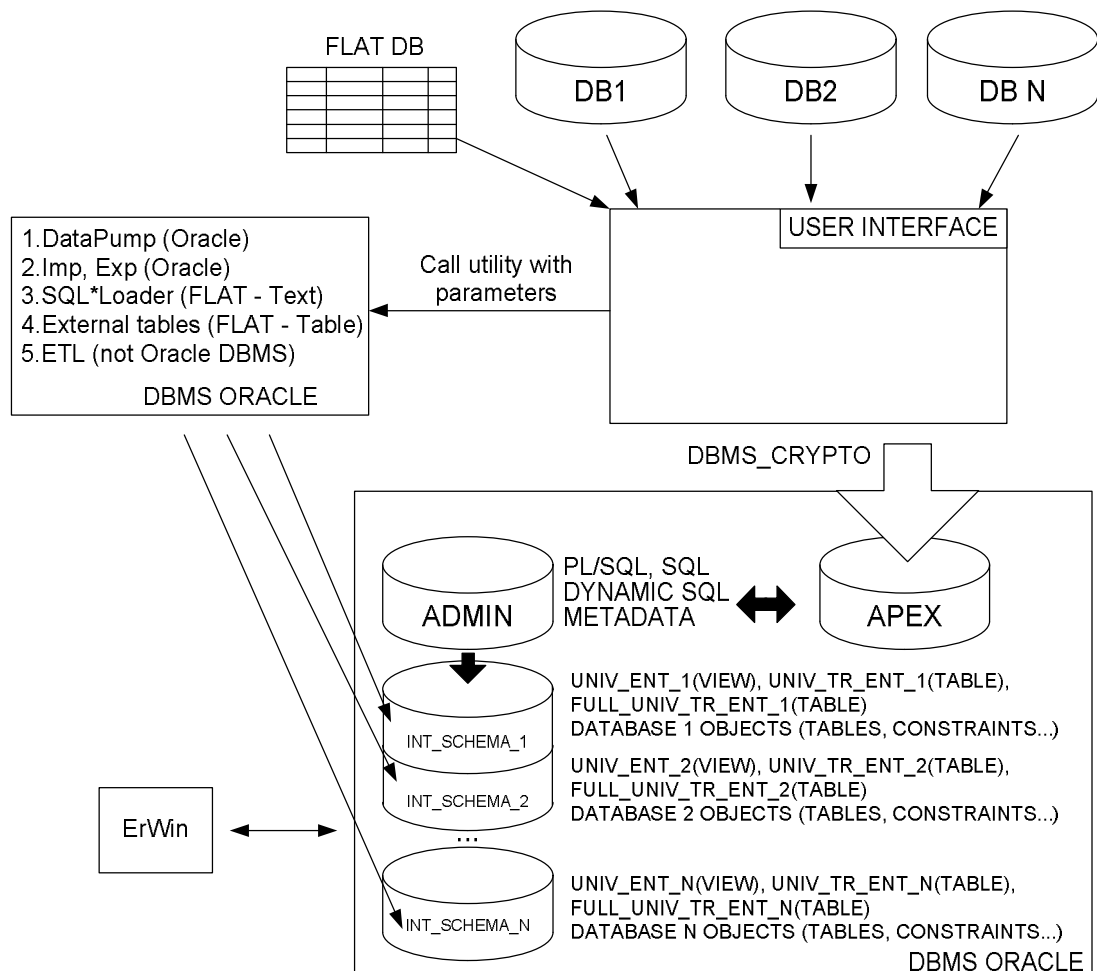


Рисунок 7 – Структура інструментального засобу інтеграції БД

Результати дисертаційної роботи впроваджені в Харківській обласній прокуратурі, Харківському обласному управлінні МНС, Харківському медичному університеті, що підтверджено відповідними актами впровадження.

У додатках до дисертації наведено акти про впровадження результатів дисертаційної роботи, приклади універсальних відношень, реалізація процедур горизонтального та вертикального фрагментування даних, опис атрибутів, псевдокод інформаційної технології.

ВИСНОВКИ

У дисертаційній роботі вирішена актуальна науково-практична задача розробки методів, інформаційної технології та інструментального засобу інтеграції БД, що дозволяють реалізувати інтеграцію реляційних, об'єктно-реляційних і плоских БД успадкованих ІС, а також знижують витрати праці і часу на проведення робіт, які пов'язані з процесом інтеграції БД. В результаті проведених досліджень отримані такі результати:

1. Удосконалено метод побудови узгодженого універсального відношення бази даних за рахунок узагальнення операцій реляційної алгебри: природного з'єднання за відповідними проекціями та зовнішнього природного з'єднання за відповідними проекціями, які дозволяють виконувати з'єднання по атрибутам із різними схемами іменування з виключенням надлишкових даних. Це дозволило зменшити часові витрати на формування універсального відношення бази даних за рахунок особливостей зазначених операцій.

2. Вперше запропоновано метод синтезу єдиної схеми інтегрованої бази даних, яка описує сумісне представлення розподіленої інформації з урахуванням структурних властивостей даних, що надало можливість реалізації інтеграції різнорідних баз даних в єдину реляційну базу даних із збереженням накопиченої інформації.

3. Удосконалено метод перетворення плоских баз даних до виду реляційного відношення, який, на відміну від існуючих, дозволяє не тільки завантажити дані, але і отримати в результаті нормалізоване універсальне відношення за рахунок застосування методів перетворення даних в ході їх завантаження.

4. Вперше запропоновано інтегральний критерій оцінки доцільності інтеграції баз даних на основі функціональної забезпеченості систем управління базами даних, який базується на використанні показників нормованих трудових витрат, що дозволяє зменшити кількість операцій в процесі побудови схеми інтегрованої бази даних.

5. Запропоновано прикладну інформаційну технологію інтеграції БД, яка дозволяє формалізувати процес інтеграції БД, визначити логічні та функціональні зв'язки між розробленими та вдосконаленими методами і алгоритмами. Представлено схематичний опис ІТ та опис її у вигляді псевдокоду.

6. Розроблено інструментальний засіб інтеграції БД (ІЗІБД), який підтримує інтеграцію реляційних, об'єктно-реляційних і плоских БД (БД OLTP-систем середніх і великих ІС). В ході інтеграції БД знижуються часові та трудові витрати і ймовірність отримання некоректних результатів (ключові операції виконуються автоматизовано). ІЗІБД дозволяє зберігати історію процесу інтеграції БД, мінімально збі-

льшуючи необхідний простір для її зберігання за рахунок використання представлень (VIEW).

7. Основні результати досліджень апробовані, застосовані та впроваджені в Центрі інформаційних систем та технологій та науково-дослідної лабораторії системних технологій при виконанні проектних робіт для Харківської обласної ради; Харківської обласної прокуратури; Харківського обласного управління МНС, Харківського медичного університету, що підтверджено відповідними актами впровадження.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових фахових виданнях

1. Левыкин, В.М. Метод синтеза единой структурной составляющей реляционной модели данных [Текст] / В.М. Левыкин, В.А. Филатов, Н.В. Черненко // Международный научный журнал: «Управляющие системы и машины». – 2011, № 6 – С. 10-13.

2. Левыкин, В.М. Интегральный критерий оценки целесообразности интеграции нескольких баз данных на основе функциональной обеспеченности СУБД [Текст] / В.М. Левыкин, Н.В. Черненко, А.В. Петриченко // Научно-технический журнал: «Радиоэлектронные и компьютерные системы» – 2011, № 4 – С. 99-103.

3. Черненко, М.В. Інформаційна технологія інтеграції БД [Текст] // «Східно-Європейський журнал передових технологій» – 2012, №6/2. – С.56-59.

4. Левыкин, В.М. Модель перехода от макро- к микропроектированию специализированных ИС [Текст] / В.М. Левыкин, Н.В. Черненко // Научный вестник КУЕИТУ: «Нові технології». – Кременчук: КУЕИТУ, 2009. – № 4 (26). – С. 40-47.

5. Левыкин, В.М. Метод интеграции реляционных и плоских БД [Текст] / В.М. Левыкин, Н.В. Черненко // Збірник наукових праць / Системи управління, навігації та зв'язку – К:ДП «Центральний науково-дослідний інститут навігації та управління» – 2012, № 1 – С. 122-126.

Друковані праці наукових конференцій

6. Левыкин, В.М. Повышение эффективности ИС посредством интеграции БД [Текст] / В.М. Левыкин, Н.В. Черненко // Сборник научных трудов VI Международной научно-практической конференции «Наука и социальные проблемы общества: информатизация и информационные технологии» – Харьков: ХНУРЭ, 2011. – С. 305-306.

7. Черненко, Н.В. Применение технологии ORACLE SPATIAL для организации доступа и управления информационными ресурсами [Текст] / Н.В. Черненко, П.А. Петров // Матер. XI Межд. молод. форума «Радиоэлектроника и молодежь в XXI веке». – Харків: ХНУРЕ, 2007. – С. 381.

8. Черненко, Н.В. Эффективное управление информационными ресурсами на основе технологии ORACLE [Текст] / Н.В. Черненко, П.А. Петров // Збірник наукових робіт / Управління розвитком – Харків: ХНЕУ, 2007. – випуск 3 – С. 82-83.

9. Черненко, Н.В. Технология автоматизированного формирования типовых документов в информационной системе [Текст] / Н.В. Черненко, А.А. Оноприенко // Збірник наукових робіт / Управління розвитком – Харків: ХНЕУ, 2008. – випуск 3 – С. 33.

10. Черненко, Н.В. Подход к автоматизированному формированию типовых

документов в информационной системе [Текст] // Матер. XII Міжнародного молод. форуму «Радиоэлектроника и молодежь в XXI веке». – Харків: ХНУРЕ, 2008. – С.202.

11. Черненко, Н.В. Технология перехода от этапа макропроектирования к этапу микропроектирования при реализации проекта информационной системы [Текст] // Материалы первой факультетской научно-практической молодежной школы семинара «Информационные интеллектуальные системы» – Харьков: ХНУРЭ, 2009. – С.73-76.

АНОТАЦІЯ

Черненко М.В. Методи та інформаційна технологія інтеграції баз даних корпоративних систем. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Харківський національний університет радіоелектроніки, Міністерство освіти та науки України. – Харків, 2013.

Дисертаційна робота присвячена вирішенню актуальної науково-практичної задачі розробки нових методів та інформаційної технології інтеграції баз даних.

Запропоновано моделі операцій природного з'єднання реляційних відношень за відповідними проекціями та зовнішнього природного з'єднання реляційних відношень за відповідними проекціями, які дозволили удосконалити метод побудови узгодженого універсального відношення бази даних та реалізувати метод синтезу інтегрованої реляційної бази даних з декількох реляційних баз даних. Запропоновано формальний опис плоских баз даних. Удосконалено метод перетворення плоских БД до виду реляційного відношення на базі методів завантаження даних в СУБД. Запропоновано критерій оцінки доцільності інтеграції баз даних на основі функціональної забезпеченості СУБД.

Запропоновано прикладну інформаційну технологію інтеграції баз даних з використанням запропонованих методів. Розроблено програмну реалізацію інструментального засобу інтеграції баз даних. Зазначена технологія та інструментальний засіб використані в процесі реінженірингу успадкованих інформаційних систем, впроваджені в декількох організаціях.

Ключові слова: метод, інформаційна технологія, інтеграція баз даних, реляційна модель даних, плоскі бази даних, критерій оцінки доцільності інтеграції баз даних, функціональне забезпечення систем управління базами даних.

АННОТАЦИЯ

Черненко Н.В. Методы и информационная технология интеграции баз данных корпоративных систем. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии. – Харьковский национальный университет радиоэлектроники, Министерство образования и науки Украины. – Харьков, 2013.

Диссертационная работа посвящена решению актуальной научно-практической задачи разработки новых и усовершенствованию существующих методов интеграции баз данных (БД), а также разработке информационной технологии интеграции БД.

Первый раздел посвящен анализу современного состояния проблемы лоскутной автоматизации организации, недостатков, вызванных ею. В качестве ее решения предложено провести интеграцию БД унаследованных информационных систем (ИС) в ходе реинжиниринга. Показана актуальность решения задач интеграции БД, проведено исследование современных подходов, методов, инструментальных средств интеграции БД. Обоснована необходимость разработки новых, усовершенствования уже существующих методов и информационной технологии интеграции БД.

Во втором разделе проведено исследование реализации проектов ИС в разрезе применимости моделей данных. Установлено, что наиболее востребованными являются реляционные и объектно-реляционные системы управления базами данных (СУБД), также часто используются плоские БД. Разработаны модели двух операций: естественного соединения реляционных отношений по соответствующим проекциям и внешнего естественного соединения реляционных отношений по соответствующим проекциям. Эти операции, в отличие от существующих, позволяют произвести соединение реляционных отношений без необходимости полного совпадения имен атрибутов и устраняя избыточные данные. Важной особенностью также является задание структурного ограничения при выполнении операций.

Введенные операции позволили усовершенствовать метод построения согласованного универсального отношения базы данных, а также разработать метод синтеза интегрированной реляционной БД из множества реляционных БД. Разработанный метод позволяет произвести интеграцию реляционных и объектно-реляционных баз данных, нормализованных до 3 нормальной формы (3НФ), с обеспечением сохранности накопленных данных.

В третьем разделе представлена модель плоских БД. Усовершенствован метод преобразования плоских БД к виду реляционного отношения, который в отличие от существующих, позволяет не только загрузить данные из плоских БД, но и привести исходные данные к виду реляционного отношения, за счет горизонтальной и вертикальной дефрагментации данных. Данный метод позволил расширить область применения метода, разработанного во втором разделе за счет возможности интеграции реляционных, объектно-реляционных и плоских БД. Разработана модификация метода синтеза интегрированной реляционной БД из множества реляционных БД для случаев интеграции БД, нормализованных до третьей нормальной формы, что позволило исключить затратные этапы базовой модификации при интеграции БД одной степени нормализации (3НФ).

Разработан критерий оценки целесообразности интеграции БД на основе функциональной обеспеченности СУБД, что позволило на ранних этапах интеграции БД оценить трудозатраты, связанные с использованием особенностей СУБД, которые могут определять успех всего процесса интеграции.

Разработана прикладная информационная технология интеграции БД на основе предложенных методов, критерия и операций. Представлено вербальное поэтап-

ное описание, схематическое описание и описание в виде псевдокода.

В четвертом разделе выполнена практическая реализация разработанной информационной технологии, которая получила свое развитие в инструментальном средстве интеграции БД. Описана программная реализация инструментального средства интеграции БД.

Применение предложенных методов и информационной технологии позволило разработать инструментальное средство интеграции БД, исключающее недостатки классических ETL-средств.

Анализ внедрения результатов диссертационных исследований показал, что использование разработанных методов, информационной технологии и инструментального средства интеграции БД позволило: сократить трудовые и временные затраты на выполнение процесса интеграции БД, при этом снизить вероятность формирования некорректного результата; снизить стоимость процесса интеграции за счет снижения временных ресурсов, а также отказа от дорогостоящих и малоэффективных инструментальных средств; уменьшить количество ошибок при интеграции БД за счет автоматизации процессов формирования структуры целевой БД и ее наполнения.

Ключевые слова: метод, информационная технология, интеграция баз данных, реляционная модель данных, плоские базы данных, критерий оценки целесообразности интеграции баз данных, функциональная обеспеченность систем управления базами данных.

ABSTRACT

Chernenko N.V. Methods and information technology for databases integration of corporate systems. – Manuscript.

The thesis for Candidate's degree in engineering science in specialty 05.13.06 – Information technologies. – Kharkov National University of Radioelectronics, Ministry of education and science of Ukraine. – Kharkov, 2013.

The study in this thesis concerns solving an actual scientific and practical problem of developing new information technology methods for databases integration.

Two natural join relation models were designed. (1) Natural join relations on conforming projections and (2) outer natural join relations on conforming projections. These operations allow implementing a method of integrated database synthesis from a variety of legacy information system databases. Formal flat databases description is included. The study involves improving method of transforming flat databases into relational form using methods of loading data into DBMS.

An information technology of database integration based on developed methods is proposed. A software tool for database integration is implemented. The information technology and the tool were used in the process of legacy information systems re-engineering in several organizations.

Keywords: method, information system, database integration, relational data model, flat database, database integration expediency estimation criterion, functional providing of DBMS.

Підп. до друку _____.2013 р.

Формат 60x84 ¹/₁₆.

Умов. друк. арк. 0,9

Зам. № Тираж 100 прим.

Папір друк. офсетний

Облік. вид. арк. 1,8

Надруковано в навчально-науковому видавничо-поліграфічному центрі
Харківського національного університету радіоелектроніки
Україна, 61166, м. Харків, пр. Леніна, 14