

Міністерство освіти і науки, молоді та спорту України
Харківський національний університет радіоелектроніки

На правах рукопису

ТАНЯНСЬКИЙ СЕРГІЙ СТАНІСЛАВОВИЧ

УДК 004.475 : [004.652 : 004.655]

МОДЕЛІ, МЕТОДИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ІНТЕГРАЦІЇ
ГЕТЕРОГЕННИХ РОЗПОДІЛЕНИХ БАЗ ДАНИХ

05.13.06 – інформаційні технології

Автореферат дисертації на здобуття наукового ступеня доктора
технічних наук

Харків – 2011

Дисертацією є рукопис.

Робота виконана в Харківському національному університеті радіоелектроніки Міністерства освіти і науки, молоді та спорту України.

Науковий консультант доктор технічних наук, професор
Філатов Валентин Олександрович,
Харківський національний університет
радіоелектроніки,
професор кафедри штучного інтелекту

Офіційні опоненти: доктор технічних наук, професор
Жолткевич Григорій Миколайович,
Харківський національний університет ім.
В.Н. Каразіна, завідувач кафедри теоретичної та
прикладної інформатики

доктор технічних наук, професор
Каргін Анатолій Олексійович,
Донецький національний університет,
завідувач кафедри комп'ютерних технологій

доктор технічних наук, професор
Пасічник Володимир Володимирович,
Національний університет "Львівська політехніка",
завідувач кафедри інформаційних систем та мереж

Захист відбудеться “ _____ ” _____ 2011 р. о _____ годині на засіданні спеціалізованої вченої ради Д 64.052.08 Харківського національного університету 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 “ _____ ” _____ 2011 р.

Учений секретар
спеціалізованої вченої ради

І.П. Плісс

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Засоби автоматизації проектування баз даних (БД) мають здебільшого частині допоміжний характер і не гарантують ефективності обробки даних і їхньої вірогідності. Навіть незначні зміни в поданні інформаційної моделі предметної області (ПрО) можуть викликати необхідність виконання великого обсягу робіт по перепроєктуванню та перепрограмуванню прикладних задач. Особливо гостро постають питання ефективності та надійності функціонування розподілених систем БД, створення яких стало можливим з розвитком як технічних засобів комунікаційних мереж, так і з розвитком програмних засобів підтримки спільної обробки даних.

Практичні задачі визначають необхідність проведення наукових досліджень з широкого кола питань проектування та організації функціонування розподілених систем БД. Зокрема, деякі задачі, які розглядаються в дисертації, ґрунтуються на результатах досліджень теорії реляційних БД, що стала надійним інструментом вирішення багатьох важливих проблем. Серед них - побудова формального апарату опису обмежень цілісності у формі правил (залежностей), які наведено в роботах В. Армстронга, К. Делобеля, Е. Кодда, К. Бірі, Е. Тененбаума, Дж. Фрідла, В. Пасічника, побудова ефективних логічних схем БД (наведено в роботах Ф. Бернштейна, Д. Мейера, Т. Лінга, А. Голосова, Е. Неклюдової, О. Горчинської), організація функціонування БД із неповною інформацією – в роботах Дж. Біскупа, В. Ліпського, К. Заніоло, автоматизація програмування БД – в роботах Б. Ігнатенко, О. Вейнерова, Г. Фурсіна, А. Замуліна, Д. Мінкера та Х. Фернадеса, оптимізація виконання запитів у середовищі інтегрованих і розподілених БД – в роботах Л. Калиніченко, С. Кузнєцова, К. Чері, А. Карденаса, Дж. Ульмана, П. Хола, Е. Дедікова, М. Бусліка, В. Філатова та ін.

Проведені в дисертаційній роботі дослідження спрямовані на створення систем інтеграції та управління інформаційними ресурсами розподілених обчислювальних систем. Під інтеграцією розуміється управління гетерогенною інформацією, що дозволить організувати доступ до неоднорідних даних, що втримуються в сформованих структурах – файлах даних і базах даних.

З іншого боку, розробка будь-якої інформаційної системи (ІС) починається не тільки з побудови сховища даних, але й із створення інформаційної, онтологічної, структурної або іншої моделі ПрО, на роботу з якою орієнтована ІС. На жаль, побудована модель не буває абсолютно точною, абсолютно повною або абсолютно коректною. Зокрема, при побудові ІС, в основі яких лежать БД або інформаційні сховища, розробники зіштовхуються із проблемою вибору сутностей, що повинні бути включені в інформаційну модель БД. На сьогоднішній день цей вибір здійснює розробник інтуїтивно на основі знань про ПрО.

У процесі роботи з моделлю ПрО, а також у процесі роботи вже побудованої ІС виникає необхідність уточнення або розширення моделі ПрО, розширення інформації про об'єкти, які втримуються в ПрО і відбиваються у відповідній моделі, тобто виникає необхідність маніпулювання моделлю і виконання різних операцій над елементами цієї моделі.

Консолідація бізнес-процесів вимагає розв'язання задач інтеграції ІС, створення єдиного інформаційного простору, в основі якого перебуває розподілене сховище даних. Об'єднання відповідних інформаційних моделей на сьогоднішній день виконується розробниками емпірично та інтуїтивно.

У процесі функціонування ІС або в процесі модифікації моделі ПрО, що лежить в основі ІС, виникає необхідність урахувувати взаємодію ПрО із зовнішнім середовищем і, відповідно, виникає проблема обліку в моделі ПрО цієї взаємодії. Саме семантична неповнота моделей БД і відсутність операційної специфікації маніпулювання моделями ПрО може порушувати коректність інтеграції інформації в сховищах даних.

Таким чином, існує комплекс проблем, нерозв'язаність яких приводить до неповноти баз даних, що лежать в основі ІС. Відсутність відповідних математичних засобів приводить до необхідності вирішення подібних проблем, оперативність яких сильно залежить від суб'єктивного фактора.

Відсутність єдиного підходу до вирішення подібних проблем особливо серйозно впливає на якість управління розподіленими гетерогенними ІС. Некоректність подання інформаційних об'єктів ПрО, отриманих у результаті маніпулювання елементами моделі ПрО, неповнота або суперечливість інформації, витрати часу на перетворення інформації приводить, в загальному підсумку, до неадекватних управлінських рішень.

На сьогоднішній день існуючі підходи в роботі із БД і моделями ПрО не орієнтовані на вирішення виділених проблем. Можна відзначити, що тільки в рамках реляційної моделі існує розвинений математичний апарат, що дозволяє формалізувати роботу з інформаційними моделями та забезпечує маніпулювання даними і метаінформацією ПрО (хоча семантика даних реалізована не досить повно). Це правила нормалізації відносин, формалізації зв'язків, реляційна алгебра та реляційне числення. Однак ці засоби не призначені для маніпулювання безпосередньо інформаційними моделями ПрО. Крім цього, ні реляційні, ні об'єктні, ні навіть онтологічні моделі, які одержали значну популярність, не враховують задачі, що розв'язуються у відповідних ПрО при динамічних змінах інформаційних об'єктів і схем БД.

Таким чином, розробка інформаційних і математичних засобів, які забезпечують підвищення ефективності функціонування інтегрованих інформаційних систем організаційного управління, є актуальним напрямком і потребує подальшого дослідження, що й визначило тему даної роботи.

Дисертаційна робота є продовженням досліджень, які виконуються в Харківському національному університеті радіоелектроніки під керівництвом доктора технічних наук, професора В.О. Філатова.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційну роботу виконано в рамках держбюджетних НДР, що здійснювались згідно з наказами Міністерства освіти і науки України за результатами конкурсного відбору проектів таких наукових досліджень:

- НДР № 173 «Розробка методологічних основ та інструментальних засобів створення просторових систем підтримки прийняття рішень» (№ ДР0103U001533);

- НДР № 200 «Методи та технології створення інформаційно-освітнього середовища з метою інтеграції у загальноєвропейський простір вищої освіти» (№ ДР0106U003152);

- НДР № 214 «Синтез методів обробки інформації за умов невизначеності на основі самонавчання та м'яких обчислень» (№ ДР0107U003028);

- НДР № 245 «Еволюційні гібридні системи обчислювального інтелекту зі змінною структурою для інтелектуального аналізу даних» (№ ДР0110U000458).

При виконанні зазначених НДР здобувачем як виконавцем синтезовано моделі предметних областей і методи спільної обробки інформації в умовах неоднорідності структур баз даних для систем підтримки аналізу даних у задачах інтеграції розподілених обчислювальних систем, що дозволяє покращити гнучкість інтегрованої системи в цілому та забезпечити формування динамічних запитів безпосередньо до локальних даних.

Мета та задачі дослідження. Метою дисертаційної роботи є розвиток теорії інтеграції розподілених гетерогенних ІС з локально незалежними структурами БД, побудова моделей визначення інформаційних об'єктів ПрО, розробка методів та інформаційних технологій підтримки цілісності інтегрованих даних при модифікації стану локальних БД; створення нового підходу для розв'язання проблеми ефективної обробки гетерогенної інтегрованої інформації зі структурною нестабільністю локальних даних.

Для досягнення поставленої мети розв'язуються такі наукові задачі:

- розробка моделі предметної області на підставі семантичних властивостей і зв'язків елементів даних;
- розробка ефективної технології інтеграції інформаційних ресурсів гетерогенних обчислювальних систем на основі технології управління системами баз даних;
- дослідження і формалізація класів неоднорідності структур даних, спрямованих на розв'язання задачі визначення типів інформаційних об'єктів;
- розробка і дослідження архітектури і технологічних рішень організації баз даних з урахуванням неоднорідності структур;
- розробка і дослідження засобів логічного опису властивостей інформаційних об'єктів для визначення межі розглянутої предметної області;
- розробка теоретико-доказової системи виведення інформаційних об'єктів предметної області на підставі заданої множини правил логічного існування;
- розробка методів побудови еквівалентних множин інформаційних об'єктів для коректної логічної трансформації структур баз даних;
- розробка і дослідження методів підтримки структурної цілісності при модифікації схеми бази даних;
- дослідження критеріїв припустимої модифікації структур баз даних щодо заданих обмежень;
- розробка методу переходу станів баз даних щодо обмежень цілісності, дослідження умов коректної модифікації даних;
- дослідження критеріїв спільності модифікації даних і визначення обмежень для множини припустимих станів бази даних;

- розробка і дослідження методу часткового відображення моделей даних (без обліку відображення операційної специфікації);
- розробка і дослідження методу доступу до табличних даних ієрархічної та мережевої структури засобами числення зі змінними на доменах;
- дослідження математичних засобів опису дедуктивних баз даних для формулювання запитів до неоднорідних даних спільного використання.

Вирішення розглянутих у дисертаційній роботі задач необхідно для побудови інтегрованого інформаційного простору, в якому допускається локальна незалежність баз даних при забезпеченні узгодженості та актуальності інформації.

Об'єктом дослідження є інформаційні технології управління спільним доступом до розподілених БД з локально незалежними структурами (моделями).

Предметом дослідження є моделі і методи опису та обробки неоднорідних БД в інтегрованих інформаційних системах організаційного управління.

Методи дослідження. Для інформаційного моделювання предметних областей використана теорія множин, загальна теорія систем, теорія баз даних, теорія графів. Для розробки методів підтримки цілісності даних використана теорія логічних залежностей і правила дедуктивного виведення. Для трансформації структур даних використана теорія міжмодельних відображень і теорія типів. Для розробки методів обробки неоднорідних даних використаний апарат теорії числення зі змінними на доменах і виразів логіки предикатів першого порядку. Для розробки засобів реалізації управління розподіленими даними використані технології аналізу і проектування інформаційних систем, реляційних і дедуктивних БД.

Наукова новизна отриманих результатів роботи полягає в узагальненні та розвитку теорії інтеграції гетерогенних інформаційних систем з локально незалежними БД та створення на цій основі підходу до розв'язання задач обробки розподілених даних без побудови єдиної інтегрованої схеми. Основні результати, які визначають наукову новизну дисертаційної роботи, полягають у такому:

вперше:

- запропоновано інфологічну модель предметної області у вигляді множини інформаційних об'єктів, на якій задається логічна функція існування інформаційних об'єктів; визначено властивості основних компонент моделі предметної області, які характеризуються семантикою відповідної бази даних; на відміну від існуючих підходів така формалізація надає можливість використовувати теоретико-доказовий метод обробки даних;
- сформульовано систему аксіом для опису властивостей предметної області, що дає змогу розглядати проблему моделювання відношень її елементів у вигляді набору правил, які визначають існування елементів даних; на цій підставі запропоновано загальний підхід до побудови інформаційної технології, орієнтованої на розв'язання задач управління гетерогенними інформаційними ресурсами обчислювальних систем;
- введено клас правил - обмежень структурної цілісності, що визначає спільність даних при модифікації структури БД; запропоновано та теоретично обґрунтовано інформаційну технологію підтримки структурної цілісності інтегрованої схеми

даних при припущенні змін у схемах локальних даних, яка дозволяє будувати еквівалентні стани баз даних при модифікації семантики предметної області;

- запропоновано метод, який забезпечує цілісність інформації, що модифікується, при неповній відповідності обмеженням цілісності; визначено умови, при яких модифікація даних сумісна з обмеженнями, тобто може бути виконана зміна стану бази даних без порушень відповідних обмежень; на відміну від існуючих методів, такий підхід дозволяє виконувати запити до баз даних, типи даних яких мають різні властивості;
- запропоновано метод відображення моделей даних без урахування відображення операційної специфікації; сформульовано умови, при яких відображення є комутативним; це дає змогу при інтеграції використовувати операційну специфікацію тієї моделі, яка відповідає локальним системам управління базами даних;

дістали подальший розвиток:

- метод перевірки належності правила до деякої повної множини, що опирається на послідовність алгоритмів обчислення замикання, тобто розв'язано проблему членства, що дає підставу для розв'язання задачі побудови множини еквівалентних правил без використання безпосередньо замикання, що значно скорочує загальний час обчислення семантики бази даних;
- метод обчислення мінімальної еквівалентної множини правил, що базується на побудові ненадлишкового покриття, який дозволяє за прийнятний час визначати ступінь перетинання предметних областей, описуваних множинами інформаційних об'єктів і правилами логічного існування;
- метод відображення реляційної моделі в дедуктивну модель на підставі бієктивної відповідності компонентів моделі: даних, відношень, операційної специфікації, що дає можливість інтегрувати неоднорідні бази даних, використовуючи систему визначення даних, не пов'язаних зі структурою відношень;
- метод підтримки цілісності даних при зміні стану бази даних, використовуючи властивості зведення обмежень, при яких стан бази даних залишається спільним, що забезпечує обчислення найменшої нерухомої точки, тобто бажаний результат виводиться за кінцеве число кроків;

удосконалено:

- метод аналізу семантичних властивостей інформаційних об'єктів предметної області за рахунок введення двох типів правил логічного існування: функціональних і структурних, які дозволяють аналізувати не тільки явно задані, а і виведені інформаційні об'єкти;
- метод визначення мінімальних змін семантики вихідної структури даних, при яких структурні обмеження залишаються спільними, тим самим підвищується ефективність побудови еквівалентних станів бази даних.

Практичне значення одержаних результатів дисертаційної роботи підтверджено поліпшенням якості функціонування реляційних систем управління БД як складової частини інтегрованих ІС. Зокрема, практичні застосування теоретичних досліджень полягають у такому:

- розроблено і реалізовано процедури моніторингу активних серверів баз даних для виявлення змін властивостей елементів даних;
- розроблено процедуру для реалізації методу формування та обробки вільних транзакцій, що забезпечують доступ до розподілених даних, розташованих на активних серверах інтегрованої системи;
- удосконалено технологію формування динамічних запитів з використанням виразів реляційного числення зі змінними на доменах;
- автоматизовано процедуру редукції операцій реляційної алгебри до предикатних виразів мови запитів дедуктивних баз даних;
- удосконалено технологію використання дедуктивних баз даних для організації доступу до реляційних баз даних;
- реалізовано метод подання інтенціонала даних і процедур виведення кінцевого результату засобами реляційної системи управління БД;
- розроблено інтерфейс зв'язку реляційної системи управління БД і компілятором дедуктивного виведення, причому передача даних здійснюється в обох напрямках: «таблиця → інтенціонал», «інтенціонал → таблиця».

Практичні результати дисертаційної роботи підтверджуються позитивним ефектом від застосування розроблених засобів у процесі побудови модулів управління інтегрованими інформаційними ресурсами в різних областях господарської діяльності: в системі інформаційної підтримки технологічної підготовки виробництва ДП «Завод ім. В.А. Малишева» в рамках науково-технічної роботи «Розробка інформаційного і програмного забезпечення системи інтеграції логічно незалежних неоднорідних баз даних». Розроблена методика опису інтегрованих задач доступу до даних, на базі яких формуються об'єкти інформаційної системи, дозволили більш ефективно використовувати засоби обробки інформації, що зберігається в локальних базах даних (акт від 10.01.2007 р.); при вдосконаленні розподіленої регіональної інформаційно-аналітичної системи Північного територіального управління ЗАТ «Український мобільний зв'язок» для обліку статистичних даних завантаженості ліній зв'язку, а також кількості успішних (не успішних) викликів. Для підвищення ефективності вирішення поставленої задачі використовувалася модифікована архітектура системи інтеграції БД, заснована на міжмодельних відображеннях при спільній обробці інформації (акт від 22.12.2008 р.); в корпоративній багаторівневій інтегрованій автоматизованій системі управління інформаційними потоками ДК «Укртрансгаз». Підтримка узгодженості інформації здійснюється за допомогою програмного шлюзу, що забезпечує виконання як безпосередньо SQL запитів, так і підключення (при необхідності) додаткових процедур управління, що зберігаються на сервері (акт від 20.11.2009 р.); при вдосконаленні інформаційної системи обліку товарів виробництва та формування звітної документації на ДП ХМЗ «ФЕД». Програмна реалізація підтримки інтенціонала БД забезпечила незалежність ведення локальних даних у підрозділах (акт від 18.12.2010 р.); при виконанні держбюджетних науково-дослідних робіт відповідно до тематичного плану Харківського національного університету радіоелектроніки; у навчальному процесі Харківського національного університету радіоелектроніки при проведенні лекційних, практичних і

лабораторних занять по дисциплінах, спрямованим на вивчення методів організації і підтримки БД, в курсових і дипломних проектах, а також в спецкурсах для магістрів, здобувачів і аспірантів (акт від 27.09.2011 р.).

Особистий внесок здобувача. Всі наукові результати дисертаційної роботи, що виносяться на захист, отримані автором самостійно. У роботах, опублікованих у співавторстві, авторові належать такі результати: [1] - розробка моделі семантики БД, визначення класів неоднорідних таблиць; [2] - побудова структури розподіленої структури БД для проведення комплексного аналізу даних; [3] - визначення послідовності таблиць, що з'єднуються, при динамічних змінах стану БД; [4] - розробка методу обробки невизначених значень при виконанні запитів, забезпечення підтримки цілісності при допущенні невизначеностей в таблицях БД; [5] - розробка моделі інтегрованої інформаційної системи з урахуванням семантичної непогодженості локальних структур БД; [6] - розробка методу вибору послідовності з'єднань при забезпеченні найменшої потужності проміжного відношення; [7] - розробка методу розрахунку кількості ітерацій при виконанні складної композиції операцій над розподіленими даними; [8] - класифікація запитів по організації їхнього виконання (прості, багатотабличні, складені, вкладені, корельовані, квантифіковані тощо), визначення характеристичних властивостей оцінки виконання запиту; [11] - розробка моделі динамічної транзакції для обробки незалежних потоків даних у розподілених системах БД; [13] - класифікація та опис засобів подання неоднорідних структур у табличних документах, підтримуваних реляційними системами управління БД; [15] - опис властивостей обмежень цілісності при з'єднанні розподілених БД; [17] - визначення критеріїв оцінки інтегрованої системи на підставі характеристик локальних систем БД; [21] - розробка програмного засобу обробки неоднорідних даних на основі розширення операційної специфікації реляційної алгебри; [24] - визначення поняття часткового міжмодельного відображення з урахуванням збереження структурних інваріантів БД; [26] - визначення властивостей реляційної моделі при допущенні невизначеностей в екземплярі відношення; [27] - розробка методу декомпозиції даних у відношеннях розподіленої БД з підтримкою цілісності при множинній асоціації зв'язних атрибутів; [28] - розширення операційної специфікації об'єктно-реляційної моделі з обліком складених кортежів об'єктних відношень; [30] - розробка методу редукції операцій реляційної алгебри до виразів дедуктивної БД; [32] - подання складених алгебраїчних формул у вирази реляційного числення зі змінними на доменах; [33] - розробка методу визначення неоднорідних БД у термінах реляційної моделі даних; [34] - опис властивостей предметної області засобами логічних правил існування інформаційних об'єктів; [35] - опис структури БД, при якій можливе дублювання даних ключа для ідентифікованих користувачів різного рівня доступу; [36] - розробка засобів підтримки неузгодженої структури розподіленої БД для аналітичної обробки інформації; [37] - розробка методу узгодження даних при об'єднанні даних динамічних серверів БД; [38] - модифікація мовних засобів обробки даних з використанням розширеної операційної специфікації реляційної моделі даних; [40] - опис методу відображення структурних компонент моделей даних, зокрема реляційної моделі і моделі,

заснованої на правилах (дедуктивної); [41] - розробка методу еквівалентного перетворення запиту при семантичній неузгодженості схем БД; [42] - визначення властивостей неузгоджених структур БД і критеріїв еквівалентності схем даних; [43] - розробка системи побудови ациклічної схеми з урахуванням перевірки узгодженості даних шляхом попарного з'єднання відношень БД; [44] - визначення критеріїв для оцінки якості структури БД на підставі встановлення додаткових обмежень на структурні зміни відношень БД; [45] - розробка моделі багаторівневої транзакції для обробки розподілених даних з невизначеними значеннями; [47] - розробка засобів контролю послідовності з'єднань при реалізації багатотабличних запитів; [48] - опис властивостей інтегрованого інформаційного простору, визначення основних компонентів управління потоками транзакцій; [49] - розробка засобів доступу до розподілених даних, заснованих на операції q - з'єднання з використанням диз'юнктивної умови; [50] - розробка засобів опису властивостей моделі предметної області з урахуванням заданих обмежень; [51] - опис алгоритму побудови мінімальної семантики БД як множини зв'язних інформаційних об'єктів ПрО; [52] - опис моделі і властивостей інформаційних об'єктів ПрО з обліком їхньої структурної нестабільності в процесі життєвого циклу ІС; [53] - розробка моделі інтегрованої ІС із розподіленою підтримкою систем БД; [54] - розробка методу підтримки цілісності в неузгоджених БД із циклічною схемою; [55] - розробка методу формування сильно зв'язних блоків у циклічних схемах розподілених БД; [56] - опис методу застосування операцій реляційної алгебри для БД із неоднорідною структурою; [57] - розробка методу порівняння графічних даних по їхньому цифровому коду; [58] - розробка методу формування динамічного запиту з урахуванням наявності невизначених значень в атрибутах відношення; [61] - подання компонентів моделі даних і побудова діаграм покомпонентного відображення; [62] - опис властивостей обмежень цілісності, що визначають структуру схеми БД; [63] - розробка методу трансформації структури БД при збереженні еквівалентності залежностей даних; [64] - розробка моделі подання інформаційних об'єктів ПрО; [65] - визначення структури підпорядкованого запиту, що реалізує доступ до даних з використанням умови, що корелює; [67] - опис компонентів логічної мови маніпулювання даними реляційної структури; [68] - розробка методу побудови схеми БД на основі множини незалежних таблиць, що представляють сутності однієї ПрО; [69] - подання об'єктно-реляційної схеми в термінах множин функціональних залежностей; [70] - опис властивостей функціональних залежностей для пошуку неповторюваної послідовності значень у відношенні БД; [71] - розробка методу формування складної транзакції для обробки розподілених даних з динамічним визначенням активного серверу БД; [72] - розробка методу визначення залежностей між даними з урахуванням порожніх значень в атрибутах відношення; [73] - визначення структурних компонент реляційної моделі для визначення семантичних властивостей ПрО; [74] - розробка засобів оцінки якості схеми БД; [75] - опис послідовності виконання правил виводу факту для різних операцій реляційної алгебри; [76] - опис обмежень цілісності даних у термінах предикатних правил зі змінними на доменах; [78] - побудова схем виконання складних запитів для їхньої трансформації в множину логічних правил

виводу факту; [79] - розробка методу побудови однозначних обмежень, при яких існує єдина семантика БД; [80] - доказ існування границі послідовності рекурсивної логічної програми обчислення семантики бази даних.

Апробація результатів дисертації. Основні положення та результати дисертаційної роботи були представлені, доповідалися і обговорювалися на міжнародних і всеукраїнських наукових конференціях і семінарах, зокрема на: 3-й, 4-й, 6-й, 8-й, 9-й, 10-й Міжнародних конференціях «Теорія й техніка передачі, прийому і обробки інформації» (Харків-Туапсе, 1997, 1998, 2000, 2001, 2002, 2003, 2004 рр.); науково-методичній конференції «Використання комп'ютерних технологій у навчальному процесі» (Харків, 1997 р.); 10-й Міжнародній науковій конференції «Інформатизація правоохоронних систем» (Москва, Росія, 2001 р.); науково-практичному семінарі «Сучасні проблеми інформатизації органів внутрішніх справ» (Київ, 2001 р.); 10-й, 15-й, 16-й, 17-й Міжнародних конференціях по автоматичному управлінню «Автоматика - 2003», «Автоматика - 2008», «Автоматика - 2009», «Автоматика - 2010» (Севастополь, 2003 р., Одеса, 2008 р., Чернівці 2009 р., Харків 2010 р.); міжнародній науковій конференції «Освіта, наука, виробництво і управління в XXI столітті» (Старий Оскол, Росія, 2004, 2005 рр.); 7-й, 10-й, 11-й, 12-й Міжнародних науково-технічних конференціях «Системний аналіз і інформаційні технології. SAIT» (Київ, 2005, 2008, 2009, 2010 рр.); 14-й, 15-й, 16-й, 18-й Міжнародних науково-практичних конференціях «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я. MicroCAD» (Харків, 2006, 2007, 2008, 2010 рр.); 1-й, 2-й Міжнародних наукових конференціях «Сучасні інформаційні системи. Проблеми й тенденції розвитку» (Харків-Туапсе, 2006, 2007 рр.); міжнародній науковій конференції «Складні системи керування і менеджмент якості» (Старий Оскол, Росія, 2007 р.); науково-практичній конференції «Інформатизація вищих навчальних закладів МВС України» (Харків, 2008 р.); 8-й, 9-й Міжнародних науково-технічних конференціях «Проблеми інформатики і моделювання» (Харків, 2008, 2009 рр.); 3-й, 4-й Всеукраїнських науково-практичних конференціях «Сучасні тенденції розвитку інформаційних технологій у науки, освіті і економіці» (Луганськ, 2009, 2010 рр.); 10-й Міжнародній науково-практичній конференції «Сучасні інформаційні і електронні технології» (Одеса, 2009 р.); 4-й Міжнародній науково-технічній конференції «Комп'ютерні науки та інформаційні технології. CSIT» (Львів, 2009 р.); 1-й Міжнародній науково-практичній конференції «Проблеми і перспективи розвитку ІТ-індустрії» (Харків, 2009 р.); 7-й Міжнародній науково-практичній конференції «Математичне і програмне забезпечення інтелектуальних систем» (Дніпропетровськ, 2009 р.); всеукраїнській науково-практичній конференції «Інформатика і системні науки. ІСН» (Полтава, 2010 р.); 1-й, 2-й Міжнародних науково-технічних конференціях «Інформаційні технології в навігації і управлінні: стан і перспектива розвитку» (Київ, 2010, 2011 рр.); 1-й науково-технічній конференції «Сучасні напрямки розвитку інформаційно-комунікаційних технологій і засобів управління» (Харків, 2010 р.).

Публікації. За результатами дисертаційних досліджень опубліковано 80 наукові праці: 40 стаття, з них 34 у наукових фахових виданнях України серед яких 13 статей без співавторів, і 6 у наукових не фахових виданнях України; 40

публікація матеріалів і тез доповідей на міжнародних науково-практичних конференціях.

Структура дисертації. Дисертація складається зі вступу, 6 розділів, висновків і списку використаних джерел. Повний обсяг дисертації складає 462 сторінки, що включає 64 рисунка (на 26 сторінках), 2 таблиці (на 1 сторінці), 5 додатків, оформлених у вигляді окремої книги (на 116 сторінках), список використаних джерел з 360 найменувань (на 34 сторінках).

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність обраної теми дисертації, сформульовано мету та задачі дослідження, визначено об'єкт, предмет і методи досліджень, визначені наукова новизна та практичне значення отриманих результатів, а також особистий внесок автора в роботах, виконаних у співавторстві, наведено відомості про апробацію результатів дисертації та кількість публікацій за темою дисертаційної роботи.

У першому розділі на базі вивчення літературних джерел проведено аналіз методологій побудови інтегрованих інформаційних систем, визначено, що інтеграція інформаційних ресурсів на рівні єдиного простору даних є стратегічним напрямком в галузі управління розподіленими системами БД. Визначено основні концепції побудови єдиного інформаційного простору, які стосуються опису та дослідження методів взаємодії ІС в межах єдиного інформаційного кола, для чого доцільно використовувати розподілені однорангові архітектури з проміжним програмним рівнем. Такий підхід дозволяє досягти необхідного рівня гнучкості, відкритості і продуктивності розподілених інформаційних систем. Розглянуто питання інтеграції даних, виділено основні види інтеграції: консолідація, розподілення та федералізація даних. Надано формальну модель федеративних даних як відображення множини інформаційних об'єктів в множину правил, що визначають відношення між елементами даних. Виділено властивості, при яких розв'язання задачі інтеграції не є тривіальною: властивість однозначності обмежень цілісності $I_i \cap I_j = \emptyset$ і властивість спільності БД $\neg \exists (o_j^i, \neg o_j^i) \in DB_i$, де $\{DB_1, \dots, DB_n\}$ – множина баз даних, що інтегруються, $\{I_1, \dots, I_n\}$ – множина відповідних обмежень, $DB_i = \{o_1^i, \dots, o_m^i\}$ – множина інформаційних об'єктів i -ї бази даних. Розглянуто труднощі побудови моделі ПрО, яка, в свою чергу, визначає властивості БД, показано, що неповнота, розмитість, невизначеність і суперечливість апріорних представлень про ПрО приводить до неточностей, а іноді і до помилок при побудові БД на перших етапах її проектування.

При поданні моделі ПрО, що лежить в основі ІС, використовується замкнуте різноманіття предикатів, як одномісних, так і предикатів іншої розмірності, що визначається верхньою межею їхньої обчислювальної складності.

Наведено огляд публікацій з методів побудови моделей ПрО, а також проектування ІС і розподілених БД. Аналіз проблем інтеграції гетерогенних інформаційних систем дозволяє зробити висновок, що загальна задача поділяється

на підзадачі: розробку моделі та методів визначення даних ПрО з урахуванням їх структурної неоднорідності; розробку моделі та методів підтримки цілісності структури БД при модифікації схеми даних; розробку методів підтримки цілісності стану БД при маніпулюванні даними; розробку методів і засобів управління інтегрованими даними.

На основі наведених висновків сформульовано загальну проблему дослідження: на основі аналізу моделей і засобів подання об'єктів предметної області розробити та обґрунтувати єдиний підхід до побудови, формалізації і опису інформаційних об'єктів, що мають нереляційну структуру, для моделювання та підтримки неоднорідних розподілених даних.

У другому розділі на основі змістовного аналізу можливих підходів до побудови моделі ПрО визначено місце концепції єдиного інформаційного простору як найбільш актуальної і перспективної.

Досліджено характеристики ПрО для побудови ІС із можливістю накопичувати і обробляти дані. Відзначено, що множина інформаційних об'єктів як образ ПрО може бути використана для визначення динаміки зміни властивостей її моделі. Розглянуто методи побудови моделі ПрО, що узгоджуються з формальними вимогами побудови БД. Опис ПрО включає такі основні компоненти: функції та задачі (процедури) обробки даних і їхні характеристики, інформаційні об'єкти і відношення між ними, характеристики інформаційних об'єктів і процедур обробки даних, відношення між інформаційними об'єктами та процедурами. Таким чином, модель ПрО представляється у вигляді кортежу (1)

$$M^{PrO} = \langle F, Z, O, X, Y, C, R, Q \rangle, \quad (1)$$

де $F = \{f_i | i = \overline{1, n}\}$ - множина функцій; $Z = \{z_i | i = \overline{1, m}\}$ - множина задач (процедур) обробки даних; $O = \{o_i | i = \overline{1, k}\}$ - множина інформаційних об'єктів; $X = \{X_i | i = \overline{1, p}\}$ - множина вхідних даних; $Y = \{Y_i | i = \overline{1, l}\}$ - множина вихідних даних; $C = \{c_i | i = \overline{1, q}\}$ - повна множина інформаційних об'єктів ПрО; $R = \{R_i | i = \overline{1, g}\}$ - множина відносин (взаємозв'язків) між інформаційними об'єктами; $Q = \{Q_i | i = \overline{1, v}\}$ - множина процедур обробки даних.

Модель специфікацій інформаційних вимог представляється у вигляді відношення (2)

$$M^{Cn} = \langle \alpha R \beta \rangle, \quad (2)$$

де α, β – структурні елементи ПрО з множин O, X, Y, R – відношення між елементами ПрО.

В роботі докладно досліджено методи проектування інтегрованих БД, визначено, що процес аналізу ПрО і побудови канонічної структури інтегрованої БД складається з таких етапів: аналіз ПрО і побудова зовнішніх моделей; формування узагальненої зовнішньої моделі; побудова канонічної структури інтегрованої ПрО.

Під зовнішньою моделлю розуміється формалізований опис множини характеристик і властивостей даних, процедур їхньої обробки та відношень між ними, які в сукупності становлять ПрО.

Структуру зовнішньої моделі визначають відношення між інформаційними об'єктами, що задані в специфікаціях вимог $\langle o_i^i R o_i^j \rangle$, де $o_i^i, o_i^j \in O_i$. Формально структуру зовнішньої моделі деякої k -ої ПрО можна описати за допомогою матриці семантичної суміжності $SM^k = \| sm_{ij}^k \|$, елементи якої $sm_{ij}^k = 1$, якщо існує відношення $\langle o_i^i R o_i^j \rangle$ і $sm_{ij}^k = 0$, у протилежному випадку.

Тип семантичного зв'язку між інформаційними елементами O_i задається матрицею $BM^k = \| bm_{ij}^k \|$, елементи якої приймають значення: $bm_{ij}^k = 11$, якщо $sm_{ij}^k = 1$ і тип відношення між інформаційними об'єктами 1:1; $bm_{ij}^k = 10$, якщо $sm_{ij}^k = 1$ і тип відношення 1: M ; $bm_{ij}^k = 01$, якщо $sm_{ij}^k = 1$ і тип відношення M :1; $bm_{ij}^k = 00$, якщо $sm_{ij}^k = 1$ і тип відношення M : M ; $bm_{ij}^k = 0$, якщо $sm_{ij}^k = 0$.

Важливим компонентом зовнішньої моделі k -ої ПрО є опис процедур обробки даних, що задається у вигляді двох множин: множини запитів $Q^k = \{ Q_i^k \}$ ($i = \overline{1, n}$), де Q_i^k – формалізований вигляд запиту (формула, правило, предикатне вираз тощо) і множини обмежень $L^k = \{ L_i^k \}$ ($i = \overline{1, m}$) на відновлення та модифікацію інформації.

Вид інформаційних об'єктів O задається в зовнішній моделі у вигляді векторів, обумовлених на множині задач Z . Елемент вектора запитів Q_i^k приймає одиничне значення, якщо задача z_j використовує i -й інформаційний об'єкт у запиті, в протилежному випадку $Q_i^k = 0$, і елемент вектора обмежень L_i^k приймає одиничне значення, якщо в j -му запиті використовується i -е обмеження при відновленні або модифікації, у протилежному випадку $L_i^k = 0$.

Для множини запитів узагальненої зовнішньої моделі задаються відповідні матриці $QM = \| qm_{ij} \|$. При цьому елементи матриць проіндексовані по осях множиною задач і множиною модифікацій, де елементи qm_{ij} приймають одиничне значення, якщо i -а задача виконує j -у модифікацію, у протилежному випадку $qm_{ij} = 0$.

Топологія розподіленої ІС задається у вигляді бінарної матриці $TM = \| tm_{ij} \|$, проіндексованої по осях множиною вузлів обчислювальної мережі. Елементи матриці приймають значення одиниці, якщо i -й вузол обчислювальної мережі пов'язаний з вузлом j , у протилежному випадку $tm_{ij} = 0$.

Таким чином, формалізований опис узагальненої зовнішньої моделі ПрО представляється у вигляді сукупності множин (Z, O_i, C, Q^k, L^k) , сукупності векторів (z_i, q_i^k, l_i^k) і сукупності матриць (SM^k, BM^k, QM^k, TM) , тобто модель можна представити у вигляді (3)

$$M^{Bn} = \langle (Z, O_i, C, Q^k, L^k), (z_i, q_i^k, l_i^k), (SM^k, BM^k, QM^k, TM) \rangle. \quad (3)$$

Таке представлення забезпечує необхідною інформацією етап проектування структури інтегрованої ПрО.

Розглянуті матричні моделі дозволяють підвищити ефективність і якість побудови структури ПрО, знизити витрати на її розробку. На підставі введених методів отримано методику аналізу ПрО і побудови її канонічної структури в рамках єдиного підходу до проектування інтегрованої БД. З іншого боку, матричні моделі мають строгу структуру, робота з якою обмежує дії над схемою даних. При розгляді схем даних з неоднорідної структурою такі моделі не забезпечують динамічного режиму формування віртуального інформаційного простору. Це, у свою чергу, знижує якісні характеристики інтегрованих систем і розподілених БД.

У третьому розділі розглянуто питання розробки та дослідження моделей і методів розробки математичних засобів опису властивостей інформаційних об'єктів відносно вимог до ІС і БД відповідної ПрО.

В силу того, що задачі, які розв'язуються в роботі, направлені на дослідження гетерогенних систем, необхідно побудувати строгу математичну послідовність перетворень щодо обробки неоднорідних даних.

Для приведення даних до єдиного виду необхідно виконати низки перетворень. Під перетворенням даних розуміється зміна семантичного типу даних, злиття або декомпозиція змінних відношення БД для формування необхідного представлення. Перетворення структури БД і самих даних повинні розглядатися як етапи єдиного процесу.

Для формалізації побудови моделі ПрО введено низку позначень та визначено їх властивості. При цьому основою для визначення елементів ПрО є задачі, які задають структурні компоненти БД.

Кожний елемент ПрО представляємо парою (4)

$$o_i = \langle A_{i_j}, \rho_j(A_{i_j}) \rangle, \quad (4)$$

де A_{i_j} – ім'я елемента даних ПрО, $\rho_j(A_{i_j})$ – множина значень ρ_j для заданого елемента з іменем A_{i_j} .

Якщо об'єкти утворюють деякий інформаційний простір ПрО, то сукупність взаємозв'язків відображає структуру фрагмента реального світу або стан ПрО.

З часом множина об'єктів O може змінюватися, а також можуть змінюватися їхні властивості та взаємозв'язки. При цьому одержувані нові стани вважаються однією й тією ж ПрО. Таким чином, ПрО розглядається як система, що складається з певної послідовності станів $\tau_0, \tau_1, \dots, \tau_t$, у яких перебувають зв'язні інформаційні об'єкти в моменти часу $0, 1, \dots, t$. У загальному вигляді стан ПрО, пов'язний за часом, можна представляти як множину можливих станів $T = \{\tau_0, \tau_1, \dots, \tau_t\}$.

Для аналізу семантичних властивостей ПрО введено правила, що дозволяють визначити взаємозв'язок між об'єктами ПрО. Задамо правила у вигляді логічної функції на множині об'єктів O (у позначенні " $o_j \downarrow o_i$ "). Правила визначають той факт, що існування одного набору об'єктів впливає (або не впливає) з існування іншого (читається як "існує – якщо").

В загальному вигляді правило представляє формулу як вираз вигляду (5)

$$\{o_1, \dots, o_m\} \leftarrow \{o_1, \dots, o_n\}, \quad (5)$$

яке означає, що будь-яка множина об'єктів може породжувати також будь-яку множину інших інформаційних об'єктів.

Множину правил вигляду (5) будемо називати правилами логічного існування або правилами існування (L – правила) у позначенні (6)

$$L = \{L_1, \dots, L_n\}, \quad (6)$$

де L_i - правило вигляду (5), для якого декларативний вигляд має вид $l_{1\dots m} \leftarrow l_{1\dots n}$.

Будемо розглядати два класи правил:

1. Функціональні правила L^f , у вигляді (7), які породжують об'єкти на підставі застосування деяких операцій (як правило, арифметичних) до існуючого набору інформаційних об'єктів.

$$L^f = (\{o_1, \dots, o_m\} \leftarrow f \{o_1, \dots, o_n\}), \quad (7)$$

де f - множина припустимих операцій та функцій.

2. Структурні правила L^r , у вигляді (8), що деталізують семантично складні об'єкти. Розрізняються два види деталізації:

- узагальнення: дозволяє співвіднести множину інформаційних об'єктів з одним загальним;
- декомпозиція: зворотна узагальненню дія, причому властивості, визначені для об'єктів-джерел, можуть не зберігатися при їхній декомпозиції для кожного окремого об'єкта.

$$\forall L^r = (\{o_k \leftarrow o_1, \dots, o_n\} \exists o_p \mid o_p \notin O \wedge o_p \leftarrow o_k). \quad (8)$$

Для детального аналізу та визначення ступеня схожості інформаційних просторів розглянутих ІС введено поняття межі ПрО, у позначенні (9), як множини інформаційних об'єктів, що визначають поточний стан ПрО, отриманий на підставі заданої множини O із застосуванням правил L (включаючи як L^f , так і L^r) і базису, у позначенні (10), як елементи множини O , для яких має місце правило виду $\{o_i\} \leftarrow \emptyset$. Таким чином, модель ПрО можна представити у вигляді (11)

$$C = \langle O, L^f(\{o_j \leftarrow o_i\}), L^r(\{o_j \leftarrow o_i\}) \rangle, \quad (9)$$

де $o_i, o_j \in O$.

$$B = \{o_1, \dots, o_p\} \subseteq O = \{o_1, \dots, o_n\}, \quad (10)$$

де $p \leq n$.

$$M^{PrO} = \langle B, L, \{Ci\} \rangle. \quad (11)$$

Для порівняння семантики предметних областей необхідно провести аналіз декількох (можливо, різних) множин O на предмет їхньої відповідності правилам $L = L_1 \cup \dots \cup L_n$, отриманих у результаті об'єднання правил із всіх розглянутих ПрО.

Очевидно, що один стан може задовольняти такій множині L , а інший – ні. У загальному випадку не завжди ясно, що є первинним: множина припустимих станів ПрО, що визначає L – правила, або L – правила накладають обмеження на стани ПрО. Таким чином, потрібно виявити сімейство L – правил, якому задовольняють всі припустимі стани ПрО.

Введено аксіоми:

L1. Псевдорефлексивність:

$$[l_n] \leftarrow [l_n, l_m] \text{ (або } [l_m] \leftarrow [l_n, l_m]).$$

L2. Розширення:

$$\text{якщо виконується } [l_n] \leftarrow [l_m], \text{ то виконується } [l_n, l_m] \leftarrow [l_m].$$

L3. Згортання:

$$\text{якщо виконується } [l_n] \leftarrow [l_m] \text{ і } [l_p] \leftarrow [l_m], \text{ то виконується } [l_n, l_p] \leftarrow [l_m].$$

L4. Декомпозиція:

$$\text{якщо виконується } [l_n, l_m] \leftarrow [l_k], \text{ то виконується } [l_m] \leftarrow [l_k] \text{ (і, відповідно, } [l_n] \leftarrow [l_k]). \text{ Як окремий випадок, можна затверджувати, якщо } [l_p] \leftarrow [l_k] \text{ і } [l_p] \subseteq [l_m], \text{ то } [l_m] \leftarrow [l_k].$$

L5. Псевдотранзитивність:

$$\text{якщо виконується } [l_n] \leftarrow [l_m] \text{ і } [l_w] \leftarrow [l_p, l_n], \text{ то виконується } [l_w] \leftarrow [l_p, l_m],$$

де $[l_n]$ - елемент вигляду $l_{1\dots n}$, а $[l_n, l_m]$ – елемент вигляду $l_{1\dots n} \cup l_{1\dots m}$.

Лема 1. Система аксіом L1 – L5 є надійною. Тобто, якщо правило $l_y \leftarrow l_x$ виведене з L з використанням аксіом L1 – L5, то воно справедливе для будь-якого стану ПрО, у якій справедливі правила з L .

Введена аксіоматика дозволяє побудувати послідовність вигляду (12)

$$\Pi = \left| \Rightarrow L_1 \right| \Rightarrow \dots \left| \Rightarrow L_n, \quad (12)$$

де кожне правило належить заданій множині правил L або виводиться за допомогою аксіом з попередніх правил.

Множину правил, яку можна вивести із заданої множини L , будемо називати замиканням L , у позначенні L^+ . З множини L – правил можна вивести правило $l_y \leftarrow l_x$, якщо це правило належить до L^+ .

Очевидно, що в загальному випадку обчислення L^+ для множини правил L є трудомісткою задачею тому, що множина L^+ має кількість елементів 2^n , де n – кількість елементів L .

Нехай L – множина правил на множині об'єктів O і нехай $l_{1\dots x} \subseteq O$. Тоді $l_{1\dots x}^+$, замикання $l_{1\dots x}$ відносно L , є множина об'єктів $\{l_i\}$, таких, що правило $l_i \leftarrow l_{1\dots x}$ може бути виведене з L по аксіомах L1 – L5.

Основна властивість $l_{1\dots x}^+$ полягає в тому, що замикання дозволяє визначити, чи витікає деяке правило $l_{1\dots y} \leftarrow l_{1\dots x}$ з L по аксіомах L1 – L5. Наступна лема підтверджує цей факт.

Лема 2. Правило вигляду $l_{1\dots y} \leftarrow l_{1\dots x}$ витікає з аксіом $L1 - L5$ тоді і тільки тоді, коли $l_{1\dots y} \subseteq l_{1\dots x}^+$.

Цей факт дозволяє розв'язати так звану проблему членства, тобто для визначення, чи породжує множина L правило $l_{1\dots y} \leftarrow l_{1\dots x}$, досить перевірити, чи належить $l_{1\dots y} \leftarrow l_{1\dots x}$ до L^+ без побудови самого L^+ .

Пропонований підхід до опису властивостей елементів ПрО з використанням L -правил є математичним фундаментом, на якому будуються оцінки порівняності ІС у цілому і БД зокрема. При цьому аксіоми виведення дозволяють маніпулювання множинами L -правил без втрати їхньої еквівалентності.

Визначення 1. Нехай L і \hat{L} - дві множини L -правил. Будемо говорити, що L і \hat{L} еквівалентно, у позначенні $L \equiv \hat{L}$, якщо $L^+ = \hat{L}^+$.

Крім цього, в роботі розв'язно задачу побудови мінімальної (ненадлишкової) еквівалентної множини правил.

Визначення 2. Множина L -правил є ненадлишковою, якщо не існує такого L' , що $L' \subseteq L$ і $L' \equiv L$.

Розроблені в третьому розділі методи та засоби визначають інформаційну технологію приведення неоднорідних структур даних до єдиного семантичного виду. Технологія включає послідовне перетворення множини інформаційних об'єктів деякої ПрО до множини інформаційних об'єктів, що відповідає іншій ПрО, або буде перетин предметних областей відносно спільних інформаційних об'єктів.

Показано, що запропонована теорія L -правил є розв'язуваною, несуперечливою та незалежною.

В четвертому розділі розроблено та досліджено методи підтримки цілісності структур БД при модифікації схеми даних. При цьому необхідно розв'язати задачу побудови еквівалентної множини обмежень цілісності при модифікації структури БД як основи визначення семантичних властивостей ПрО.

При моделюванні ПрО існують деякі аспекти дійсності, для адекватного подання яких розглянутих раніше інструментальних можливостей недостатньо. Для розв'язання цієї задачі додаткова інформація про об'єкти реального світу представлена у вигляді певних обмежень, які зв'язані зі структурними БД, що відповідають опису конкретних інформаційних об'єктів і зв'язків.

Введено поняття семантики БД як множини інформаційних об'єктів S , що відповідають виразу (13)

$$S = \langle O, I(\{o_i\}) \rangle, i = \overline{1, n}, \quad (13)$$

де I – множина правил, яка визначає спільні інформаційні об'єкти або обмеження цілісності ПрО.

Основна концепція пропонованого підходу полягає в тому, що і БД, і обмеження цілісності повинні буди спільними. Виходячи із цього, можна зробити висновок про те, що семантика S може виявитися неспільною, навіть якщо O і L спільні. В подальшому під модифікацією БД будемо розуміти додавання або видалення інформаційного об'єкта, при виконанні якого множина O залишається спільною.

Проведено подальше дослідження семантики БД і визначено, що при модифікації структури БД кількість семантик може бути більше однієї. Таким чином, необхідно виявити семантику, яка має найменшу кількість модифікованих інформаційних об'єктів, для чого в роботі використовується симетрична різниця семантик $S \setminus S' = (S - S') \cup (S' - S)$, де S – вихідна семантика, S' – модифікована семантика.

При побудові S' необхідно розв'язати задачу виведення множини інформаційних об'єктів, відповідних S' . Нехай I – фіксована множина обмежень, $I \subseteq L$. Через $O^I(L)$ позначимо множину всіх екземплярів правил з L , в яких всі складені об'єкти замінені інформаційними об'єктами з I . Зв'яжемо з I семантичний оператор η^I , що визначимо як вираз (14)

$$\eta^I(O) = \{l \mid l \leftarrow l_{1,\dots,n} \in L \text{ и } l_{1,\dots,n} \in O^I(L)\} \cup I. \quad (14)$$

Оскільки η^I – монотонний оператор по відношенню включення множин (якщо $I_1 \subseteq I_2$, то $\eta^{I_1} \subseteq \eta^{I_2}$), то послідовність (15) має границю

$$\eta_0(O) = O, \eta_1(O) = \eta(\eta_0(O)) \dots \eta_n(O) = \eta(\eta_{n-1}(O)), \forall n > 0. \quad (15)$$

Позначимо відповідну границю як η^O і назовемо найменшою нерухомою точкою η відносно O , тобто $\eta^O(O) = O$.

Визначення 3. Семантика множини інформаційних об'єктів S визначається як найменша нерухома точка η відносно S , у позначенні η^S . Базу даних будемо називати спільною, якщо її семантика спільна.

Визначення 4. Множина обмежень I є спільною, якщо для всякого інформаційного об'єкта $l \in O$ існує множина інформаційних об'єктів O' така, що $\eta^{O' \cup \{l\}}$ є спільною множиною.

В конкретний момент часу семантика БД є виділеною сукупністю інформаційних об'єктів і зв'язків, виражених правилами логічного існування та визначаючих стан ПрО. Якщо інформаційні об'єкти утворюють множину припустимих значень, то сукупність правил відображає структуру фрагмента дійсності. З часом одні об'єкти зникають, інші з'являються, змінюються їхні властивості і взаємозв'язки. Проте, отримані нові стани вважаються станами однієї і тієї ж ПрО.

Таким чином, подалі семантика розглядається як система, що складається з певної послідовності станів, що залежать від модифікацій (як множини інформаційних об'єктів, так і множини їх властивостей).

Для того, щоб модифікація була припустима, семантика S' модифікованої БД повинна задовольняти таким умовам:

1. Множина S' повинна бути спільна.
2. Множина S' повинна включати l при $ins(l)$ і не повинна включати l у випадку $del(l)$.
3. Множина S' повинна бути дедуктивно замкнута відносно I .
4. По змісту множина S' визначається набором інформаційних об'єктів, які були в S (і доданих об'єктів l у випадку $ins(l)$).

5. Семантика S' повинна бути отримана за допомогою мінімальної зміни семантики S вихідної БД.

Визначення 5. Нехай l – інформаційний об'єкт і $DB = \langle O, I \rangle$ – спільна БД із семантикою S . Множину інформаційних об'єктів S^{ins} будемо називати семантикою з доданим елементом, а S^{del} будемо називати семантикою з вилюченим елементом для l і O .

Твердження 1. Для якої-небудь спільної БД і для будь-якого додавання або видалення інформаційного об'єкта l існує хоча б одна S^{ins} і S^{del} семантика.

С другого боку, очевидно, що не будь-які зміни в множинах інформаційних об'єктів БД і не будь-які правила можуть забезпечити спільність, що в підсумку приведе до протиріччя в БД, що інтегруються.

Надалі розглянуто задачу пошуку таких обмежень I , для яких існує єдина семантика S^{ins} і/або S^{del} при модифікації бази даних.

Визначення 6. Множина обмежень I називається однозначною щодо додавання (видалення), якщо для будь-якої спільної бази даних DB і для будь-якого інформаційного об'єкта l існує єдина S^{ins} (S^{del}) семантика.

Для забезпечення ефективності функціонування системи інтегрованих БД розглядаються додаткові властивості однозначності обмежень щодо додавання і/або видалення.

Визначення 7. Нехай I – множина спільних обмежень, тоді $SAT_{DB}(I)$ – множина інформаційних об'єктів всіх баз даних DB , що задовольняють всім обмеженням з I . (Якщо DB мається на увазі неявно, замість $SAT_{DB}(I)$ будемо використовувати позначення $SAT(I)$).

Відповідно введеним позначенням еквівалентність обмежень можна охарактеризувати такими умовами:

1. $I_1 \mid \Rightarrow I_2$ ($I_2 \mid \Rightarrow I_1$) тоді, коли $SAT(I_1) \subseteq SAT(I_2)$ ($SAT(I_2) \subseteq SAT(I_1)$).

2. $I_1 \equiv I_2$ тоді, коли $SAT(I_1) = SAT(I_2)$.

Визначення 8. Нехай I_1 і I_2 – множини обмежень для баз даних DB_1 і DB_2 (для DB_1 і DB_2 множина O та сама), яким відповідають семантичні оператори η' і η'' . Нехай також P_1 і P_2 – інформаційні простори баз даних DB_1 і DB_2 , відповідно, таких, що $P_1 \subseteq P_2$. Будемо говорити, що I_1 зводиться до I_2 , у позначенні $I_1 \sim I_2$, якщо $SAT(I_1) = SAT(I_2) \pm P_1$ (де $\pm P$ – операція модифікації множини $SAT(I)$).

Показано, що множина O спільна відносно I_1 , якщо вона спільна відносно I_2 , і її семантика відносно I_1 може бути отримана модифікацією на елементи з P_1 її семантики відносно P_2 .

Сформульовано та доведено твердження, що визначає властивості спільності БД при її модифікації. Для цього позначено через $\mathbf{S}^{ins} = \{ \eta^{\varpi \cup \{l\}} \mid \max(\varpi) \text{ при } \varpi \subseteq O \text{ і } \eta^{\varpi \cup \{l\}} \text{ – спільно} \}$ і $\mathbf{S}^{del} = \{ \eta^{\varpi} \mid \max(\varpi) \text{ при } \varpi \subseteq O \text{ і } l \notin \eta^{\varpi} \}$, де $\varpi = S_1^{ins} \cap O$ для \mathbf{S}^{ins} і $\varpi = S_1^{ins} \cap O$ для \mathbf{S}^{del} . Для різниці семантики відносно правил їхнього формування записуватимемо $\mathbf{S}^{ins}(O, I_1)$ або $\mathbf{S}^{ins}(O, I_2)$, а також $\mathbf{S}^{del}(O, I_1)$ або $\mathbf{S}^{del}(O, I_2)$.

Для підтвердження того, що множини обмежень I_1 і I_2 поведуться однаково стосовно модифікацій БД, сформульовано та доведено теорему.

Теорема 1. Нехай I_1 і I_2 – обмеження, що зводяться ($I_1 \sim I_2$). Тоді для будь-якої

множині інформаційних об'єктів O , для якої база даних $DB = \langle O, I_1 \rangle$ спільна, виконуються такі рівняння:

1. $\mathbf{S}^{ins}(O, I_1) = \mathbf{S}^{ins}(O, I_2) \pm P_1$;
2. $\mathbf{S}^{del}(O, I_1) = \mathbf{S}^{del}(O, I_2) \pm P_1$.

Показано, що для досягнення еквівалентності однозначної і простої множини обмежень (просте правило має вигляд $l_i \leftarrow l_{1\dots n}$) досить виконати деяке розширення набору інформаційних об'єктів. При зведенні обмежень нові інформаційні об'єкти можна вивести по L – правилам, що входять в I_2 .

Теорема 2. Для будь-якої однозначної множини обмежень I_1 існує проста множина обмежень I_2 і при цьому $I_1 \sim I_2$.

Таким чином, розв'язувані в розділі задачі визначають інформаційну технологію підтримки структури даних при переході БД до еквівалентного виду. Розроблена технологія складається з послідовності методів модифікації обмежень, що описують структурні характеристики вихідного стану БД.

В п'ятому розділі пропонуються нові і розвиваються існуючі методи підтримки цілісності даних при реалізації запитів до БД.

При побудові інтегрованої ІС обмеження цілісності локальних БД можуть не збігатися, при цьому виконання запитів не повинне порушувати ці обмеження. При виконанні запитів, які порушують обмеження, необхідно розв'язати задачу, з одного боку коректних, а з іншого боку - мінімальних змін даних, так, щоб запит завершився успішно.

В результаті виконання запиту початковому стану τ_0 і зовнішньому запиту на модифікацію потрібно так мінімально змінити τ_0 на новий стан τ_1 , щоб виконати запит і забезпечити коректність обмежень I . Тобто, необхідно забезпечити коректне виконання відображення $\delta: \tau_0 \rightarrow \tau_1$.

Нехай O' – множина всіх інформаційних об'єктів поточного стану τ_0 і O'' – множина всіх інформаційних об'єктів нового стану τ_1 . Тоді для відображення δ і деякого інформаційного об'єкта $l_o \in O'$ будемо говорити, що l_o залишається в стані τ_1 (тобто виводиться при переході від τ_0 до τ_1 , у позначенні $\delta \mid \Rightarrow l_o$), якщо $l_o \in \tau_0$ при $l_o \in O'$ і $l_o \in \tau_1$ при $l_o \in O''$. При цьому обмеження цілісності $I_i \in I$ вигляду $l_i \leftarrow l_{1\dots n}$ виконується при відображенні δ , якщо $\delta \mid \Rightarrow l$, коли $\delta \mid \Rightarrow l_i$ ($i = \overline{1, n}$). Відображення δ задовольняє множині обмежень I , якщо кожне правило з I виконується на δ , у позначенні $\delta \mid \Rightarrow I$.

Під модифікацією стану БД розуміємо трійку $\Lambda = \langle O^{ins}, O^{del}, O^{upd} \rangle$, де $O^{ins} \subseteq O'$, $O^{del} \subseteq O'$, $O^{upd} \subseteq O' \times O'$. Позначимо через Λ^{ins} , Λ^{del} і Λ^{upd} елементи з множин O^{ins} , O^{del} і O^{upd} відповідно. При цьому $\Lambda^{upd} \Leftrightarrow (\Lambda_{ins}^{upd} = \{l'_i \mid \exists l''_i ((l'_i, l''_i) \in \Lambda^{upd})\})$ і $\Lambda_{del}^{upd} = \{l''_i \mid \exists l'_i ((l'_i, l''_i) \in \Lambda^{upd})\}$.

Визначення 9. Модифікація Λ виконується при відображенні $\delta: \tau_0 \rightarrow \tau_1$, якщо $\Lambda^{ins} \in \tau_1$, $\Lambda^{del} \cap \tau_1 = \emptyset$ і для будь-якої пари інформаційних об'єктів $(l'_i, l''_i) \in \Lambda^{upd}$, якщо $l'_i \in \tau_0$, то $l'_i \notin \tau_1$, а $l''_i \in \tau_1$.

У загальному випадку запити на модифікацію можуть виявитися неспільними з обмеженнями цілісності, тобто може бути порушена істинність правил з I . Для розв'язання цієї проблеми в роботі визначені критерії спільності модифікацій Λ і множини обмежень I .

Для обмежень I множина всіх відображень $\delta: \tau_0 \rightarrow \tau_1$, на яких виконана модифікація Λ , позначається як $\delta^\tau(I, \Lambda)$. Для вихідного стану τ_0 нові стани представимо множиною $\dot{\tau}(I, \Lambda) = \{\tau_i \mid (\tau_0 \rightarrow \tau_i) \in \delta^\tau(I, \Lambda)\}$.

Визначення 10. Модифікація Λ спільна з обмеженнями I , якщо а) $\delta^\tau(I, \Lambda) \neq \emptyset$ і б) $\dot{\tau}(I, \Lambda) \neq \emptyset$.

Це визначення означає, що множина $\delta^\tau(I, \Lambda)$ містить хоча б одне відображення $\delta: \tau_0 \rightarrow \tau_1$, і є хоча б один стан τ_1 , такий, що $(\tau_0 \rightarrow \tau_1) \in \delta^\tau(I, \Lambda)$.

Досліджуючи питання зміни даних, в роботі визначено умови коректності станів БД при виконанні запитів на модифікацію даних. Подальші викладки ґрунтуються на принципах збереження інформації. Тобто, при виконанні змін у БД необхідно максимально зберегти існуючі дані, а потім додати нові.

Визначення 11. Нехай τ_0, τ_1, τ_2 , – деякі стани БД. Стан τ_1 “ближче” (в розумінні мінімальності змін елементів множини τ_1) до стану τ_0 , чим τ_2 , якщо виконуються умови:

- 1) $\tau_0 \cap \tau_2 \subset \tau_0 \cap \tau_1$, або
- 2) $(\tau_0 \cap \tau_2 = \tau_0 \cap \tau_1)$ і $(\tau_1 - \tau_0) \subset (\tau_2 - \tau_0)$.

Два стани τ_0 і τ_1 називаються суміжними, якщо немає стану τ_2 , що перебуває до τ_0 ближче, чим τ_1 .

Визначено дії, які по вихідному стані БД τ_0 і запиту на відновлення ставлять у відповідність новий стан τ_1 , що задовольняє обмеженням цілісності I і є найближчим до вихідного стану τ_0 .

Визначення 12. Нехай відновлення Λ і обмеження I спільні. Оператор $\mathfrak{G}^{(I, \Lambda)}$ на множині станів T називається оператором модифікації, якщо для кожного стану τ_i ($i = \overline{1, n}$), такого, що відновлення Λ разом з I , виконуються умови:

- 1) $\mathfrak{G}^{(I, \Lambda)} \in \dot{\tau}(I, \Lambda)$;
- 2) після застосування $\mathfrak{G}^{(I, \Lambda)}$ стан τ_1 і початковий стан τ_0 є суміжними по відношенню $\dot{\tau}(I, \Lambda)$.

Змістовно перша умова визначає той факт, що запит повинен бути виконаний і отриманий стан повинний задовольняти обмеженням цілісності. Друга умова фіксує безпосередньо модифікацію, тобто виконуються тільки необхідні (мінімальні) зміни в початковому стані, щоб суміжний стан задовольняв першій умові.

Для визначення властивості повноти оператора модифікацій для неоднорідних БД сформульовано і доведено теорему.

Теорема 3. Нехай $\mathfrak{G}^{(I, \Lambda)}$ – оператор модифікації і нехай τ_0 і τ_1 – стани БД, такі, що $(\tau_0, \tau_1) \mid \Rightarrow I$. Тоді існує відновлення Λ' , спільне з I , таке, що $\tau_1 = \mathfrak{G}^{(I, \Lambda')}(\tau_0)$.

З цього твердження випливає, що для заданої множини обмежень I в деякому стані τ_0 область значень оператора модифікації $\mathfrak{G}^{(I, \Lambda)}$ при відображенні $\delta: \tau_0 \rightarrow \tau_1$ збігається з множиною інформаційних об'єктів стану τ_1 .

Для побудови алгоритмів пошуку елементів з $T^{I,\Lambda}(\tau_0)$ і побудови всіх можливих станів неоднорідних БД при модифікації даних визначено властивості обмежень цілісності у вигляді такого твердження.

Лема 3. Для неоднорідних БД і для будь-яких обмежень цілісності виконуються такі умови:

- 1) якщо для деякого відображення $\hat{\delta}: \hat{\tau}_0 \rightarrow \hat{\tau}_1$ існує замикання η^{+I} для станів $(\hat{\tau}_0 \rightarrow \hat{\tau}_1)$, то $\eta^{+I}(\hat{\tau}_0 \rightarrow \hat{\tau}_1) \mid \Rightarrow I$;
- 2) якщо існує відображення $\delta: \tau_0 \rightarrow \tau_1$ і $(\tau_0 \rightarrow \tau_1) \mid \Rightarrow I$, то для інформаційних об'єктів переходу $(\hat{\tau}_0 \rightarrow \hat{\tau}_1) \subseteq (\tau_0 \rightarrow \tau_1)$ справедливо $\eta^I(\hat{\tau}_0 \rightarrow \hat{\tau}_1) \subseteq (\tau_0 \rightarrow \tau_1)$ і $\eta^I(\hat{\tau}_0 \rightarrow \hat{\tau}_1) \mid \Rightarrow I$;
- 3) якщо модифіковані інформаційні об'єкти Λ спільні з обмеженнями I для деякого стану БД τ_0 і $\eta^{I(\Lambda)}(\tau_0 \rightarrow \emptyset) = (\tau_0 \rightarrow \tau_1)$, то τ_1 є станом з найменшою кількістю змін серед $\dot{\tau}(I, \Lambda)$;
- 4) якщо стан $\tau' \in T^{I,\Lambda}(\tau_0)$, то $(\tau_0 \rightarrow \tau') = \eta^{I(\Lambda)}(\tau_0 \rightarrow \tau_0 \cap \tau')$.

Показано, що за допомогою розроблених алгоритмів для будь-якого стану τ_0 , представленого впорядкованою множиною інформаційних об'єктів $\{l_{o_1}, \dots, l_{o_m}\} \subseteq \tau^\lambda$, можна побудувати стан $\tau_1 \in T^{I,\Lambda}(\tau_0)$.

Використовуючи розроблені методи підтримки цілісності даних далі в роботі розглядаються формальні засоби обробки даних з неоднорідною структурою. Досліджено низку підходів до сумісного доступу к даним розподілених ІС.

Одержали подальший розвиток методи використання реляційного числення зі змінними на доменах. Ґрунтуючись на введеній специфікації побудови моделі ПрО, розширено сигнатуру мови предикатів для формування виразів числення введенням логічного символу існування та функціонального квантору. Визначені їх відповідні властивості.

Якщо таблиця R має арність n зі змінними $l_{o_1}^1, l_{o_2}^1, \dots, l_{o_n}^1$, то вираз числення зі змінними на доменах буде мати вигляд $\{l_{o_1}^1, l_{o_2}^1, \dots, l_{o_n}^1 \mid \mathcal{P}(l_{o_1}^1, l_{o_2}^1, \dots, l_{o_n}^1)\}$, де \mathcal{P} - предикат, у якому будь-який атом має вигляд $l_{o_1}^1 = l_{o_1}^2 \wedge l_{o_2}^1 = l_{o_2}^2 \wedge \dots \wedge l_{o_n}^1 = l_{o_n}^2 \wedge l_{o_{n+1}}^1 = a_1^i$.

У запропонованій нотації до таблиць мережевої структури можуть бути сформульовані запити у вигляді виразів числення зі змінними на доменах (16), де в якості доменів виступають елементи даних таблиці

$$\left\{ l_{o_1}^1, l_{o_2}^1, \dots, l_{o_n}^1 \mid \exists (l_{o_1}^2, l_{o_2}^2, \dots, l_{o_n}^2) \left(R(l_{o_1}^2, l_{o_2}^2, \dots, l_{o_n}^2) \wedge l_{o_1}^1 = l_{o_1}^2 \wedge \dots \wedge l_{o_n}^1 = l_{o_n}^2 \right) \wedge \right. \\ \left. \neg \exists (l_{o_1}^3, l_{o_2}^3, \dots, l_{o_n}^3) \left(S(l_{o_1}^3, l_{o_2}^3, \dots, l_{o_n}^3) \wedge l_{o_1}^2 = l_{o_1}^3 \wedge \dots \wedge l_{o_n}^2 = l_{o_n}^3 \right) \right\} \quad (16)$$

Визначено характеристики компонентів обмежень цілісності для формування формул числення. Крім традиційних обмежень на значення і ключ, у роботі визначені обмеження на залежність (17) і рекурсивне обмеження на посилання (18)

$$\begin{aligned} \mathcal{FD}(R, l_{o_i}^1, l_{o_i}^2) &\equiv \forall l_{o_i}^1 (R(l_{o_i}^1) \supset \neg \exists l_{o_i}^2 (R(l_{o_i}^2) \wedge \langle l_{o_i}^1, D_1(a_i) \rangle = \\ &= \langle l_{o_i}^1, D_2(b_i) \rangle \wedge \langle l_{o_i}^2, D_1(a_i) \rangle \neq \langle l_{o_i}^2, D_2(b_i) \rangle)), \end{aligned} \quad (17)$$

$$\begin{aligned} \mathcal{JN}(R_1, R_2, l_{o_i}^i) &\equiv \exists R_1, \exists R_2, \forall l_{o_i}^i (R_1, \mathcal{FK}(R_1, R_2, l_{o_i}^i), R_2) \supset \\ &\supset \mathcal{JN}(R_1, R_2, l_{o_i}^i), \end{aligned} \quad (18)$$

де \mathcal{FK} – предикат обмеження зовнішнього ключа.

Таким чином, при побудові єдиного інформаційного простору узгодженість даних встановлюється за допомогою еквівалентних перетворень залежностей, що, в свою чергу, забезпечує коректність переходу станів баз даних.

У шостому розділі досліджено комплекс логічних засобів, орієнтований на розробку прикладних програм, незалежних від СУБД при одночасній взаємодії з декількома, можливо, неоднорідними БД.

Оскільки введені L -правила описують Про на рівні відношень між інформаційними об'єктами БД, то при реалізації запитів до неоднорідних структур даних використовується логічне програмування.

Прикладним інструментом для реалізації логічних запитів є компілятор дедуктивної бази даних (ДБД) як засіб трансляції логічної мови обробки даних. Модель ДБД є трійкою (19)

$$M^{DDB} = \langle EDB, IDB, I \rangle, \quad (19)$$

де EDB - екстенсіональна база даних (ЕБД): відповідні відношення зберігаються в БД, IDB - інтенсіональна база даних (ІБД): відношення обчислюються за допомогою застосування одного або декількох правил, I - множина обмежень цілісності.

Для побудови запитів до ДБД реалізована редукція операцій реляційної алгебри до правил дедуктивного виведення. Так, для запиту алгебраїчного виду (20) логічне правило буде мати вигляд (21)

$$R = \pi_{l_{o_1}, l_{o_2}} (\sigma_{l_{o_3} \leq N} (R(l_{o_1}, l_{o_2}, l_{o_3}, l_{o_4}, l_{o_5}))), \quad (20)$$

$$\mathcal{P}(x_1, x_2) \leftarrow R(x_1, x_2, x_3, x_4, x_5) \wedge x_3 \leq N, \quad (21)$$

де $(x_1, x_2, x_3, x_4, x_5)$ – предикатним змінним відповідають інформаційні об'єкти $(l_{o_1}, l_{o_2}, l_{o_3}, l_{o_4}, l_{o_5})$.

Для визначення семантики БД використовується поняття найменшої нерухомої точки η як монотонного оператора, застосовуваного до множини інформаційних об'єктів БД і перевіряючих їх коректність щодо множини обмежень цілісності I .

Застосування оператора найменшої нерухомої точки до рівняння, що виражає відношення R , можна сформулювати у вигляді виразу (22)

$$\eta(R = \pi_{A,B}(R \triangleright \triangleleft S) \cup \pi_{A,B}(S)), \quad (22)$$

де A і B – атрибути відношень R і S зі значеннями $\{x_1, x_2, x_3, \dots\}$.

Як відомо, засобами реляційної алгебри не можна здійснити об'єднання нескінченної послідовності виразів, що повертають множину i -х ($i = \overline{1, \infty}$) продовжень.

У контексті мови логічного програмування такі питання можуть бути розв'язані. Використовуючи один и той же предикат, як у заголовку, так і в тілі правила, можна виразити багаторазову операцію об'єднання.

На цій підставі вираз (22) записується у вигляді двох правил (23)

$$R(x_1, x_2) \leftarrow S(x_1, x_2); \quad R(x_1, x_2) \leftarrow S(x_1, x_3) \text{ AND } R(x_3, x_2). \quad (23)$$

Сформульовано твердження, що дозволяє спростити рекурсивні правила для обчислення результату запиту.

Твердження 2. Нехай задано деякий інтенціонал БД, тоді при обчисленні запиту нові кортежі, що підлягають включенню в результуюче відношення, можуть бути отримані тільки в результаті застосування правил, у яких хоча б одна з підцілей зводиться до кортежу, доданому у відношення на попередній ітерації.

Досліджувані задачі обробки даних розглядаються в контексті табличних моделей, які можуть мати деякий ступінь неоднорідності, тому необхідно знати, як зіставити елементи реляційної і дедуктивної моделі для коректної обробки даних.

Відображення реляційної моделі даних M в модель даних M^{DDB} , в позначенні $\xi : M \rightarrow M^{DDB}$, будемо представляти сукупністю відображень:

- $\varphi : \tau \rightarrow EDB$ – припустимі стани БД моделі M в припустимі стани БД моделі M^{DDB} ;
- $\psi : Sch \rightarrow IDB$ – схем БД моделі M в схеми БД моделі M^{DDB} ;
- $\nu : \theta \rightarrow P$ – операційні специфікації моделі M в послідовність виведення (програму) P .

Для коректного відображення моделі M в модель M^{DDB} сформульовано відповідні критерії.

Дослідження відображень розглянутих моделей визначило деякі труднощі, які виникають при побудові ξ , зокрема, при відображенні операційної специфікації. Для подолання цих труднощів в роботі операційна специфікація реляційної моделі розглядається як об'єднання операцій $\theta = \theta^{Sch} \cup \theta^\tau$, де $\theta^{Sch} : Sch \rightarrow \tau$ $O^i: S^i \rightarrow V^i$ – функція визначення даних в M , θ^τ – функція маніпулювання даними в M .

В дедуктивній моделі поділ на компоненти типу ‘значення’, ‘схема’, ‘запит’ виконується умовно тому, що при побудові послідовності виводу для організації доступу до даних використовуються всі ці компоненти. Таким чином, функція відображення схеми в стан в явному вигляді не задається.

Для розв'язання загальної задачі побудови переходу стану БД при відображенні моделей розглядається часткове відображення без урахування функції ν . Для цього будується модель M' , для якої функції мають вигляд $\xi' : M \rightarrow M'$:

- $\varphi' : \tau \rightarrow EDB'$ – простір станів БД M в простір станів БД моделі M' ;
- $\psi' : Sch \rightarrow IDB \cup A'$ – схем БД M – в схему БД моделі M' , де A' – множина імен елементів даних, виведених на підставі запропонованих аксіом.

На підставі розроблених методів у роботі показано, що модель даних M інтегрується з моделлю M^{DDb} , якщо існує розширення M' моделі M^{DDb} , що забезпечує комутативне відображення $\xi' : M \rightarrow M'$.

Розглянутий підхід до побудови відображення моделей є основою для розробки інформаційної технології приведення різних моделей до виду, що задовольняє критеріям інтеграції БД.

На основі результатів, одержаних в попередніх розділах, досліджено низку практичних задач інтеграції розподілених БД.

У відділі АСУП ДП “Завод імені В.О. Малишева” удосконалено систему обробки розподілених даних за рахунок впровадження методів побудови динамічних запитів дає можливість реалізувати транзакції за принципом вільного розподілення черги. Для організації доступу к даним використовуються принципи формування запитів за допомогою виразів числення зі змінними на доменах, що дозволяє згладити структурні неоднорідності схем розподілених даних.

При таких діях все навантаження по обробці запиту та перетворенню результату покладається на сервер БД, на якому формується запит, що дозволяє виконувати структурні зміни даних без додаткових перетворень глобальної схеми інтегрованої системи БД.

Використані методи для опису предметної області локальних систем дозволяють в лаконічній формі представити інформаційні об'єкти та зв'язки між ними. Така форма представлення інформаційних об'єктів забезпечує можливість порівнювати різні інформаційні системи та отримувати відповідь на запитання про їх інтеграцію.

Розроблена алгоритмічна база, що реалізує методи адаптивної підтримки обробки розподілених даних на інтегрованому інформаційному просторі, дозволяє не контролювати структурні зміни в БД кожної локальної інформаційної системи, а управляти доступом до розподілених даних, використовуючи транзакції, які формуються і управляються (модифікуються) кожною системою окремо, що забезпечує незалежне функціонування інтегрованої системи в цілому.

Проведене тестування програмних засобів інтеграції розподілених даних показало, що використання методів адаптивної інтеграції, розроблених у ході виконання роботи, дозволяє маніпулювати локальними даними з зовнішнього рівня без порушення цілісності інформації.

Показано, що використання розроблених методів адаптивного управління розподіленими даними однаково ефективно може бути застосовано як в інформаційних системах з неоднорідною структурою баз даних, так і в системах з однорідною структурою. Супроводження програмного забезпечення визначило, що затрати на доробку та реорганізацію системних компонент та метаданих при розширенні інтегрованого інформаційного простору займає значно менш часу, ніж при реорганізації та реінжинірингу структури програмного забезпечення, що використовує єдине представлення інтегрованих даних.

В інституті транспорту газу ДК «Укртрансгаз» на основі запропонованих методів і інформаційних технологій розроблено засоби підтримки системи сумісного доступу до розподілених даних з незалежною структурою з

використанням методів логічного перетворення типів даних. Побудовано програмний шлюз, який забезпечує підтримку відображень моделей даних і реалізує редукцію реляційних операцій до логічних правил дедуктивної БД.

Отримані результати, що реалізовані у вигляді програмних елементів, які об'єднують засоби проміжного рівня (медіатора), дали можливість вирішити такі задачі:

- реалізувати ефективне управління доступом до даних, підтримуваних різними серверами (Oracle 10x, MS SQL server 8x, систем реального часу в складі АСК ТП на базі сучасних SCADA InTouch, Saitek);
- забезпечити надійну підтримку процесів синхронізації (реплікації) даних між серверами розподіленої БД інтегрованої автоматизованої системи управління ДК «Укртрансгаз».

Описано результати, що отримані під час проведеного дослідження, які дозволяють забезпечити ефективне формування запитів за рахунок використання принципів і синтаксису мови дедуктивних БД. Розроблено метод виведення, заснований на статичних правилах і безпосередньо правилах запиту, який дозволяє забезпечити вирішення задачі інтеграції даних з неоднорідною структурою без побудови глобальної (інтегрованої) схеми.

Показано, що використана технологія обробки інтегрованих даних дозволила, з одного боку, забезпечити гнучку підтримку розподіленої інформаційної системи із можливістю управління даними в локальних БД і, з іншого боку, організувати запити до БД із неоднорідною структурою, що значно знизило витрати на супровід інтегрованої системи в цілому.

У додатках наведено акти про впровадження результатів дисертаційної роботи у виробництво, держбюджетні роботи та навчальний процес, доведення тверджень, методи реалізації алгоритмів та приклади застосування розроблених методів.

ВИСНОВКИ

В дисертаційній роботі одержано результати, які в сукупності є подальшим узагальненням і розвитком теорії інтеграції гетерогенних інформаційних систем та фундаментальною основою загального підходу до моделювання даних і розв'язання задач управління інформаційними ресурсами розподілених систем БД. Результати роботи є теоретичною основою розв'язання важливої наукової проблеми: побудови єдиного інформаційного простору без визначення глобальної схеми даних. На основі теоретичних і експериментальних досліджень, застосування системного підходу та методів сучасних інформаційних технологій в роботі створені математичні моделі, методи і алгоритмічні засоби інтеграції та управління розподіленими даними, використання яких дозволило суттєво поліпшити якісні та кількісні показники функціонування корпоративних та регіональних інформаційних систем і баз даних.

У рамках даного напрямку отримані такі наукові результати, що мають істотні переваги перед існуючими рішеннями:

1. Досліджено різноманітні методи інтеграції інформації. Виділено три основних підходи: консолідація, розподіл і федералізація даних. Досліджено методи опису предметних областей. Визначено властивості предметних областей і недоліки методів їхнього опису традиційними засобами для подання інформації на даталогічному рівні.
2. Дано класифікацію неоднорідностей в БД. Обґрунтовано доцільність застосування підходу до управління неоднорідними даними, що базується на методах побудови розподілених інформаційних систем. Такий підхід є досить універсальним у теоретичному аспекті і конструктивним із прикладної точки зору.
3. Запропоновано формалізовану інфологічну модель ПрО, що орієнтована на семантичні відношення між інформаційними об'єктами БД. Виділено основні компоненти моделі ПрО і надані формальні визначення її базовим сутностям. Використання такої моделі дозволяє обійти структурну строгість алгебраїчних засобів при формулюванні запитів до табличних документів гетерогенних БД.
4. Удосконалено методи опису семантичних властивостей інформаційних об'єктів ПрО. Введено два типи правил: функціональні та структурні, які дозволяють аналізувати не тільки явно задані, але й логічно виведені інформаційні об'єкти, тобто визначати межу предметної області. Інформація про всі можливі інформаційні об'єкти ПрО дає можливість визначити область перетину семантики даних, що інтегруються.
5. Сформульовано аксіоматичний підхід до опису ПрО, що дає можливість розглянути проблему моделювання відношень елементів ПрО у вигляді набору правил. Такий підхід дозволяє аналізувати «неявні» семантичні властивості інформаційних об'єктів, які виводяться із заданої множини елементів даних.
6. Virішено задачу перевірки належності деякого правила до замикання множини правил L^+ (проблема членства). Розроблено відповідний метод, що опирається на послідовне застосування алгоритмів обчислення замикання. Цей метод вирішує задачу побудови мінімальної множини еквівалентних правил без використання замикання, що значно скорочує загальний час обчислення семантики БД.
7. Virішено задачу побудови мінімальної еквівалентної множини правил. Сформульовано і доведено необхідні умови редукції правил для побудови базису. Сформульовано умови оптимальної множини правил щодо числа інформаційних об'єктів. Таке вирішення дозволяє за прийнятний (поліноміальний) час визначати ступінь перетинання предметних областей, описуваних множинами інформаційних об'єктів.
8. Введено клас правил - обмежень структурної цілісності, що визначає спільність даних при модифікації структури БД. Запропоновано і теоретично обґрунтовано інформаційну технологію підтримки структурної цілісності інтегрованої схеми даних при допущенні змін у схемах локальних БД. Обмеження структурної цілісності забезпечують коректність модифікації елементів БД без порушення її семантики.

9. Запропоновано метод, що забезпечує цілісність модифікованої інформації при частковій відповідності обмеженням цілісності. Визначено умови, при яких модифікація даних спільна з обмеженнями, тобто може бути виконана зміна стану БД без порушень цілісності. Даний метод дає можливість контролювати часткове виконання обмежень, коли умови цілісності порушені при трансформації таблиці (дані недоступні).
10. Для побудови еквівалентних станів баз даних удосконалено метод визначення мінімальних змін семантики вихідної структури БД. Визначено умови модифікації БД, при яких структурні обмеження залишаються спільними. Запропонований метод визначає найбільш ефективну схему БД (щодо умов обмежень).
11. Побудовано відображення табличної моделі в дедуктивну модель на підставі бієктивної відповідності її компонентів: даних, відношень, операційної специфікації. Міжмодельне відображення визначає можливість інтегрувати неоднорідні БД, використовуючи більш гнучку, чим алгебраїчну, систему опису даних.
12. Розроблено метод підтримки цілісності даних при зміні стану БД, використовуючи властивості зведення обмежень, при яких стан БД залишається спільним. Показано, що для спільної БД завжди можна обчислити найменшу нерухому точку, тобто за кінцеву кількість кроків можна вивести бажаний результат.
13. Побудовано неповне комутативне відображення моделей даних. Визначено відповідність операцій реляційної алгебри та предикатних виразів мови запитів дедуктивних БД. Використання операційної специфікації моделі, що реалізується прикладною СУБД, дає можливість не реорганізовувати існуючу систему запитів при інтеграції БД.
14. Побудовані в роботі моделі, розроблені методи та алгоритми були використані як ядро інформаційних технологій формування динамічних запитів з використанням виразів реляційного числення зі змінними на доменах, використання дедуктивних баз даних для організації доступу до табличних структур даних тощо. Сформульовані умови еквівалентності запитів визначають припустимі транзакції, тобто перехід з одного спільного стану БД в інший спільний стан. Запропоновано використовувати вирази, що описують інтенціонал БД при виведенні кінцевого факту (результату), що дозволяє згладити структурні неоднорідності таблиць на логічному рівні (при виконанні запиту).

Достовірність отриманих результатів дисертаційної роботи підтверджена експериментальними дослідженнями, результатами впровадження на реальних об'єктах (акти впровадження наведені в Додатку до дисертації).

Напрямки теоретичних і практичних досліджень дисертаційної роботи доцільно розвивати в області проектування сховищ даних для підтримки оперативного аналізу розподіленої інформації.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Танянский С.С. Семантика реляционных баз данных с неопределёнными значениями / С.С. Танянский, И. В. Кобзев, И.А. Яковлева // Радиоэлектроника и информатика. – 1998. – № 1(02) – С. 51–53.
2. Танянский С.С. Об одном вопросе построения системы поддержки принятия решений на основе распределенных систем обработки данных / С.С. Танянский, Д.А. Руденко, Е.С. Яковлева // Радиоэлектроника и информатика. – 1999. – № 4(9) – С. 72 – 76.
3. Танянский С.С. Оптимальный план соединения в интегрированных системах совместной обработки данных / С.С. Танянский, Д.А. Руденко, Е.С. Самойлова // Інтегровані технології та енергозбереження. – 2000. – № 1. – С. 98 – 106.
4. Певнев В.Я. Семантика неопределенных значений в моделях безопасности баз данных / В.Я. Певнев, С.С. Танянский, Д.А. Руденко // Системи обробки інформації. – 2000. – Випуск 3(9). – С. 41 – 46.
5. Чирун В.А. Повышение эффективности работы интегрированных информационных систем подразделений ОВД Украины / В.А. Чирун, С.С. Танянский, Д.А. Руденко // Право і безпека. – 2003. – № 2’2. – С. 188 – 190.
6. Пономаренко Л.А. Построение оптимальной последовательности соединения отношений в запросах реляционной базы данных / Л.А. Пономаренко, С.С. Танянский, В.А. Филатов // Системні дослідження та інформаційні технології. – 2003. – № 2. – С. 53 – 58.
7. Танянский С.С. Оценка вычислительной сложности операций реляционной алгебры в системах управления базами данных / С.С. Танянский, Д.А. Руденко, В.В. Тулупов // Вестник Национального технического университета «ХПИ». – 2003. – № 19. – С. 161 – 166.
8. Танянский С.С. Сравнительная характеристика показателей сложности выполнения запросов в реляционных СУБД / С.С. Танянский, В.А. Филатов, В.В. Тулупов // Системи обробки інформації. – 2004. – Випуск 2. – С. 91 – 95.
9. Танянский С.С. Модель поддержки целостности в интегрированных базах данных неоднородной структуры / С.С. Танянский // Вестник Херсонского государственного технического университета. – 2004. – №1(19). – С. 230 – 235.
10. Танянский С.С. Управление интегрированной неоднородной системой баз данных на основе модели поведения автоматов / С.С. Танянский // Искусственный интеллект. – 2004. – № 4. – С. 456 – 461.
11. Танянский С.С. Децентрализованное управление потоками транзакций / С.С. Танянский, В.В. Тулупов, Д.А. Руденко // Вестник Национального технического университета «ХПИ». – 2004. – № 34. – С. 179 – 184.
12. Танянский С.С. Семантическая модель предметной области в задачах интеграции неоднородных информационных систем / С.С. Танянский // Вестник Херсонского Национального технического университета. – 2005. – №1(21). – С. 52 – 59.
13. Танянский С.С. Семантическая эквивалентность слабоструктурированных баз данных / С.С. Танянский, Д.А. Руденко, В.В. Тулупов // Вестник Национального технического университета «ХПИ». – 2005. – № 46. – С. 160 – 165.

- 14.Танянский С.С. Поддержка приведенной модели предметной области при интеграции слабоструктурированных данных / С.С. Танянский // Вестник Херсонского Национального технического университета. – 2006. – №1(24). – С. 216 – 222.
- 15.Танянский С.С. Модификация ограничений на ведение данных для обеспечения целостности крупномасштабных информационных систем / С.С. Танянский, В.В. Тулупов, Д.А. Руденко // Вестник Национального технического университета «ХПИ». – 2006. – № 23. – С. 137 – 144.
- 16.Танянский С.С. Модель согласования данных при интеграции информационных систем / С.С. Танянский // Радиоэлектроника и информатика. – 2006. – № 2 – С. 84 – 91.
- 17.Танянский С.С. Оценка качества функционирования интегрированных информационных систем / С.С. Танянский, Е.Е. Гринева // Системні дослідження та інформаційні технології. – 2006. – № 3. – С. 37 – 47.
- 18.Танянский С.С. Характеристические свойства объектов информационных систем / С.С. Танянский // Искусственный интеллект. – 2007. – № 1. – С. 78 – 89.
- 19.Танянский С.С. Технология построения крупномасштабной базы данных / С.С. Танянский // Бионика интеллекта. – 2007. – № 2(65). – С. 53 – 56.
- 20.Танянский С.С. Управление потоками транзакций в системах распределенной обработки данных / С.С. Танянский // Вестник Херсонского Национального технического университета. – 2007. – № 27. – С. 252 – 258.
- 21.Танянский С.С. Поиск частично совпадающих данных в задачах идентификации информационных объектов / С.С. Танянский, О.В. Вахета // Системи обробки інформації. – 2007. – Випуск 4(62). – С. 118 – 120.
- 22.Танянский С.С. Межмодельное отображение неоднородных баз данных / С.С. Танянский // Системні технології. – 2008. – Т.1. - Вип. 3(56). – С. 158 – 162.
- 23.Танянский С.С. Операционная спецификация несогласованных баз данных / С.С. Танянский // Вестник Херсонского Национального технического университета. – 2008. – № 30. – С. 203 – 209.
- 24.Пономаренко Л.А. Интеграция информационных систем при частичном отображении моделей данных / Л.А. Пономаренко, С.С. Танянский, В.А. Филатов // Проблеми системного підходу в економіці. – 2008. – Випуск 26. – С. 33 – 44.
- 25.Танянский С.С. Модель процессов выполнения множества запросов на основе сетей Петри / С.С. Танянский // Вестник Херсонского Национального технического университета. – Херсон, ХНТУ – 2009. – № 34. – С. 298 – 302.
- 26.Касаткина Н.В. Построение системы функциональных ассоциативных правил на основе свойств реляционной модели данных / Н.В. Касаткина, Л.А. Пономаренко, С.С. Танянский, В.А. Филатов // Вестник Киевского Национального университета им. Т. Г. Шевченко. – 2009. – Выпуск 23. – С. 34 – 38.
- 27.Касаткина Н.В. Методы хранения и обработки нечетких данных в среде реляционных систем / Н.В. Касаткина, С.С. Танянский, В.А. Филатов // Автоматика. Автоматизація. Електротехнічні комплекси та системи. – 2009. – Випуск 2(24). – С. 80 – 86.

28. Касаткина Н.В. Об одном подходе к построению объектно–реляционной модели данных / Н.В. Касаткина, С.С. Танянский, Е.Б. Чапланова // Збірник наукових праць Військового інституту Київського національного університету ім. Т. Шевченка. – 2009. – Випуск № 20. – С. 141 – 147.
29. Танянский С.С. Поддержка логической целостности в задачах преобразовании структур баз данных / С.С. Танянский // Вестник Херсонского Национального технического университета. – 2010. – № 2(38). – С. 227 – 231.
30. Танянский С.С. Организация запросов к распределенным данным средствами логического программирования / С.С. Танянский, Ю.А. Мальков // Бионика интеллекта. – 2010. – № 1(72). – С.118 – 121.
31. Танянский С.С. Поддержка целостности баз данных при динамических структурных изменениях / С.С. Танянский // Бионика интеллекта. – 2010. – № 3(74). – С.112 – 119.
32. Пономаренко Л.А. Обработка данных произвольной структуры языковыми средствами реляционной модели / Л.А. Пономаренко, С.С. Танянский // Управляющие системы и машины. – 2010. – № 6. – С. 47 – 53.
33. Танянский С.С. Реинжиниринг логической схемы реляционной базы данных с использованием методов выявления функциональных зависимостей / С.С. Танянский, В.А. Радченко // Вестник Херсонского Национального технического университета. – 2011. – № 2(41). – С. 197 –200.
34. Филатов В.А. Модель «сущность–связь» в задачах представления объектно - реляционных свойств предметной области / В.А. Филатов, С.С. Танянский, Е.Б. Чапланова, А.И. Сизов // Управляющие системы и машины. – 2011. – № 3. – С. 73 – 78.
35. Таняньський С.С. Проблеми захисту даних в інформаційних системах органів внутрішніх справ / С.С. Таняньський, Д.О. Руденко // Актуальні проблеми сучасної науки в дослідженнях молодих вчених: зб. наук. праць. – Х. – 1997. – Випуск 3 – 4. – С. 492 – 494.
36. Руденко Д.А. Использование систем поддержки принятия решений на основе крупномасштабных баз данных / Д.А. Руденко, С.С. Танянский // Зб. наук. праць Української Академії державного управління при Президенті України. – Х., УАДУХФ. – 2001.– Випуск 1. – Част. 5. – С. 64 – 67.
37. Танянский С.С. Автоматизация распределения учебной нагрузки в высших учебных заведениях / С.С. Танянский, М.А. Колтунов, В.С. Павлов // Актуальні проблеми державного управління. – Зб. наук. праць. – Х., 2002. – №2(13). – С.297 – 299.
38. Таняньський С.С. Організація доступу до баз даних у сфері послуг / С.С. Таняньський, В.О. Філатов, К. В. Глушко // Вісник Київського національного торговельно–економічного університету – К., 2004. – № 1. – С. 112–119.
39. Танянский С.С. Языковая поддержка манипулирования данными, основанная на логике предикатов первого порядка / С.С. Танянский // Управління розвитком. – Х., ХНЕУ, 2007 – № 7. – С. 63 – 64.

- 40.Танянский С.С. Отображение баз данных в задачах интеграции информационных систем / С.С. Танянский, Д.А. Руденко // Управління розвитком. – Х., ХНЕУ, 2008. – № 15. – С.117 – 118.
- 41.Танянский С.С. Об одном алгоритме преобразования логической схемы несогласованной реляционной базы данных / С.С. Танянский, В.А. Филатов, И.В. Кобзев // Теория и техника передачи, приема и обработки информации: 3-я междунар. конф.: тезисы докл. – Туапсе, 1997. – С. 275 – 276.
- 42.Танянский С.С. Система автоматизации проектирования схемы и процессов ведения реляционной базы данных / С.С. Танянский, И.В. Кобзев, Д.А. Руденко // Використання комп'ютерних технологій у навчальному процесі: науково-метод. конф.: тези доп. – Х., 1997. – С. 154 – 155.
- 43.Руденко Д.А. Критерии оценки надежности и средства защиты систем управления базами данных / Д.А. Руденко, С.С. Танянский // Теория и техника передачи, приема и обработки информации. Новые информационные технологии: сб. научн. работ по материалам 4-й междунар. научн. конф. – Х., 1998. – С. 501 – 502.
- 44.Танянский С.С. Об одном подходе к использованию неопределенных значений в многоуровневой модели безопасности баз данных / С.С. Танянский, Д.А. Руденко // Теория и техника передачи, приема и обработки информации. Новые информационные технологии: сб. научн. работ по материалам 4-й междунар. научн. конф. – Х., 2000. – С. 44 – 46.
- 45.Танянский С.С. Об одном подходе построения модели несогласованной реляционной базы данных / С.С. Танянский // Информатизация правоохранительных систем: сб. работ 10-й междунар. научн. конф. – М., 2001. – С. 217 – 219.
- 46.Танянский С.С. К вопросу об эффективности выполнения запросов в реляционных базах данных / С.С. Танянский, Д.А. Руденко // Теория и техника передачи, приема и обработки информации. Интегрированные информационные системы, сети и технологии: сб. научн. работ по материалам 8-й междунар. научн. конф. – Х., 2002. – С.480 – 481.
- 47.Филатов В.А. Модель информационного пространства мультиагентной системы управления потоками работ / В.А. Филатов, С.С. Танянский, Д.А. Руденко // Автоматика – 2003: материалы 10-й междунар. конф. по автоматическому управлению. – Севастополь, СевНТУ, 2003. – Т. 3. – С. 113 – 114.
- 48.Руденко Д.А. Преобразование запросов в распределенных системах обработки данных / Д.А. Руденко, С.С. Танянский // Теория и техника передачи, приема и обработки информации: сб. тезисов докладов по материалам 9-й междунар. научн. конф. – Харьков – Туапсе, 2003. – С. 264 – 265.
- 49.Руденко Д.А. Семантическая структура ограничений целостности баз данных / Д.А. Руденко, С.С. Танянский, Л.И. Маркова // Теория и техника передачи, приема и обработки информации: сб. тезисов докладов по материалам 10-й междунар. научн. конф. – Харьков – Туапсе, 2004. – С. 185 – 186.
- 50.Танянский С.С. Об одном алгоритме определения семантики неоднородных баз данных / С.С. Танянский, О.Ф. Козыр // Образование, наука, производство и

- управление в XXI веке: сб. работ междунар. научн. конф. – Старый Оскол, 2004. – Т.1. – С. 320 – 322.
- 51.Танянский С.С. Формальная система построения объектов предметной области в задачах интеграции информационных систем / С.С. Танянский, Д.А. Руденко // Образование, наука, производство и управление в XXI веке: сб. работ регион. научн. конф. – Старый Оскол, 2005. – Т.1. – С. 276 – 280.
 - 52.Гринева Е.Е. Об одном подходе к оценке качества функционирования интегрированных информационных систем / Е.Е. Гринева, С.С. Танянский // Системний аналіз та інформаційні технології САІТ 2009: матеріали VII міжнар. науково – техн. конф.– К., ІПСА, 2005. – С. 185.
 - 53.Танянский С.С. Об одном подходе к решению задачи поддержки целостности крупномасштабных баз данных / С.С. Танянский, В.В. Тулупов, Д.А. Руденко // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я MicroCAD 2006: матеріали XIV міжнар. науково-практич. конф. – Х., НТУ “ХПІ”, 2006. – С. 50.
 - 54.Маркова Л.И. Метод построения блочно-ациклической схемы реляционной базы данных / Л.И. Маркова, С.С. Танянский, Д.А. Руденко // Глобальные информационные системы. Проблемы и тенденции развития: материалы I междунар. научн. конф. – Харьков –Туапсе, ХНУРЭ, 2006. – С.101 – 102.
 - 55.Танянский С.С. Операционная спецификация управления информационными объектами в корпоративных системах с динамической структурой / С.С. Танянский, Д.А. Руденко // Сложные системы управления и менеджмент качества: сб. материалов междунар. научн. конф. – Старый Оскол, СТИ, 2007. – С.221 – 224.
 - 56.Танянский С.С. Об одном методе поиска данных в задачах идентификации информационных объектов / С.С. Танянский, О.В. Вахета // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я MicroCAD 2007: матеріали XV міжнар. науково-практич. конф. – Х., НТУ “ХПІ”, 2007. – С.51.
 - 57.Танянский С.С. Организация запросов к базе данных с неоднородной структурой / С.С. Танянский, Д.А. Руденко // Современные информационные системы. Проблемы и тенденции развития: материалы II междунар. научн. конф. – Харьков – Туапсе, ХНУРЭ, 2007. – С.245 – 246.
 - 58.Танянский С.С. Модель представления данных в задачах управления учебным процессом / С.С. Танянский // Інформатизація вищих навчальних закладів МВС України: матеріали науково – практ. конф. – Х., НУВС, 2007. – С.106 – 111.
 - 59.Танянский С.С. Поддержка функциональных зависимостей в системах совместной обработки данных / С.С. Танянский // Системний аналіз та інформаційні технології: матеріали X міжнар. науково – техн. конф. – К., ІПСА, 2008. – С.411.
 - 60.Таняньський С.С. Властивості операційної специфікації при відображенні реляційних моделей даних / С.С. Таняньський, В.С. Левченко // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я MicroCAD 2008: матеріали XVI міжнар. науково-практич. конф. – Х., НТУ “ХПІ”, 2008. – С.331.

- 61.Танянский С.С. Эквивалентность схем баз данных при наличии ограничений / С.С. Танянский, Е.Е. Гринева // Автоматика-2008: материалы XV междунар. конф. по автоматическому управлению. – Одесса, Одесская национальная морская академия, 2008. – С.584 – 585.
- 62.Танянский С.С. Метод реструктуризации канонической структуры базы данных с учетом динамики изменений предметной области / С.С. Танянский, В.С. Левченко // Проблемы информатики и моделирования: материалы 8-й междунар. научно-техн. конф. – Х., НТУ «ХПИ», 2008. – С.27.
- 63.Танянский С.С. Структуризация предметной области для построения пользовательской модели данных / С.С. Танянский, Д.А. Руденко // Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці: матеріали III Всеукр. науково-практ. конф. – Луганськ, ЛНУ ім. Т. Шевченка, 2009. – С.98 – 100.
- 64.Танянский С.С. Об одном подходе к организации предикатных запросов / С.С. Танянский, Д.А. Руденко // Современные информационные и электронные технологии: труды 10-й междунар. научно-практ. конф. – Одесса, ОНПУ, 2009. – С.66.
- 65.Танянский С.С. Организация запросов к таблицам сетевой структуры языковыми средствами реляционной структуры данных / С.С. Танянский // Системний аналіз та інформаційні технології САІТ 2009: матеріали XI міжнар. науково – техн. конф.– К., ІПСА, 2009. – С.572.
- 66.Танянский С.С. Использование методов логического проектирования при организации запросов к реляционным базам данных / С.С. Танянский, Ю.А. Мальков // Автоматика-2009: материалы XV междунар. конф. по автоматическому управлению. – Черновцы, Черновицкий национальный университет им. Ю. Федьковича, 2009. – С.342 – 344.
- 67.Касаткіна Н.В. Синтез логічних структур великомасштабних баз даних / Н.В. Касаткіна, С.С. Таняньський, В.О. Філатов // Комп'ютерні науки та інформаційні технології CSIT-2009: матеріали 4-ї між нар. науково-техн. конф. – Львів, Національний університет «Львівська політехніка», 2009. – С.374 – 377.
- 68.Танянский С.С. Поддержка объектных технологий в реляционных базах данных / С.С. Танянский, Е.Б. Чапланова // Проблемы информатики и моделирования: материалы 9-й междунар. научно-техн. конф. – Х., НТУ «ХПИ», 2009. – С.49.
- 69.Танянский С.С. Определение функциональных зависимостей в системах реляционных баз данных / С.С. Танянский, В.А. Радченко // Проблемы информатики и моделирования: материалы 9-й междунар. научно-техн. конф. – Х., НТУ «ХПИ», 2009. – С.49 – 50.
- 70.Танянский С.С. Использование многоуровневых транзакций в системах обработки данных реального времени / С.С. Танянский, Д.А. Руденко // Проблеми й перспективи розвитку ІТ-індустрії: матеріали 1-ї міжнар. науково-практ. конф. – Х., ХНЕУ, 2009. – С.348 – 350.
- 71.Касаткіна Н.В. Поиск функциональных зависимостей средствами интеллектуального анализа данных / Н.В. Касаткіна, С.С. Таняньський, В.А. Філатов // Математичне та програмне забезпечення інтелектуальних систем:

- матеріали 7-й міжнар. науково-практ. конф. – Дніпропетровськ, ДНУ ім. Олеса Гончара, 2009. – С.106 – 107.
- 72.Танянський С.С. Структурна специфікація семантичної моделі баз даних / С.С. Танянський, Д.О. Руденко // Інформатика та системні науки (ІСН-2010): матеріали Всеукр. науково-практ. конф. – Полтава, Полтавський університет споживчої кооперації України, 2010. – С. 182 – 185.
- 73.Танянський С.С. Об одном алгоритме синтеза рациональной схемы базы данных / С.С. Танянский, Д.А. Руденко, А.В. Белущенко // Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці: матеріали III Всеукр. науково-практ. конф. – Луганськ, ЛНУ ім. Т. Шевченка 2010. – С.124 – 127.
- 74.Танянский С.С. Редукция правил Datalog к операциям реляционной алгебры / С.С. Танянский, Ю.А. Мальков // Системний аналіз та інформаційні технології, САІТ 2010: матеріали XII міжнар. науково– техн. конф. – К., ІІСА, НТУУ «КПІ», 2010. – С.493.
- 75.Мальков Ю.А. Використання логічних правил для забезпечення обмежень цілісності в дедуктивних базах даних / Ю.А. Мальков, С.С. Танянский // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я. МісгоСАД 2010: матеріали XVIII між нар. науково-практ.конф. – Х., НТУ «ХПІ», 2010. – С.35.
- 76.Танянский С.С. Свойство семантики динамически изменяемых баз данных / С.С. Танянский // Інформаційні технології в навігації і управлінні: стан і перспективи розвитку: матеріали 1-ї міжнар. науково-техн. конф. – К., ДП «ЦНДІ НіУ», 2010. – С.49.
- 77.Мальков Ю.А. Редукция операций реляционной алгебры к правилам логического программирования / Ю.А. Мальков, С.С. Танянский // Автоматика – 2010: матеріали 17-ї міжнар. конф. по автоматичному управлінню – Х., ХНУРЕ, 2010. – Т.1. – С.132 -134.
- 78.Танянский С.С. Условия сохранения однозначности ограничений целостности при модификации базы данных / С.С. Танянский, Д.А. Руденко // Сучасні напрямки розвитку інформаційно-комунікаційних технологій та засобів управління: матеріали 1-ї науково-техн. конф. – Х., ДП «ХНДІТМ», Київ, ДП«ЦНДІ НіУ», 2010. – С.73 – 74.
- 79.Танянский С.С. Вычисление наименьшей неподвижной точки в рекурсивных запросах к базам данным / С.С. Танянский, Д.А. Руденко // Інформаційні технології в навігації і управлінні: стан і перспективи розвитку: матеріали 2-ї міжнар. науково-техн. конф. – К., ДП «ЦНДІ НіУ», 2011. – С.43.
- 80.Танянский С.С. Эквивалентность запросов к данным с неоднородной структурой / С.С. Танянский, Ю.А. Мальков // Информатика, математическое моделирование, экономика: междунар. научно–практ. конф. – Смоленск, Смоленский филиал АНО ВПО ЦС РФ «Российский университет кооперации», 2011. – С.150 – 154.

АНОТАЦІЯ

Таняньський С.С. Моделі, методи та інформаційні технології інтеграції гетерогенних розподілених баз даних. - Рукопис.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 - інформаційні технології. - Харківський національний університет радіоелектроніки, Харків, 2011.

Дисертаційна робота присвячена рішенням науково-практичної проблеми створення на єдиній методологічній основі нових моделей і методів інтеграції розподілених баз даних з неоднорідною структурою, методів і засобів аналізу взаємозв'язку моделей даних.

Запропоновано класифікацію неоднорідних структур баз даних, виділено класи фізичної і логічної неоднорідності. Показано, що для інтеграції неоднорідних баз даних найбільш прийнятною є технологія федералізації даних.

Введено правила (L - правила), які формують повну, у деякому сенсі, множину інформаційних об'єктів, що визначає інформаційний простір розглянутої предметної області. Введено систему аксіом для побудови замикання правил L^+ . Вирішено задачу перевірки належності деякого правила до множини L^+ .

На основі логічних правил існування введено окремий клас правил - обмежень цілісності. Досліджено поводження семантики при модифікації бази даних. Розроблено метод мінімальних змін семантики бази даних при модифікації її структури. Досліджено можливості маніпулювання даними, структурованими з порушеннями вимог реляційної моделі. Формально визначено мову запитів у вигляді виразів числення зі змінними на доменах. Показано, що запропоноване формулювання мови може бути застосовано не тільки до даних реляційної структури.

Отримані теоретичні результати використані при розробці та впровадженні інформаційних систем різного функціонального призначення, а також використовувалися в навчальному процесі, що підтверджується актами та довідками про впровадження.

Ключові слова: предметна область, інформаційна система, модель даних, база даних, розподілена база даних, правила логічного існування, інтеграція даних, неоднорідні структури даних, цілісність даних, гетерогенна інформаційна система.

АННОТАЦИЯ

Таняньский С.С. Модели, методы и информационные технологии интеграции гетерогенных распределенных баз данных. – Рукопись.

Диссертация на соискание ученой степени доктора технических наук по специальности 05.13.06 – информационные технологии. – Харьковский национальный университет радиоэлектроники, Харьков, 2011.

Диссертационная работа посвящена решению научно-практической проблемы создания на единой методологической основе новых моделей и методов интеграции распределенных баз данных с неоднородной структурой, вычислительных

механизмов, методов и инструментальных средств анализа взаимосвязи моделей данных с целью применения их в интегрированных информационно-аналитических системах и системах поддержки принятия решений.

Предложена классификация неоднородных структур баз данных, выделены классы физической и логической неоднородности. Показано, что для интеграции неоднородных баз данных наиболее приемлемой является технология федерализации данных.

Формально описана расширенная модель предметной области в виде набора множеств информационных задач Z , решаемых в информационной системе, множества входных и выходных информационных объектов O , а также множества взаимосвязей между ними. Построена модель спецификаций информационных требований для каждой возможной структуры предметной области и определена последовательность действий для решения задач формирования спецификаций информационных требований.

Введены правила (L - правила), которые формируют полное, в некотором смысле, множество информационных объектов, определяющее информационное пространство рассматриваемой предметной области. Определены свойства L - правил, разработан алгоритм проверки согласованности правил относительно множества O .

Введена система аксиом для построения полного множества L - правил. Доказано, что система аксиом может строить замыкание L^+ для любого множества L - правил. Решена задача проверки принадлежности некоторого правила $l_{1\dots y} \leftarrow l_{1\dots x}$ к множеству L^+ . Разработан соответствующий метод, опирающийся на последовательное применение алгоритмов вычисления замыкания и собственно самой проверки принадлежности правила к L^+ . Для сравнения различных множеств L - правил введено понятие покрытия, доказано условие эквивалентности множества правил L и его покрытия \hat{L} .

На основе логических правил существования введен отдельный класс правил – ограничений целостности. Введены понятия совместности базы данных и совместности ограничений и на их основании расширено понятие семантики базы данных. Исследовано поведение семантик при модификации базы данных. Определены условия допустимой модификации базы данных. Разработан метод минимальных изменений семантики базы данных при модификации ее структуры.

Определены свойства совместности базы данных при ее модификации. Сформулировано и доказано утверждение, определяющее эквивалентность семантик базы данных при модификации ограничений целостности.

Сформулированы условия корректного перехода состояний базы данных $\delta: \tau_0 \rightarrow \tau_1$ как отображение состояний, при котором ограничения целостности из всего множества возможных отображений определяют те, на которых они выполняются. Определены условия, при которых модификация базы данных совместна с ограничениями, то есть может быть выполнено изменение состояния базы данных без нарушения ограничений целостности. Определена функция вывода информационных объектов $M^I(\tau_0 \rightarrow \tau_1)$ для различных состояний базы данных.

Сформулирована и доказана теорема, подтверждающая совместность выполняемых модификаций с заданными ограничениями целостности.

Исследованы возможности манипулирования данными, структурированными с нарушениями требований реляционной модели. Сформулированы условия эквивалентности запросов. Формально определен язык запросов в виде выражений исчисления с переменными на доменах. Показано, что предложенная формулировка языка может быть применена не только к данным реляционной структуры. Определены характеристики ограничений целостности для формирования формул исчисления.

Определены основные свойства и характеристики декларативного описания данных. Показано, что использование синтаксических конструкций дедуктивных баз данных хорошо сопрягаются с введенными логическими правилами существования при построении интенционала базы данных. Трансляция реляционного языка в язык логики предикатов реализована посредством редукции реляционных операций в правила DataLog. Определены свойства редукции для каждой операции.

Для поддержки реляционной базы данных средствами логического программирования решена задача отображений моделей данных $\xi : M \rightarrow M^{DDB}$. Для установления корректного отображения предлагается рассматривать частичное отображение, при котором сохраняется операционная спецификация исходной модели. Показано, что отображение между табличной и дедуктивной моделью является односторонним, так как редукция правил к операциям алгебры не является однозначной.

Полученные теоретические результаты использованы при разработке и внедрении систем различного функционального назначения – от корпоративных систем организационного управления до региональных распределенных информационных систем, а также использовались в учебном процессе, что подтверждается актами и справками о внедрении.

Ключевые слова: предметная область, информационная система, модель данных, база данных, распределенная база данных, правила логического существования, интеграция данных, неоднородные структуры данных, целостность данных, гетерогенная информационная система.

ABSTRACT

Tanyanskiy S.S. Models, methods and information technologies for integration of heterogeneous distributed databases. – Manuscript.

A Thesis for a Doctor of Technical Sciences degree in the speciality 05.13.06 – information technologies. – Kharkiv National University of Radio Electronics, Kharkiv, 2011.

The thesis is devoted to solving of scientific and practical problem of creation on unitary methodological base of new models and methods for integration of distributed databases with heterogeneous structure, methods and tools of analysis of data models association.

A classification of heterogeneous structures of databases is proposed, a classes of physical and logical heterogeneity are extracted. It is shown that the technology of data federalization is most acceptable for integration of heterogeneous data.

L-regulations which form complete in one sense set of information objects that defines an information space of the knowledge domain are introduced. A system of axioms for constructing of the regulations L^+ is proposed. A problem of checking of belonging of a regulation to the set L^+ is solved.

A separate class of the regulations – constructions of integrity is introduced on the ground of the logical existence regulations. A behavior of semantics under modification of the database is investigated. A method of minimal changes of database semantics under modification of the database structure is worked out. Possibilities of manipulation of the data which are structured with violation of requirements of relational model are investigated. A query language in the form of expressions of calculus with variables on domains is formally defined. It is shown that proposed formulation of the language may be applied not only to data having relational structure.

Obtained theoretical results are used under construction and implementation of information systems of different functional purpose and training that is confirmed by acts and certificates about implementation.

Key words: knowledge domain, information system, data model, database, distributed database logical existence regulations, data integration, heterogeneous data structures, data integrity, heterogeneous information system.