

Київська міська державна адміністрація
Науково-виробниче об'єднання "Міськсистемотехніка"

На правах рукопису
УДК 681.3

БЕРКО АНДРІЙ ЮЛІАНОВИЧ

МЕТОДИ І ЗАСОБИ ПРЕДСТАВЛЕННЯ ЗАЛЕЖНОСТЕЙ
В РЕЛЯЦІЙНИХ БАЗАХ ДАНОК ЯК ФОРМИ ЗНАТЬ

Спеціальність 05.13.17

Теоретичні основи інформатики

*Автореферат дисертації на здобуття вченого ступеня
кандидата технічних наук*



Київ 1983

ТВ 28.28

Робота виконана в НВО "Міськсистемотехніка"

Науковий керівник: чл.-кор. АН України, д.ф.-м.н., професор
Стогній Анатолій Олександрович.

Офіційні опоненти: д.т.н., проф. Г.П. Кожевнікова
к.ф.-м.н., с.н.с. М. В.Оленін

Провідна організація - Інститут програмних систем
науково-технічного комплексу "Інститут кібернетики
ім. В.М. Глушкова" АН України (м.Київ)

Захист відбудеться 10 листопада 1993 р. об 15⁰⁰ год.
на засіданні Спеціалізованої Ради Д 166.01.01 в науково-виробничому
об'єднанні "Міськсистемотехніка" за адресою:
252004, м. Київ, вул. Червоноармійська 23-б.

З дисертацією можна ознайомитись в бібліотеці
НВО "Міськсистемотехніка".

Автореферат розісланий 12 листопада 1993 р.

Вчений секретар
Спеціалізованої Ради
Д 166.01.01

Гладун В.П.

ЛННБ України ім.В.Стефаніка



00810561 (L)

ЛННБ ім. В. Стефаніка
АН України

AB-28,297

Загальна характеристика роботи.

Актуальність роботи - в даний час можливості апарату залежностей в проектуванні, моделюванні, управлінні та експлуатації баз даних використовуються недостатньо, а в багатьох випадках просто ігноруються. Причини цього автор бачає у відсутності однозначного підходу до розуміння, визначення, обробки та представлення залежностей даних. Різноманітність типів, форм, характеристик та методів роботи з залежностями створює цілий ряд труднощів, зокрема таких як повне використання переваг та властивостей тої чи іншої залежності при створенні бази даних, комплексне застосування залежностей різних типів, формалізація процесів роботи з залежностями та ін. Серед засобів формального маніпулювання залежностями даних найбільш широко вживаними є пропозиційна логіка, логіка першого порядку, реляційні числення, алгебраїчний апарат, графічні та табличні методи. Кожен з цих засобів придатний і має переваги в певних застосуваннях, і, в той же час, має цілий ряд обмежень як по типах залежностей, так і по широті використання.

Тому проблема створення єдиного підходу до способів обробки та представлення залежностей, як специфічної форми знань бачиться авторові актуальною, так як і проблема створення засобів обробки даних, що базуються на знаннях. Інтеграція досвіду, нагромадженого за час дослідження різних типів залежностей даних і способів їх застосування, з технологіями штучного інтелекту може служити основою створення інструментальних засобів нового типу.

Вказані вище фактори визначили напрямок, тему та задачі досліджень.

Ціль та задачі роботи. Основною ціллю роботи є розробка методів та засобів представлення і обробки залежностей в реляційних базах даних і створення інструментальних засобів застосування залежностей, які базуються на методах та технологіях штучного інтелекту, а також застосування отриманих результатів і програмних засобів у наукових та прикладних розробках і навчальному процесі.

У відповідності до поставленої цілі визначаються такі основні напрямки досліджень:

- аналіз способів представлення залежностей, методів і алгоритмів маніпулювання ними в різних застосуваннях;

- дослідження природи залежностей даних їх властивостей, джерел утворення, структури та семантики;
- розробка узагальненого методу представлення та обробки залежностей даних;
- розробка структури та специфікація задач управління даними за допомогою знань утворених на основі залежностей;
- розробка інструментальних програмних засобів на основі технологій штучного інтелекту для апробації та реалізації отриманих результатів.

Наукова новизна роботи полягає в наступному:

- реалізується спосіб представлення та обробки залежностей, який дозволяє інтегрувати різні типи та класи залежностей даних в єдине поняття залежностей даних та об'єднати при цьому їх властивості;
- залежності даних розглядаються, як специфічна форма знань про проблемну область, структуру та властивості реляційної бази даних;
- застосовується технологія штучного інтелекту для обробки та представлення залежностей даних;
- розроблений спосіб комп'ютерного представлення та обробки залежностей даних як форми знань, а також алгоритми автоматизованого розв'язку ряду задач управління реляційними базами даних на основі залежностей.

Практична цінність роботи полягає в тому, що розроблений інтегрований підхід до залежностей в реляційних базах даних, метод представлення залежностей як специфічної форми знань і застосування технології штучного інтелекту для їх обробки може служити основою для створення систем і засобів управління базами даних принципово нового типу. В СУБД, що використовують залежності даних, можуть практично вирішуватись задачі підтримки цілісності даних та побудови семантично правильних структур на основі різних типів обмежень.

Апробація результатів роботи. Окремі положення роботи за час її виконання доповідались на IV Всесоюзній конференції "Управление большим городом" (Москва, 1989 р.), Міжнародній конференції з інтелектуальних систем управління (Conference on Intelligent Management Systems, Варна, 1989 р.), Всесоюзній школі-симпозиумі "Системология: логика междисциплинарных исследований" (Львів, 1989 р.), III Всесоюзній школі-семінарі "Базы данных и знаний для

персональних ЕВМ" (Львів-Славсько, 1989 р.), наукових семінарах та конференціях у Львівському політехнічному інституті.

Ряд досліджень, виконані і апробовані в рамках науково-дослідних робіт по держбюджетній та госпдоготорній тематиці, що проводились протягом 1987-1992 рр. на кафедрі АСУ Львівського політехнічного інституту.

Окремі практичні результати, отримані в процесі підготовки роботи використані в розробках автоматизованих систем управління та автоматизованих робочих місць впроваджених в обчислювальному центрі Львівського політехнічного інституту, Рівненському обласному госпіталі інвалідів війни, спеціальному конструкторському бюро електромеханічних систем ЛПІ, а також в навчальному процесі на кафедрі АСУ ЛПІ, що підтверджується відповідними актами впровадження.

Основні положення та результати виконаних досліджень опубліковані в 6 наукових статтях та матеріалах наукових конференцій та 3 звітах по науково-дослідній роботі.

До захисту пропонуються:

- 1) аналітичний огляд методів представлення та обробка залежностей в реляційних базах даних, висновки, оцінки узагальнення та пропозиції щодо застосування основних способів аналітичного маніпулювання залежностями даних;
- 2) результати дослідження природи, семантики, властивостей і способів продукування залежностей даних, критерії та методика визначення різних типів залежностей в реляційних базах даних;
- 3) метод представлення залежностей в реляційних базах даних як форми знань у вигляді інтегрованого комплексного поняття з застосуванням форм і способів представлення знань, механізм та алгоритм переходу від залежностей даних до метазнань реляційної бази даних;
- 4) практична реалізація запропонованого методу представлення знань, утворених на основі залежностей, постановка та характеристика задач управління і обробки баз даних реляційного типу із застосуванням знань такого типу.

Структура роботи. Робота складається з вступу, чотирьох розділів, підсумків, списку літератури та додатків.

Зміст роботи.

У вступі обґрунтовується важливість та актуальність питань, що підлягають розробці в даній роботі. Сформульовані цілі та задачі досліджень, їх наукова новизна, основні положення та практичні результати, досягнуті в ході виконання роботи, дається коротка характеристика змісту розділів.

У першому розділі роботи даються визначення і формалізації основних понять та термінів, які використовуються в роботі, узгоджується система позначень. Дається загальна характеристика і обґрунтування проблеми, вирішенню якої присвячена робота. Базові поняття стосуються реляційної теорії баз даних і, зокрема, таких понять як обмеження та залежності даних. Термінологія та визначення використовуються у формулюванні, прийнятому в роботах Кодда, Фейджина, Бірі, Делобеля. Другою групою понять, які використовуються в роботі є поняття пов'язані з комп'ютерним представленням знань, базами знань і технологіями штучного інтелекту. Подаються визначення основних способів представлення і обробки знань, структури баз знань і підходи до їх створення.

Проблема, що вирішується в даній роботі стосується способів поєднання реляційного підходу до баз даних з технологіями штучного інтелекту. В більшості існуючих сучасних інформаційних технологій та систем не використовується широкий спектр властивостей та можливостей, які надає застосування апарату залежностей. Це призводить до механізації процесів проектування та використання баз даних обмежує їх інтелектуальні і функціональні можливості, звужує коло застосування систем на основі баз даних реляційного типу. Причини цього автор вбачає у відсутності єдиного підходу до поняття залежностей в реляційних базах даних, неоднозначності у визначенні їх семантики та розумінні ролі і місця залежностей, великій кількості видів і класів залежностей та способів їх застосування, складності процесів маніпулювання та підтримки систем залежностей. Вирішення цілого ряду проблем пропонується на основі застосування для обробки залежностей технологій штучного інтелекту, що робить доступним та можливим використання великих переваг, які дають залежності, в роботі з базами даних. Крім того вирішення даної проблеми може стати основою для принципово нового підходу до

побудови засобів управління даними та інформаційних систем.

У другому розділі проводиться аналіз ряду методів формального маніпулювання залежностями в реляційних базах даних. Метою проведення аналізу є визначення закономірностей, спільних рис та відмінностей в різних типах залежностей даних і способах їх обробки. Отримані в результаті аналізу узагальнення використовуються для вироблення пропозицій та рішень по проблемі створення єдиного підходу і методу обробки та опису залежностей даних як форми знань. В роботі розглядаються аналітичні методи формального маніпулювання залежностями:

- метод булевого перетворення залежностей даних;
- алгебраїчний метод на основі апарату решіток;
- логіка першого порядку;
- реляційні числення.

Графічні та табличні методи в роботі не розглядаються, оскільки складають принципово інший напрямок досліджень.

Метод булевого перетворення залежностей зводиться до наступних формалізацій:

- кожному імені атрибута відношення ставиться у відповідність деяка логічна змінна, тобто вводиться однозначне відображення множини атрибутів відношення U на певну множину логічних змінних X

$$F: U \rightarrow X;$$

- кожній множині атрибутів співставляється кон'юнкція відповідних змінних;
- кожному виразу функціональної залежності - булева імплікація змінних;
- в якості булевого еквівалента виразу функціональної залежності виду $X \rightarrow Y$ застосовується проста імпліканта виду $x \cdot y'$, яка називається функцією булевого перетворення - btf , де x і y , відповідно представляють ліву і праву частини залежності:

$$btf(\alpha) = (x \rightarrow y)' = (x' + y)' = x \cdot y'.$$

- системі функціональних залежностей відповідає булева функція, представлена набором імплікант, які є булевим перетворенням виразів функціональних залежностей, що входять в неї:

$$f(x) = BTF(\Sigma) = \bigvee_{\alpha \in \Sigma} btf(\alpha).$$

Аналогічно визначається булева перетворююча функція для багатозначної залежності виду $X \leftrightarrow Y|Z$:

$$\text{btI}(X \leftrightarrow Y|Z) = (x \Rightarrow (y+z))' = (x' + y+z)' = xy'z'.$$

Аксиоматика функціональних, багатозначних залежностей, а також правила їх взаємного виводу мають свої аналоги в термінах булевих функцій, що показується в роботі. Булевське перетворення функціональних та багатозначних залежностей дозволяє формальний розв'язок таких задач як пошук ключів, логічний вивід залежностей, перевірка входження залежностей в систему тощо.

Алгебраїчний метод базується на подібності властивостей функціональних залежностей з відношенням часткового порядку, а багатозначних з відношенням рівності в часткових решітках. Якщо R відношення на множині атрибутів U , то для нього може бути визначена функція розбиття $e: 2^U \rightarrow \Pi(R)$, така що для множини $X \subseteq U$

$$e: X \rightarrow e(X) = \{(t_1, t_2) \in R \times R \mid t_1[X] = t_2[X]\}.$$

Множина образів функції відображення $e(X) = \text{Im}(e)$ називається реляційною решіткою. Якщо у ній встановлене відношення порядку $e(X) \leq e(Y)$ то це відповідає функціональній залежності $X \rightarrow Y$ у відношенні R .

Умова існування багатозначної залежності в термінології решіток формулюється як умова декомпонованості двох підмножин U_1 і U_2 множини атрибутів U :

$$e(U_1 + U_2) = e(U_1) \cdot e(U_2) = e(U_2) \cdot e(U_1) = e(U_1) + e(U_2).$$

За допомогою засобів апарату теорії решіток можуть бути сформульовані також умови існування узагальненої багатозначної залежності, з'єднувальної залежності, умови ациклічності схем відношень, а також вирішуватись проблеми логічного виводу залежностей, побудови базису системи залежностей, входження залежностей в систему.

В основу методу маніпулювання даними, за допомогою апарату логіки першого порядку покладений підхід запропонований для функціональних, багатозначних, взаємних і надалі узагальнений на інші типи залежностей даних. Зміст його полягає в тому, що готові формальні системи різноманітних типів залежностей даних можуть бути замінені, без втрат для їх властивостей, виразами теорії першого порядку.

Метод представлення залежностей даних в формі речень першого порядку придатний до застосування для будь яких інших відомих типів залежностей. Вводиться загальне поняття "залежність даних", яке визначається як речення першого порядку вигляду:

$$(\forall x_1 x_2 \dots x_n)((A_1 \wedge A_2 \wedge \dots \wedge A_n) \Rightarrow \exists y_1 y_2 \dots y_r (B_1 \wedge B_2 \wedge \dots \wedge B_r)),$$

де кожна з A_1, A_2, \dots, A_n є реляційною формулою,

кожна з B_1, B_2, \dots, B_r – реляційною формулою або рівністю,

x_1, x_2, \dots, x_n та y_1, y_2, \dots, y_r – індивідні змінні, що позначають значення атрибутів певних відношень.

Порівняно з апаратом логіки першого порядку реляційне числення є значно компактнішим засобом, оскільки в ньому застосовуються лише інструментарій відношень та операцій над ними.

Автор пропонує ряд власних формул реляційного числення із змінними-кортежами, які описують загальну формулу відношення, на якому підтримується певний тип залежності даних. Зокрема сформульовані умови підтримки функціональної, багатозначної, з'єднувальної, транзитивної, шаблонної залежностей та зроблені узагальнення на випадок підтримки довільного типу залежності. Для побудови формул, які описують умови існування залежностей даних в якості базового пропонується вираз виду

$$\{t_1, t_2, \dots, t_n \mid (\forall t_1 t_2 \dots t_n) \psi(t_1, t_2, \dots, t_n)\},$$

де t_1, t_2, \dots, t_n – змінні кортежі однакової розмірності і структури,

$\psi(t_1, t_2, \dots, t_n)$ – формула, $(\forall t_1 t_2 \dots t_n) \psi(t_1, t_2, \dots, t_n) \equiv (\forall t_1)(\forall t_2) \dots (\forall t_n)$.

В загальному випадку формул відношення на якому підтримується залежність даних має вигляд

$$\{t_1, t_2, \dots, t_n \mid (\forall t_1 t_2 \dots t_n) R(t_1) \wedge R(t_2) \wedge \dots \wedge R(t_n) \wedge \psi(t_1, t_2, \dots, t_n)\}.$$

Пропонується загальна форма виразу числення із змінними на доменах, який задає умови існування на відношенні певної залежності даних:

$$\{x_1 x_2 \dots x_n \mid L(x_1, x_2, \dots, x_n)\},$$

де x_1, x_2, \dots, x_n – змінні на доменах,

$L(x_1, x_2, \dots, x_n)$ – першопорядкова форма відповідного виразу.

В третьому розділі розробляється методика представлення залежностей даних як форми знань. Проводиться дослідження та природи залежностей в рамках якого даються пропозиції автора по способах визначення залежностей, аналізу їх формальних особливостей та семантики. В рамках дослідження природи залежностей даних виконана класифікація способів виникнення та визначення залежностей. Розглядаються три шляхи виникнення залежностей даних, і відповідно, три способи їх визначення:

- на основі аналізу проблемної області, її особливостей, зв'язків

та фактологічної структури:

- спосіб що базується на формулюванні обмежень, які виникають і вводяться при проектуванні схеми бази даних на основі вивчення потреб та вимог до бази даних;

- спосіб утворення залежностей як властивостей інформаційного наповнення відношень реляційної бази даних.

Розробляється класифікація та характеристика шляхів визначення залежностей даних в межах кожного із способів, формулюються необхідні та достатні умови існування залежностей на кортежах відношень, наводяться алгоритми та схеми процесів утворення залежностей.

Пропонується для визначення достатньо повної та достовірної системи залежностей реляційної бази даних, оптимальної з точки зору її застосування при проектуванні та експлуатації, комплексне застосування запропонованих підходів. Система залежностей бази даних - Σ_{db} при цьому може бути одержана як

$$\Sigma_{db} = \Sigma_r \cup \Sigma_c \cup \Sigma_e,$$

де Σ_r - система залежностей, утворених на основі аналізу предметної області бази даних;

Σ_c - система залежностей, утворена як обмеження проектування;

Σ_e - система залежностей, породжених інформаційним наповненням бази даних.

Суттєвим фактором представлення залежностей даних як форми знань є їх зв'язок з Хорнівськими виразами. Подібність до Хорнівських виразів найкраще прослідковується у випадку функціональних залежностей. В булевському представленні вираз функціональної залежності може бути зведений до Хорнівського диз'юнкта $x_1 \vee x_2 \vee \dots \vee x_m \vee y'$. Цей факт, зокрема, використовується для побудови алгоритмів логічного виводу залежностей в булевому вигляді. Представлення різних типів залежностей у вигляді формул першого порядку показує, що широкий клас, так званих, повних залежностей описується реченнями виду

$$(\forall x_1 x_2 \dots x_m) ((A_1 \wedge A_2 \wedge \dots \wedge A_n) \rightarrow B),$$

які можуть зводитись до Хорнівських. Перехід до клаузальної нотації шляхом опущення квантифікованих змінних і заміни імплікації дає Хорнівське речення виду $(\neg A_1 \vee \neg A_2 \vee \dots \vee \neg A_n \vee B)$. В загальному випадку довільна залежність даних може бути зведена до форми

$$Q_1 x_1 Q_2 x_2 \dots Q_n x_n (N_1 \wedge N_2 \wedge \dots \wedge N_m),$$

де $Q_i, i=1, 2, \dots, n$ - квантор,

H_1, H_2, \dots, H_n - Хорнівські речення.

Хорнівські властивості залежностей відображають факт існування зв'язків між атрибутами, обумовлених підтриманням певного типу залежності на інтенціоналі відношення бази даних.

Семантика ряду типів залежностей даних таких як функціональні та багатозначні є достатньо прозорю і природною. В той же час семантика ряду інших типів залежностей не має прямих аналогій в предметних областях. Пропонується для визначення семантики залежностей даних враховувати два аспекти: природу - особливості реального світу, пов'язані з її існуванням в базі даних, і структурну семантику - набір властивостей відношень, які забезпечуються підтриманням залежності певного типу на цих відношеннях.

В основі природного фактору семантики залежності лежить поняття зв'язку. В цьому розумінні залежність є обмеженням, накладеним на відношення виходячи з усталених існуючих форм і співвідношень. Так функціональна залежність є реалізацією направленою зв'язку типу "один до одного", багатозначна - двох направлених зв'язків типу "один до багатьох", транзитивна - двох направлених зв'язків типу "один до одного", взаємна - двох ненаправлених типу "один до багатьох" і т.д.

Під структурною семантикою розуміється такий спосіб побудови відношень бази даних, який впливає з існування на ньому залежності певного типу. Найчастіше вживаним структурним аспектом залежностей даних є форма безвтратних декомпозицій. Така декомпозиція, в свою чергу забезпечується структурою і вмістом кортежів відношення. Тобто між природним і структурним факторами семантики залежностей існує прямий зв'язок.

Аналіз природи залежностей даних, особливостей їх представлення та інтерпретації дозволяє обґрунтувати існування ряду формальних і природних аналогій між поняттями залежностей в реляційних баззах даних та знаннями, які дозволяють розглядати залежності як специфічну форму знань.

1. Залежності, як і знання описують узагальнені якісні, процесуальні і структурні характеристики певного класу об'єктів.
2. Залежності характеризують наявність стійких природних зв'язків між певними об'єктами та їх атрибутами.

3. Залежності даних мають семантику, яка в певних інтерпретаціях може сприйматись як знання про процеси та форми, що підтримуються в базі даних.

4. Залежність в реляційній базі даних несе інформацію про причинно-наслідковий зв'язок між фактом її існування та результатами, які можуть бути отримані на основі застосування властивостей залежності.

Пропонується задавати інтерпретацію залежності в реляційній базі даних тетрадою $\langle \text{dep}, \kappa, \sigma, I \rangle$.

де dep - вираз залежності даних, κ - область інтерпретації виразу залежності - відношення бази даних, σ - множина правил побудови та обробки відношень, що співставляються даному виразу залежності, I - функція трансляції, що відображає множину властивостей виразу залежності даного типу P_{dep} у множину σ - правил обробки відношення, на якому визначена залежність.

Перелічені фактори дозволяють в певних випадках розглядати і представляти залежності як специфічну форму знань про проблемну область, структуру бази даних, способи її побудови та використання.

Знання, утворені на основі залежностей даних, мають ряд специфічних рис, обумовлених проблемною областю, в якій вони визначаються та застосовуються, способами продукування, методами і формам представлення.

В даному випадку об'єктом знань виступає реляційна база даних і її структурні компоненти та особливості. По відношенню до реляційної бази даних як об'єкта знання, утворені на основі залежностей є інтенціоналом, визначеним над даними. Таким чином, даний клас знань утворює над об'єктом - базою даних реляційного типу - своєрідну надструктуру, яку будемо називати метазнаннями про реляційну базу даних, а базу знань, яку вони утворюють базою метазнань реляційної бази даних. Оскільки сама база даних є інформаційною моделлю певної проблемної області реального світу, то метазнання по відношенню до такої проблемної області є вторинними, як явище, що продукується об'єктом, який в свою чергу є похідним від іншого.

Суб'єктом метазнань реляційної бази даних є проектувальники, Адміністратор, користувачі бази даних і програмні комплекси підтримки та управління даними інформаційних систем. Специфіка знань випливає з відносин між об'єктом та суб'єктами знань. В даному випадку має місце безпосередній вплив споживача на предметну область знань. Проектувальники, Адміністратор безпосередн формує

структуру, вимоги та обмеження для бази даних, впливають на зміст та форму її метазнань. Користувач в процесі роботи, застосовуючи метазнання, виконує над базою даних певні операції, які можуть вплинути на дані та параметри бази даних в цілому. Таке взаємодія представляється схемою:

об'єкт знань → метазнання → суб'єкти знань → об'єкт знань,

згідно з якою кожен суб'єкт - споживач знань - через свій вплив на предметну область і об'єкт знань одночасно виступає в ролі джерела і продуцента знань.

Характерною особливістю знань, утворених на основі залежностей даних є спосіб продукування. На відміну від загальноприйнятих методів - навчання, експертиза, індуктивний вивід, розпізнавання образів, тощо, в роботі для отримання спеціальних знань пропонується застосувати ізоморфне перетворення певного прототипу до відповідної форми представлення знань. Як прототип знань застосовуються залежності даних в сукупності з властивостями, якими вони володіють.

Запропонований підхід до залежностей реляційної бази даних як форми знань дає можливість перейти від чисте аналітичного їх застосування до створення бази метазнань реляційної бази даних на основі системи залежностей. База метазнань дозволяє використовувати властивості залежностей даних безпосередньо у процесах управління та обробки реляційних баз даних. Знання, які входять до бази метазнань утворюються на основі залежностей у відповідності до їх семантики і є системою фактів та правил. Фактами є предикати залежностей даних виду $dep(U_1, U_2, \dots, U_n)$, де U_1, U_2, \dots, U_n - підмножини множини атрибутів відношення. Під правилами розуміється формалізований опис способів побудови і обробки відношень бази даних у відповідності із системою визначених на них обмежень. До них належать:

- 1) правила виводу типів залежностей, які визначені на відношеннях бази даних $M=(M^k)$, де M^k - набір аксіом виводу залежностей типу k ;
- 2) правила декомпозиції відношень згідно з виразами залежностей $D^k=(D_k^*)$, де D_k^* - схема декомпозиції відношення згідно з виразом залежності типу k ;
- 3) множина загальних формул залежностей різних типів - $\Phi=(\phi(dep^k))$, де $\phi(dep^k)$ - загальна формула залежності типу k ; під загальною формулою залежності будемо розуміти структуру та співвідношення

елементів кортежів відношення, на якому підтримується залежність заданого типу.

4) правила оновлення відношень згідно з вимогами існування на них залежностей різних типів.

В четвертому розділі розробляється загальна структура бази метазнань реляційної бази даних, формат та способи комп'ютерного представлення метазнань, утворених на основі залежностей даних, принципи та схема застосування метазнань, чотановка основних задач управління реляційними базами даних за допомогою метазнань.

База метазнань являє собою комп'ютерну реалізацію виразів залежностей та їх властивостей у вигляді знань. Структура та вигляд бази метазнань визначаються особливостями середовища реалізації, однак функціональна структура не залежить від засобів реалізації і включає такі компоненти: базу фактів, систему баз правил. База фактів призначається для зберігання виразів конкретних залежностей, які підтримуються на відношеннях певної реляційної бази даних. База правил представляє наведені вище знання, які отримані на основі залежностей даних. Кожна компонента бази правил реалізує певну групу правил, яка описує певні процеси і способи застосування властивостей залежностей. Структурний поділ бази правил обумовлюється різноманітністю функцій і задач, які вирішуються за допомогою залежностей даних, а також різними напрямками їх застосування. Модульний принцип побудови бази знань, дозволяє створити на її основі підсистему з гнучкою структурою і конфігурацією, яка легко адаптується до існуючих умов та вимог.

База метазнань включає до свого складу дві функціональні компоненти – спеціальну і загальну. Спеціальною компонентою – є база фактів, що описує конкретні предикати залежностей, які підтримуються на відношеннях бази даних. Загальна компонента складає базу правил, що визначаються для різних типів залежностей даних і не пов'язані зі структурою відношень бази даних. Пропонується використовувати реляційний підхід для побудови бази метазнань. Для цього розроблений формат внутрішнього представлення метазнань. Основні принципи представлення фактів є такими. Кожна вираз залежності розглядається як група множин атрибутів, а співвідношення між цими множинами визначається типом залежності. Крім того суттєвими показниками є ім'я відношення та імена атрибутів для яких визначається залежність.

Для представлення залежностей одного відношення бази даних будується одне відношення бази фактів. Це дає можливість організувати її за модульним принципом. Для реалізації запропонована наступна структура відношень бази фактів:

- ідентифікатор виразу залежності;
- ідентифікатор типу залежності;
- номер елементу залежності;
- атрибути, що входять в елемент залежності.

Входження атрибута в певну складову частину залежності задається значеннями "true" логічних елементів у відповідному кортежі. Вираз залежності зображається множиною кортежів із спільним ідентифікатором. В загальному випадку кількість відношень які реалізують базу фактів відповідає кількості відношень бази даних. Крім цього будується відношення відповідності складових бази даних і бази фактів.

База правил також реалізується за реляційним принципом. Атрибути кожного відношення бази правил діляться на дві групи: група посилок (антецедент) і група висновків (сукцедент). Програмно реалізується механізм трансляції правил загального виду в конкретні вирази у відповідності з заданими фактами залежностей даних.

Підсистема управління реляційною базою даних за допомогою бази метазнань на основі залежностей виконується у вигляді відкритого програмно-інформаційного комплексу, який функціонує в середовищі СУБД реляційного типу. Використання комплексу передбачається шляхом включення окремих його компонентів до складу інформаційних систем, а також шляхом автономного застосування на етапах проектування баз даних, розробки інформаційних систем, АРМ тощо. Функціонально до складу підсистеми входять:

- база метазнань;
- засоби продукування та підтримки знань на основі залежностей;
- засоби контролю коректності виразів залежностей та знань;
- засоби підтримки семантичної та структурної цілісності бази даних за допомогою метазнань;
- засоби реалізації декомпозиційних та синтетичних властивостей залежностей даних.

Структурно підсистема реалізується з врахуванням особливостей середовища СУБД та загальних принципів запропонованого підходу. Основними структурними елементами є база знань, побудована згідно

вимог та можливостей середовища СУБД, та набір програмних засобів. Програми, призначені для реалізації певної групи функцій об'єднуються в інтегровані програмні модулі. Кожен з цих модулів являє собою функціонально закінчену одиницю і може використовуватись самостійно сумісно з базою знань. Ряд програмних модулів, що реалізують відповідні функції, можуть використовуватись в якості компонентів програмного забезпечення інформаційних систем. Засоби, які реалізують функції проектування передбачають автономне використання. Для цього до їх складу включаються засоби управління та діалогового інтерфейсу.

Основними функціями підсистеми управління даними на основі залежностей є вирішення наступних груп проблем:

1. Продуктування та підтримка метазнань на основі залежностей. Задачі даної групи призначаються для вводу, розпізнання, пошуку виразів залежностей та генерації знань на їх основі, підтримки діалогових режимів вводу і актуалізації знань, загального управління базою метазнань.
2. Контроль коректності знань, отриманих на основі залежностей. Вирішення задач цього напрямку включає перевірку відповідності між залежностями та структурою відношень, а також перевірка протиріч між існуючими залежностями та виразами залежностей і правилами їх використання.
3. Управління процесами актуалізації баз даних на основі властивостей залежностей. Задачі цієї групи призначаються для підтримки обмежень на відношеннях бази даних, що визначаються системою залежностей. Реалізація цих функцій передбачає виконання операцій оновлення та доповнення за певними правилами та алгоритмами у відповідності з заданим комплексом вимог.
4. Задачі зміни структури бази даних передбачають розв'язання таких проблем як збереження цілісності даних при розширенні чи скороченні схеми відношення, модифікація системи залежностей бази даних, згідно залежностей, що на ній підтримуються побудова нових структур, які впливають з існуючих та ін.

На основі запропонованого методу з метою практичної апробації та реалізації поставлених вище задач автором розроблені інструментальні програмні засоби управління реляційною базою даних за допомогою метазнань на основі залежностей.

В підсумковому розділі формулюються основні результати досягнуті в процесі виконання роботи, висновки, рекомендації по практичному застосуванню.

В додатках до роботи подаються:

- приклади реалізації бази метазнань реляційної бази даних;
- опис інструментальних програмних засобів управління та підтримки баз даних реляційного типу за допомогою метазнань, утворених на основі залежностей даних;
- акти впровадження результатів дисертаційної роботи.

Основні результати роботи.

1. Проведений аналіз і вивчення проблеми застосування на практиці такої складової реляційної моделі бази даних як система обмежень, зокрема, залежностей даних. Внаслідок відсутності подібних засобів в сучасних інформаційних системах та СУБД значно обмежуються можливості баз даних реляційного типу, знижується достовірність та надійність інформації, не гарантується цілісність даних, коректність спроектованих структур, захищеність даних від аномалій вставки/знищення.

2. Існуючі методи та засоби представлення та обробки залежностей в реляційних базах даних такі як метод булевого перетворення, алгебраїчний метод, методи застосування логіки першого порядку і реляційних числень поряд з позитивними моментами мають ряд недоліків, які функціонально та предметно обмежують, а в багатьох випадках унеможливають, можливості їх практичного застосування. Це такі недоліки як недостатня загальність по відношенню до різних типів даних, обмежене коло задач, які вирішуються в рамках методу, складність та громіздкість реалізації, відсутність належних інструментальних засобів для практичного втілення.

3. В роботі виконаний аналіз та дослідження природи, семантики та структурних особливостей залежностей в реляційних базах даних, що дало можливість обґрунтувати теоретичні основи методики представлення залежностей як форми знань, визначити способи і шляхи їх утворення, сформулювати особливості знань, отриманих на основі залежностей даних, розробити механізм, алгоритми та процедури переходу від звичайної форми залежностей даних до форми

представлення знань. Введено поняття метазнань реляційної бази даних як знань про структуру, особливості, методи проектування і обробки відношень реляційної бази даних отриманих на основі залежностей.

4. Запропонована нестандартна інтерпретація залежностей даних як форми знань, яка полягає у формулюванні виразів залежностей як деякого інтенціоналу над даними, що є певною інтегрованою інформацією. Така форма представлення залежностей включає до свого складу не тільки вираз залежності даних певного типу, а й множину правил, пов'язаних з цим типом залежностей, область та функції інтерпретації правил на конкретному відношенні бази даних. Такий підхід до залежностей даних дозволяє пов'язати їх загальні властивості із способами і процедурами обробки відношень.

5. Розроблена загальна структура та формат представлення залежностей як метазнань реляційної бази даних в середовищі СУБД реляційного типу. Запропонована структура може використовуватись як дистрибутивна оболонка бази метазнань і настраиватись на конкретні застосування в інформаційних системах та автоматизованих робочих місцях на персональних ЕОМ. В рамках запропонованих форматів розроблені структури представлення правил для найбільш широко вживаних типів залежностей даних.

6. Виконана постановка та розробка загальної структури комплексів задач управління та обробки баз даних реляційного типу за допомогою метазнань на основі залежностей даних. Зокрема визначені такі основні комплекси задач:

- продукування та підтримка знань на основі залежностей в реляційній базі даних;
- контроль коректності метазнань реляційної бази даних;
- забезпечення цілісності бази даних при оновленні;
- декомпозиція та синтез відношень бази даних допомогою метазнань.

В рамках визначених комплексів задач специфіковані процедури і програмні одиниці, що реалізують основні функції управління та обробки реляційних баз даних.

7. Розроблені методи, алгоритми і програмні та інформаційні засоби апробовані на практиці і можуть бути використані в практичних розробках автоматизованих робочих місць, інформаційних систем та технологій, які використовують бази даних реляційного типу на персональних ЕОМ сумісних з ІВМ РС та середовище dBase-сумісних

Результати, отримані в процесі виконання роботи пропонується використовувати в трьох основних напрямках:

- в науково-дослідних розробках по проблематиці проектування та використання реляційних баз даних;
- в промислових розробках автоматизованих робочих місць, інформаційних систем, засобів адміністрування баз даних на основі персональних ЕОМ та відповідних СУБД;
- в навчальному процесі для вивчення окремих розділів, лабораторних та практичних занять з курсів "Бази даних та знань", "Інтелектуальні інформаційні системи", "Системи управління базами даних", і т.п.

Основні результати, отримані в процесі виконання роботи опубліковані у виданнях:

1. Берко А.Ю., Ковальчук Л.Е., Шумков Ю.М. Базы данных неразрушающего контроля. Контрольно-измерительная техника: респ. сб. №43, Львов, 1988.- с.44-47.
2. Берко А.Ю. Автоматизированное рабочее место руководителя на базе СМ ЭВМ. Тезисы докладов IV Всесоюзной конференции "Управление большим городом". М.:НПО АСУ-Москва, 1989.- с.110-112.
3. Берко А.Ю., Пасичник В.В. Применение знаний в проектировании реляционных баз данных. Conference on Intelligent Control Systems, Sept.26-Oct.2, 1989, Varna, Bulgaria, Vol.2.- pp.362-368.
4. Берко А.Ю. Зависимости в реляционных базах данных и возможность представления знаний. Вестник Львовского политехнического института № 248, Львов, 1990.- с.11-14.
5. Берко А.Ю., Пасичник В.В. Методи визначення залежностей в реляційних базах даних. Вісник Львівського політехнічного інституту № 257, Львів, 1991.- с.12-14.
6. Берко А.Ю., Пасичник В.В. Обобщенный подход к зависимостям в реляционных базах данных. Проблемы информатики города: Сб. научных трудов / под. ред. Дробышева Ю.П.- Киев, Наукова думка, 1991.- с.248-258.
7. Реляційні моделі баз даних: Звіт по НДР (заверш.) / ЛП; керівник Пасичник В.В. - № 01880057829.-Львів, 1990.-327 с. -Відп.викон Берко А.Ю. та ін.

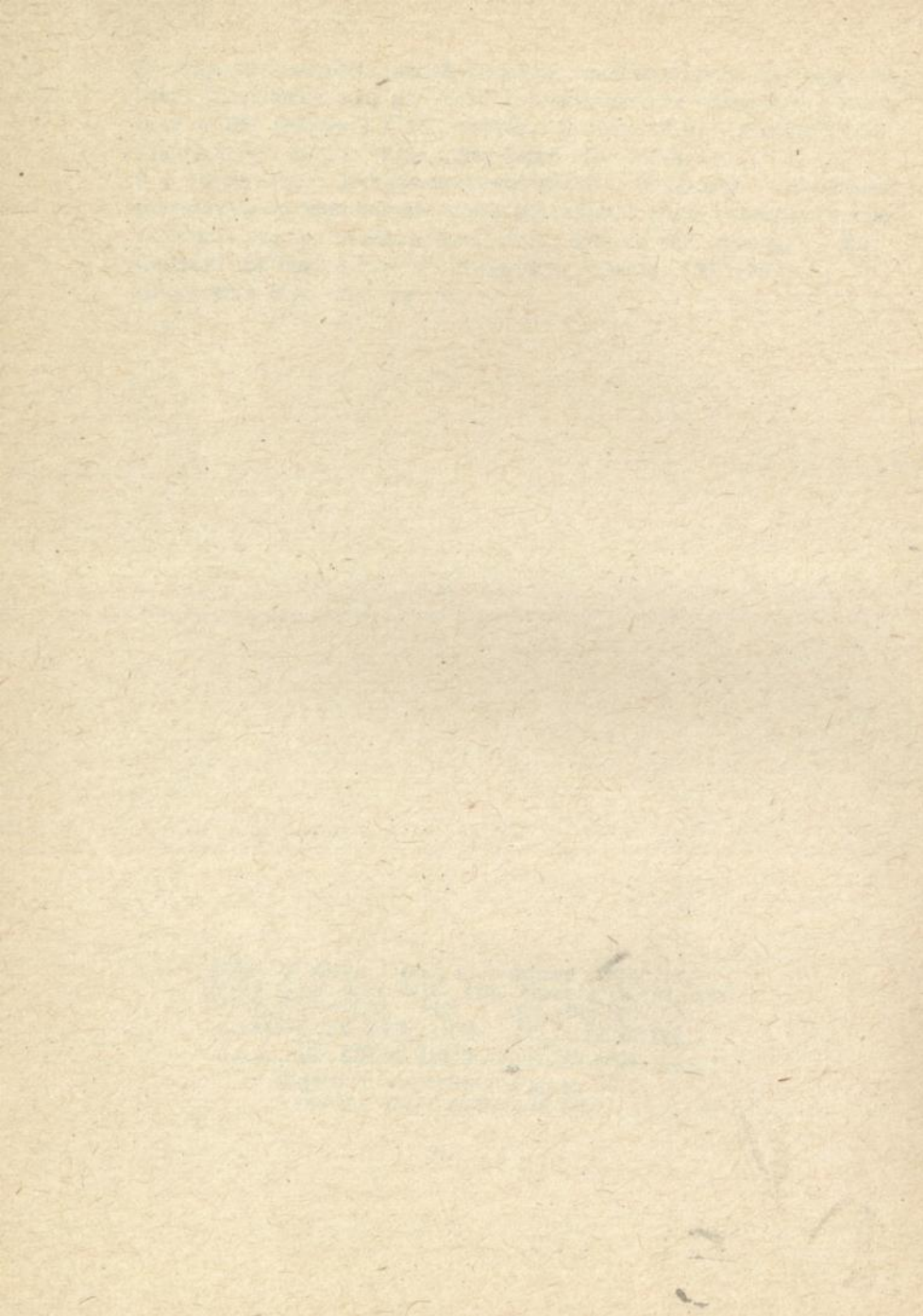
8. Розробка елементів автоматизованої системи обробки баз даних та знань реляційного типу на основі персональних ЕСМ професійного класу: Звіт по НДР (заверш.) / ЛПІ; керівник Пасічник В.В. - № 01890057488. - Львів, 1991. - 320 с. - Відп. викон Берко А.Д. та ін.

9. Створення програмно-математичного комплексу розширення функціональних можливостей систем баз даних і знань реляційного типу на основі апарату нетрадиційних логік: Звіт по НДР (заверш.) / ЛПІ; керівник Пасічник В.В. - № 01910041765. - Львів, 1991. - 257 с. - Відп. викон Берко А.Д. та ін.

Підп. до друку 11.03.93. Формат 60x84¹/16.
Папір друк. № 2. Друк. офс. Умовн. друк. арк. 1.25
Умовн. фарб.-відб. 1.15. Обл.-вид. арк. 1.12
Тираж 100 прим. Зам. 78. Безплатно

ЛПІ 290646 Львів-13, Ст. Бандери, 12

Дільниця оперативного друку ЛПІ
Львів, вул. Городоцька 286



463328

AB 28.297

AB 28.297