

КИЇВСЬКА МІСЬКА ДЕРЖАВНА АДМІНІСТРАЦІЯ
ІНСТИТУТ ПРИКЛАДНОЇ ІНФОРМАТИКИ

На правах рукопису
УДК 681.3

ЛІБІНЕЦЬ ЯРОСЛАВ ВОЛОДИМИРОВИЧ

МЕТОДИ ТА ЗАСОБИ ПРЕДСТАВЛЕННЯ ТА ОБРОБКИ
ІНФОРМАЦІЇ З ЧАСОВИМ ПАРАМЕТРОМ В РЕЛЯЦІЙНИХ БАЗАХ ДАНИХ

Спеціальність 06.ІЗ.І7 - Теоретичні основи інформатики

Автореферат дисертації на здобуття вченого ступеня
кандидата фізико-математичних наук

Київ -- 1994

АВ 30.999

Робота виконана в державному університеті
"Львівська політехніка"

Науковий керівник:

д.т.н. В.В.Пасічник

Офіційні опоненти:

д.ф.-м.н., проф. Г.Г. Цегелик (м.Львів)

к.ф.-м.н., В.В.Колінько (м.Київ)

Провідна організація - Інститут програмних систем АН України
(м.Київ)

Захист відбудеться 25 10 1994 р. о 16 год.
на засіданні Спеціалізованої Ради Д І66.01.01 в Інституті
прикладної інформатики (ІПРІН) за адресою: 252004, м. Київ, вул.
Червоноармійська 23-б.

З дисертацією можна ознайомитись в бібліотеці
інституту прикладної інформатики.

Автореферат розісланий 23 09 1994 р.

Вчений секретар

Спеціалізованої Ради

Д І66.01.01

Міхалевський П.В.

ЛНБ ім. В. Стефаника
АН України

ЛНБ України ім. В. Стефаника



00801435 (L)

AB - 30.999

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ.

Актуальність проблеми. Реляційна модель даних протягом останніх років здобула загальне визнання як зручний засіб для інформаційного моделювання дійсності. Однак, поряд з багатьма перевагами, реляційна модель в її традиційному розумінні та СКБД, побудовані на її основі мають ряд недоліків. Зокрема, виникають труднощі при моделюванні еволюції об'єктів реальної дійсності, тобто відсутні ефективні методи та засоби представлення часових параметрів. Явне введення часового контексту та його відображення в конструкціях інформаційних систем передбачає розробку спеціальних засобів, оскільки часова розмірність, часові атрибути відношень бази даних мають особливу семантику. Введення часових моделей означає необхідність збереження в базі даних значень властивостей об'єктів, які втратили достовірність. Концепція несуперечливості реляційної бази даних, яка означає недопустимість представлення в ній одночасно деякого факта і його заперечення, повинна бути розширена можливістю представлення станів бази даних, які відповідають різним дискретним моментам часу.

В результаті аналізу літературних джерел можна виділити два основні підходи до представлення часу в реляційних базах даних: введення часового параметру на рівні атрибутів та часова параметризація кортежів. Кожен з вказаних підходів має ряд невирішених питань, наприклад:

- проблема надлишковості представлення даних та часової параметризації багатоатрибутих властивостей;

- реалізація навігації на рівні історії об'єкта та на рівні еволюції предметної області.

Отже проблема розширення традиційної реляційної моделі

для адекватного маніпулювання часовою інформацією є актуальною як в теоретичному, так і в прикладному аспекті.

Мета дисертаційної роботи полягає в розробці теоретичних та прикладних аспектів проектування, представлення та обробки еволюції об'єктів реальної дійсності в базах даних, побудованих на основі відповідного розширення реляційної моделі даних. Виходячи з цього, дослідження проводились в наступних основних напрямках:

- аналітична оцінка основних результатів в області реляційних моделей даних, розширених для представлення інформації з часовими параметрами;

- інтеграція різноманітних підходів до моделювання часу в інформаційних системах та побудова часової реляційної моделі даних. Розробка уніфікованої системи основних понять, визначень та позначень;

- побудова розширеної реляційної алгебри, мови опису даних та мови запитів для часової реляційної моделі даних (ЧРМД), як теоретичного фундаменту для побудови СКБД на її основі;

- побудова дослідницького прототипу СКБД та аналіз можливих практичних застосувань БД на основі ЧРМД.

Наукова новизна полягає в наступному:

- розроблена та представлена теорія моделювання семантики інформації з часовими параметрами в базах даних;

- введено поняття часових нормальних форм, досліджено співвідношення між ними та побудовано алгоритми нормалізації;

- сформульовано принципи побудови часових реляційних алгебр, досліджені властивості та умови коректності розширень реляційних операторів;

- побудовані SQL-подібні мови опису та маніпулювання

даними для ЧРМД, розроблені принципи побудови дослідницького прототипу СКБД на основі ЧРМД.

Методи досліджень базуються на основних положеннях теорії відношень, формальної алгебри, математичної логіки, часової логіки, математичної теорії баз даних.

Достовірність основних наукових результатів забезпечується строгим математичним доведенням сформульованих тверджень та теорем, шляхом розв'язання практичних задач та їх впровадженням в конкретних інформаційних системах.

Практична цінність дисертації. Результати дисертаційних досліджень перевірялись на практиці в процесі виконання договірних та держбюджетних робіт. Теоретичні та методологічні результати, розроблені алгоритми і програми можуть використовуватися:

- при побудові узагальненої методики проектування БД;
- при розробці промислових СКБД на основі ЧРМД;
- в навчальному процесі при підготовці спеціалістів за спеціальностями прикладної математики, АСУ, систем штучного інтелекту;
- при розробці прикладних інформаційних систем для промислової експлуатації.

Реалізація результатів досліджень. В процесі впровадження основних результатів дисертаційних досліджень в конкретних інформаційних системах розроблена узагальнена методика проектування ЧБД, елементи прототипу СКБД на основі ЧРМД з автономними модулями розширення бібліотеки базових процедур для маніпулювання часовими параметрами та ведення архівів. Ці результати використовувалися при виконанні держбюджетних та господарських робіт для Бориславської

Фабрики нетканних матеріалів, Івано-Франківського заводу КСМ, заводу збірних конструкцій №769 (м. Львів), Бориславського експериментального ливарно-механічного заводу, Львівського домобудівного комбінату, автомобільної бази №2440 (м. Львів), спільного українсько-німецького підприємства "ІСМ" по виробництву бурового інструменту (м. Дрогобич) та ін.

Апробація роботи. Результати дисертаційної роботи доповідались на всесоюзних школах-семінарах з баз даних та знань (м. Львів - Славсько), на щорічних наукових конференціях професорсько-викладацького складу Львівського політехнічного інституту (Ф-т комп'ютерної техніки та інформаційних технологій, 1990-1993 роки), на міжнародній конференції з інформаційних технологій та систем (Львів, 4.10-9.10.1993 р.), на семінарі "Новые информационные технологии и инструментально-технологические средства поддержки принятия решений" (Ін-т кібернетики ім.В.М.Глушкова АН України)

Структура та об'єм роботи. Дисертація складається з вступу, чотирьох розділів, висновку, списку основної використаної літератури та додатку. Робота містить III сторінок основного тексту, список літератури із 81 найменувань, 5 сторінок додатку. Загальний об'єм дисертації 124 сторінки.

У вступі обґрунтована важливість та актуальність питань представлення інформації з часовим параметром в базах даних реляційного типу. Проведено короткий аналіз досліджень в області часових реляційних баз даних. Сформульована мета досліджень, наукова новизна, основні положення, що виносяться на захист.

Перший розділ присвячений питанням моделювання змінної

в часі інформації в базах даних. Розглядається поняття часового куба даних, досліджуються його властивості. Наведено формалізми структурування часових доменів. Проведена типізація часових елементів. Досліджено різні види операцій над ними.

В якості фундаментального концепта моделювання залежних від часу даних використовується поняття події. Проведено детальний аналіз цього поняття, зроблено визначення атомарної події, елементарної події та події, показано їх взаємовідношення.

Означення. Еволюція предметної області (об'єкта) – це послідовність станів предметної області (об'єкта). Предметна область (об'єкт) переходить з попереднього стану в наступний в результаті здійснення події (елементарної події). Аналогічно виконання атомарних подій формує еволюцію атрибута об'єкта. До моделювання еволюції предметної області поставлено вимоги часової повноти, часової щільності та часового ізоморфізму.

Для представлення атомарних подій використано трійки даних (o, u, t) , де o – об'єкт, u – значення деякої характеристики цього об'єкта, t – часовий момент, що визначає час реалізації даної події. Сукупність триплетів з рівними часовими елементами представляє деяку подію. Зафіксувавши об'єкт та атрибут, отримуємо сукупність триплетів, впорядкованих згідно з значеннями часових елементів. Назвемо таку послідовність простою часовою послідовністю (ПЧП). Проста часова послідовність об'єкта служить для представлення еволюції атрибута об'єкта. Сукупність ПЧП всіх атрибутів об'єкта називається часовою послідовністю (ЧП) цього об'єкта. ЧП об'єкта служить для

представлення еволюції об'єкта. Часові послідовності всіх об'єктів предметної області формують сукупність часових послідовностей (СЧП). СЧП представлятиме еволюцію предметної області.

Для моделювання часових даних, враховуючи їх трьохмірність, на концептуальному рівні зручно використовувати просторову побудову - часовий куб даних. Одну розмірність куба даних становлять об'єкти предметної області O , другу розмірність - атрибути об'єктів U , третя розмірність куба - часова вісь, на яку відображаються часові елементи. Часовий куб даних адекватно відображає сукупність часових послідовностей.

Показано взаємозв'язок різних компонент та зрізів куба з часовими послідовностями, досліджено їх семантику. Описано властивості синхронності атрибутів об'єктів, відмінність атрибутів по способу виведення проміжних значень. Виділено три семантичні різновиди часової параметризації даних: час дійсності події, час транзакції та час, визначений користувачем. Розглянуто види підтримки актуальності часового куба та способи задання його часового інтервалу існування.

При розробці методів представлення часових елементів в інформаційних системах базовим є поняття структурування.

Розглянемо множину $S = (\dots, s_0, s_1, \dots, s_n, \dots)$, на якій задано відношення порядку $<$, тобто $\forall s_i, s_j \in S$ або $s_i < s_j$, або $s_i > s_j$, або $s_i = s_j$.

Якщо множина S є скінченною, то можна визначити наступні оператори :

$$\text{first}(S) = s_i \in S, \forall s_j \in S: s_i \leq s_j$$

$$\text{last}(S) = s_i \in S, \forall s_j \in S: s_i \geq s_j$$

Скінченим закритим інтервалом множини S є множина S' , яка задовільняє наступним умовам:

- 1) S' - скінченна;
- 2) $S' \subseteq S$;
- 3) $\forall s \in S : \text{first}(S') \leq s \leq \text{last}(S') \Rightarrow s \in S'$

Означення. Нехай $(S_0 = \{s_{0_0}, s_{0_1}, \dots, s_{0_n}\})$ множина в відношенням порядку $<$. Тоді структуруванням множини S_0 називається пара $\langle S_1, f_0^1 \rangle$, де

1. $S_1 = \{s_{1_0}, s_{1_1}, \dots, s_{1_k}\}$, $k \leq n$
2. $\forall x \in S_0 \Rightarrow x \notin S_1$
3. $f_0^1: S_1 \rightarrow 2^{S_0}$ є відображенням, яке володіє властивостями:

- a) f_0^1 відображає кожен елемент S_1 в скінченний закритий інтервал множини S_0 .
- б) $\cup f_0^1(S_1) = S_0$.

Згідно з структуруванням, множина S_0 розбивається на суміжні інтервали, кожному з яких відповідає єдиний елемент з S_1 , і навпаки, кожному елементу множини S_1 ставиться у відповідність єдиний інтервал множини S_0 . Множина S_0 повністю покривається даним відображенням. На множині S_1 ми можемо ввести відношення порядку $<'$ наступним чином:

$$s_{1_i} <' s_{1_j} \Leftrightarrow \text{last}(f_0^1(s_{1_i})) < \text{first}(f_0^1(s_{1_j}))$$

На базі структурування розроблено способи представлення часових моментів та досліджено питання часових доменів.

Часовий момент служить для представлення найменшої часової одиниці, що використовується в базі даних.

Крім часового моменту базовими конструкціями для представлення часових елементів в інформаційних системах є наступні типи даних: часовий інтервал, термін часу та часовий ланцюг. Часовий інтервал задається за допомогою двох моментів часу - початкового та кінцевого, і складається з

множини всіх моментів часу, які лежать між початковим та кінцевим. Початковий момент включається в інтервал, а кінцевий не включається. Термін часу використовується для задання тривалості часового інтервалу, або віддалі між двома моментами часу. Часовий ланцюг - це множина часових інтервалів, які не мають спільних точок.

Операції над часовими елементами можна поділити на наступні групи: операції відношення, часові операції, арифметичні операції, функції перетворення. Наведено опис операцій, досліджено їх повноту.

В другому розділі дається визначення часового інформаційного відношення, реляційної схеми, кретежа, часового значення атрибута, розглядаються операції на часових значеннях. Введено часові нормальні форми відношення, розроблено алгоритми нормалізації. Досліджено поняття ключа, залежностей даних в часовій реляційній моделі. Значна увага приділена властивості інформаційної рівності відношень.

Означення. Реляційною схемою часового відношення \mathcal{R} називається трійка $\mathcal{R}=(R, V, \Sigma)$, де $R \subseteq U$, U - це множина атрибутів всіх об'єктів предметної області, V - система твірних залежностей, Σ - множина залежностей даних.

В часовому відношенні значенням атрибута є не окремий елемент області допустимих значень атрибута, як в традиційній реляційній моделі, а множина елементів, кожен з яких відповідає стану атрибута протягом деякого проміжку часу.

Означення. Часовим значенням (надалі просто значенням) атрибута A є функція $f=t(A)$, визначена на часовому елементі t , така, що для кожного $te t$, $f(t)$ є елементом області

визначення атрибута A . Часовий елемент τ називається часовим доменом значення f і позначається через $\text{tdom}(f)$.

З множини залежностей, які підтримуються на часовому відношенні, автором вперше виділяється система твірних залежностей τ . На відміну від залежностей з множини Σ , твірні залежності визначаються співвідношенням атрибутів, а не їх значень. Система твірних залежностей задається розбиттям схеми R на класи атрибутів та заданням виділеного ключа K відношення r . Твірні залежності можуть прямо не залежати від семантики предметної області, а визначатись проектувальником в залежності від потреб представлення та обробки даних. Тобто система твірних залежностей в загальному випадку є привнесеною системою залежностей.

Означення. Кортежем τ часового відношення зі схемою R називається функція над R така, що для кожного атрибута $A \in R$ $\tau(A)$ є часовим значенням A .

Означення. Інформаційним часовим відношенням (далі відношенням) зі схемою $R = \{R, \tau, \Sigma\}$ називається скінченна множина кортежів зі схемою R , що задовільняє залежностям Σ та τ .

Часове відношення можна трактувати, як послідовність скінченного числа статичних відношень. Тому доцільно ввести функцію стану часового відношення $\text{Temp}(r, t): r \times T \rightarrow \text{Rel}(R)$, яка для кожного моменту часу $t \in \text{tdom}(r)$ визначає статичне відношення $r(t)$ зі схемою R .

На часовому відношенні r підтримується функціональна залежність $X \rightarrow Y$, якщо $\forall t \in \text{tdom}(r), \forall \tau_1 \in r, \forall \tau_2 \in r$:

$$(\tau_1(X, t) \neq \tau_2(X, t)) \vee (\tau_1(X, t) = \tau_2(X, t) \wedge \tau_1(Y, t) \neq \tau_2(Y, t))$$

Конструктивне визначення залежності впливає з наступного твердження.

Твердження. Нехай γ — часове відношення з схемою R , $X, Y \subseteq R$. $X \rightarrow Y$ тоді і тільки тоді, коли $\forall \tau_1 \in \gamma, \forall \tau_2 \in \gamma$:

$$(\tau_1(X) \neq \tau_2(X)) \vee (\tau_1(X) = \tau_2(X) \wedge \tau_1(Y) = \tau_2(Y))$$

Означення. На часовому відношенні γ підтримується глобальна функціональна залежність $X \rightarrow Y$, якщо $\forall t_1, t_2 \in \text{tdom}(\gamma), \forall \tau_1, \tau_2 \in \gamma$:

$$(\tau_1(X, t_1) \neq \tau_2(X, t_2)) \vee (\tau_1(X, t_1) = \tau_2(X, t_2) \wedge \tau_1(Y, t_1) = \tau_2(Y, t_2))$$

Очевидно є властивість $X \rightarrow Y \Rightarrow X \rightarrow Y$.

Традиційне визначення ключа через функціональні залежності має ряд небажаних для часової моделі властивостей, зокрема значення ключових атрибутів може мінятися з часом. Тому такий ключ називається слабим ключем, а для ключа наводиться інше визначення.

Означення. Ключем відношення γ зі схемою R називається підмножина $K \subseteq R$, якщо виконуються наступні умови:

- 1) $\forall A \in K, |\tau(A)|$ містить єдиний елемент;
- 2) Нехай τ_1 та τ_2 є кортежі відношення γ , тоді $\forall A \in K, |\tau_1(A)| = |\tau_2(A)| \Leftrightarrow \tau_1 = \tau_2$;
- 3) $\exists X \in K$ такої, що X є ключем.

Серед специфічних типів залежностей часової моделі автором вперше виділяється такі:

- функціональні залежності з затримкою. Нехай $\tau, \tau' \in \gamma, \mu, \nu \in \text{dom}(\tau), \mu', \nu' \in \text{dom}(\tau'), \text{begin}(\nu) = \text{end}(\mu), \text{begin}(\nu') = \text{end}(\mu')$, тоді $\tau(X, \mu) = \tau'(X, \mu') \Rightarrow \tau(Y, \nu) = \tau'(Y, \nu')$

- функціональні залежності тривалості. При збереженні попередніх припущень виконується:

$$\tau(X, \mu) = \tau'(X, \mu') \wedge \text{duration}(\mu) = \text{duration}(\mu') \Rightarrow \tau(Y, \nu) = \tau'(Y, \nu')$$

Значну роль в часових базах даних відіграють залежності, пов'язані з поняттям синхронності атрибутів. Автором досліджено властивості, дано конструктивно

визначення цього поняття та його основі розроблено алгоритмів, які використовуються в процесах нормалізації часових відношень.

Означення. Нехай K ключ відношення r зі схемою R , A_1 та A_2 атрибути схеми R ; A_1 та A_2 є асинхронними $A_1 \neq A_2$, якщо і тільки якщо існують два суміжні часові моменти μ_1 та μ_2 та кортеж t , такі, що

$$1) \tau(A_1)|\mu_1 = \tau(A_1)|\mu_2 \wedge \tau(A_2)|\mu_1 \neq \tau(A_2)|\mu_2, \text{ або}$$

$$2) \tau(A_2)|\mu_2 = \tau(A_2)|\mu_1 \wedge \tau(A_1)|\mu_2 \neq \tau(A_1)|\mu_1.$$

В роботі розвинено теорію нормалізації часових відношень, яка знаходиться в зародковій стадії.

Означення. Часове відношення r перебуває в еволюційній формі, якщо і тільки якщо всі атрибути схеми R належать до одного класу розбиття.

Кортеж часового відношення в еволюційній формі представляє елементарну подію предметної області, а схема містить єдиний часовий атрибут.

Означення. Часове відношення r перебуває в еволюційній нормальній формі, якщо і тільки якщо воно перебуває в еволюційній формі та в схемі відношення не існує асинхронних атрибутів.

Вводяться також історична та історична нормальна форми часового відношення.

Розглянуто поняття інформаційної рівності, яке дозволяє в гнучкій манері оперувати інформаційним наповненням відношень.

Означення. Нехай r і s - часові відношення. Будемо говорити, що r і s інформаційно рівні $r \sim s$, якщо виконуються умови: 1) $\text{tdom}(r) = \text{tdom}(s)$

$$2) \forall t \in \text{dom}(r) \text{ Temp}(r, t) = \text{Temp}(s, t)$$

Хоча інформаційно рівні відношення не є еквівалентними, вони несуть однакове інформаційне навантаження. Відношення інформаційної рівності є відношенням еквівалентності, тому для кожного відношення r існує клас $[r]$ відношень, інформаційно рівних до r .

Нехай $r_1 \dot{=} r_2$ та $s_1 \dot{=} s_2$. Бінарний оператор \circ називається інформаційно інваріантним, якщо $r_1 \circ s_1 \dot{=} r_2 \circ s_2$. Аналогічно можна визначити інформаційно інваріантний уварний оператор.

Досліджено операції на часових значеннях атрибутів, уточнено поняття предиката на часових значеннях та операторів відношення між часовими значеннями. Наприклад, відношення рівності часових значень має ряд семантичних трактувань:

- рівні часові значення $f_1 = f_2$: $\text{tdom}(f_1 = f_2) = \{t \in (\text{tdom}(f_1) \cap \text{tdom}(f_2)) : f_1(t) = f_2(t)\}$

- повністю рівні $f_1 = f_2$: $\text{tdom}(f_1 = f_2) = \text{tdom}(f_1) = \text{tdom}(f_2)$

- рівні на спільному часовому домені $f_1 =_i f_2$: $\text{tdom}(f_1 = f_2) = (\text{tdom}(f_1) \cap \text{tdom}(f_2))$

- часові значення, рівні по даних $f_1 =_d f_2$: $|f_1| = |f_2|$

- рівні по даних та їх порядку $f_1 =_o f_2$: Нехай k - кількість елементів $|f_1|$, тоді повинні виконуватися наступні умови:

$(|f_1| = |f_2|) \wedge (\forall n, 0 < n \leq k : \text{Order}(n, f_1) = \text{Order}(n, f_2))$

- часові значення, рівні з точністю до зсуву часового домена $f_1 =_s f_2$: $\exists n : \text{Shift}(n, f_1) = f_2$

Твердження. Між різними трактуваннями відношення рівності виконуються наступні співвідношення.

1) $(f_1 = f_2) \Rightarrow (f_1 =_i f_2)$, $(f_1 =_o f_2) \Rightarrow (f_1 =_d f_2)$,

$(f_1 = f_2) \Rightarrow (f_1 =_o f_2)$, $(f_1 = f_2) \Rightarrow (f_1 =_s f_2)$

2) $(f_1 =_o f_2) \Rightarrow (f_1 =_d f_2)$

$$3) (f_1 =_{a_1} f_2) \Rightarrow (f_1 =_d f_2), (f_1 =_{a_1} f_2) \Rightarrow (f_1 =_o f_2)$$

В третьому розділі розроблено принципи та побудовано розширення операторів часової реляційної алгебри. Досліджено взаємовідношення прямих та часових розширень алгебраїчних операторів і їх зв'язок з операторами традиційної реляційної алгебри. Сформульовано властивості введених операторів, зокрема властивість інформаційної інваріантності.

Часова реляційна модель дає велику різноманітність в побудові операторних розширень. Розглянемо деякий бінарний оператор γ традиційної реляційної моделі даних.

Означення. Нехай r та s - часові відношення. Прямим розширенням оператора γ називається оператор γ' часової реляційної алгебри такий, що $\forall t \text{ tdom}(r) \cup \text{tdom}(s)$ виконується

$$\text{Temp}(r\gamma's, t) = \text{Temp}(r, t) \gamma \text{Temp}(s, t)$$

Аналогічно, прямим розширенням унарного оператора γ назвемо оператор γ' , такий що $\forall t \text{ tdom}(r)$ виконується

$$\text{Temp}(\gamma'(r), t) = \gamma(\text{Temp}(r, t))$$

Прямі розширення реляційних операторів володіють рядом бажаних властивостей, зокрема, їхня сукупність складає реляційно повну систему, але вони не дають можливості адекватно маніпулювати з еволюцією значень атрибутів і не враховують всіх особливостей реляційної схеми часового відношення. Тому для "часової" повноти слід розглянути такі розширення операторів реляційної алгебри, для яких не виконуються вимоги означення. Назвемо такі розширення часовими. Часові розширення реляційних операторів в загальному випадку не можна виразити через прямі розширення. Такий підхід до побудови часової реляційної алгебри, а також введення часових розширень реляційних операторів вводяться

автором вперше.

Формально визначені прямі та часові розширення операторів проєкції, селекції, поєднання, різні види операторів об'єднання та доповнення. Крім прямого розширення введено слабку та сильну форми операторів різниці та перетину. Зокрема, результат виконання прямого розширення оператора перетину представляється формулою:

$$rns = (\tau: \exists \tau_1 \in \tau \exists \tau_2 \in s,$$

$$\tau_1(K) = \tau_2(K) \wedge \tau = \tau_1 \uparrow \text{tdom}(\tau_1) \cap \text{tdom}(\tau_2) = \tau_2 \uparrow \text{tdom}(\tau_1) \cap \text{tdom}(\tau_2))$$

"Строгий" перетин включає в результат кортежі тільки при повній їх рівності: $rn_s = (\tau: \tau \cap s)$

"М'який" перетин включає в результат розширення рівних по ключу сумісних кортежів до спільного часового домену: $rn_s = (\tau: \exists \tau_1 \in \tau \exists \tau_2 \in s, \tau_1(K) = \tau_2(K) \wedge \tau = \tau_1 \cup \tau_2)$

Виконується властивість $rn_s \subseteq rns \subseteq rn_s$, причому рівність досягається при $R=K$.

Наводяться алгоритми обчислення результатів виконання введених операторів. Доведена наступна теорема.

ТЕОРЕМА. Для відношень в еволюційній формі різновиди кожного оператора часової реляційної алгебри співпадають.

Вводиться принципово новий оператор форми $Perfom$, який оперує відношеннями шляхом зміни твірної системи.

Нехай задано часове відношення γ з схемою $R = (A_1, A_2, A_3, \dots, A_n)$. В результаті виконання оператора $Perfom(\gamma; K; (X_1) \dots (X_m))$ отримуємо відношення γ' , яке є інформаційно-рівним відношенню γ . Виділивши ключем відношення γ' є K , атрибути формуль m класів синхронності $(X_1) \dots (X_m)$.

Виконання оператора $Perfom(\gamma; K; (X_1) \dots (X_m))$ проводиться наступним чином:

1) Для всіх кортежів τ , для яких $|\tau|$ містить більше одного елемента, проводиться розбиття $\text{tdom}(\tau) = \mu_1 + \mu_2 + \dots + \mu_n$, таке що $\forall i \ 0 < i \leq n$ виконується $|\tau| \mu_i|$ містить єдиний елемент. Розбиваємо кортеж τ на n кортежів: $\tau_1 = \tau| \mu_1$, $\tau_2 = \tau| \mu_2$, ..., $\tau_n = \tau| \mu_n$.

2) Впорядкуємо кортежі утвореного відношення r за значеннями атрибута K . Якщо існують кортежі τ_1 та τ_2 з різними K -значеннями, виконуємо операцію $\tau = \tau_1 \cup \tau_2$. Кортеж τ додаємо до відношення r , а кортежі τ_1 та τ_2 виключаємо.

3) На кожному кортежі відношення для кожного класу $X_i = (A_{i,1}, A_{i,2}, A_{i,3}, \dots, A_{i,n})$ знаходимо $\Sigma(X_i) = \Sigma(A_{i,1}) + \Sigma(A_{i,2}) + \Sigma(A_{i,3}) + \dots + \Sigma(A_{i,n})$. Нехай $\Sigma(X_i) = (t_1, t_2, t_3, \dots, t_k)$. Часове значення кожного атрибута $A_{i,j}$, $1 \leq j \leq n$ визначається на інтервалах $[t_1, t_{1,k})$, $1 \leq k$. Для кожного інтервала записуємо конфігурацію значень атрибутів класу X_i .

Принципову роль відіграє наступна теорема:

ТЕОРЕМА. Якщо оператор часової реляційної алгебри є прямим розширенням відповідного традиційного оператора, то він є інформаційно інваріантним.

В четвертому розділі розроблено мову опису та маніпулювання даними TSQL, яка є розширенням SQL на випадок відношень з часовими параметрами. Формально описано основні синтаксичні конструкції мови, введено нові семантичні типи атрибутів. Серед нововведень також є ряд часових типів даних, ключових слів мови, агрегуючих функцій. Показано зв'язок запропонованої мови з реляційною алгеброю.

Серед основних принципів побудови TSQL можна виділити такі:

- TSQL повинен бути природним розширенням SQL. 208

ЛНБ ім. В. Стефаніка
АН України

довільний запит, сформульований на SQL, повинен вірно інтерпритуватися засобами TSQL, а обробка запитів до традиційних баз даних об'ємними мовами повинна давати однаковий результат;

- додаткові конструкції TSQL, розроблені для маніпулювання часовими параметрами, повинні по замовчуванню встановлюватися в значення, що відповідають традиційним статичним даним, а також повинні бути синтаксично близькими до відповідних зворотів звичайної мови;

- часове відношення - результат запиту "живе" в базі даних, тобто підтримується його актуальність в кожен момент часу.

Мова опису даних доповнена специфічними можливостями опису відношень, зокрема введено такі ключові слова:

TIME TYPE - для задання семантичних типів часових атрибутів, які включаються в часове відношення та мають вбудований механізм підтримки;

REFRESH - задає порядок актуалізації відношення;

LIFESPAN, PERSISTENCY - описують проміжок часу, який охоплюється.

Для представлення часових параметрів введено нові типи даних: **TIMECLOCK** (часовий момент), **INTERVAL** (часовий інтервал), **SPAN** (проміжок часу).

Значно змінена семантика операцій оновлення відношень, оскільки в часових базах даних виконується вимога збереження значень атрибутів, які втратили свою достовірність. Виходячи з цього, в TSQL відсутня операція знищення кортежа в традиційному розумінні.

Блок **SELECT** містить нові конструкції **WHEN** для виконання часової селекції відношення та **AS-OF...ALONG**, яка служить

для часового обмеження / відносно вибраної часової розмірності. Потужність отриманих запитів не нижча за потужність реляційної алгебри, що показано шляхом вираження алгебраїчних операторів через запити розробленої мови.

Фраза GROUP BY доповнена ключовим словом TIME, що дозволяє формулювати складні запити з скануванням часового інтервалу заданої тривалості вздовж часової осі. Наведено алгоритм виконання запитів такого типу.

Набір агрегуючих функцій розширений агрегатами, які не мають семантичних аналогів в традиційних мовах запитів, наприклад:

ACC - функція "акумулявання" для отримання наростаючих значень числових атрибутів за певні проміжки часу;

P_AVG - обчислення середнього терміну сталості значень;

I_MAX - повертає найбільший часовий інтервал дійсності значень атрибутів;

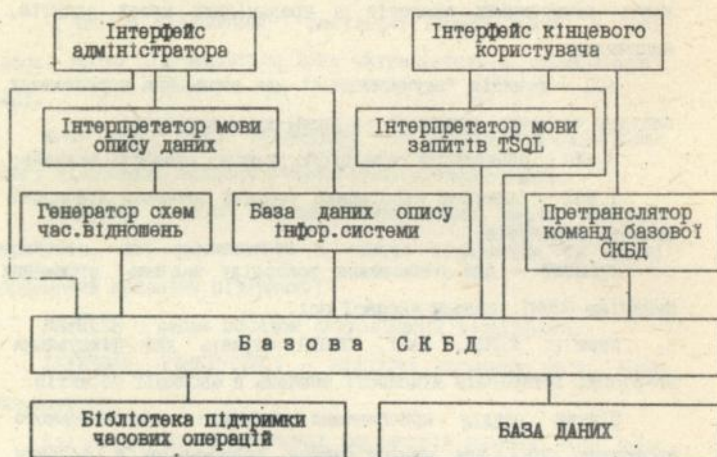
T_COUNT - для обчислення розподілу значень, отриманих функцією COUNT, вздовж часової осі.

Функції FIRST, LAST, ORDER служать для фіксування потрібних інтервалів дійсності значень в еволюції об'єктів.

П'ятий розділ присвячений побудові дослідницького прототипу СКЕД для маніпулювання інформацією з часовим параметром на базі промислової СКЕД dbase-сімейства. Наведено опис мови представлення даних прототипу, яка включає механізм семантичного контролю даних. Розглянуто питання представлення та підтримки часового відношення через систему відношень базової СКЕД.

При розробці дослідницького прототипу створено автономну бібліотеку функцій обробки часових типів даних, в

якій реалізовано базовий набір часових предикатів, операторів, арифметичних функцій та функцій перетворення. Значну увагу присвячено розвитку методики створення інформаційних систем для динамічних та історичних областей застосування на базі традиційних СКБД. Прототип СКБД включає програмний модуль забезпечення зручного користувацького інтерфейсу для роботи з інформацією, змінною в часі. Виклад розділу проілюстровано прикладами з прикладних розробок автора.



мал. I.

Функціональна схема дослідницького прототипу СКБД зображена на мал. I.

В висновках сформульовані основні результати, отримані в період виконання дисертаційних досліджень. Запропоновані та проаналізовані можливі напрямки подальших досліджень.

В додатку наведені документи, які підтверджують впровадження результатів дисертації.

ОСНОВНІ РЕЗУЛЬТАТИ ДИСЕРТАЦІЇ

1. Запропонована та обґрунтована узагальнена методика моделювання змінних в часі предметних областей. Проаналізовано способи представлення та типізації часових елементів.

2. Розроблено уніфікований підхід до формалізації часового інформаційного відношення. Введено поняття часових нормальних форм, проаналізовані питання залежностей даних в часовій реляційній моделі.

3. Запропоновано принципи побудови та розроблено часову реляційну алгебру. Досліджено властивості часових реляційних операторів.

4. Побудовано мови опису та маніпулювання даними для часової реляційної моделі даних, які є розширеннями SQL, представлено зв'язок основних конструкцій мови маніпулювання даними з операторами часової реляційної алгебри.

5. Розроблені принципи, методи, алгоритми побудови програмного комплексу дослідницького прототипу СКБД на базі часової реляційної моделі даних.

Основні результати дисертації викладені в роботах:

1. Любінець Я.В. Моделювання часочної інформації в базах даних.- Вісник Львівського політехнічного інституту, збірник "Комп'ютерна техніка".

2. Любінець Я.В. Временная реляционная модель данных.- Тез. док. на семинаре "Новые информационные технологии и инструментально-технологические средства поддержки принятия решений" пос. Кацивели, 21-26 дек., - Киев, Ин-т кибернетики им.В.М.Глушкова АН Украины, 1992.

3. Розробка та аналіз методів і засобів обробки ненормалізованих реляційних баз даних і знань (розробка дослідницького прототипу СКБД на основі розширеної ненормалізованої реляційної моделі даних): Звіт про НДР (заверш.) ЛПІ; Керівник Пасічник В.В. - ОІ9U0061893. - Львів, 1994. - Відп. викон.: Грабовецький Ю.В., Любінець Я.В., Тавпаш Ю.А.

4. Любінець Я.В. Временная реляционная модель данных.- Тез. док. на II международном семинаре "Теоретические и прикладные проблемы моделирования предметных областей в системах баз данных и знаний" /Под. ред. Игнатенко Б.К. - Киев: Concept Ltd.

5. Любінець Я.В. Представлення часових параметрів в реляційних базах даних. Тези доповіді на I-й міжнародній конференції з інформаційних технологій і систем, 4- 9 жов. 1993 р., - м. Львів.

6. Створення програмно- математичного комплексу розширення функціональних можливостей систем баз даних і знань реляційного типу на основі апарату нетрадиційних логік: Звіт про НДР (заверш.) ЛПІ; Керівник Пасічник В.В. - ОІ9U0041765. - Львів, 1991. - 257 с. - Відп. викон.: Грабовецький Ю.В., Любінець Я.В., Тавпаш Ю.А.

7. Любінець Я.В., Пасічник В.В. Україномовні бази даних. Тези доповіді на I міжнародній конференції з проблем україномовних комп'ютерних систем. 28.09-2.10.1992 р.,

АНОТАЦІЯ.

Любинец Я.В. Методы и средства представления и обработки информации с временным параметром в реляционных базах данных.

Диссертация на соискание ученой степени кандидата физико-математических наук по специальности 06.13.17 - теоретические основы информатики, Институт прикладной информатики, Киев, 1994.

Диссертационная работа содержит теоретические исследования расширения возможностей реляционных баз данных для адекватного манипулирования информацией с временным параметром, а также результаты их практического применения. Разработана обобщенная временная реляционная модель данных, которая учитывает зависимости синхронности атрибутов, и SQL-подобный язык описания и манипулирования данными на ее основе. Предложены принципы расширения и построено соответствующее расширение реляционной алгебры. Осуществлен ряд промышленных внедрений информационных систем с использованием технологии временных баз данных.

Lubinetc Ya.V. Methods and tools for representation and processing of information with temporal parameters in relational databases.

The candidat's thesis on physics and mathematics in 05.13.17 - Theoretical foundations of informatics, Institute of applied informatics, Kiev.

The thesis contains theoretic research, which expanding the capability of relational databases for adequate information processing with temporal parameters, as well as the practical application of that theory. Developed the generalized temporal relational data model, which takes into account the dependencies on synchronized attributes, SQL-like language of data describing and manipulation based on this theory. Principles of relational algebra extending are proposed and respective relational algebra is designed. Implemented several instances of applied information system using temporal database technology.

Ключові слова: реляційні бази даних, часові моделі даних, мови запитів.

459536

АВ 30.999

АВ 30.999

Підп. до друку 10.09.94. Формат 60x84^I/16
Папір друк. № 2. Офс. друк. Умовн. друк. арк. 1.5
Умовн. фарб.-відб. 1.5 Умовн. видав. арк. 1.32
Тираж 110 прим. Зам. 477, безплатно

ДУЛП 290646 Львів-ІЗ, Ст. Бандери, 12

Дільниця оперативного друку ДУЛП
Львів, вул. Городоцька, 286