

ГБ ОЯ
21 ДЕК 1998

На правах рукописи

БЕЛОУСОВ Алексей Леонидович

**МЕТОДЫ ОРГАНИЗАЦИИ СИСТЕМ УПРАВЛЕНИЯ ДАННЫМИ
НА ОСНОВЕ НУМЕРАЦИОННЫХ МЕТОДОВ
И ИНТЕРВАЛЬНЫХ ВЫЧИСЛЕНИЙ**

**05.13.11 – Математическое и программное обеспечение вычислительных машин,
комплексов, систем и сетей**

**Автореферат
диссертации на соискание ученой степени
кандидата технических наук**

Пенза 1999

Работа выполнена в Пензенском государственном педагогическом университете.

Научный руководитель: д.т.н., профессор Линьков В.М.

Официальные оппоненты:

доктор технических наук, профессор Левин В.И.;

кандидат технических наук, доцент Мякишев Д.Б.

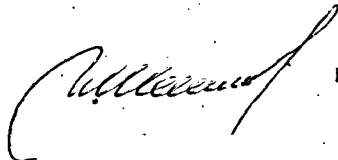
Ведущая организация: Пензенский научно-исследовательский электротехнический институт.

Защита состоится «21» января 1999 г., в 14 часов, заседании диссертационного совета Д 063.18.02 в Пензенском государственном университете по адресу: 440017, г. Пенза, Красная, 40, ПГУ.

С диссертацией можно ознакомиться в библиотеке Пензенского государственного университета.

Автореферат разослан «7» декабря 1998 г.

Ученый секретарь
диссертационного совета



Шашков Б.Д.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Диссертационная работа посвящена вопросам повышения эффективности систем управления данными на основе применения мерационных методов и интервальных вычислений.

Актуальность работы. В последние годы значительное внимание уделяется решению вопросов, связанных с проектированием систем управления данными, которые работают с большим объемом хранимой информации. В связи с этим важной задачей является поддержание структуры системы эффективной обработки различных запросов. Этому при проектировании систем управления данными необходимо уделять вопросы поддержания целостности информации, устранения избыточности данных, повышения общей производительности системы. Решение этих задач позволит сократить объемы используемой памяти, повысить скорость обработки запросов, обеспечить высокий уровень достоверности информации, защиту от несанкционированного доступа. При этом для улучшения этих характеристик системы применяется несколько подходов, для которых разрабатываются оптимизаторы запросов. Эффективность системы в целом во многом определяется качеством системы оптимизации запросов. Во всех существующих направлениях, связанных с оптимизацией, остаются нерешенные проблемы. Большинство из них имеет переборный характер и требует развития эвристических решений.

Другая специфическая проблема оптимизации запросов и структур хранения и стратегий доступа относится к системам управления базами данных в оперативной памяти. Такие системы становятся все более актуальными в связи с постоянным увеличением объемов доступной в ЭВМ оперативной памяти и ее удешевлением.

Проблемы, связанные с эффективной обработкой запросов в системах управления данными, рассматриваются в работах отечественных и зарубежных специалистов в области проектирования баз данных (Д. Эйер, К. Дейт, Д. Ульман, С. Кузнецов и др.). Линьковым В.М. рассматривалось применение нумерационных методов для решения задач эффективного хранения и управления данными. Использование монотонных отображений в нумерации дает возможность производить интервальные вычисления над числовыми номерами для эффективной обработки данных.

Применение нумерационных методов и интервальных вычислений в эвристически-ориентированной системе управления данными позволяет решить проблемы, связанные с сокращением затрат на модификацию баз данных (ЭД) в случае изменения отдельных полей или связей, поддерживать

семантику при работе с данными, решить вопросы организации эффективных методов хранения и доступа к данным.

Предметом исследования являются вопросы повышения эффективности работы информационных систем путем применения нумерационных методов интервальных вычислений при обработке запросов для домен ориентированной организации данных.

Цель работы. Целью работы является разработка внутренней модели данных, обеспечивающей наиболее полное использование преимуществ нумерационных методов и методов эффективной организации систем управления базами данных (СУБД) в реализации запросов на основе использования интервальных вычислений.

Задачи исследования. Для достижения поставленной цели решаются следующие задачи:

1. Разработка формальной модели базы данных на основе домен ориентированной структуры системы с использованием нумерационных методов.
2. Разработка методов преобразования запроса в интервалы уравнения.
3. Разработка механизмов применения нечетких интервалов и оптимизации процесса обработки запроса.
4. Разработка эффективных методов решения интервальных уравнений.
5. Разработка архитектуры системы управления данными на основе интервальных вычислений и методов проектирования СУБД.
6. Разработка алгоритмов реализации функций обработки запросов в системе управления данными.

Научная новизна.

1. Обосновано применение нумерационных методов для повышения эффективности обработки запросов в системах управления данными.
2. Разработана внутренняя модель данных и архитектура СУБД на основе нумерационных методов с использованием домен ориентированного подхода к организации хранения и обработки данных, что позволило применить интервальные вычисления в реализации эффективных методов доступа к данным.
3. Предложены методы преобразования запросов в интервалы уравнения, позволяющие эффективно решать задачу оптимизации запросов.
4. Для повышения эффективности в использовании ресурсов ЭВМ при обработке запросов предложены методы, основанные на работе нечеткими интервалами.
5. На основе интервальных вычислений разработаны алгоритмы обработки запросов, максимально использующие преимущества предложенной внутренней модели данных.

Основные результаты и положения, выносимые на защиту:

1. Внутренняя модель данных в виде совокупности бинарных отношений.

2. Метод преобразования реляционной модели во внутреннюю модель данных.

3. Методы преобразования запросов, представленных в виде эдиката, в систему интервальных уравнений.

4. Метод интерпретации результатов решения интервальных уравнений, позволяющий однозначным образом осуществить выборку ртежей отношений в доменно-ориентированной базе данных.

5. Метод применения показателя нечеткости для оптимизации пользования имеющихся ресурсов.

6. Архитектура системы управления данными и алгоритмы поиска вета на запрос, позволяющие эффективно использовать предложенную модель данных.

Практическая ценность заключается в следующем:

- предложена архитектура системы, в которой для хранения и работки данных применяются нумерационные методы;

- определены основные требования, предъявляемые к базам данных на нове доменно-ориентированного подхода;

- разработаны принципы построения системы управления данными на нове применения нумерационных методов, которые нашли свое именение в реализации ряда информационных систем.

Апробация работы. Основные результаты работы докладывались:

- на международной выставке-конференции "Новые компьютерные хнологии в учебном . процессе и научных исследованиях" (г. сква, 1995г.);

- на II международной конференции "Новые информационные хнологии и системы:" (г. Пенза, 1996);

- на Международной методической конференции "Университетское разование в условиях формирования рыночных отношений (г. Пенза, 97г.);

- на научно-практическом семинаре "Применение баз данных" (г. нза, 1997);

- на всероссийской научно-технической конференции "Непрерывная и ежные логики в информатике, экономике и социологии" (Пенза, 1997).

Реализация работы. Основные результаты, изложенные в ссертационной работе, использованы в НИР, проводимой в рамках анта «Доменно-ориентированная нумерационная система управления

базами данных» программы «Конверсия и высокие технологии. 1997-2000 гг.». Кроме того, результаты работы использованы в НИР, выполнены для КПО "Пенэснаб" (г. Пенза) и РКК "Энергия" (г. Москва) и внедрены:

- в автоматизированную информационную систему "Пенэснаб", что подтверждено соответствующим актом;

- в автоматизированную обучающую систему для кафедры спортивной игры ПГПУ, что подтверждается соответствующим актом.

Публикации. Основные результаты диссертации опубликованы в печатных работах.

Структура и объем работы. Диссертация состоит из введения, 7 глав, заключения, списка литературы из 83 наименований и 2 приложений. Работа содержит 142 страницы текста, 7 рисунков, таблиц, 7 страниц библиографии, 8 страниц приложений.

СОДЕРЖАНИЕ РАБОТЫ

Во введении отражены: актуальность темы, цели и задачи диссертации, новизна и практическая ценность полученных результатов, структура диссертации и положения, выносимые на защиту.

В первой главе рассмотрены основные способы оптимизации запросов в системах управления данными.

Проведенный анализ основных методов оптимизации запросов информационным системам позволяет сделать вывод, что решение проблем повышения производительности имеет актуальное значение для СУБД различных типов. Существенное повышение эффективности можно добиться за счет оптимизации запросов. Целью оптимизации запросов является уменьшение количества элементов, участвующих в реализации запроса.

Традиционные СУБД решают проблемы оптимизации путем логической физической оптимизации запросов. На этапе логической оптимизации происходит преобразование запроса в форму, содержащую минимальное количество элементов запроса и оптимальные планы его выполнения. Здесь можно использовать некоторую семантическую информацию. Такой подход в СУБД System R и Ingres базируется на использовании представлений или ограничений целостности. При физической оптимизации учитывается и формируется среда, обеспечивающая высокую скорость выполнения запросов. Так, для ускорения поиска в FoxPro используется

текстные файлы, в системе ADABAS - инвертированные файлы. В ADABAS тексты хранятся сжатыми, в виде строк байтов переменной длины.

Анализ методов оптимизации запросов с учетом физической организации БД позволяет сделать вывод, что при выборе методов хранения данных приходится принимать компромиссные решения.

Методы, обеспечивающие высокую плотность данных, приводят к увеличению времени, затрачиваемого на адресацию и поиск данных в файле. Средства, обеспечивающие высокую гибкость поиска, как правило, связаны с увеличением объема занимаемой памяти и времени доступа.

Средства, обеспечивающие возможность добавления записей в большой объем времени, требуют использования большей памяти для хранения данных. Средства эффективного использования памяти часто оказываются удовлетворительными при эксплуатации, так как приводят к необходимости реорганизации хранения данных при добавлении новых записей.

Вопросы оптимизации запросов актуальны в распределенных реляционных системах, а также в системах с использованием микропроцессорной архитектуры (Informix, Oracle7). В этих системах необходимо решать комбинаторно сложные проблемы обеспечения масштабируемости и уменьшения числа элементов, участвующих в запросе.

Для решения вопросов повышения производительности систем управления данными предлагается использовать нумерационные методы и менно-ориентированную внутреннюю модель, которые являются новыми решениями в области обработки запросов, открывающие новые возможности повышения эффективности за счет применения нумераций, внутренней декомпозиции и применения интервальных вычислений.

Во второй главе предлагается модель внутренней организации реляционных систем, которая позволяет эффективно применять нумерационные методы и более полно учитывает особенности доменно-ориентированной топологии.

Пусть имеется реляционное отношение r со схемой $(Attr_1, \dots, Attr_n)$. Каждому имени атрибута $Attr_i$ ставится в соответствие множество D_i , называемое доменом атрибута $Attr_i$, $1 \leq i \leq n$. Домен атрибута

Atr_1 будем также обозначать как $dom(Atr_1)$ ($dom: \{Atr_1\} \rightarrow \{D_1\}$). Доме являются произвольными непустыми конечными или счетными множествами

Для построения внутренней модели предлагается выполнить уникальную идентификацию кортежей отношений в БД. С этой целью рассмотрим множество отношений B в реляционной БД и построим множество кортежей E путем объединения кортежей отношений r БД:

$$E = \bigcup_{r \in B} r$$

Для множества кортежей вводится нумерация v :

$$v: N_0 \rightarrow E, \text{ где } N_0 - \text{номерное множество.}$$

Таким образом, каждый кортеж в БД получает уникальный номер. Генерация уникальных номеров для включаемых кортежей является функцией системы управления данными. Номер кортежа следует рассматривать как внутрисистемный ключ кортежа отношения.

Значение любого атрибута в кортеже зависит от ключевых значений кортежа, то есть от его внутрисистемного номера (VSN). Таким образом имеет место ряд функциональных зависимостей для каждого атрибута:

$$A_1: N_0 \rightarrow D_1$$

$$A_j: N_0 \rightarrow D_j.$$

Исходя из этого, каждому атрибуту можно сопоставить бинарное отношение. Тогда БД будет представлять собой совокупность бинарных отношений:

$$B = \{A_1, A_2, \dots, A_n\},$$

где n - число атрибутов в базе данных.

Наряду с нумерацией кортежей предлагается выполнить кодирование значений кортежа путем замены значения его номером.

Пусть $D = \{D_1, D_2, \dots, D_n\}$. Зададим для каждого множества нумерацию вида:

$$v_1: N_1 \rightarrow D_1, \text{ где } N_1 - \text{номерное множество.}$$

Нумерация v_1 определяется так, чтобы задать монотонное отображение множества N_1 в D_1 , т.е.

$$n_x \leq n_y \Leftrightarrow v_1(n_x) \leq v_1(n_y), \quad n_x, n_y \in N_1.$$

Таким образом, пронумерованные домены можно рассматривать как отношения R_{0i} , которые представляются как множества кортежей:

$$\langle n_x, d \rangle \mid d \in D_i, n_x \in N_i, v_i(n_x) = d \rangle.$$

Эти отношения образуют множество $R_0 = \{R_{0i}\}$. Используя нумерацию, можно конструировать отношения, содержащие вместо значений доменов номера значений. Информация о самих значениях может быть тогда легко восстановлена при помощи операции реляционной алгебры – естественного соединения.

Тогда модель внутренней организации связей будет иметь вид:

$$M = \langle S, R_0, v, v_0, \Sigma \rangle,$$

где $v_0 = \{v_i \mid 1 \leq i \leq n\}$,

S – множество бинарных отношений $S = \{S_i \mid i \in I\}$,

$S_i = \{ \langle s_i, n_i \rangle \mid (v_{0i}(n_i) \in D_i) \text{ и } (v(s_i) \in E) \}$,

Σ – набор теоретико-множественных операций.

Рассмотрим отображения δ и δ^{-1} , позволяющие соответственно из отношений S_i сконструировать $\{S_i\}$ и наоборот:

$$\delta: B \rightarrow T(S),$$

$$\delta^{-1}: T(S) \rightarrow B,$$

$T(S)$ – система подмножеств множества бинарных отношений S .

Номер кортежа s однозначно определяет значение каждого атрибута данного кортежа. Поэтому имеет место функциональная зависимость \bar{f} :

$$f: s \rightarrow n_x,$$

где $v_0(n_x) = d_x$, $d_x \in D_x$ и атрибут с номером a_j имеет область значений D_k .

Поэтому мы можем провести декомпозицию отношения R , при этом исключаются возможности дублирования и потери информации; даже если при нумерации v_{0i} использовались одинаковые номерные множества N_i для различных доменов или два атрибута имеют одинаковую область значений D_k .

Суть операции δ^{-1} состоит в том, что для каждого атрибута кортежа с номером s в отношении S_i находятся кортежи $\{s, n\}$, затем

для каждого атрибута с областью значений D_i по номеру n , исполняя нумерацию v_{n_i} , восстанавливаем его значение d .

С использованием нумераций v_{n_i} мы можем получить еще функциональных зависимостей:

$$A_1^v: E \rightarrow E_1,$$

которые каждому номеру кортежа ставят в соответствии ВСН атрибут, причем операция A_1^v является композицией операций A_1 и v_{n_1} .

Рассмотрим пример преобразования стандартной реляционной модели в предложенную бинарную модель.

Пусть имеется отношение: «Студент» (таблица 1).

Таблица 1. Отношение «Студент»

№	ФИО ст-та	Год рожд.	Пол	Группа
1	Иванов	1974	М	91Ф2
2	Петров	1973	М	90Ф3
3	Сидоров	1962	М	84Е1
4	Кривошеев	1968	М	84Е1
5	Андрянова	1967	Ж	84Е1
6	Сидоренко	1968	Ж	84Е1
7	Ревунов	1971	М	90Ф3
8	Матросов	1973	М	90Ф3

Первая колонка - номер кортежа - в традиционных реляционных отношениях не используется и введена для удобства нумерации связей.

Введем домены для данного отношения

Домен «ФИО» (D_1)

ВСН	ФИО
...	...
101	Алексеева
102	Андрянова
103	Борисов
...	...
301	Иванов
302	Кривошеев
303	Куроедов
304	Матросов
305	Петров
...	...
500	Ревунов
501	Сидоренко

502	Сидоров
...	...

Домен «Название группы» (D_2)

ВСН	Название группы
...	...
500	84Е1
...	...
600	90П1
...	...
1001	90Ф3
...	...
1500	91Ф2
...	...

Домен «Пол» (D_3)

1	М
---	---

ВСН	Пол
0	Ж

С использованием нумераций составляются таблицы связей:

Связь Студент-ФИО

№ Кортажа	ВСН ФИО
1	301
2	305
3	502
4	302
5	102
6	501
7	500
8	304

Связь Студент - Год рожд.

№ Кортажа	Год рожд.
1	1974
2	1973
3	1962
4	1968
5	1967
6	1968
7	1971
8	1973

Связь Студент - Пол

№ Кортажа	Пол
1	1
2	1
3	1
4	1
5	0
6	0
7	1
8	1

Связь Студент - Группа

№ Кортажа	Группа
1	1500
2	1001
3	500
4	500
5	500
6	500
7	1001
8	1001

По номеру кортежа можно однозначно определить ВСН атрибута, а через домены - и само его значение. Таким образом, можно восстановить любой кортеж исходного отношения.

Так как отношения данной модели состоят из чисел, эффективно применить к этой модели процедуры обработки числовой информации, в частности, интервальные вычисления.

Введем определенный набор символов операций Ω , принятый в рассматриваемой информационной системе.

Множество Ω включает в себя

$$\Omega = \{ +, -, *, /, =, >, <, \geq, \leq, \vee, \neg, \exists, \forall \}$$

Запрос для баз данных в указанной схеме представляется соответствующими формулами исчисления предикатов первого порядка (отношения, логические операции, кванторы общности и существования).

Использование доменно-ориентированной методологии и нумерации с отмеченными выше свойствами дает возможность организовать поиск данных, удовлетворяющих запросу, следующим образом:

-определение интервалов возможных значений $I(D_1)$ переменных, входящих в запрос;

-преобразование системы интервалов $I(D_1)$ в систему числовых интервалов $I(N_1)$, используя нумерации v_{D_1} ;

-преобразование исходного запроса в систему интервальных уравнений;

-решение интервальных уравнений относительно систем интервалов $I(N_1)$ и получение нового множества интервалов $I'(N_1)$;

-определение множеств кортежей $I(s_1)$ для интервалов $I'(N_1)$ каждого атрибута, которые удовлетворяют соответствующему условию предиката;

-получение общего набора кортежей, удовлетворяющих условию путем проведения теоретико-множественных операций над множествами $I(s_1)$;

-восстановление по номерам кортежей номеров и значений атрибутов, входящих в данный запрос, то есть получение элементов и D_1 , используя нумерации v_{D_1} .

Данный алгоритм осуществляет операции над числами, которые требуют меньшего времени обработки. Кроме того, использование чисел позволяет в большинстве случаев сжать исходную информацию.

Будем рассматривать множество возможных значений R как частичн упорядоченное множество, на котором задан линейный порядок \leq . Элементы множества R будем обозначать строчными буквами $a...z$.

Определим на множестве R понятие замкнутого интервала

$$I = [a, b] = \{ x \in R \mid a \leq x \leq b \}.$$

Множество всех замкнутых интервалов в R обозначим через $I(R)$. Всякое вещественное число x из R может считаться особым элементом и $I(R)$, имеющим вид $[x, x]$.

Два интервала A и B называются равными (записывается $A=B$), если они равны в теоретико-множественном смысле. Для замкнутых интервалов $A=[a_1, a_2]$ и $B=[b_1, b_2]$ из определения равенства следует:

$$A=B \Leftrightarrow a_1=b_1, \quad a_2=b_2.$$

Наряду с замкнутыми интервалами существуют открытые полуоткрытые интервалы:

$$I_1 = (a, b] = \{x \mid a < x \leq b, a, b \in D\},$$

$$I_2 = [a, b) = \{x \mid a \leq x < b, a, b \in D\},$$

$$I_3 = (a, b) = \{x \mid a < x < b, a, b \in D\}.$$

Для открытого (полуоткрытого) интервала всегда можно указать замкнутый интервал, содержащий все его элементы.

Одним из этапов реализации запроса является вычисление области значений переменных, входящих в запрос. Эта задача может быть сведена к задаче представления запроса в виде интервальных уравнений. Решение этой задачи состоит в вычислении по исходному предикату $P(x_1, x_2, \dots, x_n)$ соответствующей системы интервалов.

Рассмотрим переменные a и b , которые имеют одну и ту же область возможных значений D . Область значений этих переменных характеризуется некоторыми интервалами I_a и I_b из D . Предикат $P(a, b)$ накладывает дополнительные ограничения на области значений переменных. Обозначим интервалы, при которых предикат истинен, I'_a и I'_b , $I'_a \subseteq I_a$, $I'_b \subseteq I_b$. Наряду с операцией вхождения \subseteq определим отношение $<$ на множестве интервалов:

$$I_a < I_b \Leftrightarrow (\inf(I_a) < \inf(I_b)) \& (\sup(I_a) < \sup(I_b)),$$

$$\text{где } I_a, I_b \in I(D).$$

В диссертации приведены методы преобразования простейших предикатов в интервальные уравнения. В интервальном уравнении в качестве неизвестных используются интервалы. В результате решения интервальных уравнений получаются новые интервалы, например:

- для предиката $a \leq t, t$ -константа:

$$I'_a = I_a \cap (\inf(D), t];$$

- для предиката $a \leq b$:

$$I'_a = \begin{cases} I_a, & \text{если } I_a < I_b \text{ или } I_a \subseteq I_b, \\ I_a \cap I_b, & \text{если } I_a < I_b, \\ [\inf(I_a), \sup(I_b)], & \text{если } I_b \subseteq I_a. \end{cases}$$

$$I'_b = \begin{cases} I_b, & \text{если } I_a < I_b \text{ или } I_b \subseteq I_a, \\ I_a \cap I_b, & \text{если } I_b < I_a, \\ [\inf(I_a), \sup(I_b)], & \text{если } I_a \subseteq I_b. \end{cases}$$

Для интервалов I'_a и I'_b в последнем случае в системе происходит сужение области значений переменных, удовлетворяющих запросу. Но полученные интервалы не означают, что для каждой пары

$\langle a, b \rangle$, такой, что $a \in I'_a$ и $b \in I'_b$, предикат $a \leq b$ будет истинен. Для нахождения номеров кортежей, удовлетворяющих запросу, необходимо для каждого кортежа, в котором значение a принадлежит интервалу I' проверить истинность условия ($b \in I'_b$) и $(a \leq b)$.

Рассмотрим, как вычисляются области значений переменных предикатах, содержащих операции конъюнкции и дизъюнкции:

$$P_1(a) \& P_2(a).$$

Пусть J_1 и J_2 - области значений переменной a в предикатах P_1 и P_2 . Тогда интервал I'_a вычисляется следующим образом:

$$a_{\min} = \max\{\inf(J_1), \inf(J_2)\},$$

$$a_{\max} = \min\{\sup(J_1), \sup(J_2)\},$$

$$I'_a = \begin{cases} [a_{\min}, a_{\max}] & \text{с } a_{\min} \leq a_{\max}, \\ \emptyset, & a_{\min} > a_{\max}. \end{cases}$$

При использовании рациональных функций в запросах вычисленные функции можно заменить на переменные, сведя вычисления вышеперечисленным случаям.

Рассмотренные выше преобразования ориентированы на переменные одной областью значений. Однако реальные информационные системы характеризуются несколькими областями значений предметных переменных.

Для более эффективной реализации запроса следует (где это возможно) привести запрос к запросу с разделенными переменными:

$$P(x_1, x_2, \dots, x_n) = P_1(x_1) \theta P_2(x_2) \theta \dots \theta P_n(x_n),$$

где вместо символа θ можно подставить операцию конъюнкции ($\&$) и дизъюнкции (\vee).

Приведем примеры запросов, приводящиеся к запросам с разделенными переменными:

$$\neg P_1(x_1) \& (P_2(x_1) \vee P_3(x_2)).$$

Если выполняется условие $P_1(x_1) \& P_2(x_1) = \text{False}$, то исходный запрос можно преобразовать в эквивалентный запрос с разделенными переменными $P_1(x_1) \& P_3(x_2)$.

$$\neg P_1(x_1) \vee (P_2(x_1) \& P_3(x_2)).$$

Если $P_1(x_1) \vee P_2(x_1) = \text{True}$, то исходный запрос эквивалентен запросу с разделенными переменными $P_1(x_1) \vee P_3(x_2)$.

Как было сказано ранее, при использовании интервалов вычислений задача поиска на заключительных этапах состоит

нахождении пересечения и объединения интервалов. Однако в некоторых случаях из-за ограниченности имеющихся ресурсов или с целью оптимизации процесса обработки запроса следует использовать покрывающее множество интервалов, которое содержит меньше элементов. Возникает задача: перейти от множества интервалов $\{I\}$, удовлетворяющих запросу, к покрывающему множеству $\{I^{\circ}\}$, такому, что $\{I\} \subseteq \{I^{\circ}\}$. Во множестве $\{I^{\circ}\}$ будут находиться значения, которые не удовлетворяют запросу. Будем называть систему интервалов правильной, если все значения в этой системе удовлетворяют запросу. В противном случае система интервалов называется неправильной. Для неправильной системы интервалов на заключительном этапе необходимо для каждого ее элемента осуществить проверку, удовлетворяет ли он исходному запросу. Для эффективной обработки запросов для каждого интервала вводится числовой параметр p , показывающий, какая часть элементов интервала удовлетворяет запросу. Этот параметр равен единице, если все элементы интервала удовлетворяют запросу, и нулю, если не один элемент не удовлетворяет запросу.

Пусть имеются два правильных интервала $A=[a_1, a_2]$ и $B=[b_1, b_2]$ с параметрами p_A и p_B , равными единице (это означает, что все элементы этих интервалов удовлетворяют запросу), причем

$$a_1 \leq a_2 \leq b_1 \leq b_2.$$

Составим покрытие этих интервалов одним интервалом C таким, что $\{A, B\} \subseteq C$, $C=[c_1, c_2]$.

Если $a_2 \leq b_1$, то не все элементы интервала C удовлетворяют исходному запросу. Поэтому введем характеристику p_C для интервала C , применяя геометрическое определение вероятности (отношение длины отрезков, которым принадлежат элементы, удовлетворяющие запросу $/(b_2 - a_1) - (b_1 - a_2)/$, к общей длине интервала $C / (b_2 - a_1) /$):

$$p_C = 1 - (b_1 - a_2) / (b_2 - a_1)$$

Вероятность p_C показывает, что наудачу выбранный элемент интервала C удовлетворяет запросу.

Рассмотрим, как вычисляются вероятностные характеристики для операций множества Ω над интервалами $A^{\circ} = \{A=[a_1, a_2], p_A\}$ и $B^{\circ} = \{B=[b_1, b_2], p_B\}$:

-пересечение: $A^{\circ} \cap B^{\circ} = \{A \cap B, p_{A \cap B}\}$, $p_{A \cap B} = p_A * p_B$.

-дополнение: данная операция с нечеткими множествами может привести к тому, что в конечной системе интервалов будут отсутствовать элементы, удовлетворяющие запросу. Поэтому необходимо при реализации операций над интервалами перейти к эквивалентному выражению, в котором не требуется проводить операцию дополнения;

-арифметические операции: $P_A * P_B$;

-предикаты вида $a \leq t$, t -константа - $P'_a = P_a$;

-предикаты вида $a \leq b$, $P'_a = P'_a$ и $P'_b = P'_b$.

Использование числа как меры появления элемента удовлетворяющего запросу, дает возможность изменения стратегии поиска в процессе изменения вероятностных характеристик интервалов. Так уменьшение вероятностной характеристики интервала до определенного предела может привести к решению о пересмотре этого интервала уточнения его вероятностных характеристик. Это позволит ускорит поиск путем уменьшений области перебора.

В третьей главе рассматриваются вопросы проектирования систем управления данными. При этом необходимо решать вопросы разработку эффективных методов хранения и доступа к данным, а также обработку запросов пользователя. Чтобы эффективно использовать для решения эти задачи доменно-ориентированный подход и интервальные вычисления при реализации запросов к системе, предлагается структура, приведенная на рис. 1.

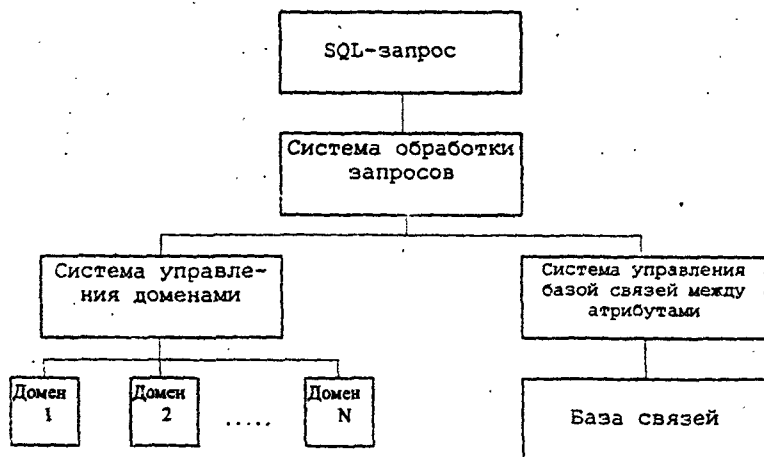


Рисунок 1 Структура системы обработки запросов

В данной системе связи между атрибутами хранятся отдельно от амых значений атрибутов.

Процесс обработки запросов в предложенной системе организуется ледующим образом. SQL-запрос поступает в систему обработки запросов. а первом этапе производится его лексический и синтаксический анализ. здесь выявляются ошибки и производится логическая оптимизация словий, входящих в запрос.

На второй стадии производится преобразование условия запроса в истему интервальных уравнений. При этом осуществляется обращение к истеме управления доменами с целью определения области возможных начений переменных, входящих в запрос. На третьем этапе роизводится решение интервального уравнения с целью сужения области оиска элементов, удовлетворяющих исходному условию. На четвертой тадии производится выборка интервалов номеров кортежей для каждого лементарного условия. При этом возможно создание покрывающих ножеств с целью повышения эффективности обработки. На пятом этапе роизводится пересечение/объединение множеств номеров кортежей, олученных на четвертой стадии. В результате находится покрывающее ножество для множества, удовлетворяющего общему условию. На аключительном (пятом) этапе производится выборка "правильных" лементов покрывающего множества и восстановление исходных значений трибутов путем обращения к системе управления доменами.

Интервальные вычисления могут применяться не только в редложениях запроса SELECT языка SQL, но и в других его командах. ак, в операции удаления элементов отношений DELETE и модификации уществующих элементов UPDATE, как и в предложении SELECT, входит лчечное слово WHERE, определяющее условие выполнения соответствующей перации.

В результате экспериментов была построена зависимость времени еакции системы на запрос от количества обрабатываемых атрибутов. Она имеет линейный характер для доменно-ориентированной системы, азработанной автором, и относительно постоянный для аписеориентированных систем. Сравнение разработанной системы с имеющимися СУБД показало преимущество разработанных методов при еализации запросов определенных классов. Использование в SQL-запросе

до пяти атрибутов обеспечивает более высокую скорость по сравнению с записеориентированной СУБД FoxPro.

В Приложении А приводятся основные функции, используемые при реализации системы управления данными на основе интервальных вычислений.

В Приложении В представлены документы, подтверждающие внедрение результатов диссертационной работы.

ОБЩИЕ ВЫВОДЫ ПО РАБОТЕ

1. Для повышения эффективности реализации запросов к информационной системе на основе применения нумерационных методов (в частности, интервальных вычислений) предложена внутренняя модель данных. Предложен механизм преобразования отношений реляционных БД во внутреннюю модель данных.

2. Предложены методы преобразования запросов к системе управления данными из предикатной формы в систему интервальных уравнений. Показано, что любое выражение, представленное в виде предиката, можно преобразовать в интервальное уравнение.

3. Рассмотрены вопросы решения интервальных уравнений над переменными с одним и тем же доменом и над переменными с различными областями значений. Показано, что для эффективной обработки запроса его необходимо на ранних этапах преобразовать к запросу с разделенными переменными.

4. Для оптимизации использования имеющихся ресурсов ЭВМ предложен метод оптимизации, основанный на использовании нечетких интервалов. Использование числовых параметров нечеткости множества позволяет гибко менять стратегию поиска в зависимости от изменения этих параметров.

5. Для применения предложенной модели разработана архитектура и выполнена реализация системы управления данными с использованием доменно-ориентированного подхода. Реализованы методы решения интервальных уравнений.

6. Использование независимых методов хранения и обработки информации позволяет эффективно применять интервальные вычисления для реализации запросов по выборке элементов БД.

7. Для оптимизации запросов предложены и реализованы на основе интервальных вычислений алгоритмы, позволяющие сузить количество элементов, участвующих в запросе, и, следовательно, область поиска. Сравнение разработанной системы с имеющимися СУБД показало преимущество разработанных методов при реализации запросов определенных классов. Использование в SQL-запросе до пяти атрибутов обеспечивает более высокую скорость по сравнению с записеориентированной СУБД FoxPro.

Основные положения диссертации изложены в следующих публикациях:

1. Белоусов А.Л. Применение электронных таблиц для преподавания и изучения экономико-математических моделей и методов // Тезисы доклада. Материалы международной выставки-конференции "Новые компьютерные технологии в учебном процессе и научных исследованиях". - М.: МАДИ, 1995г.

2. Белоусов А.Л., Линьков В.М. Применение электронных таблиц при решении статистических задач // Тезисы доклада. Материалы международной выставки-конференции "Новые компьютерные технологии в учебном процессе и научных исследованиях". - М.: МАДИ, 1995г.

3. Белоусов А.Л., Линьков В.М. Применение электронных таблиц при преподавании математических дисциплин. // Материалы IV Международной конференции-выставки «Информационные технологии в образовании». - М.: 1995г.

4. Линьков В.М., Белоусов А.Л. Применение электронной таблицы TELES для преподавания и изучения математических дисциплин // Тезисы доклада. Материалы VII Международной конференции «Применение новых технологий в образовании», 29 июня-2 июля 1996г., Троицк.- Фонд новых технологий в образовании «Байтик», 1996.

5. Белоусов А.Л. Применение интервальных вычислений для реализации запросов языка SQL // Новые информационные технологии и системы: Материалы II международной конференции - Ч.1. Пенза, ПГТУ, 1996. - С. 39-40.

6. Линьков В.М., Белоусов А.Л. Применение электронных таблиц для

решений математических задач. // Университетское образование условиях формирования рыночных отношений: Материалы Международной методической конференции - Ч. II, Пенза, ПГТУ, 1997г. - С.136-138.

7.V.M.Linkov, V.N.Nedoshivin, V.V.Drozhdin, E.V.Ogurechniko A.L.Belousov, D.V.Majorov. The Intellectual System of Automat Design of Schemes of Measuring Systems (SAPR SI). EAST-WE International Conference INFORMATION TECHNOLOGY IN DESIGN EWITD'9 PROCEEDINGS, Moscow, Russia, 1996. p.66-68.

8.Линьков В.М., Белоусов А.Л., Майоров Д.В., Огуречников Е.В. Дружаев А.А. Применение доменно-ориентированного подхода при разработке систем автоматизированного проектирования средств измерения с поддержкой временной семантики // Применение баз данных Материалы научно-практического семинара. - Пенза, 1997. - С. 5-6.

9.Линьков В.М., Дрождин В.В., Майоров Д.В., Белоусов А.Л. Огуречников Е.В., Недошивин В.Н. Принципы реализации доменно ориентированной методологии построения баз данных //Непрерывная смежные логики в информатике, экономике и социологии: Материалы всероссийской научно-технической конференции. - Пенза, 1997. - С.39.

10.Белоусов А.Л., Огуречников Е.В. Применение нумерационных методов для реализации запросов к информационной системе с доменно ориентированной организацией данных // Математика и информатика Межвузовский сборник. - Пенза: ПГПУ им. В.Г.Белинского, 1996., с. 61-80.

11.Белоусов А.Л., Мокшанина М.А. Использование компьютера на курсе высшей математики //Математика и информатика: Межвузовский сборник. - Пенза: ПГПУ им. В.Г.Белинского, 1996., с.113-116.

12.Белоусов А.Л., Долгарев А.И., Линьков В.М., Самуйлова С.И. "Решение статистических задач с использованием электронных таблиц" Учебное пособие/ Пенз. Гос. пед. ун-т им. В.Г.Белинского.- Пенза 1996.- 44с.

13.Белоусов А.Л., Дружаев А.А., Иванцов М.А. «Применение временных баз данных при многоэтапном проектировании сложных систем»//Сучасні проблеми математики: Матеріали міжнародної наукової конференції. Частина 4.- Чернівці: Рута, 1998., С. 9-11.