

Елисеев Дмитрий Владимирович

**МЕТОДИКА ОБРАБОТКИ ТЕМПОРАЛЬНОЙ РЕЛЯЦИОННОЙ БАЗЫ
ДАННЫХ В МИВАРНОМ ПРОСТРАНСТВЕ**

**Специальность 05.13.17 - Теоретические основы информатики
(технические науки)**

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

Москва, 2011

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Московский государственный технический университет имени Н.Э. Баумана»

Научный руководитель: доктор технических наук, профессор
Балдин Александр Викторович

Официальные оппоненты: доктор технических наук,
старший научный сотрудник
Варламов Олег Олегович

кандидат технических наук, доцент
Рудаков Игорь Владимирович

Ведущая организация: ОАО «Научно-исследовательский институт систем автоматизации»

Защита диссертации состоится «8» декабря 2011 года в 16:30 на заседании диссертационного совета Д212.141.10 в Московском государственном техническом университете им. Н.Э. Баумана по адресу: 105005, г. Москва, 2-я Бауманская ул., д. 5, стр. 1

С диссертацией можно ознакомиться в библиотеке Московского государственного технического университета им. Н.Э. Баумана.

Автореферат разослан «__» _____ 2011 года

Ученый секретарь
Диссертационного совета Д 212.141.10
к.т.н., доцент

 С.Р. Иванов



ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время сложной задачей является адаптация класса информационных систем, обрабатывающих темпоральные данные посредством реляционной модели и функционирующих в динамично меняющихся предметных областях. Особенностью таких информационных систем является то, что они работают с темпоральными данными, структура которых также изменяется со временем. Описание темпоральных данных с изменяющейся структурой существующими методами является не всегда эффективным, а иногда невозможным, так как это приводит к перепроектированию базы данных. Процесс адаптации такой базы данных к новым условиям является сложным, приводит к необходимости перекачки данных и требует добавления новых объектов поверх старых с сохранением истории изменения темпоральных данных. Примерами рассматриваемого класса информационных систем являются: система управления предприятием, система управления персоналом, финансовые приложения, страховые приложения и ряд других.

В разработку методов хранения и обработки темпоральных данных внесли вклад учёные, такие как Р. Снодграс, К. Дженсен, Дж. Бен-Зви, С. Гадия, Е. МакКензи, А. Стейнер.

Известные темпоральные модели, расширяющие реляционную, обладают рядом недостатков, которые сдерживают их применение на практике. Использование этих моделей в системах, в которых со временем необходимо изменять структуру данных, приводит к увеличению избыточности хранения информации, а также к разрастанию базовой реляционной модели, в которой усложняется составление, выполнение запросов и адаптация базы данных к новым условиям.

В настоящее время существует разработанный Варламовым О.О. новый миварный подход к описанию изменяемых, динамических предметных областей, который позволяет устранить избыточность хранения данных. Миварный подход используется для класса обучаемых систем, задачей которых является изучение и моделирование сложных динамических предметных областей. Основой миварного подхода к представлению данных является целостное, единое описание изучаемой предметной области с различных точек зрения посредством многомерного пространства. Однако, использование этого подхода для работы с темпоральной реляционной базой данных невозможно, так как отсутствует механизм обработки многомерного представления изменяющихся отношений.

Таким образом, актуальной задачей является разработка методики обработки реляционной базы данных с изменяемой структурой на основе миварного пространства, позволяющей описывать темпоральность структуры данных и самих данных унифицированным способом и устраняющей вышеуказанные недостатки. Использование такой методики позволит увеличить длительность жизненного цикла посредством адаптации базы данных информационной системы, функционирующей в динамично изменяющейся предметной области с темпоральными данными.

В.Э. Билин
Библиотека

Цель работы. Целью данной работы является исследование и разработка методики обработки и адаптации темпоральной реляционной базы данных с изменяемой структурой на основе миварного подхода.

В работе решаются следующие задачи:

1. Исследование и анализ основных методов хранения и обработки темпоральных данных, с целью формирования новых принципов построения темпоральных реляционных моделей с изменяемой структурой данных.
2. Разработка структуры миварного пространства для темпоральной реляционной базы данных.
3. Разработка методики обработки темпоральной базы данных с изменяемой структурой.
4. Разработка алгебры многомерных пространств для работы с миварным пространством.
5. Разработка операторов преобразования из реляционной модели в многомерное пространство и обратно.
6. Разработка грамматики и интерпретатора языка запросов к многомерному пространству.
7. Разработка операторов изменения реляционной базы данных через миварное пространство.
8. Оценка границ применимости новой методики обработки темпоральной реляционной базы данных с изменяемой структурой.

Объект исследования. Объектом исследования настоящей работы являются темпоральные реляционные базы данных.

Предмет исследования. Предметом исследования являются процессы хранения и обработки темпоральных данных посредством использования реляционной СУБД и миварного пространства.

Научная новизна. В работе получены следующие новые научные результаты:

1. Разработана новая методика обработки темпоральной реляционной базы данных с изменяемой структурой посредством использования миварного пространства.
2. Разработана алгебра многомерных пространств и доказаны свойства операций для определения части миварного пространства, содержащего результат запроса.
3. Разработаны алгоритмы операторов преобразования реляционной модели в многомерное пространство и обратно для использования стандартных SQL запросов к реляционной базе данных.
4. Разработаны алгоритмы операторов изменения темпоральной базы данных в миварном пространстве, которые модифицируют данные и их структуру одновременно.
5. Разработана грамматика языка запросов к миварному пространству на базе языка SQL.

Методы исследования. Исследования проводились на основе математического аппарата теории множеств, реляционной алгебры, теории формальных языков.

Практическая ценность полученных результатов. В данной работе для практического использования предложенной методики обработки темпоральной реляционной базы данных с изменяемой структурой разработано программное средство на базе технологической платформы «1С: Предприятие 8.2». Оно позволяет сформировать многомерное представление темпоральной реляционной модели, вносить изменения как в саму многомерную модель, так и в структуру многомерного пространства и динамически формировать отношения из многомерных структур данных при выполнении запросов. Таким образом, данную методику можно использовать с базами данных уже функционирующих информационных систем, чтобы не перекачивать накопленный архив изменений состояний темпоральных объектов, и разрабатывать новые информационные системы на базе многомерного представления реляционной модели данных.

Внедрение результатов исследований. Методика обработки реляционной базы данных с изменяемой структурой и разработанное программное средство внедрены для базы данных информационной системы кадрового учёта Московского государственного технического университета имени Н.Э. Баумана.

Публикации по теме. По материалам настоящей работы опубликовано 4 научные работы в журналах, рекомендуемых ВАК.

Апробация работы. Основные результаты работы докладывались на IV Всероссийской конференции молодых учёных и специалистов «Будущее машиностроения России» (Москва, 2011 г), на XIII Международной научно-практической конференции «Актуальные вопросы современной науки» (Москва, 2011 г).

Материалы работы были изложены автором на семинарах кафедры «Системы обработки информации и управления» МГТУ им. Н.Э. Баумана в 2009 - 2011 годах.

Объем работы. Диссертационная работа содержит 149 страниц, 47 рисунков и список литературы из 89 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность проблемы. Формулируются цель и задачи, решаемые в диссертации, раскрывается научная новизна и практическая ценность работы, а так же излагается краткое содержание глав диссертации.

В первой главе приведено описание темпоральных систем, выделены их особенности. В рамках реляционных баз данных рассмотрена концепция хранения и обработки темпоральных данных. Подробно проанализированы существующие подходы к обработке темпоральной базы данных и выделен ряд существенных недостатков при изменении структуры данных. На основе проведенного анализа сформулирована концепция и требования к новой

методике обработки темпоральных реляционных баз данных с изменяемой структурой.

Проведен анализ и классификация методов построения темпоральных реляционных баз данных по ключевым понятиям, таким как представление времени, тип отметки времени, количество осей времени, уровень темпоральности данных и показаны их преимущества и недостатки по каждой классификации. Выделены ключевые особенности темпоральных баз данных:

1. В темпоральных моделях факты представлены как блоки данных, имеющие отметку времени, которая показывает, когда они были действительны в предметной области и хранились в базе данных.
2. Темпоральные модели данных хранят информацию об эволюции объектов предметной области: в базе данных сохраняются все состояния объекта предметной области в течение его времени жизни.

При анализе существующих методов хранения и обработки темпоральных данных было выявлено, что они обладают рядом недостатков, ограничивающих их применение на практике при использовании в информационных системах, функционирующих в динамично меняющихся предметных областях. Показано, что каждое изменение структуры данных в темпоральной модели увеличивает избыточность хранения информации в отношениях и приводит к разрастанию исходной реляционной модели, в которой усложняется составление, выполнение запросов и адаптация базы данных к новым условиям.

Обоснована практическая целесообразность применения нового миварного подхода к построению темпоральных систем, который используется для разработки моделей данных с динамической, изменяемой структурой. Этот подход основан на самоорганизующемся динамическом многомерном объектно-системном дискретном пространстве унифицированного представления данных и правил (миварном пространстве). В настоящее время миварное пространство используется в системах искусственного интеллекта для описания изменяющихся знаний о предметной области и организации на них процедур логического вывода. Представление реляционной модели в миварном пространстве позволит устранить рассмотренные недостатки существующих темпоральных моделей на базе реляционной. При этом темпоральность модели данных достигается не за счёт применения реляционных конструкций, а является свойством реляционной модели и определяется структурой многомерного пространства.

Таким образом, для применения миварного подхода с темпоральными реляционными базами данных возникла необходимость разработки методов обработки и модификации многомерного представления изменяющихся отношений реляционной модели.

Во второй главе предложена новая методика обработки темпоральной базы данных с изменяемой структурой в миварном пространстве. Определена структура миварного пространства для описания темпоральной реляционной базы данных, для которого была разработана алгебра многомерных пространств, позволяющая выделять подпространство, содержащее результат

запроса, и доказаны основные свойства её операций. Разработаны алгоритмы преобразований из реляционной модели в многомерное пространство и обратно, чтобы использовать эксплуатируемые нетемпоральные базы данных и написанные для них запросы.

Миварное пространство состоит из множества осей, множеств элементов осей, множества точек пространства и множества значений точек. Формальное описание миварного пространства приведено ниже:

Пусть $A = \{a_n\}$, $n = \overline{1, N}$, где A – множество названий осей миварного пространства, N – количество осей миварного пространства.

Тогда $\forall a_n \exists B_n = \{b_{n_i}\}$, $n = \overline{1, N}$, $i_n = \overline{1, I_n}$, где B_n – множество элементов оси a_n , i_n – идентификатор элемента множества B_n , $I_n = |B_n|$.

Множества B_n образуют многомерное пространство: $M = B_1 \times B_2 \times \dots \times B_N$.

$m = (i_1, i_2, \dots, i_N)$, $m \in M$, m – точка многомерного пространства, $(i_1, i_2, \dots, i_N)$ – координаты точки m .

Существует множество значений точек многомерного пространства M :

$C_M = \{c_{i_1, i_2, \dots, i_N} \mid i_1 = \overline{1, I_1}, i_2 = \overline{1, I_2}, \dots, i_N = \overline{1, I_N}\}$, где $c_{i_1, i_2, \dots, i_N}$ – значение точки многомерного пространства M с координатами $(i_1, i_2, \dots, i_N)$.

Для каждой точки пространства M не существует или существует единственное значение из множества C_M . Таким образом, C_M – множество изменений состояний модели данных, представляемой в многомерном пространстве. Для перехода между многомерным пространством и множеством значений точек было введено отношение μ : $C_X = \mu(M_X)$, где $M_X \subseteq M$, $M_X = B_{1X} \times B_{2X} \times \dots \times B_{NX}$.

Для описания реляционной модели данных в миварном пространстве необходимо выделить три оси: ось отношений V , ось атрибутов отношения S и ось идентификаторов записей отношения ID как показано на рис. 16. Данные множества являются независимыми в отличие от реляционной модели данных.

В многомерном пространстве каждому значению атрибута кортежа отношения соответствует точка с определёнными координатами. Множество всех точек многомерного пространства соответствует реляционной модели данных. При миварном подходе структура данных определяется точками пространства, которые хранят соответствующие значения атрибутов отношения реляционной модели. Первичность данных создаёт новые возможности по изменению схемы реляционной базы данных и предпосылки для разработки новых методов обработки данных с изменяемой структурой.

Свойства реляционной модели данных при миварном подходе зависят от структуры пространства. Для того чтобы описывать темпоральные данные необходимо к рассмотренным трём осям добавить ось времени T .

Атрибут 1(s_1)	Атрибут 2(s_2)	...	Атрибут $n(s_n)$
Значение 1.1	Значение 1.2	...	Значение 1.n
Значение 2.1	Значение 2.2	...	Значение 2.n
...	...	...	...
Значение k.1	Значение k.2	...	Значение k.n

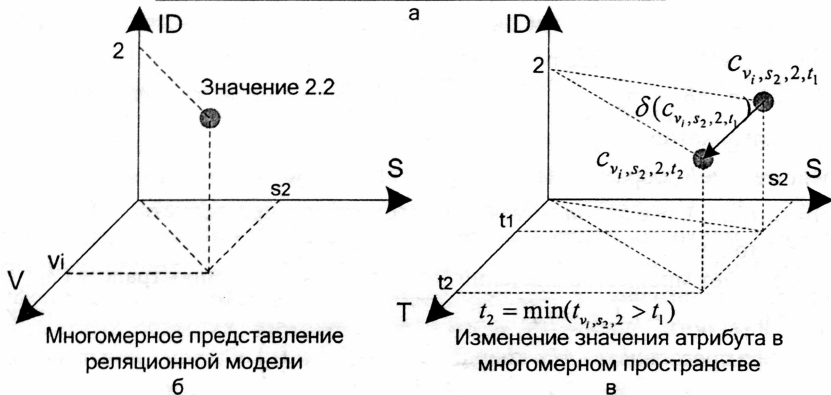


Рис. 1. Реляционное и многомерное представление данных

Определение 1. Пространство $\langle \text{отношение, атрибут, идентификатор} \rangle$ называется *определяющим*, а образующие это пространство оси – *определяющими осями*.

Определяющее пространство описывает реляционную модель данных. Значение каждой точки в определяющем пространстве зависит от параметров - других осей, образующих многомерное пространство.

Определение 2. Последнее (первое) состояние модели данных в момент времени t – совокупность точек многомерного пространства C_i , значения которых в определяющем пространстве по параметру оси времени являются последними (первыми) до заданного момента времени t .

Для реляционной модели состоянием модели данных является совокупность отношений в определённый момент времени. Таким образом, состояние реляционной модели изменяется, когда изменяется хотя бы одно значение атрибута в каком-либо отношении или изменяется схема какого-либо отношения.

На рис. 1в показано изменение значения атрибута в многомерном пространстве. Для описания этого процесса была введена функция δ , которая по заданной точке определяет другую точку многомерного пространства, имеющую такие же координаты в определяющем пространстве, и координату по оси времени, равную следующему изменению значения этой точки в определяющем пространстве. На рис. 1в $t_2 = \min(t_{v_i, s_2, 2} > t_1)$, где $t_{v_i, s_2, 2}$ - времена изменения значения атрибута s_2 кортежа 2 отношения ν_i .

Тогда многомерное пространство для описания темпоральной реляционной базы данных будет иметь следующий вид: $M = V \times S \times ID \times T$.

На многомерном пространстве, описывающем темпоральную реляционную базу данных, была разработана алгебра, которая позволяет выделить часть общего пространства, содержащую результат запроса к базе данных.

Носителем алгебры является следующее множество: $E = \{M_X : M_X \cup_M M = M\}$, где M – общее многомерное пространство темпоральной реляционной базы данных, $\cup_M$ – операция многомерного объединения пространств, рассмотренная ниже.

Сигнатура алгебры состоит из 4 операций: операции многомерного объединения ($\cup_M$), пересечения ($\cap_M$), разности ($\setminus_M$) пространств и операция многомерного среза ($\psi_M^{f(v,s,id,t)}$). Данные операции выполняются на множествах, образующими оси пространства, не анализируя множество значений точек C_M . Пусть даны два многомерных пространства:

$$M_A = V_A \times S_A \times ID_A \times T_A, V_A \subseteq V, S_A \subseteq S, ID_A \subseteq ID, T_A \subseteq T,$$

$$M_B = V_B \times S_B \times ID_B \times T_B, V_B \subseteq V, S_B \subseteq S, ID_B \subseteq ID, T_B \subseteq T.$$

Операции многомерного объединения пространств

$$M_D = M_A \cup_M M_B = V_D \times S_D \times ID_D \times T_D, V_D = V_A \cup V_B, S_D = S_A \cup S_B, ID_D = ID_A \cup ID_B,$$

$$T_D = T_A \cup T_B.$$

Операции многомерного пересечения и разности пространств определяются аналогично.

Операция многомерного среза пространства

$$M_D = \psi_M^{f(v,s,id,t)}(M_A) = V_D \times S_D \times ID_D \times T_D, \text{ где}$$

$$f(v_A, s_A, id_A, t_A) = f_V(v_A) \wedge f_S(s_A) \wedge f_{ID}(id_A) \wedge f_T(t_A), v_A \in V_A, s_A \in S_A, id_A \in ID_A,$$

$$t_A \in T_A.$$

$f(v,s,id,t)$ – общее логическое условие для операции многомерного среза. Это условие состоит из конъюнкции логических условий для каждого множества, образующего ось многомерного пространства M_A .

$$V_D = \{v_A : f_V(v_A) = \text{истина}\}, S_D = \{s_A : f_S(s_A) = \text{истина}\},$$

$$ID_D = \{id_A : f_{ID}(id_A) = \text{истина}\}, T_D = \{t_A : f_T(t_A) = \text{истина}\}.$$

Для введённых операций алгебры многомерных пространств были выявлены и доказаны шесть основных свойств.

На множестве значений точек многомерного пространства были разработаны операторы преобразования, которые позволяют получать определённые состояния реляционной модели данных и преобразовывать множество значений точек многомерного пространства в реляционную модель. Данные операторы не включены в алгебру многомерных пространств, так как получаемый результат выходит за рамки носителя алгебры.

Для темпоральной реляционной базы данных необходима возможность получения состояний реляционной модели в определённые моменты времени.

Чтобы выполнять указанную операцию были разработаны операторы преобразования «Срез первых» и «Срез последних».

Преобразование «Срез первых» («Срез последних»)

Преобразование «Срез первых» формирует первое состояние реляционной модели данных с заданного момента времени t («Срез последних» - последнее состояние реляционной модели данных до заданного момента времени t):

$$C_{\tau_F} = \tau_F(C, t_0) = \{\min_T [c_{m_i}]_{\rho_{VSId}}\}, \text{ где } c_{m_i} \in C_{i \geq t_0}, C_{i \geq t_0} = \{c_m : t \geq t_0, c_m \in C\}.$$

$$(C_{\tau_L} = \tau_L(C, t_0) = \{\max_T [c_{m_i}]_{\rho_{VSId}}\}, \text{ где } c_{m_i} \in C_{i \leq t_0}, C_{i \leq t_0} = \{c_m : t \leq t_0, c_m \in C\}.)$$

ρ_{VSId} - отношение эквивалентности на множестве значений точек многомерного пространства C_M :

$$m_1 = \langle v_1, s_1, id_1, t_1 \rangle, m_2 = \langle v_2, s_2, id_2, t_2 \rangle,$$

$$c_{m_2} = \rho_{VSId}(c_{m_1}) \Leftrightarrow v_1 = v_2 \text{ и } s_1 = s_2 \text{ и } id_1 = id_2.$$

Множество $[c_{m_0}]_{\rho_{VSId}} = \{c_m : c_m = \rho_{VSId}(c_{m_0})\}$, где $c_{m_0}, c_m \in C_M$ - класс эквивалентности по отношению ρ_{VSId} . Множество классов эквивалентности по отношению ρ_{VSId} на множестве значений точек многомерного пространства C_M образует разбиение этого множества и обозначается C_M / ρ_{VSId} .

Так как множество T является упорядоченным, элементы множества $[c_{m_0}]_{\rho_{VSId}}$ можно упорядочить по оси T и найти максимальный ($\max_T [c_{m_0}]_{\rho_{VSId}}$) или минимальный ($\min_T [c_{m_0}]_{\rho_{VSId}}$) элемент.

При выполнении операции «Срез первых» («Срез последних») сначала отбираются все значения точек, которые имеют координату по оси времени больше $C_{i \geq t_0} = \{c_m : t \geq t_0, c_m \in C\}$ (меньше $C_{i \leq t_0} = \{c_m : t \leq t_0, c_m \in C\}$) заданного значения t_0 . Затем выбранные значения точек разбиваются на классы эквивалентности по отношению ρ_{VSId} , и в каждом классе эквивалентности ищется наименьшая (наибольшая) точка по отношению порядка, заданного на оси времени.

В диссертации был разработан оператор перехода ϕ от реляционной базы данных в миварное пространство для выполнения операций с темпоральными данными. Переход в миварное пространство позволяет изменять структуру темпоральных данных без потери их истории и модификации исходной реляционной модели.

$\phi : R_D \rightarrow C_M$. При этом выполняются следующие действия:

1. $V = \{vr_1, vr_2, \dots, vr_K\}$, где vr_i - название отношения r_i , $i = \overline{1, K}$.
2. $S = \bigcup_{i=1}^K R_i$, где R_i - схема отношения r_i .
3. Dr_i - конечное множество кортежей отношения r_i , значит существует взаимно однозначное соответствие $f : Dr_i \rightarrow N_{Dr_i}$, где

$N_{Dr_i} = \{1, 2, \dots, |Dr_i|\}$, $|Dr_i|$ - мощность множества Dr_i . Тогда

$$ID = \bigcup_{i=1}^K N_{Dr_i}.$$

4. $T = \{t_0\}$.

5. $c_{vr_i, R_i A_j, n, t_0} = Dr_{in} R_i A_j$, $i = \overline{1, K}$, $n = \overline{1, |Dr_i|}$, $j = \overline{1, Pi}$, где $Dr_{in} R_i A_j$ - значение атрибута $R_i A_j$ в кортеже n отношения r_i , Pi - количество атрибутов в схеме отношения R_i .

Многомерное пространство, описывающее темпоральную реляционную модель данных, накапливает изменения исходной реляционной модели. Чтобы динамически формировать изменённую реляционную модель данных, были разработаны операторы преобразования значений точек многомерного пространства α и β .

Оператор преобразования α получает из определённого множества значений точек многомерного пространства соответствующую ему реляционную модель данных: $\alpha: C_M \rightarrow R_D$. При этом выполняются следующие действия:

1. $C_\eta = \{c_m : v = vr_i, c_m \in C_M\}$, $i = \overline{1, |V_{C_M}|}$.
2. $R_i = S_{C_\eta} \cup \{\text{Идентификатор}\}$, $i = \overline{1, |V_{C_M}|}$.
3. $Dr_{in} = \{c_m : id = id_n, c_m \in C_\eta, c_m \neq null\}$, $(\forall Dr_{in} \neq \emptyset) \Rightarrow Dr_{in} = Dr_{in} \cup \{id_n\}$,
 $Dr_i = \{Dr_{in}\}$, $i = \overline{1, |V_{C_M}|}$, $ID_{C_\eta} = \{id_n\}$, $n = \overline{1, |ID_{C_\eta}|}$.
4. $r_i = (R_i, Dr_i)$, $R_D = \{r_i\}$, $i = \overline{1, |V_{C_M}|}$.

Оператор преобразования β получает из определённого множества значений точек многомерного пространства соответствующую ему реляционную модель данных, состоящую из отношений с темпоральными кортежами: $\beta: C_M \rightarrow H_D$. При этом выполняются следующие действия:

1. $C_h = \{c_m : v = vr_i, c_m \in C_M\}$, $i = \overline{1, |V_{C_M}|}$.
2. $H_i = S_{C_h} \cup \{\text{Период, Идентификатор нетемпорального кортежа}\}$, $i = \overline{1, |V_{C_M}|}$.
3. $\forall id_k, t_p (\{c_m : id = id_k \wedge t = t_p, c_m \in C_h\} \neq \emptyset) \Rightarrow$
 $\Rightarrow Dh_{in} = \{c_m : id = id_k \wedge t = t_p, c_m \in C_h\} \cup \{id_k, t_p\}$,
 $Dh_i = \{Dh_{in}\}$, $i = \overline{1, |V_{C_M}|}$, $ID_{C_h} = \{id_k\}$, $k = \overline{1, |ID_{C_h}|}$, $T_{C_h} = \{t_p\}$, $p = \overline{1, |T_{C_h}|}$.
4. $h_i = (H_i, Dh_i)$, $H_D = \{h_i\}$, $i = \overline{1, |V_{C_M}|}$.

На основе многомерного пространства, описывающего темпоральную реляционную базу данных, разработанных на нём алгебры многомерных пространств и соответствующих преобразований, в диссертационной работе предложена новая методика обработки темпоральной реляционной базы

данных с изменяемой структурой в миварном пространстве, которая представлена на рис. 2.

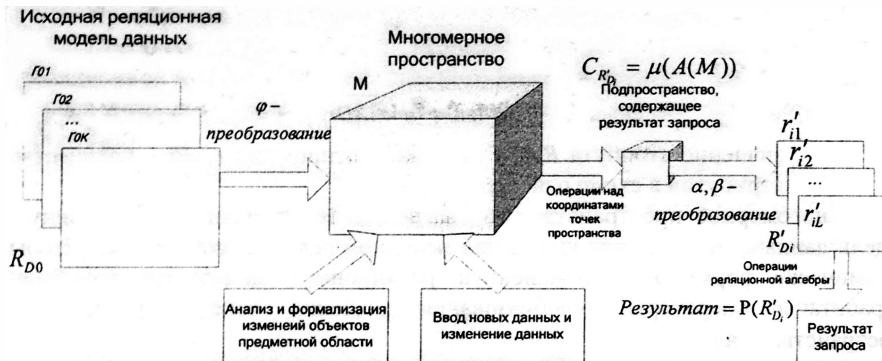


Рис. 2. Методика обработки темпоральной реляционной базы данных с изменяемой структурой в миварном пространстве

Методика состоит из трёх этапов:

1 этап. Преобразование исходной реляционной модели данных в многомерное пространство

Исходными данными для построения темпоральной реляционной базы данных в многомерном пространстве является реляционная модель данных $R_{D0} = (r_{01}, r_{02}, \dots, r_{0k})$. R_{D0} посредством оператора φ преобразуется в многомерную модель данных: $C_0 = \varphi(R_{D0})$.

2 этап. Изменение модели предметной области в многомерном пространстве

Изменение реляционной модели проводится в многомерном пространстве. Процесс изменения реляционной модели происходит посредством добавления новых точек в многомерное пространство. Координаты точек определяют структуру данных.

На рис. 2 изменение многомерного пространства изображено двумя блоками:

1. «Анализ и формализация изменений объектов предметной области». Результатом выполнения данной функции является изменение многомерного пространства: либо изменяется структура пространства (добавляются новые оси), либо происходит изменение множеств, определяющих оси пространства (добавляются новые элементы в эти множества).
2. «Ввод новых данных и изменение данных». Результатом выполнения данной функции является сохранение значений новых точек в многомерном пространстве.

Процесс изменения реляционной базы данных является итерационным (рис. 3), т.е. каждое следующее состояние базы данных зависит от предыдущего: $C_i = \delta(C_{i-1})$, где δ – функция, задающая изменения текущего состояния от предыдущего.

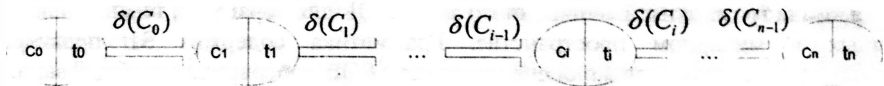


Рис. 3. Процесс изменения модели предметной области

Кроме того, рассматриваемый процесс протекает во времени, т.е. каждое состояние базы данных действительно в определённом интервале времени: $C_i, t \in [t_i, t_{i+1})$. Преобразование реляционной базы данных в многомерное пространство позволяет описать процесс её изменения во времени. При этом состояние базы данных получается из многомерного пространства посредством использования оператора преобразования «Срез последних»: $C_i = \tau_L(C_M, t_i)$.

3 этап. Выполнение запросов к многомерному пространству

Данный этап состоит из трёх подпунктов.

3.1. Выделение подпространства, содержащего результат запроса. На данном этапе выполняются операции над координатами точек в многомерном пространстве (операции алгебры многомерных пространств), не анализируя значения самих точек. В результате из всего многомерного пространства выделяется подпространство, содержащее результат запроса: $C_{R_{D_i}} = \mu(A(M))$, где $A(M)$ - выражение алгебры многомерных матриц.

3.2. Преобразование выделенного подпространства $C_{R_{D_i}}$ в реляционную модель данных. С помощью операторов α и β выделенное подпространство преобразуется в реляционную модель данных $R'_{D_i} = \alpha(C_{R_{D_i}}) = (r'_{i1}, r'_{i2}, \dots, r'_{ik})$, либо в реляционную модель данных с темпоральными кортежами $H'_{D_i} = \beta(C_{R_{D_i}}) = (h'_{i1}, h'_{i2}, \dots, h'_{ik})$.

3.3. Получение окончательного результата запроса. Окончательный результат запроса вычисляется из полученной реляционной модели данных посредством операций реляционной алгебры: *Результат* = $P(R'_{D_i})$ или *Результат* = $P(H'_{D_i})$, где $P(R'_{D_i})$ и $P(H'_{D_i})$ - выражение реляционной алгебры.

Таким образом, методика включает однократное преобразование начальной реляционной базы данных в многомерное пространство и дальнейшую работу с её многомерным представлением: ввод новых данных, изменение структуры данных и выполнение запросов к многомерному представлению темпоральной реляционной базы данных.

В третьей главе разрабатывается грамматика расширенного языка SQL для выполнения запросов к миварному пространству и описывается метод реализации многомерного пространства на реляционных СУБД. Для предложенной структуры базы данных проводится реализация операций алгебры многомерных пространств и соответствующих преобразований, а также проводится разработка и реализация операторов изменения темпоральной реляционной базы данных через миварное пространство.

На основе алгебры многомерных пространств и операторов преобразования значений точек пространства в реляционную модель

разработана грамматика расширенного языка SQL для обработки темпоральных данных в миварном пространстве. Грамматика содержит 31 правило, описывающие стандартные операции языка SQL, операции преобразования многомерного пространства в реляционную модель и операции алгебры многомерных пространств. В результате становится возможным работать в одном запросе и с отношениями реляционной модели, и с точками многомерного пространства, из которых динамически формируются реляционные отношения.

В диссертационной работе было предложено использовать ROLAP технологии для реализации многомерного пространства темпоральной реляционной базы данных. При таком подходе применяется схема «звезда», которая состоит из центральной таблицы, описывающей факты (точки многомерного пространства) и нескольких связанных с ней таблиц, описывающие измерения (элементы осей многомерного пространства).

Для предложенной структуры базы данных были разработаны способы реализации операций алгебры многомерных пространств на языке SQL и алгоритмы преобразования значений точек многомерного пространства в реляционную модель данных. На базе реализации операций алгебры и преобразований значений точек был разработан интерпретатор для предложенного языка, который преобразует запрос к многомерному пространству в запрос языка SQL.

В диссертационной работе для модификации отношений темпоральной реляционной базы данных по аналогии с языком SQL были разработаны и реализованы пять операторов изменения темпоральной реляционной базы данных через миварное пространство: добавление новых записей в отношение (*MININSERT*), обновление существующих записей в отношении (*MUPDATE*), удаление существующих записей из отношения (*MDELETE*), удаление атрибутов из отношения (*MDROP COLUMN*), изменение многомерного пространства (*MDIMENTION*).

Разработанные операторы выполняют различные операции с темпоральной базой данных унифицированным способом – записывают новые значения точек многомерного пространства. При таком подходе изменение данных и схемы темпоральной реляционной базы данных выполняется одновременно посредством применения одного из рассмотренных операторов, что снижает затраты времени на адаптацию информационной системы, так как не нужно отдельно изменять схему базы данных. Реализация рассмотренных операторов была выполнена на языке SQL для предложенной структуры базы данных многомерного пространства.

В четвертой главе выполнена реализация многомерного пространства и интерпретатора грамматики языка запросов средствами технологической платформы «1С: Предприятие 8.2». Проведена практическая апробация предложенной методики и операторов языка запросов на примере кадровой системы Московского государственного технического университета имени Н.Э.Баумана и выполнена оценка эффективности методики по обработке

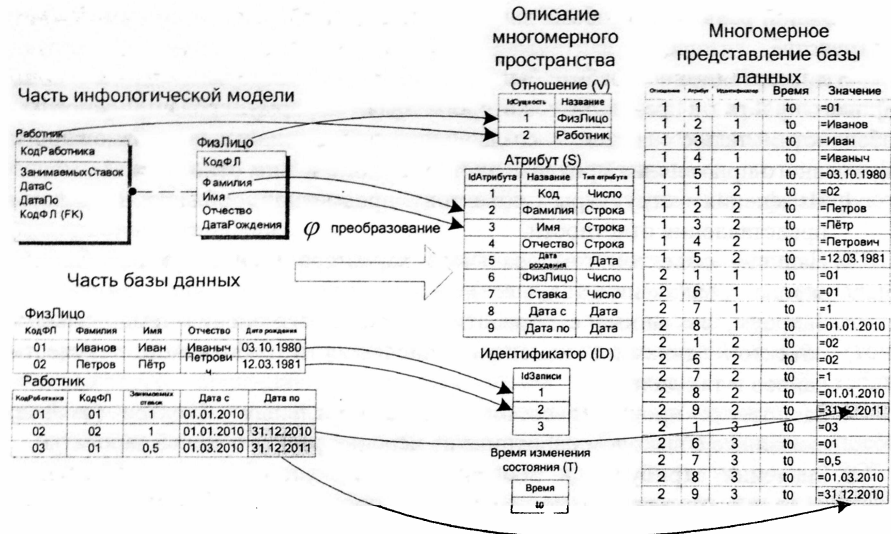
темпоральной реляционной базы данных с изменяемой структурой в миварном пространстве.

В диссертационной работе был проведён анализ информационных систем кадрового учёта, в результате которого были сформулированы основные особенности таких систем, создающие принципиальные возможности эффективного применения предложенной методики. К ним относятся:

1. Изменения структуры объектов предметной области и форм представления отчётности.
2. Наличие объектов темпорального характера (данные о работниках, зарплата, вакансии и так далее).
3. Запросы, связанные с обработкой истории изменения темпоральных объектов (динамика изменения численности работников, текучести кадров и так далее).
4. Запросы к прошлым состояниям темпоральных объектов (отчёты, связанные с выходом работника на пенсию, расчёт средней зарплаты).
5. Значения атрибутов в последовательных состояниях темпоральных объектов отличаются незначительно (продление трудовых договоров, групповое назначение надбавок, реорганизация структурных подразделений и так далее).

В работе была выполнена практическая реализация предложенной структуры многомерного пространства для темпоральной реляционной базы данных и разработана программная реализация интерпретатора и операторов работы с темпоральными данными на миварном пространстве в рамках технологической платформы «1С:Предприятие 8.2». Для практической проверки предложенной методики была использована база данных системы кадрового учёта МГТУ им. Н.Э. Баумана, которая посредством разработанного оператора φ была преобразована в многомерное представление. Фрагмент φ -преобразования базы данных системы кадрового учёта представлен на рис. 4.

В результате, на примере кадровой базы данных было выполнено практическое использование предложенной методики для обработки темпоральных данных в миварном пространстве. Была показана возможность получения истории работы сотрудника в условиях изменения места работы и структуры организации, а также возможность оперативной адаптации такой системы к новым требованиям.



Р

ис. 4. Пример φ -преобразования части реляционной модели системы кадрового учёта в многомерное пространство

Для определения границ применимости предложенной методики были разработаны модель оценки размера отношения с темпоральными кортежами и модель оценки размера этого отношения в многомерном пространстве (рис. 5а), которые зависят от пяти параметров:

1. o – количество объектов предметной области, хранимое в отношении. Диапазон изменения этого параметра $[1, n]$.
2. p – частота изменения значения атрибута в двух кортежах, определяющих последовательные состояния объекта предметной области. Диапазон изменения этого параметра $(0, 1]$.
3. c – частота изменения схемы отношения при допущении, что схема отношения изменяется равномерно, каждые c кортежей. Диапазон изменения этого параметра $(1, \infty)$.
4. a – начальное количество атрибутов в схеме отношения. Диапазон изменения этого параметра $(1, \infty)$.
5. n – количество кортежей в отношении.

Размер отношения вычисляется как количество хранимых значений атрибутов.

Результаты анализа эффективного применения методики приведены в виде графиков: зависимость минимального количества кортежей в отношении (N_{min}) от частоты изменения значения атрибута при различных значениях параметра c (рис. 5б) и при различных значениях параметра o (рис. 5г). Размер реляционного отношения, описывающего работника, уменьшился при $n=50000$ на 26%, а при $n=100000$ на 48%. На рис. 5в представлены графики времени

адаптации информационной системы при пяти изменениях схемы БД для реляционной базы данных и многомерного пространства. Время адаптации информационной системы при использовании разработанной методики в среднем уменьшилось на 53%.

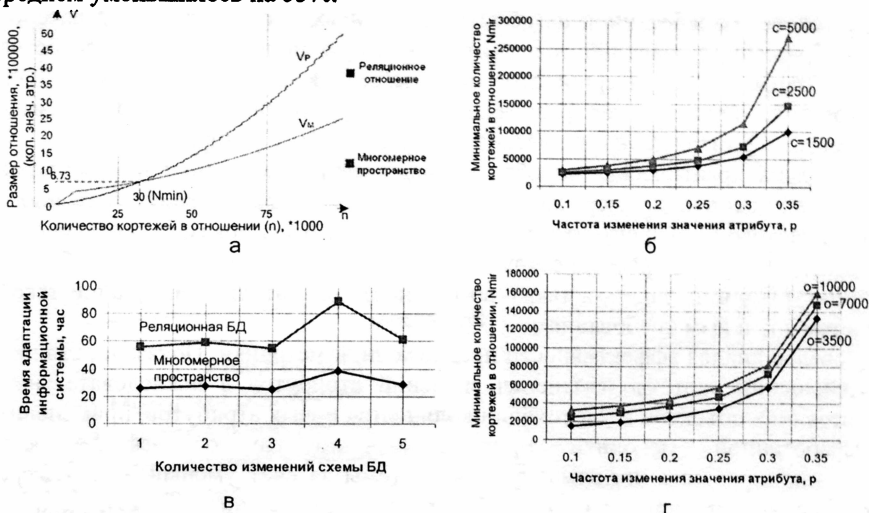


Рис. 5. Анализ результатов моделирования

Таким образом, результаты проведённых экспериментов показывают, что применение разработанной методики позволяет уменьшить размер отношений реляционной модели в среднем на 32% и сократить время адаптации информационной системы в среднем на 53%.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. В результате исследования и анализа принципов построения информационных систем, обрабатывающих темпоральные данные, предложена принципиально новая методика работы с темпоральными данными, основанная на использовании миварного информационного пространства.
2. Предложен способ преобразования реляционных баз данных в многомерное миварное пространство для эффективного выполнения операций над темпоральными данными с изменяющейся структурой.
3. Разработана методика обработки темпоральной реляционной базы данных в миварном пространстве с обратным преобразованием в реляционную модель. Предложены процедуры прямого и обратного преобразования реляционной модели в миварное пространство. Предложенная методика не требует модификации исходной реляционной базы данных для выполнения запросов к темпоральным данным с изменяющейся структурой.
4. Разработана алгебра многомерных пространств для определения подпространства, содержащего результат запроса, и преобразования его во множество значений точек для последующего формирования реляционной модели.

5. Разработаны операторы изменения темпоральной базы данных через миварное пространство, при которых изменение данных и схем отношений происходит одновременно, что сокращает затраты времени на адаптацию информационной системы.
6. На основе разработанной алгебры многомерных пространств и операторов преобразования значений точек в реляционную модель данных разработан язык запросов и интерпретатор для него, как расширение языка SQL. Проведена практическая реализация интерпретатора и операторов языка запросов в рамках технологической платформы «1С:Предприятие 8.2».
7. Проведена практическая апробация предложенной методики и операторов языка запросов на примере кадровой системы Московского государственного технического университета имени Н.Э.Баумана. Была показана возможность получения истории работы сотрудника в условиях изменения места работы и структуры организации, а также возможность оперативной адаптации такой системы к новым требованиям.
8. Проведен анализ эффективности предложенной методики. Показано, что при уменьшении частоты изменения значений атрибутов в последовательных состояниях объектов и увеличении количества новых атрибутов, применение разработанной методики приводит к повышению эффективности функционирования информационной системы за счёт уменьшения размера отношений реляционной модели и времени модернизации базы данных.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Елисеев Д.В. Модель представления знаний при создании адаптивной информационной системы [Электронный ресурс], - Электрон. журн. – М.: «Наука и образование: электронное научно-техническое издание», 2010 – Режим доступа: <http://technomag.edu.ru/doc/139659.html>, свободный (Дата обращения: 06.10.2011).
2. Елисеев Д.В., Балдин А.В. Адаптируемая модель данных на основе многомерного пространства [Электронный ресурс], - Электрон. журн. – М.: «Наука и образование: электронное научно-техническое издание», 2010 – Режим доступа: <http://technomag.edu.ru/doc/161410.html>, свободный (Дата обращения: 06.10.2011).
3. Елисеев Д.В., Балдин А.В. Алгебра многомерных матриц для обработки адаптируемой модели данных [Электронный ресурс], - Электрон. журн. – М.: «Наука и образование: электронное научно-техническое издание», 2011 – Режим доступа: <http://technomag.edu.ru/doc/199561.html>, свободный (Дата обращения: 06.10.2011).
4. Елисеев Д.В., Балдин А.В. Адаптация темпоральной реляционной модели данных в многомерном пространстве [Электронный ресурс], - Электрон. журн. – М.: «Наука и образование: электронное научно-техническое издание», 2011 – Режим доступа: <http://technomag.edu.ru/doc/225699.html>, свободный (Дата обращения: 06.10.2011).