На правах рукописи

Полтавцева Мария Апатольевпа

ХРАНЕНИЕ СЛОЖНЫХ СТРУКТУР ДАННЫХ В РЕЛЯЦИОННЫХ БАЗАХ ДАННЫХ

Специальность 05 13 01 - Системный анализ, управление и обработка информации (в промышленности)

АВТОРЕФЕРАТ диссертации на соискание ученой степени кандидата технических наук



На правах рукописи

Полтавцева Мария Анатольевна

ХРАНЕНИЕ СЛОЖНЫХ СТРУКТУР ДАННЫХ В РЕЛЯЦИОННЫХ БАЗАХ ДАННЫХ

Специальность 05 13 01 - Системный анализ, управление и обработка информации (в промышленности)

АВТОРЕФЕРАТ диссертации на соискание ученой степени кандидата технических наук

Работа выполнена в Тверском государственном техническом университете

Научный руководитель

Доктор технических наук, профессор,

Григорьев Вадим Алексеевич

Официальные оппоненты

Доктор технических наук, профессор,

Семенов Николай Александрович

Кандидат технических наук Орлов Юрий Петрович

Ведущая организация

ОАО «Редкинское опытно-конструкторское

бюро автоматики»

Защита состоится 27 июня 2008 г в 14-00 на заседании диссертационного совета Д 212 262 04 в Тверском государственном техническом университете по адресу 170026, г Тверь, наб Аф Никитина, 22 в аудитории 212

С диссертацией можно ознакомиться в библиотеке Тверского государственного технического университета

Автореферат размещен на сайте ТГТУ, адрес http://www.tstu.ru/new_struct/phd/27062008 zip

Автореферат разослан *17 шая* 2008 г

Ученый секретарь диссертационного совета

de la companya dela companya dela companya dela companya de la companya de la companya de la companya dela companya de la companya dela companya

ФилатоваН Н

ОБШАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Сегодня организации не могут функционировать без централизованных ресурсов основных данных, интегрированных с архитектурой предприятия, в которой решения по управлению основными данными могут обеспечивать не только представление данных для бизнес-процессов, но и выполнение нормативных требований и подготовку достоверных отчетов для руководства организаций и компаний, регулирующих органов и акционеров

Особое место занимает так называемая нормативно-справочная информация, или основные данные (master data) словари, справочники и классификаторы описывающие ключевые понятия бизнеса (объекты и субъекты деятельности компании или организации) От их точности и согласованности зависит функционирование практически всех бизнес приложении и аналитических систем Например, американские компании ежегодно тратят ботьше 600 млрд долл на обеспечение качества данных и, по оценкам Gartner, к 2010г более 70% компаний из списка Fortune 1000 реализуют программы по управлению основными данными как часть корпоративной стратегии управления информацией

Разработаны и существуют наборы алгоритмов для работы в информационных системах со сложно структурированьой информацией, описываемой в общем случае графами, но на сегодня не является достаточно исследованным вопрос эффективного хранения подобных структур данных

Для хранения информации с развитием компьютерной техники применятся базы данных Каждая из них построена на основании той или иной модели данных, опредетяющий способ формализации информации С 80гг двадцатого века и по сеи день наиболее широко используемой и единственной имеющей математический аппарат для описания операций (реляционную алгебру и реляционное исчисление) является реляционная модель данных

Реляционная модель данных обладает при всех достоинствах одним существенным недостатком - она является плоской Все данные в ней хранятся в виде связанных таблиц Это вызывает существенные трудности при хранении в реляционной базе данных боле сложных информационных структур очередей и списков, деревьев и графов, сетей и г д

Поэтому насущной задачей является разработка архитектур, методик проектирования и реализации систем хранения сложно структурирозанной информации, опирающихся на резяционные СУБД Ключевыми проблемами в реализации данного подхода являются

- Анализ и систематизация основных методов структурирования информации, предлагаемой для хранения в реляционной СУБД Выбор методов структурирования для исследования
- 2 Разработка методов и алгоритмов принятия решений и обработки информации при представлении структурированной информации в реляционных слемах

3 Исследование вопросов создания специального математического и программного обеспечения систем, обеспечивающих их реализацию, в том числе архитектур хранения и методов преобразования

<u>Пель работы</u> Работа посвящена интеграции методов представления структурированной информации и реляционных серверов баз данных на основе отражения сложно структурированных типов данных в реляционную модель Цель достигается через отбор оптимальных схем хранения и создание методик проектирования информационных систем, использующих структурированную информацию при их разработке и реализации, и имеющих в своем составе реляционный сервер базы данных

В соответствии с указанной целью определены следующие <u>задачи</u> исследований:

- 1 Проанализировать существующие способы структурирования информации Выделить наиболее типовые структуры и построить обобщенную классификацию, которая может быть использована для выбора реляционных схем хранения
- 2 Исследовать задачи хранения классифицированной и иерархически организованной информации в реляционных серверах и возможности взаимного отражения соответствующих схем
- 3 Выполнить исследование по поиску наилучших схем хранения, предложив количественные и качественные критерии сравнения и отбора оптимальных На основании результатов исследований отобрать оптимальные (рациональные) методы хранения перархически структурированной информации
- 4 Использовать полученные результаты для предложения рациональных схем хранения графов и сетей
- 5 Объединить выбранные методы и схемы в методику принятия решений и обработки информации при отображении структурированной информации в реляционные схемы
- 6 Сделать предложения по построению и функциональным особенностям специального программного обеспечения для поддержки полученных методов и алгоритмов принятия решений и обработки информации Обосновать его архитектуру и функциональную организацию

Методы исследования включают

- 1 Аналитические методы теорию множеств и математическую логику, графов, сетей и языков программирования, теории моделей данных, нормализации, основные принципы проектирования информационных систем
- 2 Аналитико-экспериментальные методы вычислительный эксперимент в виде имигационного моделирования на ЭВМ объектов и задач исследования

<u>Достоверность</u> полученных результатов определяется корректным применением использованных методов исследования Она подтверждается совпадением результатов вычислительных экспериментов для тех данных, которые имеют аналоги в литературе, что позволяет сделать вывод об адекватности разработанных способов и моделей

<u>Научная новизиа</u> Сонскателем получены следующие результаты, имеющие научную повизну

- 1 Предложены метрики для оценки качества результатов отображения способов структурирования информации в реляционную модель данных и сформулирована задача по поиску наилучших решений В качестве целевого функционала выбрано количество обрашений к диску при выполнении типовой операции над структурой Функционал минимизируется на множестве схем хранения структуры и типовых операций над ней
- 2 Впервые проведены комплексные исследования и сравнительный анализ схем отображения выбраны оптимальные (или рациональные) методы отображения иерархически организованных структур в реляционные схемы Определены принципы выбора схемы хранения в зависимости от преобладающей группы операции над хранимыми данными
- 3 Предложены новые ал оритмы хранения информации, представленной в виде графов и сетей Схему хранения предлагается создавать на основе алгоритма представления графа в виде леса деревьев с независимой и съвозной нумирацией на основе метода вложенных множеств
- 4 Построена обобщенная методика моделирования структурированной информации данными методами Методика предназначена для отображения семантической модели в логическую модель данных при проектировании корпоративных информационных систем, опирающихся на реляционные сервера

На защиту выпосятся:

- Методика исследования, показатели и результаты сравнительного анализа методов хранения иерархически организованной информации в реляционной модели
- 2 Методика выбора рациональной схемы хранения данных со структурой ациклического направленного графа в реляционном сервере базы данных
- 3 Результаты исследований и предложения по хранению в реляционных базах данных графовых и сетевых структур
- 4 Элементы специального программного обеспечения виде набора модулей на языке хранимых процедур для принятия решений и опгимизации схем хранения структурированной информации
- 5 Результаты практического применения предложенных методик, алгоритмов и структур хранения при разработке информационной системы управления взаимосвязанным электронным документооборотом

<u>Практическая ценность</u> Разработанные методики позволяют снизить сроки разработки информационных систем повысить их качество и эффективность, обеспечить семантическую, сущностную и ссылочную целостность хранения сложно структурированных типов данных Результаты исследований важны прежде всего организациям, занимающимися созданием

корпоративных информационных систем Они могут быть также использованы разработчиками СУБД – реляционных, ХМL, объектных

Реализация результатов работы На основе разработанных в диссертации положений получены решения по созданию программного обеспечения для построения реляционных баз данных Разработанные методики использованы в разработке программного комплекса управления взаимосвязанным электронным документооборотом в Тверском государственном университете на кафедре документационного обеспечения управления

Апробация работы Работа докладывалась на XVII Международной научно-технической конференции «Математические методы и информационные технологии в экономике, социологии и образовании», Пенза, май 2006г, XX Международной научно-технической конференции «Математические методы и информационные технологии в экономике, социологии и образовании», Пенза, декабрь 2007г, юбилейной научнопрактической конференции ТГТУ "Региональная система профессионального технического образования", Тверь, ТГТУ, декабрь 2007г

<u>Публикации</u> Основные положения диссертации опубликованы в 8-ми печатных работах, в том числе 1статья в журнале из перечня ВАК

Структура и объемы работы. Диссертационная работа состоит из введения, четырех глав, заключения, двух приложений и списка литературы Общий объем диссертации 181 страньца, в том числе 61 рисунка, 19 таблиц, список литературы из 127 наименований

СОДЕРЖАНИЕ РАБОТЫ

<u>Во</u> введении обоснована актуальность темы диссертации, сформулированы цель, основные задачи, методы проведения исследовании, определены научная новизна и практическая ценность результатов работы, приведены сведения об их реализации и апробации и представлены основные положения, выносимые на защиту

<u>В первой главе</u> на основе литературных данных анализируются требования, выдвигаемые к хранению данных современными корпоративными информационными системами, а также предоставляемые для этого возможности и ограничения существующих реляционных СУБД

Показано, что существует широкий круг задач, сталкивающийся с необходимостью хранения структурированных данных обладающих сложной внутренней структурой (деревьев, графов, сетей)

Дтя того чтобы дать возможность приложениям иметь доступ к получающимся комбинированным источникам данных некоторым унифицированным способом и с помощью средств высокого уровня, эти разнообразные форматы должны быть интегрированы Необходимы средства управления схемой, которые могут адаптироваться к динамической природе внешних данных, т є необходима компьютерная поддержка концептуального

проектирования, дающая семантически насыщенный и адаптирующийся по форме и по уровню абстрагирования инструмент При этом должны быть решены вопросы, являющиеся традиционными для систем управления базами данных, а именно проблемы производительности, корректности, надежности

В тоже время реляционные системы имеют относительно слабые возможности по части представления семантики приложения обусловленные равноправием атрибутов отношений, ограниченностью набора базовых типов и отсутствием возможности его расширения, отсутствием списочных структур и иерар\ической подчиненности типов данных, а также фиксированностью методов хранения и доступа к данным Реляционная модель требует полной нормализации данных, однако, всякая используемая на ЭВМ модель необходимо имеет некоторый определенный уровень абстракции по отношению к моделируемой области, а это противоречит нормализации

Поэтому существующие решения не отвечают полностью требовапиям компьютерной поддержки проектирования и использования семантически насыщенных структур ланных Вопрос хранения связных ациклических графов рассматривался в ряде работ, однако сравнительного анализа этих способов, позволившего бы выбрать оптимально решение не производилось Поэтому отсутствует систематическии подход к выбору схемы хранения структурированных данных в РСУБД Все это приводит к тому, что в задачах используются различные схемы хранения без рассмотрения вопросов выбора оптимального решения, обоснования и выбора той или иной реляционной схемы

Вторая глава посвящена анализу структур данных, их программной реализации и алгоритмов работы с ними в РСУБД Рассмотрены типовые задачи с использованием анализируемых структур и операции над ними В результате предложена классификация структур данных на основании способа программного представления, пригодная для использования в задачах выбора схем представления данных

При проектировании информационной системы строятся модели для внешних схем и концептуальной схемы Концептуальные модели отображаются в табличные Габличные схемы преобразуются во внутренные Однако, если для хрансния данных используется реляционный сервер, то применимость технологии проектирования зависит от возможности перенесения ее базовых принципов и понятий на реляционную модель При отображении проектной модели в реляционную необходимо выбрать одну из множества возможных аргют правил отображения

Базовыми понятиями семантических модечей явчяются понятия объектов и связей между объектами Выделение данных понятий и конструирование модели основывается на абстракциях обобщение, агрегация, ассоциация Реляционные модечи строятся из посылки обеспечения целостности данных, а также эффективности выполнения запросов Обеспечение целостности достигается ликвидацией избыточности через

процедуру нормализации Реляционная модель не поддерживает сколь-нибудь развитую семантику межобъектных отношений Поэтому на данном этапе происходит искажение модели прикладного домена, и появляется несоответствие между моделью предметной области и моделью данных (т н "mismatching impedance")

Отражение проектной схемы в, или эксграгирование ее из хранимых в виде регяционной модели данных — это определение какие проектные конструкции должны быть отражены в (получены из) реляционную схему Конструкциями, представляющими наибольший интерес, являются статические структуры данных (векторы, массивы, множества, записи, таблицы), полустатические структуры (стеки, очереди, деки, строки), динамические структуры (линейные связные списки, разветвленные связные списки, графы и сети) Отдельно выделяется подкласс связных ациклических графов — т н деревьев как наиболее широко применяемый в семантических моделях тип данных (рисунок 1)



Рисунок 1 Классификация структур данных в соответствии с используемой схемой хранения

Над любыми структурами данных могут выполняться четыре общие операции создание, уничтожение, выбор (доступ), обновление Помимо этих обязательных для всех структур и типов данных общих операций для каждой структуры данных могут быть определены операции специфические, работающие только с данными данного типа (данной структуры)

Например, важнейшей операцией для записи является операция доступа к выбранному полю записи - операция квалификации Основной операцией при работе с таблицами является операция доступа к элементу по ключу, реализуемой процедурой поиска Операции над деком включение элемента справа, включение элемента справа, исключение элемента справа, исключение элемента справа, определение размера, очистка

В работе в качестве предмета исследования выделены графы и сети, и, в частности, важнейший их подкласс – деревья, как структуры представляющие наибольший интерес в семантических моделях Типовыми задачами над графами являются добавление вершины, удаление вершины, добавление дуги, удаление дуги, определение смежности вершин, определение инцидентности узла ребру, нахождение пути между двумя вершинами, определение наличия

циклов в графе и нахождение их, нахождение подграфа обладающего заданными свойствами (например, остовного дерева)

Леревья используются для описания наиболее часто встречающейся в семантических моделях структуры - иерархии При работе с деревьями возникает четко очерченный круг задач, выполняющихся с разной частотой в зависимости от придожения в целом Приложение по работе с деревьями (в том числе моделирование в РСУБД) должно предоставлять возможность работы со всеми из них, с той или иной эффективностью в зависимости от задачи в целом добавить/удалить узел дерева, найти узел внутри дерева, определить, является ли данный элемент терминальным (листом), определить, на каком уровне иерархии находится заданный элемент, подсчитать количество всех потомков у данного элемента, получить список всех потомков какого-нибудь элемента, при этом с разными условиями например, только терминальных, или, находящихся только на определенном уровне или всех потомков заданного элемента, у которых нет детей, получить всех предков данного элемента Распространенная задача - "выполнение заданной операции Р с каждым элементом дерева" Здесь Р рассматривается как параметр более общей задачи посещения всех узлов, или, как это обычно называют обход дерева В том или ином виде операция обхода дерева присутствует во всех задачах, работающих с деревьями

<u>В третьей главе</u> исследованы методы отражения выделенных семантических конструкций (графов и прежде всего, деревьев) в реляционную модель Предложены критерии оценки качества методов отображения, произведен анализ по данным критериям и выбраны оптимальные (рациональные) методы отражения композитных структур в реляционные отношения

Анализируются методики хранения деревьев в РБД, как с помощью рекурсивных моделей, так и по методу вложенных множеств Рассмотрена возможность использования для построение схем хранения нумерации элементов дерева на основании алгоритма обхода в ширину Рассмотрены возможные представления графов в реляционной модели, определены способы выполнения операции над графами в РСУБД, поведен сравнительный анализ методик

Хранение иерархических данных в реляционной базе данных является общеизвестной проблемой Есть два главных подхода модель списка смежных вершин (модель рекурсии — рисунок 1) и модель основанная на понятии вложенных множеств И хотя по данной теме написано не мало, до сих пор основным способом является хранение с помощью самосвязанной таблицы Данная модель называется также моделью списка смежных вершин следуя терминологии теории графов пары узлов смежны друг с другом

К недостаткам рекурсивного подхода к получению данных из модели списков смежности является медленность и неэффективность Для каждого узла, функция запускает отдельный экземпляр себя Следовательно для

каждого узла в дереве необходим отдельный запрос к базе данных Поскольку каждый запрос занимает время, это делает процедуру медленной при работе с деревьями с большим количеством узлов Кроме того, каждый экземпляр функции запимает место в памяти и требует время на ее инициализирование, те тратится много дополнительной памяти, а также времени на ее выделение/освобождение Все это и делает рекурсию очень загратной по временным и техническим ресурсам при обработке больших иерархических структур

Рисунок 2 Схема хранения нерархии согласно модели списка смежных вершин

Предложены многочисленные модификации основной рекурсивной модели Все очи, упрошая и улучшая одни операции ухудшают и усложняют другис и по совокупности показателей не дают существенного улучшения относительно базовой модели

Другой путь представления деревьев состоит в том, чтобы показывать их как вложенные множества, Это более подходящая модель, т к SQL - язык работы с РСУБД - язык, ориентированный на множества При реализации в реляционной базе данных модель вложенных множеств трансформируется в модель вложенных интервалов на основе алгоритма обхода графа в глубину в прямом порядке

Критерием сравнения моделей хранения выбрано число обращений к диску (дисковых операций), соответственно определяющее и скорость работы приложения Сравнение данных методов для операции определения терминальности узла представлено в таблице 1

Таблица 1 Сравнение схем хранения

Матрица смежности-	Метод вложенных множеств	Метод	
Определение терминальности узла		Название	
Лист яв тяется терминальным	Описание		
~∃С (C parent = "Узел ID")	~∃D(D parent = "Узел ID")	Реляционное выражение	
Select C ID From Tree	Select D ID From Tree	Запись на языке запросов	
where No Exist C parent = ID	where No Exist D parent = ID		
σ~3C ID(C Parent=C ID)(C)	σ-3D ID(D Parent=D ID)(D)	Операцы режилонией - ангебры — —	
Считывание табт С в память	Считывание табл D в память	Операции с диском-	
$U_1=M_2$	$U_2=M_1$	Число операции с диском	
Матрица смежности выгодне	Результат		

В работе показано, предпочтительными способами хранения древовидных структур являются схема в виде матрицы смежности с использованием вспомогательной таблицы, и схема полученная по методу вложенных множеств (вложенных интервалов)

К произвольному графу данные схемы напрямую не применимы, так как в них используются ограничения иерархической структуры именно апикличность и связность

Используя матрицу смежности для хранения графов в РСУБД представим ее связным списком вершин Ключом такого реляционного идентификаторов исходящей сочетание вершины является отношения (родителя) входящей (потомка) Для ненаправленного графа, И подразумевающего возможное повторение этих сочетаний, введем отдельное идентификационное поле

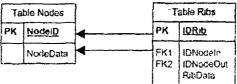


Рисунок 3 Схема хранения графа списком узлов и списком ребер

В этом случае ID_Out и ID_In соответствующие обозначениям входящих и исходящих вершин по матрице соответствуют ID_el соответствующих узлов графа в основной таблице с данными (рисунок 3)

Случаи хранения графа с использованием связного (списочного) представления сводится к хранению структуры, напоминающей нелинейный разветвленный список Получаемая структура имеет два типа элементов (элемент графа — ссылка) и (ссылка-ссылка) Тем не менее, полученная структура нелинейным разветвленным списком не является, так как в ней могут наличествовать петли и обратно направленные связи (рисунок 4)

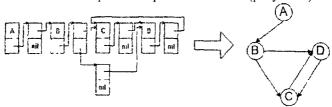


Рисунок 4 Представление графа нелинейным списком

Связи B->D и D->C не соответствуют представлению в виде дерева и не соответствуют теории нелинейных разветвленных списков

Реляционная схема хранения приведена на рисунке 5 Основные данные (элементы списка) хранятся в таблице «Граф» Связанная таблица «Граф элементы графа» используется совместно с основной для отдельного хранения элементов, отсутствующих в случае сочетаний ссылка-ссылка Поле «Туре» имеет бинарное значение 0- ссылка-элемент, 1- ссытка-ссылка

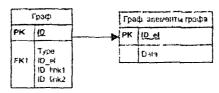


Рисунок 5 Схема хранения графа в виде нелинейного списка

Наконец еще одним вариантом является использование для хранения представление графа в виде леса связанных деревьев (рисунок 6) В случае такого представления рационально использовать схему хранения деревьев, усовершенствовав ее для представления нескольких связанных иерархий

В работе выявлены две наиболее рациональные схемы хранения иерархических структур, имеющие разную эффективность, в зависимости от типа решаемой задачи В первом случае, для хранения дерева, используются две таблицы основная и вспомогательная В основной таблице хранится только непосредственный предок узла, а во вспомогательной каждому узлу соотнесены все его предки по направлению к вершине Для хранения графа расширим эту схему на идентификатор дерева, использующийся совместно с номером узла, к которому он принадлежит Таким образом, применяемая для хранения графов схема будет иметь вид представленный на рисунке 4

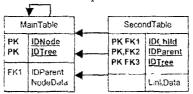


Рисунок 6 Схема хранения графа в виде леса деревьев по методу матрицы смежности

Во втором случае, согласно методу вложенных множеств (МВМ) каждому элементу иерархии в зависимости от его места в ней присваиваются лва числовых значения левое и правое Каждое дерево, входящее в граф, получает эти значения независимо от структуры самого графа (рисунок 7)

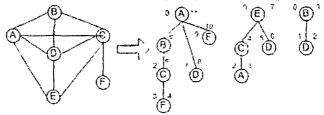


Рисунок 7 Независимая нумерация поддеревьев графа

В этом случае, модернизация схемы хранения иерархий сводится к добавлению нового поля - идентификатора дерева (рисунок 8)

- <u>I</u> :	Structure
PK PK	ID Tree ID Node
FK1	ID PNode Left Right

Рисунок 8 Схема хранение графа в виде леса деревьев по методу вложенных множеств

Логично использовать для всех деревьев графа единую, сквозную нумерацию, учитывающую все деревья его составляющие (рисунок 9)

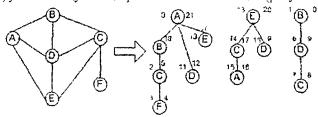


Рисунок 9 Сквозная нумерация поддеревьев графа

Представления матрицей смежности и списками ребер и вершин рационально объединить в одно ибо для этих представлений используется одна и та же схема хранения Назовем этот объединенный способ "хранение при помощи списков ребер и вершин", что наиболее точно отвечает сути схемы хранения Исключим из окончательного рассмотрения связное (списочное) представление графа из-за выявленной особой сложности работы с ним Таким образом, список представлений графа, которые необходимо подвергнуть сравнительному анализу будет представление графа списками ребер и вершин, представление в виде леса связанных деревьев на основании метода матрицы смежности со вспомогательной таблицей, представление в виде леса связанных деревьев на основании метода вложенных множеств с независимой нумерацией, представление в виде леса связанных деревьев на основании метода вложенных множеств со сквозной нумерацией

Критерием сравнения служит число обращений к диску, определяющее соответственно и компактность хранения и скорость работы приложения

Результаты проведенного сравнения представлены в сводной таблице Выделены наилучшие варианты хранения на каждом этапе (таблица 2)

Проведенный анализ выявил полезность с точки зрения составления запросов метода представления в виде леса деревьев по методике матрицы смежности, и наилучшее поведение в задаче поиска пути в графе метода представления в виде леса деревьев с использованием метода вложенных множеств и сквозной нумерации

Таблина 2 Сравнение схем хранения графов

Метод	Списки ребер и вершин	Лес деревьев и метод матрицы смсжности	Лес дер и метод вл мн (независ	Лес дер и метод вл ми (сквозная
Добавление вершины	 U ₁ =1	$U_2 = M_3 + 1$	нумерация) U3=M3+1	пумерация) U ₄ =M ₃ +1
Удаление вершины	$U_1=1+2*M_2$	$U_2 = M_3 + 2 M_3 + 1$	U,=2*M ₃	$U_4=2*M_3$
Добавление дуги	U ₁ =1	$U_2 = M_3 + M_3 + 1$	$U_3=2*M_1$	U ₄ =2*M ₃
Удаление дуги	U ₁ =1	$U_2 = M_3 + M_3$	$U_2 = 2*M_3+2$	$U_4 = 2*M_3 + 2$
Опредет смежности вершин	$U_i=M_2$	U ₂ =M,	U ₃ =M ₃	U ₄ =M ₃
Определ инцидентности узла к ребру	U ₁ =M ₂	U ₂ =0	U,=0	U ₄ =0
Выделение вершин с тотько входными дугами	$U_1 = M_{1+}M_2$	$U_2=M$,	U3=M3	U ₄ =M ₃
Выделение вершин с только выходными дугами	$U_1 = M_1 - M_2$	$U_2 = M_3$	<i>U</i> ₃≈ <i>M</i> ₃	U ₄ =M ₃

<u>В четвертой главе</u> даны примеры практического использования хранения деревьев, лесов деревьев графов в реляционной базе данных Даны готовые решения по работе со структурированными данными на языке SQL Описана работа конкретных составляющих, показана эффективность решений принятых в предыдущих главах работы Реализована задача электронного документооборота в организации, одна из функциональных составляющих которой показана на диаграмме (рисунок 10)

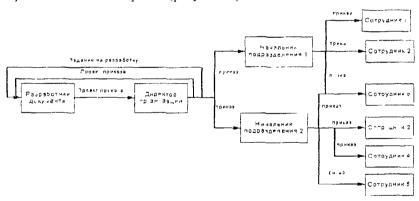


Рисунок 10 Диаграмма движения документа

Типовой круг задач, возникающих при работе с диаграммой добавление и удаление участника (сотрудника), добавление и удаление пути документа, определение передается ли документ между двумя сотрудниками или конкретно от одного сотрудника к другому, определение кому передается документ от конкретного сотрудника, определение от кого передается документ конкретному сотруднику, определение кто имеет доступ к документу

определенного типа (например, проекту приказа или опубликованному приказу), нахождение пути документа определенного типа

Это основные типовые задачи, реализация которых требуется для используемых схем хранения Все они, так или иначе, сводятся к задачам работы с графами

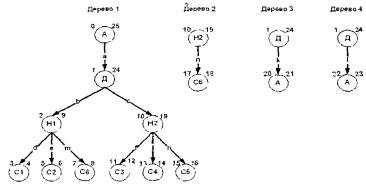


Рисунок 11 Представление графа движения документа в виде леса деревьев

Прежде чем записать данные в таблицу следует представить искомый граф в виде леса деревьев (рисунок 11) Также после преобразования необходимо пронумеровать вершины полученных деревьев согласно модернизированному для графов алгоритму нумерации вершин на основании обхода дерева в глубину

При использовании хранения в виде леса деревьев с применением для иерархий метода вложенных множеств со сквозной нумерацией проектируются таблицы, оперирующие всеми необходимыми данными

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

- 1 Проанализированы существующие способы структурирования информации и классифицированы по существующим методам хранения в реляционных СУБД, построена обобщенная классификация, которая может быть использована для выбора реляционных схем хранения
- 2 Предложены критерии качества и выполнены исследования и сравнительный анализ возможных методов отражения композитных структур в реляционные схемы
- 3 На основании результатов исследований выделены и обоснованы оптимальные (рациональные) схемы хранения классифицировациой и иерархически организованной информации
- 4 На основании полученных результатов по обработке информации в виде связных ациклических графов предложены алгоризмы хранения графов и сетей произвольной структуры
- 5 Полученные результаты сведены в обобщенную методику принятия решений и обработки информации при отображении структурированной информации в реляционные схемы

- 6 Предложено функциональное наполнение реализующей системы, представляющей собой специальное программное обеспечение поддержки соответствующих методов и алгоритмов структурного моделирования сложно структурированных объектов в реляционной базе данных
- 7 Проведена экспериментальная проверка адекватности и состоятельности алгоритмов и методик

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ ОТРАЖЕНЫ В СЛЕДУЮЩИХ ПУБЛИКАЦИЯХ

- 1 Полтавцева М А Программные реализации схем представления структурированных данных в реляционной базе данных // Международный журнал "Программные продукты и системы" №1, Тверь, 2008 С 20-22
- 2 Полтавцева М А Сравнение методов вспомогательной таблицы и матрицы смежности при хранении иерархий в РСУБД // Сборник статей XIV Международной научно-технической конференции «Математические методы и информационные технологии в экономике, социологии и образовании» (МК-81-94) Пенза, 2004 –С 394-396,
- 3 Поттавцева М А Храненые и обработка классификаторов и классифицированной информации // Вестник Тверского государственного технического университета Научный журнал Вып 7 Тверь ТГТУ, 2005 -С 143-144,
- 4 Полтавцева М А Хранение иерархических структур в РСУБД описание Паттерна проектирования // Сборник статей XVI Международной научно-технической конференции «Математические методы и информационные технологии в экономике, социологии и образовании» Пенза, 2005 –С 423-426,
- 5 Полтавцева М А Применение методов хранения связанных иерархий к представлению сетевых структур в РСУБД // Сборник статей XVII Международной научно-технической конференции «Математические методы и информационные технологии в экономике, социологии и образовании» Пенза, 2006 –С 209-212,
- 6 Полтавцева М А Применение реляционных схем хранения слабоструктурированных данных в задачах автоматизации управленческой деятельности // Становчение и развитие системы управления в России Сборник научных статей Вып 1 Сыктывкар КРАГСиУ, 2007 С 112-116,
- 7 Полтавцева М А Хранение в реляционной базе данных полустатических структур данных и списочных структур // Вестник Тверского государственного технического университета Научный журнал Вып 10 Тверь ТГТУ, 2007 С 239-241,
- 8 Полтавцева М А Использование алгоритма обхода графа в глубину при хранении графов в реляционнои базе данных в виде леса деревьев // Сборник статей XX Международной научно-технической конференции «Математические методы и информационные технологии в экономике, социологии и образовании» Пенза, 2007 С 265-267

Подписано в печать 21 05 08 Заказ № 51 Тираж 100 экз Физ печ л 1,0
Типография Тверского государственного технического университета
170026, г Тверь, наб А Никитина, 22