

75 03
21 ДЕК 1993

На правах рукописи

Чанхиева Фарид Юсуповна

**Методы и модели проектирования
тематических баз данных
при автоматизации деятельности
архивных учреждений**

05.25.02. - Документалистика, документоведение и архивоведение

Автореферат

диссертации на соискание ученой степени
кандидата технических наук

Москва 1998

Работа выполнена в Российском государственном гуманитарном университете

Научный руководитель - доктор экономических наук, профессор
Архинова Н.И.

Официальные оппоненты:

1. Доктор технических наук, профессор - Квасницкий В.И.
2. Кандидат технических наук - Пелехов В.И.

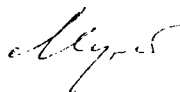
Ведущая организация: Российский государственный архив
научно-технической документации

Защита диссертации состоялась 29 декабря 1998 года в
"15" часов на заседании диссертационного совета К.Д. 064 19 03 в
Российском государственном гуманитарном университете по адресу
125267, Москва, Миусская пл., д.6

С диссертацией можно ознакомиться в библиотеке Российского
государственного гуманитарного университета.

Автореферат разослан 26 ноября 1998 г.

Ученый секретарь
диссертационного совета



В.И.Меркулов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Переход к широкомасштабному применению современных информационных технологий (ИТ) и систем во всех сферах деятельности общества с целью обеспечения широкого круга пользователей наиболее полной, достоверной и своевременной информацией приобретает в настоящее время особую актуальность в связи с происходящими кардинальными преобразованиями в России. Информатизация является основой развития всех сфер общественной жизни и должна рассматриваться не только как информационный и технологический процесс, но прежде всего как процесс, имеющий социальное, экономическое, организационно-правовое, историческое и другие значения.

На современном этапе развития информатизации, когда рынок информационных продуктов и услуг представляет достаточно широкие возможности оснащения средствами технического, программного и информационного обеспечения предприятий и организаций всех форм собственности, основной проблемой информатизации является переход от создания различных инструментариев информационных технологий (программных и технических средств, локальных сетей и систем), к созданию и последующей реализации концепции комплексной информационной поддержки различных видов деятельности, имеющих общегосударственную значимость, к которым, безусловно, принадлежит архивное дело.

В архивных учреждениях России сосредоточен огромный информационный потенциал. Они организованы в сеть государственных, ведомственных и других архивов. Сеть государственных архивов Российской Федерации включает свыше 2000 архивных учреждений, в которых хранится более 200 тыс архивных фондов, содержащих свыше 180 млн. единиц хранения.

Архивные документы содержат уникальную информацию по различным аспектам общественно-политического, государственного, экономического, социального, культурного развития народов, населяющих Российскую Федерацию; в совокупности эти документы являются составной частью национального и культурного наследия страны, задача сохранения которого закреплена в законодательном порядке.

В настоящее время хранящиеся в архивах документы используются исследователями как в своем первичном виде, так и для создания вторичных, производных источников. При этом исходные данные интегрируются во вновь создаваемых каталогах, таблицах и так называемых тематических базах данных (ТБД). ТБД определяется как отобранная из архивных

источников и организованная определенным образом коллекция тематически связанных данных, созданная для решения определенной научной или прикладной проблемы и позволяющая в процессе ее решения рациональным способом извлекать, просматривать и редактировать необходимые данные. ТБД в настоящее время интенсивно создаются государственными, общественными и частными организациями, архивными структурами, а также различными научными группами и отдельными исследователями, что существенно повышает скорость и качество обслуживания пользователей.

Для снижения общей трудоемкости процесса концептуального проектирования ТБД необходимо использовать методы формализованного анализа информационных требований пользователей, для последующего формирования канонических структур ТБД и построения логических структур БД.

В этой связи, наряду с общетеоретическими, методологическими и прикладными проблемами проектирования баз данных (БД), возникает связанная с ними проблема обеспечения максимальной полноты удовлетворения информационных потребностей широкого круга пользователей при создании и эксплуатации тематических баз данных по материалам архивных источников. Ее решение базируется, прежде всего, на качественном проведении этапа концептуального проектирования ТБД, связанного с описанием и синтезом информационных требований в рамках некоторой достаточно общей и безызыточной структуры, объединяющей требования полного множества пользователей ТБД.

Большие масштабы работ по созданию и использованию ТБД в различных архивных организациях, и в то же время отсутствие единой методологии проектирования ТБД обуславливают высокую актуальность проведенных в работе научных исследований.

Цель работы и задачи исследования Целью работы является разработка и использование взаимосвязанных моделей и методов анализа требований пользователей архивных данных и построение безызыточных канонических структур тематических баз данных при автоматизации деятельности архивных учреждений.

Для достижения этой цели потребовалось:

- провести анализ современного состояния процессов информатизации архивного дела в РФ и за рубежом, определить основные направления информатизации деятельности архивных учреждений в РФ;

- провести анализ существующих методов и форм координационной, методической, информационно-справочной и публикационной деятельности государственных ар-

хивных учреждений, а также научно-исследовательской работы пользователей и обосновать необходимость и эффективность использования ТБД.

- разработать методы анализа и структуризации информационных требований пользователей архивных данных и построения конкретной канонической структуры ТБД;
- разработать модели и методы реструктуризации канонической модели ТБД;
- использовать предложенные модели и методы для построения канонической структуры ТБД.

Научная новизна состоит в том, что в результате проведенных исследований, анализа и обобщения опыта проектирования и эксплуатации систем автоматизации деятельности архивных учреждений РФ:

- разработаны общий подход и предложения по оценке стоимости информационных услуг для пользователей автоматизированных архивных систем государственных архивных учреждений;
- разработана единая методология проектирования оптимальных структур тематических баз данных, заключающаяся в последовательном решении задач анализа и структуризации предметной области пользователей и формирования безызбыточной канонической структуры, используемой для создания логических и физических структур ТБД;
- разработана совокупность графовых и матричных моделей и процедур их преобразования, обеспечивающая формальный анализ и структуризацию предметных областей пользователей, выявление и анализ дублируемых информационных элементов и избыточных взаимосвязей, выделение групп данных, ключей и зависимых атрибутов, а также построение рациональных канонических структур тематических баз данных на этапе, предшествующем их техническому проектированию;
- разработаны формализованные процедуры анализа и модификации информационных структур пользователей тематических баз данных при изменении требований на обработку запросов, а также формализованные процедуры анализа вновь поступающих требований пользователей и построения обобщенной внешней модели тематических баз данных.

Практическая ценность. Предложенные модели и методы позволяют создавать эффективные локальные и распределенные тематические базы архивных данных с целью их многоцелевого использования в архивных учреждениях. Использование указанных мето-

дов, алгоритмов и программ проектирования ТБД позволяет снизить трудоемкость и общие затраты на разработку, внедрение и модификацию соответствующих автоматизированных архивных систем в 2 - 3 раза.

С использованием предложенной методологии формирования и реструктуризации канонических структур ТБД разработана каноническая структура тематической базы архивных данных “Контингент студентов РГГУ”.

Разработанные методы, алгоритмы и программные средства могут быть использованы при создании автоматизированных архивных систем широкого спектра и назначения в архивных учреждениях, научно-исследовательских институтах, ВУЗах, организациях и вычислительных центрах, разрабатывающих, внедряющих и эксплуатирующих подобные системы.

Апробация результатов диссертационного исследования. Основные положения и результаты диссертации докладывались и обсуждались на II, III и IV Международных конференциях “Проблемы управления в чрезвычайных ситуациях”(Москва, 1994 г., 1995 г., 1997 г.), на российско-французской конференции “Управление высшим образованием” (Москва, 1994 г.), на международных конференциях “Проблемы управления безопасностью сложных систем” (Москва, 1998 г.) и “Управление в XX веке: итоги и перспективы” (Москва, 1998 г.).

Публикации. По теме диссертационной работы опубликовано 11 печатных работ.

Структура диссертационной работы.

Диссертация состоит из введения, четырех глав, заключения, списка цитируемой литературы и приложений.

Текст работы содержит 160 страниц машинописного текста, 38 рисунков и 32 таблицы.

Содержание работы.

Во введении обоснована актуальность темы, сформулированы цель исследования и задачи, рассматриваемые в работе. Представлена структура диссертационной работы, взаимосвязь и краткое содержание ее разделов.

В первой главе проводится анализ организации архивного дела в Российской Федерации, его состава, характеристик государственной и негосударственной систем Архивного Фонда РФ, а также представлены основные направления и формы всестороннего

использования информационного потенциала Архивного фонда Российской Федерации, результаты деятельности архивных учреждений Государственной архивной службы России.

Архивный фонд Российской Федерации - совокупность документов, отражающих материальную и духовную жизнь страны, имеющих историческое, научное, социальное, экономическое, политическое и культурное значение и являющихся неотъемлемой частью историко-культурного наследия народов Российской Федерации.

Специфическими особенностями функционирования архивной отрасли РФ являются: большой объем с постоянным накоплением массива архивных документов, поливидовой состав архивного фонда, территориальная распределенность архивных учреждений, широкие хронологические рамки архивных источников, многоаспектность тематики, исторически и логически сложившаяся внутренняя дифференциация на документы, многообразие форм и способов описания информации в архивных учреждениях, широкий круг потребителей, большое число нерегламентированных запросов, многоцелевое использование информации и др.

В настоящее время, когда реформированию подвергаются практически все стороны общественного устройства, многократно возрастает ценность имеющейся архивной информации, а также многократно повышается роль одной из основных функций государственной архивной службы - обеспечение потребностей общества в ретроспективной информации, заключенной в архивных документах.

В этой связи одной из важных проблем в деле совершенствования архивной службы РФ является повышение степени адекватности отражения всех наиболее существенных сторон исторического процесса в принимаемых на долговременное или постоянное хранение архивных материалов. Эта проблема может быть всесторонне и кардинально решена лишь путем разработки научно-методологических основ информатизации ведения архивного дела в условиях рыночной экономики и использования современных информационных технологий.

Основной целью информатизации архивного дела является комплексная автоматизация на базе высокoeffективной технологии информационной поддержки деятельности Архивной службы РФ по созданию, хранению, ведению и использованию архивного фонда РФ, направленная на повышение активности его информационных ресурсов, превращение архивных данных в информацию, доступную для автоматизированного поиска, хранения, обработ-

ки и многоаспектного использования в процессе обслуживания информационных запросов широкого круга пользователей.

Анализ требований Государственной целевой программы “Информатизация России” и “Концепции формирования и развития единого информационного пространства в России и соответствующих государственных информационных ресурсов” позволил выделить основные направления информатизации архивного дела, включающие:

- развитие фундаментальных, методологических и прикладных исследований и разработок в области технологий информационной поддержки основных видов деятельности архивных учреждений;
- развитие инфраструктуры информатизации архивного дела в РФ;
- разработка методов и средств маркетинга информационных продуктов (ИП) архивной отрасли в условиях рынка;
- развитие методов прогнозирования состава предметных областей и содержания запросов пользователей, представляющих центральные и региональные органы законодательной и исполнительной власти;
- развитие методов и средств типизации проектирования функциональных и обеспечивающих подсистем автоматизированных архивных систем различного назначения;
- комплексная автоматизация информационной поддержки комплектования архивных фондов, экспертизы и отбора хранимых в них документальных материалов;
- комплексная автоматизация процедур хранения архивной информации в локальных и региональных базах данных;
- комплексная автоматизация информационной поддержки учета и обеспечения сохранности документальных материалов в Архивном фонде РФ;
- комплексная автоматизация процедур создания, представления и использования научно-справочного аппарата в Архивном фонде РФ;
- комплексная автоматизация процедур предоставления архивной информации пользователям различного уровня (государственным учреждениям, руководству и специалистам Росархива, исследовательским учреждениям и другим юридическим и физическим лицам);
- комплексная автоматизация подготовки к изданию архивных справочников и отдельных документальных материалов;

- комплексная автоматизация информационной поддержки перспективного и текущего планирования, учета деятельности и формирования отчетности архивных учреждений;

- комплексная автоматизация информационной поддержки научно-исследовательской и методической работы архивных учреждений;

- комплексная автоматизация информационной поддержки, основанной на материалах архивных источников;

- комплексная автоматизация информационной поддержки административно-управленческой деятельности органов управления архивным делом и архивных учреждений

Выделенные направления комплексной информатизации архивного дела и перспективные информационные технологии автоматизированной поддержки деятельности архивных учреждений базируются на новейших достижениях в области разработки диалоговых систем “человек - ЭВМ”, средств создания и ведения распределенных баз данных, содержащих архивную информацию, способов и методов ее многоаспектного использования заинтересованными учреждениями, организациями и исследовательскими группами.

Во второй главе изложены результаты проведенного исследования существующих методов и форм координационной, методической, информационно-справочной и других видов деятельности государственных архивных учреждений, рассмотрены типовые нормативы времени, предусмотренные для выполнения определенных видов работ. Показана зависимость научно-исследовательской работы пользователей от состояния архивов и, прежде всего, их научно-справочного аппарата. Определены основные направления разработки типовых программных комплексов информационной поддержки деятельности государственных архивных учреждений. Приведены разработанные предложения по оценке информационных услуг архивных учреждений в условиях рынка

Важнейшим направлением реализации информационного потенциала отечественных архивов является использование хранящихся в них документальных материалов в научно-исследовательской работе пользователей для решения как фундаментальных теоретических, так и конкретных прикладных задач. Особенно большое значение использование документов архивов имеет для исторической науки, базирующей свои выводы на основе ретроспективной информации. Архивы хранят зафиксированную на материальных носителях информацию о прошлом человеческого общества, сохраняют документальные

свидетельства о наиболее существенных проявлениях его жизнедеятельности. Поэтому очевидно, что значительную часть пользователей информации архивных документов традиционно составляют историки, накопившие к настоящему времени наибольший опыт работы в архивах. Но к документам архивов обращаются также представители многих других отраслей знаний: филологи, искусствоведы, медики, геологи и многие другие.

Статистические исследования показали, что в период 1994-1997 г.г. в среднем ежегодно пользователями документальных материалов, хранящихся в российских архивах, были 45,7 тыс. исследователей, в том числе 2 тыс. из ближнего и дальнего зарубежья. Им были выданы в пользование 1,2 млн. дел, что составляет около 1% от общего количества дел, находящихся в отечественных архивах (без учета возможности повторного использования одних и тех же материалов).

Достаточно низкая производительность выполнения работ, связанных с информационно-справочной и публикационной деятельностью государственных архивов объясняется.

1) большим объемом рутинных операций поиска информации вследствие ориентированности научно-справочного аппарата, в первую очередь, на учет сохранности и местонахождения документов и, лишь во вторую очередь, на раскрытие их содержания,

2) отсутствием единой технологии и комплекса средств информационной поддержки проведения работ данного вида.

Эти же обстоятельства объясняют большие плановые сроки (3-5 лет) выполнения научных исследований на базе информации, зафиксированной в хранящихся в архивах документальных материалах, и ограниченный объем использования последних (в целом лишь несколько процентов от общего объема Архивного фонда РФ). В то же время и зарубежный и отечественный опыт свидетельствует о том, что задача повышения эффективности работы пользователей вполне разрешима.

Переход к автоматизированной обработке информации делает возможным введение в научный оборот значительно большего объема документации Архивного фонда РФ. Это достигается за счет переноса сведений многочисленных первичных документов, хранящихся в архивах, в тематические базы данных, что дает возможность решать научно-познавательные задачи, перед которыми вынуждены были отступать исследователи, оперировавшие исключительно описательными методами анализа.

Эффективное проектирование тематических баз данных требует от исследователей и разработчиков профессиональной источниковедческой работы по оценке информации

онного потенциала источника, возможностей сохранения этого потенциала в максимально адекватной форме и последующего его многоаспектного использования при проведении научных и прикладных исследований. Создание тематических баз данных позволяет всему научному сообществу России пользоваться более широким кругом источников, причем не только исходных, но и вторичных, структурированных, созданных на основе архивных данных.

Проектирование локальных и распределенных тематических баз архивных данных (ТБД, РБАД) сопряжено с большими трудностями, связанными с отсутствием единых формальных методов их проектирования, многообразием характеристик решаемых задач, необходимостью учета различного рода требований и ограничений, накладываемых на разрабатываемую структуру БД программными средствами систем управления базами данных (СУБД) и операционных систем, трудностью формализации и оценки качества проектных решений.

Основными пользователями автоматизированных архивных систем (ААС) государственных архивных учреждений являются:

- правительственные учреждения и организации;
- государственные учреждения, организации и предприятия;
- государственные музеи и выставки;
- библиотеки и читальные залы;
- самостоятельные научно-исследовательские организации, отдельные исследовательские группы, общественные организации и движения;
- коммерческие фирмы и отдельные граждане РФ.

Цели, содержание и способы информационного обслуживания указанных групп пользователей существенно различны.

В этой связи представляется нецелесообразным использование единого технологического, стоимостного или временного критерия оценки эффективности создания и функционирования разрабатываемой автоматизированной системы государственных архивных учреждений (ГАУ).

Оценку эффективности информационного обслуживания правительственных организаций, государственных учреждений и предприятий целесообразно проводить с использованием критерия минимума суммарного времени обработки поступающих запросов к информационной базе и выдачи полных и точных результатов в удобной для дальнейшего использования форме.

Оценку эффективности информационного обслуживания государственных музеев и выставок, библиотек и читальных залов целесообразно проводить по критерию максимума покрытия их информационных потребностей документальными материалами, хранящимися в базе данных системы.

Оценку эффективности информационного обслуживания самостоятельных научно-исследовательских организаций, отдельных исследовательских групп, общественных организаций и движений, коммерческих фирм и отдельных граждан целесообразно проводить с использованием стоимостного критерия, максимизирующего суммарный доход, получаемый архивными учреждениями при предоставлении платных информационных услуг выделенным категориям пользователей.

В силу специфики автоматизированных архивных систем основной экономический эффект достигается не столько в самой системе, сколько у ее пользователей. При этом экономический эффект от функционирования ААС у таких пользователей превышает последний в самой системе (в сравнении с рядом действующих в настоящее время информационных систем) на один-два порядка.

В условиях рынка косвенный экономический эффект функционирования ААС ГАУ должен привлекать достаточное число клиентов, увеличивая тем самым объем запрашиваемых и оплачиваемых услуг и, соответственно, общий доход системы при практически неизменных капиталовложениях и незначительном увеличении эксплуатационных расходов.

В третьей главе представлена разработанная совокупность взаимосвязанных графовых и матричных моделей и формализованных процедур, обеспечивающих проектирование, формализованное описание, анализ, структуризацию и нормализацию внешних моделей предметных областей пользователей распределенных баз данных архивных данных. Показана целесообразность использования тематических баз данных. Создание ТБД по материалам архивных источников является весьма мощным средством активизации национальных информационных ресурсов за счет перевода значительной их части на машиночитаемые носители и последующего многоаспектного индивидуального и коллективного использования в процессе исследований.

Показано, что наряду с теоретическими, методологическими и прикладными проблемами проектирования баз данных (БД), возникает связанная с ними проблема обеспечения максимальной полноты удовлетворения информационных потребностей широкого

круга исследователей при создании и эксплуатации тематических баз данных по материалам архивных источников. Ее решение базируется, прежде всего, на качественном проведении этапа концептуального проектирования ТБД, связанного с описанием и интеграцией информационных требований пользователей в рамках некоторой достаточно общей и избыточной структуры, объединяющей требования полного множества пользователей ТБД.

Таким образом, начальным этапом проектирования ТБД является этап предпроектного анализа информационных требований и структуризации предметных областей пользователей. На данном этапе осуществляется изучение и анализ требований к разрабатываемой структуре ТБД, определяется состав информационных элементов и групп определенной предметной области, информационной моделью которой будет служить создаваемая ТБД.

Требования отдельных пользователей-исследователей объединяются далее в единое глобальное представление в рамках канонической структуры ТБД. При этом выявляются и исключаются противоречивые и избыточные данные (информационные элементы и их группы) и связи между ними.

Процесс анализа информационных требований разбивается на три взаимосвязанных этапа:

1 Анализ информационных структур, представляющих требования пользователей, и формирование графов информационных структур.

2. Выделение ключей и атрибутов в группах элементов данных

3 Построение канонической структуры БД, являющейся результатом операции объединения графов информационных структур в один общий граф структуры данных.

Пусть $S = \{S_k | k = \overline{1, k_n}\}$ - множество информационных структур пользователей ТБД. $D_k = \{d_i | i \in L_k \subseteq L\}$ - множество структурных элементов, входящих в состав k -й структуры. L - множество структурных элементов конкретной предметной области. Исходной информацией для процедур данного этапа является информация о парных отношениях между структурными элементами предметной области, задаваемая в информационных требованиях пользователей и формализуемая в виде матриц семантической смежности $B_k = \|b_{ij}^k\|, k = \overline{1, k_n}$.

Если между структурными элементами d_i и d_j существует отношение R такое, что элемент d_i составляет (расширяет, дополняет и т.д.) смысловое содержание элемента d_j ,

то элемент b_{ii}^k матрицы B_k равен единице, в противном случае - $b_{ii}^k = 0$. Матрицы B_k ставятся в соответствие орграф $G_k = (D_k, U_k)$, множеством вершин которого являются структурные элементы множества D_k , а множеством дуг - отношения смежности между ними. С использованием матрицы B_k формируется матрица семантической достижимости $A_k = \left\| \|d_{ij}^k\| \right\|$. Матрицы B_k и A_k связаны булевым уравнением

$$B_k^{\alpha-1} \neq B_k^\alpha \approx B_k^{\alpha+1} = A_k$$

где α - целое положительное число, $\alpha \leq L_k - 1$. Матрица A_k дает возможность определения множества предшествования $C(d_i)$ и достижимости $F(d_i)$ по всем элементам $d_i \in D_k$. Анализ множеств $C(d_i)$ позволяет выделить базовые типы структур, из которых конструируется данная информационная структура - элементы данных и группы. Элементам данных соответствуют те элементы $d_i \in D_k$, для которых $C(d_i) = \emptyset$, и на графе G_k им соответствуют висячие вершины. Множество элементов данных обозначим через D_k^d . Тогда множество групп D_k^g определится из выражения $D_k^g = D_k \setminus D_k^d$. Элементы-группы будем обозначать через d_i^g . Из матрицы A_k выделим подматрицу $A_k^g = \left\| \|d_{ij}^{kg}\| \right\|$, которая индексировается элементами $d_i^g \in D_k^g$, и вдоль главной диагонали A_k^g введем записи $a_{ii}^{kg} = 1$. На основании подматрицы A_k^g проводится упорядочение групп по уровням иерархии $P_m (m = \overline{1, M_n})$, начиная с наивысшего ($m = 1$) уровня P_1 , на основании рекуррентного соотношения

$$P_m = \left\{ d_i^g \in D_k^g \setminus P_1 \setminus \dots \setminus P_{m-1} / F_{m-1}(d_i^g) \cap C_{m-1}(d_i^g) \right\}$$

где $F_{m-1}(d_i^g)$ и $C_{m-1}(d_i^g)$, соответственно, множества достижимости и предшествования групп $d_i^g \in D_k^g$ (с учетом $a_{ii}^{kg} = 1$) на подмножестве $D_k^g \setminus P_1 \setminus \dots \setminus P_{m-1}$. Упорядочение позволяет выделить группы, являющиеся возможными точками входа в структуру (группы уровня P_1), а также промежуточные (P_2, \dots, P_{M_n-1}) и конечные (P_{M_n}) группы, лежащие на тех или иных путях доступа к информационным элементам структуры. Состав элементов данных $d_i \in D_k^d$, образующих некоторую группу d_i^g , обозначаемый $H(d_i^g)$, определится из выражения:

$$H(d_i^r) = \{d_i / b_i^k = 1, \forall d_i \in D_i^d\}.$$

Процедуры выделения ключей и атрибутов в группах элементов данных применяются для каждой группы информационной структуры S_k .

При анализе групп информационной структуры различается два типа ключей: основной ключ группы и вспомогательный.

Основным ключом группы d_i^r является элемент данных d_i^d , входящий в состав $H(d_i^r)$ данной группы $d_i^d \in H(d_i^r)$, который полностью и однозначно идентифицирует остальные элементы данной группы, т.е. для него выполняются следующие условия

$$d_i^l \neq d_i^2 \neq \dots \neq d_i^m \neq d_i^{m+1},$$

где m_n - максимальная мощность множеств областей определения информационных элементов группы.

Вспомогательным ключом (или ключами) группы является элемент данных, который однозначно идентифицирует некоторое подмножество элементов данных (атрибутов) данной группы. Для него выполняется приведенное выше условие по отношению к данному подмножеству атрибутов.

Исходной информацией, требуемой для построения канонической структуры БД, являются матрицы V_k^* и орграфы G_k^* , $k = \overline{1, k_n}$. На данном этапе решаются следующие задачи: формирование безыбыточных множеств ключей W_1^n , атрибутов W_2^n и полного безыбыточного множества информационных элементов W_n ; построение обобщенной матрицы смежности $B_n = \|b_{ij}^n\|$ и обобщенного графа структуры данных G_n , выявление и анализ пересекающихся атрибутов и избыточных взаимосвязей, выявление и анализ циклических участков структуры данных; упорядочение графа G_n по уровням иерархии, что обеспечивает построение матрицы смежности B и орграфа G канонической структуры БД (93-99).

Множество W_1^n определяется из выражения

$$W_1^n = W_1^1 \cup W_1^2 \setminus W_1^{1,2} \cup W_1^3 \setminus (W_1^{1,3} \cup W_1^{2,3}) \cup \dots \cup W_1^k \setminus \bigcup_{i=1}^{k-1} W_1^{i,k}$$

где $W_1^{i,k}$ - подмножество пересечения множеств W_1^i и W_1^k . Таким образом, W_1^n представляет собой объединение множества W_1^1 со всеми остальными множествами

$W_1^k (k = \overline{2, k_n})$, взятых без подмножеств пересечения множеств W_1^j и W_1^k и всех $(k-2)$ предшествующих W_1^k множеств.

Формирование множества W_2^0 осуществляется аналогично формированию W_1^0 и проводится на множествах W_2^k :

В случае, когда $W_1^0 \cap W_2^0 = W_{12}^0 \neq \emptyset$, элементы $d_i \in W_{12}^0$ исключаются из состава групп информационных структур, в которые они входили в качестве атрибутов, и устанавливается взаимосвязь между ключом данной группы и ключом $d_i \in W_1^0$

Определение $W_n = W_1^0 \cup W_2^0$ дает возможность построения обобщенной матрицы смежности B_n .

Сформированная каноническая структура предметной области пользователей ТБД должна отражать наиболее характерные и устойчивые свойства данных и их взаимосвязей, быть полной и независимой от программных и аппаратных средств ведения и обслуживания БД и использоваться на последующих этапах ее логического и физического проектирования.

Аналогично процедурам формирования канонической структуры локальных ТБД анализ предметных областей распределенных баз архивных данных (РБАД) включает четыре взаимосвязанных этапа.

1. Предпроектный анализ предметных областей пользователей РБАД
2. Анализ предметных областей и построение внешних моделей пользователей
3. Построение обобщенной внешней модели РБАД
4. Нормализация обобщенной внешней модели и построение канонической структуры РБ АД

Этап предпроектного анализа позволяет на ранних стадиях разработки оценить целесообразность проектирования РБАД на основе анализа общности предметных областей рассматриваемого множества пользователей. В соответствии со степенью общности предметных областей пользователей проводится их классификация и определяются множества пользователей, для которых целесообразно проектирование РБАД. Целесообразность определяется заданной степенью общности между предметными областями территориально распределенных пользователей. Анализ общности предметных областей основан на оценке их функций подобия.

Предложенные методы анализа информационных требований достаточно широкого множества потенциальных пользователей ТБД, построения их рациональных структур

позволяют существенно сократить общее время их разработки, повысить информационный потенциал и эксплуатационные характеристики информационно-поисковых систем, созданных на основе тематических баз данных.

В четвертой главе исследованы задачи актуализации и реорганизации ТБД, т.е. эффективного ведения ТБД в условиях изменения информационных требований пользователей. При расширении знаний о предметной области пользователей ТБД, вовлечении в процесс автоматизированной обработки новых пользователей и задач могут потребоваться существенные изменения первоначального проекта ТБД, которые могут распространяться как на состав, так и на структуру ТБД. Эти изменения вызываются добавлением новых или удалением существующих информационных элементов, групп данных, ключей и атрибутов данных, взаимосвязей и отношений между группами и логическими записями, изменением программного обеспечения ТБД (СУБД и операционных систем), требований к обработке данных и т.д.

Обеспечение требуемой эффективности работы с изменяющейся во времени информационной моделью предметной области пользователей требует периодической реорганизации ТБД.

Реорганизация БД является длительным и трудоемким процессом, в течение которого вся база или часть ее (в зависимости от выбранной стратегии) может быть выведена из режима нормального функционирования. Основным критерием эффективности реорганизации является минимизация стоимости либо времени реорганизации ТБД.

В настоящее время исследованию проблемы реорганизации БД уделяется большое внимание как у нас в стране, так и за рубежом. Большинство работ посвящено разработке методов и процедур реорганизации ТБД на физическом уровне: перераспределению данных по устройствам памяти, уплотнению и сжатию данных, выбору эффективных стратегий реорганизации и методов доступа и др. Значительно меньшее количество работ связано с исследованием проблем реорганизации на логическом и особенно на концептуальном уровне.

В то же время изменения предметной области, программного и технического обеспечения ТБД обычно требуют реорганизации различных уровней абстракции представления данных.

Под реорганизацией ТБД понимается процесс динамического изменения состава и (или) канонической, логической и физической структуры ТБД с целью повышения экс-

платационных характеристик банков данных в связи с изменением информационных требований пользователей. Изменение канонической и (или) логической структуры называется реструктуризацией, а изменение физической структуры - реформатизацией ТБД.

Разработанный комплекс процедур реструктуризации ориентирован на режим диалога пользователей и разработчиков с ЭВМ, что позволяет рассматривать его как основу автоматизации процесса реорганизации ТБД на концептуальном уровне.

Выделим следующие основные типы изменений предметной области.

- изменения информационных требований пользователей, заключающиеся в добавлении и (или) удалении информационных элементов предметной области;
- поступление новых информационных требований пользователей;
- удаление существующих информационных требований пользователей.

В соответствии с данной классификацией основных типов изменений предметной области процесс их анализа разбивается на три этапа.

1. Анализ и модификация информационных структур пользователей при изменении требований на обработку.
2. Анализ вновь поступающих требований пользователей и построение обобщенной внешней модели ТБД
3. Анализ и реструктуризация канонической структуры ТБД при удалении информационных требований отдельных пользователей.

Разработанный комплекс процедур реструктуризации ориентирован на режим диалога пользователей и разработчиков с ЭВМ, что позволяет рассматривать его как основу автоматизации процесса реорганизации ТБД на концептуальном уровне

В работе представлена разработанная каноническая структура тематической базы архивных данных "Контингент студентов РГГУ" с использованием предложенной методологии формирования и реструктуризации канонических структур моделей данных.

В приложении к диссертации приведены материалы, подтверждающие практическое использование и внедрение полученных автором научно-методических результатов.

Выводы и заключение. В диссертационной работе получены следующие основные результаты:

1. Проведен анализ организации архивного дела в Российской Федерации, его основных особенностей, состава и характеристик государственной и негосударственной

частей Архивного Фонда РФ. Представлены основные направления и формы всестороннего использования информационного потенциала Архивного Фонда Российской Федерации, а также некоторые результаты деятельности архивных учреждений Государственной архивной службы России. Проведен обзор современного состояния процессов информатизации архивного дела в Российской Федерации и за рубежом. На основе результатов анализа определены цели, основные задачи и показатели эффективности процессов информатизации деятельности архивных учреждений РФ.

2. На основе изучения требований Государственной целевой программы "Информатизация России" и "Концепции формирования и развития единого информационного пространства России и соответствующих государственных информационных ресурсов" проанализированы основные направления и задачи комплексной информатизации архивного дела в Российской Федерации. Предложен и использован ряд перспективных принципов построения интегрированных информационно-управляющих систем и технологий обработки данных, в том числе формирования структур тематических баз данных.

3. Проведен детальный анализ существующих методов и форм деятельности государственных архивных учреждений, а также научно-исследовательских работ пользователей архивов. Показано, что наличие единой нормативно-методической базы обеспечивает необходимую взаимосвязь и преемственность в решении общих задач работы с документами в делопроизводстве в ведомственных и государственных архивах, которые проявляются в применении единых методик классификации, учета, экспертизы ценности, составления научно-справочного аппарата к архивным документам планирования комплектования государственных архивов документами учреждений и др.

4. Разработаны общий подход и предложения по оценке стоимости информационных услуг для пользователя автоматизированных архивных систем.

5. Разработана единая методология проектирования рациональных структур ТБД, заключающаяся в последовательном решении задач анализа и структуризации предметной области пользователей и формирования без избыточной канонической структуры ТБД, используемой для формирования логических и физических структур.

6. Разработана совокупность графовых и матричных моделей и процедур их преобразования, обеспечивающая формальный анализ и структуризацию предметной области пользователей, выявление и анализ дублируемых информационных элементов и избыточных взаимосвязей, выделение групп данных, ключей и зависимых атрибутов, а также по-

строение рациональных канонических структур ТБД на этапе, предшествующем их техническому проектированию. Разработаны взаимосвязанные графовые и матричные модели и формализованные процедуры, обеспечивающие проектирование, формализованное описание, анализ, структуризацию и нормализацию внешних моделей предметных областей пользователей РБАД

7. Предложена методология нормализации информационных структур предметных областей пользователей, описываемых в виде иерархических и сетевых моделей данных, отличающаяся от известных методов универсальностью, экономичностью, простотой алгоритмизации и автоматизации. Формализованные процедуры нормализации обобщенной внешней модели РБАД обеспечивают выделение базовых структурных элементов, анализ дублируемых информационных элементов и транзитивных взаимосвязей, структуризацию информационных элементов по уровням иерархий, выделение ключевых атрибутов, то есть проектирование канонической структуры РБАД

8. Разработан комплекс формализованных моделей и методов анализа и реструктуризации канонической модели архивных данных в процессе учета поступающих изменений. Разработаны формализованные процедуры анализа и модификации информационных структур пользователей ТБД при изменении требований на обработку запросов

9. Разработана каноническая структура тематической базы архивных данных "Контингент студентов РГГУ" с использованием предложенной методологии формирования и реструктуризации канонических структур моделей данных, что позволило сократить общие затраты на разработку ТБД на 40% по сравнению с традиционными способами проектирования, а эксплуатационные расходы снизить на 25%

Основные публикации по теме диссертации.

1. Чанхнева Ф.Ю. К оценке стоимости предоставления справочных услуг архивными учреждениями // Сборник докладов конференции "Управление в XX веке: итоги и перспективы" М., 1998. С. 35.

2. Архипова Н.И., Чанхнева Ф.Ю. Организация и направления развития архивного дела в современных условиях // Пятая международная конференция "Управление безопасностью сложных систем": Тезисы докладов. М., 1998. С. 102-104

3. Архипова Н.И., Кульба В.В., Косяченко С.А., Чанхнева Ф.Ю., Щелков А.Б. Введение в системы организационного управления. М. Приор. 1998 - 446 с.

4. Архипова Н.И., Голиков А.Г., Росс Г.В., Чанхиева Ф.Ю., Шикло А.Е. Отчет по проекту 64.3. Исследование и анализ методов и средств оценки эффективности информационной поддержки координационной, методической, информационно-справочной, публикационной деятельности государственных архивных учреждений, а также научно-исследовательской работы пользователей. М., МГУ., 1994. - 65 с.

5. Архипова Н.И., Бородкин Л.И., Власов А.В., Голиков А.Г., Чанхиева Ф.Ю., Чернов И.В., Швецов А.Р., Шелков А.Б. Разработка методов и средств автоматизации документооборота и информационной поддержки деятельности государственных архивных учреждений, а также научно-исследовательской работы пользователей. Отчет по проекту 324.3. М., МГУ., 1996. - 39 с.

6. Косяченко С.А., Чанхиева Ф.Ю. Модели и методы реструктуризации канонической структуры распределенных баз архивных данных // Сборник докладов конференции "Управление в XX веке: итоги и перспективы". М., 1998. С. 64.

7. Ужастов И.А., Чанхиева Ф.Ю. Модели и методы анализа, структуризации и реструктуризации информационных требований пользователей и построения канонической структуры тематических баз данных // Пятая международная конференция "Управление безопасностью сложных систем": Тезисы докладов. М., 1998. С. 107-110.

8. Островский В.А., Чанхиева Ф.Ю., Швецов А.Б. Особенности разработки системы регистрации архивной информации в АСУ ЧС // Четвертая международная конференция "Проблемы управления в чрезвычайных ситуациях": Тезисы докладов. М., 1997. С. 171-172.

9. Архипова Н.И., Косяченко С.А., Чанхиева Ф.Ю. Определение функций и концептуальной структуры информационной базы для системы управления риском на региональном уровне // Третья международная конференция "Проблемы управления в чрезвычайных ситуациях": Тезисы докладов. М., 1995. С. 72-73

10. Архипова Н.И., Курочкин Е.П., Чанхиева Ф.Ю. Компьютерное моделирование в университетском образовании // Российско-французская конференция "Управление высшим образованием": Тезисы докладов. М., 1994. С. 56-57.

11. Архипова Н.И., Росс Г.В., Чанхиева Ф.Ю. Методика исследования организации труда программистов и проектировщиков // Вторая международная конференция "Проблемы управления в чрезвычайных ситуациях": Тезисы докладов. М., 1994. С. 75-76.