

На правах рукописи

ЛЕВКОВ Александр Александрович

**МЕТОДОЛОГИЯ ПОСТРОЕНИЯ СТРУКТУРЫ
СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ
НА ОСНОВЕ РАСШИРЕННОЙ РЕЛЯЦИОННОЙ
МОДЕЛИ ДАННЫХ И АЛГОРИТМОВ
(на примере информационной поддержки
деятельности предприятия)**

**Специальности 05.13.01 – Системный анализ, управление
и обработка информации (в промышленности)**

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

Уфа – 2013

Работа выполнена в ФГБОУ ВПО
«Уфимский государственный авиационный технический университет»
на кафедре технической кибернетики

Научный консультант	заслуж. деятель науки и техники РФ д-р техн. наук, проф. ИЛЬЯСОВ Барый Галеевич
Официальные оппоненты	д-р техн. наук, МИНАКОВ Игорь Александрович Федеральное государственное бюджетное учреждение науки «Институт проблем управ- ления сложными системами РАН», ст. науч. сотр.
	д-р техн. наук, проф. ПАВЛОВ Сергей Владимирович ФГБОУ ВПО «Уфимский государственный авиационный технический университет», зав. кафедрой геоинформационных систем
	д-р техн. наук, проф. ФЕДУНОВ Борис Евгеньевич Федеральное государственное унитарное предприятие «Государственный научно- исследовательский институт авиационных систем», нач. лаборатории
Ведущая организация	ФГБУН «Санкт-Петербургский институт информатики и автоматизации РАН»

Защита диссертации состоится 18.10.2013
на заседании диссертационного совета Д-212.288.03
при Уфимском государственном авиационном техническом университете
по адресу: 450000, Республика Башкортостан, г. Уфа, ул. К. Маркса, 12

С диссертацией можно ознакомиться в библиотеке университета

Автореферат разослан

2013 г.

Ученый секретарь
диссертационного совета
д-р техн. наук, проф.



В. В. Миронов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Развитие промышленности и повышение эффективности промышленных предприятий являются одной из наиболее важных задач современности. Выполнение данных задач является базисом развития государства в целом и роста благосостояния его граждан. Повышение эффективности промышленных предприятий и общее развитие промышленности базируется на внедрении инновационных методов и способов производства, но не может быть достигнуто без эффективной информационной поддержки деятельности предприятия.

Системы информационной поддержки деятельности предприятия реализуют выполнение широко круга бизнес-функций предприятия, таких как: управление складом, управление запасами и заказами, управление персоналом, управление продажами, управление отношениями с клиентами, управление документооборотом и т. д. В основе современных систем информационной поддержки деятельности предприятия лежат системы обработки информации, от структуры и способов организации которых напрямую зависит эффективность и надежность выполнения бизнес-задач.

Степень разработанности темы исследования. Решению вопросов логической структуризации данных в системах обработки информации посвящены исследования отечественных ученых С. Кузнецова, М. Когаловского, А. Голосова, А. Сахарова, С. Павлова, И. Галахова, И. Блудова, Е. Григорьева, В. Кириллова, О. Христодуло, а также таких известных зарубежных ученых, как С. Date, E. Codd, R. Boyce, I. Heath, E. McCreight, A. Maule, M. Fotache, J. Martin, H. Darwen и других. Теоретическими результатами данных исследований являются формальные логические модели данных, а прикладными – системы управления базами данных, позволяющие осуществлять высокоуровневое управление информацией, независимо от формы и способа её хранения.

Область эффективного управления алгоритмами обработки информации глубоко проработана в трудах таких ученых, как Д. фон Неймана, Г. Буча, Е. Григорьева, Б. Страуструпа, P. Wegner, A. Simonet, M. Simonet, M. Roger, U. Keller, А. Степанова и других. Основными результатами данных работ явились объектно-ориентированная (и субъектно-ориентированная), аспектно-ориентированная и обобщенная (А. Степанов) парадигмы обработки информации.

В области формального описания знаний и метаданных работали такие зарубежные ученые, как Б. Рассел, M. Schmidt, G. Smolka, M. Milici, D. Verzati, F. Vaader и др., и такие отечественные ученые как М. Черняховская, А. Костюченко, Б. Майер, А. Бездушный, работы которых составили фундамент онтологического описания информации в виде дескриптивных

логик и широко находят применения в построение Symantec Web, прикладных онтологий и т. д.

Несмотря на актуальность данных работ, посвященных различным аспектам проектирования и построения структур систем обработки информации, существует ряд принципиальных проблем, неразрешённых на концептуальном уровне, наличие которых затрудняет построение эффективных, согласованных, целостных и выразительных систем.

Современные системы обработки информации, сталкиваясь с возрастающей интенсивностью потоков данных, увеличением их количества, возрастающей связностью информации, усложнением методов обработки данных и расширением областей своего применения в целом, требуют использования новых, комплексных средств организации структуры данных и средств их обработки.

Современные системы обработки информации уже не могут быть описаны на основе нестрогих и фрагментарных моделей, независимо друг от друга определяющих различные компоненты системы (факты, ограничения целостности, алгоритмы и знания). Для построения эффективных систем информационной поддержки деятельности предприятия необходима новая методология построения структуры системы обработки информации, позволяющая формально описать данные и алгоритмы их обработки в рамках единого неразрывного пространства абстракций.

Целью диссертационной работы является разработка методологии, теоретических основ и прикладных методов построения структуры системы обработки информации на основе расширенной реляционной модели данных и алгоритмов, применение их в составе систем информационной поддержки деятельности предприятия и оценка их эффективности.

Основные задачи, которые потребовалось решить для достижения цели:

1. Разработать информационную модель структуры системы обработки информации на основе системной интеграции данных и алгоритмов.
2. Разработать методы и принципы структуризации реляционных данных, направленные на построение иерархий реляционных отношений и повышение согласованности информации и разработать способы их реализации в рамках существующих систем управления базами данных.
3. Разработать методы описания и структуризации алгоритмов обработки информации в рамках расширенной реляционной модели данных, основанные на иерархической декомпозиции и нормализации алгоритмов и разработать способы их реализации в рамках существующих систем управления базами данных.
4. Разработать методы отображения пространства определения онтологий (концепции открытого мира) в пространство определения баз данных

(концепцию закрытого мира) и способы прямой и обратной трансформации расширенных реляционных схем данных в онтологии дескриптивных логик.

5. Разработать прикладную систему обработки информации на основе расширенной реляционной модели данных и алгоритмов, применить её в составе системы информационной поддержки деятельности предприятия и оценить её эффективность.

Научная новизна диссертационной работы заключается в следующем:

1. Предложена информационная модель структуры системы обработки информации на основе триады «Данные»-«Знания»-«Действия», отличающаяся системной интеграцией элементов в рамках единого базиса описания информации и позволяющая формулировать задачи хранения данных, обработки информации и манипуляции знаниями в рамках единой модели данных.

2. Предложена расширенная реляционная модель данных, которая отличается:

- предложенной структурообразующей функцией семантической эквивалентности, позволяющей создавать целостные иерархии реляционных отношений;
- типизацией неопределенных значений, что позволяет выявлять и устранять скрытые аномалии в структурах реляционных схем данных;
- предложенными правилами совместимости реляционных отношений по теоретико-множественным операторам и новой интерпретацией области определения атрибутов реляционных отношений, что позволяет повысить логическую целостность данных.

3. Предложены методы описания и структуризации алгоритмов в рамках расширенной реляционной модели данных, отличающиеся учетом алгоритмов как элементов данной модели, их структуризацией на основе устранения повторяющихся инструкций и семантических эквивалентностей, что позволяет минимизировать количество алгоритмов в системе, повышая целостность и согласованность всей системы обработки информации в целом.

4. В рамках интеграции онтологий и расширенной реляционной модели данных предложены:

- метод отображения концепции открытого мира в концепцию закрытого мира, основанный на покрытии пространства определения открытого мира суперпозицией реляционных выражений; данный метод отличается использованием неопределенных значений и возможностью выбора совокупности реляционных выражений, имеющих минимальную пространственную сложность, что позволяет производить трансляцию выражений дескриптивных логик в выражения реляционной модели данных;
- вариант дескриптивной логики, соответствующий иерархически-реляционным структурам в рамках расширенной реляционной модели дан-

ных, отличающийся использованием оператора наследования концептов и типизацией ролей с помощью концептов, что позволяет транслировать расширенные реляционные схемы данных в терминологические аксиомы дескриптивной логики и обратно.

Практическую значимость работы составляют:

1. Методы универсальной и специальной структуризации реляционных отношений по критерию семантической эквивалентности, методы структуризации отношений на основе типизированных неопределенных значениях, что позволяет создавать таксономии реляционных отношений, повышает централизацию данных и увеличивает их согласованность.

2. Методы структуризации алгоритмов на основе инструкций и методов как особого вида атрибутов расширенной модели данных и алгоритмов, их нормализации и структуризации на основе отношения семантической эквивалентности, что позволяет повысить целостность алгоритмов обработки данных и согласованность данных и алгоритмов.

3. Методы трансформации выражений в рамках концепции открытого мира в выражения в рамках концепции закрытого мира и способы трансляции онтологий в расширенные реляционные схемы данных, что позволило реализовать расширенную реляционную модели данных в рамках существующих систем управления базами данных и обеспечить использование предложенных способов структуризации данных в прикладных системах обработки информации без необходимости разработки новых средств управления данными на уровне внутренней и внешней памяти.

4. Программная реализация системы обработки информации на основе расширенной реляционной модели данных и алгоритмов, использование которой позволяет снизить структурную сложность системы обработки информации и повысить целостность данных и алгоритмов, а также их согласованность.

В основу диссертации положены научные результаты, полученные лично автором при непосредственной разработке прикладных информационных систем и технологий для предприятий России:

- система информационной поддержки деятельности промышленного предприятия ОАО «НПП «Мотор»;
- система информационной поддержки деятельности промышленного предприятия ОАО «УЗМ «Магнетрон»;
- система информационной поддержки деятельности фармацевтического торгово-дистрибьюторского предприятия ООО «Илья» (ЗАО «36,6»);
- система информационной поддержки деятельности фармацевтического торгово-дистрибьюторского предприятия ООО «Башмедснаб» (ЗАО «36,6»).

Основные теоретические положения диссертации использованы в учебном процессе УГАТУ: на кафедре Информатики по дисциплине «Организация баз данных» для студентов специальности 230301 «Моделирование и исследование операций в организационно-технических системах».

Эффективность реализованных методов, алгоритмов и программного обеспечения подтверждена 4 актами о внедрении.

Методология и методы исследования.

Объектом исследования являются системы обработки информации, их концептуальные и логические модели, операции над ними, правила композиции их структур. Предметом исследования являются методология и методы построения эффективных систем обработки информации. Для решения поставленных задач в работе использовались методы онтологического и объектно-ориентированного проектирования и программирования, аппарата реляционной алгебры, теории нормализации, теории множеств и дескриптивные логики.

Положения, выносимые на защиту:

1. Информационная модель структуры системы обработки информации основанная на системной интеграции данных и алгоритмов (*соответствует п.2. паспорта специальности*).

2. Методы и принципы структуризации реляционных данных, основанные на построении иерархии реляционных отношений, типизированных неопределенных значениях и активных доменах (*соответствует пп. 4, 5 и 8 паспорта специальности*).

3. Методы описания и структуризации алгоритмов обработки информации, основанные на нормализации алгоритмов и их консолидации по семантической эквивалентности (*соответствует пп. 4, 5 и 8 паспорта специальности*).

4. Методы отображения концепции открытого мира в концепцию закрытого мира и способы прямой и обратной трансформации расширенных реляционных схем данных в онтологии, основанные на дескриптивной логике (*соответствует пп. 4, 5 и 8 паспорта специальности*).

5. Программная реализация системы обработки информации для поддержки деятельности предприятия (на примере системы обработки информации промышленного предприятия ОАО УЗМ «Магнетрон») на основе предложенных методов структуризации данных и алгоритмов их обработки, а также результаты исследования эффективности предложенных методов (*соответствует пп. 3, 4, 5, 8 и 12 паспорта специальности*).

Степень достоверность и апробация результатов.

Основные результаты научных разработок, выполненных автором по теме диссертации, представлены на международных и российских конференциях:

- Международная конференция CSIT-2002, Греция, Университет г. Патра, 18–24 сентября 2002 г.
- Седьмая международная конференции “Проблемы техники и технологий телекоммуникаций”, Самара, 2006.
- Международная научно-технической конференция «Информационно-математические технологии в экономике, технике и образовании», Екатеринбург, 2007.
- Международная научная конференция «Решетневские чтения», Красноярск, 2007.
- Международная научно-техническая конференция “Проблемы техники и технологии телекоммуникаций”, Уфа, 2007.
- Международная научно-технической конференция «Проблемы управления и моделирования в сложных системах», Самара, 2012;
- Международная научно-технической конференция «Современные сложные системы управления», С. Оскол, 2012.
- Конференция «Информационные технологии в управлении», Санкт-Петербург, 2012.
- Конференция «Информационные технологии и системы», Челябинск, 2013

Публикации. Основные результаты диссертационной работы отражены в 48 научных трудах, в том числе в 15 статьях в рецензируемых журналах из перечня ВАК, 1 монографии, 30 статьях и трудах конференций, 2 свидетельствах об официальной регистрации программы для ЭВМ, 16 работ без соавторов.

Структура и объем работы. Диссертация включает введение, 6 глав, заключение, библиографию из 146 наименований и 12 приложений. Диссертация изложена на 330 страницах текста, из них основное содержание работы составляет 264 страницы, в т. ч. 8 таблиц, 113 рисунков.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность проблемы построения даталогической и алгоритмической структуры системы обработки информации для информационной поддержки деятельности предприятия. Сформулирована цель диссертационной работы, перечислены задачи исследований, научная новизна, апробация, практическая ценность и реализация основных результатов работы.

В первой главе рассмотрены существующие подходы к построению структуры системы обработки информации в рамках систем поддержки деятельности предприятия, основанные на механистическом объединении разных информационных моделей, что приводит к внутренним противоречиям и

несогласованностям таких систем и невозможности использования существенной части возможностей каждой из моделей.

Определены основные черты системы обработки информации для управления предприятием:

- Большой объем обрабатываемых данных при интерактивном режиме работы, что требует использования только масштабируемых, ресурсно-эффективных решений.
- Сложная схема данных, состоящая из большого количества абстракций и связей между ними (10^3 – 10^6 элементов), что не позволяет эффективно использовать экспертный анализ и требует применения адекватных исходной системе формальных методов структуризации.
- Сложная семантика данных, требующая применения множества сложных алгоритмов обработки данных и организации формальной связи данных и метаданных, что определяет необходимость нового подхода к организации структуры системы обработки информации.
- Высокая динамика структуры системы, обусловленная влиянием как внешних факторов, так и внутренним развитием системы, и требующая одновременно обеспечения надежности и оперативности выполнения изменений.

Рассмотрены наиболее распространенные подходы к проектированию и построению структуры системы обработки информации для поддержки деятельности предприятия:

- проектирование на основе логической модели сущность-связь с трансляцией в физическую реляционную схему данных и хранимые процедуры обработки информации;
- объектно-ориентированное моделирование системы на основе реляционного хранилища данных и поэлементной обработки объектов.

Первый подход позволяет использовать формальные средства структуризации данных и получать высокую ресурсную эффективность прикладных реализаций, но не позволяет структурировать алгоритмы обработки данных, являющиеся существенной частью системы обработки информации, что приводит к необходимости экспертного контроля за целостностью и согласованностью алгоритмов. Кроме того, данный подход предлагает слишком примитивные средства структуризации реляционных отношений, неадекватные структурам реальной системы. В целом, многие задачи проектирования решаются за счет наращивания организационной сложности проекта, что ведет к существенному росту его стоимости и затрудняет развитие и поддержку системы.

Второй подход отличается высокой структурной выразительностью и общим системным взглядом на структуризацию данных и алгоритмов как единого целого. Существенными недостатками является отсутствие строго

формализованного базиса структуризации абстракций, что не позволяет в полной мере автоматизировать проектирование и сопровождение системы, и низкая ресурсная эффективность, обусловленная использованием баз данных только как хранилища объектов и поэлементной их обработкой в рамках парадигмы объектно-ориентированного программирования.

Также рассмотрены вопросы применимости многомерных и онтологических (дескриптивно-логических) моделей данных. Показано, что обе эти модели не предназначены для потоковой интерактивной обработки оперативной информации и не позволяют управлять алгоритмами обработки данных.

Таким образом, вышеперечисленные проблемы существенно затрудняют построение эффективных, надежных и масштабируемых систем обработки информации. Для преодоления данных проблем требуется разработка новых концепций и методов построения структур систем обработки информации, интегрировано отражающих вопросы организации данных и алгоритмов в рамках систем информационной поддержки деятельности предприятия.

Во второй главе предложена информационная модель структуры системы обработки информации системно интегрирующая следующие ключевые абстракции:

- Данные – первичная абстракция, являющаяся отражением внешних и внутренних процессов в рамках системы обработки информации. Сами данные могут быть определены как «информация, в любой форме ее проявления, представленная в пригодном для машинной обработке виде», где информация – «любые сведения». На физическом уровне, в современной информационной среде основанной на двоичной логике, данные являются совокупностью бинарных состояний памяти (как внутренней, так и внешней) вычислительной машины.

- Знания – особый вид данных (данные второго порядка, метаданные), которые определяют тип, структуру, способы и принципы организации данных первого порядка, без которых данные первого порядка не несут смысловой нагрузки. Знания могут быть встроены в систему, обрабатывающую и интерпретирующую данные, могут храниться совокупно с данными, но независимо от конкретных форм и способов своего представления образуют самостоятельную абстракцию: одни и те же данные могут быть по-разному сформулированы, зафиксированы, обработаны и интерпретированы, в зависимости от различных знаний.

- Действия – активная составляющая, выражаемая через алгоритмы данных (именно полная совокупность всех алгоритмов обработки данных реализует данную абстракцию) и реализующая функцию движения информации. Действия обеспечивают трансформацию данных – на основе первичных, сырых данных генерируются вторичные данные. Если первичные реализуют отражение реального мира в информационной системе, то на основе вторич-

ных данных происходит отражение моделируемых процессов на реальные системы (вторичные данные являются основой для управляющего воздействия).

Информационная модель структуры простой системы обработки информации, ориентированной на потоковую обработку транзакций в рамках жёстко определенной бизнес-логики функционирования представлена в виде триады элементов «данные» (D – data), «знания» (K – knowledge) и «действия» (A – activity) (рисунок 1).

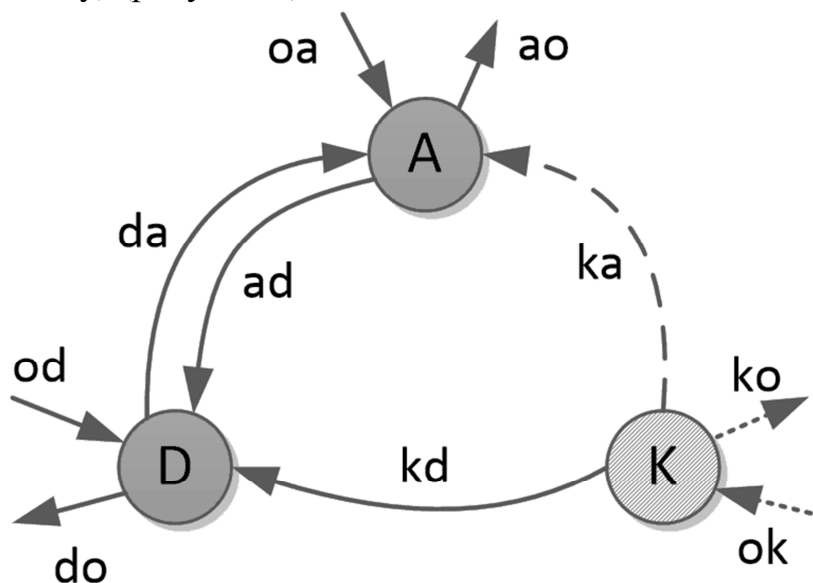


Рисунок 1 – Информационная модель структуры простой системы обработки информации

Связи в данной модели имеют следующую семантику:

- (od), (oa), (ok) – отражает поступление новых данных, знаний и действий извне;
- (do), (ko), (ao) – отражает доступность данных, действий и знаний внешнему наблюдателю;
- (da) – отражает выполнение определенных алгоритмов при поступлении новых данных;
- (ad) – отражает генерацию новых данных действиями (в результате работы алгоритмов обработки информации);
- (kd) – отражает структуризацию данных на основе соответствующих знаний;
- (ka) – отражает соответствие действий имеющимся знаниям;

Таким образом, информационная модель структуры простой системы обработки информации является датацентричной, т. е. ориентированной только на систематизацию, хранение и изменение данных, но предлагает минимальный функционал по управлению алгоритмами их обработки и управлению знаниями. Знания в данной модели не имеют полного отражения (что

показано светлой заливкой элемента) и присутствуют только в объеме, необходимом для описания данных в рамках выбранной формальной модели, но не могут сами быть объектом манипулирования и анализа. Знания статичны и не могут существенно изменяться без реинжиниринга всей системы. Алгоритмы в рамках данной модели имеют минимальную экспертно-определяемую структуризацию, основанную на процедурном, или, в лучшем случае, на объектно-ориентированном подходе, но также не могут быть объектом манипуляции и анализа со стороны самой системы. Отсутствие глубокого представления знаний и отсутствие явной связи действий и знаний приводят к потере согласованности абстракций в системе, невозможности автоматической адаптации алгоритмов обработки информации и реструктуризации системы без привлечения экспертов и реинжиниринга всей системы.

В работе предложена двухуровневая модель информационной структуры комплексной системы обработки информации (рисунок 2), свободная от данных недостатков и обеспечивающая явное сохранение знаний и действий.

Дополнительные связи в двухуровневой модели структуры системы обработки информации имеют следующую семантику:

- (dk) – отражает представление знаний в системе в виде данных второго порядка (метаданных);
- (ak) – отражает генерацию новых знаний через действия.

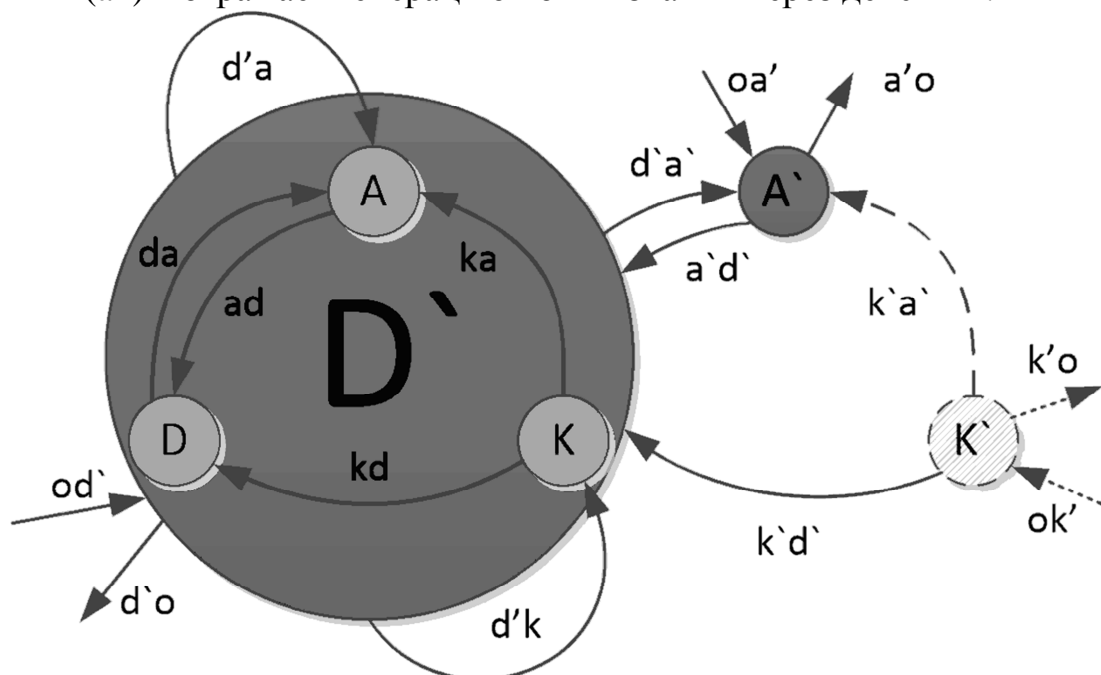


Рисунок 2 – Информационная модель структуры комплексной системы обработки информации

В модели комплексной системы обработки информации светлым цветом выделены элементы модели первого порядка (D , A , K), а темным – элементы модели второго порядка (D' , A' , K'), обеспечивающие управление элементами знаний и действий модели первого порядка, т. е. рассматриваю-

щие их как данные второго порядка. Такая модель позволяет на основе операций с данными второго порядка гибко изменять поведение системы – генерировать новые действия и знания. При соблюдении условия совместимости средств описания разных уровней модели системы обработки информации на основе единой модели данных, двухуровневая модель может быть «свернута» до одноуровневой полносвязной триадной модели.

В данной работе в качестве единого базиса описания информации предлагается использовать реляционную модель данных, расширенную: возможностью формального определения иерархий, для отражения высокоподобных структур данных; формальными средствами описания и структуризации алгоритмов и способами трансляции абстракций единого базиса описания информации в абстракции дескриптивных логик.

В третьей главе предложен вариант расширения реляционной модели данных новыми абстракциями и приемами структуризации, направленными на построение иерархий реляционных отношений и повышении согласованности информации.

Для расширения существующих возможностей структуризации в рамках реляционной модели данных предлагается использовать новый вид отношения между атрибутами – отношение семантической эквивалентности, определяемое на атрибутах (x, y) реляционных сущностей (R, P) следующим образом: $R.x \sim P.y$. Семантическая эквивалентность говорит об одинаковой смысловой нагрузке, которую несут атрибуты (т. е. семантически эквивалентные атрибуты отличаются только названием). В реляционных системах выполняется условие равенства доменов (D) семантически эквивалентных атрибутов: $R.x \sim P.y \Rightarrow D(R.x) = D(P.y)$.

Структуризация на основе семантической эквивалентности приводит к построению иерархий реляционных отношений. Универсальный вариант структуризации определяется следующим образом: пусть заданы отношения R с заголовком $\{A, X, Z\}$ и P с заголовком $\{B, Y, Q\}$, где A и B – первичные ключи отношений R и P ($A \rightarrow XZ, B \rightarrow YQ$) и выполняются следующие семантические эквивалентности: $A \sim B, X \sim Y$. Тогда процесс структуризации по критерию семантической эквивалентности может быть определен с помощью формирования вспомогательного отношения C с первичным ключом W , содержащей кортежи $C(W) = \{\{1\}, \{2\}\}$ следующим образом:

$$\begin{aligned} R^* &= \pi_{A,Z}(R) \times \sigma_{W=1}(C), \\ P^* &= \pi_{B,Q}(P) \times \sigma_{W=2}(C), \\ O^3 &= \pi_{A,X}(R) \times \sigma_{W=1}(C), \\ O^6 &= \rho_{X/Y} \left(\rho_{A/B} \left(\pi_{B,Y}(P) \times \sigma_{W=2}(C) \right) \right), \\ O &= O^3 \cup O^6. \end{aligned}$$

Процесс обратной композиции в этом случае может быть представлен следующим образом:

$$R = \pi_{A,X,Z}(R^* \bowtie O),$$

$$P = \rho_{Y/X} \left(\rho_{B/A} \left(\pi_{A,X,Q}(P^* \bowtie O) \right) \right).$$

Чтобы подтвердить отсутствие потерь и артефактов при выполнении данных преобразований доказаны:

- теорема «о разбиении», которая устанавливает, что для любого отношения $R\{A, X, Z\}$, $A \rightarrow XZ$, выполняется равенство $R = \pi_{A,X,Z} \left(\left(\pi_{A,X}(R) \times C \right) \bowtie \left(\pi_{A,Z}(R) \times C \right) \right)$ для любого $C \neq \emptyset$;
- теорема «о соединении с объединением», которая устанавливает истинность $R \bowtie T = \emptyset \Rightarrow R \bowtie S = R \bowtie (S \cup T)$.

Если рассмотреть процесс декомпозиции и синтеза отношений при построении иерархии на основе отношения семантической эквивалентности, объединив формулы, то выполняемые преобразования можно записать следующим образом:

$$R = \pi_{A,X,Z} \left(\left(\pi_{A,Z}(R) \times \sigma_{W=1}(C) \right) \bowtie \left(\left(\pi_{A,X}(R) \times \sigma_{W=1}(C) \right) \cup O^6 \right) \right),$$

$$R = \pi_{A,X,Z} \left(\left(\left(\pi_{A,Z}(R) \times \sigma_{W=1}(C) \right) \bowtie \left(\pi_{A,X}(R) \times \sigma_{W=1}(C) \right) \right) \cup \left(\pi_{A,Z}(R) \times \sigma_{W=1}(C) \right) \bowtie O^6 \right).$$

Так как W является первичным ключом отношения C и входит в состав первичного ключа отношений O^6 и $\pi_{A,Z}(R) \times \sigma_{W=1}(C)$, то $\left(\pi_{A,Z}(R) \times \sigma_{W=1}(C) \right) \bowtie O^6 = \emptyset$, так как: $O^6.W = 2$, $\pi_{A,Z}(R) \times \sigma_{W=1}(C).W = 1$.

Отсюда, согласно теореме «о соединении с объединением», в данном контексте

$$R = \pi_{A,X,Z} \left(\left(\pi_{A,Z}(R) \times \sigma_{W=1}(C) \right) \bowtie \left(\pi_{A,X}(R) \times \sigma_{W=1}(C) \right) \right),$$

что согласно теореме «о разбиении» является тождеством.

Таким образом, в работе доказано, что декомпозиция отношений и последующий их синтез при построении иерархий на основе семантической эквивалентности атрибутов отношений произведены без потерь и артефактов.

Также в работе предложен упрощенный способ построения иерархии реляционных отношений, основанный на требовании сквозной уникальности первичных ключей всех отношений в пределах всей схемы данных, т. е. требование определение всех первичных ключей в рамках одного домена в виде атомарного суррогатного ключа. В этом случае декомпозиция и синтез не требуют вспомогательного отношения C :

$$R^* = \pi_{A,Z}(R),$$

$$P^* = \pi_{A,Q}(P),$$

$$O = \pi_{A,X}(R) \cup \pi_{A,X}(P).$$

Процесс обратной композиции в этом случае может быть представлен следующим образом:

$$R = R^* \bowtie O,$$

$$P = P^* \bowtie O.$$

Построение иерархии реляционных сущностей на основе критерия семантической эквивалентности позволяет избежать аномалии неполного покрытия и аномалии соединения по омонимам.

Аномалия неполного покрытия определяется следующим образом: пусть дан запрос $K\{x\}$, для которого определена семантика x ($sem(x) = s$). И пусть существует множество отношений E , содержащих атрибуты с данной семантикой ($\exists E\{e_1, e_2, \dots, e_n\}, \forall e_i \in E s \in e_i$). Тогда невозможно на уровне реляционной модели данных гарантировать целостное выполнение запроса, применительно ко всему множеству отношений E (если $E' \subset E$, то $K\{x\}(E') \neq K\{x\}(E)$). На практике это означает, что при выполнении консолидирующих запросов (связанных с объединением данных разных отношений) существует вероятность реализации запроса не на всех необходимых отношениях, а только на их подмножестве. Актуальность данной аномалии особенно высока для баз данных с изменяющейся структурой, для которых необходимо постоянное отслеживание и изменение используемых запросов, с целью их актуализации и приведения в соответствие со схемой базы данных. Для сложных баз данных такие операции отличаются высокой трудоемкостью. Построение иерархии по критерию семантической эквивалентности позволяет избежать данной аномалии, так как выполняется условие $|E| = 1$.

Аномалия соединения по омонимам определяется следующим образом: пусть существует отношение R с заголовком $\{A, X\}$ и P с заголовком $\{B, Y\}$, где A и B – первичные ключи отношений R и P ($A \rightarrow X, B \rightarrow Y$) соответственно, и $D(A) = D(B)$. И пусть существует отношение $N \{C, Z\}$, где C – первичный ключ отношения N ($C \rightarrow Z$), такое, что $Z = FK(A)$. Тогда существует возможность построения нового отношения $M = N \bowtie P$, для которого $|M| \neq \emptyset$ (тело отношения M будет непустым). Возможность получения непустого тела такого отношения является аномалией соединения по омонимам, так как не несет полезной информации и тело отношения является артефактом. Построение иерархии по критерию семантической эквивалентности позволяет избежать данной аномалии, так как все первичные ключи определяются на одном домене и не могут содержать одинаковых значений.

Для повышения целостности и согласованности данных проведен анализ проблемы неопределенных значений и определены три их подвида.

Неизвестные значения (UNKN). Используются, когда значения атрибута какого-либо кортежа имеет семантический смысл, но точное значение ат-

рибута в данный момент неизвестно. Такое использование не несет опасности потери согласованности данных в структуре данных системы обработки информации.

Неиспользуемые значения (NIL). Значение NULL в данном случае интерпретируется как известное отсутствие значения для атрибута в кортеже (атрибут является *возможной*, но *необязательной* характеристикой объекта) и может быть заменено при изменении схемы отношений на иерархическую. В рамках реляционной теории эта операция может быть сформулирована следующим образом. Пусть задано отношение $R \{A, B, X\}$, где A – первичный ключ, B – атрибут (возможно составной), не допускающий значений NIL и X – атрибут (возможно составной) допускает значения NIL , тогда возможно произвести декомпозицию отношения R на два различных отношения (основное R^* и уточняющее R') по следующему правилу: $R^* = \pi_{A,B}(R)$, $R' = \sigma_{X \neq NIL}(\pi_{A,X}(R))$.

Обратный синтез исходного отношения можно получить при помощи операции соединения и использованию вспомогательного отношения $C\{X\} = \{\{NIL\}\}$: $O = (\pi_A(R^*) - \pi_A(R')) \bowtie R^*$, $R = (R' \bowtie R^*) \cup (O \times C)$.

Семантически запрещенные значения (NOP). В данном контексте значение NULL трактуется как отсутствие семантического смысла значения атрибута для данного кортежа и говорит об объединении сущностей по семантически эквивалентным атрибутам, что может быть разрешено построением явной иерархии на основе семантической эквивалентности, как это было рассмотрено выше.

Типизация неопределенных значений позволяет на этапе проектирования схемы данных системы обработки информации производить правильное структурирование данных, не позволяющее появлению неопределенных и семантически запрещенных значений, что повышает общую согласованность данных в системе.

Предложена новая формулировка требования совместимости реляционных отношений по теоретико-множественным операторам, заключающееся в прямом использовании активных доменов (AD) Кодда – особого вида доменов, которые определяются на основе явно используемых в отношении значениях атрибутов в соответствии со следующим определением: пусть дано отношение $R \{r_1, \dots, r_n\}$, тогда активный домен атрибута $r_i, i \in [1, n]$ $AD(r_i) = \pi_{r_i}(R)$.

Новый вид требования совместимости отношений сформулирован следующим образом: пусть даны отношения $A(a_1, \dots, a_n)$ и $B(b_1, \dots, b_n)$, $\forall a_i \in A \exists$ такой $b_j \in B$, что $a_i = b_j \ i \in [1, n], j \in [1, n]$, тогда результатом операции $C = A \cup B$ будет отношение $C(c_1, \dots, c_n)$, такое, что $\forall c_k \in C \exists$ такая пара a_i, b_j , что $c_k = a_i = b_j \ i \in [1, n], j \in [1, n], k \in [1, n]$ и $AD(c_k) =$

$AD(a_i) \cup AD(b_j)$ (активный домен атрибута результирующего отношения является объединением активных доменов атрибутов исходных отношений). В прикладных системах данное определение может быть дополнительно ограничено требованием совместимости типов одноименных атрибутов в отношениях, либо разрешено при помощи неявного преобразования типов, что является тривиальной технической задачей.

Предлагаемый подход, основанный на использовании активных доменов, позволяет снять ограничение на вычисление только доменно-независимых запросов и явно использовать оператор реляционного дополнения, а также позволяет осуществлять преобразование доменов на основе выполнения стандартных операций реляционной алгебры.

В рамках систем, основанных на активных доменах и новой формулировке требования совместимости, устранена аномалия семантической недоуверности дополнений сложных отношений, которая определяется следующим образом: пусть задано отношение $R\{A, X, Y\}$, определенное на доменах $D(A) = \{a1, a2\}$, $D(X) = \{x1, x2\}$, $D(Y) = \{y1, y2\}$. Если реальная предметная область имеет область определения $O = \{\{a1, x1, y1\}, \{a2, x2, y2\}\}$ то реляционное дополнение $\langle NOT \rangle \{a1, x1, y1\}$ не совпадет с логическим дополнением относительно предметной области $\neg \{a1, x1, y1\}$. Такое несоответствие связано с тем, что теория доменов устанавливает область определения любого отношения как декартово произведение доменов, в то время как область определения реальных систем не обязательно соответствует такому допущению. В общем виде, область определения реальной системы является произвольным подмножеством области определения реляционного отношения.

Для устранения данной аномалии предлагается различать область определения для атрибутов отношения, не являющихся внешними ключами отношений (простых атрибутов отношения), и атрибутов, которые связаны внешними ключами (ссылочными ограничениями целостности) с другими отношениями (ссылочных атрибутов).

Состав области определения атрибута отношения определяется:

- для простых атрибутов отношений область определения ограничена активным доменом данного атрибута;
- для ссылочных атрибутов область определения задается возможными значениями первичного (потенциального) ключа отношения, на которое ссылается внешний ключ (главного отношения).

Соблюдение данных требований при использовании иерархической организации реляционных сущностей на основе отношения семантической эквивалентности атрибутов отношений позволяет определять отношения в качестве области определения для атрибутов других отношениях, создавать иерархии областей определений, что обеспечивает возможность получения

правильных логических дополнений реляционных отношений на основе оператора взятия реляционной разности.

В четвертой главе рассмотрены методы описания и структуризации данных и алгоритмов их обработки в рамках интеграции расширенной реляционной модели данных и объектно-ориентированного подхода.

Предлагается следующий структурный изоморфизм расширенной реляционной и объектно-ориентированной моделями данных:

- отношение – дифференциальная часть класса;
- класс – производное (полное) отношение, которое может быть получено по следующей формуле: $E^{full} = \begin{cases} \text{если } \exists E_{пред}, \text{ то } E \\ E \bowtie E_{пред}^{full} \end{cases}$;

- объект – кортеж полного отношения;
- идентификатор объекта – первичный ключ отношения;
- поле класса – атрибут отношения.

Установление структурного изоморфизма между моделями позволяет перенести способы и приемы построения структур, используемые в рамках объектно-ориентированного подхода, на способы построения структур в расширенной реляционной модели данных.

В работе предлагается использовать свойство полиморфизма применительно к атрибутам реляционных отношений. Полиморфизм атрибута – это ограничение целостности на атрибуте отношения, которое может быть уточнено в рамках иерархии наследования реляционных отношений (рисунок 3).

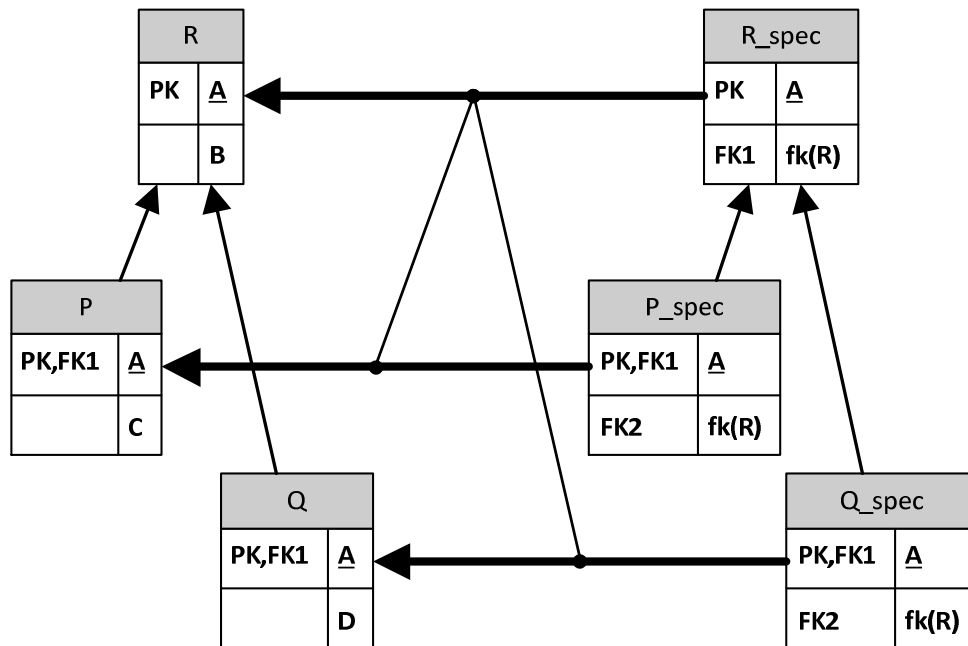


Рисунок 3 – Реализация полиморфизма атрибутов в расширенной реляционной схеме данных ($R_spec.fk(R)$ полиморфно в P_spec и Q_spec)

Реляционная модель данных не предлагает средств описания и структуризации алгоритмов обработки информации. В диссертации предлагается использовать объектно-ориентированную парадигму описания алгоритмов в виде иерархически-дифференциальной декомпозиции методов, так как построение иерархии реляционных сущностей на основе отношения семантической эквивалентности аналогично таксономии классов объектно-ориентированного подхода.

В работе предлагается определить реляционную операцию (инструкцию) в дифференциальной форме как атомарный элемент воздействия на данные. Каждая инструкция принадлежит тому отношению, кортежи которого изменяются при ее выполнении. Совокупность инструкций образует множество подобное атрибутам отношения.

Использование дифференциальной формы инструкций позволяет определить отношение семантической эквивалентности между ними и адекватно реалиям прикладной *sql*-систем, в которых операции разделены подобным же образом, на операции вставки, удаления и модификации данных. При этом инструкция является структурно более простым компонентом модели – область определения инструкции ограничена метаданными и не имеет отражения на уровне кортежей отношений, в отличие от атрибутов реляционных сущностей.

Семантика	Дифференциальная форма	Алгебраическая форма
Добавление	$R: I^+ = P$	$R = R \cup P$
Удаление	$R: I^- = P$	$R = R - P$
Изменение	$R: I_{\{B\}}^{\&} = P$	$R = (R - P) \cup \left(\pi_{A,C}(R) \bowtie \pi_{A,B}(P) \right)$

Отношение семантической эквивалентности для инструкций может быть сформулировано следующим образом: если типы инструкций совпадают и правые части инструкций равны, то инструкции семантически эквивалентны (для $I^{*1} = Op_1$ и $J^{*2} = Op_2$; $Op_1 = Op_2$ и $*_1 = *_2 \Rightarrow I^{*1} \sim J^{*2}$). На основе данного определения возможно проведение структуризации системы обработки информации не только по атрибутам, но и по инструкциям реляционных отношений (рисунок 4).

Дифференциальные инструкции являются простейшими элементами, реализующими поведение системы, однако они не могут выразить сложное поведение, связанное с согласованным внесением изменений в множество различных отношений.

Дальнейшую структуризацию алгоритмов предложено проводить на основе методов в объектно-ориентированном понимании данного термина – особого вида исполняемых процедур, наследуемых по иерархии наследования реляционных сущностей. Метод является именованной упорядоченной последовательностью инструкций и/или методов обработки данных. Методы,

подобно инструкциям, принадлежат классу, но эта принадлежность устанавливается экспертно, в зависимости от смысла выполняемых методом операций с данными, также как и в объектно-ориентированном подходе.

Структуризацию методов предлагается проводить на основе их нормализации и консолидации по семантической эквивалентности.

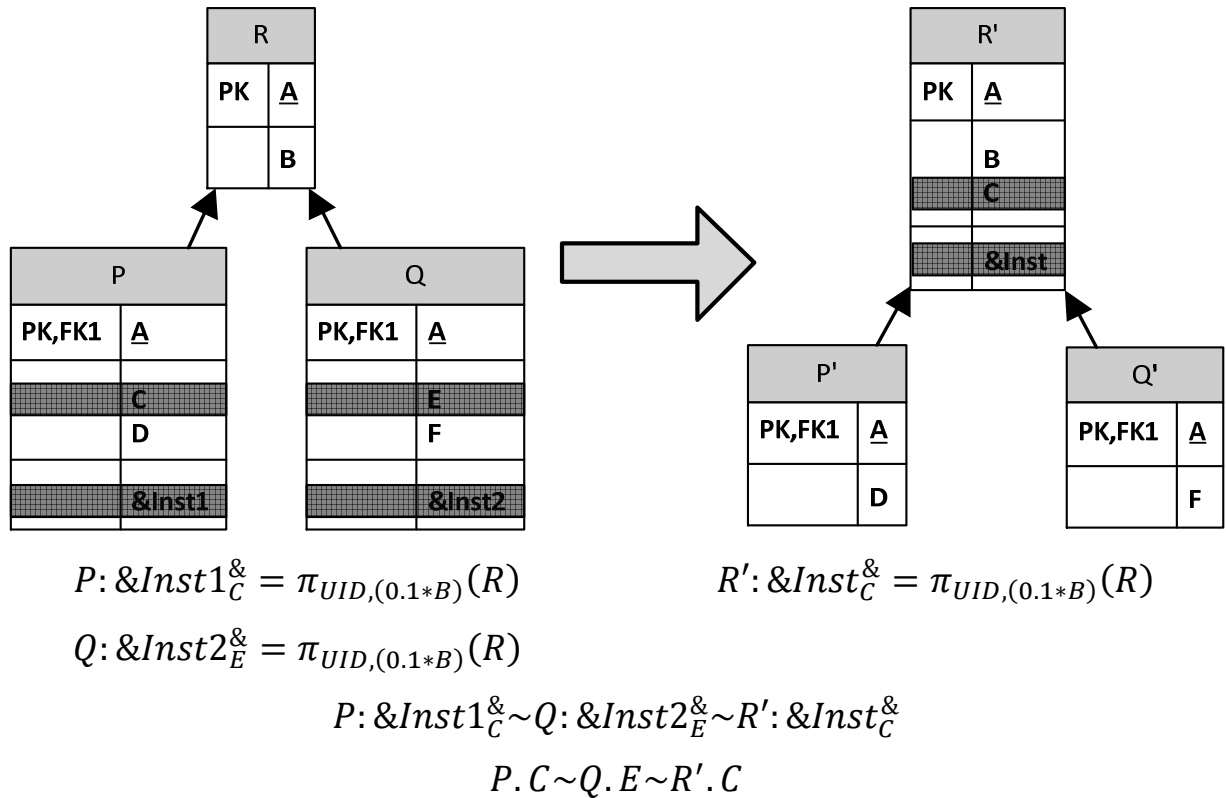


Рисунок 4 – Структуризация схемы данных и алгоритмов по семантически эквивалентным методам

Нормализация методов заключается в определении скрытых и вложенных методов (цепочек инструкций) и их консолидации в производных методах на родительском уровне иерархии. Таким образом, если в разных методах существуют одинаковые последовательности инструкций, то такая последовательность должна быть выделена в отдельный метод и вызов последовательностей данных инструкций в исходных методах должен быть заменен на вызов нового метода.

Формально данная операция может быть представлена следующим образом: пусть заданы методы $G\{g_1, g_2, \dots, g_m\}$; $F\{f_1, f_2, \dots, f_k\}$; и $\exists F'\{f_{i+1}, f_{i+2}, \dots, f_{i+n}\}$; $F' \subset F$; $\exists G'\{g_{j+1}, g_{j+2}, \dots, g_{j+n}\}$; $G' \subset G$ такие, что $F' = G' \Rightarrow$ необходимо создание $H\{h_1, h_2, \dots, h_n\}$, такого что $H = F' = G'$ и переопределения $F\{f_1, \dots, f_i, H, f_{i+n+1}, \dots, f_k\}$ и $G\{g_1, \dots, g_j, H, g_{j+n+1}, \dots, g_m\}$. Предлагаемая структуризация не зависит от семантики инструкций и методов и в прикладных системах может производиться автоматически, без необходимости привлечения экспертов.

Консолидация методов по семантической эквивалентности заключается в формировании полиморфной группы методов с одним именем в рамках иерархии наследования реляционных сущностей.

Формально данная операция может быть определена следующим образом: пусть заданы методы $G\{g_1, \dots, g_m, H, g_{m+1} \dots, g_{|G|}\}$, $F\{f_1, \dots, f_k, H, f_{k+1} \dots, f_{|F|}\}$ (возможна ситуация, когда $H = \emptyset$) и выполняется условие $G \sim F \sim H$, тогда можно реализовать вызов иерархии семантически эквивалентных методов на основе следующего правила:

$H^{full} = \begin{cases} \text{если } \nexists H_{пред}, \text{ то } H \\ \{H^-, H_{пред}, H^+\} \end{cases}$. В этом случае метод H определяется как родительский метод, тогда как методы G и F определяются как пары пред- и пост-

методов $G^-\{g_1, \dots, g_m\}$, $G^+\{g_{m+1} \dots, g_{|G|}\}$, $F^-\{f_1, \dots, f_k\}$, $F^+\{f_{k+1} \dots, f_{|F|}\}$. Подобное определение методов позволяет соблюдать их строгую инкапсуляцию с одной стороны и обеспечивает их полиморфизм с другой за счет уточняющего описания методов в иерархии наследования.

Применение предложенных методов структуризации данных и алгоритмов их обработки позволяет существенно уменьшить структурную сложность системы обработки информации.

В диссертации структурная сложность системы обработки информации рассчитывается на основании агрегатной сложности Н. П. Бусленко и определяется количеством составляющих систему элементов и связей. Структура элементов расширенной реляционной модели данных и алгоритмов имеет иерархическую природу и выражается через:

- реляционные сущности (E), состоящие из:
 - скалярных атрибутов (A);
 - ссылок на другие сущности – комбинация атрибута и внешнего ключа (C);
 - ссылок иерархии (H) – ссылок один-ко-одному, определенных на первичных ключах отношений;
 - реляционных инструкций (I), определенных на:
 - ◆ реляционных сущностях (Ie);
 - ◆ атрибутах (Ia);
 - ◆ ссылках (Ic);
 - методов (M), которые состоят из:
 - ◆ инструкций (Mi)

Таким образом, структурная сложность (O) определяется как вектор:

$$O \langle |E|, |A|, |C|, |H|, |I|, |Ie|, |Ia|, |Ic|, |M|, |Mi| \rangle.$$

Если считать дерево иерархии реляционных сущностей сбалансированным, а структуру системы обработки информации полносвязной (т. е. системой, которая строго описана иерархией по семантическим эквивалентностям

и не имеет корректных запросов, основанных на объединении отношений с разных ветвей иерархии) и весовые коэффициенты для каждого показателя структурной сложности считать равными (0.1), то относительная структурная сложность расширенной реляционной системы будет иметь вид, представленный на рисунке 5.

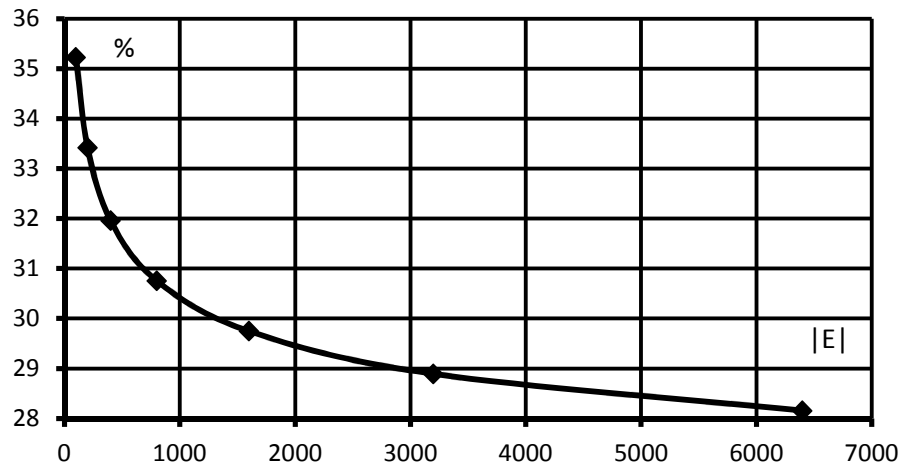


Рисунок 5 – Относительная структурная сложность расширенной реляционной схемы данных и алгоритмов их обработки к классической реляционной схеме

Как видно из данного графика, структурная сложность расширенной реляционной модели данных существенно ниже классической реляционной модели и её эффективность повышается с ростом количества моделируемых сущностей, что особенно актуально для больших и растущих систем.

Пятая глава посвящена разработке методов отображения пространства определения онтологий (концепции открытого мира) в пространство определения баз данных (концепцию закрытого мира) и способам прямой и обратной трансформации расширенных реляционных схем данных в онтологию дескриптивных логик.

Одна из основных проблем отображения онтологий на реляционную модель является высокая пространственная сложность такого отображения, связанная с тем, что в реальной системе для сложного предиката истинных высказываний существенно меньше чем ложных.

В данной работе предлагается способ отображения: выражение в рамках КОМ в виде n -местного предиката $M\{m_1, m_2, \dots, m_n\}$, который транслируется в $(n+1)$ -местное отношение $R\{r_1, r_2, \dots, r_n, r_{n+1}\}$, где $r_i \sim m_i$ ($i = \overline{1..n}$), а $r_{n+1} \sim M\{m_1, m_2, \dots, m_n\}$ в рамках концепции КЗМ. В этом случае отношение R , выступающее в роли проекции предиката, может быть представлено как объединение значимой тройки отношений, характеризующий возможные состояния предиката (истинные, ложные, и неизвестные):

$$R = (R_{true}) \cup (R_{false}) \cup (R_{unkn}),$$

$$\begin{aligned} R_{true} &= \sigma_{r_{n+1}=1}(R), \\ R_{false} &= \sigma_{r_{n+1}=0}(R), \\ R_{unkn} &= \sigma_{r_{n+1}=UNKN}(R). \end{aligned}$$

Так как в расширенной реляционной модели данных каждый атрибут отношения R может быть явно определен на домене-отношении (для $\forall r_i (i = \overline{1..n}) \exists E^{r_i}$ такое, что $r_i = FK(E^{r_i})$), то для R как для полного отображения всей области значения предиката M выполняется следующее условие:

$$R' = \pi_{r_1, r_2, \dots, r_n}(R) = \prod_{i=1}^n E^{r_i}.$$

При определении вспомогательного отношения C , задающего область определения для атрибута r_{n+1} $C\{\{0\}, \{1\}, \{UNKN\}\}$, можно вывести следующие эквивалентные отношения:

$$\begin{aligned} R_{true} &= \left(R' - \pi_{r_1, r_2, \dots, r_n}(R_{UNKN} \cup R_{false}) \right), \\ R_{false} &= \left(R' - \pi_{r_1, r_2, \dots, r_n}(R_{UNKN} \cup R_{true}) \right), \\ R_{UNKN} &= \left(R' - \pi_{r_1, r_2, \dots, r_n}(R_{true} \cup R_{false}) \right). \end{aligned}$$

Таким образом, при помощи неизвестных значений и трансляции всей области определения предиката в реляционное отношение становится возможным преобразовывать без потери предикаты, сформулированные в рамках КОМ в отношения расширенной реляционной модели данных, сформулированные в рамках КЗМ. Также на основе данных отношений можно определить явно хранимой любую пару отношений из тройки R_{true} , R_{false} и R_{UNKN} , исходя из соображений минимизации использования ресурсов, так как третье отношений может быть получено как разность пространства определения предиката и пары хранимых отношений, согласно приведенным формулам.

Для точного отображения таксономии классов в плоскость дескриптивных логик предлагается новый диалект дескриптивной логики, основанный на явных таксономиях (*ALT*-диалект). В рамках данного диалекта вводятся следующие новые понятия – типизация и наследование концептов.

Типизация концептов заключается в их разделении на базовые и производные концепты, где базовый концепт является средством определения таксономии понятий и отражает реальные типы сущностей в онтологии, в то время как производный концепт выполняет роль фильтра на множестве индивидов базового концепта и выделяет их некоторое подмножество по ролям (т. е. «человек» является базовым концептом, а «инженер» - производным концептом). Базовые концепты организуются в структуры при помощи нового оператора наследования («<»), определяющего расширение описания концепта и более строгого, нежели стандартный оператор поглощения. Произ-

водные концепты организуются при помощи стандартного оператора поглощения.

Предлагаемый подход позволяет формально отличать структурообразующие концепты, от всевозможных вторичных их проекций и пересечений по ролям, что позволяет упростить описание полновязных и близких к ним онтологий.

Для базовых концептов, в рамках предложенного *ALT*-диалекта, разработана матрица замены структурных конструкторов онтологии на структурные конструкторы расширенной реляционной модели данных (рисунок 6).

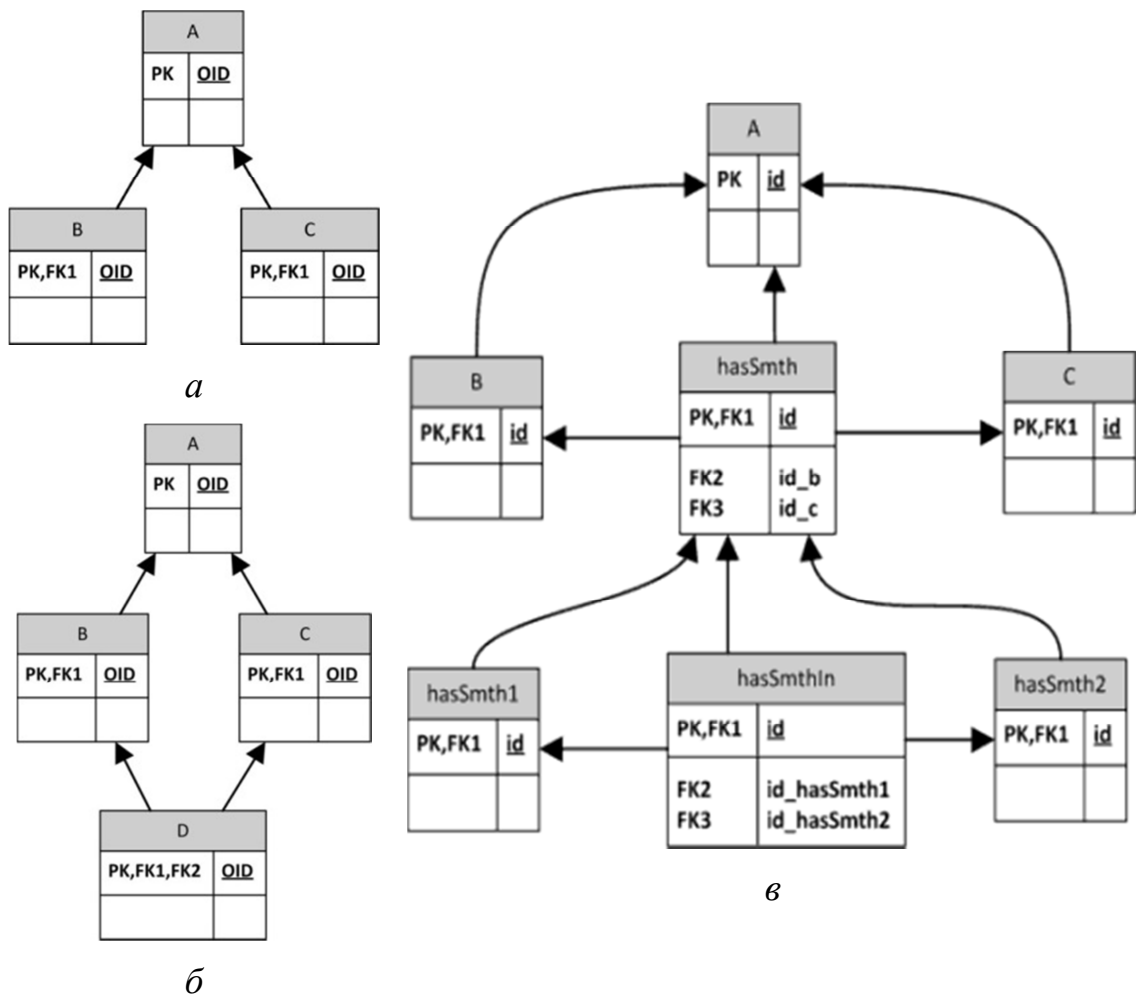


Рисунок 6 – Отображение *TBox* и *RBox* на расширенную реляционную модель данных (*a* – дизъюнкция, *б* – конъюнкция, *в* – иерархия, инверсия и вложенность ролей)

Транзитивные роли на уровне реляционной базы данных могут быть реализованы через представления на основе общих табличных выражений. Расширение языка формальными типами неотъемлемо присутствует в любой прикладной системе управления базами данных, а ограничения кардинальности ролей и функциональность полей могут быть реализованы на основе триггеров базы данных.

В шестой главе на основе предложенной методологии и прикладных методов разработана система обработки информации для поддержки деятельности предприятия ОАО УЗМ «Магнетрон». Данная система обработки информации представляет собой интегрированную среду ввода, хранения, обработки и визуализации оперативных и первичных аналитических данных о детальном состоянии происходящих и произошедших на предприятии учетных событиях и самих автоматизируемых процессах.

Логическая структура системы обработки информации для поддержки деятельности промышленного предприятия ОАО УЗМ Магнетрон представлена на рисунке 7.

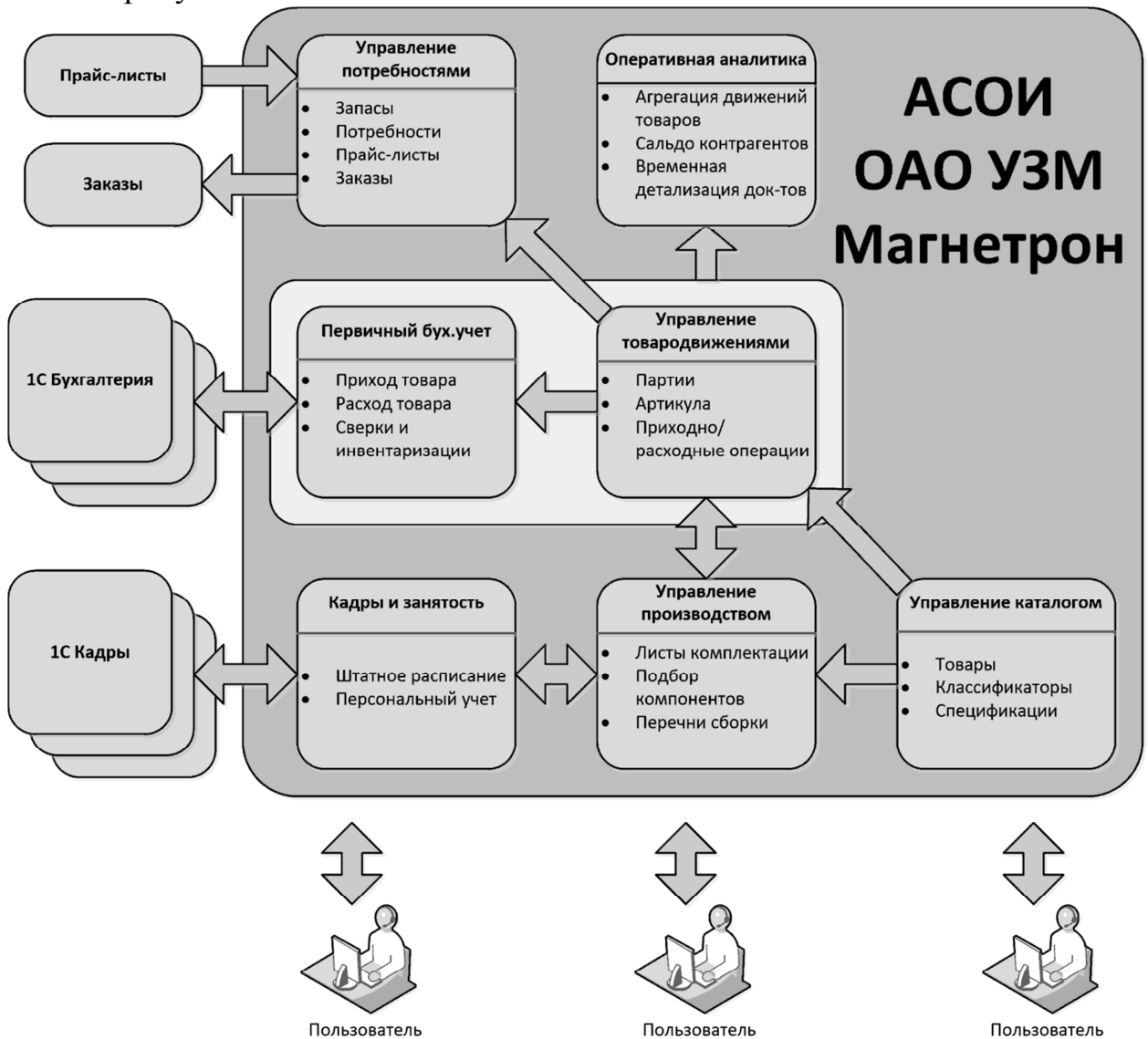


Рисунок 7 – Логическая структура системы обработки информации ОАО УЗМ Магнетрон

В рамках созданной системы обработки информации, исходя из требований бизнес-заказчика, были реализованы следующие подсистемы информационной поддержки деятельности предприятия:

- подсистема управления каталогом элементов, узлов, изделий и их спецификаций, позволяющая производить учет существующей номенклатуры МТР, осуществлять её экспорт и импорт, группировку и классификацию.
- подсистема управления внутренними и внешними товарными движениями компонентов, что позволяет производить точный партионный учет товародвижения на складах и цехах сборки изделий, автоматизировать процесс приема-отпуска МТР со склада, их тестирования, проверок, возврата поставщикам и отпуска заказчикам
- подсистема управления изготовлением узлов и изделий и контроля их качества, позволяющая осуществлять адресный подбор необходимых элементов изделий, учет производственных операций и контроль их результатов.
- подсистема управления элементными потребностями предприятия, планирование заказов и запасов элементов и узлов, основанная на анализе производственных планов, существующих ресурсов, планов выпуска продукции и предлагаемой поставщиками номенклатуры элементов и узлов и позволяющая осуществлять автоматическое формирование графиков производства и заказов по критерию минимизации себестоимости продукции;
- подсистема первичного бухгалтерского учета предприятия, позволяющая осуществить учет экономических операций в разрезе финансовых показателей;
- подсистема учёта кадров и управление рабочим временем работников предприятия, позволяющая вести учет и контроль их деятельности и оптимизировать выполнение бизнес-операций за счет правильного планирования и назначения заданий работникам;
- подсистема консолидированной оперативной аналитики по движениям товаров в разрезе партий и артикулов и движениям финансовых потоков в разрезе документов и контрагентов, позволяющая пользователям получать быстрый доступ к консолидированным финансовым и товарным данным.

На основе расширенной реляционной модели данных и алгоритмов была разработана структура системы обработки, фрагмент которой представлен на рисунке 8.

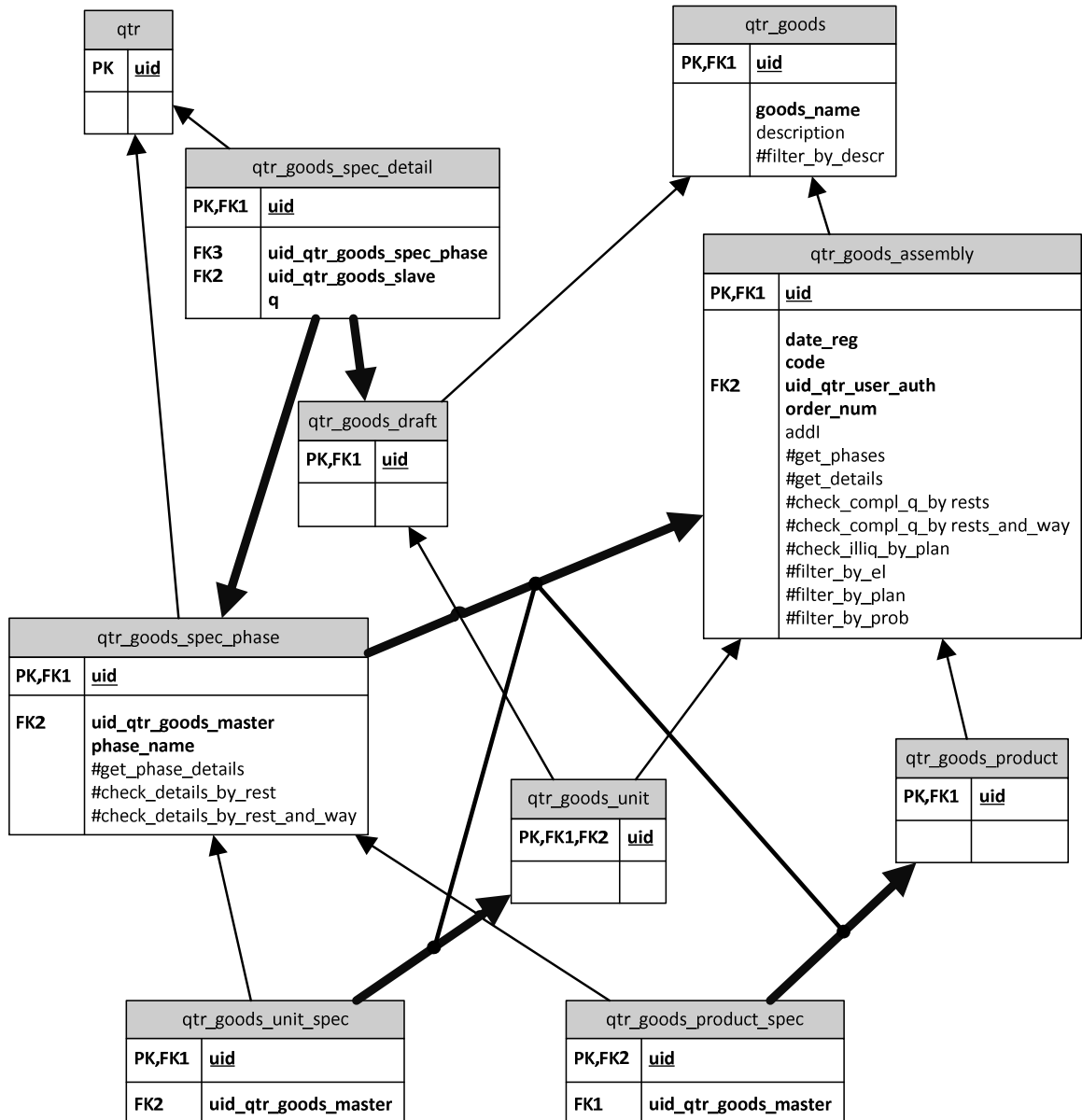


Рисунок 8 – Расширенная реляционная схемы данных и алгоритмов спецификации узлов и изделий (методы отмечены символом «#»)

Экспериментальное исследование эффективности разработанной системы обработки информации для поддержки деятельности предприятия показало:

1. Адекватность разработанных моделей, методов и алгоритмов требованиям, предъявляемым к системе обработки информации со стороны специалистов по бухгалтерскому, складскому, кадровому учету и высшего менеджмента организации.

2. Существенное снижение структурной сложности системы обработки информации (структурная сложность предложенной структуры данных и алгоритмов составила 122.5 пункта, что более чем в 2 раза ниже структурной сложности аналогичной структуры на основе классического реляционного базиса, которая составила 285.3 пункта).

3. Повышение целостности и согласованности данных и алгоритмов их обработки, вызванное объединением семантически эквивалентных абстракций расширенной реляционной модели данных и устранения неточностей и разночтений при операциях с ними, что выразилось в:

- отсутствию аномалии неполного покрытия пространства определения атрибута для 85 используемых в системе обработке информации методов (75 % методов);
- отсутствию аномалии соединения по омонимам для всех сущностей в структуре системы обработки информации;
- отсутствию семантически запрещенных и неиспользуемых значений во всех сущностях структуры системы обработки информации;
- выявлении и консолидации 78 вложенных и скрытых методов, охватывающих 70 сущностей в структуре иерархии (70 % сущностей);
- создании 7 полиморфных методов, охватывающих 33 сущности в структуре иерархии (33 % сущностей).

В заключении подчеркивается прогрессивность предложенных методов и технологий проектирования структуры системы обработки информации, когда последние интенсивно развиваются и усложняются. Сформулированы актуальные направления дальнейших научных исследований в области структурирования систем обработки информации.

ЗАКЛЮЧЕНИЕ

Основные научные положения диссертационной работы и их практическая реализация представляют собой дальнейшее развитие реляционной теории, теории объектно-ориентированного проектирования и дескриптивных логик. Прикладные аспекты работы посвящены решению научной проблемы – построению интегрированной структуры данных и алгоритмов системы обработки информации для поддержки деятельности предприятия.

В диссертации получены следующие новые теоретические и практические результаты:

1. Предложена информационная модель структуры системы обработки информации на основе триады «Данные»-«Знания»-«Действия», отличающаяся системной интеграцией элементов в рамках единого базиса описания информации и позволяющая формулировать задачи хранения данных, обработки информации и манипуляции знаниями в рамках единой модели данных.
2. Предложена расширенная реляционная модель данных, которая отличается:
 - предложенной структурообразующей функцией семантической эквивалентности, позволяющей создавать целостные иерархии реляционных отношений;

- типизацией неопределенных значений, что позволяет выявлять и устранять скрытые аномалии в структурах реляционных схем данных;
- предложенными правилами совместимости реляционных отношений по теоретико-множественным операторам и новой интерпретацией области определения атрибутов реляционных отношений, что позволяет повысить логическую целостность данных.

3. Предложены методы описания и структуризации алгоритмов в рамках расширенной реляционной модели данных, отличающиеся:

- учетом алгоритмов как элементов данной модели, что позволяет интегрировано описывать данные и алгоритмы в рамках одной структуры;
- структуризацией алгоритмов на основе устранения повторяющихся инструкций и семантических эквивалентностей, что позволяет минимизировать количество алгоритмов в системе, повышая целостность и согласованность структуры системы обработки информации в целом.

4. В рамках интеграции онтологий и расширенной реляционной модели данных предложены:

- метод отображения концепции открытого мира в концепцию закрытого мира, основанный на покрытии пространства определения открытого мира суперпозицией реляционных выражений; данный метод отличается использованием неопределенных значений и возможностью выбора совокупности реляционных выражений имеющих минимальную пространственную сложность, что позволяет производить трансляцию выражений дескриптивных логик в выражения реляционной модели данных;
- вариант дескриптивной логики, соответствующий иерархически-реляционным структурам в рамках расширенной реляционной модели данных, отличающийся использованием оператора наследования концептов и типизацией концептов, что позволяет транслировать расширенные реляционные схемы данных в терминологические аксиомы дескриптивной логики и обратно.

5. Разработана прикладная программная реализация системы обработки информации для поддержки деятельности предприятия, основанная на предложенной расширенной реляционной модели данных и алгоритмов их обработки. В рамках системы обработки информации разработаны подсистемы: управления каталогом, управления товародвижением, управления производством, управления кадрами, первичного бухгалтерского учета, управления потребностями и заказами предприятия. Разработанная система обработки информации внедрена в промышленных и торговых предприятиях России.

Экспериментальные исследования эффективности предложенных моделей и алгоритмов показали их адекватность требованиям бизнес- и менеджмент-специалистов. Применение расширенной реляционной модели данных и алгоритмов позволило существенно увеличить их целостность и со-

гласованность, что выразилось в устранении: аномалии неполного покрытия пространства определения атрибута, актуальной для 75 % методов; аномалии соединения по омонимам, актуальной для всех методов в системе; семантически запрещенных и неиспользуемых значений для всех сущностей в структуре системы обработки информации; а также в выявлении и консолидации вложенных и скрытых методов, которые охватывают 70 % сущностей в разработанной структуре и создании полиморфных методов, охватывающих 33 % сущностей в системе обработки информации.

Рекомендации и перспективы дальнейшей разработки темы. Важным направлением дальнейшего развития данной тематики является реализация контрактного исполнения методов агентами активности, что позволит в условиях динамического информационного окружения реализовать субоптимальный выбор прикладных алгоритмов обработки информации. Другим перспективным направлением дальнейших исследований является определение дополнительных метрик структурной сложности, связанных с выявлением циклов в рамках методов и инструкций, что позволит оценить потенциальную устойчивость системы обработки информации.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

В рецензируемых журналах, входящих в перечень ВАК

1. Модели представления данных в информационных системах реального времени / А. А. Левков, В. А. Масленников // Вестник УГАТУ. 2005. № 1 (12). С. 121–134.
2. Проблемы организации структуры данных в сверхбольших базах данных / А. А. Левков, В. А. Масленников // Системы управления и информационные технологии: науч.-техн. журн. 2007. № 3.1 (29). С. 153–156.
3. Методы использования классовых алгоритмов и структур в реляционных базах данных / А. А. Левков // Системы управления и информационные технологии. № 4.1 (30). 2007. С. 159–162.
4. Методики формального описания классово-ориентированной модели данных / А. А. Левков // Системы управления и информационные технологии. № 4.1 (30). 2007. С. 162–166.
5. Методы оптимизации хранения информации об испытаниях авиационных двигателей / А. А. Левков, С. П. Павлинич // Нефтегазовое дело. № 2. 2007. С. 20–25.
6. Методы повышения эффективности систем управления организационно–техническими объектами / А. А. Левков // Мехатроника, автоматизация, управление. № 7. 2008. С. 42–47.
7. Построение эффективных онтологий в больших системах управления предприятием / А. А. Левков // Искусственный интеллект и принятие решений. № 4. 2011. С. 92–101.

8. Организация эффективной системы хранения фактов в онтологиях / А. А. Левков // Информационные технологии и вычислительные системы. № 4. 2011. С. 3–9.

9. Управление данными в корпоративных ИС / А. А. Левков // Вестник УГАТУ. № 1 (41). 2011. С. 176–181.

10. Организация эффективных полиморфных алгоритмов в сложных системах управления информацией / А. А. Левков // Прикладная информатика. № 5 (35). 2011. С. 127–133.

11. Оценка структурно–алгоритмической сложности реляционных схем данных / А. А. Левков, Б. Г. Ильясов // Вестник компьютерных и информационных технологий. № 4. 2011. С. 28–32.

12. Структурная оптимизация реляционных моделей сложных иерархических систем / А. А. Левков, Б. Г. Ильясов // Информационные технологии. № 3. 2011. С. 50–54.

13. Триадный подход к управлению интеллектуальными информационными системами (теоретические основы) / А. А. Левков, Б. Г. Ильясов // Вестник компьютерных и информационных технологий. № 4. 2011. С. 3–6.

14. Триадный подход к управлению интеллектуальными информационными системами (практические аспекты) / А. А. Левков, Б. Г. Ильясов // Вестник компьютерных и информационных технологий. № 1. 2012. С. 3–7.

15. Интеграция знаний и данных в больших информационных системах / А. А. Левков // Информационные технологии. № 2. 2012. С. 29–33.

Монография

16. Структурная оптимизация многомерных систем хранения данных / А. А. Левков, Ю. С. Кабальнов, В. А. Масленников. М.: “Изд. МАИ”, 2009. 140 с.

Зарегистрированные программы для ЭВМ

17. Свид. об офиц. рег. программы для ЭВМ № 2013612328. Программа для автоматизированного структурирования реляционных схем данных и алгоритмов их обработки по отношению семантической эквивалентности / А. А. Левков. Зарег. 21.02.2013.

18. Свид. об офиц. рег. программы для ЭВМ № 2013612325. Программа для обеспечения деятельности кадрового подразделения предприятия / А. А. Левков, В. А. Масленников. Зарег. 21.02.2013.

Другие публикации

19. Модуль адаптации многомерных структур хранения данных / А. А. Левков, Д. В. Ивлев // Информационные и кибернетические системы управления и их элементы: матер. всеросс. молодежн. науч.-техн. конф., 25–26 ноября 1997. Уфа: УГАТУ, 1997. С. 49–50.

20. Система предварительной обработки информации в составе интегрированной АСУТП нефтегазодобычи / А. А. Левков, Ю. И. Зозуля, С. И. Братцев, Ю. С. Кабальнов, Р. Р. Каримов, Д. В. Ивлев // Вычислитель-

ная техника и новые информационные технологии: межвуз. науч. сб. Уфа: изд. УГАТУ, 1998. С.59–65.

21. Автоматизированная система хранения и обработки диагностической информации основе концепции многомерных баз данных применительно к объектам нефтедобычи / А. А. Левков, Ю. И. Зозуля, Ю. С. Кабальнов, Р. Р. Каримов, Д. В. Ивлев // Проблемы нефтегазового комплекса России: междунар. науч.-техн. конф. Уфа: УГНТУ, 1998. С. 143–147.

22. Анализ достоверности данных в системах реального времени / А. А. Левков, Ю. И. Зозуля, С. И. Братцев, Ю. С. Кабальнов, Р. Р. Каримов, Д. В. Ивлев // Новые технологии управления движением технических объектов: 1-я междунар. конф. 13–15 января, 1999. Ставрополь. С. 174–175.

23. Информационная система реального времени в составе интегрированной АСУТП нефтегазодобычи / А. А. Левков, Ю. И. Зозуля, С. И. Братцев, Ю. С. Кабальнов, Р. Р. Каримов, Д. В. Ивлев // Управление сложными системами: межвуз. науч. сб. Уфа: изд. УГАТУ, 1999. С. 37–44.

24. Оценка эффективности индексных структур в многомерных базах данных / А. А. Левков, Ю. И. Зозуля, С. И. Братцев, Ю. С. Кабальнов, Р. Р. Каримов, Д. В. Ивлев // Интеллектуальные системы управления и обработки информации: междунар. науч. техн. конф. Уфа: УГАТУ, 1999. С. 69.

25. Хранение и обработка данных в информационных системах в реальном времени, используя многомерные технологии / А. А. Левков, Ю. С. Кабальнов, Д. В. Ивлев, О. И. Христодуло // Информатика и информационные технологии (CSIT'2000): тр. 2-й междунар. конф., 18–23 сентября, 2000. Уфа, изд. УГАТУ, 2000. Т. 1. С. 49–50.

26. Применение многомерных технологий при создании информационных систем в образовании / А. А. Левков, Б. Г. Ильясов, Ю. С. Кабальнов, Д. В. Ивлев // Телематика 2001: междунар. науч. методич. конф. Санкт-Петербург: СПбГУ ИТМО, 2001. С. 62.

27. Применение многомерных технологий в информационных системах реального времени / А. А. Левков, Ю. С. Кабальнов, Д. В. Ивлев, О. И. Христодуло // Вопросы управления и проектирования в информационных кибернетических системах: межвуз. науч. сбор. Уфа: изд. УГАТУ, 2001. С. 141–151.

28. Применение бит–логической модели информационного представления для организации эффективного взаимодействия пользователя с многомерными базами данных / А. А. Левков, Д. В. Ивлев // Информатика и информационные технологии (CSIT'2002): тр. 4-й междунар. конф., 18–24 сентября, 2002. Греция, университет г. Патры. С. 39.

29. Применение многомерных информационных технологий в системах открытого образования / А. А. Левков, Д. В. Ивлев // Информатика и информационные технологии (CSIT'2002): тр. 4-й междунар. конф., 18–24 сентября, 2002. Греция, университет г. Патры. С. 42.

30. Методы математического описания информационных объектов гиперкуба многомерной базы данных / А. А. Левков, Д. В. Ивлев // Проблемы техники и технологии телекоммуникаций: матер. 3-й междунар. науч.-техн. конф., 26–27 ноября, 2002. Уфа: УГАТУ. С. 57–58.

31. Применение циклических структур для организации эффективного размещения информации на физических носителях в СУБД / А. А. Левков, О. И. Христодуло // Телекоммуникации и информатизация образования. № 1. 2003. С. 43–52.

32. Применение синтеза многомерного гиперкуба для увеличения централизации данных / А. А. Левков, О. И. Христодуло, Д. В. Ивлев // 5-й международный симпозиум по информатике и информационным технологиям. Уфа: УГАТУ, 2003. С. 169–174.

33. Увеличение производительности многомерных баз данных, применяя древовидные структуры поиска для систем дистанционного образования / А. А. Левков, Д. В. Ивлев // 5-й международный симпозиум по информатике и информационным технологиям. Уфа: УГАТУ, 2003. С. 75–78.

34. Объектно-ориентированные модели для хранения данных / А. А. Левков, В. А. Масленников // Компьютерные науки и информационные технологии: матер. IX междунар. сем. Уфа: УГАТУ, 2007. С. 87–89.

35. Повышение эффективности реинжиниринга сложных БД при помощи использования объектно-ориентированных БД / А. А. Левков, В. А. Масленников // Проблемы техники и технологий телекоммуникаций: 7-я междунар. конф. Самара: ПГУТИ, 2006. С. 34–36.

36. Реализация наследования при оптимизации объектно-ориентированной модели данных / А. А. Левков, В. А. Масленников // Проблемы техники и технологий телекоммуникаций: 7-я междунар. конф. Самара: ПГУТИ, 2006. С. 48–50.

37. Оптимизация структуры базы данных оперативного учета и документооборота / А. А. Левков, В. А. Масленников // Информационно-математические технологии в экономике, технике и образовании: II междунар. науч.-техн. конф. Екатеринбург: УГТУ–УПИ, 2007. С. 54–55.

38. Методология построения системы оперативного учета и документооборота / А. А. Левков, В. А. Масленников // Мавлютовские чтения: Всероссий. молодежн. науч. конф., посвященная 75-летию УГАТУ. Уфа: УГАТУ, 2007. С. 41–43.

39. Способы оптимизации построения логических моделей данных ОО БД / А. А. Левков, В. А. Масленников // Проблемы техники и технологии телекоммуникаций: VIII междунар. науч.-техн. конф. Уфа, 2007. С. 144–146.

40. Использование классовых алгоритмов и структур в реляционных базах данных / А. А. Левков, В. А. Масленников // Решетневские чтения: XI междунар. науч. конф. Красноярск: СГАУ, 2007. С. 36–37.

41. Управление структурой информации в сложных реляционных системах / А. А. Левков // Проблемы управления и моделирования в сложных системах: междунар. науч. техн. конф. Самара: ИПУСС РАН, 2012. С. 227–231.

42. Управление информацией в сложных системах на основе семантики / А. А. Левков, Б. Г. Ильясов // Современные сложные системы управления: междунар. науч. техн. конф. С. Оскол: Тонкие наукоемкие технологии, 2012. С. 48–51.

43. Построение комплексных информационных систем управления данными и знаниями в системах расширенного управления предприятием / А. А. Левков // Информационные технологии в управлении: междунар. науч. техн. конф. С.-Петербург: СПИИРАН, 2012.

44. Классификация неопределенных значений для правильной структуризации реляционных отношений / А. А. Левков // Информатика и информационные технологии (CSIT'2012): тр. 14-й междунар. конф., 2012. Уфа: УГАТУ. № 2. С. 57-60. (опубликовано на английском языке).

45. Определение операций в расширенной реляционной модели данных / А. А. Левков // Информатика и информационные технологии (CSIT'2012): тр. 14-й междунар. конф., 2012. Уфа: УГАТУ. № 2. С. 74-77. (Опубликовано на англ. языке).

46. Построение эффективных онтологий в больших корпоративных информационных системах / А. А. Левков // Научно-техническая информация. 2012. № 6 (39). С. 328-335. (Опубликовано на англ. языке).

47. Концептуальная модель современной системы обработки информации для систем управления предприятием / А. А. Левков // Информационные технологии и системы: междунар. науч. техн. конф. Челябинск: ЧелГУ, 2013. С. 165–170.

48. Повышение согласованности данных в современных информационных системах / А.А. Левков // Информационные технологии и системы: междунар. науч. техн. конф. Челябинск: ЧелГУ, 2013. С. 139–142.

ЛЕВКОВ Александр Александрович

МЕТОДОЛОГИЯ ПОСТРОЕНИЯ СТРУКТУРЫ
СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ
НА ОСНОВЕ РАСШИРЕННОЙ РЕЛЯЦИОННОЙ
МОДЕЛИ ДАННЫХ И АЛГОРИТМОВ
(на примере информационной поддержки
деятельности предприятия)

Специальности 05.13.01 – Системный анализ, управление и обработка
информации (в промышленности)

Автореферат диссертации на соискание ученой степени
доктора технических наук

Подписано к печати 26.08.2013. Формат 60x84 1/16.
Бумага офсетная. Печать плоская. Гарнитура Times New Roman Cyr.
Усл. печ. л. 2,0. Усл. кр.-отт. 1,0. Уч.–изд. л.2,0.
Тираж 100 экз. Заказ № 443

ФГБОУ ВПО Уфимский государственный авиационный технический
университет
Центр оперативной полиграфии
450000, Уфа-центр, ул. К.Маркса, 12