

На правах рукописи

Дейнеко Александр Олегович

**МОДЕЛИ, МЕТОДЫ И ПРОГРАММНЫЕ СРЕДСТВА  
РАСПРЕДЕЛЕННОГО ПРИОБРЕТЕНИЯ ЗНАНИЙ ДЛЯ  
АВТОМАТИЗИРОВАННОГО ПОСТРОЕНИЯ БАЗ  
ЗНАНИЙ ИНТЕГРИРОВАННЫХ ЭКСПЕРТНЫХ  
СИСТЕМ**

Специальность 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

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

Автор:

Москва – 2011

Работа выполнена в Национальном исследовательском ядерном университете «МИФИ».

Научный руководитель: доктор технических наук, профессор  
Рыбина Галина Валентиновна

Официальные оппоненты: доктор технических наук, профессор  
Еремеев Александр Павлович  
кандидат технических наук, доцент  
Тарасов Валерий Борисович

Ведущая организация: Институт системного анализа  
Российской академии наук

Защита состоится «    » \_\_\_\_\_ 2011 г. в \_\_ час. \_\_ мин. на заседании диссертационного совета Д 212.130.03 при Национальном исследовательском ядерном университете «МИФИ» по адресу: Москва, Каширское ш., 31 (конференц-зал, 2 этаж главного корпуса)

С диссертацией можно ознакомиться в библиотеке Национального исследовательского ядерного университета «МИФИ».

Отзывы в двух экземплярах, заверенные печатью, просьба направлять по адресу: Каширское ш., 31, Ученый Совет МИФИ.

Автореферат разослан « \_\_ » \_\_\_\_\_ 2011 г.

Учёный секретарь  
диссертационного совета

Леонова Н.М.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования.** Приобретение знаний является центральной проблемой, возникающей при разработке современных интеллектуальных систем, в частности, традиционных экспертных систем и более сложных – интегрированных экспертных систем (ИЭС), обладающих масштабируемой архитектурой и расширяемой функциональностью. Однако, несмотря на значительное число исследований и разработок в этой важнейшей области искусственного интеллекта вопросы практического использования традиционных методов приобретения знаний и создания технологии автоматизированного приобретения знаний по-прежнему являются актуальной проблемой.

Существенный дефицит экспертов, возрастание объемов противоречивых и «зашумленных» данных и знаний, нехватка специальных компьютерных систем, имитирующих искусство эксперта/экспертов, слабая теоретическая и технологическая исследованность методов приобретения знаний из электронных источников знаний и другие проблемы наиболее остро ощущаются при создании интегрированных экспертных систем, предназначенных для решения сложных практических задач, особенно в таких областях, как медицина, энергетика, космос, экология и др., что приводит к значительному удорожанию стоимостных и временных параметров разработки систем подобного класса сложности.

Опыт практического использования целого ряда прикладных ИЭС, разработанных на основе задачно-ориентированной методологии (ЗОМ), предложенной Г.В. Рыбиной в середине 90-х годов, и поддерживающего её инструментального комплекса АТ-ТЕХНОЛОГИЯ (в том числе для диагностики сложных технических систем, проектирования уникальных объектов машиностроения, комплексных экологических задач, экспресс-диагностики крови и др.) показал необходимость *мониторинга*, т.е. проведения регулярных проверок и подтверждений накапливаемых и формализуемых знаний в соответствующих базах знаний (БЗ), чтобы их качество не отразилось на функционировании ИЭС в целом.

Соответственно возрастает актуальность и роль разработки методов и средств автоматизации труда экспертов и создания специальных программных средств, направленных на компьютерную поддержку процессов получения знаний от эксперта (или групп экспертов), являющихся основным источником знаний. С другой стороны, типология источников знаний уже не ограничивается только экспертами, поскольку значительные объемы экспертных знаний накоплены в текстах на естественных языках, а в последние годы – в информации, аккумулирующейся в современных информационных бизнес-системах, в частности, большие информационные ресурсы накоплены в базах данных (БД).

Проблеме автоматизированного извлечения знаний из БД в искусственном интеллекте посвящены такие новые направления как Data Mining и Knowledge Discovery in Databases (KDD). Однако эти технологии возникли и развивались независимо от технологий автоматизированного приобретения знаний от экспертов, и сегодня подобная *автономность и распределенность* не позволяет осуществлять эффективный мониторинг всех информационных ресурсов (БЗ, БД, а в последние годы и онтологии), которыми обладают интеллектуальные системы, в частности ИЭС. В настоящее время, практически, отсутствуют исследования в области создания инструментальных средств и технологий *распределенного приобретения знаний из источников различной типологии*.

Значительный вклад в исследование и решение проблем приобретения знаний внесли отечественные учёные О.И.Ларичев, Д.А.Поспелов, Э.В.Попов, Г.С.Осипов, Г.В.Рыбина, А.П.Еремеев, Т.А.Гаврилова, И.Б.Фоминых, В.Ф.Хорошевский, В.Л.Стефанюк, Н.Г. Загоруйко, В.Б.Тарасов, Б.А. Кобринский, А.С.Нариньяни, О. К. Подлипский, А.Б. Петровский, В.К. Финн и др., а также целый ряд зарубежных учёных S. Easterbrook., J. Quinlan, K. Sreeraman, P. Gregory, R. Dieng, D. Waterman, T. Nguyen, H. Toivonen, F. Coenen, J. Tsai, P. Meseguer, K. Wang и др.

Таким образом, актуальность темы диссертации определяется возрастающей потребностью индустрии интеллектуальных систем в современных эффективных средствах автоматизированного приобретения знаний, распределенных по источникам знаний различной типологии.

**Цель работы.** Целью данной диссертационной работы является автоматизация процессов построения баз знаний ИЭС на основе разработки моделей, методов и программных средств распределенного приобретения знаний.

Для достижения поставленной цели в диссертации решены следующие задачи:

- 1) На основе анализа существующих моделей, методов и средств приобретения знаний выбор базисного метода автоматизированного приобретения знаний, на основе которого предложен новый подход к интеграции информации, полученной из источников знаний различной типологии, предложено и исследовано понятие «распределённого приобретения» знаний с целью автоматизированного построения максимально полных и непротиворечивых моделей проблемных областей.
- 2) Разработка оригинального алгоритма распределенного приобретения знаний из баз данных, основанного на построении бинарных деревьев решений, существенным отличием которого от базового алгоритма CART является ориентация на приобретение знаний из множества баз данных и оптимизация структуры и количества выведенных продукционных правил.

3) Разработка метода и алгоритмов группового приобретения знаний, основанных на вычислении вероятности принадлежности по мнению группы экспертов каждого рассматриваемого объекта определенному классу.

4) Разработка метода и алгоритмов объединения продукционных правил, полученных из источников знаний различной типологии, основанных на применении теории мультимножеств, расширенных таблиц решений и таблиц мер схожести правил, строящихся путем вычисления мер схожести Хемминга для посылок и заключений правил.

5) Разработка инструментальных программных средств приобретения знаний, распределенных по источникам знаний различной типологии, и средств объединения продукционных правил, полученных из различных источников.

6) Проведение экспериментальной апробации предложенных алгоритмов и разработанных инструментальных программных средств на задачах медицинской диагностики и задачах контроля радиационных дозовых нагрузок персонала атомных электростанций РФ.

**Объект и предмет исследования.** Объектом исследования являются БЗ интегрированных экспертных систем. Предметом исследования являются методы распределенного приобретения знаний для автоматизированного построения БЗ интегрированных экспертных систем.

**Методы исследования.** Для решения поставленных задач в диссертации использованы методы искусственного интеллекта (модели и методы представления и обработки знаний, методы инженерии знаний), методы Data Mining (методы классификации, построения деревьев решений), метод эвристической классификации, теория мультимножеств, теория множеств и отношений, технология разработки программного обеспечения.

**Научная новизна.** В диссертации получены следующие новые результаты:

1. С учетом распределенного варианта приобретения знаний предложен расширенный для источников знаний различной типологии комбинированный метод автоматизированного приобретения знаний.

2. Предложен и исследован алгоритм распределенного приобретения знаний из БД, основанный на построении бинарных деревьев решений, существенным отличием которого от базового алгоритма CART является ориентация на приобретение знаний из множества БД и оптимизация структуры и количества выведенных правил.

3. Впервые предложен и исследован метод объединения продукционных правил, полученных из источников знаний различной типологии.

4. Впервые разработаны инструментальные программные средства распределенного приобретения знаний из БД для автоматизированного построения БЗ интегрированных экспертных систем и рекомендации по их применению.

Таким образом, в результате выполненных исследований разработаны модели, методы, алгоритмы и инструментальные программные средства, позволяющие снизить трудоемкость и временные затраты на разработку поля знаний и БЗ в процессе разработки прикладных ИЭС.

**Основные научные результаты, выносимые на защиту.**

1. Расширенный для источников знаний различной типологии комбинированный метод приобретения знаний, построенный на основе выполненного анализа существующих моделей, методов и средств приобретения знаний.
2. Алгоритм распределенного приобретения знаний из БД, основанный на построении бинарных деревьев решений, существенным отличием которого от базового алгоритма CART является ориентация на приобретение знаний из множества БД и оптимизация структуры и количества выведенных правил.
3. Метод объединения продукционных правил, полученных из источников знаний различной типологии, основанный на применении теории мультимножеств, расширенных таблиц решений и таблиц мер схожести правил, строящихся путем вычисления мер схожести Хемминга для посылок и заключений правил.
4. Комплекс инструментальных программных средств распределенного приобретения знаний из БД для автоматизированного построения БЗ ИЭС.

**Практическая ценность и значимость работы** заключается в том, что созданные инструментальные программные средства в составе комплекса АТ-ТЕХНОЛОГИЯ используются в учебном процессе НИЯУ «МИФИ» на кафедре Кибернетики, а также в деятельности ОАО «Концерн Росэнергоатом» (акты о внедрении).

Апробация разработанных моделей, методов и инструментальных программных средств при создании нескольких БЗ для задач медицинской диагностики и задач контроля радиационных дозовых нагрузок персонала атомных электростанций РФ подтвердили практическую значимость результатов диссертации.

**Реализация результатов диссертации.** Результаты диссертации использовались в НИР, выполненной при поддержке РФФИ (проект № 09-01-00638) в учебно-научной лаборатории “Интеллектуальные системы и технологии” кафедры Кибернетики МИФИ. В феврале 2011 работа стала победителем конкурса НИЯУ МИФИ по программе "У.М.Н.И.К.-2011".

**Апробация результатов.** Основные результаты диссертации докладывались и обсуждались на 12-й национальной конференции по искусственному интеллекту с международным участием КИИ-2010 (Тверь, 2010), на конференции «Реинжиниринг бизнес-процессов на основе современных информационных технологий. Системы управления процессами и знаниями (РБС-СУЗ-2010)» (Москва, 2009), международном семинаре «Интегрированные модели и мягкие вычисления, вероятностные системы и комплексы программ

в искусственном интеллекте» (Коломна, 2009), 17-ом и 18-ом международных научно-технических семинарах (Алушта 2009, 2010), ежегодных «Научных сессиях МИФИ» с 2007 по 2011 гг.

**Публикации.** Основные результаты диссертационного исследования опубликованы в десяти печатных трудах, в том числе в трех статьях в журналах, включённых ВАК РФ в перечень ведущих рецензируемых научных журналов и изданий.

**Структура и объём работы.** Диссертация состоит из введения, четырёх разделов, заключения, списка использованной литературы (79 наименований) и приложений. Основная часть диссертации содержит 127 страниц машинописного текста, включая 48 рисунков, 8 таблиц.

## СОДЕРЖАНИЕ РАБОТЫ

Во *введении* обоснована актуальность темы диссертации, её научная новизна и практическая значимость, сформулирована цель работы.

В *первом разделе* проводится анализ современных подходов и методов к автоматизированному построению полных и непротиворечивых моделей проблемных областей, рассматриваются основные понятия и определения в области приобретения знаний из источников различной типологии, исследуются проблемы, характерные для этой области, проводится обзор и сравнительный анализ методов и программных средств приобретения знаний из источников различной типологии, вводится и исследуется понятие *распределённого приобретения знаний* (Рис. 1) применительно к интеграции информации, полученной из источников знаний различной типологии, с целью автоматизированного построения максимально полных и непротиворечивых моделей проблемных областей.

Значительное внимание в диссертации уделяется анализу отечественных и зарубежных программных средств поддержки процессов приобретения знаний, и построения БЗ как классических экспертных систем, так и более сложного и широкого класса - ИЭС, отличающиеся масштабируемой архитектурой, позволяющей расширять функциональность простых (продукционных) экспертных систем, предназначенных для поддержки решения неформализованных задач, с помощью дополнительных подсистем, реализующих широкий класс формализованных задач.

В диссертации показано, что несмотря на значительное число разработанных в настоящее время инструментальных средств приобретения знаний, только некоторые из них (SIMER + MIR, CLEMENTINE, КАТЕMES, инструментальный комплекс АТ-ТЕХНОЛОГИЯ) позволяют учитывать *совместное* использование источников знаний различной типологии (эксперты, проблемно-ориентированные тексты, электронные носители в виде БД).

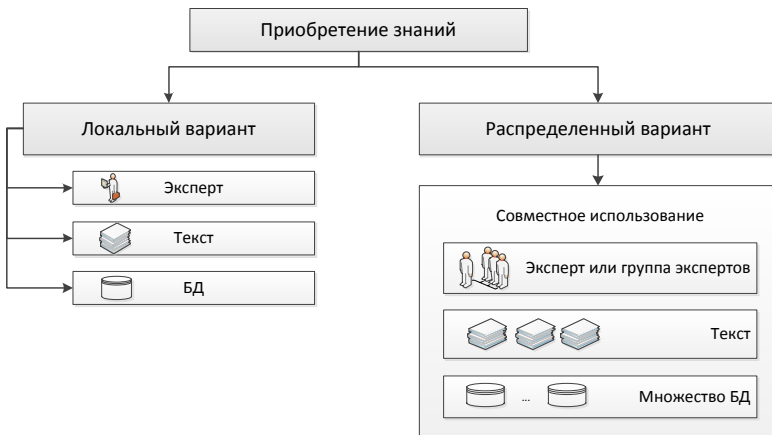


Рис. 1. Локальный и распределенный варианты приобретения знаний.

Кроме того, в диссертации показано, что как в отечественных, так и зарубежных исследованиях незначительное внимание уделяется решению проблем приобретения знаний, связанных с учетом географической распределенности источников знаний, а также возможных групп удаленных источников, например, отдельных групп экспертов. Проведен анализ некоторых подходов и методов группового приобретения знаний. На основе проведенного анализа выбран базисный метод группового приобретения знаний, основанный на вычислении вероятности принадлежности по мнению группы экспертов каждого рассматриваемого объекта определенному классу.

В результате проведенного анализа инструментальных средств приобретения знания показано, что среди отечественных систем приобретения знаний, использующих комбинацию различных методов и подходов к приобретению знаний, выделяется инструментальный комплекс АТ-ТЕХНОЛОГИЯ, созданный на основе ЗОМ построения прикладных интегрированных экспертных систем, предложенной Г.В. Рыбиной в середине 90-х годов. Неотъемлемой частью ЗОМ является ЗОМ приобретения знаний, представляющая собой совокупность комбинированного метода приобретения знаний (КМПЗ) и технологии его использования на различных стадиях жизненного цикла построения ИЭС и веб-ориентированные ИЭС (веб-ИЭС).

Следует отметить, что источники знаний 1-го и 2-го типа в контексте КМПЗ, реализованного в рамках ЗОМ построения ИЭС, рассматриваются как *совмещенные*, поскольку в КМПЗ существует совокупность хорошо апробированных технологических процедур, позволяющих дополнять информацию, полученную от эксперта/экспертов, за счет информации, выявленной из проблемно-ориентированных естественно-языковых текстов (ЕЯ-текстов) (в



данном случае – это обработка протоколов интервьюирования экспертов, сбор лексики инженера по знаниям, анализ сигнальных лексем во входных ЕЯ-текстах и др.).

Поэтому основное внимание в диссертации уделялось источникам знаний 3-его типа, т.е. БД, для чего исследовались технологии Data Mining и Knowledge Discovery in Databases (KDD).

В диссертации рассмотрены основные методы Data Mining, среди которых выделены методы, основанные на построении деревьев решений, обладающие такими достоинствами как хорошая интерпретируемость и способность к анализу больших объемов информации. Приведена классификация методов, основанных на построении деревьев решений, а также на основе проведенного анализа выбран базисный метод Data Mining, основанный на построении бинарных деревьев решений.

Показано, что одной из наиболее сложных и трудоёмких задач, затрудняющей приобретение знаний из источников различной типологии с учетом географической удаленности источников и наличия групп источников знаний, является задача объединения полученных продукционных правил в единую БЗ, для решения которой предложено использование *теории множеств* для представления *многопризнаковых объектов*, что значительно упрощает решение известных задач классификации, сортировки, ранжирования многопризнаковых объектов. Проведен анализ современных методов группировки и кластеризации продукционных правил, среди которых выделены методы, основанные на вычислении *мер близости* между утверждениями правил.

На основе проведенного в диссертации анализа показано, что в настоящее время не существует универсальных методов и методологий, позволяющих эффективно решать проблему автоматизированного приобретения знаний и учитывающих совместное использование источников знаний различной типологии, а также географическую распределенность источников и наличие возможных групп удаленных источников (например, отдельных групп экспертов).

В связи с этим делается вывод об актуальности темы диссертационного исследования, направленного на разработку эффективных методов и инструментальных программных средств распределенного приобретения знаний, повышающих степень автоматизации процессов построения БЗ для широкого класса прикладных ИЭС, в том числе проектируемых на основе ЗОМ.

Сформулирована цель и поставлены конкретные задачи диссертационного исследования.

Во *втором разделе* диссертации рассматриваются теоретические аспекты разработки методов и алгоритмов распределенного приобретения знаний.

В соответствии с концептуальными основами ЗОМ приобретения знаний в рамках базового КМПЗ и средств его реализации рассматривается так называемый *локальный вариант* приобретения знаний. Однако, при переходе к веб-версии комплекса АТ-ТЕХНОЛОГИЯ стал возможен другой вариант автоматизированного приобретения знаний на основе КМПЗ – *распределенный*, обеспечивающий в рамках клиент-серверной архитектуры, интеграцию всех рассмотренных выше типов источников знаний, учет их географической распределённости, а также возможность работы с группами удаленных источников знаний.

Теоретической основой базового КМПЗ является задача экспертной классификации *многомерных объектов*, которая формулируется следующим образом: пусть имеется эксперт, являющийся специалистом в рассматриваемой проблемной области; требуется на основе его знаний построить классификацию объектов. В основу подхода к решению задачи экспертной классификации положена гипотеза о различной степени характерности отдельных значений каждого признака для каждого свойства, что даёт возможность вынесения заключений по ряду состояний без их непосредственной идентификации экспертом, позволяя организовать *рациональную процедуру экспертного опроса*, целью которого является классификация всех возможных состояний при уменьшении числа вопросов к эксперту.

В рамках базового КМПЗ, ориентированного на поддержку процессов приобретения знаний на всех этапах жизненного цикла разработки ИЭС, а также на возможность итеративного пополнения БЗ ИЭС, для решения проблемы *неполноты* БЗ ИЭС предусмотрено локальное использование источников знаний различной типологии.

В общем виде теоретико-множественную модель расширенного для источников знаний различной типологии КМПЗ с учетом распределенного варианта приобретения знаний можно представить в виде:  $M_{км} = \langle N, S, F, K, Z \rangle$ , где  $N = (M_{мртз}, I, T, E)$ , где  $M_{мртз}$  – совокупность моделей решения типовых задач (МРТЗ) диагностики, проектирования, планирования, обучения, и управления;  $I = (I^{авт}, I^{групп})$  – знания, полученные от экспертов путем применения МРТЗ, где  $I^{авт}$  – знания, полученные от автономного эксперта, а  $I^{групп}$  – знания от групп экспертов;  $E$  – знания, приобретенные из БД;  $T = (Q, D, W)$  – знания, выявленные в результате лингвистической обработки ЕЯ-текстов, где  $Q$  – знания, получаемые в результате “понимания” ЕЯ-текстов, описывающих решаемую проблему;  $D = \{D_{ij}\}$  – знания, приобретенные от эксперта/экспертов;  $W$  – лексические знания, получаемые на этапе анализа системных требований пользователя. Компонент  $S = \{S_m\}$ ,  $m=1, \dots, m_m$  – есть множество *структурированных описаний проблемной области* в виде *поля знаний*, выполняющего важную функцию в процессе структурирования полученной от эксперта информации, обеспечивая единое внут-

реннее представление и унификацию основных понятий и отношений проблемной области, выявленных из различных источников знаний. Соответственно, с учетом особенностей распределенного приобретения знаний обобщенную модель поля знаний можно представить в виде:  $S_m = \langle IS_m, SS_m, O_m, R_m \rangle$ , где  $IS_m$  – порядковый номер структурированного описания проблемной области;  $SS_m = \langle T_{SS}, I_{SS} \rangle$  – источник, откуда получено описание, где  $T_{SS}$  – тип источника (эксперт, группа экспертов, БД, текст),  $I_{SS}$  – идентификатор источника знаний;  $O_m = \{O_{mj}\}$ ,  $j=1, \dots, n$  – множество объектов;  $R_m = \{R_{mk}\}$ ,  $k=1, \dots, p$  – множество правил.

Компонент  $K$  – множество процедур конвертации сформированного поля знаний в форматы языков представления знаний различных инструментальных средств (зарегистрированных в комплексе АТ-ТЕХНОЛОГИЯ), а  $Z$  – фрагменты БЗ в форматах языков представления знаний.

Компонент  $F$  определяет множество процедур внутреннего малого жизненного цикла формирования фрагментов БЗ на основе поля знаний. Разработанный в диссертации распределенный вариант приобретения знаний на основе КМПЗ позволяет строить более *полные и непротиворечивые* БЗ за счет использования как географически удаленных источников знаний, так и наличия групп источников знаний различной типологии. При переходе от локального варианта приобретения знаний к распределенному в диссертации решались такие задачи как *групповое приобретение знаний*, приобретение знаний из *множества БД*, а также автоматизация *объединения* множества полученных из источников разной типологии продукционных правил в единую БЗ. Таким образом, при переходе от локального варианта приобретения знаний к распределенному, компонент  $F$  представляется в виде:

$$F = F_{БАЗ} \cup F_{РАСПР},$$

где  $F_{БАЗ}$  – множество базовых процедур, а  $F_{РАСПР}$  – множество процедур, добавленных при переходе от локального варианта приобретения знаний к распределенному:  $F_{РАСПР} = \{F_{P1}, \dots, F_{P4}\}$ , где  $F_{P1}$  – процедуры группового приобретения знаний;  $F_{P2} = \{P_{P2-1}, P_{P2-2}, P_{P2-3}, P_{P2-4}, P_{P2-5}\}$  – процедуры построения хранилищ данных, где  $P_{P2-1}$  – процедура генерации SQL запроса;  $P_{P2-2}$  – процедура извлечения данных из БД в соответствии с запросом, сформированным процедурой  $P_{P2-1}$ ;  $P_{P2-3}$  – процедура фильтрации некоторого подмножества данных, которое будет использоваться для построения фрагмента поля знаний;  $P_{P2-4}$  – процедура выделения зависимых и независимых атрибутов в анализируемой выборке;  $P_{P2-5}$  – процедура конвертации данных;  $F_{P3} = \{P_{P3-1}, P_{P3-2}, P_{P3-3}, P_{P3-4}\}$  – процедуры распределенного приобретения знаний из БД, где  $P_{P3-1}$  – процедура построения деревьев решений,  $P_{P3-2}$  – процедура генерации наборов правил,  $P_{P3-3}$  – процедура упрощения и сокращения количества правил,  $P_{P3-4}$  – процедура конвертации полученных правил в формат поля знаний;  $F_{P4} = \{P_{P4-1}, P_{P4-2}, P_{P4-3}\}$  – процедуры объединения

фрагментов полей знаний, полученных из источников разной типологии, где  $P_{P4-1}$  – процедура объединения типов атрибутов,  $P_{P4-2}$  – процедура объединения объектов,  $P_{P4-3}$  – процедура объединения продукционных правил.

Перейдем к детальному рассмотрению предложенных и реализованных в диссертации методов и алгоритмов. Для решения задач группового приобретения знаний в диссертации был разработан метод группового приобретения знаний, суть которого заключается в следующем: рассматривается случай, когда в задаче классификации только два класса решений:  $C_1 = C$  и  $C_2 = \bar{C}$ , т.е. в класс  $C$  попадают объекты, обладающие, по мнению эксперта, свойством  $G$ , а в класс  $\bar{C}$  — не обладающие свойством  $G$ . На этих множествах решений с помощью эксперта строится отображение  $F : Y \rightarrow \{Y_1, Y_2\}$ , такое, что  $Y = Y_1 \cup Y_2$ ;  $Y_1 \cap Y_2 = \emptyset$  ( $Y_i$  — множество векторных оценок, принадлежащих классу  $C_i$ ), удовлетворяющее свойству непротиворечивости:  $\forall x, y \in Y: x \in Y_i, y \in Y_j, xPy \Rightarrow i \geq j$ . Функция  $f(x) : Y \rightarrow \{0, 1\}$  определяется так, что:

$$f(x) = \begin{cases} 1, & x \in Y_1 \\ 0, & x \in Y_2 \end{cases}$$

В случае, если рассматриваемую задачу решают  $Z$  экспертов, то в результате решения каждый из экспертов построит отображение  $F_i$  и соответствующую ему функцию  $f_i(x)$  ( $i = 1, \dots, Z$ ). Величина:

$$W(x) = \frac{\sum_{i=1}^Z f_i(x)}{Z}$$

является вероятностью принадлежности объекта  $x \in Y$  классу  $C_1$  по мнению группы экспертов. В рассматриваемом случае эта величина также называется *вероятностью* (по мнению группы экспертов) того, что объект  $x \in Y$  обладает свойством  $G$ .

Аналогичным образом определяется понятие вероятности (по мнению группы экспертов) в задачах классификации с более чем двумя классами решений. Однако, в этом случае вероятность вводится по отношению к набору классов решений, либо начиная/заканчивая некоторым классом. Разработанный для реализации предлагаемого метода алгоритм описывается следующим образом: на этапе экспертной классификации каждый эксперт решает упрощенные задачи, после чего один из них решает задачу в полном пространстве решений; для построенной БЗ выделяются границы классов решений; каждый из оставшихся экспертов начинает работу в полном пространстве решений с решения задач, лежащих на границах классов решений первого эксперта. Особое внимание было уделено случаю, когда отдельные сочетания значений признаков не могут быть реализованы на практике, и тогда

подобные объекты исключаются из проблемной области и не предъявляются эксперту. Классификация заканчивается если всем не исключенным объектам будет назначен класс. Завершающий этап методики — проверка границ классов, причем проверяется назначенный экспертом класс, а затем полученные границы классов преобразуются в продукционные правила. Таким образом, за счет реализации описанной методики создаётся достаточно полный и непротиворечивый фрагмент поля знаний.

Рассмотрим разработанные в диссертации методы и алгоритмы, реализующие распределенное приобретение знаний из множества БД, для чего в диссертации применялись технологии KDD и Data Mining, обеспечивающие выявление скрытых закономерностей в больших объемах информации. Следует отметить, что в соответствии с ЗОМ под KDD подразумевается *весь процесс* извлечения знаний, начиная от соединения с БД, заканчивая представлением полученных результатов, а Data Mining являешься лишь некоторым *этапом* общего процесса KDD.

В базовом КМПЗ для приобретения знаний из БД использовались два алгоритма построения деревьев решений ID3 и C4.5. При переходе к распределенному варианту КМПЗ предпочтение было отдано концепции алгоритма CART, позволяющего строить бинарные деревья решений, что более удобно при визуализации и возможной постобработке продукционных правил.

Предложенный в диссертации алгоритм распределенного приобретения знаний из БД ориентирован на работу с предварительно подготовленными данными, причем вне зависимости от модели БД, предварительно проделанных операций, выборка данных должна представляться в виде плоской таблицы и содержать следующую информацию: множество независимых атрибутов выборки  $\{A_1, \dots, A_n\}$ , а также один выделенный зависимый атрибут  $A^*$ , значение которого определяет класс, к которому принадлежит пример.

При переходе от локального варианта приобретения знаний к распределенному возникает необходимость приобретения знаний из множества БД, для чего в диссертации применялся подход, суть которого заключается в построении из каждой БД  $db_k$  одного хранилища данных  $s_k$  (Рис. 2).

Предложенный в диссертации алгоритм распределенного приобретения знаний из БД представляет собой алгоритм построения бинарных деревьев решений: в каждом узле дерева осуществляется разбиение множества атрибутов, ассоциированного с узлом, на два подмножества по следующему правилу - выбранный атрибут должен разбивать множество на подмножества так, чтобы количество объектов из других классов было минимальным. Для оценки качества разбиения использовался статистический критерий, основанный на индексе *Gini*.

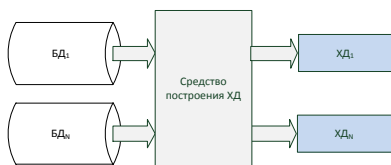


Рис. 2. Построение множества хранилищ данных из множества БД

Рассмотрим подробнее предложенный в диссертации алгоритм распределенного приобретения знаний из БД:

**Шаг 1.** Производится загрузка множества анализируемых выборок и выделение первой из них. Данные выборки записываются в матрицу и проводится разбиение матрицы на левую

(не выполняется правило разбиения) и правую (выполняется правило разбиения). Разбиение происходит по следующему алгоритму. Сортируются все столбцы матрицы  $S$  по убыванию, кроме первого и последнего, по последнему столбцу определяется принадлежность к классу, а по первому - название правила разбиения. Выбирается второй столбец, формируются все правила разбиения. Для каждого правила матрица разбивается на левую и правую (левый и правый лист), подсчитывается значение Gini. Если произведено разбиение по всем правилам, то выбирается следующий столбец. Определяется правило разбиения соответствующее максимальному значению  $\tilde{G}_{split}$ . Производится разбиение по выбранному правилу, получившееся разбиение считается лучшим.

**Шаг 2.** Если примеры на некотором листе принадлежат одному классу, то лист помечается именем класса. Если не все листы дерева помечены именами классов, то производится проверка, все ли примеры в листе принадлежат одному классу, если нет, то проводится дальнейшее разбиение.

**Шаг 3.** Если все листы помечены именами классов, то дерево считается построенным, производится генерация правил и, в случае наличия других анализируемых выборок, строится следующее дерево решений.

**Шаг 4.** По окончании построения всех деревьев решений путем обхода каждого дерева генерируется набор продукционных правил: каждый путь от корня до листа дерева даёт одно правило, а условиями правила являются проверки из узлов, принадлежащих пути.

**Шаг 5.** Упрощение и сокращение числа выведенных правил для каждого полученного фрагмента поля знаний.

Важно отметить, что по завершении работы алгоритма распределенного приобретения знаний из БД генерируется большое количество простых правил, которые могут быть преобразованы в одно или несколько более сложных, но в то же время более наглядных и удобных для использования правил. Для этого в рассматриваемом алгоритме применена постобработка выведенных правил, направленная на сокращение их количества. Способы упрощения правил можно разделить на следующие группы:

Способ 1. Некоторый набор выведенных правил имеет одинаковое заключение, тогда посылка нового правила является дизъюнкцией посылок рассматриваемого набора правил.

Способ 2. Некоторый набор выведенных правил имеет одинаковое заключение, тогда посылка нового правила является конъюнкцией посылок рассматриваемого набора правил, причем результирующее правило является более сильным, чем исходные (посылки первоначального набора правил могут быть взаимоисключающими, в таком случае выполнение преобразования не возможно).

Способ 3. Некоторый набор выведенных правил имеет одинаковые посылки, тогда заключение нового правила является дизъюнкцией заключений рассматриваемого набора правил (способ носит теоретический характер, т.к. при решении задач классификации необходимо находить условия, четко относящие пример к тому или иному классу).

Способ 4. Некоторый набор выведенных правил имеет одинаковое заключение, и в посылках правил рассматриваются числовые атрибуты, тогда возможно объединение условий, налагаемых на числовые атрибуты (способ требует дополнительных знаний, в частности, о дискретных значениях атрибутов).

Перечисленные способы в некоторых случаях позволили существенно уменьшить число сгенерированных изначально правил.

Решение задачи объединения фрагментов полей знаний, полученных из источников знаний различной типологии, существенно затрудняется задачей *объединения объектов*, описанных в каждом фрагменте поля знаний. Для решения этой проблемы в диссертации предложен метод, основанный на теории мультимножеств. Рассмотрим совокупность  $A = \{A_1, \dots, A_k\}$  объектов, которые описываются  $m$  атрибутами  $Q_1, \dots, Q_m$ , имеющими конечное число значений. Также рассмотрим вариант, когда объект  $A_i$  получен из  $k$  независимых источников и описывается  $m$  атрибутами. Очевидно, что объект  $A_i$  представляется в  $m$ -мерном пространстве  $Q$  уже не одной точкой  $q_i$ , а группой (“облаком”), состоящим из  $k$  точек. В таком случае удобно представить объект в виде мультимножества: пусть количество фрагментов полей знаний равно  $k$ , суммарное количество всех уникальных атрибутов равно  $m$ ; введем обобщенную шкалу признаков – множество  $G = \{Q_1, \dots, Q_m\}$ , состоящую из  $m$  групп признаков, и представим объект  $A_i$  в символическом виде:

$$A_i = \{k_{A_i}(q_1^1) \circ q_1^1, \dots, k_{A_i}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{A_i}(q_m^1) \circ q_m^1, \dots, k_{A_i}(q_m^{h_m}) \circ q_m^{h_m}\},$$

где число  $k_{A_i}(q_s^{es})$  указывает, сколько раз атрибут  $q_s^{es} \in Q_S$  встречается в описании объекта  $A_i$ , знак  $\circ$  обозначает кратность вхождения атрибута  $q_s^{es}$ . Например, при получении объекта  $A_i$  из нескольких источников число  $k_{A_i}(q_s^{es})$  равно числу источников, в которых объекту  $A_i$  присвоен атрибут  $q_s^{es}$  по критерию  $Q_S$ .

Объекты группируются в мультимножества по именам, либо синонимам имен. Объекты возможно объединить, либо заменить объект  $A_m$  объектом  $A_k$  в случае совпадения имен объектов, совпадения синонима имени объекта  $A_m$  с именем объекта  $A_k$ , а также в случае совпадения синонимов имён объектов. В случае, если объединение объектов по словарю синонимов не может гарантировать приемлемых результатов необходимо воспользоваться подсчетом расстояния между мультимножествами. Рассмотрим *метрическое пространство* между двумя объектами, представленными в виде мультимножеств:  $d_{1p}(A,B) = [m(A\Delta B)]^{1/p}$ , где  $p$  – целое число,  $m$  – мера мультимножества, действительная неотрицательная функция, заданная на алгебре мультимножеств  $L(Z)$ , причем  $m(A) = \sum_i w_i k_A(x_i)$ ,  $w_i > 0$ . В таком случае  $d_{1p}(A,B)$  можно представить в виде:

$$d_{1p}(A,B) = \left( \sum_{x_i \in G} w_i |k_A(x_i) - k_B(x_i)| \right)^{1/p}.$$

Метрика  $d_{1p}(A,B)$  аналогична *мере схожести Хемминга* для мультимножеств. Для объединения нескольких фрагментов полей знаний в один необходимо привести множество многомерных точек, образующих мультимножество к единственной многомерной точке. В случае, если сделать это невозможно, каждая точка выделяется в виде отдельного многомерного объекта. На Рис. 3, на примере двухмерного пространства, приведена последовательность преобразований мультимножеств  $A$  и  $B$ , а также разбиение мультимножества  $B$  на две многомерных точки.

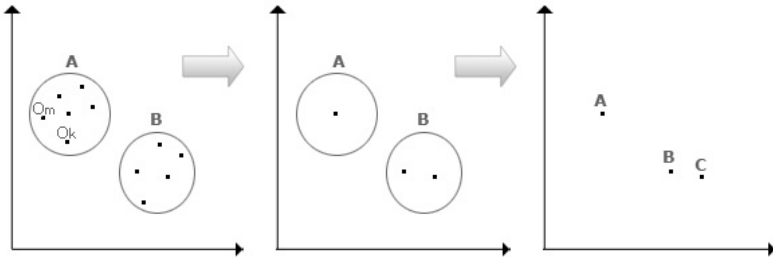


Рис. 3. Приведение множеств многомерных точек к единой точке и разбиение мультимножества на множество многомерных точек.

Для подсчета расстояния между каждой парой точек  $O_m$  и  $O_k$  внутри мультимножества применяется квадрат геометрического расстояния:

$$(d_{O_m O_k})^2 = \sum_{i=1}^n (x_{O_m i} - x_{O_k i})^2,$$

значение которого является количеством атрибутов, различных для объектов.



Обозначим  $A_m$  – не пустое множество атрибутов объекта  $O_m$ , а  $A_k$  – не пустое множество атрибутов объекта  $O_k$ . Объекты  $O_m$  и  $O_k$  имеют одинаковые имена, либо совпадающие синонимы имён. Рассмотрим состояния:

- $(d_{O_k O_m})^2 = 0$ , т.е.  $A_k = A_m$ , в этом случае будем называть объекты  $A_k$  и  $A_m$  идентичными, возможно их автоматическое объединение;
- $(d_{O_k O_m})^2 = |A_k| + |A_m|$  и  $A_k \cap A_m = \emptyset$ , в этом случае, либо допущена ошибка при сопоставлении имен и синонимов имен объектов, либо имена объектов полностью идентичны, но объекты различны. Для устранения данного совпадения необходимо уточнение экспертом, либо инженером по знаниям имени объекта;
- $(d_{O_k O_m})^2 > 0$  и  $A_k \neq A_m$ , в этом случае возможны следующие состояния:
  - $A_k \subset A_m$ , при данном условии возможна автоматическая замена объекта  $A_k$  объектом  $A_m$ , т.к. объект  $A_m$  является более детально описанным;
  - $A_k \not\subset A_m$  и  $A_k \cap A_m \neq \emptyset$  при данных условиях необходима ручная обработка объектов экспертом, либо инженером по знаниям. Возможно как объединение объектов с частичным или полным объединением атрибутов, так и замена выбранных атрибутов, либо декомпозиция объектов.

Следующим после завершения объединения объектов этапом объединения фрагментов полей знаний, является этап объединения наборов правил. Для объединения наборов правил в диссертации использовались *расширенные таблицы решений*, представляющие собой набор строк и столбцов, где каждая ячейка строки таблицы хранит данные о вхождении и параметрах вхождения *утверждения*, характеризующегося заголовком строки, в конкретное *правило*. По мере рассмотрения правил, входящих в состав поля знаний, расширенная таблица решений пополняется новыми строками, однозначно идентифицирующимися парой «объект - атрибут объекта». Правила представляются в расширенной таблице решений ее столбцами. В каждую ячейку расширенной таблицы решений записывается «тип» утверждения, принимающий следующие значения: 0 - утверждение отсутствует в правиле; 1 – утверждение присутствует в правиле.

На основе анализа расширенной таблицы решений осуществляется построение *таблицы мер схожести* правил (Рис. 4) путем подсчета числа совпадающих атрибутов, участвующих в правилах  $R_i$  и  $R_k$ .

	$R_1$		...	$R_N$	
$R_1$	$\mu_{11}^{IR} (1)$	$\mu_{11}^{IL} (1)$	...	$\mu_{N1}^{NR}$	$\mu_{N1}^{NL}$
...	...		...	...	
$R_N$	$\mu_{1N}^{IR}$	$\mu_{1N}^{IL}$	...	$\mu_{NN}^{NR} (1)$	$\mu_{NN}^{NL} (1)$

Рис. 4. Общий вид таблицы схожести правил.

Для левой и правой частей правил подсчитывается мера сходства Хем-

минга ( $\mu_{ik}^{NL}$  и  $\mu_{ik}^{NR}$ ):  $\mu_{ik}^N = n_{ik}/N$ , где  $n_{ik}$  – есть число совпадающих признаков у образцов  $R_i$  и  $R_k$ ,  $\mu_{ik}^{NL}$  – есть отношение количества совпавших атрибутов посылок правил  $R_i$  и  $R_k$  к количеству всех атрибутов, участвующих в заключениях правил. На пересечении каждого столбца и строки таблицы мер схожести имеется две ячейки, предназначенные для хранения мер схожести посылок и заключений правил.

Для определения последовательности вывода правил устанавливается *контрольная зона* мер схожести. В случае попадания текущих меры схожести *посылки* и меры схожести *заключения* в заданный интервал, пары правил, образующие пересечение столбца и строки таблицы мер схожести могут быть выведены для дальнейшего анализа инженером по знаниям. Затем полученный фрагмент поля знаний подвергается завершающей обработке – все объекты и правила собираются в единый XML файл и проходят финальную перенумерацию и верификацию.

**В третьем разделе** диссертации рассматриваются вопросы реализации в комплексе АТ-ТЕХНОЛОГИЯ инструментальных программных средств распределенного приобретения знаний из БД. В процессе разработки были учтены требования последних версий комплекса АТ-ТЕХНОЛОГИЯ, основным из которых является реализация программных средств в виде СОМ-объектов, выполняющих установленные правила обмена сообщениями с остальными компонентами комплекса, а также содержащие функции, обеспечивающие связь с типовыми проектными процедурами, которые активизируются специальными средствами интеллектуальной поддержки разработки ИЭС.

Для поддержки приобретения знаний из источников знаний различной типологии применяется типовая проектная процедура «Приобретение знаний из БД», на Рис. 5 приведена диаграмма последовательности действий инженера по знаниям при её реализации. На первом этапе реализации типовой проектной процедуры «Приобретение знаний из БД» инженером по знаниям производится выбор зарегистрированных БД. При помощи средств поддержки проектирования хранилищ данных производится формирования набора хранилищ данных, являющихся анализируемыми выборками для алгоритма распределенного приобретения знаний из БД. Вторым этапом является настройка алгоритма распределенного приобретения знаний из БД и генерация с его помощью набора фрагментов полей знаний, полученных из каждой зарегистрированной БД. Третий этап - объединение всех фрагментов полей знаний, полученных из источников знаний различной типологии. Основными шагами этапа являются загрузка и объединение объектов, построение расширенной таблицы решений и таблицы схожести правил. На этапе объединения фрагментов полей знаний экспертом задаются контрольные зоны и величины совпадения числовых атрибутов, производится объединение объектов, типов

атрибутов объектов и объединение правил. Четвертым (завершающим) этапом является верификация итогового фрагмента поля знаний.



Рис. 5. Последовательность действий инженера по знаниям при реализации типовой проектной процедуры «Приобретение знаний из БД».

Общая архитектура разработанных инструментальных средств поддержки типовой проектной процедуры «Приобретение знаний из БД» представлена на Рис. 6, в соответствии с чем основными средствами поддержки процессов распределенного приобретения знаний являются следующие компоненты: компонент диалога с экспертом с целью выяснения требуемых ему параметров, компонент поддержки построения структур БД, компонент сопоставления объектов хранилищ данных объектам поля знаний, компонент заполнения хранилищ данных, компонент загрузки фрагментов полей знаний, компонент построения расширенной таблицы решений и таблицы мер схожести, компонент объединения объектов и правил, компонент анализа расширенной таблицы решений, таблицы мер схожести и итогового фрагмента поля знаний, компонент формирования отчетов.

Базовые средства распределенного приобретения знаний реализованы как отдельные компоненты, поскольку активизация каждого из них определяется различными условиями.

При выполнении задач, соответствующих типовой проектной процедуре «Приобретение знаний из БД», поле знаний обрабатывается с помощью компонента формирования расширенной таблицы решений, который обеспечивает нормализацию правил и заполнение расширенной таблицы решений.

Расширенная таблица решений служит в качестве исходных данных, на основе заключенной в ней информации, при помощи компонента построения таблицы мер схожести рассчитывается мера сходства Хемминга для посылок и заключений пар правил.

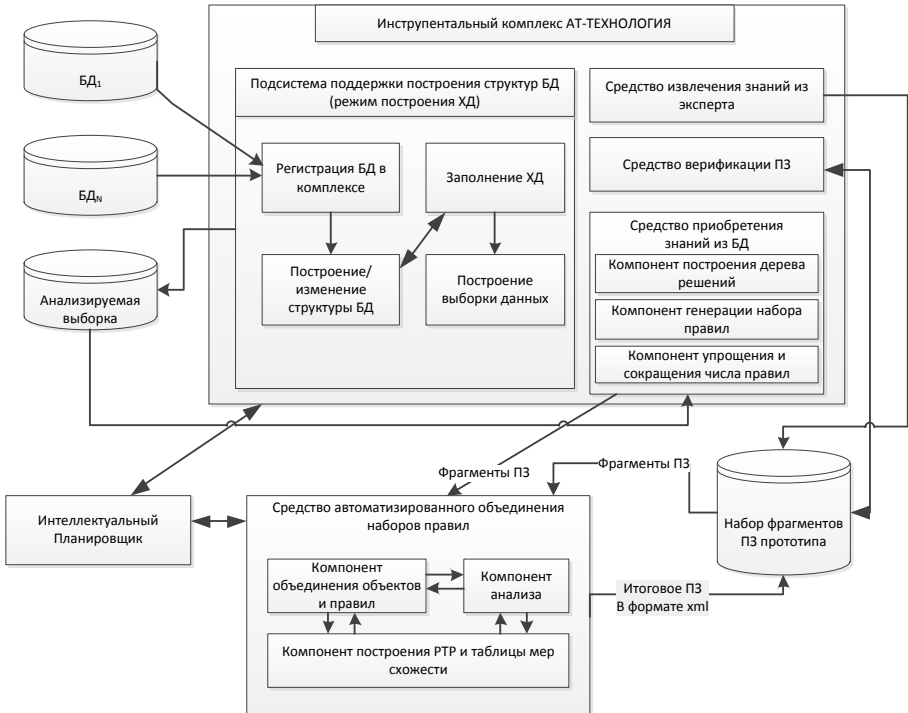


Рис. 6. Архитектура средств поддержки типовой проектной процедуры «Приобретение знаний из БД».

Компонент объединения объектов и правил является визуальным компонентом, предоставляющим эксперту возможность указать на равенство равноименных атрибутов различных правил. Приравненным атрибутам и объектам присваивается имя одного из них. Присваиваемое имя определяется экспертом. Компонент анализа построенных таблиц и итогового фрагмента поля знаний также является визуальным компонентом, предоставляющим эксперту возможность создания и упорядочения контрольных зон, определения действий над правилами, определения порядка вывода правил по контрольным зонам. Ускорение процесса построения БЗ обеспечивается за счет частичной автоматизации этапов сравнения и объединения объектов и наборов правил.

**Четвертый раздел** диссертации посвящен экспериментальной проверке разработанных программных средств распределенного приобретения знаний из БД. Экспериментальная проверка выполнялась на БД, полученных при разработке нескольких приложений, связанных с задачами проектирования, медицинской диагностики, определения географического положения IP-адреса, контроля радиационных дозовых нагрузок персонала атомных электростанций РФ и др. Кроме того, эксперименты проводились с различными специально подготовленными тестовыми БД и тестовыми анализируемыми выборками, опубликованными Center for Machine Learning and Intelligent Systems Университета Калифорнии.

В качестве анализируемых БД использовались фрагменты БД географического местонахождения IP-адресов и БД по диагностике заболеваний дыхательных путей. Полученная в первом случае анализируемая выборка включала 1000 записей и 12 полей. Итоговые наборы правил и скорость работы разработанных программных средств приобретения знаний из БД сравнивалась с реализацией алгоритма CART средства Answer Tree компании SPSS, Inc. При применении алгоритма распределенного приобретения знаний из БД, реализованного в рамках диссертации из БД местонахождения IP-адресов было выведено 417 правил, а при применении алгоритма CART средства Answer Tree было получено 439 правил (Рис. 7а.). Анализируемая выборка, полученная из БД по диагностике заболеваний дыхательных путей, содержала 500 записей. Из данной выборки было выведено 236 правил при применении реализованного в диссертации алгоритма распределенного приобретения знаний из БД и 238 правил при применении реализации алгоритма CART средства Answer Tree (Рис. 7б.).

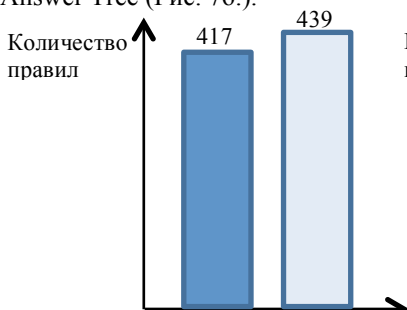


Рис. 7а. Количество правил, выведенных из БД IP-адресов.

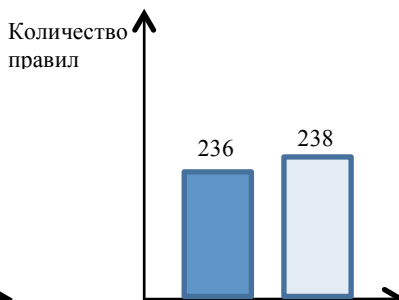


Рис. 7б. Количество правил, выведенных из БД по диагностике заболеваний дыхательных путей.

Разница в количестве выведенных правил обусловлена наличием этапа постобработки в реализованном в диссертации алгоритме распределенного приобретения знаний из БД, направленного на упрощение и сокращение числа выведенных правил.

Тестирование разработанных программных средств приобретения знаний из БД показало (Рис. 8), что при увеличении количества записей в анализируемой выборке в 2 раза изменение времени обработки незначительно превышает 2 раза, что связано с различным числом полей в различных анализируемых выборках, а также с наличием этапа постобработки выведенных наборов правил.

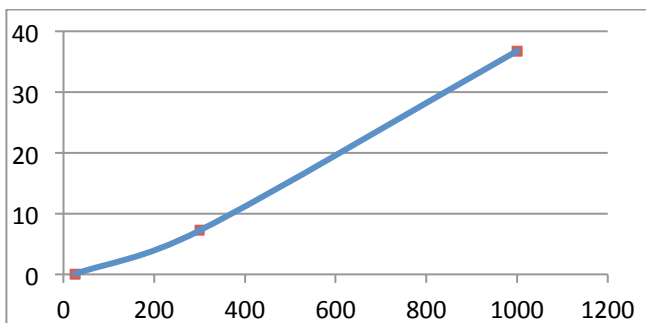


Рис. 8. График зависимости времени, затраченного на приобретение знаний от объема контрольной выборки

Проверка работы предложенного алгоритма распределенного приобретения знаний из БД оценивалась на фрагментах БД по дозовым нагрузкам сотрудников атомных электростанций РФ. Для работы были использованы фрагменты 10 БД атомных электростанций (Балаковской, Белоярской, Билибинской, Калининской, Кольской, Курской, Ленинградской, Нововоронежской, Ростовской и Смоленской). Объем контрольных выборок, полученных из каждой БД составил от 2500 до 4000 записей.

На Рис. 9 приведена диаграмма зависимости времени обработки каждой контрольной выборки (в секундах) от её объема. Суммарное время работы алгоритма распределенного приобретения знаний на десяти контрольных выборках (общий объем: 35072 записи) составило 126.2 секунды. Объединение всех десяти фрагментов полей знаний, полученных из различных БД было осуществлено в полностью автоматическом режиме, и заняло менее 10 минут. Суммарное количество правил, выведенных из десяти контрольных выборок составило 10350 продукционных правил. Объем итогового поля знаний составил 1674 правила. Кроме того, в диссертации были проведены эксперименты, в ходе которых были объединены два фрагмента полей знаний: фрагмент, полученный путем интервьюирования экспертов и содержит 340 правил и фрагмент, полученный путем приобретения знаний из БД и содержит 236 правил.

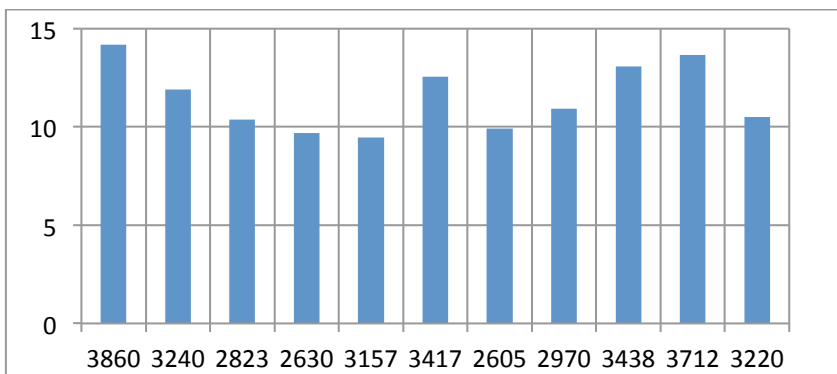


Рис. 9. Время обработки каждой контрольной выборки в зависимости от её объёма.

При объединении фрагментов поле знаний было пополнено на 42 правила, что соответствует приросту объёма фрагмента поля знаний приблизительно на 12%.

В *заключении* приводятся основные выводы и результаты диссертации.

В *приложении* вынесены: акты об использовании результатов диссертационного исследования; примеры фрагментов БД, использовавшихся для экспериментальной проверки разработанных инструментальных средств распределенного приобретения знаний из БД; примеры фрагментов полей знаний и БЗ, использованные для экспериментальной проверки разработанных инструментальных средств распределенного приобретения знаний из БД; тестовые примеры работы средств распределенного приобретения знаний из БД.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

- 1) На основе анализа существующих моделей, методов и средств приобретения знаний выбран базисный метод автоматизированного приобретения знаний, на основе которого предложен новый подход к интеграции информации, полученной из источников знаний различной типологии, и введено понятие «распределённого приобретения» знаний с целью автоматизированного построения максимально полных и непротиворечивых моделей проблемных областей.
- 2) На основе выполненного анализа существующих моделей, методов и средств приобретения знаний построена модель расширенного для источников знаний различной типологии комбинированного метода приобретения знаний.

- 3) Разработан метод и алгоритм группового приобретения знаний, основанный на вычислении вероятности принадлежности по мнению группы экспертов каждого рассматриваемого объекта определенному классу.
- 4) Разработан оригинальный алгоритм распределенного приобретения знаний из БД, основанный на построении бинарных деревьев решений, существенным отличием которого от базового алгоритма CART является ориентация на приобретение знаний из множества БД и оптимизация структуры и количества выведенных продукционных правил.
- 5) Разработан метод и алгоритмы объединения продукционных правил, полученных из источников знаний различной типологии, основанный на применении теории мультимножеств, расширенных таблиц решений и таблиц мер схожести правил, строящихся путем вычисления мер схожести Хемминга для посылок и заключений правил.
- 6) Разработаны инструментальные программные средства приобретения знаний, распределенных по источникам знаний различной типологии, и средства объединения продукционных правил, полученных из различных источников, включенные в состав комплекса АТ-ТЕХНОЛОГИЯ третьего поколения.
- 7) Проведена экспериментальная апробация предложенных алгоритмов и разработанных инструментальных программных средств на задачах медицинской диагностики и задачах контроля радиационных дозовых нагрузок персонала атомных электростанций РФ.

## **ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### **Публикации в журналах, входящих в перечень ВАК:**

- 1) Рыбина Г.В., Дейнеко А.О. Распределенное приобретение знаний для автоматизированного построения интегрированных экспертных систем // Искусственный интеллект и принятие решений, 2010. №4. С. 56-62.
- 2) Рыбина Г.В., Дейнеко А.О., Нистратов О.В. Автоматизированное построение полных и непротиворечивых баз знаний в интегрированных экспертных системах // Приборы и системы. Управление. Контроль. Диагностика. 2010. №5. С. 21-24.
- 3) Рыбина Г.В., Нистратов О.В., Дейнеко А.О. Особенности автоматизации и интеллектуализации процессов подготовки и переподготовки персонала на промышленных предприятиях // Промышленные АСУ и контроллеры, 2011. №. 3. С. 10-14

### **Работы, опубликованные в сборниках трудов конференций:**

- 1) Дейнеко А.О., Никулин Г.Н. Прототип динамической интеллектуальной системы управления ядерным реактором «СУЗ РБМК-1000» // Научная сес-



сия МИФИ-2007. Интеллектуальные системы и технологии. Т.3. – М.: МИФИ, 2007. – С. 210-211.

2) Рыбина Г.В., Дейнеко А.О., Нистратов О.В. Особенности построения полных и непротиворечивых баз знаний в интегрированных экспертных системах // Интегрированные модели, мягкие вычисления, вероятностные системы и комплексы программ в искусственном интеллекте. Сборник научных трудов V-й международной научно-практической конференции. Т2. – М.: Физматлит, 2009 – С. 760-768.

3) Рыбина Г.В., Дейнеко А.О. Распределенное приобретение знаний для автоматизированного построения баз знаний интегрированных экспертных систем // Двенадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2010 (20 - 24 сентября 2010 г., г. Тверь, Россия): Труды конференции. Т2. – М.: Физматлит, 2010 – С. 240-247.

4) Рыбина Г.В., Дейнеко А.О. Интеллектуальная технология распределенного приобретения знаний для автоматизированного построения баз знаний интегрированных экспертных систем // Реинжиниринг бизнес-процессов на основе современных информационных технологий. Системы управления процессами и знаниями (РБС-СУЗ-2010). Сборник научных трудов. М.: МЭСИ, 2010 – С. 260-264.

5) Дейнеко А.О., Нистратов О.В. Об одном подходе к преодолению неполноты баз знаний в интегрированных экспертных системах. // Научная сессия НИЯУ МИФИ-2010. XIII Международная телекоммуникационная конференция студентов и молодых ученых «Молодежь и наука». Тезисы докладов в 3-х частях. Ч.2. М.: НИЯУ МИФИ, 2010. – С. 101-102.

6) Дейнеко А.О., Рыбина Г.В. Распределенное приобретение знаний для автоматизированного построения баз знаний интегрированных экспертных систем // Современные технологии в задачах управления, автоматики и обработки информации: Труды XIX Международного научно-технического семинара. Алушта, сентябрь 2010 г. – М.: МИРЭА, 2010. – С. 125-126.

7) Дейнеко А.О. Распределенное приобретение знаний для автоматизированного построения интегрированных экспертных систем // Научная сессия НИЯУ «МИФИ»-2011. XIV Международная телекоммуникационная конференция студентов и молодых ученых «Молодежь и наука». Тезисы докладов в 3-х частях. Ч.2. М.: НИЯУ МИФИ, 2011. – С. 93-94.