На правах рукописи

Бездушный Алексей Анатольевич

Математическая модель интеграции данных на основе дескриптивной логики

Специальность 05.13.18 – математическое моделирование, численные методы и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени кандидата физико-математических наук

Москва – 2008

Работа выполнена на кафедре
математического моделирования сложных процессов и систем
Московского физико-технического института
(государственного университета)

Научный руководитель:

доктор физико-математических наук,
профессор
Серебряков Владимир Алексеевич

Официальные оппоненты:

доктор физико-математических наук,
профессор,
член-корреспондент РАН
Флёров Юрий Арсениевич

доктор технических наук,
профессор
Кузнецов Сергей Дмитриевич

Ведущая организация:
Новосибирский государственный университет

Защита состоится «19» декабря 2008 года в 10 час. 30 мин. на заседании диссертационного совета Д 212.156.05 в Московском физико-техническом институте (государственном университете) по адресу: 141700, г. Долгопрудный Московской обл., Институтский пер., д. 9, ауд. 903 КПМ.

С диссертацией можно ознакомиться в библиотеке Московского физико-технического института (государственного университета).

Автореферат разослан «18» ноября 2008 г.

Ученый секретарь
диссертационного совета Федько О.С.

Общая характеристика работы

Актуальность темы

Интеграция данных является одним из наиболее востребованных направлений в современной информационной индустрии. Интенсивное развитие информационных технологий и сети Интернет привело к накоплению огромных объемов данных в различных источниках, разнородных, автономно разработанных, представляющих информацию различными способами, содержащих взаимосвязанные и взаимно противоречивые сведения. Интеграция и совместное использование информации из множества таких источников данных является сложной задачей, остающейся неизменно актуальной на протяжении последних десятилетий.

Интеграция данных необходима для крупных организаций, в которых информация разбросана по различным специализированным системам, построенным в разное время и для разных целей, для повышения эффективности внутриведомственного и межведомственного взаимодействия государственных органов, для предоставления более качественных поисковых сервисов в сети Интернет, обеспечивающих получение согласованной информации из множества структурированных источников данных. Интеграция данных играет сегодня ключевую роль и для научной деятельности. В настоящее время всё большие объемы научной информации становятся в том или ином виде доступны в сети Интернет. В то же время, возможности существующих поисковых систем общего назначения не позволяют обеспечить эффективный поик научной информации, что ставит вопрос о разработке специализированных поисковых систем, интегрирующих интересующие научных сотрудников сведения.

Задача интеграции данных в настоящее время в той или иной степени исследована для различных условий, преимущественно, в контексте реляционных баз данных. В то же время предложенные подходы к решению задачи имеют недостатки и ограничения, и многие актуальные вопросы остаются открытыми.

Актуальным направлением исследований в этой области является применение к задаче интеграции данных аппарата дескриптивной логики, прежде всего в контексте Семантического Веб (Semantic Web).

Технологии Семантического Веб являются молодым и перспективным направлением развития современной информационной индустрии. Утвержденные World Wide Web Консорциумом (W3C) в 2004 году модель описания информационных ресурсов RDF (Resource Description Framework) и язык веб-онтологий OWL (Web Ontology Language) определили стандартный способ семантически богатого описания распределенной в сети Интернет информации. В этой связи представляется целесообразным

рассматривать их применение в контексте современных систем интеграции распределенных данных.

Формальной основой языка веб-онтологий OWL является так называемая дескриптивная логика – математический аппарат, предназначенный для представления терминологического знания о предметной области. Применение в системе интеграции данных аппарата дескриптивной логики вместо реляционной модели данных позволяет существенно расширить выразительные возможности системы. Онтологии позволяют специфицировать структуру и семантику терминов системы интеграции данных и информационных источников, выразить различные формы сложных ограничений целостности в системе интеграции данных, правила логического вывода.

Ключевой проблемой при рассмотрении задачи интеграции данных в контексте дескриптивной логики является ее трудноразрешимость или неразрешимость для достаточно выразительных диалектов дескриптивной логики. В то же время на практике важно сочетать выразительные возможности выбранного диалекта дескриптивной логики с эффективной работой с большими объемами данных.

В данной работе рассмотрен вопрос построения систем интеграции данных с применением аппарата дескриптивной логики и предложен выбор диалекта дескриптивной логики, который целесообразно использовать при интеграции больших объемов данных, хранимых в реляционных базах данных. Рассмотрен метод вычисления ответа на запрос к такой системе интеграции данных, предполагающий предварительную переформулировку исходного запроса, и предложен алгоритм переформулировки запроса для выбранного диалекта дескриптивной логики.

Таким образом, работа посвящена актуальной задаче интеграции данных с применением дескриптивной логики и технологий Семантического Веб, а предложенные в ней математическая модель, методы и алгоритмы формируют прочный фундамент для построения таких систем интеграции данных на практике.

Цель работы

Целью работы является разработка математической модели системы интеграции данных, основанной на применении аппарата дескриптивной логики, и исследование методов вычисления ответа на запрос к такой системе при условии интеграции больших объемов данных.

В работе исследованы и решены следующие задачи:

1. Проведено сопоставление выразительных возможностей и вычислительных характеристик различных диалектов дескриптивной логики.
2. Предложена методика интеграции данных, основанная на применении аппарата дескриптивной логики, разработана математическая модель системы интеграции данных на основе онтологий, формали-

зованы понятия ответа на запрос и переформулировки запроса в такой системе интеграции данных.

3. Предложен и обоснован выбор максимального, в определенном смысле, диалекта дескриптивной логики, для которого возможна эффективная интеграция больших объемов данных.
4. Предложен и обоснован алгоритм построения точной переформулировки запроса для выбранного класса систем интеграции данных на основе онтологий.
5. Разработан прототип системы исполнения распределенных запросов в среде Единого Научного Информационного Пространства РАН (ЕНИП РАН).

Научная новизна

В работе рассмотрен перспективный класс систем интеграции данных, отличительной особенностью которого является применение аппарата дескриптивной логики для более гибкого описания семантической взаимосвязи терминов, ограничений целостности, правил логического вывода.

В отличие от предшествующих работ по интеграции данных, полученный в данной работе результат имеет следующие особенности:

1. В основу рассматриваемого класса систем интеграции данных положен мощный математический аппарат дескриптивной логики, что является ключевым отличием от большинства смежных работ, рассматривающих интеграцию данных на основе реляционной модели данных и других семантически более бедных моделей данных.
2. В работе предложена оригинальная математическая модель системы интеграции данных, основанная на аппарате дескриптивной логики.
3. В работе рассматриваются выразительные системы интеграции данных, в которых отображения онтологий задаются парами конъюнктивных запросов с ограничениями, несмотря на допущение в онтологиях достаточно сложных ограничений целостности. Более того, показывается, что рассматриваемые системы в определенном смысле обладают максимально допустимыми выразительными возможностями для эффективного использования на практике. В предшествующих работах, посвященных применению дескриптивной логики к задаче интеграции данных, рассматривались существенно более ограниченные по выразительным возможностям отображения, позволяющие устранить меньшее число семантических конфликтов между информационными источниками. Таким образом, полученные в работе результаты представляют собой существенный шаг вперед по расширению систем интеграции данных аппаратом дескриптивной логики.
4. Для выбранных условий задачи предложен алгоритм переформулировки запросов в системе интеграции данных на основе онтологий,

представляющий собой новый существенный вклад в технологии интеграции данных, а также позволяющий непосредственно использовать полученный результат для практических задач.

Кроме того, разработан прототип системы исполнения распределенных запросов в среде Единого Научного Информационного Пространства РАН (ЕНИП РАН), позволяющий обеспечить виртуальную интеграцию данных различных научных учреждений в ЕНИП. Такая система позволяет расширить ЕНИП новым сервисом ответа на поисковые запросы с учетом разнородности информационных источников ЕНИП, при этом, в отличие от предшествующей реализации поисковых сервисов ЕНИП, не требуется предварительной репликации или индексации информации из источников.

Практическая ценность

Непосредственное применение полученные в работе теоретические результаты нашли в проекте «Единое Научное Информационное Пространство РАН» (ЕНИП РАН). Работа расширяет полученные ранее результаты по ЕНИП новыми функциональными возможностями. Предложенные в диссертационной работе математическая модель системы интеграции данных на основе онтологий и практический алгоритм переформулировки запросов в такой системе представляют собой фундамент для виртуальной интеграции данных различных научных учреждений в рамках ЕНИП.

На основе полученных в диссертационной работе теоретических результатов разработан прототип системы исполнения распределенных запросов в среде ЕНИП. Такой поисковый сервис позволяет динамически получать ответы на поисковые запросы, выраженные в терминах OWL онтологий ЕНИП. При исполнении запроса в системе обеспечивается соединение сведений из релевантных информационных источников ЕНИП, и на основе таких сведений формируется интегрированный согласованный ответ. При этом система позволяет преодолеть семантическую разнородность информационных источников, то есть, различие схем данных (онтологий) источников. В отличие от предшествующей реализации поисковых сервисов ЕНИП, не требуется предварительной репликации или индексации сведений из информационных источников – вычисляемый системой ответ включает исключительно актуальные сведения, полученные непосредственно из источников данных.

Помимо ЕНИП, полученные в работе результаты могут быть использованы при построении других распределенных информационных систем, предполагающих виртуальную интеграцию данных из разнородных источников. В частности, в настоящее время широко востребованы специализированные поисковые системы, интегрирующие информацию из различных Интернет-сайтов и систем, по некоторой тематике. Полученный в

работе результат представляет метод построения таких поисковых систем на основе технологий Семантического Веб.

Апробация работы

Основные результаты работы докладывались и обсуждались на следующих научных конференциях и семинарах:

- Всероссийская научная конференция "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" (Санкт-Петербург, 2003; Пущино, 2004).
- Научно-практический семинар "Новые технологии в информационном обеспечении науки" (Таруса, 2003-2005).
- Международная конференция The 8th World Multi-Conference on Systemics, Cybernetics and Informatics – SCI 2004 (Orlando, Florida, 2004).
- Международный коллоквиум Spring Young Researcher's Colloquium On Database and Information Systems – SYRCoDIS (Санкт-Петербург, 2004).
- Всероссийская научная конференция "Научный сервис в сети Интернет" (Новороссийск, 2004).
- Международная конференция "Порядковый анализ и смежные вопросы математического моделирования" (Владикавказ, 2006).
- Научная конференция МФТИ (Долгопрудный, 2005-2007).
- Научные семинары отдела Систем математического обеспечения Вычислительного Центра им. А.А. Дородницына РАН (Москва, 2003-2008).
- Научные семинары кафедры математического моделирования сложных процессов и систем МФТИ (ГУ) (Москва, 2005-2008).

Публикации

По теме диссертации опубликовано 20 работ, в том числе две [1, 2] из списка изданий, рекомендованных ВАК РФ.

Структура и объем работы

Диссертация состоит из введения, шести глав, заключения, списка использованных источников, включающего 70 наименований. Работа изложена на 100 страницах.

Краткое содержание работы

Введение

Во введении обоснована актуальность темы исследования, описаны решаемые проблемы. Рассмотрена задача интеграции данных, вопрос применения математического аппарата дескриптивной логики и технологий

Семантического Веб в системах интеграции данных. Введение дает характеристику основных проблем и задач, возникающих при этом.

Глава 1. Обзор методов интеграции данных

В первой главе охарактеризовано текущее состояние отрасли, приведен обзор методов предоставления интегрированного доступа к данным, указаны области применимости, преимущества и недостатки различных подходов. В частности, рассматриваются архитектурные принципы распределенных СУБД, «хранилищ данных», федеративных БД, систем интеграции данных по принципу посредников, по принципу взаимодействия равноправных узлов (P2P). Рассматривается способ классификации такого рода систем. Поясняются принципы проектирования и разработки систем интеграции данных «сверху вниз» и «снизу вверх».

Рассматривается архитектура централизованной системы интеграции данных по принципу посредников, на которой акцентируется внимание в работе. Задача такой системы, называемой также *посредником*, заключается в том, чтобы предоставить интегрированный доступ к множеству распределенных, разнородных, автономно разработанных источников, без необходимости централизовано хранить всю информацию из источников. Система предоставляет пользователю возможность формулировать запросы на выборку информации из таких источников в терминах глобальной схемы данных (общей системы понятий), которая проектируется «сверху» исходя из интересующих пользователя аспектов предметной области.

Для того чтобы абстрагироваться от разнообразия возможных видов информационных источников, предполагается, что каждый источник «обернут» так называемым *адаптером*, отвечающим за выборку сведений из источника в рамках принятой в системе единой модели данных, за предоставление стандартного технического интерфейса и стандартного языка запросов. Задача системы интеграции данных сводится к тому, чтобы обеспечить возможность динамически получить запрошенные пользователем данные через адаптеры информационных источников.

При этом в каждом источнике информация может представляться в терминах собственной схемы данных (системы понятий), соответственно, при включении источника в систему указывается некоторое семантическое отображение между терминами глобальной схемы данных и терминами различных схем данных источников. Выбор методики спецификации таких отображений определяет типы семантических конфликтов, которые могут быть разрешены с помощью системы интеграции данных, а также определяет алгоритмы, используемые для ответа на запросы в такой системе интеграции данных.

В завершение главы рассматриваются различные подходы к спецификации семантического отображения терминов. Дается сравнение подходов

Local-as-view (LAV), Global-as-view (GAV), описываются их преимущества и недостатки, рассматривается также гибридный подход (GLAV).

Глава 2. Технологии Семантического Веб и дескриптивная логика

Глава 2 дает обзор стандартов Семантического Веб (Semantic Web): унифицированной модели данных RDF (Resource Description Framework), языка веб-онтологий OWL (Web Ontology Language), языка запросов SPARQL, и рассматривает математическую основу языка веб-онтологий OWL – дескриптивную логику (Description logics). В этой главе вводятся необходимые для дальнейшего изложения определения.

Дескриптивная логика – это семейство языков представления знаний, предназначенных для выражения терминологического знания о предметной области. Дескриптивная логика оперирует двумя видами отношений – унарными, называемыми *концептами*, и бинарными – *ролями*. Различают *абстрактные роли*, связывающие объекты, и *атрибуты*, связывающие объекты со значениями примитивного типа данных.

В главе вводится формальная система определений, унифицирующая понятия различных диалектов дескриптивной логики, а также проводится систематизация и сравнение выразительных возможностей ряда диалектов дескриптивной логики. Специфицируется методика трансляции онтологий, выраженных на языке OWL, в формальную систему.

Онтологией в работе называется пара $\mathcal{O} \stackrel{\text{def}}{=} \{\mathcal{T}, \mathcal{A}\}$, где:

- $\mathcal{A}$ – *множество фактов*, высказываний об объектах онтологии в форме *C(a)* или *R(a,b)*, где *C* – концепт, *R* – роль, *a*, *b* – константы. Первый вид аксиом указывает принадлежность объекта *a* к концепту *C*, второй указывает, что объект *a* связан ролью *R* с объектом или значением *b*.
- $\mathcal{T}$ – *терминология*, множество терминологических аксиом, форма которых варьируется в различных диалектах дескриптивной логики.

Традиционно выделяются аксиомы вложения концептов $C_1 \sqsubseteq C_2$, где концепты C_i могут определяться на основе атомарных концептов A с помощью ряда конструкторов. В дескриптивной логике $\mathcal{SHOIN}^{(\mathcal{D})}$, соответствующей языку OWL-DL, концепты определяются следующей нотацией:

$$C \to \top \mid \bot \mid A \mid \neg C \mid C_1 \sqcap C_2 \mid C_1 \sqcup C_2 \mid \forall R_o.C \mid \exists R_o.C \mid \geq n\ R_s \mid \leq n\ R_s$$
$$\mid \{a_1,..., a_n\} \mid \geq n\ R_d \mid \leq n\ R_d \mid \forall R_d..R_{dn}.D \mid \exists R_{d1}..R_{dn}.D$$

Также в современных диалектах дескриптивной логики выделяют аксиомы вложения ролей $R_1 \sqsubseteq R_2$ и аксиомы транзитивности ролей Trans(R).

В работе дается подробное описание и пояснение семантики аксиом и конструкторов дескриптивной логики. Семантика аксиом определяет, как на основе исходного множества фактов $\mathcal{A}$ и аксиом $\mathcal{T}$ могут быть дедуцированы производные факты. Вводится понятие *интерпретации* $\mathcal{I}$, как функции, сопоставляющей каждому концепту онтологии некоторое мно-

жество объектов, и каждой роли – декартово произведение таких множеств. Интерпретация $\mathcal{I}$ называется *моделью* онтологии $\mathcal{O} = \{\mathcal{T}, \mathcal{A}\}$ ($\mathcal{I} \in \mathcal{M}(\mathcal{O})$), если она удовлетворяет всем высказываниям в $\mathcal{T}$ и $\mathcal{A}$. Высказывание *логически следует* из онтологии, если оно истинно для всех моделей онтологии. Онтология, не имеющая моделей, называется *противоречивой*.

В завершение главы вводится нотация для запросов на выборку данных и определяется семантика ответа на запросы относительно онтологий дескриптивной логики. Введенная формальная нотация для языка запросов ставится в соответствие синтаксической нотации языка SPARQL.

В частности, рассматривается класс *конъюнктивных запросов с простыми ограничениями* $\mathcal{CQ}_S$-$\mathcal{L}$ над некоторым диалектом дескриптивной логики $\mathcal{L}$, задаваемых в форме:

$$Q(\underline{X}) \leftarrow \bigwedge_{i=1..n} p_i(\underline{Z}_i) \ \wedge \ \bigwedge_{i=0..m} c_i(\underline{Z}_i),$$

где $Q(\underline{X})$ – заголовок запроса, а справа указано определение запроса, некоторая конъюнктивная формула $Q^{def}(\underline{X},\underline{Y})$ от переменных из векторов $\underline{X}$ и $\underline{Y}$. Переменные из вектора $\underline{X}$, указанного в заголовке запроса, называются свободными переменными запроса. Первая конъюнкция в определении запроса содержит атомы концептов $C(a)$ или ролей $R(a, b)$, а вторая – ограничения $op_1(u)$ или $(u\ op_2\ v)$, где op_1, op_2 – встроенные предикаты (например, операторы сравнения), a, b – переменные или константы, u – переменная, v – константа.

Множеством ответов на запрос Q *относительно интерпретации* $\mathcal{I}$ называется множество векторов констант $\underline{t}$, таких, что при подстановке их вместо свободных переменных в определении запроса Q, формула $\exists\underline{Y}\ Q^{def}(\underline{t},\underline{Y})$ является истинной в $\mathcal{I}$:

$$Q(\mathcal{I}) \stackrel{def}{=} \{\underline{t} \mid \mathcal{I} \vDash \exists\underline{Y}\ Q^{def}(\underline{t},\underline{Y}),\ \underline{t} = (c_1,\ldots,c_n),\ c_i \in \mathbb{C}\}$$

Множеством ответов на запрос Q *относительно онтологии* $\mathcal{O}$ называется множество векторов констант $\underline{t}$, которые являются ответами на запрос Q относительно любой ее модели:

$$Q(\mathcal{O}) \stackrel{def}{=} \{\underline{t} \mid \mathcal{I} \vDash Q(\underline{t})\ \forall \mathcal{I} \in \mathcal{M}(\mathcal{O}),\ \underline{t} = (c_1,\ldots,c_n),\ c_i \in \mathbb{C}\}$$

Здесь и далее $\mathcal{I} \vDash Q(\underline{t})$ по определению означает $\underline{t} \in Q(\mathcal{I})$.

Глава 3. Математическая модель системы интеграции данных на основе онтологий

В третьей главе вводится математическая модель системы интеграции данных по принципу посредников, особенностью которой является применение онтологий дескриптивной логики, и рассматривается семантика ответа на запросы в такой системе.

Система интеграции данных на основе онтологий для m источников данных $\Delta_1 \ldots \Delta_m$ формально определяется как $\Psi \stackrel{def}{=} \{\mathcal{O}_\Gamma, \{\mathcal{O}_{\Delta i}\}_{i=1..m}, \mathcal{F}\}$, где:

– $\mathcal{O}_\Gamma = \{\mathcal{T}_\Gamma, \mathcal{A}_\Gamma\}$ – *глобальная онтология*, непротиворечивая и выраженная на языке дескриптивной логики $\mathcal{L}_\Gamma$, без ограничения общности $\mathcal{A}_\Gamma$

$= \varnothing$. Глобальная онтология проектируется «сверху» исходя из аспектов предметной области, которые должна представлять система интеграции данных, и содержит термины, в которых пользователь может формулировать запросы к системе.

– $\{\mathcal{O}_{\Delta i}\}_{i=1..m}$ – конечное множество *онтологий источников* $\mathcal{O}_{\Delta i} = \{\mathcal{T}_{\Delta i}, \mathcal{A}_{\Delta i}\}$, непротиворечивых и выраженных на языке дескриптивной логики $\mathcal{L}_\Delta$. Физически информация, моделируемая множеством фактов $\mathcal{A}_{\Delta i}$, некоторым образом хранится в источнике данных Δ_i, и доступна через интерфейс запросов соответствующего адаптера.

– $\mathcal{F}$ – конечное множество *отображений* между глобальной онтологией $\mathcal{O}_\Gamma$ и множеством онтологий источников $\{\mathcal{O}_{\Delta i}\}_{i=1..m}$.

Рассматриваются отображения, задаваемые формулами $q_\Delta \rightsquigarrow q_\Gamma$, где q_Δ и q_Γ – некоторые запросы с одинаковым числом свободных переменных в терминах $\{\mathcal{O}_{\Delta i}\}_{i=1..m}$ и $\mathcal{O}_\Gamma$ соответственно, а знак $\rightsquigarrow$ обозначает одно из отношений $\{\subseteq, \supseteq, \equiv\}$ (семантика таких отображений вводится ниже понятием глобальной модели). Ключевым классом отображений, для которого в работе предлагается эффективный алгоритм ответа на запросы, являются так называемые корректные конъюнктивные GLAV-отображения с простыми ограничениями: $q_\Delta \subseteq q_\Gamma$, где $q_\Delta, q_\Gamma \in \mathcal{CQ}_S$-$\mathcal{L}$.

Запросы к системе интеграции данных на основе онтологий формулируются в терминах глобальной онтологии $\mathcal{O}_\Gamma$, и основной задачей системы является вычисление ответов на такие запросы на основе информации в источниках данных, а также правил отображения и аксиом глобальной онтологии.

Семантика ответа на запросы относительно системы интеграции данных на основе онтологий Ψ определяется следующим образом. Рассматриваются возможные модели $\mathcal{I}_\Gamma$ глобальной онтологии $\mathcal{O}_\Gamma$, *корректные* относительно всех отображений в Ψ, то есть такие, что для каждого отображения $q_\Delta \rightsquigarrow q_\Gamma$ в Ψ верно $q_\Delta(\mathcal{O}_\Delta) \rightsquigarrow q_\Gamma(\mathcal{I}_\Gamma)$, где $\mathcal{O}_\Delta \stackrel{\text{def}}{=} \{\bigcup_{i=1..m}\mathcal{T}_{\Delta i}, \bigcup_{i=1..m}\mathcal{A}_{\Delta i}\}$ – объединение онтологий источников, $\rightsquigarrow$ обозначает одно из отношений $\{\subseteq, \supseteq, \equiv\}$ между указанными множествами ответов на запрос. Такие $\mathcal{I}_\Gamma$ называются *глобальными моделями* системы интеграции Ψ, множество всех глобальных моделей Ψ обозначается $\mathcal{M}(\Psi)$.

Множеством точных ответов на запрос Q относительно системы интеграции данных Ψ называется множество векторов констант $\underline{t}$, которые являются ответами на запрос Q относительно любой глобальной модели:

$$Q(\Psi) \stackrel{\text{def}}{=} \{\underline{t} \mid \mathcal{I}_\Gamma \vDash Q(\underline{t})\ \forall \mathcal{I}_\Gamma \in \mathcal{M}(\Psi)\}$$

Это определение означает, что такие ответы логически следуют из фактов и высказываний онтологий источников, отображений, а также высказываний глобальной онтологии.

В следующем разделе рассматривается теоретическая сторона ответа на запросы относительно системы интеграции данных на основе онтоло-

гий и предлагается способ сведения этой задачи к известной задаче ответа на запросы относительно отдельной онтологии. Для этого предлагается алгоритм вычисления *множества извлеченных фактов* $\mathcal{A}_{\text{ret}}(\{\mathcal{O}_{\Delta i}\}_{i=1..m}, \mathcal{F})$, содержащего все высказывания, получаемые из источников данных на основе отображений $\mathcal{F}$, и вводится понятие *извлеченной онтологии*:

$$\mathcal{O}_{\text{ret}}(\Psi) = \{\mathcal{T}_\Gamma, \mathcal{A}_{\text{ret}}(\{\mathcal{O}_{\Delta i}\}_{i=1..m}, \mathcal{F})\}$$

Доказана следующая теорема. Пусть задана система интеграции данных $\Psi = \{\mathcal{O}_\Gamma, \{\mathcal{O}_{\Delta i}\}_{i=1..m}, \mathcal{F}\}$, где отображения онтологий $\mathcal{F}$ заданы в форме $q_\Delta \subseteq q_\Gamma$, где $q_\Gamma \in \mathcal{CQ}_S\text{-}\mathcal{L}_\Gamma$. В таком случае множество глобальных моделей системы интеграции данных Ψ совпадает с множеством моделей извлеченной онтологии:

$$\mathcal{M}(\Psi) = \mathcal{M}(\mathcal{O}_{\text{ret}}(\Psi))$$

Таким образом, для вычисления множества точных ответов на запрос Q относительно системы интеграции данных достаточно вычислить ответ на этот запрос относительно извлеченной онтологии: $Q(\Psi) = Q(\mathcal{O}_{\text{ret}}(\Psi))$.

Глава 4. Интеграция больших объемов данных на основе онтологий

В главе 4 рассматривается вопрос поиска эффективных на практике методов ответа на запросы в случае интеграции источников, содержащих большие объемы информации, прежде всего реляционных баз данных. Очевидно, в таких случаях построение извлеченной онтологии потребует чрезмерных вычислительных и сетевых ресурсов и рассмотренный «прямой» метод ответа на запросы не может быть приемлем на практике. В этой связи предлагается разбить задачу на два этапа следующим образом:

1) На первом этапе на основе определений системы интеграции данных производится *переформулировка* исходного запроса Q, заданного в терминах глобальной онтологии, в запрос Q', заданный в терминах онтологий источников данных, который может быть эффективно исполнен системой-посредником.
2) На втором этапе соответственно полученный запрос исполняется системой-посредником.

При этом под запросами, которые могут быть эффективно исполнены относительно источников, представленных объемными реляционными базами данных, понимаются запросы, выразимые в виде формулы реляционного исчисления (такие запросы могут быть представлены на языке SQL и эффективно исполнены РСУБД).

Вводятся формальные определения, связанные с переформулировкой запросов относительно систем интеграции данных на основе онтологий.

Запрос Q' называется *точной переформулировкой* запроса Q на основе системы интеграции данных Ψ, если:

1) Все концепты или роли, используемые в запросе Q', являются терминами онтологий источников $\{\mathcal{O}_{\Delta i}\}_{i=1..m}$ системы интеграции Ψ;
2) Множество ответов на запрос Q' относительно объединения всех онтологий источников $\mathcal{O}_\Delta \stackrel{\text{def}}{=} \{\bigcup_{i=1..m}\mathcal{T}_{\Delta i}, \bigcup_{i=1..m}\mathcal{A}_{\Delta i}\}$ совпадает с множеством точных ответов на запрос Q относительно системы интеграции данных Ψ: $Q'(\mathcal{O}_\Delta) = Q(\Psi)$.

Вводятся понятия *частичной* и *максимальной переформулировки* запроса на основе системы интеграции данных *на заданном языке запросов*, который обозначим $\mathcal{QL}_R$. Максимальной называется переформулировка, вычисляющая, в определенном смысле, наиболее близкий к исходному запросу ответ при любых данных в источниках.

Рассматривается вопрос, в каких случаях точная переформулировка может быть выражена на заданном языке запросов (на практике этот язык должен быть фиксирован и реализован в адаптерах информационных источников).

Доказаны леммы, определяющие необходимые условия существования точной переформулировки на заданном языке запросов. Пусть в системе интеграции данных Ψ возможна точная переформулировка запроса на языке $\mathcal{QL}_\Gamma$ над дескриптивной логикой $\mathcal{L}_\Gamma$ в запрос на языке $\mathcal{QL}_R$ над дескриптивной логикой $\mathcal{L}_\Delta$, где $\mathcal{L}_\Gamma$ и $\mathcal{L}_\Delta$ – диалекты глобальной онтологии и онтологий источников, соответственно. В таком случае, сложность ответа на $\mathcal{QL}_R$ запросы для онтологий на дескриптивной логике $\mathcal{L}_\Delta$ относительно объема множества фактов онтологии $|\mathcal{A}|$ не ниже сложности ответа на запросы на языке $\mathcal{QL}_\Gamma$ для системы интеграции данных Ψ относительно суммарного объема фактов всех источников в системе $|\bigcup_{i=1..m}\mathcal{A}_{\Delta i}|$, которая в свою очередь не ниже сложности ответа на запросы на языке $\mathcal{QL}_\Gamma$ для онтологий на дескриптивной логике $\mathcal{L}_\Gamma$ относительно объема множества фактов онтологии $|\mathcal{A}|$.

Как следствие, сформулирована теорема, задающая необходимое условие существования точной переформулировки конъюнктивного запроса в виде формулы реляционного исчисления. Показано, что для того, чтобы всегда существовала точная переформулировка $\mathcal{CQ}$-$\mathcal{L}_\Gamma$ запроса относительно Ψ в виде формулы реляционного исчисления, необходимо, чтобы сложность ответа на $\mathcal{CQ}$-$\mathcal{L}_\Gamma$ запросы для онтологий на языке дескриптивной логики $\mathcal{L}_\Gamma$ относительно объема фактов онтологии $|\mathcal{A}|$ лежала в классе $\mathcal{LOGSPACE}$.

На основе анализа вычислительных характеристик различных диалектов дескриптивной логики (сложности ответа на запросы относительно объема множества фактов онтологии) делается вывод о том, какие диалекты и конструкции дескриптивной логики не могут быть использованы в системе интеграции данных на основе онтологий, если такая система должна обеспечивать эффективную интеграцию объемных реляционных БД.

Рассматривается вопрос поиска максимально выразительного диалекта дескриптивной логики, допускающего точную переформулировку конъюнктивных запросов в реляционное исчисление. Вводится диалект $\mathcal{DL}_{\text{trio}}$, обладающий указанным свойством (аббревиатура trio – от tractable integration of ontologies, т.е. диалект, допускающий интеграцию онтологий с полиномиальным временем ответа).

Онтология на диалекте $\mathcal{DL}_{\text{trio}}$ может включать следующие формы терминологических аксиом, семантика которых рассмотрена в работе:

- $C_L \sqsubseteq C$ – ограниченная аксиома вложения концептов, где:
  $C_L \rightarrow A \mid \exists R_o \mid \exists R_d \mid C_{L1} \sqcap C_{L2}$
  $C \rightarrow \top \mid \bot \mid A \mid \neg A \mid C_1 \sqcap C_2 \mid \exists R_o \mid \exists R_d \mid \exists R_o.C \mid \exists R_{d1}..R_{dn}.D \mid \neg\exists R_o \mid \neg\exists R_d$
- $R_{o1} \sqsubseteq R_{o2}$, $R_{d1} \sqsubseteq R_{d2}$ – аксиомы вложения ролей;
- $R_{o1} \sqsubseteq \neg R_{o2}$, $R_{d1} \sqsubseteq \neg R_{d2}$ – аксиомы различия ролей;
- $\rho(R_d) \sqsubseteq d$ – аксиома типа значений атрибута.

Показано, что диалект $\mathcal{DL}_{\text{trio}}$ – максимальный в том смысле, что расширение этого диалекта рядом других конструкторов или видов аксиом приводит к невозможности точной переформулировки конъюнктивных запросов в реляционное исчисление.

Приводится спецификация соответствия диалекта $\mathcal{DL}_{\text{trio}}$ конструкциям языка веб-онтологий OWL. Рассматривается метод вычисления множества ответов на запрос относительно онтологии для дескриптивной логики $\mathcal{DL}_{\text{trio}}$, используемый в дальнейшем для доказательства корректности основного предлагаемого в работе алгоритма построения переформулировки запроса.

Сформулирована и доказана также используемая в дальнейшем теорема о существовании точной переформулировки. Пусть система интеграции данных $\Psi = \{\mathcal{O}_\Gamma, \{\mathcal{O}_{\Delta i}\}_{i=1..m}, \mathcal{F}\}$ такова, что $\mathcal{O}_\Gamma$ и $\mathcal{O}_{\Delta i}$ выражены на некотором диалекте дескриптивной логики $\mathcal{L}$, все отображения онтологий $\mathcal{F}$ заданы в форме $q_\Delta \subseteq q_\Gamma$, где q_Δ, $q_\Gamma \in \mathcal{CQ}_S\text{-}\mathcal{L}$. Пусть пользовательский запрос Q к системе Ψ задается в форме объединения конъюнктивных запросов $Q \in \mathcal{UCQ}_S\text{-}\mathcal{L}$. Пусть Q' есть объединение максимальных переформулировок запроса Q в форме конъюнктивных запросов с простыми ограничениями $\mathcal{CQ}_S\text{-}\mathcal{L}$. Тогда Q' является максимальной переформулировкой запроса Q на языке $\mathcal{UCQ}_S\text{-}\mathcal{L}$, и Q' является точной переформулировкой запроса Q относительно системы интеграции данных Ψ.

Глава 5. Алгоритм переформулировки запросов для систем интеграции данных на основе онтологий

В пятой главе рассматривается задача построения точной переформулировки запроса относительно системы интеграции данных, построенной на основе онтологий $\mathcal{DL}_{\text{trio}}$.

Доказана следующая теорема. Пусть система интеграции данных $\Psi = \{\mathcal{O}_\Gamma, \{\mathcal{O}_{\Delta i}\}_{i=1..m}, \mathcal{F}\}$ такова, что:

- Глобальная онтология $\mathcal{O}_\Gamma$ непротиворечива и выражена на языке дескриптивной логики $\mathcal{L}_\Gamma = \mathcal{DL}_{\text{trio}}$;
- Онтологии источников данных $\mathcal{O}_{\Delta i}$ непротиворечивы и выражены на языке дескриптивной логики $\mathcal{L}_\Delta = \mathcal{DL}_{\text{trio}}$;
- Отображения $\mathcal{F}$ заданы в форме $q_\Delta \subseteq q_\Gamma$, где $q_\Delta, q_\Gamma \in \mathcal{CQ}_S\text{-}\mathcal{DL}_{\text{trio}}$ – конъюнктивные запросы с простыми ограничениями над дескриптивной логикой $\mathcal{DL}_{\text{trio}}$.

Пусть пользовательский запрос Q к системе Ψ задается в форме объединения конъюнктивных запросов $\mathcal{QL}_\Gamma = \mathcal{UCQ}_S\text{-}\mathcal{DL}_{\text{trio}}$. Тогда существует точная переформулировка запроса Q на основе системы интеграции данных Ψ на языке запросов $\mathcal{QL}_R = \mathcal{UCQ}_S$, то есть, представимая в виде объединения реляционных конъюнктивных запросов в терминах источников.

Для указанного в теореме класса систем интеграции данных на основе онтологий предлагается алгоритм построения точной переформулировки. Основная идея алгоритма заключается в разделении на отдельные этапы переформулировки запроса относительно отображений онтологий и относительно аксиом онтологий. При этом предлагается способ сведения подзадачи переформулировки относительно отображений онтологий к аналогичной задаче для реляционных систем интеграции данных, алгоритмы решения которой известны.

Предварительными условиями для предлагаемого алгоритма переформулировки запросов являются:

1) Нормализация $\mathcal{DL}_{\text{trio}}$ терминологий $\mathcal{T}_\Gamma$ и $\{\mathcal{T}_{\Delta i}\}_{i=1..m}$ применением к аксиомам ряда правил нормализации.
2) Проверка непротиворечивости системы интеграции данных Ψ, то есть существования глобальных моделей Ψ, что эквивалентно непротиворечивости извлеченной онтологии $\mathcal{O}_{\text{ret}}(\Psi)$. В случае, если данные в источниках в какой-то степени противоречат аксиомам глобальной онтологии, ответ на запрос относительно системы интеграции данных является по определению бессмысленным.

Предлагаемый алгоритм переформулировки запросов включает следующие основные этапы:

1) Переформулировка относительно аксиом глобальной онтологии;
2) Переформулировка относительно отображений онтологий;
3) Переформулировка относительно аксиом онтологий источников;
4) Минимизация полученного запроса.

На первом этапе алгоритма производится построение промежуточной переформулировки пользовательского запроса $Q \in \mathcal{UCQ}_S\text{-}\mathcal{DL}_{\text{trio}}$ с учетом аксиом глобальной онтологии $\mathcal{T}_\Gamma$. Основная идея этапа заключается в том, чтобы «закодировать» в запрос необходимые аксиомы терминологии $\mathcal{T}_\Gamma$.

«Прямой» метод ответа на запрос относительно системы Ψ предполагает предварительное проведение логического вывода на основе фактов и аксиом извлеченной онтологии $\mathcal{O}_{\text{ret}}(\Psi)$, в результате которого вычисляются производные факты. Предлагаемый алгоритм позволяет исключить необходимость проведения такого логического вывода, вместо этого переформулировав исходный запрос таким образом, чтобы он учитывал все необходимые производные факты. По сути, построение такого переформулированного запроса является своего рода логическим выводом относительно исходного запроса и аксиом глобальной онтологии. Алгоритм производит построение альтернативных формулировок запроса на основе аксиом терминологии исчерпывающим применением к запросу ряда правил замены предикатов. Помимо расширения запроса альтернативными формулировками, алгоритм производит отсеивание подзапросов, заведомо пустых согласно аксиомам, и промежуточную минимизацию запросов.

На следующем этапе вычисляется переформулировка полученного запроса относительно правил отображения онтологий $\mathcal{F}$, уже без необходимости учитывать аксиомы глобальной онтологии. Абстрагируясь от способа ответа на запросы относительно онтологий источников, такая задача может быть сведена к аналогичной задаче для реляционной модели данных. В работе описывается адаптированный к рассматриваемой задаче алгоритм переформулировки запросов относительно отображений.

На третьем этапе алгоритма производится сведение полученного после предыдущих этапов $\mathcal{UCQ}_{\text{S}}$-$\mathcal{DL}_{\text{trio}}$ запроса, сформулированного в терминах онтологий источников, в реляционный запрос из класса $\mathcal{UCQ}_{\text{S}}$ (подкласс реляционного исчисления). Для этого требуется «закодировать» в запрос аксиомы онтологий источников данных, что делается полностью аналогично первому этапу, относительно объединения всех аксиом источников $\bigcup_{i=1..m} \mathcal{T}_{\Delta i}$.

На последнем этапе производится минимизация итогового запроса, то есть удаление из запроса избыточных целей путем применения к нему ряда преобразований. Этот этап обеспечивает частичную оптимизацию запроса.

В результате выполнения всех этапов алгоритма вычисляется точная переформулировка исходного $\mathcal{UCQ}_{\text{S}}$-$\mathcal{DL}_{\text{trio}}$ запроса относительно системы интеграции Ψ в виде объединения реляционных конъюнктивных запросов с ограничениями ($\mathcal{UCQ}_{\text{S}}$), в которых упоминаются только атомарные концепты и роли онтологий источников данных. Для исполнения полученного запроса могут быть применены известные методы, используемые для реляционной модели данных.

В работе приводится доказательство корректности предложенного алгоритма, анализируются его характеристики и вычислительная сложность.

Глава 6. Анализ и применение полученных результатов

В главе 6 приведен анализ полученных результатов, выразительных возможностей рассмотренного класса систем интеграции данных на основе онтологий. Показано, каким образом такие системы интеграции данных позволяют устранить различные виды семантических конфликтов. Приведено сравнение со смежными работами, анализируются направления дальнейших исследований.

Описывается методология построения программных систем интеграции данных, соответствующих предложенной математической модели.

В предлагаемой архитектуре всякий адаптер к информационному источнику сопровождается онтологией на языке OWL (на диалекте $\mathcal{DL}_{\text{trio}}$), содержащей семантическое описание источника. Адаптер реализует внешний интерфейс ответа на SPARQL запросы, и обеспечивает трансляцию таких запросов во внутренний язык запросов источника данных, например SQL.

Глобальная онтология системы интеграции данных также формулируется на языке веб-онтологий OWL, в рамках диалекта $\mathcal{DL}_{\text{trio}}$.

Отображения онтологий источников в глобальную онтологию могут задаваться различными способами, в т.ч.:

- В онтологиях источников с помощью конструкций языка OWL (позволяют выразить терминологические отображения);
- С помощью правил вывода, в том числе, на языке SWRL (могут использоваться для спецификации конъюнктивных отображений);
- С помощью отображений в виде пары SPARQL запросов $q_{\Delta} \subseteq q_{\Gamma}$, где запрос в терминах источников q_{Δ} в общем случае может быть произвольным запросом, который может быть исполнен адаптерами, а форма запроса в терминах глобальной онтологии q_{Γ} ограничена сводимостью к классу $\mathcal{CQ}_{S}$.
- С помощью расширенных отображений, в которых помимо пары SPARQL запросов указывается программная функция преобразования значений переменных.

Все указанные способы сводятся к предложенной в работе математической модели отображений онтологий.

Запросы к системе интеграции данных (посреднику) формулируются на ограниченном языке SPARQL, в терминах глобальной онтологии. Система обеспечивает динамическое исполнение таких запросов в соответствии с предложенным алгоритмом.

В завершение рассмотрено практическое применение полученных результатов в контексте Единого Научного Информационного Пространства РАН. Приводится описание прототипа системы исполнения распределенных запросов в среде Единого Научного Информационного Пространства РАН (ЕНИП РАН). Рассматривается также спектр других актуальных за-

дач, для решения которых могут быть применены полученные результаты.

Заключение

В заключении приведены основные результаты диссертационной работы.

Основные результаты работы

1. Предложена математическая модель систем интеграции данных на основе онтологий, введена система определений на базе математического аппарата дескриптивной логики, формализованы понятия ответа на запрос и переформулировки запроса в системах интеграции данных на основе онтологий.

2. В рамках предложенной модели исследованы условия существования точной переформулировки запроса на выбранном языке запросов. Предложен и обоснован выбор диалекта дескриптивной логики, который целесообразно использовать при интеграции больших объемов данных, хранимых в реляционных базах данных.

3. Разработан алгоритм построения точной переформулировки запроса для выбранного класса систем интеграции данных на основе онтологий.

4. Предложена методология разработки систем интеграции данных на основе онтологий, в соответствии с формальной моделью.

5. На основе полученных теоретических результатов разработан прототип системы и комплекс программ исполнения распределенных запросов в среде Единого Научного Информационного Пространства РАН (ЕНИП РАН), предназначенной для виртуальной интеграции данных различных научных учреждений в ЕНИП РАН.

Список публикаций по теме диссертации

1. *Бездушный А.А.* Математическая модель системы интеграции данных на основе онтологий // Журнал «Вестник НГУ», серия «Информационные технологии» – Новосибирск, 2008. – Т.6, вып.2. – С 15-40.

2. *Бездушный А.Н., Кулагин М.В., Серебряков В.А., Бездушный А.А., Нестеренко А.К., Сысоев Т.М.* Предложения по наборам метаданных для научных информационных ресурсов // Журнал «Вычислительные Технологии» – Новосибирск, 2005 – Т.10, вып.7. – С. 29-48.

3. *Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И.* Интеграция метаданных Единого Научного Информационного Пространства РАН. – М.: Вычислительный Центр им. А.А. Дородницына РАН, 2006. – 238 с.

4. *Бездушный А.А.* Распределенное исполнение SPARQL-запросов в гетерогенной среде // Моделирование и обработка информации: Сборник научных трудов / Моск. физ.-тех. ин-т. – М., 2008. – С. 230-235.

5. *Bezdushny A.A., Bezdushny A.N., Nesterenko A.K., Serebriakov V.A., Sysoev T.M.* Integrated System of Information Resources of the Russian Academy of Sciences // Proceedings of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics SCI 2004, Orlando, Florida – 2004. – P. 462-467.

6. *Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М.* Архитектура RDFS-системы. Практика использования открытых стандартов и технологий Semantic Web в системе ИСИР // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды V всероссийской научной конференции / Изд-во СПбГУ. – СПб., 2003. – С. 45-60.

7. *Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М.* Java и XML технологии новой версии ИСИР // Современные технологии в информационном обеспечении науки (ред. Н. Е. Калёнов) – М., 2003. – С. 182-205.

8. *Бездушный А.А., Бездушный А.Н., Жижченко А.Б., Кулагин М.В., Серебряков В.А.* RDF схема метаданных ИСИР // Современные технологии в информационном обеспечении науки (ред. Н. Е. Калёнов) – М., 2003. – С. 141-159.

9. *Bezdushny A.A., Nesterenko A.K.* ISIR Architecture for Web-Repository Integration // Сборник докладов Первого весеннего коллоквиума молодых исследователей в области баз данных и информационных систем (SYRCoDIS'2004) – СПб., 2004. – С. 60-66.

10. *Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М.* Возможности технологий ИСИР в поддержке Единого Научного Информационного Пространства РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды VI всероссийской научной конференции – М., 2004. – С. 254-262.

11. *Бездушный А.А., Бездушный А.Н., Жижченко А.Б., Калёнов Н.Е., Кулагин М.В., Серебряков В.А.* Предложения по наборам метаданных для научных информационных ресурсов ЕНИП РАН // Электронные биб-

лиотеки: перспективные методы и технологии, электронные коллекции: Труды VI всероссийской научной конференции – М., 2004. – С. 277-284.

12. *Бездушный А.А., Нестеренко А.К, Сысоев Т.М. , Бездушный А.Н., Серебряков В.А.* Архитектурные решения ИСИР на платформах Java и XML // Интегрированная система информационных ресурсов: архитектура, реализация, приложения: Сборник трудов / Вычислительный Центр им. А.А. Дородницына РАН. – М., 2004. – С. 78-95.

13. *Бездушный А.А.* Роль технологий Semantic Web в решениях ИСИР // Интегрированная система информационных ресурсов: архитектура, реализация, приложения: Сборник трудов / Вычислительный Центр им. А.А. Дородницына РАН. – М., 2004. – С. 36-55.

14. *Бездушный А.А., Нестеренко А.К., Сысоев Т.М., Кулагин М.В.* Semantic Web и OWL-онтологии в разработке ИСИР-систем // Научный сервис в сети Интернет: Труды Всероссийской научной конференции. / Изд-во МГУ. – М., 2004. – С. 188-191.

15. *Бездушный А.А., Бездушный А.Н., Серебряков В.А.* Схемы метаданных ЕНИП: практика применения OWL в ЕНИП // Информационное обеспечение науки: новые технологии (ред. Н. Е. Калёнов) – М., 2005. – С.155-182.

16. *Бездушный А.А.* Применение технологий Semantic Web для обеспечения интероперабельного обмена научной информацией // Современные проблемы фундаментальных и прикладных наук: Труды 48-й научной конференции МФТИ. Часть VII. / Моск. физ.-тех. ин-т. – М., 2005. – С. 209-211.

17. *Бездушный А.А.* Схемы метаданных для научных информационных ресурсов ЕНИП РАН // Порядковый анализ и смежные вопросы математического моделирования: Труды IV международной научной конференции. / Институт прикладной математики и информатики. – Владикавказ, 2006. – С. 260 - 271.

18. *Бездушный А.А.* Архитектура интеграции данных ИСИР // Современные проблемы фундаментальных и прикладных наук. Часть VII: Труды 49-й научной конференции МФТИ. / Моск. физ.-тех. ин-т. – М., 2006. – С. 230-231.

19. *Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М., Теймуразов К.Б., Филиппов В.И.* Информационная Web-система «Научный институт» на платформе ЕНИП. – М.: Вычислительный Центр им. А.А. Дородницына РАН, 2007. – 248 с.

20. *Бездушный А.А.* RQuery - язык запросов к источникам данных Semantic Web // Современные проблемы фундаментальных и прикладных наук. Часть VII: Труды 50-й научной конференции МФТИ. / Моск. физ.-тех. ин-т. – М., 2007. – Т.2 – С. 57-59.

В работах с соавторами личный вклад автора заключается в создании методов разработки распределенных систем и интеграции данных на основе OWL-онтологий и дескриптивной логики, в соответствии с формальной моделью. Автором предложен основанный на применении OWL-онтологий подход к интеграции данных в Интегрированной Системе Информационных Ресурсов (ИСИР), Едином Научном Информационном Пространстве РАН (ЕНИП РАН), создан соответствующий комплекс программных модулей.