

*На правах рукописи*



**КРОПОТИН Александр Александрович**

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ  
ПРОЦЕССА АНАЛИЗА РЕЛЯЦИОННЫХ  
БАЗ ДАННЫХ ПРИ ИНТЕГРАЦИИ  
ИНФОРМАЦИОННЫХ СИСТЕМ**

**05.13.18 – «Математическое моделирование,  
численные методы и комплексы программ»**

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата физико-математических наук

**Тюмень – 2017**

Работа выполнена на кафедре программной и системной инженерии института математики и компьютерных наук ФГАОУ ВО «Тюменский государственный университет».

**Научный руководитель:** доктор технических наук, профессор  
**Ивашко Александр Григорьевич**

**Официальные оппоненты:** **Инютин Сергей Арнольдович,**  
доктор технических наук, профессор,  
ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)», профессор кафедры управления информационной структурой бизнеса

**Кикоть Stanisлав Павлович,**  
кандидат физико-математических наук,  
ФГБУН Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), старший научный сотрудник

**Ведущая организация:** **ФГБОУ ВПО «Московский государственный университет имени М.В. Ломоносова»**

Защита состоится 07 декабря 2017 года в 15-00 часов на заседании диссертационного совета Д 212.274.14 при ФГАОУ ВО «Тюменский государственный университет» по адресу: 625003, г. Тюмень, ул. Перекопская, 15А, ауд. 410.

С диссертацией можно ознакомиться в библиотеке ФГАОУ ВО «Тюменский государственный университет» и на сайте <https://diss.utmn.ru/sovet/diss-sovet-212-274-14/zashchita/327341/>.

Автореферат разослан « \_\_\_\_ » \_\_\_\_\_ 2017 г.

*Ученый секретарь  
диссертационного совета*

*Е.А. Оленников*

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования.** Развитие технологий создания интеллектуальных систем и Semantic Web во многом обуславливает возросший интерес к исследованиям, направленным на разработку технологий и инструментов интеграции и обеспечения интероперабельности информационных систем.

Среди многих научных школ во всем мире, которые участвуют в исследованиях, связанных с интеграцией информационных систем, можно выделить Department of Informatics – Database Technology, University of Zurich, Wright State University, Dublin City University School of Computing. Кроме того, ежегодно организуется множество конференций на тему интеграции информационных систем, в числе которых проводятся следующие: Интеграция интеллектуальных систем, Международная конференция по корпоративным информационным системам, Международная конференция по интеграции данных в науках о жизни.

Особое место в процессе интеграции и обеспечения интероперабельности совокупности информационных систем занимают задачи интеграции данных, которые характеризуются необходимостью обеспечения обмена данными между системами, исключая её дублирования, потерю и несогласованности, при возрастающем её объеме.

Традиционные методы решения задач интеграции данных в основном направлены на решение задач интеграции каждого конкретного случая и характеризуются построением связей между элементами систем вручную. Такие методы, как правило, требуют выполнения следующих шагов для каждой из систем, в случаях реализации метода или изменения структуры одной из систем: проведение детального анализа каждой системы для выявления связей и зависимостей, поиск источников информации, семантическая обработка запросов, построение правил и ограничений целостности для запросов, обрабатывающих несколько источников данных одновременно, обеспечение ограничения доступа, производство работы с метаданными. Кроме того, готовые технологии по интеграции данных сложны и далеки от решения семантических проблем.

Один из наиболее перспективных на данный момент методов интеграции основан на применении онтологических спецификаций для описания информационных ресурсов. Этот метод тесно связан с технологией Semantic Web и одним из ее самых перспективных и развивающихся направлений – технологией связанных данных. Большой вклад в разработку таких методов внесли российские и зарубежные ученые А. Ф. Тузовский, Л. В. Найханова, А. Н. Бездушный, А. С. Клещев, В. А. Серебряков, И. С. Михайлов, Ю. А. Загорулько, К. И. Шахгельдян, N. Guarino, N. Noy, M. Ehrig, G. Nenadic и A. Maedche.

**Целью** исследования является разработка способа применения формализма дескрипционных логик для автоматизации процесса выявления конфликтов концептуальных схем сущность-связь реляционных баз данных.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Разработать математическую модель онтологического представления реляционных баз данных на основе формализма дескрипционных логик, который обеспечивает необходимые синтаксис, семантику и механизмы логического вывода, достаточные для выявления семантических конфликтов метаданных и несоответствия данных, при интеграции этих баз данных.

2. На основании математической модели онтологического представления концептуальных объектов и численных алгоритмов анализа таких баз знаний для выявления и разрешения структурных и семантических конфликтов интегрируемых данных разработать комплекс программ проверки согласованности концептуальных схем сущность-связь, который реализует функции отображения метаданных и данных в базу знаний, описанную в формате языка OWL.

3. С целью апробации комплекса программ проверки согласованности концептуальных схем сущность-связь провести серию экспериментов на опытных данных, которые представляют различные случаи возникновения таких конфликтов.

**Научная новизна исследования** заключается в получении следующих результатов:

1. Разработана онтологическая модель концептуальных объектов, которая предоставляет правила представления концептуальных схем сущность-связь реляционных баз данных в виде терминологических аксиом и утверждений дескрипционной логики. Расширение онтологической модели концептуальных объектов добавляет правила слияния терминологических аксиом и утверждений онтологий концептуальных схем, которые направлены на обеспечение возможности выявления семантических конфликтов метаданных и несоответствия данных этих схем.

2. Разработан метод проверки согласованности концептуальных схем сущность-связь реляционных баз данных, который позволяет свести процесс выявления их семантических конфликтов метаданных и несоответствия данных к решению алгоритмических проблем согласованности и классификации терминологии и согласованности онтологии этих схем.

3. Разработан алгоритм построения терминологических аксиом и утверждений дескрипционной логики, в соответствии с правилами онтологической модели представления концептуальных объектов и ее расширения, исходя из семантического представления реляционных баз данных.

4. Разработан комплекс программ, который реализует предложенный алгоритм преобразования концептуальных схем реляционных баз данных и метод выявления их семантических конфликтов и несоответствия данных.

**Теоретическая значимость работы** заключается в разработке новых математических моделей, методов и алгоритмов, и архитектуры комплекса программ, которые позволяют идентифицировать семантические конфликты метаданных и несоответствия интегрируемых реляционных баз данных.

**Практическая значимость работы** заключается в реализации разработанных математических моделей, методов и алгоритмов в виде кроссплатформенного комплекса программ, который предлагает функции преобразования концептуальных схем реляционных

баз данных и метод выявления их семантических конфликтов и несоответствия данных.

**Объектом исследования** является процесс интеграции данных информационных систем.

**Предметом исследования** являются основанные на онтологическом моделировании данных методы и алгоритмы интеграции реляционных баз данных.

**Методология и методы исследования**, которые применялись при решении поставленных задач, основаны на методах математического моделирования, теориях искусственного интеллекта, графов и множеств, парадигмы представления знаний Semantic Web, формализме дескрипционных логик, а также технологии объектно-ориентированного программирования и стандартах OWL 2.0, UML 2.5 и XMI (ISO/IEC 19509).

**На защиту выносятся** следующие результаты, соответствующие трем пунктам паспорта специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ по физико-математическим наукам:

*Пункт 1. Разработка новых математических методов моделирования объектов и явлений.*

1. Предложена новая математическая модель представления концептуальных схем сущность-связь реляционных баз данных в виде семантической сети, узлами которой являются концепты и индивиды, обозначающие сущности и значения, а дугами – роли, обозначающие ассоциации между сущностями, атрибуты и ограничения. Такая модель предоставляет правила описания концептуальных схем в виде терминологических аксиом и утверждений дескрипционной логики  $SHOIF(D)$ , которые основаны на синтаксическом сходстве баз данных и онтологий, обоснованы принятым предположением о замкнутости мира в базах данных против принятого предположения об открытости мира в онтологиях.

2. Предложена новая математическая модель объединения семантических сетей концептуальных схем сущность-связь реляционных баз данных. Такая модель предоставляет правила объединения сетей в виде терминологических аксиом и утверждений деск-

рипционной логики, которые направлены на обеспечение возможности выявления семантических конфликтов метаданных и несоответствия данных концептуальных схем через проверку согласованности их онтологии.

*Пункт 3. Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий.*

3. Предложен новый метод проверки согласованности концептуальных схем сущность-связь реляционных баз данных, направленный на выявление семантических конфликтов метаданных и несоответствия интегрируемых реляционных баз данных посредством проверки согласованности их семантических сетей. Такой метод основан на применении формализма дескрипционных логик для выявления несоответствий терминологических аксиом и/или утверждений семантической сети и, в частности, решения алгоритмических проблем согласованности и классификации терминологии, и согласованности онтологии.

*Пункт 4. Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента.*

4. Предложен комплекс программ, который реализует алгоритм построения терминологических аксиом и утверждений дескрипционной логики, по правилам онтологической модели концептуальных объектов, исходя из семантического описания реляционных баз данных и метод выявления семантических конфликтов метаданных и несоответствия данных этих баз.

5. Проведена серия вычислительных экспериментов на опытных данных, включающих 902 базы данных. Результаты экспериментов показали адекватность разработанных математических моделей, методов и алгоритма.

Таким образом, в соответствии с формулой специальности 05.13.18 в диссертации представлены оригинальные результаты одновременно из трех областей: математического моделирования, численных методов и комплексов программ.

**Достоверность** изложенных в работе результатов подтверждается использованием известных математических аппаратов, техно-

логий разработки программных продуктов, их тестирования вычислительными экспериментами, а также положительной оценкой на научных конгрессах, конференциях и семинарах.

**Апробация результатов.** Основные результаты диссертационной работы докладывались на конференциях, таких как:

1. 27th DAAM International Symposium on Intelligent Manufacturing and Automation. – 2016. Mostar, BiH. – 26-29 October.

2. 10th IEEE International Conference on Application of Information and Communication Technologies AICT2016. – 2016. Baku, Azerbaijan. – 12-14 October.

3. 3rd International Workshop on OWL Reasoner Evaluation (ORE 2014) co-located with the Vienna Summer of Logic (VSL 2014). – 2014. Vienna, Austria. 13 July.

4. Международная научно-практическая конференция «Современные проблемы и пути их решения в науке, транспорте, производстве и образовании'2012». – 2012. Одесса, Украина. – 18-27 декабря.

**Структура и объем диссертации.** Диссертация состоит из введения, 4 глав, выводов, библиографического списка из 208 наименований и 9 приложений. Общий объем работы составляет 333 страницы, включает 40 рисунков, 3 листинга и 9 таблиц.

**Благодарности.** Автор выражает глубокую благодарность научному руководителю профессору, д. т. н. А.Г. Ивашко, к. ф.-м. н. А.В. Григорьеву за наставление и помощь при выполнении диссертационных исследований; к. ф.-м. н. Е.Е. Золину за помощь и разъяснение тонкостей работы логического вывода дескрипционных логик.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** раскрывается актуальность темы диссертационного исследования, ставятся его цель, а также его задачи, решение которых необходимо для достижения поставленной цели, указывается его научная новизна и практическая значимость.

**Первая глава** содержит обзор литературных источников, посвященных проблеме интеграции данных информационных систем, в котором рассматриваются наиболее широко используемые мето-



ды и модели представления баз знаний, которые применяются для решения задачи интеграции данных.

**Классификация способов интеграции информационных систем** наглядно показала актуальность задачи интеграции данных, за счет необходимости ее решения в случаях: применения систем межведомственной интеграции приложений, интеграции корпоративных приложений и технологий управления бизнес процессами; применения принципов процессно-ориентированной, сервисно-ориентированной и информационно-ориентированной интеграции; реализации интеграции на уровнях бизнес-процессов, приложений, стандартов, платформ и данных.

**Обзор способов интеграции данных** наглядно показал актуальность задачи выявления семантических конфликтов метаданных и конфликтов данных с применением онтологических спецификаций, в виду того, что: структурные конфликты метаданных не обязательно приводят к семантическим и, тогда, могут быть разрешимы; конфликты данных могут возникать и при отсутствии конфликтов метаданных и приводить к семантическим противоречиям; технологии онтологических спецификаций предоставляют механизмы для описания семантики данных и наиболее эффективной их интеграции.

**Обзор способов представления знаний** наглядно показал перспективность применения формализма дескрипционных логик и его компьютерной реализации в виде языка OWL для представления и анализа онтологических спецификаций реляционных баз данных, в виду того, что: формализмы дескрипционных логик представляют собой компромисс между выразительностью и разрешимостью, и обладают гарантирующими получение результата алгоритмами вывода знаний; язык спецификации онтологий OWL 2 DL является рекомендованным стандартом представления онтологических баз знаний, обладает наивысшей выразительностью среди разрешимых диалектов OWL; язык спецификации онтологий OWL поддерживается большим количеством специализированного программного обеспечения, которое реализует функции редактора онтологических спецификаций и/или алгоритмы логического вывода из них. **Существующие методы онтоло-**

**гического представления данных** могут быть применены лишь для решения задачи интеграции гетерогенных данных, и их применение не позволит выявить большинство семантических конфликтов метаданных и несоответствия данных.

**Инструменты онтологической интеграции данных основаны на применении базы знаний**, онтологического представления баз данных, которая не предоставляет правила выявления конфликтов данных.

Такой обзор обосновывает наличие задач диссертационного исследования и необходимость:

а) в разработке нового способа применения формализма дескрипционных логик для идентификации семантических конфликтов метаданных и несоответствия интегрируемых данных;

б) в разработке нового комплекса программ, реализующего функции выявления и разрешения семантических конфликтов метаданных и несоответствия интегрируемых данных, и построения базы знаний, которая может быть использована в современных инструментах интеграции.

**Вторая глава** содержит описание способа применения формализма дескрипционных логик для решения поставленных в первой главе задач и предлагает описание онтологической модели представления концептуальных объектов, её расширения и метода анализа, включая доказательство их применимости.

Предложена **онтологическая модель концептуальных объектов**, которая предоставляет правила описания концептуальных схем сущность-связь реляционных баз данных в виде аксиом и утверждений дескрипционной логики  $\mathcal{SHOIQ}(\mathcal{D})$ . Такие правила основаны на синтаксическом сходстве реляционных баз данных и онтологий, обоснованы их семантическим различием и направлены на обеспечение возможности выявления семантических конфликтов метаданных и несоответствий интегрируемых данных через проверку выполнимости их онтологий.

Каждый **регулярный тип сущности** представляется через набор терминологических аксиом, который определяет атомарный концепт  $A_n$  (1).

$$\{A_n \sqsubseteq T \sqcap \neg(A_{n+1} \sqcup \dots \sqcup A_m) \sqcap \forall E. \perp\} \quad (1)$$

**Слабый тип сущности и тип значения** представляются подобным образом (1), с отличием в определении для каждого концепта  $A_n$  (1) локального экзистенциального ограничения и ограничения на значения роли  $E$  (1), которое представляет зависимость существования. Тогда концепты будут принадлежать одной области интерпретации и их семантика будет сопоставима. Каждая **сущность** или **значение атрибута** представляется через утверждение, которое определяет индивид атомарного концепта (1), причем первые дополнительно объявляются как попарно различные.

Каждый **тип связи с отображением 1:n** представляется через набор терминологических аксиом, который определяет две абстрактные роли  $R_i, R_j^i$  (2) такие, что  $R_i$  сопоставляется отображению 1:1, а роль  $R_j^i$  – обратному отображению n:1.

$$\begin{aligned} & \{R_i \subseteq U, R_i \subseteq \neg R_i^-, \\ & R_j^i \subseteq U, R_j^i \subseteq \neg R_j^{i-}, R_i \equiv R_j^{i-} \\ & \top \subseteq \leq 1R_i \cap \forall R_i. A_m \cap \forall R_j^i. A_n, \\ & A_n \equiv \exists R_i. \top, A_m \equiv \exists R_j^i. \top \} \end{aligned} \quad (2)$$

**Тип связи с отображением 1:1** представляется подобным образом (2), с отличием в определении одной симметричной абстрактной роли  $R_i \subseteq R_i^-$  (2), которая в совокупности с остальными аксиомами представляет **отображение 1:1**.

Каждый **не уникальный атрибут** представляется через набор терминологических аксиом, который определяет абстрактные роли  $F_i, F_j^i$  и конкретную роль  $T_h^i$  (3), причем роли  $F_i, F_j^i$  определяются аналогично  $R_i, R_j^i$  (2).

$$\begin{aligned} & \{F_i \subseteq U, F_i \subseteq \neg F_i^-, T_h^i \subseteq U, \\ & F_j^i \subseteq U, F_j^i \subseteq \neg F_j^{i-}, F_i \equiv F_j^{i-}, \\ & \top \subseteq \leq 1F_i \cap \leq 1T_h^i \cap \forall F_i. A_m \cap \forall F_j^i. A_n \cap \forall T_h^i. D^h, \\ & A_n \equiv \exists F_i. \top, A_m \equiv \exists F_j^i. \top \cap \exists T_h^i. \top_D \} \end{aligned} \quad (3)$$

**Уникальный атрибут** представляется подобным образом (3), с отличием в определении одной симметричной абстрактной роли

$F_i \subseteq F_i^-$  (3), которая в совокупности с остальными аксиомами представляет **ограничение на уникальное отображение атрибута**.

**Ключ сущности** представляется набором терминологических аксиом, который определяет локальное ограничение на ключевые роли (2)-(3) атомарного концепта (1):  $\{HasKey(A_n F_i \sqcap \dots \sqcap F_h^i)\}$ .

**Транзитивные связи** представляются через набор терминологических аксиом, который определяет транзитивную абстрактную роль  $S$  как суперроль для всех остальных абстрактных ролей (2)-(3):  $\{S \subseteq U, S \circ S \subseteq S, R_i \subseteq S, \dots F_j \subseteq S\}$ .

Описанную в терминах дескрипционной логики совокупность терминологических аксиом  $\mathcal{T}$  и утверждений об индивидах  $\mathcal{A}$ , предлагается ставить в соответствие **концептуальной схеме сущность-связь** и называть **онтологией концептуальных объектов**  $\mathcal{K}$ :  $\mathcal{K} := \mathcal{T} \cup \mathcal{A}$ .

В качестве метода объединения онтологий концептуальных объектов, предложено **расширение онтологической модели представления концептуальных объектов**. Расширенная онтологическая модель концептуальных объектов предоставляет правила слияния терминологических аксиом и утверждений дескрипционной логики, которыми представлены концептуальные схемы для того, чтобы их можно было сравнивать. Атомарные **концепты**, атомарные **роли** и **индивиды**, соответствующие эквивалентным элементам концептуальных схем сущность-связь, **предлагается объявить эквивалентными**.

Разработан **алгоритм** построения терминологических аксиом и утверждений дескрипционной логики  $\mathcal{K}$  исходя из семантического представления схем реляционных баз данных  $CS_l, \dots, CS_h$ . Такой алгоритм в общем случае представлен на рисунке 1.

Алгоритм (рис. 1) позволяет преобразовывать схемы сущность-связь реляционных баз данных в терминах предложенной онтологической модели представления концептуальных объектов. Наихудшее время работы такого алгоритма, если не учитывать реализацию переборных и присвоить каждому его шагу стоимость и количество добавляемых для выполнения одной итерации терминологических аксиом или утверждений, составляет:

$T_1 = (a + b \times j) \times n + c \times i + (d + (e \times 2 \times j) + f \times i) \times k + 1$ , где  $a, b, c, d, e, f$  – константы, которые определяются выбранными значениями стоимости шагов алгоритма;  $n, i, j, k$  – количество типов сущности, атрибутов, типов связи и сущностей, соответственно.

**ДЛЯ КАЖДОГО** типа сущности  $\forall B_n^l \in CS_l \triangleright$

○○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарный концепт  $A_n$ , через набор терминологических аксиом вида (1), соответствующий *регулярному или слабому типу сущности*

○○○○ **ДЛЯ КАЖДОГО** атрибута типа сущности  $\forall H_i^l: B_n^l \rightarrow V_m^l \in CS_l \triangleright$

○○○○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарный концепт  $A_m$ , через набор терминологических аксиом вида (1), соответствующий *типу значения*

○○○○○○ **ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарные роли  $F_i, \dots, T_h^i$ , через набор терминологических аксиом вида (3), соответствующий *уникальному или не уникальному атрибуту*  $\triangleleft$

○○○○ **ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** ключевые роли  $F_i, \dots, T_h^i$  концептов  $A_n, A_m$ , через терминологические аксиомы вида  $HasKey(A_n F_i \sqcap \dots \sqcap F_j^i)$ ,  $HasKey(A_m T_h^i)$  соответствующие ключу сущности  $H_k^l = \{H_i^l, \dots, H_j^l\} \triangleleft$

**ДЛЯ КАЖДОГО** типа связи  $\forall G_i^l: B_n^l \times B_m^l \in CS_l \triangleright$

○○○○**ВЫПОЛНИТЬ ОПРЕДЕЛИТЬ** атомарные роли  $R_i, \dots, R_j^i$ , через набор терминологических аксиом вида (2), соответствующий *отображению 1:1 или 1:n*  $\triangleleft$

*Рис. 1. Псевдокод алгоритма построения терминологических аксиом и утверждений дискреционной логики исходя из семантического представления схемы реляционной базы данных*

Разработан метод проверки согласованности концептуальных схем, который позволяет свести процесс выявления семантических конфликтов метаданных и несоответствия данных, возникающих при их интеграции, к решению алгоритмических проблем согласованности и классификации терминологии, и согласованности онтологии. *Следует отметить, что под согласованностью онтологии или терминологии предлагается понимать наличие*

*модели данной онтологии или терминологии, в которой все концепты и роли не пусты.* Такой метод предлагает выявление **конфликтов атрибут-атрибут, атрибут-сущность, сущность-связь, поведения, наследования и ключа**, которые содержат концептуальные схемы  $CS_l, \dots, CS_h$ , через решение алгоритмической проблемы согласованности их терминологии  $\mathcal{T}_g \supseteq \mathcal{T}_l \cup \mathcal{T}_h$ , описанной в терминах предложенной модели.

Одним из возможных проявлений **конфликта атрибут-сущность**, по определению онтологической модели концептуальных объектов, может быть несоответствие терминологических аксиом, при котором в качестве доменов эквивалентных атомарных ролей  $R_i^l, R_i^h : \mathcal{T}_g \cup \{\exists R_i^l. T \sqsubseteq C_i^l, \exists R_i^h. T \sqsubseteq C_i^h, R_i^l \equiv R_i^h\}$  в неявном виде заданы непересекающиеся концепты  $C_i^l, C_i^h : C_i^l \equiv A_n^l \in \mathcal{T}_l$  и  $C_i^h \equiv A_m^h \in \mathcal{T}_h$ , или  $C_i^l \equiv A_n^l \sqcup A_u^l \in \mathcal{T}_l$  и  $C_i^h \equiv A_m^h \in \mathcal{T}_h$ , или  $C_i^l \equiv A_n^l \sqcup A_u^l \in \mathcal{T}_l$  и  $C_i^h \equiv A_m^h \sqcup A_w^h \in \mathcal{T}_h$ , при условии:  $A_m^l \sqsubseteq \neg A_n^l \sqcap \neg A_u^l \sqcap \neg A_w^l \in \mathcal{T}_l$  и  $A_w^l \sqsubseteq \neg A_n^l \sqcap \neg A_m^l \sqcap \neg A_u^l \in \mathcal{T}_l$ ,  $A_m^l \equiv A_m^h \in \mathcal{T}_g$  и  $A_w^l \equiv A_w^h \in \mathcal{T}_g$ .

**Предложение.** Пусть в терминологии  $\mathcal{T}_g \supseteq \mathcal{T}_l \cup \mathcal{T}_h$  определены концепты  $A_n^l, A_m^l, A_u^l, A_w^l, C_i^l, \exists R_i^l. T \in \mathcal{T}_l$  и  $A_n^h, A_m^h, A_u^h, A_w^h, C_i^h, \exists R_i^h. T \in \mathcal{T}_h$ :  $A_m^l \sqsubseteq \neg A_n^l \sqcap \neg A_u^l \sqcap \neg A_w^l \in \mathcal{T}_l$  и  $A_w^l \sqsubseteq \neg A_n^l \sqcap \neg A_m^l \sqcap \neg A_u^l \in \mathcal{T}_l$ ,  $A_m^l \equiv A_m^h \in \mathcal{T}_g$  и  $A_w^l \equiv A_w^h \in \mathcal{T}_g$ ,  $\exists R_i^l. T \sqsubseteq C_i^l \in \mathcal{T}_l$  и  $\exists R_i^h. T \sqsubseteq C_i^h \in \mathcal{T}_h$ , роли  $R_i^l \in \mathcal{T}_l$  и  $R_i^h \in \mathcal{T}_h$ :  $R_i^l \equiv R_i^h \in \mathcal{T}_g$  и имеет место одно из следующих определений:

- а)  $C_i^l \equiv A_n^l \in \mathcal{T}_l$  и  $C_i^h \equiv A_m^h \in \mathcal{T}_h$ ;
- б)  $C_i^l \equiv A_n^l \sqcup A_u^l \in \mathcal{T}_l$  и  $C_i^h \equiv A_m^h \in \mathcal{T}_h$ ;
- в)  $C_i^l \equiv A_n^l \sqcup A_u^l \in \mathcal{T}_l$  и  $C_i^h \equiv A_m^h \sqcup A_w^h \in \mathcal{T}_h$ ;

тогда терминология  $\mathcal{T}_g$  не согласована.

Подобным образом, сформулирована и доказана применимость метода для выявления следующих типов конфликтов: **атрибут-атрибут, атрибут-сущность, сущность-связь, поведения, наследования, ключа, типов данных, допустимых значений, разрешенных значений, формата значений, представления значений и актуальности значений.**

**Третья глава** содержит описание комплекса программ  $C_3S$ , которое реализует предложенные модель, метод и алгоритм, что позволяет выявлять проблемы интеграции концептуальных схем реляционных баз данных.

Разработана программа онтологического представления концептуальных схем сущность-связь в качестве **программного компонента  $C_2O$** , который реализует алгоритм преобразования таких схем реляционных баз данных в терминологические аксиомы и утверждения по правилам онтологической модели представления концептуальных объектов.  $C_2O$  принимает в качестве входных данных представленные по стандарту XMI (ISO/IEC 19509) UML диаграммы, конвертирует их и выдает, в качестве выходных данных, представленные по стандарту OWL 2.0 онтологии.  $C_2O$  выполнено на языке Java и использует технологии из стека Java SE8, а также фреймворк для работы с онтологиями OWL API 3.4.10. Разработан плагин RCP к Eclipse, в качестве **программного компонента  $C_2OP$** , который позволяет выполнять множественное преобразование нескольких концептуальных схем реляционных баз данных.  $C_2OP$  позволяет пользователю взаимодействовать с компонентом  $C_2O$ , за счет предоставления интерфейсов взаимодействия с ним. Спроектирована **архитектура комплекса программ  $C_3S$** , которая реализует предложенные алгоритм преобразования концептуальных схем реляционных баз данных и метод выявления их семантических конфликтов и несоответствия данных, посредством интеграции программных компонентов  $C_2O$ ,  $C_2OP$  Parugus, OWL API, Protégé и Hermit в Eclipse Modelling Tools. Все эти компоненты распространяются в свободно в соответствии с лицензиями MPL, CPL и EPL. Архитектура  $C_3S$  отображена на рисунке 2.

Проверка согласованности концептуальных схем комплексом программ  $C_3S$  может быть представлена четырьмя ключевыми этапами (рис. 2):

1. Получение концептуальных схем сущность-связь интегрируемых реляционных баз данных или, в качестве альтернативы, описание таких схем при помощи среды моделирования Parugus.

2. Конвертация концептуальных схем сущность-связь в онтологию посредством выбора таких схем и выполнения программного компонента  $C_2O$  в исполняемой среде  $C_2OP$ .

3. Интеграция нескольких онтологий в одну по правилам расширенной онтологической модели представления концептуальных объектов при помощи редактора онтологий Protégé.

4. Анализ онтологии, в соответствии с предложенным методом применения решения алгоритмических проблем согласованности и классификации терминологии, и согласованности онтологии для выявления семантических конфликтов метаданных и несоответствия данных, посредством вызова соответствующих функций Hermit в Protégé и анализа результатов.

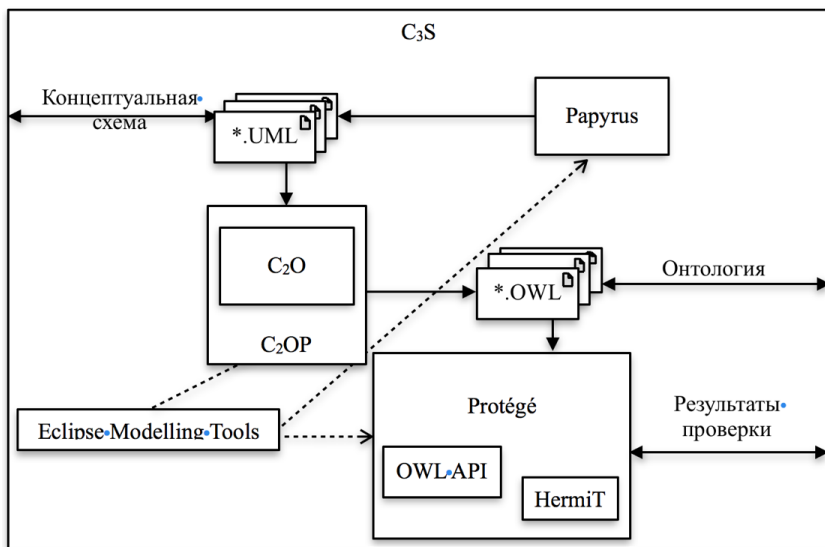


Рис. 2. Архитектура программного обеспечения  $C_3S$

**В четвертой главе** описывается экспериментальная оценка применимости предложенной онтологической модели представления концептуальных объектов и метода проверки согласованности концептуальных схем сущность-связь реляционных баз данных, для идентификации семантических конфликтов и несоответствия



интегрируемых данных., которая проведена в рамках апробации программного обеспечения C<sub>3</sub>S.

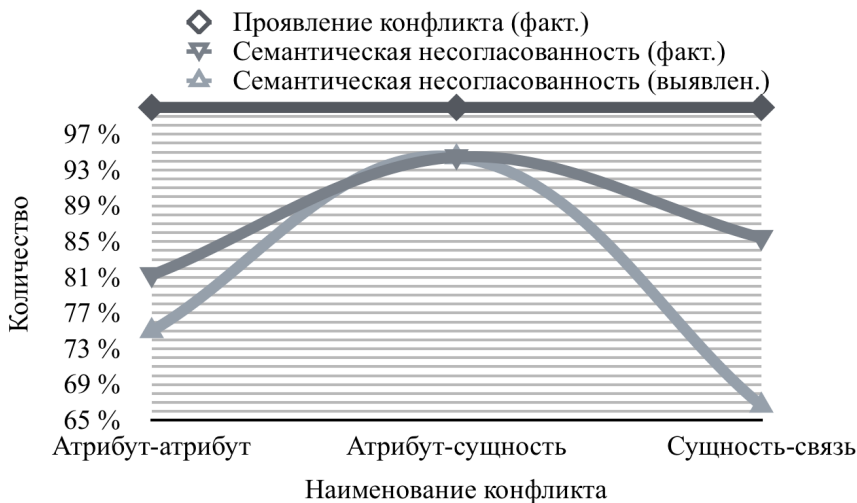
**Экспериментальная оценка** применимости предложенных моделей, алгоритма и метода проведена на основе 902 концептуальных схем баз данных, которые представляют 364 возможных случая возникновения конфликтов: атрибут-атрибут, атрибут-сущность, сущность-связь, зависимости, поведения, наследования, ключа, типов данных, допустимых значений, разрешенных значений и актуальности значений. Такая оценка включала три этапа:

1. **Подготовка опытных данных.** Такие данные представлены наборами концептуальных схем сущность-связь реляционных баз данных, каждый из которых представляет различные случаи возникновения соответствующего конфликта двух таких схем. На основании предложенных онтологической модели представления концептуальных объектов и алгоритма построения терминологических аксиом и утверждений исходя из семантического представления реляционных баз данных каждого из такого набора построены онтологии по стандарту OWL 2.0.

2. **Проведение вычислительных экспериментов.** Такие эксперименты направлены на выполнение задач табличного алгоритма на опытных данных и выявление моделируемых ими конфликтов, в соответствии с предложенным способом применения формализма дескрипционных логик для выявления семантических конфликтов и несоответствия данных при их интеграции: выполнение задач проверки согласованности и классификации терминологии для выявления конфликтов атрибут-атрибут, атрибут-сущность, сущность-связь, поведения, наследования и ключа; выполнение задачи согласованности онтологии целиком для выявления конфликтов формата значений, представления значений и актуальности значений; выполнение задач согласованности терминологии и согласованности онтологии для выявления конфликтов зависимости, типов данных, допустимых значений и разрешенных значений.

3. **Оценка результатов вычислительных экспериментов.** Такая оценка основана на сопоставлении количества реально существующих, в наборах опытных данных, семантических конфликтов

концептуальных схем сущность-связь и выявленных в ходе вычислительных экспериментов соответствующих несоответствий терминологических аксиом и/или утверждений. Результаты вычислительного эксперимента на наборах опытных данных, соответствующих конфликтам **атрибут-атрибут**, **атрибут-сущность** и **сущность-связь** показали, что в 120 возможных случаях их возникновения верно выявлена семантическая несогласованность в 81 из 99 случаев. Результаты такого вычислительного эксперимента представлены на рисунке 3.



*Рис. 3.* Результаты вычислительного эксперимента на моделях конфликтов: атрибут-атрибут, атрибут-сущность и сущность-связь

Такая оценка (рис. 3) показала, что:

1. Разработанная модель онтологического представления концептуальных объектов и алгоритм построения терминологических аксиом и утверждений исходя из семантического представления реляционных баз данных ориентированы на онтологическое моделирование только семантической составляющей таких баз, исключая структурное их представление.

2. Предложенный метод применения дескрипционного исчисления для выявления семантических конфликтов схем баз данных и несоответствия данных позволяет выявлять только семантическую несогласованность таких схем, игнорируя их структурные различия. Такая применимость метода обусловлена поставленными задачами и результатами исследования проблем интеграции, которые показали, что в виду семантической согласованности структурная гетерогенность разрешима стандартными инструментами интеграции данных.

3. Предложенный метод не может быть применен при жесткой типизации данных и не распространяется на конфликты, образованные различиями только ограничений кардинальности эквивалентных множеств связей. Такое ограничение метода обусловлено различным предположением об открытости и замкнутости мира, которое принято в онтологиях и базах данных, соответственно.

Подобным образом проведены вычислительные эксперименты на моделях конфликтов: **ключа, зависимости, наследования, формата значений, представления значений, актуальности значений, типов данных, допустимых значений и разрешенных значений.**

**В заключении** сформулированы и представлены в виде выводов основные результаты диссертационной работы.

## **ВЫВОДЫ ПО ДИССЕРТАЦИОННОЙ РАБОТЕ**

1. На основе формализма дескрипционной логики  $SHOIF(D)$  построена математическая модель онтологического представления баз данных, позволяющая выявлять их семантические конфликты и несоответствия данных. Для предложенной математической модели, разработан метод, который позволяет свести процесс выявления семантических конфликтов метаданных и несоответствия данных, возникающих при их интеграции, к решению алгоритмических проблем согласованности и классификации терминологии, и согласованности онтологии. Разработан алгоритм построения терминологических аксиом и утверждений дескрипционной логики, в соответствии с правилами онтологической модели концептуальных объектов, исходя из семантического представления схем реляцион-

ных баз данных; и рассчитано его время работы. Доказана применимость предложенных модели, алгоритма и метода для выявления следующих типов конфликтов: атрибут-атрибут, атрибут-сущность, сущность-связь, поведения, наследования, ключа, типов данных, допустимых значений, разрешенных значений, формата значений, представления значений и актуальности значений.

2. Разработан и апробирован программный компонент  $C_2O$ , который реализует алгоритм преобразования концептуальных схем сущность-связь реляционных баз данных, представленных по стандарту XMI (ISO/IEC 19509), в терминологические аксиомы и утверждения по правилам онтологической модели представления концептуальных объектов, представленных по стандарту OWL 2.0. Разработан и апробирован плагин к Eclipse  $C_2OP$ , который позволяет выполнять множественное преобразование нескольких концептуальных схем реляционных баз данных посредством выполнения  $C_2O$  и предоставляет интерфейсы взаимодействия с ним. Спроектирована архитектура комплекса программ  $C_3S$ , которая реализует предложенный алгоритм преобразования концептуальных схем реляционных баз данных и метод выявления их семантических конфликтов и несоответствия данных, посредством интеграции расширяемых свободно компонентов  $C_2O$ ,  $C_2OP$  и Protégé на базе платформы Eclipse.

3. Проведена серия вычислительных экспериментов на 902 базах данных по выявлению конфликтов: атрибут-атрибут, атрибут-сущность, сущность-связь, зависимости, поведения, типов данных, допустимых и разрешенных значений, формата значений, представления значений и актуальности значений. Результаты экспериментов показали, что предложенный способ выявления семантических конфликтов и несоответствия интегрируемых данных, основанный на решении алгоритмических проблем согласованности и классификации терминологии, и согласованности онтологии, позволяет выявить 95% случаев различного возникновения семантической несогласованности. Экспериментальным путем определена применимость предложенного способа выявления семантических конфликтов и несоответствия интегрируемых данных, которая ограничена 346 из 364 видов конфликтов, невозможна при жесткой

типизации данных и не распространяется на конфликты, образованные различиями только ограничений кардинальности эквивалентных множеств связей.

## **ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### ***Публикации в рецензируемых журналах SCOPUS***

1. Kropotin, A. A. Evaluating OWL 2 Reasoners in the Context Of Checking Entity-Relationship Diagrams During Software Development [Текст] // Informal Proceedings of the 3rd International Workshop on OWL Reasoner Evaluation (ORE 2014) co-located with the Vienna Summer of Logic (VSL 2014). – 2014. Vienna, Austria. – 13 July. – P. 44-50, – URL: <http://ceur-ws.org/Vol-1207>.

2. Kropotin, A. A. Realization of the Ontologically Based Method for Checking Structural Inconsistences of Relational Databases [Текст] / A. A. Kropotin, A. V. Grigoryev, Y. V. Bidulya, A. G. Ivashko, N. S. Durynin // Proceedings of the 27th International DAAAM Symposium on Intelligent Manufacturing and Automation 2016. – 2016. – vol. 27. Mostar, Bosnia and Herzegovina. – 26-29 October. – P. 762-767. – DOI 10.2507/27th.daaam.proceedings.110.

### ***Публикации в изданиях, рекомендованных ВАК***

3. Кропотин, А. А. Применение табличного алгоритма для верификации бизнес-процессов [Текст] / А. А. Кропотин, А. Г. Ивашко, А. В. Григорьев, Е. О. Овсянникова // Вестник Тюменского государственного университета. – 2013. – № 7. С. 202-213.

4. Кропотин, А. А. Применение формализма дескрипционной логики для выявления семантических конфликтов концептуальных схем сущность-связь [Текст] / А. А. Кропотин // Журнал: ИНТЕЛЛЕКТ. ИННОВАЦИИ. ИНВЕСТИЦИИ. Оренбургский государственный университет. – 2016. – Вып. 7. – С. 93-98. – ISSN 2077-7175.

### ***Публикации в прочих изданиях***

5. Кропотин, А. А. Проблема поиска противоречий на диаграммах классов UML [Текст] / А. А. Кропотин, А. В. Григорьев, Е. О. Овсянникова // Научный журнал. Сборник научных трудов SWorld. Современные проблемы и пути их решения в науке, транспорте, производстве и образовании'2012. – 2012. – Том 14. Одесса. – С. 3-10.

6. Кропотин, А. А. Проблема формирования онтологии терминов бизнес – процесса организации [Текст] / А. А. Кропотин, С. И. Коломиец // Современные проблемы математического и информационного моделирования. Перспективы разработки и внедрения инновационных IT – решений. Издательство Тюменского государственного университета. – 2012. – С. 103-107.

7. Kropotin, A. A. Database Schema Method for Automatic Semantic Errors Resolving During Information Systems Integration [Текст] / A. A. Kropotin, A. V. Grigoryev, A. G. Ivashko // Informal Proceedings of the 2016 10th IEEE International Conference on Application of Information and Communication Technologies, AICT2016, – 2016. Azerbaijan, Baku. – 12-14 October.

### ***Свидетельства о регистрации программ для ЭВМ***

8. Кропотин А. А. Свидетельство о государственной регистрации программы для ЭВМ №2017610866 «Программа онтологического представления концептуальных схем сущность-связь» от 18.01.2017.

9. Кропотин А. А. Свидетельство о государственной регистрации программы для ЭВМ №2017614587 «Плагин RCP Eclipse ‘С2OP’» от 19.04.2017.

Подписано в печать 28.09.2017. Тираж 130 экз.  
Объем 1,0 уч.-изд. л. Формат 60×84/16. Заказ 627.

---

Издательство Тюменского государственного университета  
625003, г. Тюмень, ул. Семакова, 10.  
Тел./факс (3452) 59-74-81; 59-74-68  
E-mail: izdatelstvo@utmn.ru