

*На правах рукописи*



**ГРИГОРЬЕВ Андрей Викторович**

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ И АЛГОРИТМЫ ОПРЕДЕЛЕНИЯ  
СОГЛАСОВАННОСТИ БАЗ ЗНАНИЙ**

**Специальность 05.13.18 - Математическое моделирование,  
численные методы и комплексы программ**

**АВТОРЕФЕРАТ**

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

Тюмень - 2013

Работа выполнена на кафедре информационных систем Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Тюменский государственный университет»

**Научный**

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

**Официальные  
оппоненты:**

**Гуц Александр Константинович** доктор физико-математических наук, профессор, ФГБОУ ВПО «Омский государственный университет им. Ф.М. Достоевского», заведующий кафедрой кибернетики  
**Захаров Сергей Дмитриевич** кандидат физико-математических наук, доцент, ФГБОУ ВПО «Тюменский государственный университет», доцент кафедры программного обеспечения

**Ведущая**

**организация:** ФГБОУ ВПО «МАТИ – Российский государственный технологический университет имени К.Э. Циолковского»

Защита диссертации состоится «20» декабря 2013 года, в «16-00» на заседании диссертационного совета Д 212.274.14 при ФГБОУ ВПО «Тюменский государственный университет» по адресу 625003, г. Тюмень, ул. Перекопская 15а, ауд. 410.

С диссертацией можно ознакомиться в Информационно-библиотечном центре ФГБОУ ВПО «Тюменский государственный университет».

Автореферат разослан «19» ноября 2013 года.

Ученый секретарь

диссертационного совета



Е.А. Оленников

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования.** Исследования в области разработки методов определения согласованности баз знаний обусловлены необходимостью решения таких практически значимых задач как семантический поиск, классификация понятий предметной области или поддержка принятия решений в экспертных системах. Для решения задач такого рода успешно применяются методы, основанные на автоматах (D. Calvanese, KRDB Research Center в Университете г. Бозен-Больцано; T. Eiter, Институт информационных систем университета г. Вена), методы, основанные на правилах вывода (E. Kazakov, Вычислительная лаборатория Оксфордского университета), метод резолюций (B. Motik, университет г. Манчестер) и табличные методы (I. Horrocks, университет г. Манчестер). Основная проблема таких методов заключается в том, что все они имеют экспоненциальную вычислительную сложность, поэтому их применение в задачах реального времени весьма затруднительно для больших баз знаний (более 5000 аксиом).

Для того, чтобы снизить время работы представленных методов на практике, учеными предпринимается огромное количество попыток разработки новых алгоритмов. Такие разработки ежегодно представляются на международных конференциях по применению логических формализмов в искусственном интеллекте «Description Logic Workshop» и «International Joint Conference On Automated Reasoning». Наиболее значимыми разработками, представленными на данных форумах, являются: алгоритм кэширования (Y. Ding, V. Haarslev), backJumping-алгоритм (I. Horrocks), алгоритм глобального кэширования (L. A. Nguyen), которые, зачастую, основываются на методах теории множеств и методах логики первого порядка, кроме того ни одна из этих разработок не направлена на снижение теоретической оценки вычислительной сложности.

**Целью** исследования является разработка алгоритмов проверки согласованности баз знаний, сокращающих вычислительную сложность табличных методов на определенных классах баз знаний.

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

1. Построение сетевой модели базы знаний, описанной выражениями дескрипционной логики.
2. Разработка метода нахождения максимального потока минимальной стоимости в разработанной сетевой модели базы знаний.

3. Реализация программного комплекса для поддержки деятельности инженера по знаниям, реализующего разработанные методы и алгоритмы.
4. Проведение вычислительного эксперимента для проверки адекватности разработанных моделей и методов.

**Объектом исследования** диссертационного исследования являются базы знаний.

**Предметом исследования** являются процессы определения согласованности описания баз знаний.

**Методология и методы исследования.** При исследовании применялись методы теории множеств, теории графов, теории методов оптимизации.

**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. Разработана новая математическая модель баз знаний, в которой для моделирования понятий баз знаний впервые были использованы сетевые модели, что позволяет применять алгоритмы поиска максимального потока минимальной стоимости для задач проверки согласованности баз знаний.
2. Разработан новый метод проверки согласованности базы знаний для классов изоморфных концептов, позволяющий сократить количество переборов в табличном методе.

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

3. Предложен новый численный метод нахождения максимального потока минимальной стоимости в разработанных сетевых моделях баз знаний. Вычислительная сложность разработанного алгоритма составила  $O(V^2)$ .

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

4. Разработан программный комплекс для поддержки деятельности инженера по знаниям, позволяющий проводить классификацию, проверку согласованности и выполнимость концептов. Описание базы знаний реализовано с помощью языка OWL 2.
5. Проведен вычислительный эксперимент, использующий 200 опубликованных баз знаний, для которых выполнялись задачи классификации и проверки согласованности. Результаты компьютерных экспериментов показали адекватность разработанных в диссертации математических моделей и методов.

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.

**Научная новизна исследования** заключается в следующем:

1. Разработана и обоснована сетевая модель базы знаний, представленной аксиомами предметной области, что дает возможность использовать математические методы и алгоритмы, применяемые в теории графов;
2. Для сетевой модели базы знаний разработан новый численный метод нахождения максимального потока минимальной стоимости с вычислительной сложностью  $O(V^2)$ . Доказано, что если максимальный поток минимальной стоимости графа, ассоциированного с конъюнкцией двух концептов, имеет положительный вес, то эта конъюнкция не согласована.
3. Модифицирован табличный метод проверки согласованности базы знаний, в котором обосновано, что для изоморфных концептов допускается не проводить повторную проверку;
4. Модифицирован алгоритм нахождения автоморфизмов в деревьях (J. Edmonds) для поиска изоморфных концептов базы знаний.

**Практическая значимость работы.** Для применения разработанных методов и алгоритмов был реализован кроссплатформенный программный комплекс поддержки деятельности инженера по знаниям. Программный комплекс позволяет создавать, импортировать и интегрировать базы знаний

на языке OWL 2, выполнять проверку согласованности базы знаний, классификацию понятий и проверку их выполнимости.

**Достоверность** изложенных в работе результатов подтверждается корректным использованием математического аппарата дескрипционной логики, теории графов, языка OWL 2 и сравнением полученных результатов с результатами авторов исследований в этой области.

**Апробация результатов.** Основные результаты исследования докладывались на:

1. II международной конференции «OWL Reasoner Evaluation Workshop (ORE 2013)» (Германия, Ульм, 22 июля 2013);
2. Международной научно-практической конференции «Современные проблемы и пути их решения в науке, транспорте, производстве и образовании'2012» (Украина, Одесса, 18-27 декабря 2012);
3. XIV международной научно-практической конференции «Современные проблемы гуманитарных и естественных наук» (Москва, 26-27 марта 2013 г.);
4. Научно-технической конференции «Современные проблемы математического и информационного моделирования. Перспективы разработки и внедрения инновационных IT-решений» (Тюмень, 13-15 апреля 2012 г.)

**Публикации.** Основные результаты диссертации опубликованы в 9 работах, список которых приведен в конце автореферата, в т.ч. 3 статьи в рецензируемых журналах, а также получено 1 свидетельство о государственной регистрации программ для ЭВМ.

**Структура и объем диссертации.** Диссертация состоит из введения четырех глав, заключения и списка литературы. Общий объем работы составляет 115 страниц, включает 26 рисунков и 18 таблиц. Список литературы содержит 111 наименований.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

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

В **первой главе** рассматриваются наиболее широко используемые методы и модели представления знаний в информационных системах, которые делятся на два класса: подходы, основанные на графах и подходы, основанные на формальной логике.

Обзор подходов основанных на графах, показал, что два наиболее эффективных, с точки зрения осуществления вывода знаний, формализма: семантические сети и фреймы имеют ряд недостатков при их использовании на практике, основной из которых заключается в том, что они не всегда однозначно отражают представляемые знания. Кроме того, для описания знаний необходимо использовать специализированные инструменты. В связи с этим существует немного разработок использующих данные формализмы. В отличие от подходов, основанных на графах, в подходах, основанных на формальной логике, существует множество формализмов, которые успешно применяются в различных предметных областях. Формализм дескрипционных логик представляющий основу языка OWL 2, который является стандартом описания баз знаний, рекомендованным консорциумом W3C, успешно применяется при решении задач семантического поиска и классификации понятий предметных областей; решение таких задач получило практическое применение в медицинских экспертных системах. Язык OWL 2 характеризуется однозначностью представления знаний и использованием точных и быстрых методов вывода знаний. На основе этих фактов делается вывод о том, что наиболее перспективным подходом для решения практических задач является язык OWL 2 и дескрипционные логики.

На основе сделанного вывода в оставшейся части главы рассматриваются методы осуществления знаний, на основе которых будут разрабатываться методы проверки согласованности баз знаний. Наиболее эффективными, с точки зрения вычислительной сложности, методами логического анализа являются табличные методы. Однако, все алгоритмы реализующие табличные методы имеют экспоненциальную вычислительную сложность, и для того чтобы сократить время работы на практике разрабатывается большое количество эвристик. Такие эвристики, зачастую, используют методы теории множеств и методы логики первого порядка, и не направлены на снижение вычислительной сложности. Эти факты позволяют сделать вывод о возможности применения методов теории графов для осуществления задач проверки согласованности баз знаний, что позволит

понизить вычислительную сложность табличных методов на определенных базах знаний.

**Глава 2** содержит описание разработанной математической модели. Математическая модель определяет схему построения сети описывающей пересечение двух классов предметной области; такая сеть строится посредством преобразования формул дескрипционной логики в термины теории графов. Схема построения графа определяется с помощью функции  $P(C, D)$ , которая ставит в соответствие двум концептам взвешенную сеть  $G < V, E, C(E), W(E) >$ . Множество вершин в сети, представленной отображением  $P(C, D)$  обозначается  $P_V(C, D)$ , множество ребер сети –  $P_E(C, D)$ , символом  $P_C(C, D)$  обозначим функцию отображающую множество ребер  $P_E(C, D)$ , во множество натуральных чисел, такую функцию назовем пропускной способностью; символом  $P_W(C, D)$  обозначим функцию отображающую множество ребер  $P_E(C, D)$ , во множество целых чисел, такую функцию назовем весовой.

Для двух концептов, являющимися простыми концептами правила определяются формулами (1)-(5), формула (5) используется вместо формулы (3) в тех случаях, когда во множестве аксиом  $T$  не существует аксиом  $C \sqsubseteq \neg D$  и  $C \equiv \neg D$ .

$$P_V(C, \neg C) = \{v_k^{C,D}, s\} \quad (1)$$

$$P_E(C, \neg C) = \{(s, v_k^{C,D})\} \quad (2)$$

$$P_C(C, \neg C) = \{(c((s, v_k^{C,D})) = 0)\} \quad (3)$$

$$P_W(C, \neg C) = \{(w((s, v_k^{C,D})) = 0)\} \quad (4)$$

$$P_C(C, D) = \{(c((s, v_k^{C,D})) = 1)\} \quad (5)$$

Формулы (6)-(9) определяют схему построения графа для двух концептов являющихся кванторами существования и всеобщности, при условии эквивалентности ролей.

$$P_V(\exists R. C', \forall R. D') = \{v_k^{C,D}\} \cup P_V(C', D') \quad (6)$$

$$P_E(\exists R. C', \forall R. D') = P_E(C', D') \cup (v_k^{C',D'}, v_k^{C,D}) \quad (7)$$

$$P_C(\exists R. C', \forall R. D') = \{(c((v_k^{C',D'}, v_k^{C,D})) = 1) \cup P_C(C', D')\} \quad (8)$$

$$P_W(\exists R. C', \forall R. D') = \left\{ \left( w \left( \left( v_k^{C', D'}, v_k^{C, D} \right) \right) = 0 \right) \cup P_C(C', D') \right\} \quad (9)$$

Формулы (10)-(13) определяют схему построения графа для двух концептов, один из которых является дизъюнкцией, а второй конъюнкцией.

$$\begin{aligned} & P_V(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcup \dots \sqcup D_M) \quad (10) \\ & = \{v_k^{C, D}\} \cup \bigcup_{i=1..N, j=1..M} (P_V(C_i, D_j) \cup v_{ij}^{C, D} \cup u_j^{C, D}) \end{aligned}$$

$$\begin{aligned} & P_E(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcup \dots \sqcup D_M) \quad (11) \\ & = \{(s, v_k^{C, D})\} \\ & \cup \bigcup_{i=1..N, j=1..M} ((P_E(C_i, D_j) \cup (v_k^{C_i, D_j}, v_{ij}^{C, D})) \\ & \cup (v_{ij}^{C, D}, u_j^{C, D}) \cup (u_j^{C, D}, t) \cup (u_j^{C, D}, v_k^{C, D})) \end{aligned}$$

$$\begin{aligned} & P_C(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcup \dots \sqcup D_M) \quad (12) \\ & = \left\{ \left( c \left( \left( v_k^{C_i, D_j}, v_{ij}^{C, D} \right) \right) = 1 \right) \right. \\ & \cup \left( c \left( \left( v_{ij}^{C, D}, u_j^{C, D} \right) \right) = 1 \right) \cup \left( c \left( \left( u_j^{C, D}, v_k^{C, D} \right) \right) = 1 \right) \\ & \cup \left( c \left( \left( u_j^{C, D}, t \right) \right) = N - 1 \right) \\ & \left. \cup \bigcup_{i=1..N, j=1..M} P_C(C_i, D_j) \right\} \end{aligned}$$

$$\begin{aligned} & P_W(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcup \dots \sqcup D_M) \quad (13) \\ & = \left\{ \left( w \left( \left( v_k^{C_i, D_j}, v_{ij}^{C, D} \right) \right) = 0 \right) \right. \\ & \cup \left( w \left( \left( v_{ij}^{C, D}, u_j^{C, D} \right) \right) = 0 \right) \\ & \cup \left( w \left( \left( u_j^{C, D}, v_k^{C, D} \right) \right) = 0 \right) \cup \left( c \left( \left( u_j^{C, D}, t \right) \right) = -1 \right) \\ & \left. \cup \bigcup_{i=1..N, j=1..M} P_C(C_i, D_j) \right\} \end{aligned}$$

Формулы (14)-(17) определяют схему построения графа для двух концептов являющимися конъюнкциями.

$$P_V(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcap \dots \sqcap D_M) \quad (14)$$

$$= \{v_k^{C,D}\} \cup \bigcup_{i=1..N, j=1..M} (P_V(C_i, D_j) \cup v_{ij}^{C,D})$$

$$P_E(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcap \dots \sqcap D_M) \quad (15)$$

$$= \bigcup_{i=1..N, j=1..M} ((P_E(C_i, D_j) \cup (v_k^{C_i, D_j}, v_{ij}^{C,D}))$$

$$\cup (v_{ij}^{C,D}, v_k^{C,D}) \cup (v_k^{C,D}, t))$$

$$P_C(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcap \dots \sqcap D_M) \quad (16)$$

$$= \left\{ \begin{aligned} & (c((v_k^{C_i, D_j}, v_{ij}^{C,D})) = 1) \\ & \cup (c((v_{ij}^{C,D}, v_k^{C,D})) = 1) \\ & \cup (c((v_k^{C,D}, t)) = NM - 1) \\ & \cup \bigcup_{i=1..N, j=1..M} P_C(C_i, D_j) \end{aligned} \right\}$$

$$P_W(C_1 \sqcap \dots \sqcap C_N, D_1 \sqcap \dots \sqcap D_M) \quad (17)$$

$$= \left\{ \begin{aligned} & (w((v_k^{C_i, D_j}, v_{ij}^{C,D})) = 0) \cup (w((v_{ij}^{C,D}, v_k^{C,D})) = 0) \\ & \cup (w((v_k^{C,D}, t)) = -1) \cup \bigcup_{i=1..N, j=1..M} P_W(C_i, D_j) \end{aligned} \right\}$$

Формулы (18)-(21) определяют схему построения графа для двух концептов являющимися дизъюнкциями.

$$P_V(C_1 \sqcup \dots \sqcup C_N, D_1 \sqcup \dots \sqcup D_M) \quad (18)$$

$$= \{v_k^{C,D}\} \cup \bigcup_{i=1..N, j=1..M} (P_V(C_i, D_j) \cup v_{ij}^{C,D})$$

$$\begin{aligned}
& P_E(C_1 \sqcup \dots \sqcup C_N, D_1 \sqcup \dots \sqcup D_M) \\
&= \{(s, v_k^{C,D})\} \\
&\cup \bigcup_{i=1..N, j=1..M} ((P_E(C_i, D_j) \cup (v_k^{C_i, D_j}, v_{ij}^{C,D})) \\
&\cup (v_{ij}^{C,D}, v_k^{C,D}))
\end{aligned} \tag{19}$$

$$\begin{aligned}
& P_C(C_1 \sqcup \dots \sqcup C_N, D_1 \sqcup \dots \sqcup D_M) \\
&= \left\{ (c((s, v_k^{C,D})) = 1) \cup (c((v_k^{C_i, D_j}, v_{ij}^{C,D})) = 2) \right. \\
&\cup \left. (c((v_{ij}^{C,D}, v_k^{C,D})) = 1) \cup \bigcup_{i=1..N, j=1..M} P_C(C_i, D_j) \right\}
\end{aligned} \tag{20}$$

$$\begin{aligned}
& P_W(C_1 \sqcup \dots \sqcup C_N, D_1 \sqcup \dots \sqcup D_M) \\
&= \left\{ (w((s, v_k^{C,D})) = 0) \right. \\
&\cup (w((v_k^{C_i, D_j}, v_{ij}^{C,D})) = 0) \\
&\cup (w((v_{ij}^{C,D}, v_k^{C,D})) = 0) \\
&\cup \left. \bigcup_{i=1..N, j=1..M} P_W(C_i, D_j) \right\}
\end{aligned} \tag{21}$$

Математической моделью двух концептов  $C$  и  $D$ , назовем взвешенную сеть  $G < V, E, C(E), W(E) >$ , для которой:

$$\begin{aligned}
& \{s, t\} \subseteq V \\
& V = V \cup P_V(C, D) \\
& E = E \cup P_E(C, D) \\
& C = C \cup P_C(C, D) \\
& W = W \cup P_W(C, D)
\end{aligned}$$

$\forall v \in V: deg^-(v) = 0$  или  $deg^-(v) = N$  и  $|e(v, t)| = N$ , добавить в граф новое ребро  $e(v, t)$  с пропускной способностью 1 и весом  $V$ , обозначим такое ребро символом  $r$ .

Для каждой вершины  $v \in V$ , определим множество  $in(v)$ , для которого  $q \in in(v)$  если  $q \in V, (q, v) \in E$ ; аналогичным образом определим  $out(v)$ , для которого  $q \in out(v)$  если  $q \in V, (v, q) \in E$ .

Корректность использования такого графа для проверки непересекаемости двух концептов доказывается в теореме 1.

**Теорема 1.** Если  $L$  максимальный поток в графе  $P(C, D)$  и он имеет отрицательный вес, то концепты  $C$  и  $D$  не пересекаются.

Задача нахождения максимального потока минимальной стоимости может быть выполнена путем последовательного решения двух задач линейного программирования. При решении первой задачи выраженной системой линейных ограничений (22) и целевой функцией (23), будет найдена величина максимального потока.

$$\left\{ \begin{array}{l} \forall e \in E: l(e) \geq 0 \\ \forall e \in E: l(e) \leq f(e) \\ \forall v \in V/\{s, t\}: \sum_{q \in \text{in}(v)} l(q) = \sum_{w \in \text{out}(v)} l(w) \\ \sum_{q \in \text{in}(t)} l(q) = \sum_{w \in \text{out}(s)} l(w) \end{array} \right. \quad (22)$$

$$V = \sum_{q \in \text{in}(t)} l(q) \rightarrow \max \quad (23)$$

Затем необходимо решить вторую задачу линейного программирования, выраженную системой линейных ограничений (24) и целевой функцией (25), в которой необходимо использовать найденную величину  $V$  максимального потока.

$$\left\{ \begin{array}{l} \forall e \in E: l(e) \geq 0 \\ \forall e \in E: l(e) \leq f(e) \\ \forall v \in V/\{s, t\}: \sum_{q \in \text{in}(v)} l(q) = \sum_{w \in \text{out}(v)} l(w) \\ \sum_{q \in \text{in}(t)} l(q) = \sum_{w \in \text{out}(s)} l(w) \\ \sum_{q \in \text{in}(t)} l(q) = V \end{array} \right. \quad (24)$$

$$\sum_{q \in E} w(q) * l(q) \rightarrow \min \quad (25)$$

Решать задачи линейного программирования можно с использованием классических методов (например, симплекс-метод) или специализированных

методов теории графов (метод Форда-Фалкерсона). Однако, минимальная вычислительная сложность среди этих методов пропорциональна полиному 4ой степени ( $O(N^4)$ ). Такая вычислительная сложность является не приемлемой для задачи проверки непересекаемости концептов. Исходя из того, что представленная схема строит сети определенной структуры, был предложен алгоритм, который осуществляет поиск максимального потока минимальной стоимости с вычислительной сложностью  $O(N^2)$ . Алгоритм нахождения максимального потока минимальной стоимости проходит следующие шаги:

**ШАГ 1.** Вычисляется величина потокового потенциала вершины  $s$ , равна сумме пропускных способностей исходящих ребер (26).

$$FP(s) = \sum_{u \in out(s)} c(s, u) \quad (26)$$

**ШАГ 2.** Вершины графа сортируются в порядке увеличения пути от вершины  $s$ .

**ШАГ 3.** В порядке определенном увеличению высоты, вычисляется потоковый потенциал каждой вершины равный (27).

$$FP(w) = \sum_{u \in in(w)} \min(c(u, w), FP(u)) \quad (27)$$

**ШАГ 4.** Для каждого ребра  $(u, w)$  инцидентного вершине  $w$ , вычисляется величина  $L$ , по формуле (28).

$$L(u, w) = \min(c(u, w), FP(u)) \quad (28)$$

**ШАГ 5.** Если существует ребро  $(w, t)$  инцидентное вершине  $w$ , то в порядке увеличения стоимости обновляются значения потоковых потенциалов вершин  $t$  и  $w$ , и поток по ребру  $(w, t)$  по формулам (29)-(31).

$$FP(t) = FP(t) + \min(c(w, t), FP(w)) \quad (29)$$

$$FP(w) = FP(w) - \min(c(w, t), FP(w)) \quad (30)$$

$$L(w, t) = \min(c(w, t), FP(w)) \quad (31)$$

**ШАГ 6.** Если ребро  $(w, t)$  имеет отрицательный вес, и  $L(w, t) > 0$ , то вес потока устанавливается равным -1.

Корректность работы такого алгоритма доказывается в теореме 2. Оценка вычислительной сложности данного алгоритма представлена в теореме 3.

**Теорема 2.**  $FP(t)$  является величиной максимального потока в графе  $P(C, D)$  и значение  $L(w, t)$  является величиной максимального потока минимальной стоимости по ребру  $(w, t)$ .

**Теорема 3.** Вычислительная сложность представленного алгоритма в худшем случае пропорциональна  $O(V^2)$ .

Разработанная математическая модель и метод нахождения максимального потока, могут быть применены для сокращения случаев недетерминированного выбора при обработке дизъюнкций; также данный метод применим непосредственно при проверке вложенности двух концептов.

Разработанная модель также применяется и для поиска сходных структур в выражениях дескрипционной логики, определяющих классы предметной области. Нахождение и установление изоморфизма двух концептов между собой или автоморфизма концепта в самого себя, позволит сократить пространство поиска интерпретаций табличным методом. Корректность данного предложения доказывается в лемме 1.

**Предложение 1.** Если для множества литералов  $C_N$  концепта  $C$  существует автоморфизм  $f$  и  $I$  - интерпретация данного концепта, то концепт  $C$  будет выполнен на интерпретации  $I^f$  тогда и только тогда, когда он будет выполнен на интерпретации  $I$ .

Поиск изоморфизмов между концептами или автоморфизмов в концепте осуществляется с помощью модифицированного метода Эдмондсона.

**Глава 3** содержит описание программного комплекса реализующего разработанные методы и алгоритмы.

Разработанные методы и алгоритмы нашли свое применение в программном комплексе поддержки деятельности инженера по знаниям TReasoner, который позволяет выполнять следующие задачи:

1. Описывать знания в виде онтологий;
2. Объединять разработанные онтологии;

3. Осуществлять логический анализ базы знаний, включающий в себя проверку выполнимости базы знаний, проверку выполнимости отдельных понятий и классификацию понятий базы знаний.

Разработанный программный комплекс имеет трехуровневую архитектуру: уровень представлений, уровень приложений, уровень хранения данных (рис. 1). Такое решение было выбрано из-за значительных требований разработанных алгоритмов к вычислительным ресурсам.

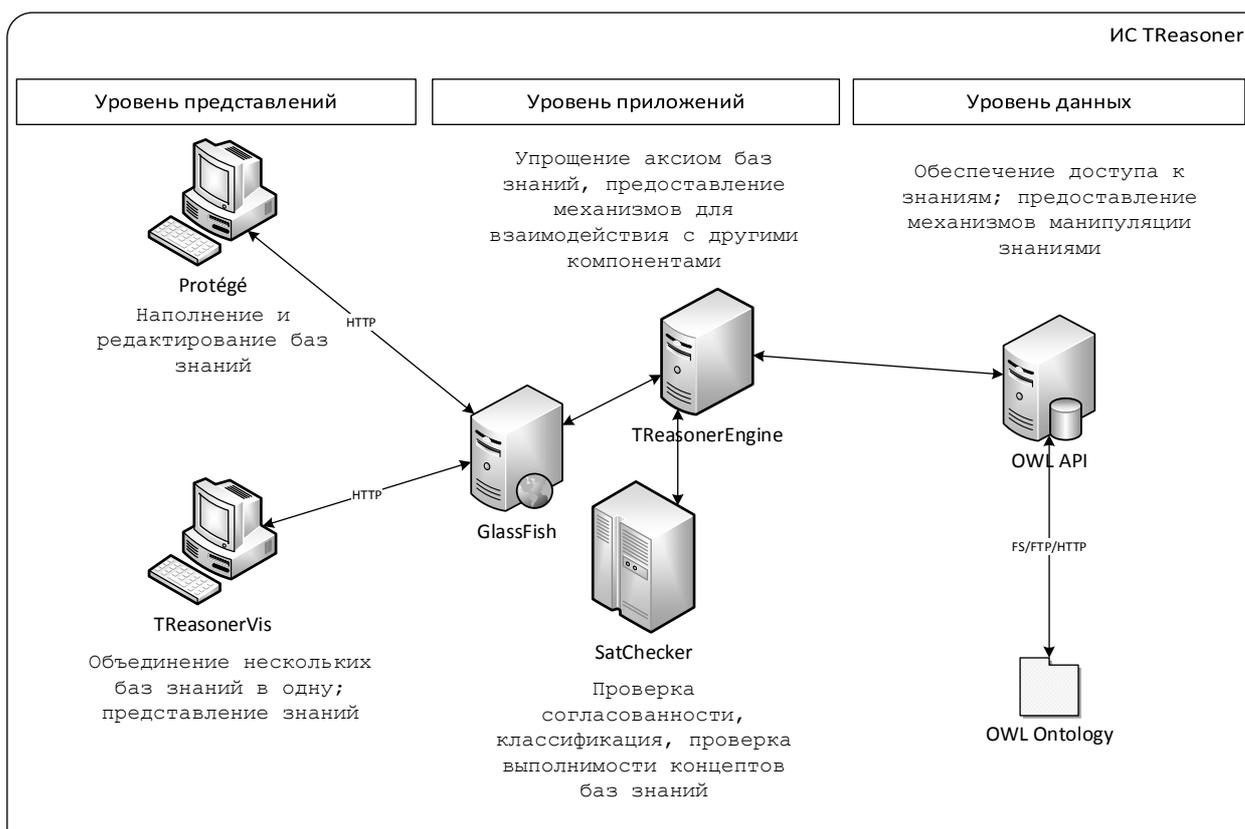


Рис. 1. Архитектура программного комплекса.

Уровень представлений отвечает за взаимодействие пользователя с системой в рамках ввода данных (формирования онтологий), запуска расчетов и отображения информации. Программы TReasonerVis и Protégé отвечающие за эти задачи разработаны на языке Java для обеспечения переносимости программного комплекса между различными системами, и взаимодействуют с уровнем представлений по протоколу HTTP. Для организации обмена данными используется web-сервер GlassFish.

На уровне приложений осуществляется выполнение разработанных алгоритмов. Все расчеты осуществляются модулем SatChecker, который реализует разработанные математические модели и методы. В рамках логического анализа база знаний проходит обработку в два этапа: первый

этап заключается в преобразовании входных данных во внутренние структуры и упрощении и сокращении аксиом, которое не повлияет на выходной результат. Для выполнения задач анализа используется алгоритм Шмидта-Шаусса-Смолки, который является реализацией табличного метода, для классификации понятий используется гибридный расширенный поперечный метод и табличный метод. Выходные данные отправляются на уровень представления в структурированном виде в формате XML-файла.

Уровень хранения данных организован файловой системой и модулем OWL API поддерживающей управление данными в формате OWL.

В главе 4 описываются результаты вычислительных экспериментов, которые позволяют оценить адекватность разработанных моделей и методов.

В качестве входных данных эксперимента были использованы базы знаний из трех репозиториях баз знаний: NCBO BioPortal, Oxford Ontology Library, Manchester Ontology Repository. Из всех онтологий данных хранилищ были отобраны только онтологии OWL 2 трех различных профилей, которые содержали не менее 100 аксиом и 10 имен концептов:

1. 50 небольших онтологий (от 100 до 499 логических аксиом);
2. 100 онтологий средних размеров (от 500 до 4999 логических аксиом);
3. 50 больших онтологий (от 5000 логических аксиом).

Для проведения эксперимента был использован компьютер Intel Xeon Quad Core CPU @2.33 GHz, 12 GB RAM (8 из которых назначались для процесса), под управлением операционной системы Fedora 12, с использованием Java-машины версии 1.6.

Суть эксперимента заключалась в определении количества правильно выполненных задач логического анализа: классификации, выполнимости концептов и согласованности баз знаний, для каждого из профилей OWL 2. Для задачи классификации были использованы 200 баз знаний, для проверки согласованности - 200 баз знаний, и для задачи проверки выполнимости были использованы 200 баз знаний и 10 концептов из каждой базы знаний. Результаты, полученные разработанным программным комплексом, сравнивались с результатами полученными другими современными системами логического анализа OWL баз знаний, такими как HermiT, FaCT++, Konclude, MORE, wsclassifier, chainsaw и др.

Результаты эксперимента показали, что разработанные методы и алгоритмы, наиболее эффективно решают задачи классификации баз знаний,

описанных с помощью языка OWL 2 RL: из 197 баз знаний, правильно смогли быть классифицированы 181 (91,9%) база знаний (рис. 2), и проверке выполнимости концептов базы знаний описанной с помощью профиля OWL 2 DL: из 2040 концептов, правильный ответ был показан для 1808 концептов (88,6%) из 204 различных баз знаний (рис. 3).

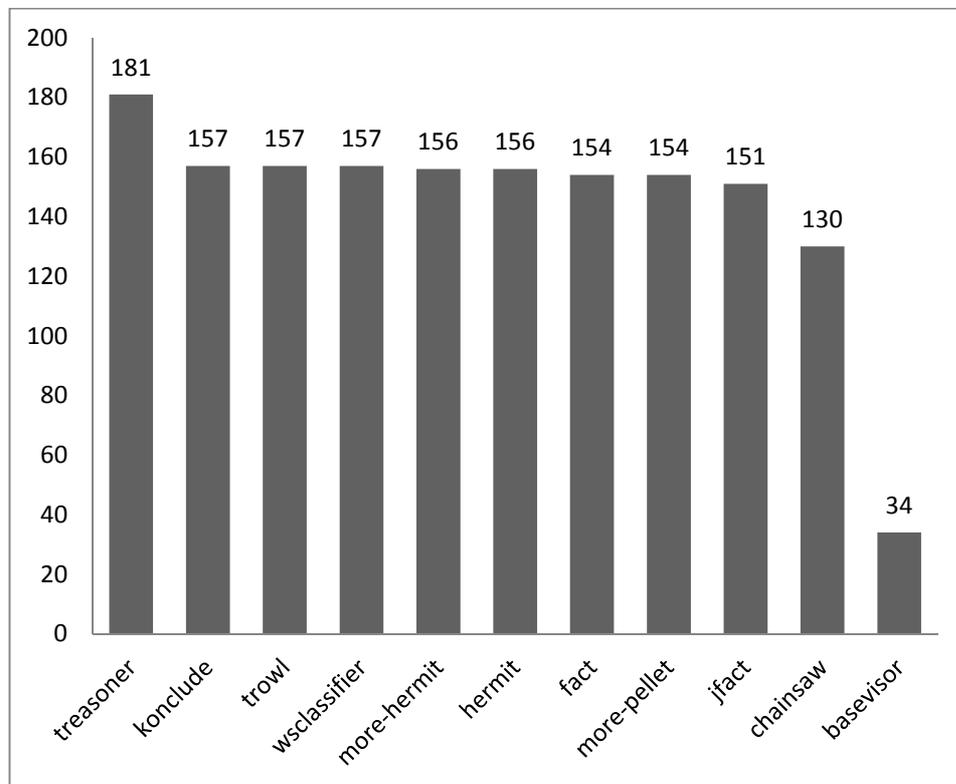
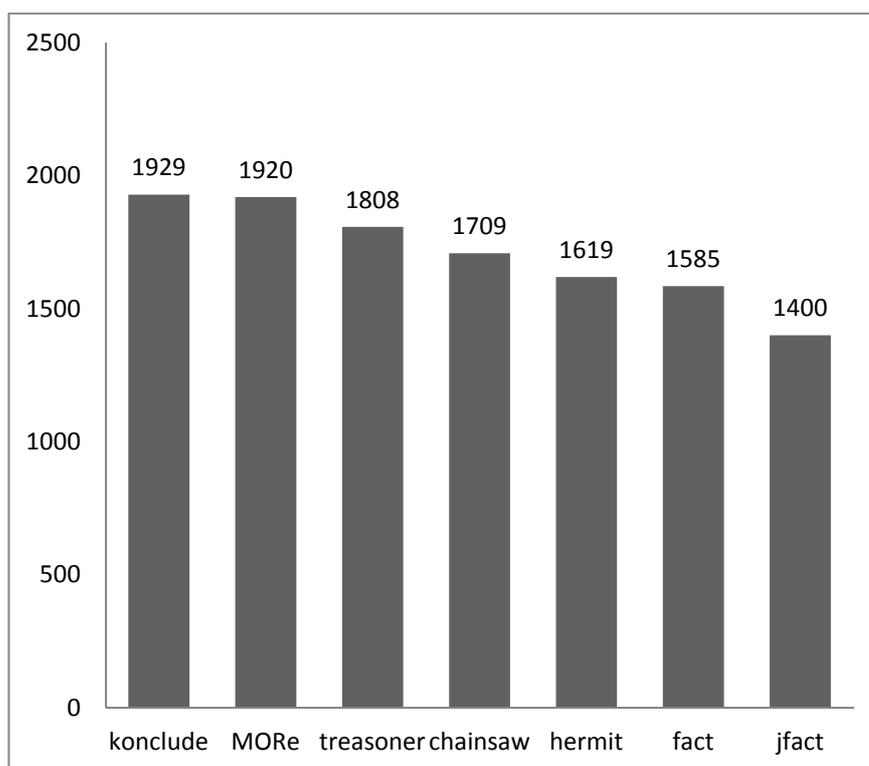


Рис. 2. Результаты эксперимента по классификации OWL 2 RL баз знаний.



*Рис. 3.* Результаты эксперимента по проверке выполнимости концептов.

Соотношение правильно выполненных задач для всех профилей OWL представлено в таблице 1.

Таблица 1. Результаты вычислительного эксперимента

Профиль OWL	Классификация	Выполнимость	Согласованность
DL	46,3%	89,5%	77,5%
EL	82,0%	97,1%	97,0%
RL	91,8%	99,5%	95,9%
Итого	73,3%	95,4%	90,1%

В **заключении** сформулированы основные результаты диссертационной работы.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработана сетевая модель представления знаний предметной области, в которой узлы графа ассоциированы с конъюнкцией концептов, а вес и стоимость ребер определяются видом операций и кванторов

дескрипционного выражения этих концептов. Доказано, что максимальный поток минимальной стоимости в этой сетевой модели позволяет определять согласованность конъюнкции концептов.

2. Разработан метод решения задачи о максимальном потоке минимальной стоимости для предложенной сетевой модели, реализованный на основе алгоритма поиска в глубину. Доказано, что вычислительная сложность созданного алгоритма согласованности концептов составила  $O(V^2)$ .

3. Модифицирован табличный алгоритм исключением дублирования проверки согласованности изоморфных концептов, который реализован на основе алгоритма Шмидта-Шаусса-Смолки и алгоритма поиска классов изоморфных концептов Эдмондсона.

4. Разработана клиент-серверная трехуровневая архитектура программного комплекса инженера по знаниям: уровень представления реализован в виде кроссплатформенного Desktop-приложения; логический уровень организован в виде серверных Java-библиотек; хранение данных включает инструменты работы с XML-файлами в стандарте OWL2.

5. Использование модуля OWL API и программного продукта стэнфордского университета Protégé в качестве модуля программного комплекса позволяет применять базы знаний, описанные с помощью стандартизированного языка OWL 2.

6. Проведение вычислительного эксперимента показало, что разработанная система показывает лучший результат по классификации концептов, в дескрипционной записи которых появляется глубокая вложенность кванторов существования. На остальных классах концептов система показывает результаты, сравнимые с результатами лучших из существующих систем. Результаты вычислительного эксперимента показали, что разработанные математические методы и алгоритмы адекватно определяют согласованность баз знаний.

## ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

### Публикации в рецензируемых журналах, рекомендованных ВАК

1. Григорьев А.В. Модификация табличного алгоритма на основе проверки непересекаемости сложных концептов [Текст] / А.Г. Ивашко, А.В. Григорьев, М.В. Григорьев. // Вестник Тюменского Государственного Университета. - 2012, №4. - С. 143-150.
2. Григорьев А.В. Сокращение пространства поиска интерпретаций за счет обнаружения изоморфных концептов [Текст] / А.Г. Ивашко, А.В. Григорьев. // Вестник Тюменского Государственного Университета. - 2013, №7. - С. 187-193.
3. Григорьев А.В. Применение табличного алгоритма для верификации бизнес-процессов [Текст] / А.Г. Ивашко, А.В. Григорьев, А.А. Кропотин, Е.О. Овсянникова // Вестник Тюменского Государственного Университета. - 2013, №7. - С. 202-213.

### Публикации в других изданиях:

4. A.V. Grigorev. TReasoner: System Description. / A.V. Grigorev, A.G. Ivashko // Proceedings of the 2nd OWL Reasoner Evaluation Workshop (ORE 2013). - 2013, С. 26-31.
5. Григорьев А.В. Извлечение классификационной информации при проверке выполнимости концептов табличным алгоритмом [Текст] / А.В. Григорьев, А.Г. Ивашко // Труды XIV международной научно-практической конференции «Современные проблемы гуманитарных и естественных наук». - Москва: ИСИ, 2013. - С. 122-125.
6. Григорьев А.В. Проблема поиска противоречий на диаграммах классов UML [Текст] / А.В. Григорьев, А.А. Кропотин, Е.О. Овсянникова // Труды международной научно-практической конференции «Современные проблемы и пути их решения в науке, транспорте, производстве и образовании'2012». – № 4. Том 14. – Одесса: КУПРИЕНКО, 2012 – 111 с.
7. Григорьев А.В. Алгоритм упрощения концептов ALC [Текст] / А.В. Григорьев // Труды научно-практической конференции «Современные проблемы математического и информационного моделирования. Перспективы разработки и внедрения инновационных IT-решений». - 2012г.
8. Григорьев А.В. Сравнение модификаций табличного алгоритма: BackJumping и семантический выбор [Текст] / А.В. Григорьев // Труды научно-практической конференции «Современные проблемы

математического и информационного моделирования. Перспективы разработки и внедрения инновационных IT-решений». - 2012г.

**Свидетельства о регистрации программ для ЭВМ:**

9. Григорьев А.В., Ивашко А.Г. Свидетельство об официальной регистрации программы для ЭВМ №2012616455 "Программный модуль для определения непротиворечивости базы знаний описанной на языке дескрипционной логики ALC табличным методом".