

МАТЕМАТИЧЕСКАЯ ЛОГИКА, АЛГЕБРА, ТЕОРИЯ ЧИСЕЛ И ДИСКРЕТНАЯ МАТЕМАТИКА

УДК 510.6, 004.6, 164.2

КОНЪЮНКТИВНЫЕ ЗАПРОСЫ К БАЗАМ ЗНАНИЙ: КАКУЮ СЕМАНТИКУ ВЫБРАТЬ?¹

Кикоть С.П.^{*}, Рыбаков М.Н.^{*,**,***}

^{*}Высшая школа современной математики МФТИ, г. Москва

^{**}НИУ ВШЭ, г. Москва

^{***}Тверской государственный университет, г. Тверь

Поступила в редакцию 05.08.2025, после переработки 25.08.2025.

В области компьютерных наук под названием «представление знаний» (knowledge representation) некоторое время назад одним из стандартов стала следующая парадигма. Под *базой знаний* (knowledge base) подразумевается архитектура, объединяющая в себе данные, знания общего характера и механизм логического вывода, позволяющий генерировать на основе базовых данных и знаний новую информацию, а также использовать её для ответов на запросы к системе. В этом обзоре мы сосредоточимся на сценарии, когда под данными подразумевается набор замкнутых атомов, под знаниями — онтология в той или иной дескрипционной логике, а в качестве языка запросов выбраны *конъюнктивные запросы* (conjunctive queries). Цель настоящей работы — собрать воедино теоретические и практические результаты о трудоёмкости процедур построения ответа в зависимости от *семантики*, то есть формального определения того, что считать ответом на запрос.

Ключевые слова: базы знаний, конъюнктивные запросы, семантика, сложность вычислений.

Вестник ТвГУ. Серия: Прикладная математика. 2025. № 3. С. 5–24.
<https://doi.org/10.26456/vtpmk754>

Введение

Предлагаемая читателю работа носит обзорный характер и может рассматриваться как краткое введение в область, связанную с представлением знаний и запросам к структурам, которые используются для представления знаний. Мы ограничимся ситуацией, когда данные представляют собой наборы замкнутых атомов, знания — онтологии в некоторой дескрипционной логике, а запросы — конъюнкции

¹Исследование поддержано Минобрнауки России, программа «Приоритет 2030».

© Кикоть С.П., Рыбаков М.Н., 2025

атомов под кванторами существования. Такие запросы мы будем называть конъюнктивными. Они соответствуют SPJ-запросам (select-project-join queries) языка SQL и играют фундаментальную роль в теории реляционных баз данных, поскольку вычисление аналитических запросов с операторами группировки, подсчёта и суммирования начинается, как правило, с соединения таблиц, т.е. с вычисления конъюнктивной части.

Наша цель состоит в том, чтобы собрать воедино основные известные теоретические и практические результаты о трудоёмкости процедур построения ответа на запрос в зависимости от выбранной семантики, так как сложность построения ответа зависит не столько от того, что считать запросом, сколько от формального определения того, что считать ответом на запрос.

Работа устроена следующим образом. Сначала мы опишем языки, которые используются для создания баз знаний, затем семантики и сложность задач, связанных с запросами. Некоторые описания будут сопровождаться примерами, цель которых — прояснить особенности возникающих понятий.

Вопрос об эффективной физической реализации хранения данных остаётся за рамками этой работы.

1. Базы знаний

Фиксируем набор $\mathcal{IN}$ *индивидуальных имён* (констант), набор $\mathcal{CN}$ *атомарных концептов* (одноместных предикатов) и набор $\mathcal{RN}$ *атомарных ролей* (двуместных предикатов). Все вместе они составляют *сигнатуру* Σ . Под *данными*² $\mathcal{D}$ подразумевается набор замкнутых атомарных формул (или просто атомов) вида $P(a)$ или $R(a, b)$, где $P \in \mathcal{CN}$, $R \in \mathcal{RN}$ и $a, b \in \mathcal{IN}$. Под *онтологией* $\mathcal{T}$ подразумевается набор аксиом некоторой дескрипционной логики; такие логики определяются ниже. Аксиомы бывают двух видов: *включение составных концептов* и *включение составных ролей*. *Составные концепты* (далее просто *концепты*) строятся при помощи *конструкторов*. Набор конструкторов и ограничений на их применение определяет *дескрипционную логику*.

Концепты логики $\mathcal{ALC}$ (attribute logic with complements) задаются грамматикой

$$C ::= \perp \mid \top \mid A \mid \neg C \mid (C \sqcap C) \mid (C \sqcup C) \mid \exists R. C \mid \forall R. C,$$

где $\perp$ — *пустой концепт*, $\top$ — *универсальный концепт*, A — атомарный концепт, R — роль, а C — нетерминальный символ. Все аксиомы в логике $\mathcal{ALC}$ — это включения концептов, т.е. выражения вида $C \sqsubseteq D$, где C и D — концепты.

Существуют многочисленные синтаксические расширения логики $\mathcal{ALC}$. Для их описания используются буквы $\mathcal{I}$, $\mathcal{O}$, $\mathcal{Q}$, $\mathcal{H}$, $\mathcal{R}$, $\mathcal{S}$, каждая из которых соответствует добавлению в язык определённых конструкций; укажем вид и название каждой:

$\mathcal{I}$: *обратные роли* вида R^- , где R — атомарная роль;

$\mathcal{O}$: *концепты-номиналы* вида $\{a\}$, где $a \in \mathcal{IN}$;

²Это соглашение удобно для изучения вычислительной сложности. В действующих системах данные (т.е. значения унарных и бинарных предикатов) обычно генерируются *отображением* из классических таблиц. Построение таких отображений, или *привязка базы данных к онтологии*, — трудоёмкий процесс, но он часто может быть автоматизирован [23].

- $\mathcal{Q}$: концепты численного ограничения вида $\geq m R.C$ и $\leq m R.C$, где R — роль, C — концепт и $m \in \mathbb{N}$;
 $\mathcal{H}$: аксиомы включения ролей вида $R \sqsubseteq S$, где R и S — роли;
 $\mathcal{R}$: аксиомы включения композиций вида $R_1 \circ \dots \circ R_n \sqsubseteq R$, где $R_1, \dots, R_n$ и R — роли;
 $\mathcal{S}$: аксиомы транзитивности вида $R \circ R \sqsubseteq R$, где R — роль.

Принято следующее использование этих букв: буквы $\mathcal{I}$, $\mathcal{O}$, $\mathcal{Q}$, $\mathcal{H}$, $\mathcal{R}$ приписываются к имени логики $\mathcal{ALC}$, а буква $\mathcal{S}$ заменяет буквы $\mathcal{ALC}$.

Чтобы добиться разрешимости получающихся систем, в присутствии конструкторов $\mathcal{S}$ и $\mathcal{Q}$ накладываются дополнительные ограничения: во-первых, выражения вида $\geq m R.C$ и $\leq m R.C$ считаются корректными концептами только в том случае, когда роль R проста, то есть не имеет транзитивных подролей, а во-вторых, множество всех слов вида $R_1 \dots R_n$, таких что включение $R_1 \circ \dots \circ R_n \sqsubseteq R$ выводится в онтологии $\mathcal{T}$ в очевидном смысле, должно быть регулярным языком.

Перейдём теперь к описанию логик, менее выразительных, чем $\mathcal{ALC}$. В логике $\text{DL-Lite}_{\text{core}}$ разрешены только включения концептов, имеющие вид $C \sqsubseteq D$ или $C \sqsubseteq \neg D$, где C и D — либо атомарные концепты, либо экзистенциальные ограничения вида $\exists R.T$, в которых допускается, чтобы роль R была обратной. В логике $\text{DL-Lite}_{\mathcal{R}}$, взятой за основу стандарта OWL 2 QL , дополнительно разрешаются аксиомы включения ролей. Логика $\mathcal{EL}$ представляет собой фрагмент логики $\mathcal{ALC}$, в котором разрешены произвольные концепты, полученные при помощи конъюнкции и экзистенциальных ограничений, но запрещены дизъюнкция, отрицание и универсальные ограничения. Кроме того, в концептах этой логики разрешается использовать $\top$, но не $\perp$. Расширение логики $\mathcal{EL}$, допускающее $\perp$, обозначается $\mathcal{EL}_{\perp}$. Как и логика $\mathcal{ALC}$, логики $\mathcal{EL}$ и $\mathcal{EL}_{\perp}$ допускают расширения, определяемые буквами $\mathcal{I}$, $\mathcal{O}$, $\mathcal{Q}$, $\mathcal{H}$, $\mathcal{R}$, как это было описано выше. Популярный стандарт OWL 2 EL основан на логике $\mathcal{ELRO}_{\perp}$. Логика Horn-SHIQ является фрагментом логики SHIQ , не содержащим явной или неявной дизъюнкции. Её упрощённое синтаксическое определение можно найти в работе [34], представляющей собой введение в хорновы дескрипционные логики, лежащие в основе стандарта OWL 2 RL .

Под базой знаний подразумевается набор $\mathcal{K} = (\mathcal{D}, \mathcal{T})$, состоящий из данных $\mathcal{D}$ и онтологии $\mathcal{T}$, сформулированной в одной из вышеперечисленных логик.

Конъюнктивный запрос — это формула языка первого порядка вида $q(\vec{x}) = \exists \vec{y} \varphi(\vec{x}, \vec{y})$, где $\varphi(\vec{x}, \vec{y})$ — конъюнкция атомов, термами которых могут быть как переменные из векторов $\vec{x}$ и $\vec{y}$, так и индивиды из $\mathcal{IN}$.

2. Интерпретации баз знаний

Интерпретацией сигнатуры Σ называется пара $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, где $\Delta^{\mathcal{I}}$ — непустое множество, называемое носителем интерпретации, а $\cdot^{\mathcal{I}}$ — функция интерпретации, определённая на сигнатуре Σ так, что

- если A — атомарный концепт, то $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$;
- если R — роль, то $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$;
- если a — константа, то $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$.

Функция интерпретации распространяется на сложные концепты и сложные роли следующим образом:

$$\begin{aligned}
\top^{\mathcal{I}} &= \Delta^{\mathcal{I}}; \\
\perp^{\mathcal{I}} &= \emptyset; \\
\{a\}^{\mathcal{I}} &= \{a^{\mathcal{I}}\}; \\
(\neg C)^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}; \\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}}; \\
(C \sqcup D)^{\mathcal{I}} &= C^{\mathcal{I}} \cup D^{\mathcal{I}}; \\
(\exists R.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} : \exists e \in \Delta^{\mathcal{I}} ((d, e) \in R^{\mathcal{I}} \ \& \ e \in C^{\mathcal{I}})\}; \\
(\forall R.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} : \forall e \in \Delta^{\mathcal{I}} ((d, e) \in R^{\mathcal{I}} \Rightarrow e \in C^{\mathcal{I}})\}; \\
(\geq m R.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} : |\{e \in C^{\mathcal{I}} : (d, e) \in R^{\mathcal{I}}\}| \geq m\}; \\
(\leq m R.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} : |\{e \in C^{\mathcal{I}} : (d, e) \in R^{\mathcal{I}}\}| \leq m\}; \\
(R^-)^{\mathcal{I}} &= \{(c, d) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} : (d, c) \in R^{\mathcal{I}}\}; \\
(R \circ S)^{\mathcal{I}} &= \{(c, d) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} : \exists e \in \Delta^{\mathcal{I}} ((c, e) \in R^{\mathcal{I}} \ \& \ (e, d) \in S^{\mathcal{I}})\},
\end{aligned}$$

где a — константа, C и D — концепты, R и S — роли, $m \in \mathbb{N}$.

В интерпретации $\mathcal{I}$ естественным образом определяется *истинность* аксиом. Именно, если C и D — концепты, а R и S — роли, то

- аксиома $C \sqsubseteq D$ истинна в $\mathcal{I}$, если $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$;
- аксиома $R \sqsubseteq S$ истинна в $\mathcal{I}$, если $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$.

Онтология $\mathcal{T}$ считается истинной в интерпретации $\mathcal{I}$, если каждая её аксиома истинна в $\mathcal{I}$. Атомы данных $A(a)$ и $P(a, b)$ считаются истинными в $\mathcal{I}$, если, соответственно, $a^{\mathcal{I}} \in A^{\mathcal{I}}$ и $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in P^{\mathcal{I}}$; истинность данных $\mathcal{D}$ в $\mathcal{I}$ означает истинность в $\mathcal{I}$ всех атомов из $\mathcal{D}$, а истинность базы знаний $\mathcal{K} = (\mathcal{T}, \mathcal{D})$ в $\mathcal{I}$ означает истинность в $\mathcal{I}$ и онтологии $\mathcal{T}$, и базы данных $\mathcal{D}$. Если база знаний $\mathcal{K}$ истинна в интерпретации $\mathcal{I}$, то пишут $\mathcal{I} \models \mathcal{K}$ и говорят, что $\mathcal{I}$ является интерпретацией базы знаний $\mathcal{K}$.

3. Замкнутый мир

Заметим, что данным $\mathcal{D}$ естественным образом соответствует интерпретация $\mathcal{I}_{\mathcal{D}}$, где в качестве $\Delta^{\mathcal{I}}$ берётся множество всех индивидных имён, входящих в $\mathcal{D}$, а истинными объявляются в точности те атомы, которые попали в $\mathcal{D}$.

В семантике замкнутого мира ответом на запрос $q(\vec{x}) = \exists \vec{y} \varphi(\vec{x}, \vec{y})$ относительно базы знаний $\mathcal{K} = (\mathcal{D}, \mathcal{T})$ является следующее множество наборов констант: $\{\vec{a} \in (\Delta^{\mathcal{I}})^{|\vec{x}|} : \mathcal{I}_{\mathcal{D}} \models q(\vec{a})\}$.

Из определения видно, что при таком подходе онтология $\mathcal{T}$, входящая в базу знаний $\mathcal{K} = (\mathcal{D}, \mathcal{T})$, при ответе на запрос просто не учитывается. Тем не менее, $\mathcal{T}$ в этой ситуации может служить для проверки корректности данных, а также может быть присоединена к данным в виде RDF-графа, позволяя формулировать запросы о её синтаксической структуре.

Этой семантике соответствуют *полные* (то есть без значений NULL) базы данных, работающие по протоколу SQL, а также многие хранилища троек (triple stores), предоставляющие сервис ответов на SPARQL-запросы.

4. Открытый мир. Семантика логического следования

Под *основной семантикой* мы понимаем *семантику логического следования*, согласно которой база знаний даёт положительный ответ в точности на те запросы, которые верны во всевозможных интерпретациях этой базы знаний.

Рассмотрим базу знаний $\mathcal{K} = (\mathcal{T}, \mathcal{D})$ и конъюнктивный запрос $q(\vec{x})$ от n свободных переменных. Набор $\vec{a}$, состоящий из n индивидов, называется *частным ответом на запрос $q(\vec{x})$* относительно $\mathcal{K}$ в основной семантике, если для любой интерпретации $\mathcal{I}$, такой что $\mathcal{I} \models \mathcal{K}$, имеет место $\mathcal{I} \models q(\vec{a})$; в этом случае пишут $\mathcal{K} \models q(\vec{a})$. *Ответом на запрос $q(\vec{x})$* называется множество всех частных ответов на этот запрос.

Пример

Приведём пример онтологии, которая может сопровождать базу данных о сотрудниках кафедры университета и читаемых ими курсах. Сотрудники кафедры — это научно-педагогические работники (НПР), среди которых выделяем профессоров и доцентов; каждый НПР владеет знаниями в областях каких-то дисциплин, т.е. потенциально может преподавать их, при этом могут быть как дисциплины кафедры, которые он действительно преподаёт, так и те, которые он не преподаёт. Поэтому считаем, что

- данные содержат имена сотрудников и названия дисциплин;
- имеются концепты Профессор, Доцент и НПР;
- имеются роли знает и преподаёт.

Тогда, понимая концепты и роли как унарные и бинарные предикаты, мы можем записать утверждение « a является доцентом» как Доцент(a), а утверждение «сотрудник a может вести занятия по дисциплине A » — как знает(a, A). Допустим, мы хотим описать онтологию, определяемую следующими условиями:

- профессора и доценты являются научно-педагогическими работниками;
- если кто-то преподаёт какую-то дисциплину, то он владеет знаниями по ней;
- любой НПР владеет знаниями по какой-то дисциплине;
- некто преподаёт какую-то дисциплину, только если он НПР.

Запишем эти условия на языке первого порядка:

$$\begin{aligned} \forall x (\text{Профессор}(x) &\rightarrow \text{НПР}(x)); \\ \forall x (\text{Доцент}(x) &\rightarrow \text{НПР}(x)); \\ \forall x \forall y (\text{преподаёт}(x, y) &\rightarrow \text{знает}(x, y)); \\ \forall x (\text{НПР}(x) &\rightarrow \exists y \text{ знает}(x, y)); \\ \forall x (\exists y \text{ преподаёт}(x, y) &\rightarrow \text{НПР}(x, y)). \end{aligned}$$

Теперь запишем их же на языке дескрипционной логики DL-Lite $\mathcal{R}$ и соберём в онтологию $\mathcal{T}$:

$$\begin{aligned} \text{Профессор} &\sqsubseteq \text{НПР}; \\ \text{Доцент} &\sqsubseteq \text{НПР}; \\ \text{преподаёт} &\sqsubseteq \text{знает}; \\ \text{НПР} &\sqsubseteq \exists \text{ знает. } \top; \\ \exists \text{ преподаёт. } \top &\sqsubseteq \text{НПР}. \end{aligned}$$

Используя онтологию $\mathcal{T}$, построим на её основе базу знаний. Для этого возьмём данные

$$\mathcal{D} = \{\text{Доцент}(a), \text{Профессор}(b), \text{преподаёт}(c, \text{Алгебра})\}$$

и положим $\mathcal{K} = (\mathcal{T}, \mathcal{D})$. Зададим запросы $q_1(x)$ и $q_2(x)$:

$$\begin{aligned} q_1(x) &= \text{НПР}(x); \\ q_2(x) &= \exists y \text{ знает}(x, y). \end{aligned}$$

Запрос $q_1(x)$ можно перевести на естественный язык как «выдать список всех научно-педагогических работников», а запрос $q_2(x)$ — как «выдать список тех, кто может преподавать какую-то определённую дисциплину». У базы знаний $\mathcal{K}$ имеется бесконечно много интерпретаций. Например, можно взять интерпретацию $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, где

$$\begin{aligned} \Delta^{\mathcal{I}} &= \{a, b, c, d, e, \text{Алгебра}, \text{Геометрия}, \text{Топология}\}; \\ \text{Доцент}^{\mathcal{I}} &= \{a\}; \\ \text{Профессор}^{\mathcal{I}} &= \{b, d\}; \\ \text{НПР}^{\mathcal{I}} &= \{a, b, c, d\}; \\ \text{преподаёт}^{\mathcal{I}} &= \{(a, \text{Геометрия}), (b, \text{Топология}), (c, \text{Алгебра})\}; \\ \text{знает}^{\mathcal{I}} &= \text{преподаёт}^{\mathcal{I}} \cup \{(d, \text{Геометрия}), (d, \text{Алгебра})\}, \end{aligned}$$

а также $a^{\mathcal{I}} = a$, $b^{\mathcal{I}} = b$, $c^{\mathcal{I}} = c$ и $\text{Алгебра}^{\mathcal{I}} = \text{Алгебра}$. Приведённая в качестве примера интерпретация $\mathcal{I}$ отражает принимаемое нами *предположение об открытости мира*: в интерпретации допускаются индивиды, не упомянутые в данных, а также более широкие объёмы предикатов. Тем не менее, в любой интерпретации, где истинна онтология $\mathcal{T}$, индивиды a, b, c существуют и принадлежат концепту НПР. Таким образом, каждый из этих индивидов будет частным ответом на запрос $q_1(x)$ в семантике логического следования. Аналогично, каждый из них будет частным ответом на запрос $q_2(x)$ в этой семантике.

Под *задачей об ответе на запрос* в зависимости от контекста подразумевается одна из двух следующих задач. Первая — это *задача получения ответа на запрос*. Она заключается в том, чтобы по базе знаний $\mathcal{K}$ и запросу $q(\vec{x})$ построить ответ на запрос $q(\vec{x})$ относительно $\mathcal{K}$. Именно на её решение должен быть нацелен разработчик, намеренный построить работающую базу знаний. Вторая — это *задача проверки частного ответа на запрос*. Она заключается в том, чтобы по $\mathcal{K}$, $q(\vec{x})$ и вектору индивидов $\vec{a}$, определить, является ли $\vec{a}$ ответом на запрос $q(\vec{x})$ относительно $\mathcal{K}$. Различные результаты о вычислительной сложности ответов на запрос относятся именно к ней.

Традиционно для задачи об ответе на запрос изучаются два вида сложности. Говоря о *комбинированной сложности*, имеют в виду время работы для задачи проверки частного ответа на запрос в зависимости от длины входа, где под входом понимается двоичное представление всего набора $(\mathcal{D}, \mathcal{T}, q(\vec{x}), \vec{a})$. Говоря о *сложности по данным*, считают $\mathcal{T}$, $q(\vec{x})$ и $\vec{a}$ фиксированными, и ставят вопрос о том, как время работы для получения решения задачи проверки ответа на запрос зависит от $\mathcal{D}$ (точнее, от длины двоичного представления $\mathcal{D}$).

Известные результаты о сложности задачи получения ответов на запросы собраны в Таблице 1; там же даны ссылки на списки работ теоретического, практического и смешанного характера, в том числе содержащих результаты экспериментов, связанных с построением и тестированием отвечающих систем; сами списки работ приведены сразу после таблицы.

Среди упомянутых работ отдельно отметим работы [47], [28] и [27]: они основаны на привлечении *теории соответствия*, в частности, на определённых связях между языком запросов и модальными формулами. С помощью методов этой теории задача проверки частного ответа на *модально определимые* запросы сводится к задаче проверки базы знаний на совместность.

ТАБЛИЦА 1: Сложность задачи получения ответов на запросы

Онтология	Сложность		Списки работ		
	по данным	комбинир.	теоретич.	смеш.	практич.
нет ($\sim$ БД)	AC^0	NP	работы по СУБД		
DL-Lite $\mathcal{R}$	AC^0	NP	[2], [4], [6], [12], [22], [24], [25]	[26]	[10], [39], [41], [45]
$\mathcal{EL}$, $\mathcal{ELH}$	P	NP	[13], [33]	[43]	—
$\mathcal{ELLI}$, $\mathcal{ELHI}$, Horn- $SHIQ$	P	EXP	[18], [38]	[44]	[37]
$\mathcal{ALC}$, $\mathcal{ALCHQ}$	coNP	EXP	[47]	—	—
$\mathcal{ALCI}$, SH , $SHIQ$	coNP	2EXP	[17], [20]	—	—
$SHOIQ$, $SROIQ$	$\geq$ coNP	$\geq$ co2NEXP	[19], [21], [28]	—	[27], [46]

Скажем ещё про две работы, не попавшие в Таблицу 1. Работа [5] представляет собой обзор задачи об ответах на запросы разного типа (атомные, конъюнктивные, путевые) в основной семантике, содержащий много примеров и описаний основных алгоритмов. Введение [32] содержит первоначальные теоретические сведения и практические указания для разработки баз знаний над DL-Lite $\mathcal{R}$.

Теперь дадим некоторые пояснения к Таблице 1. Она состоит из трёх частей. В двух верхних строчках таблицы указана одна и та же сложность задач, и это не случайно. Дело в том, что логика DL-Lite $\mathcal{R}$ изначально строилась в работах [12] и [2] как максимальная дескрипционная логика, допускающая *первопорядковую переформулировку*. Поэтому для любой онтологии $\mathcal{T}$ этого языка и запроса $q(\vec{x})$ существует такой SQL-запрос $q'(\vec{x})$, называемый *переформулировкой* запроса $q(\vec{x})$ относительно $\mathcal{T}$, что для любых данных $\mathcal{D}$ ответ на запрос $q(\vec{x})$ может быть получен путём исполнения $q'(\vec{x})$ на классической реляционной базе данных, реализующей интерпретацию $\mathcal{I}_{\mathcal{D}}$. Это наблюдение вместе с успехом реляционных баз данных с тем же самым набором сложности классов лежит в основе тезиса о жизнеспособности баз знаний на основе DL-Lite $\mathcal{R}$. Кроме того, на язык DL-Lite $\mathcal{R}$ автоматически переводятся почти все конструкторы UML-диаграмм, часто сопровождающие промышленные базы данных. Однако недавние теоретические исследования [22], [24] и [4] показали, что размер переформулировки может быть значительно больше общего размера $q(\vec{x})$ и $\mathcal{T}$ даже при очень сильных ограничениях на онтологию и форму запроса. Взгляд на теорию длины переформулировок с точки зрения схемной сложности булевых функций представлен в обзоре [40].

Первые две части таблицы объединяет еще одно свойство — существование *универсальной*, или *канонической модели*. Это значит, что для любой базы знаний $\mathcal{K}$ существует такая интерпретация $\mathcal{I}_{\mathcal{K}}$, что ответ на запрос $q(\vec{x})$ арности n может быть получен как множество $\{\vec{a} \in \text{Ind}(\mathcal{D}) : \mathcal{I}_{\mathcal{K}} \models q(\vec{a})\}$, где $\text{Ind}(\mathcal{D})$ — совокупность всех индивидов, входящих в $\mathcal{D}$. Таким образом, два основных метода построения ответа на запрос — это переформулировка запроса и генерация канонической модели для базы знаний. Последняя операция часто также называется

материализацией (materialization) или *сколемизацией* (skolemization). Иногда используется и *комбинированный подход* (combined approach): для ответа на запрос $q(\vec{x})$ относительно $\mathcal{K}$ на специальном образом сокращённой канонической модели вычисляется его переформулировка $q'(\vec{x})$ [31, 36].

Заметим, что в общем определении ответа имеется квантор по *всем* интерпретациям, и для более выразительных логик их перебор в некотором смысле неизбежен. В этом случае стараются его минимизировать и поручить *ризонерам* (reasoners) — готовому программному обеспечению для осуществления логического вывода в выразительных дескрипционных логиках, использующему *табличные алгоритмы* (tableau algorithms) или *правила* (consequence-based reasoning).

Работы из последних двух столбцов Таблицы 1 содержат результаты экспериментов, связанные с реализацией различных алгоритмов ответов на запросы в логической семантике и их тестированием на коллекциях запросов и онтологий, а также технические описания построенных систем. Эти коллекции заслуживают отдельного внимания. Они создаются искусственно, исходя из следующих требований:

- запрос должен быть разумным, то есть можно было бы себе представить ситуацию, в которой он реально может быть задан;
- запрос должен исполняться на больших естественных или искусственно сгенерированных данных.

Эксперименты показывают, что на большинство запросов из коллекций описанные в работах системы могут отвечать за разумное время.

5. Ограниченная семантика

Одним из недостатков полной логической семантики для конъюнктивных запросов является её высокая вычислительная сложность в худшем случае и потребность в разработке новых машин, осуществляющих ризонинг. С другой стороны, целесообразность рассмотрения ответов, где экзистенциально-квантифицированные переменные отображаются в анонимные узлы (а именно поиск таких ответов — причина высокой сложности), часто может быть поставлена под сомнение. Отсюда возникает *ограниченная семантика* [11], в которой вектор $\vec{a} \in \text{Ind}(\mathcal{K})^n$ считается ответом на n -арный запрос $q(\vec{x}) = \exists \vec{y} \varphi(\vec{x}, \vec{y})$ относительно базы знаний $\mathcal{K}$, если найдётся такой вектор $\vec{b} \in \text{Ind}(\mathcal{K})^{|\vec{y}|}$, что $\mathcal{I} \models q(\vec{a}, \vec{b})$ для любой интерпретации $\mathcal{I}$ базы знаний $\mathcal{K}$.

Пример

Вернёмся к примеру из предыдущего раздела. В ограниченной семантике ответом на запрос $q_1(x)$ по-прежнему будет множество $\{a, b, c\}$, а на запрос $q_2(x)$ ответом будет множество $\{c\}$, так как в указанной интерпретации при значении переменной x , выбранном из множества $\{a, b\}$, значением переменной y будет анонимный индивид. Такой ответ на запрос $q_2(x)$ является осмысленным: напомним, что запрос $q_2(x)$ мы перевели на естественный язык как «выдать список тех, кто может преподавать какую-то определённую дисциплину», и если в этом описании запроса понимать слова об определённой дисциплине так, что мы можем указать хотя бы одну такую явно, то множество $\{c\}$ и будет адекватным ответом на него.

Этот пример показывает, что бывают ситуации, когда ограниченная семантика может соответствовать определённым «ожиданиям» пользователя относительно того, как следует понимать его запрос.

В то же время, в ситуации, когда анонимные индивиды генерируются из предложений естественного языка при помощи правил онтологии, отражающей определения понятий естественного языка, в системах глубокого семантического анализа [8], игнорировать их при ответах на запросы было бы грубой ошибкой.

Достоинство этой семантики в том, что она может быть реализована на основе интерпретатора языка Дейталог ([15] — классическая ссылка по его теории; [37] — более современная по проблемам реализации). Частично поэтому эта семантика была принята за основу режима OWL Direct Semantics стандарта SPARQL 1.1, см. [29, 30].

Ещё один способ уменьшить вычислительную сложность задачи об ответах на запросы при работе с выразительными онтологиями состоит в том, чтобы использовать их приближения в менее выразительных языках из верхней части таблицы, например, DL-Lite_R (см. [1]) или $\mathcal{EL}$ (см. [42] и [14]).

6. Семантики, устойчивые к противоречию

Ещё один недостаток семантики логического следования заключается в том, что она не работает с противоречивыми базами знаний. Любой набор индивидов всегда будет ответом на произвольный запрос к противоречивой базе просто в силу отсутствия у неё интерпретаций. Однако, как показывают многочисленные эксперименты (например, в рамках соревнования ORE³), противоречивые базы знаний — скорее правило, чем исключение. В связи с этим существует ряд *семантик, устойчивых к противоречию*, основанных на понятии выправления (или починки) базы знаний. Все они принимают гипотезу о том, что онтология изначально непротиворечива, а противоречие в базе знаний возникает из-за зашумлённости данных.

Под *выправлением* (repair) базы знаний $\mathcal{K} = (\mathcal{T}, \mathcal{D})$ понимается любое подмножество $\mathcal{D}'$ множества $\mathcal{D}$, максимальное по включению во множестве

$$\{\mathcal{D}'' \subseteq \mathcal{D} : \mathfrak{I} \models (\mathcal{T}, \mathcal{D}'') \text{ для некоторой интерпретации } \mathfrak{I}\},$$

т.е. среди подмножеств множества $\mathcal{D}$, приводящих к совместной базе знаний, содержащей онтологию $\mathcal{T}$. Набор индивидов $\vec{a}$ называется частным ответом на запрос $q(\vec{x})$ в *AR-семантике*⁴, если $(\mathcal{T}, \mathcal{D}') \models q(\vec{a})$ для каждого выправления $\mathcal{D}'$ базы знаний $\mathcal{K}$. Аналогично, $\vec{a}$ считается частным ответом на запрос $q(\vec{x})$ в *IAR-семантике*⁵, если $(\mathcal{T}, \mathcal{D}') \models q(\vec{a})$, где $\mathcal{D}'$ — пересечение всех выправлений базы знаний $\mathcal{K}$.

Нетрудно понять, что AR-семантика является более дорогостоящей вычислительно, чем IAR-семантика, что связано с необходимостью рассмотрения большего количества выправлений.

³<https://www.w3.org/community/owled/ore-2015-workshop/competition/>

⁴Аббревиатура «AR» означает «all repairs».

⁵Аббревиатура «IAR» означает «intersection of all repairs».

Пример

Пусть $\mathcal{T}$ — онтология, которая задаётся представленными ниже аксиомами.

Дескрипционная логика	Язык первого порядка
Профессор $\sqsubseteq$ Сотрудник	$\forall x (\text{Профессор}(x) \rightarrow \text{Сотрудник}(x))$
Доцент $\sqsubseteq$ Сотрудник	$\forall x (\text{Доцент}(x) \rightarrow \text{Сотрудник}(x))$
Доцент $\sqsubseteq$ $\neg$ Профессор	$\forall x (\text{Профессор}(x) \rightarrow \neg \text{Доцент}(x))$

Пусть имеются данные

$$\mathcal{D} = \{\text{Доцент}(a), \text{Профессор}(a), \text{Доцент}(b)\}.$$

Тогда база знаний $\mathcal{K} = (\mathcal{T}, \mathcal{D})$ противоречива, потому что, согласно онтологии $\mathcal{T}$, сотрудник a не может быть доцентом и профессором одновременно. У базы $\mathcal{K}$ имеются два выправления:

$$\begin{aligned} \mathcal{D}'_1 &= \{\text{Доцент}(a), \text{Доцент}(b)\}; \\ \mathcal{D}'_2 &= \{\text{Профессор}(a), \text{Доцент}(b)\}, \end{aligned}$$

которые в пересечении дают

$$\mathcal{D}'' = \{\text{Доцент}(b)\}.$$

Рассмотрим запрос $q(x) = \text{Сотрудник}(x)$. Тогда в AR-семантике ответом на запрос $q(x)$ будет множество $\{a, b\}$, а в IAR-семантике — множество $\{b\}$.

С другой стороны, она позволяет получить более полный ответ на запрос. Результаты о вычислительной сложности этих семантик для логики DL-Lite $_{\mathcal{R}}$ приведены в Таблице 2.

ТАБЛИЦА 2: Сложность семантик для DL-Lite $_{\mathcal{R}}$

Семантика	Сложность по данным	Комбинированная сложность
AR	coNP	Π_2^P
IAR	$\leq P$	NP

По поводу сложности по данным в этих семантиках см. [35], а по поводу комбинированной сложности см. [7].

Существуют и менее очевидные семантики, устойчивые к противоречиям. Так, в работе [35] предлагаются *CAR-семантика* и *ICAR-семантика*⁶, в которых рассматриваются только те выправления, которые замкнуты относительно добавления логических следствий из исходной базы знаний. В работе [7] предлагаются два семейства семантик, параметризованные целым параметром k , приближающие AR-семантику сверху и снизу, сходящиеся к ней при k , стремящемся к бесконечности, и допускающие первопорядковые переписывания, то есть со сложностью по данным AC^0 и комбинированной сложностью NP для логики DL-Lite $_{\mathcal{R}}$. В работе [3] предлагается семантика для случая, когда данные ранжированы с точки зрения их достоверности (preferred repair semantics). Эта работа также интересна тем, что содержит, во-первых, практическую реализацию некоторых семантик,

⁶Здесь «C» означает «closed».

устойчивых к противоречию, на основе SAT-решателя, а во-вторых, результаты тестирования на довольно больших объёмах данных.

Заключение

Существует несколько семантик ответов на конъюнктивные запросы относительно баз знаний. Наиболее стандартная и логичная из них — семантика логического следования в открытом мире — для выразительных логик сопровождается большой вычислительной сложностью в худшем случае. Несмотря на это, на практике она может быть жизнеспособна при условии небольшого размера запросов, онтологий и данных при том предположении, что они «далеки от худшего случая». В случае средних объёмов данных основная семантика может быть реализована в менее выразительных дескрипционных логиках или может быть использована сокращённая семантика. Кроме того, существует большое число ещё более вычислительно дорогостоящих семантик для работы с противоречивыми базами знаний.

Таким образом, при построении баз знаний выбор языка онтологии и семантики ответов на запросы должен осуществляться обдуманно с учётом специфики поставленной задачи, теоретической сложности задачи об ответах на запросы в данной семантике в худшем случае и экспериментальных данных о времени работы стандартных алгоритмов на коллекциях запросов и баз знаний тех или иных предметных областей.

Список литературы

- [1] Antonioni N., Castano F., Coletta S., Grossi S., Lembo D., Lenzerini M., Poggi A., Virardi E., Castracane P. Ontology-based data management for the Italian public debt // Formal Ontology in Information Systems – Proceedings of the Eighth International Conference, FOIS 2014. Eds. by P. Garbacz, O. Kutz. Series: Frontiers in Artificial Intelligence and Applications. Brazil: IOS Press, 2014. Pp. 372–385.
- [2] Artale A., Calvanese D., Kontchakov R., Zakharyashev M. The DL-Lite family and relations // Journal of Artificial Intelligence Research. 2009. Vol. 36. Pp. 1–69.
- [3] Bienvenu M., Bourgaux C., Goasdoue F. Querying inconsistent description logic knowledge bases under preferred repair semantics // Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Quebec City, Canada: AAAI Press, 2014. Pp. 996–1002.
- [4] Bienvenu M., Kikot S., Podolskii V. Tree-like queries in OWL 2 QL: succinctness and complexity results // 30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015. Kyoto, Japan: IEEE, 2015. Pp. 317–328.
- [5] Bienvenu M., Ortiz M. Ontology-mediated query answering with data-tractable description logics // Reasoning Web. Web Logic Rules – 11th International Summer School 2015. Eds. by W. Faber, A. Paschke. Series: Lecture Notes in Computer Science. Berlin, Germany: Springer, 2015. Pp. 218–307.

- [6] Bienvenu M., Ortiz M., Simkus M., Xiao G. Tractable queries for lightweight description logics // Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Ed. by F. Rossi. Beijing, China: IJCAI/AAAI, 2013.
- [7] Bienvenu M., Rosati R. New inconsistency-tolerant semantics for robust ontology-based data access // Informal Proceedings of the 26th International Workshop on Description Logics (CEUR Workshop Proceedings). Ed. by T. Eiter. 2013. Pp. 53–64.
- [8] Boguslavsky I., Dikonov V., Iomdin L., Lazursky A., Sizov V., Timoshenko S. Semantic analysis and question answering: a system under development // Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference “Dialogue”. 2015. Pp. 62–79.
- [9] Brodley C. E., Stone P. (eds.) Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Quebec City, Canada: AAAI Press, 2014.
- [10] Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Poggi A., Rodriguez-Muro M., Rosati R., Ruzzi M., Savo D.F. The MASTRO system for ontology-based data access // Semantic Web. 2011. Vol. 2, № 1. Pp. 43–53.
- [11] Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R. EQL-Lite: Effective first-order query processing in description logics // Proceedings of the 20th International Joint Conference on Artificial Intelligence. Ed. by M.M. Veloso. Hyderabad, India: IJCAI, 2007. Pp. 274–279.
- [12] Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R. Tractable reasoning and efficient query answering in description logics: The DL-Lite family // Journal of Automated Reasoning. 2007. Vol. 39, № 3. Pp. 385–429.
- [13] Calvanese D., Franconi E., Haarslev V., Lembo D., Motik B., Turhan A.-Y., Tessaris S. (eds.) Proceedings of the 20th International Workshop on Description Logics (DL 2007). 2007.
- [14] Carral D., Feier C., Grau B.C., Hitzler P., Horrocks I. EL-ifying ontologies // Automated Reasoning – 7th International Joint Conference. Eds. by S. Demri, D. Kapur, C. Weidenbach. Series: Lecture Notes in Computer Science. Vienna, Austria: Springer, 2014. Pp. 464–479.
- [15] Ceri S., Gottlob G., Tanca L. What you always wanted to know about datalog (and never dared to ask) // IEEE Transactions on Knowledge and Data Engineering. 1989. Vol. 1, № 1. Pp. 146–166.
- [16] Eiter T., Glimm B., Kazakov Y., Krotzsch M. (eds.) Informal Proceedings of the 26th International Workshop on Description Logics. CEUR Workshop Proceedings. 2013.
- [17] Eiter T., Ortiz M., Simkus M. Conjunctive query answering in the description logic SH using knots // Journal of Computer and System Sciences. 2012. Vol. 78, № 1. Pp. 47–85.

- [18] Eiter T., Ortiz M., Simkus M., Tran T., Xiao G. Query rewriting for Horn-SHIQ plus rules // *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. Eds. by J. Hoffmann, B. Selman. Toronto, Canada: AAAI Press, 2012.
- [19] Glimm B., Kazakov Y., Lutz C. Status QIO: an update // *Proceedings of the 24th International Workshop on Description Logics*. Barcelona, Spain, 2011.
- [20] Glimm B., Lutz C., Horrocks I., Sattler U. Conjunctive query answering for the description logic SHIQ // *Journal of Artificial Intelligence Research*. 2008. Vol. 31. Pp. 157–204.
- [21] Glimm B., Rudolph S. Status QIO: conjunctive query entailment is decidable // *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*. Eds. by F. Lin, U. Sattler, M. Truszczynski. Toronto, Canada: AAAI Press, 2010.
- [22] Gottlob G., Kikot S., Kontchakov R., Podolskii V. V., Schwentick T., Zakharyashev M. The price of query rewriting in ontology-based data access // *Artificial intelligence*. 2014. Vol. 213. Pp. 42–59.
- [23] Jimenez-Ruiz E., Kharlamov E., Zheleznyakov D., Horrocks I., Pinkel C., Skjaeveland M. G., Thorstensen E., Mora J. BootOX: Practical Mapping of RDBs to OWL 2 // *The Semantic Web - ISWC 2015*. Cham: Springer International Publishing, 2015. Pp. 113–132.
- [24] Kikot S., Kontchakov R., Podolskii V. V., Zakharyashev M. On the succinctness of query rewriting over shallow ontologies // *Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. Eds. by T.A. Henzinger, D. Miller. Vienna, Austria: ACM, 2014.
- [25] Kikot S., Kontchakov R., Zakharyashev M. On (in)tractability of OBDA with OWL 2 QL // *Proceedings of the 24th International Workshop on Description Logics*. Barcelona, Spain, 2011.
- [26] Kikot S., Kontchakov R., Zakharyashev M. Conjunctive query answering with OWL 2 QL // *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference*. Eds. by G. Brewka, T. Eiter, S.A. McIlraith. Rome, Italy: AAAI Press, 2012.
- [27] Kikot S., Tsarkov D., Zakharyashev M., Zolin E. Query answering via modal definability with fact++: First blood // *Informal Proceedings of the 26th International Workshop on Description Logics*. Ed. by T. Eiter. CEUR Workshop Proceedings. 2013. Pp. 328–340.
- [28] Kikot S., Zolin E. Modal definability of first-order formulas with free variables and query answering // *Journal of Applied Logic*. 2013. Vol. 11, № 2. Pp. 190–216.
- [29] Kollia I., Glimm B. Optimizing SPARQL query answering over OWL ontologies // *Journal of Artificial Intelligence Research*. 2013. Vol. 48. Pp. 253–303.

- [30] Kollia I., Glimm B., Horrocks I. SPARQL query answering over OWL ontologies // The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference. Eds. by G. Antoniou et al.. Series: Lecture Notes in Computer Science. Heraklion, Greece: Springer, 2011. Pp. 382–396.
- [31] Kontchakov R., Lutz C., Toman D., Wolter F., Zakharyashev M. The combined approach to query answering in DL-lite. 2010.
- [32] Kontchakov R., Zakharyashev M. An introduction to description logics and query rewriting // Reasoning Web. Reasoning on the Web in the Big Data Era - 10th International Summer School. Eds. by M. Koubarakis et al.. Series: Lecture Notes in Computer Science. Athens, Greece: Springer, 2014. Pp. 195–244.
- [33] D. Calvanese et al. (eds.) Data complexity in the EL family of DLs // Proceedings of the 20th International Workshop on Description Logics. CEUR Workshop Proceedings. Brixen-Bressanone, Israel, 2007.
- [34] Krotzsch M., Rudolph S., Hitzler P. Complexities of horn description logics // ACM Transactions on Computational Logic. 2013. Vol. 14, № 1.
- [35] Lembo D., Lenzerini M., Rosati R., Ruzzi M., Savo D.F. Inconsistency-tolerant semantics for description logics // Web Reasoning and Rule Systems – Fourth International Conference. Eds. by P. Hitzler, T. Lukasiewicz. Series: Lecture Notes in Computer Science. Bressanone, Italy: Springer, 2010. Pp. 103–117.
- [36] Lutz C., Seylan I., Toman D., Wolter F. The Combined Approach to OBDA: Taming Role Hierarchies Using Filters // The Semantic Web – ISWC 2013. 2013. Pp. 314–330.
- [37] Motik B., Nenov Y., Piro R., Horrocks I., Olteanu D. Parallel materialisation of Datalog programs in centralised, main-memory RDF systems // Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Eds. by C.E. Brodley, P. Stone. Quebec City, Canada: AAAI Press, 2014. Pp. 129–137.
- [38] Ortiz M., Simkus M. Reasoning and query answering in description logics // Reasoning Web. Semantic Technologies for Advanced Query Answering - 8th International Summer School. Eds. by T. Eiter, T. Krennwallner. Series: Lecture Notes in Computer Science. Vienna, Austria: Springer, 2012. Pp. 1–53.
- [39] Perez-Urbina H., Motik B., Horrocks I. Tractable query answering and rewriting under description logic constraints // Journal of Applied Logic. 2010. Vol. 8, № 2. Pp. 186–209.
- [40] Podolskii V.V. Circuit complexity meets ontology-based data access // Computer Science - Theory and Applications - 10th International Computer Science Symposium in Russia. Eds. by L.D. Beklemishev, D.V. Musatov. Series: Lecture Notes in Computer Science. Listvyanka, Russia: Springer, 2015. Pp. 7–26.
- [41] Rodriguez-Muro M., Kontchakov R., Zakharyashev M. Ontology-based data access: Ontop of databases // The Semantic Web – ISWC 2013 – 12th International Semantic Web Conference. Eds. by H. Alani et al.. Series: Lecture Notes in Computer Science. Sydney, Australia: Springer, 2013. Pp. 558–573.

- [42] Romero A.A., Grau B.C., Horrocks I. More: Modular combination of OWL reasoners for ontology classification // The Semantic Web – ISWC 2012 – 11th International Semantic Web Conference. Eds. by P. Cudre-Mauroux et al. Series: Lecture Notes in Computer Science. Boston, USA: Springer, 2012. Pp. 1–16.
- [43] Rosati R. On conjunctive query answering in EL // Proceedings of the 20th International Workshop on Description Logics (CEUR Workshop Proceedings). Eds. by D. Calvanese et al. Brixen-Bressanone, Israel, 2007.
- [44] Trivela D., Stoilos G., Chortaras A., Stamou G.B. Optimising resolution-based rewriting algorithms for OWL ontologies // Journal of Web Semantics. 2015. Vol. 33. Pp. 30–49.
- [45] Venetis T., Stoilos G., Stamou G.B. Incremental query rewriting for OWL 2 QL // Proceedings of the 2012 International Workshop on Description Logics (CEUR Workshop Proceedings). Eds. by Y. Kazakov, D. Lembo, F. Wolter. Rome, Italy, 2012.
- [46] Zhou Y., Grau B.C., Nenov Y., Horrocks I. Pagoda: Pay-as-you-go abox reasoning // Proceedings of the 28th International Workshop on Description Logics (CEUR Workshop Proceedings). Eds. by D. Calvanese, B. Konev. Athens, Greece, 2015.
- [47] Zolin E. Query answering based on modal correspondence theory // Proc. of the 4th “Methods for modalities” Workshop. 2005. Pp. 21–37.

Образец цитирования

Кикоть С.П., Рыбаков М.Н. Конъюнктивные запросы к базам знаний: какую семантику выбрать? // Вестник ТвГУ. Серия: Прикладная математика. 2025. № 3. С. 5–24. <https://doi.org/10.26456/vtpmk754>

Сведения об авторах

1. Кикоть Станислав Павлович

старший научный сотрудник Высшей школы современной математики МФТИ.

Россия, 115184, г. Москва, Климентовский пер., д. 1, стр. 1, ВШМ МФТИ.

E-mail: staskikotx@gmail.com

2. Рыбаков Михаил Николаевич

ведущий научный сотрудник Высшей школы современной математики МФТИ;

доцент факультета математики НИУ ВШЭ;

доцент кафедры функционального анализа и геометрии, а также кафедры математического моделирования и вычислительной математики Тверского государственного университета.

Россия, 115184, г. Москва, Климентовский пер., д. 1, стр. 1, ВШМ МФТИ.

E-mail: m_rybakov@mail.ru

CONJUNCTIVE QUERIES TO KNOWLEDGE BASES: WHICH SEMANTICS SHOULD WE CHOOSE?

Kikot S.P.^{*}, Rybakov M.N.^{*,**,*}

^{*}Higher School of Modern Mathematics, MIPT, Moscow

^{**}HSE University, Moscow

^{***}Tver State University, Tver

Received 05.08.2025, revised 25.08.2025.

In the field of computer science known as “knowledge representation,” the following paradigm has become standard in recent years. A knowledge base refers to an architecture that combines data, general knowledge, and a reasoning mechanism that can generate new information from the underlying data and knowledge, using it to answer user queries. In this survey, we focus on the scenario where the data consists of a set of ground atoms, the knowledge is represented by an ontology in some description logic, and the query language is conjunctive queries. The aim of this work is to consolidate both theoretical and practical results about the computational complexity of query answering procedures, depending on the semantics, that is, the formal definition of what constitutes an answer to a query.

Keywords: knowledge bases, conjunctive queries, query answering, semantics.

Citation

Kikot S.P., Rybakov M.N., “Conjunctive queries to knowledge bases: Which semantics should we choose?”, *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2025, № 3, 5–24 (in Russian). <https://doi.org/10.26456/vtpmk754>

References

- [1] Antonioli N., Castano F., Coletta S., Grossi S., Lembo D., Lenzerini M., Poggi A., Virardi E., Castracane P., “Ontology-based data management for the Italian public debt”, *Formal Ontology in Information Systems – Proceedings of the Eighth International Conference, FOIS 2014*, Frontiers in Artificial Intelligence and Applications, eds. P. Garbacz, O. Kutz, IOS Press, Brazil, 2014, 372–385.
- [2] Artale A., Calvanese D., Kontchakov R., Zakharyashev M., “The DL-Lite family and relations”, *Journal of Artificial Intelligence Research*, **36** (2009), 1–69.
- [3] Bienvenu M., Bourgaux C., Goasdoue F., “Querying inconsistent description logic knowledge bases under preferred repair semantics”, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI Press, Quebec City, Canada, 2014, 996–1002.

- [4] Bienvenu M., Kikot S., Podolskii V., “Tree-like queries in OWL 2 QL: succinctness and complexity results”, *30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015*, IEEE, Kyoto, Japan, 2015, 317–328.
- [5] Bienvenu M., Ortiz M., “Ontology-mediated query answering with data-tractable description logics”, *Reasoning Web. Web Logic Rules – 11th International Summer School 2015*, Lecture Notes in Computer Science, eds. W. Faber, A. Paschke, Springer, Berlin, Germany, 2015, 218–307.
- [6] Bienvenu M., Ortiz M., Simkus M., Xiao G., “Tractable queries for lightweight description logics”, *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, ed. F. Rossi, IJCAI/AAAI, Beijing, China, 2013.
- [7] Bienvenu M., Rosati R., “New inconsistency-tolerant semantics for robust ontology-based data access”, *Informal Proceedings of the 26th International Workshop on Description Logics (CEUR Workshop Proceedings)*, ed. T. Eiter, 2013, 53–64.
- [8] Boguslavsky I., Dikonov V., Iomdin L., Lazursky A., Sizov V., Timoshenko S., “Semantic analysis and question answering: a system under development”, *Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference “Dialogue”*, 2015, 62–79.
- [9] Brodley C. E., Stone P. (eds.), *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI Press, Quebec City, Canada, 2014.
- [10] Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Poggi A., Rodriguez-Muro M., Rosati R., Ruzzi M., Savo D.F., “The MASTRO system for ontology-based data access”, *Semantic Web*, **2:1** (2011), 43–53.
- [11] Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R., “EQL-Lite: Effective first-order query processing in description logics”, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, ed. M.M. Veloso, IJCAI, Hyderabad, India, 2007, 274–279.
- [12] Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R., “Tractable reasoning and efficient query answering in description logics: The DL-Lite family”, *Journal of Automated Reasoning*, **39:3** (2007), 385–429.
- [13] Calvanese D., Franconi E., Haarslev V., Lembo D., Motik B., Turhan A.-Y., Tessaris S. (eds.), *Proceedings of the 20th International Workshop on Description Logics (DL 2007)*, 2007.
- [14] Carral D., Feier C., Grau B.C., Hitzler P., Horrocks I., “EL-ifying ontologies”, *Automated Reasoning – 7th International Joint Conference*, Lecture Notes in Computer Science, eds. S. Demri, D. Kapur, C. Weidenbach, Springer, Vienna, Austria, 2014, 464–479.
- [15] Ceri S., Gottlob G., Tanca L., “What you always wanted to know about datalog (and never dared to ask)”, *IEEE Transactions on Knowledge and Data Engineering*, **1:1** (1989), 146–166.

- [16] Eiter T., Glimm B., Kazakov Y., Krotzsch M. (eds.), *Informal Proceedings of the 26th International Workshop on Description Logics*, CEUR Workshop Proceedings, 2013.
- [17] Eiter T., Ortiz M., Simkus M., “Conjunctive query answering in the description logic SH using knots”, *Journal of Computer and System Sciences*, **78**:1 (2012), 47–85.
- [18] Eiter T., Ortiz M., Simkus M., Tran T., Xiao G., “Query rewriting for Horn-SHIQ plus rules”, *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, eds. J. Hoffmann, B. Selman, AAAI Press, Toronto, Canada, 2012.
- [19] Glimm B., Kazakov Y., Lutz C., “Status QIO: an update”, *Proceedings of the 24th International Workshop on Description Logics*, Barcelona, Spain, 2011.
- [20] Glimm B., Lutz C., Horrocks I., Sattler U., “Conjunctive query answering for the description logic SHIQ”, *Journal of Artificial Intelligence Research*, **31** (2008), 157–204.
- [21] Glimm B., Rudolph S., “Status QIO: conjunctive query entailment is decidable”, *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*, eds. F. Lin, U. Sattler, M. Truszczynski, AAAI Press, Toronto, Canada, 2010.
- [22] Gottlob G., Kikot S., Kontchakov R., Podolskii V. V., Schwentick T., Zakharyashev M., “The price of query rewriting in ontology-based data access”, *Artificial intelligence*, **213** (2014), 42–59.
- [23] Jimenez-Ruiz E., Kharlamov E., Zheleznyakov D., Horrocks I., Pinkel C., Skjæveland M. G., Thorstensen E., Mora J., “BootOX: Practical Mapping of RDBs to OWL 2”, *The Semantic Web - ISWC 2015*, Springer International Publishing, Cham, 2015, 113–132.
- [24] Kikot S., Kontchakov R., Podolskii V. V., Zakharyashev M., “On the succinctness of query rewriting over shallow ontologies”, *Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, eds. T.A. Henzinger, D. Miller, ACM, Vienna, Austria, 2014.
- [25] Kikot S., Kontchakov R., Zakharyashev M., “On (in)tractability of OBDA with OWL 2 QL”, *Proceedings of the 24th International Workshop on Description Logics*, Barcelona, Spain, 2011.
- [26] Kikot S., Kontchakov R., Zakharyashev M., “Conjunctive query answering with OWL 2 QL”, *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference*, eds. G. Brewka, T. Eiter, S.A. McIlraith, AAAI Press, Rome, Italy, 2012.
- [27] Kikot S., Tsarkov D., Zakharyashev M., Zolin E., “Query answering via modal definability with fact++: First blood”, *Informal Proceedings of the 26th International Workshop on Description Logics*, CEUR Workshop Proceedings, ed. T. Eiter, 2013, 328–340.

- [28] Kikot S., Zolin E., “Modal definability of first-order formulas with free variables and query answering”, *Journal of Applied Logic*, **11**:2 (2013), 190–216.
- [29] Kollia I., Glimm B., “Optimizing SPARQL query answering over OWL ontologies”, *Journal of Artificial Intelligence Research*, **48** (2013), 253–303.
- [30] Kollia I., Glimm B., Horrocks I., “SPARQL query answering over OWL ontologies”, *The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference*, Lecture Notes in Computer Science, eds. G. Antoniou et al., Springer, Heraklion, Greece, 2011, 382–396.
- [31] Kontchakov R., Lutz C., Toman D., Wolter F., Zakharyashev M., *The combined approach to query answering in DL-lite*, 2010.
- [32] Kontchakov R., Zakharyashev M., “An introduction to description logics and query rewriting”, *Reasoning Web. Reasoning on the Web in the Big Data Era - 10th International Summer School*, Lecture Notes in Computer Science, eds. M. Koubarakis et al., Springer, Athens, Greece, 2014, 195–244.
- [33] D. Calvanese et al. (eds.), “Data complexity in the EL family of DLs”, *Proceedings of the 20th International Workshop on Description Logics*, CEUR Workshop Proceedings, Brixen-Bressanone, Israel, 2007.
- [34] Krotzsch M., Rudolph S., Hitzler P., “Complexities of horn description logics”, *ACM Transactions on Computational Logic*, **14**:1 (2013).
- [35] Lembo D., Lenzerini M., Rosati R., Ruzzi M., Savo D.F., “Inconsistency-tolerant semantics for description logics”, *Web Reasoning and Rule Systems – Fourth International Conference*, Lecture Notes in Computer Science, eds. P. Hitzler, T. Lukasiewicz, Springer, Bressanone, Italy, 2010, 103–117.
- [36] Lutz C., Seylan I., Toman D., Wolter F., “The Combined Approach to OBDA: Taming Role Hierarchies Using Filters”, *The Semantic Web – ISWC 2013*, 2013, 314–330.
- [37] Motik B., Nenov Y., Piro R., Horrocks I., Olteanu D., “Parallel materialisation of Datalog programs in centralised, main-memory RDF systems”, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, eds. C.E. Brodley, P. Stone, AAAI Press, Quebec City, Canada, 2014, 129–137.
- [38] Ortiz M., Simkus M., “Reasoning and query answering in description logics”, *Reasoning Web. Semantic Technologies for Advanced Query Answering - 8th International Summer School*, Lecture Notes in Computer Science, eds. T. Eiter, T. Krennwallner, Springer, Vienna, Austria, 2012, 1–53.
- [39] Perez-Urbina H., Motik B., Horrocks I., “Tractable query answering and rewriting under description logic constraints”, *Journal of Applied Logic*, **8**:2 (2010), 186–209.
- [40] Podolskii V.V., “Circuit complexity meets ontology-based data access”, *Computer Science - Theory and Applications - 10th International Computer Science Symposium in Russia*, Lecture Notes in Computer Science, eds. L.D. Beklemishev, D.V. Musatov, Springer, Listvyanka, Russia, 2015, 7–26.

- [41] Rodriguez-Muro M., Kontchakov R., Zakharyashev M., “Ontology-based data access: Ontop of databases”, *The Semantic Web – ISWC 2013 – 12th International Semantic Web Conference*, Lecture Notes in Computer Science, eds. H. Alani et al., Springer, Sydney, Australia, 2013, 558–573.
- [42] Romero A.A., Grau B.C., Horrocks I., “More: Modular combination of OWL reasoners for ontology classification”, *The Semantic Web – ISWC 2012 – 11th International Semantic Web Conference*, Lecture Notes in Computer Science, eds. P. Cudre-Mauroux et al., Springer, Boston, USA, 2012, 1–16.
- [43] Rosati R., “On conjunctive query answering in EL”, *Proceedings of the 20th International Workshop on Description Logics* (CEUR Workshop Proceedings), eds. D. Calvanese et al., Brixen-Bressanone, Israel, 2007.
- [44] Trivela D., Stoilos G., Chortaras A., Stamou G.B., “Optimising resolution-based rewriting algorithms for OWL ontologies”, *Journal of Web Semantics*, **33** (2015), 30–49.
- [45] Venetis T., Stoilos G., Stamou G.B., “Incremental query rewriting for OWL 2 QL”, *Proceedings of the 2012 International Workshop on Description Logics* (CEUR Workshop Proceedings), eds. Y. Kazakov, D. Lembo, F. Wolter, Rome, Italy, 2012.
- [46] Zhou Y., Grau B.C., Nenov Y., Horrocks I., “Pagoda: Pay-as-you-go abox reasoning”, *Proceedings of the 28th International Workshop on Description Logics* (CEUR Workshop Proceedings), eds. D. Calvanese, B. Konev, Athens, Greece, 2015.
- [47] Zolin E., “Query answering based on modal correspondence theory”, *Proc. of the 4th “Methods for modalities” Workshop*, 2005, 21–37.

Author Info

1. Kikot Stanislav Pavlovich

Senior Researcher at Higher School of Modern Mathematics, MIPT.

Russia, 115184, Moscow, Klimentovskiy Pereulok, 1 bld. 1, Higher School of Modern Mathematics, MIPT. E-mail: staskikotx@gmail.com

2. Rybakov Mikhail Nikolayevich

Leading Researcher at Higher School of Modern Mathematics, MIPT;
Associate Professor at Faculty of Mathematics, HSE University;
Associate Professor at the Department of Functional Analysis and Geometry, as well as the Department of Mathematical Modeling and Computational Mathematics, Tver State University.

Russia, 115184, Moscow, Klimentovskiy Pereulok, 1 bld. 1, Higher School of Modern Mathematics, MIPT. E-mail: m_rybakov@mail.ru