

Г.Ю. Лобанов, Д.В. Хизанишвили
ОБЗОР ПРЕДСТАВЛЕНИЙ ОБ ОНТОЛОГИЯХ: КОМПЬЮТЕРНАЯ НАУКА, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ТЕОРИЯ АРГУМЕНТАЦИИ*

В статье представлен обзор западных исследований онтологий в области компьютерной науки, искусственного интеллекта и теории аргументации. Приведена классификация и выделены критерии конструирования онтологий. Даны примеры онтологий и информационных систем, построенных на основе онтологий.

The article presents an overview of Western studies of ontologies in the field of computer science, artificial intelligence and the theory of argumentation. It shows the classification and selects criteria for ontologies constructions and exemplifies ontologies and information systems that are based on ontologies.

Ключевые слова: *онтология, компьютерная наука, искусственный интеллект, теория аргументации, модель аргументации.*

Keywords: *ontology, computer science, artificial intelligence, argumentation theory, argumentation model*

С конца XX века онтологии становятся популярной темой исследований в таких разделах науки искусственного интеллекта, как инженерия знаний, обработка естественного языка, представление знаний, интеллектуальная интеграция информации, извлечение информации и управление знанием. Интерес обуславливается возможностями онтологий к обмену и повторному использованию совокупности знаний (см. [17]).

* Статья выполнена в рамках проекта РФФИ №12-06-00285а "Место и роль онтологий в моделировании аргументации».

Определение онтологии

В философии, термин «онтология» относится к «частной теории о природе бытия или видов существования». Это широкое определение может быть интерпретировано множеством способов, в зависимости от метафизической позиции по отношению к существующему. Поэтому исследователи инженерии знаний предложили более узкое определение онтологии, ориентированное на искусственный интеллект, избегающее обращения к вопросам об отношении к реальному, в терминах *репрезентации* и *концептуализации* (см. [10]).

В компьютерных науках широко распространено определение онтологии, предложенное Т. Грубером: «онтология суть явная спецификация концептуализации» [6, 1]. Он утверждает, что для систем, основанных на знаниях, существующее совпадает с тем, что может быть репрезентировано. Когда знание о некоторой предметной области представлено декларативным формализмом, множество объектов, которое может быть представлено, называется пространством дискурса. Множество объектов и поддающиеся описанию отношения между ними отражаются в репрезентативном словаре, при помощи которого программа, основанная на знании, представляет само знание. Таким образом, Грубер предлагает описывать онтологию программы определением множества терминов репрезентации. В такой онтологии определения соотносят имена существительных в пространстве дискурса (например, классами, отношениями, функциями и другими объектами) с текстом, доступным для человеческого понимания, описывающим, что имена призваны обозначать, и формальными аксиомами, ограничивающими интерпретации и верно сформулированные использования терминов.

Концептуализация

Понятие концептуализации, использованное Грубером в определении онтологии, впервые введено в употребление Генесеретом и Нильсоном. Под концептуализациями они понимали экстенциональные сущности, определенные в терминах наборов отношений (см. [4]). Концептуализации таким образом критически рассмотрены на том основании, что подобное экстенциональное понимание слишком отдаляет их от естественного языка, в котором превалирует интенциональный контекст (см. [9]).

Грубер концептуализацию определял как абстрактный взгляд на мир, который мы хотим представить для некоторых целей. Каждая база знаний, система, основанная на знании или knowledge-level agent связаны с некоторой концептуализацией, явно или неявно (см. [7]).

Таким образом, онтология не соотносится с вопросом онтологического реализма, раскрывающегося в вопросе об истинности концептуализации некоторой независимо существующей реальности. Скорее это результат строго прагматичного процесса, который начинается с концептуализации, и идет от нее к описанию соответствующих областей объектов, под которыми часто ошибочно понимают концепты. Эти области ни что иное, как узлы или элементы моделей данных приближенных к реальности, разработанные под конкретную цель (см. [15]).

Типы онтологий

Онтологии можно классифицировать в соответствии с двумя измерениями: размер и тип структуры концептуализации и предмет концептуализации (см. [10]). В соответствии с первым измерением различаются три категории:

- Терминологические онтологии, такие как лексиконы. Устанавливают термины, которые используются для представления знания в области дискурса. Примером подобной онтологии является семантическая сеть в UMLS (Unified Medical Language System) (см. [12]).

- Информационные онтологии, которые определяют структуру записей баз данных. Схемы баз данных являются примером таких онтологий. Первый уровень модели PEN&PAD, платформы моделирования медицинских записей пациентов, является типичным примером такой онтологии. На этом уровне модель обеспечивает платформу для записи основных наблюдений за пациентами, но не различает внешние признаки, симптомы, лечения и пр. (см. [13]).

- Онтологии моделирования знания устанавливают концептуализацию знания. По сравнению с информационными онтологиями онтологии моделирования знания обычно обладают более богатой внутренней структурой. Более того, подобные онтологии часто приведены в соответствие со специальными целями использования знания, которое они описывают. Возвращаясь к модели PEN&PAD, второй ее уровень является примером онтологии моделирования знания. На этом уровне наблюдения первого уровня сгруппированы для описания процессов принятия решений.

Другое основание, по которому можно различить онтологии, это субъект концептуализации. По этому основанию можно различить четыре категории:

- Прикладные онтологии содержат все определения, необходимые для моделирования знания, требуемого конкретным приложением. Обычно, прикладные онтологии представляют собой смесь концептов, взятых из областных онтологий и основных онтологий, описанных ниже. Более того, прикладные онтологии могут содержать расширения, характерные для

конкретного метода или конкретной задачи. Сами по себе прикладные онтологии не предполагают повторного использования вне изначально заданного контекста. Они могут быть получены извлечением теорий из библиотеки онтологий, которые в свою очередь хорошо настроены на некоторое приложение. Подобным образом термин прикладные онтологии использован в PROTEGE-II (см. [19]).

- *Областные онтологии* выражают концептуализации специальных областей. Современные методологии инженерии знаний ясно различают областные онтологии и областные знания. Тогда как областные знания описывают фактические ситуации в определенной области, областные онтологии накладывают граничные условия на структуру и содержание областного знания.

- *Общие онтологии* схожи с областными онтологиями, однако концепции, которые они определяют, предполагаются общими среди множества областей. Так, общие онтологии определяют такие концепты как состояние, событие, процесс, действие, компонент, и т. д. Несмотря на то, что граница между общими онтологиями и областными онтологиями нечеткая, различие интуитивно значимо и полезно в строительстве библиотек.

- *Репрезентативные онтологии* развертывают концептуализацию, стоящую за формализмом репрезентативного знания (см. [1]). Они предназначены быть нейтральными по отношению к сущностям мира (см. [8]). Это означает, что они обеспечивают репрезентативную платформу, не делая заявлений о мире. Областные онтологии и общие онтологии описаны с использованием примитивов репрезентативной онтологии. Примером этой категории является Frame Ontology, использованная в Ontolingua (см. [6]).

Критерии конструирования онтологий

Ниже представлен предложенный Грубером набор критериев, предваряющих конструирование онтологий, задачей которых является обеспечение обмена знаниями и взаимодействия между программами распределенной концептуализации.

В процессе принятия решения о том, как представить что-либо в онтологии, выносятся суждения конструктивного характера. Для руководства и оценки конструирования необходимы объективные критерии, основанные скорее на предназначении результата, чем на априорных представлениях о естественной Истине (см. [7]).

1. *Ясность.* Онтология должна эффективно передавать предполагаемое значение определенного термина. Определения должны быть объективными. Тогда как мотивация определения концепта может возникать из социальных ситуаций или вычислительных потребностей, определение должно быть независимо от социального или вычислительного контекста. Формализм является средством достижения этой цели. Всегда, когда это возможно, определение должно быть сформулировано по средством логически аксиом. Полное определение (предикат, определенный необходимыми и достаточными условиями) предпочтительнее частичному (определен лишь либо необходимым, либо достаточным условием). Все определения должны быть построены на естественном языке.

2. *Связность.* Онтология должна быть связной, что значит поддерживать умозаключения, согласующиеся с определениями. По крайней мере, определяющие аксиомы должны быть согласованы. Связность также должна быть применима к определенным на естественном языке концептам.

3. *Расширяемость.* Онтология должна быть сконструирована таким образом, чтобы было возмож-

но предвосхищение использования общего словаря. Она должна предлагать концептуальное основание для предполагаемого набора задач, и представления должны быть созданы так, чтобы онтологии можно было расширять и специализировать монотонно – не ясен смысл этого суждения. Другими словами, чтобы можно было, основываясь на существующем словаре, в дальнейшем определять новые термины для специального использования таким образом, чтобы не потребовался пересмотр и исправление уже существующих терминов.

4. *Минимальная погрешность кодирования.* Концептуализация должна быть специфицирована на уровне знаний без использования и зависимости от частного символического уровня кодирования. Погрешности кодирования возникают, когда выбор способа репрезентации основывается лишь только исходя из соображений удобства обозначения или реализации. Погрешность кодирования должна быть минимизирована, потому что агенты доступа к распределенному общему знанию могут быть сконструированы на основе различных репрезентативных систем и в различных стилях репрезентации.

5. *Минимальные онтологические обязательства (ontological commitment).* Онтология должна иметь минимальные онтологические обязательства, необходимые для поддержания предполагаемого обмена знанием. Онтология должна делать как можно меньше утверждений о моделируемом мире, что даст сторонам, связанным с онтологией, свободу специализировать и экземплифицировать онтологию согласно своим нуждам. Исходя из того, что онтологические обязательства основаны на согласованном использовании словаря, они могут быть минимизированы указанием слабейшей теории (позволяющей сконструировать большее моделей) и определением лишь тех терминов,

которые существенны для передачи знаний согласованных с этой теорией.

KIF

Язык KIF (Knowledge Interchange Format) разработан Майком Генесереттом и его коллегами в Стенфорде (см. [3]). Несмотря на то, что этот язык не предназначен специально для решения задач, возникающих при работе с онтологиями, KIF тем не является вехой в развитии применения онтологий в качестве метода решения проблем обмена знанием и интеграции знаний. KIF можно рассматривать как один из вариантов языка вычислений предикатов первого порядка, разработка которого имела своей целью развитие выразительного, гибкого, доступного для понимания компьютером и человеком носителя информации при обмене базами знаний.

Существование подобного языка означает, что каждая система, обеспечивающая переводимость собственного синтаксиса в синтаксис KIF, может внутренне обрабатывать данные своими способами а также общаться с пользователем-человеком иными способами, с гарантией, что результаты операций системы будут автоматически совместимы с другими системами, также структурированными для проведения обмена с KIF.

У языка KIF три основных особенности:

1. Стандартная семантика, основанная на теории множеств, которая, в терминологии компьютерных наук, может быть отнесена скорее к дескриптивным, нежели процедурным семантикам;
2. Логическая полнота, это означает, что язык обладает всеми выразительными ресурсами вычислений предикатов первого порядка;

3. Возможность поддержки представлений о представлениях, или знания о знании.

Основное влияние KIF оказал в области синтаксиса. Его система обозначений предложила удобные способы, которыми формулы логики первого порядка могут быть введены с помощью клавиатуры, без необходимости использования шрифтов со специальными символами. Влияния KIF таким образом покоится на тех частях его технического аппарата, которые прямо связаны с его синтаксисом.

Семантическая сторона KIF опирается на техническое понятие концептуализации предложенное Генесеретом и Нильсоном в 1987 году (см. [4]). Концептуализации рассматриваются как теоретико-множественные объекты, выстроенные из компонентов двух видов: пространства дискурса (множество объектов, предположительно существующих в мире) и множество релевантных свойств, отношений и функций, под которыми в свою очередь понимаются множества упорядоченных кортежей (наборов взаимосвязанных величин). А именно, отношения и функции представлены в качестве множеств конечных списков объектов.

Так, например, с некоторым грубым упрощением концептуализация ситуации, когда John целует Mary, будет выглядеть следующим образом:

$\langle \{John, Mary\}, \{male_person, female_person, kissing\} \rangle$,

где *male_person* и *female_person* – свойства объектов John и Mary, а *kissing* – их отношение.

Таким образом, концептуализация является объектом, тип которого узнаваем в стандартной теоретико-множественной модельной теории. С учетом концептуализаций, отдельные термины KIF обозначают объекты в соответствующем пространстве дискурса, термины предикатов KIF являются установленными значениями множества соответствующих свойств и от-

ношений. Семантика относительна к концептуализации в том смысле, что утверждение истинно либо ложно исходя из концептуализации, с построения которой оно и начинается.

Пространство дискурса состоит из объектов. Объекты в свою очередь состоят как из частных элементов, так и из множеств или классов. KIF в качестве составной части реализует теорию множество Неймана-Бернейса-Геделя (Neumann-Bernays-Gödel).

Каждое пространство дискурса должно включать, по крайней мере, следующие объекты:

- Комплексные числа, которые могут быть рассмотрены как пары действительных чисел, одно из которых представляет действительную часть комплексного числа, а другое мнимую. (Действительные числа суть комплексные, мнимая часть которых равна нулю. Платформа KIF таким образом позволяет хранить действительные, рациональные и целые числа, а также проводить ряд арифметических, тригонометрических и логарифмических операций).

- Объект, являющийся условным значением функции для несущественных комбинаций аргументов.

- Все конечные списки объектов и все множества объектов;

- Внутренние слова и выражения KIF, представленные в качестве списка терминов, которые в свою очередь также могут быть списками.

Последнее позволяет строить репрезентацию репрезентации посредством платформы KIF, а именно создавать выражения из объектов пространства дискурса и в то же время включать предикат истинности и средства обработки выражений, такие как операторы вынесения за скобки или приведения объема некоторого понятия. Анализ посредством списков означает, что KIF имеет также способность анализировать внут-

ренние структуры выражений. Ссылки на выражения могут быть построены с помощью оператора цитирования, их свойства могут быть исследованы, они могут быть представлены численно, позволяя тем самым формулировать схемы аксиом. Это позволяет также приведение в количественную форму частей выражений. Оператор истинности KIF может быть применен как к вынесенным, так и не вынесенным за скобки выражениям с условиями, позволяющими избежать парадоксов. Центральная роль списков в KIF придает ему сходство с такими логическими языками программирования, как LISP (см. [16]).

Ontolingua

На основе KIF Том Грубер со своими коллегами из Стенфордского исследовательского института разработали язык более пригодный для представления онтологий известный как Ontolingua (см. [5]), разработанный в качестве общего языка специалистов, работающих в области создания онтологий. Ontolingua построен на основе KIF 3.0, но имеет совсем иное предназначение. Тогда как KIF задуман как интерфейс между системами представления знаний, Ontolingua предполагался в качестве интерфейса между онтологиями. Он предоставляет рабочее окружение и набор программных инструментов разработанных для обеспечения гетерогенным онтологиям связи посредством единой платформы через перевод на единый язык.

Ontolingua добавляет в исходное ядро языка KIF возможности, которые позволяют строить представления структурных компонентов онтологии. Прежде всего, он позволяет расширять обработку отношений (с почти исключительным фокусом на бинарные отношения, которые рассматриваются как множества бинарных списков), и вводит понятие класса. Классы формально определяются как унарные отношения, то

есть являются множествами списков единичной длины. Элементы листа обязательно единичны и называются реализациями класса. В дополнение Ontolingua вводит понятие единичной сущности (Individual Thing), класса объектов, не являющихся множествами, но которые могут быть реализованы в качестве множеств. Также существует класс сущности Thing, который определен как класс, заключающий в себе все объекты, допускающие пребывание в классе, а именно единичные сущности и простые множества (Simple Set). Класс сущности соотносится с классом ограниченных сущностей в KIF.

Ни KIF, ни Ontolingua не принимают идею единой распределенной онтологии. Авторы Ontolingua и KIF не предполагают включить такие понятия, как процесс, время, материя и сознание в свои платформы. Основной целью создателей Ontolingua было скорее собрание большого количества различных, специализированных онтологий, и обеспечить лингвистическими ресурсами движение от одной из них к другой.

Дескрипционные логики

Дескрипционные логики, известные также как терминологические логики, отражают попытку найти фрагменты логики первого порядка с максимально высокой выразительной мощностью, остающиеся при том разрешимыми и выводимыми, чтобы служить основой эффективных доказательных процедур.

Дескрипционные логики фокусируются на представлении концептов и на иерархической классификации. Предоставляя достаточно богатое множество описаний концептов, доказательный аппарат, использующий дескрипционные логики произведет соответствующее дерево категоризации. Цель дескрипционных логик таким образом в выведение так называемых Is-A отношений (Персона – Сущность-с-

параметром-высота, Собака — млекопитающее). Каждая дискрепционная логика формально является подмножеством логики первого порядка с определенными особенностями второго порядка. Дескрипционная логика содержит унарные предикаты, бинарные предикаты и единичные константы. Важной особенностью дескрипционной логики является наличие классов. Это типы объектов определенные намеренно в терминах дескрипций, указывающих свойства, которыми обладает объект данного типа.

Дескриптивные логики представлены системами: BACK, CLASSIC, CRACK, FLEX, K-REP, KL-ONE, KRIS, LOOM. Одно из их основных использований в составлении обращений к базам знаний (см. [16]).

Медицинские онтологии

Информационные системы, основанные на онтологиях, нашли широкое применение в медицине. Ниже перечислены три самых влиятельных них:

- GALEN (Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine). Онтология, обеспечивающая проведение хирургических операций (см. [14]).
- UMLS (Unified Medical Language System). Синтез медицинских словарей с более чем 10 000 000 концептов (см. [20]).
- SNOMED (Systematized Nomenclature of Medicine). Всеобщий справочник для сравнения и сбора данных в течение всего процесса лечения (см. [11]).

Онтологии в аргументации

В теорию аргументации термин «онтология» был принесен из компьютерных наук, поэтому аргумента-

тивные онтологий являются частным случаем описанных в предыдущей части статьи онтологий. Задачей аргументативных онтологий является репрезентация существенных сторон процесса аргументации при помощи стандартного, как правило, формализованного, набора средств. Под аргументацией чаще всего принято понимать процесс коммуникации, в котором одна сторона пытается убедить другую в истинности/ложности некоторого суждения, называемого тезисом, при помощи других суждений, называемых аргументами. Таким образом, аргументация представляет собой речевой акт, состоящий из набора аргументов и тезиса, определенным образом связанных между собой. Ввиду этого существенными составляющими аргументации в самом общем виде становятся аргументы (и их компоненты) и связи между ними, тезис, связи между аргументами и тезисом. Данные составляющие аргументации и призваны эксплицировать такие репрезентационные системы как онтологии.

В аргументации онтологии представлены в основном так называемыми аргументационными схемами или диаграммами. Приведем примеры двух наиболее известных аргументационных онтологий.

Стандартная модель.

Стандартную модель также называют моделью Уолтона. «Под стандартным подходом мы понимаем систему репрезентации макроструктуры аргументации посредством диаграмм, в которых составляющие аргументацию высказывания представлены главным образом заключенными в круги цифрами, а логические связи представлены стрелками» [2, 3]. Таким образом, простейший вид диаграммы в стандартном подходе представляет собой направленный граф, состоящий из двух узлов, изображаемых кругами, и стрелки, соединяющей эти узлы.



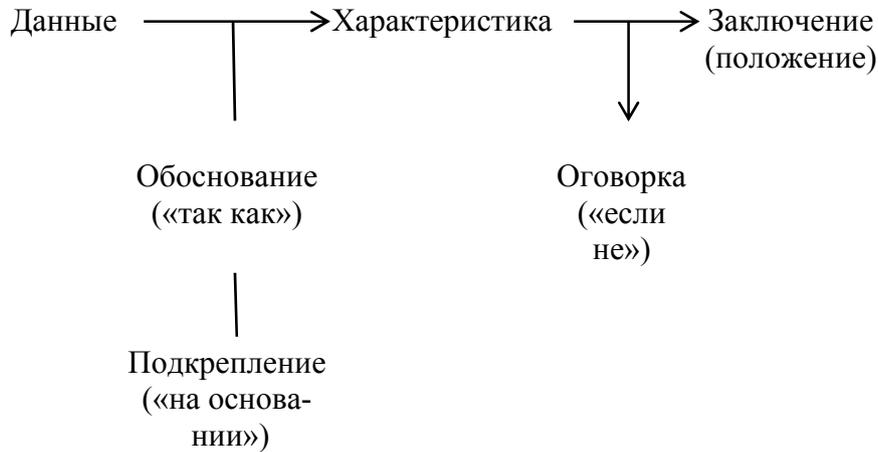
Данная схема репрезентирует единичный (single) аргумент, состоящий из одного аргумента, тезиса и связи (как правило, логической) между ними. Подобный способ построения онтологий позволяет также создавать диаграммы, репрезентирующие и другие типы аргументации, такие как последовательная (serial), связанная (linked), сходящаяся (convergent) и расходящаяся (divergent) аргументации, а также различные их комбинации.

Модель Тулмина.

Онтология, предложенная С. Тулминым в классической работе *The Uses of Argument*, имеет более сложную структуру. В ней выделяется шесть типов элементов, встречающихся в различных актах аргументации. Перечислим эти элементы.

1. Положение (Claim) - суждение, которое обосновывается в процессе аргументации;
2. Данные (Data) - суждения, обосновывающие или поддерживающие положение.
3. Обоснование (Warrant) - основания рассуждения, которое связывает данные с положением.
4. Подкрепление (Backing) - суждения, выдвигаемые в поддержку обоснования.
5. Характеристика (Qualifier) - определение характера утверждений, содержащихся в положении, обосновании и подкреплении (необходимые или возможные, достоверные или вероятные).
6. Оговорка (Rebuttal) - описание контрпримеров и контрдоводов.

Получающаяся в итоге схема аргументации выглядит следующим образом:



Подход Тулмина позволяет реконструировать аргументацию с большим приближением к естественным процессам, поскольку учитывает различные модальности и оговорки, ограничивающие условия действия аргумента.

Список литературы

1. *Davis R., Shrobe H., Szolovits P.* What Is a Knowledge Representation? // *AI Magazine* 14(1), 1993. P. 17 – 33.
2. *Freeman J.* Argument Structure. Lawrence Erlbaum Associates, 2005.
3. *Genesereth, M. R., Fikes, R. E.* KIF Version 3.0. Reference Manual // *Technical Report Logic 92*, no. 1, 1992.
4. *Genesereth, M. R., Nilsson, N. J.* Logical Foundations of Artificial Intelligence. Los Altos, California. Morgan Kaufmann, 1987.
5. *Gruber, T. R.* Ontolingua: A Mechanism to Support Portable Ontologies. Stanford Knowledge Systems Laboratory, 1992.

6. *Gruber, T. R.* Translation Approach to Portable Ontology Specifications // Knowledge Acquisition 5, no. 2, 1993. P. 199–220.
7. *Gruber, T. R.* Toward Principles for the Design of Ontologies Used for Knowledge Sharing // International Journal of Human and Computer Studies 43, no. 5/6, 1995. P. 907–928.
8. *Guarino, N., Boldrin, L.* Ontological Requirements for Knowledge Sharing. Chambery, France, 1993.
9. *Guarino, N.* Formal Ontology in Information Systems. Amsterdam, Berlin, Oxford, Tokyo, Washington, 1998.
10. *Van Heijst, G., Schreiber, A. T., Wielinga, B. J.* Using Explicit Ontologies in KBS Development // International Journal of Human-computer Studies 46, no. 2–3, 1997. P. 183–292.
11. International Health Terminology Standards Development Organisation. SNOMED CT // www.ihtsdo.org/snomed-ct
12. *Lindberg, D. A. B., Humphreys, B. L., McCray, A. T.* Unified Medical Language System // Methods of Information in Medicine 32, 1993. P. 281 – 291.
13. *Rector, A. L., Nowlan, W. A., Kay, S., Goble, C. A., Howkins, T. J.* A Framework for Modelling the Electronic Medical Record // Methods of Information in Medicine 32, 1993. P. 109 – 119.
14. OpenGALEN // www.opengalen.org
15. *Smith, B.* Ontology // Blackwell Guide to the Philosophy of Computing and Information. Oxford: Blackwell, 2003. P. 155-166
16. *Smith, B.* Ontology and Information Systems, n.d. // [http://ontology.buffalo.edu/ontology\(PIC\).pdf](http://ontology.buffalo.edu/ontology(PIC).pdf)
17. *Studer, R., Benjamins, V.R., Fensel, D.* Knowledge Engineering: Principles and Methods // Data & Knowledge Engineering 25, no. 1, 1998. P. 161–197.

18. *Toulmin, S.* The Uses of Argument. Cambridge University Press, 2003.
19. *Tu, S., Eriksson, H., Gennari, J.H., Shahar, Y., Musen, M.A.* Ontology-based Configuration of Problem-solving Methods and Generation of Knowledge PROTOGE-II to Protocol-based Decision Support. Acquisition Tools: The Application of PROTEGE-II // Artificial Intelligence in Medicine 7, 1995. P. 257 - 289.
20. U.S. National Library of Medicine. National Institutes of Health. Unified Medical Language System // www.nlm.nih.gov/research/umls/

Об авторах

Лобанов Глеб Юрьевич – аспирант кафедры философии Балтийского федерального университета им. И. Канта, lobanov.gleb@gmail.com.

Хизанишвили Давид Васильевич – ассистент кафедры философии Балтийского федерального университета имени Иммануила Канта, khizza@gmail.com.

About author

Gleb Lobanov, postgraduate student, Department of Philosophy, Immanuel Kant Baltic Federal University, lobanov.gleb@gmail.com.

David Khizanishvili, assistant lecturer, Department of Philosophy, Immanuel Kant Baltic Federal University, khizza@gmail.com.