

УДК 004.942

А.В. Прохоров

Национальный аэрокосмический университет имени Н.Е. Жуковского «ХАИ», Харьков

МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ОНТОЛОГИЧЕСКИХ ЗНАНИЙ С УЧЕТОМ НЕОПРЕДЕЛЕННОСТЕЙ И ПРОТИВОРЕЧИЙ

Проведен анализ особенностей представления в онтологиях неопределенностей и работы с противоречивыми знаниями. Рассматривается математический аппарат теории категории, который выступает в роли универсального формализма онтологического описания и является независимым от выбора конкретного языка описания онтологий и специфики реализации. Для обработки неопределенностей и противоречий, возникающих в процессе объединения онтологий и функционирования распределенной интеллектуальной системы выбрано многозначное исчисление присутствия. Для интеграции с существующими в системе знаниями предлагается конвертация онтологии на OWL в модель на многозначном исчислении присутствия, для которого реализован логический вывод.

Ключевые слова: онтология, OWL, распределенная интеллектуальная система, теория категорий, неопределенности, противоречия.

Введение

В настоящее время социальные сети, Интернет вещей и мобильные решения приводят к экспоненциальному росту объемов разнообразных данных, что в свою очередь делает актуальной и важной задачей создание новых эффективных методов и инструментов анализа, прогнозирования, извлечения знаний, машинного обучения. При этом семантические связи между данными повышают их ценность и предоставляют дополнительные возможности для информационного поиска, поскольку, интегрированные в единую семантическую паутину, представляют собой пространство знаний о некоторой предметной области. Развитие концепции Semantic Web привело к появлению связанных открытых данных LOD (Linked Open Data), где данные представлены в RDF-формате (Resource Description Framework), т.е. триплетов вида «субъект – предикат – объект». В этой связи интеллектуальные системы становятся всё более привлекательными в свете растущих вычислительных мощностей и возможностей таких систем по логической обработке уже накопленной в огромных количествах информации. Здесь следует отметить IBM Watson – суперкомпьютер фирмы IBM, оснащённый вопросно-ответной системой искусственного интеллекта в рамках проекта IBM DeepQA. Watson Services for Bluemix является средством быстрого макетирования и построения когнитивных приложений в облаке.

Важнейшим элементом распределенных интеллектуальных систем являются онтологии, которые позволяют соответствующим программным компонентам – интеллектуальным агентам – автоматически производить взаимодействие и обработку семантики информационных ресурсов и сервисов сети Интернет для эффективного решения различных прикладных задач.

Одной из задач при построении глобальной системы формализованных знаний является правильный выбор логической схемы. Разнообразные диалекты логических языков и соответствующие логические исчисления могут служить гибким инструментом для описания поведения объектов в рамках предметных областей. Для онтологий наибольшее распространение получила дескриптивная логика, которая является подмножеством логики предикатов первого порядка, ограниченным таким образом, чтобы гарантированно выполнялось любая процедура вывода (вычислительная разрешимость). Но с учетом особенностей работы с несколькими онтологиями важной задачей является обеспечение возможности представлять в онтологиях неопределенности и работать с противоречивыми знаниями.

Анализ последних исследований и публикаций. Поскольку онтология определяет термины конкретной области знаний, она должна описываться формальным языком, основанным на принципах математической логики. Тогда для классов объектов, их свойств и отношений между ними можно сформулировать четкие, детализированные и целостные определения. В свою очередь, средства обработки онтологии смогут автоматически вывести некоторые заключения, основываясь на принципах математической логики.

Сегодня наиболее распространённой онтологической моделью, рекомендуемой консорциумом W3C, является язык OWL. В нём определены три диалекта [1]: *OWL Lite* – классификационная иерархия сущностей и простые условия согласованности сущностей; *OWL DL (Description Logic* – дескриптивная логика – разрешимое подмножество логики предикатов первого порядка) – максимальная степень выразительных возможностей языка без потери вычислительной полноты и разрешимости; *OWL Full* – очень высокая вырази-

тельность (метаклассы, классы как значения) и полная синтаксическая свобода RDF (*Resource Description Framework*) с потерей гарантированности вычислительной полноты и разрешимости.

Именно уровень OWL DL ориентирован на существующие сегодня системы описания знаний и системы логического программирования и вывода, которые решают следующие задачи: проверка корректности онтологии, обработка запросов в терминах онтологии, отображение и интеграция онтологий.

Для полной поддержки вывода OWL DL необходимо использовать внешние машины вывода, такие как FaCT (бесплатная DL машина вывода – разработка профессора Манчестерского университета Яна Хоррокса), RacerPro (коммерческая разработка компании Racer Systems GmbH & Co., Германия), Pellet (открытая машина вывода – разработка лаборатории MIND LAB Мэрилендского университета) и другие.

К основным проблемам использования онтологических моделей компонентами распределенных интеллектуальных систем можно отнести неопределенности и противоречия.

Исследование основных современных логических анализаторов (Pellet, HermiT, FaCT++) показало, что они не могут справиться с онтологиями большого размера, так как изначально были рассчитаны на языки с большой экспрессивностью и выполняют процесс классификации онтологии путём итеративного построения модели для каждой пары концептов и поиск противоречий в ней.

Таким образом, применяемые сегодня машины вывода на OWL DL ориентированы на работу с классами и плохо масштабируются для поддержки сотен тысяч экземпляров классов.

В целом, неопределенность можно рассматривать:

- неопределенность в определениях, неточность терминов (в запросах пользователей, в контенте ресурсов, а также при необходимости использовать различные базы знаний с перекрывающимися, но тонко различающимися концептами). Эти проблемы частично решаются с помощью нечеткой логики [2];

- неопределенность в значениях (точные термины с неопределенными значениями). Эти проблемы частично решаются с использованием вероятностного подхода (probabilistic reasoning technics).

Логические противоречия неизбежно возникают при разработке больших онтологий и при объединении онтологий из различных источников. Дедуктивный логический вывод не может с этим справиться, потому что «из противоречия можно вывести все» [3]. Одним из вариантов решения проблемы является применения defeasible reasoning [4] и paraconsistent reasoning [5].

Сам по себе OWL не предоставляет никаких средств для представления неопределенных знаний.

Во многом это происходит потому, что на данный момент нет единого мнения насчет того, как именно они должны интерпретироваться, насколько это будет полезно в реальных задачах, а также насколько проблемной будет реализация. Одной из реализаций представление неопределенности является Pronto.

В настоящее время существуют готовые программные решения для некоторых операций объединения онтологических моделей. Одним из наиболее популярных является Jena[6], где поддерживаются базовые операции объединения онтологий. Некоторые возможности объединения онтологий также обеспечивает редактор онтологий Protégé и некоторые плагины к нему.

Постановка задачи исследования

Проведенный анализ показал, что логические исчисления для решаемых задач оказываются неприменимы в силу их неконструктивности и экстенциональной теоретико-множественной основы. Альтернативой теории множеств является логико-категорный подход [7], позволяющий рассматривать предметную область как сложную динамическую систему, состоящую из объектов различного характера, интерпретируемых на множестве их свойств. Кроме того, булевы топосы категории, как известно, могут использоваться для семантических конструкций логических теорий и именно в топосах получается очень естественно выразить многозначные логики.

Поэтому **целью данной работы** является рассмотрение задач, связанных с распределенным накоплением, обработкой и использованием знаний в интеллектуальных системах, с учетом сложности структуры и распределенности компонентов, спонтанного характера процессов, происходящих в системе, неопределенности и противоречивости знаний, что приводит к необходимости расширения классической мультиагентной архитектуры и онтологических моделей знаний и использование современных интеллектуальных информационных технологий для обеспечения гибкости, динамичности и эффективной работы системы в гетерогенной информационной среде.

Представление онтологических знаний на основе многозначного исчисления присутствия

В качестве основы для представления онтологических знаний в работе предлагается использовать интенциональное многозначное исчисление присутствия (МИП). Исчисление присутствия удовлетворяет требованиям, предъявляемым к логике процессов принятия решений и управления в распределенных интеллектуальных системах, предоставляет широкие возможности для описания связей по присутствию между объектами, их свойствами и атрибутами в предметной области, а также позволяет

учитывать неопределенность и противоречивость знаний, в отличие от используемых в настоящее время подходов.

Логические модели представления знаний, основанные на МИП, обладают рядом достоинств, основными из которых являются следующие:

- все истины в логическом исчислении присутствия рассматриваются как относительные, а не абсолютные, как в классической логике предикатов. Это связано с тем, что в физической реальности истины зависят от положения объектов в пространстве-времени, сложившихся условий и др.;

- истины в логике присутствия рассматриваются в динамике их формирования и развития, в то время как в классической логике предикатов они являются статическими, неизменными;

- истины, касающиеся текущего состояния реальности, имеют однозначную прагматическую интерпретацию, т.е. можно указать, какие действия должна предпринять интеллектуальная система в той или иной ситуации, что особенно важно для мультиагентных систем.

Чтобы иметь возможность производить операции и учитывать эволюционные изменения в онтологиях рассматривается математический аппарат теории категории, который выступает в роли универсального формализма онтологического описания и является независимым от выбора конкретного языка описания онтологий и специфики реализации. Логико-категорный подход является при этом альтернативой известной теории множеств:

- представление с единых позиций в виде объектов категорий классов абстрактных и конкретных понятий, свойств, состояний и процессов и т.д.;

- выражение морфизмами категории любых связей между объектами (вложения одних объектов в другие, а также функциональные (ролевые) отображения);

- категория является основой представления динамики функционирования системы с добавлением специальных операторов изменения значений модуса присутствия (дифференциалов изменения присутствия).

Особенность логики присутствия заключается в том, что формулы, обозначающие объекты (понятия, свойства, процессы и др.), могут принимать либо значение присутствует – Pr, либо значение отсутствует – Ab. Такая интерпретация заключается в указании свойств, которыми должен обладать экземпляр объекта. На множестве свойств объекта x задана булева алгебра присутствия:

$$\{X, \cap, \cup, \Rightarrow, \sim\},$$

где \cap – одновременное присутствие; \cup – присутствие хотя бы одного объекта.

Любой произвольный объект представляется в виде формулы на языке булевой алгебры присутствия:

$$x \equiv \bigcup_{i=1}^N \bigcap_{m=1}^{n_i} \varphi_{ij},$$

где φ_{ij} – некоторое свойство или его дополнение по присутствию; N, n_i – соответственно, количество наборов признаков и число признаков на каждом наборе. Эти свойства используются для распознавания объекта x . Это позволяет, выбрав достаточно полный набор свойств или признаков, образовывать новые объекты из простых. В общем случае не все признаки являются доступными, поэтому их можно разделить на два класса: доступные для системы свойства π_{ij} и недоступные ξ_{ir} .

В условиях неопределенности относительно значений присутствия части свойств на множестве формул, описывающих объекты, может быть задана топология, что позволяет перейти от булевой алгебры формул к топологической булевой алгебре:

$$\{x, \cap, \cup, \Rightarrow, \sim, I, C\},$$

где внутренность I и замыкание C обладает всеми свойствами аналогичной операции в топологии множеств. Данная алгебра лежит в основе четырехзначной или многозначной логики присутствия, в которой любому объекту x ставится в соответствие вектор вида

$$\vec{x} = \langle Ix, C \sim x, Cx, I \sim x \rangle,$$

который характеризует модус присутствия и имеет четыре значения: Pr - присутствие, Ab - отсутствие, Un - неопределенность, Cn - противоречие.

Для построения глобальных формализованных аксиоматических теорий (ФАТ) для описания сложных динамических объектов предметной области, можно использовать многозначную логику присутствия, дополненную рядом операций и правил построения формул, используемых в алгебраической топологии, теории категорий и топосов.

Определение категории K включает: класс $Ob(K)$ – множество объектов категории; набор множеств $Hom(x, y)$, по одному для каждой пары объектов $x, y \in Ob(K)$, элементы которых называются морфизмам и обозначаются $\mu: x \rightarrow y$, где стрелка означает отображение объекта x в объект y . В этом случае каждое свойство или понятие рассматривается в качестве объекта некоторой категории. В аксиоматической теории объекта, определяющие его свойства x_i замещаются их морфизмами, входящими в данный объект. Эти морфизмы образуют покрывающее семейство морфизмов для объекта y вида:

$$\Phi(y) = \{\mu_i: x_i \rightarrow y / i \in I\},$$

где x_i – объекты, соответствующие определяющим понятиям (свойствам) объекта y . Это $\Phi(y)$ покрытие

объекта y называется решетом над y , если любая композиция морфизмов $\mu_i \rho_i \in \Phi(y)$, где $\rho_i : \pi_{ij} \rightarrow x_i$, то есть объект x_i , в свою очередь, имеет существенные свойства π_{ij} .

Проделав указанные операции для всех объектов категории K , получим семейство решет $C(y)$, образованных композициями морфизмов различной кратности и называемых покрывающими решетками топологии, которые задают топологию Гротендика на объектах категории K . Категория вместе с заданной на ней топологией Гротендика $C(y)$ называется сайтом. В связи с тем, что экземпляры объекта x_i могут иметь различные значения модуса присутствия (Pr, Ab, Un, Cn) , то морфизм $\mu : x_i \rightarrow y$ можно представить в виде четверки взаимоисключающих по присутствию морфизмов:

$$\mu_i : \begin{cases} Ix_i \rightarrow y = Pr \mu_i : x_i \rightarrow y; \\ I \sim x_i \rightarrow y = Ab \mu_i : x_i \rightarrow y; \\ (Cx_i \cap C \sim x_i) \rightarrow y = Un \mu_i : x_i \rightarrow y; \\ (Ix_i \cap I \sim x_i) \rightarrow y = Cn \mu_i : x_i \rightarrow y, \end{cases}$$

которые можно рассматривать как значения присутствия морфизма μ_i .

Каждому из четверки морфизмов $\mu_i : x_i \rightarrow y$ ставится в соответствие множество образующих (аксиом), содержащих соответствующее значение присутствия объекта x_i в качестве условия и задается отображение элементов покрытия x_i в теорию объекта y . Кроме того, каждой формуле теории объекта y соответствует множество значение присутствия объекта x_i , встречающееся в этой формуле в качестве следствия. Таким образом, задается ограничение теории объекта y на объект x_i . Таким образом, образуется пучок теорий над сайтом, называемый топосом Гротендика.

Особенностью такого топоса, дополненного логикой присутствия, заключается в том, что он позволяет строить иерархические многоуровневые теории, соответствующие иерархическим системам объектов предметной области, объединенных между собой. При этом можно осуществить вложение теории объекта подчиненного уровня в теорию данного объекта и использовать полученную теорию при поиске доказательства теорем.

Теорию динамического объекта можно представить путем введения дифференциалов присутствия $\Delta_{\alpha\beta} \mu x_i$. Дифференциал $\Delta_{\alpha\beta} x_i$ принимает значение присутствия – "присутствует", если значение присутствия μx_i изменяется с α на β , в результате шага логического вывода. Алгебра таких дифференциалов с операцией композиции морфизмов

$$\{X, \Delta_{\alpha\beta}, \cap, \cup, \Rightarrow, \sim, I, C, \circ\}$$

образует исчисление присутствия.

Рассмотрим пример описания модели представления знаний на МИП об объектах некоторой предметной области (рис. 1). Здесь Q – есть любое из значений (Pr, Ab, Un, Cn) .

$$\Phi AT_C : \Delta_{Pr} \mu_A \cap \Delta_{Pr} \mu_B \Rightarrow \Delta_{QR} \mu_C;$$

$$\Delta_{Pr} \mu_{A_1} \cup \Delta_{Pr} \mu_{A_2} \cup \Delta_{Pr} \mu_{A_3} \cup \Delta_{Pr} \mu_{A_4} \Rightarrow \Delta_{QR} \mu_A;$$

$$\Delta_{Pr} \mu_{B_1} \cap \Delta_{Pr} \mu_{B_2} \cap \Delta_{Pr} \mu_{B_3} \cap \Delta_{Pr} \mu_{B_4} \Rightarrow \Delta_{QR} \mu_B;$$

$$\Delta_{Pr} \mu_{B_3} \cup \Delta_{Pr} \mu_{B_4} \Rightarrow \Delta_{QR} \mu_{B_5};$$

а также формулы вложенных концептов

$$\bigcap_i \bigcup_j \Delta_{Pr} \mu_{\pi_{ij}} \Rightarrow \Delta_{QR} \mu_{x_k},$$

где

$$\mu_{\pi_{ij}} : \pi_{ij} \rightarrow x_k$$

и $\{x_k\} = \{A_1, A_2, A_3, A_4, B_1, B_2, B_3, B_4\}$.

Таким образом, исчисление присутствия и приемы построения ФАТ составляют теоретическую основу представления онтологических знаний.

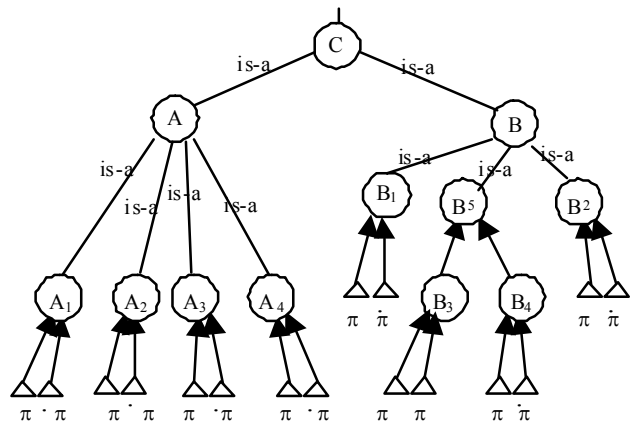


Рис. 1. Фрагмент онтологии для распознавания присутствия объекта C

Прагматические знания должны обеспечивать построение структур целевых установок и синтез (реализацию) программ их достижения при работе мультиагентной системы.

Таким образом, для уровня прагматики формируется набор программных модулей, которые будут использованы агентами в той или иной ситуации. Программный модуль можно представить как отображение (преобразование)

$$\mu PM_i : f_i^{BX}(X_1^{kl_1}, \dots, X_n^{kl_j}) \rightarrow f_i^{BYX}(Y_1^{kl_1}, \dots, Y_m^{kl_j}),$$

где f_i^{BX} - совокупности входных данных, соответствующих классов kl_1, kl_j в f_i^{BYX} - совокупность выходных данных, соответствующих классов.

Итак, в процессе объединения онтологических моделей с использованием МИП и логико-категорного подхода возможно возникновение противоречивых и неопределенных аксиом. При объе-

динении онтологий эти все аксиомы помечаются специальными тегами Un и Cn. Un – недостаточно информации о свойствах для определения присутствия объекта. Cn – может служить признаком несоответствия описания одного и того же концепта в различных моделях. При объединении онтологий также могут возникать ситуации: класс не может быть противоположен сам себе, два класса не могут быть одновременно противоположны и эквивалентны, не все экземпляры найдены, не все связи между экземплярами подтверждены и др. Каждый класс объектов предметной области описывается некоторым набором аксиом, при этом OWL позволяет указать аксиомы класса, которые определяют необходимые и достаточные условия для определения принадлежности классу. В случае, если аксиомы, описывающие эти условия удовлетворяются частично, возможно рассмотрение ситуации неопределенности. Т.е. некоторого атрибута для класса (объекта), который указывает на неопределенную ситуацию. При сохранении новой информации в онтологии агенты получают разнородные данные из различных источников. В случае если эти данные распознаются как неопределенные и/или противоречивые, они также помечаются тегами Un и Cn.

Для получения подобного представления используется специальный конвертер из OWL. При конвертировании онтологий производится поиск отношений между онтологическими понятиями разных онтологических контекстов, выясняется, какие понятия другой онтологии могут быть эквивалентны, какие являются суперпонятиями или подпонятиями онтологии, которая уже загружена в базу знаний интеллектуальной системы.

Заключение

Таким образом, была предложена модель представления онтологических знаний на основе многозначного исчисления присутствия, которая в отличие

от существующих, обеспечивает учет неопределенностей в онтологиях и работу с противоречивыми знаниями, что позволяет повысить эффективность решения задачи объединения онтологических моделей. Исчисление присутствия удовлетворяет требованиям, предъявляемым к логике процессов принятия решений и управления в распределенных интеллектуальных системах, предоставляет широкие возможности для описания связей по присутствию между объектами, их свойствами и атрибутами в предметной области, а также позволяет учитывать неопределенность и противоречивость знаний.

Список литературы

1. МакГиннес Дебора Л. OWL, язык веб-онтологий. Руководство [Электронный ресурс] / Дебора Л. МакГиннес, Фрэнк ван Хармелен. – Режим доступа: <https://www.w3.org/TR/2004/REC-owl-features-20040210/>.
2. Biacino L. Fuzzy logic, continuity and effectiveness [Text] / L. Biacino, G. Gerla // *Archive for Mathematical Logic*. – 2002. – № 7. – P. 643–667.
3. Principle of explosion [Электронный ресурс] – Режим доступа: http://en.wikipedia.org/wiki/Principle_of_explosion - Назв. с титул. экрана.
4. Defeasible reasoning [Электронный ресурс] – Режим доступа: http://en.wikipedia.org/wiki/Defeasible_reasoning - Назв. с титул. экрана.
5. Paraconsistent logic [Электронный ресурс] – Режим доступа : http://en.wikipedia.org/wiki/Paraconsistent_logic - Назв. с титул. экрана.
6. A free and open source Java framework for building Semantic Web and Linked Data application [Электронный ресурс] – Режим доступа: <http://jena.apache.org/#ch-Operations%20on%20Models> – Назв. с титул. экрана.
7. Прохоров А. В. Логико-категорный подход в онтологическом моделировании распределенных интеллектуальных систем [Текст] / А.В. Прохоров, Е.Н. Владимирская // *Радиоелектронні комп'ютерні системи*. – 2010. – № 4 (45). – С. 91-99.

Надійшла до редколегії 30.07.2015

Рецензент: д-р техн. наук, проф. Є.А. Дружинін, Національний аерокосмічний університет ім. М.С. Жуковського «ХАІ», Харків.

МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ОНТОЛОГИЧЕСКИХ ЗНАНИЙ С УРАХОВАНИЕМ НЕВИЗНАЧЕНОСТИ ТА ПРОТИРІЧ

О.В. Прохоров

Проведено аналіз особливостей подання в онтологіях невизначеностей і роботи з протиріччями. Розглядається математичний апарат теорії категорій, який виступає в ролі універсального формалізму онтологічного опису і є незалежним від вибору конкретної мови опису онтологій і специфіки реалізації. Для обробки невизначеностей і суперечностей, що виникають в процесі об'єднання онтологій і функціонуванні розподіленої інтелектуальної системи використовується багатозначне числення присутності. Для інтеграції з існуючими в системі знаннями пропонується конвертація онтології OWL в модель на багатозначному численні присутності, для якого реалізований логічний висновок.

Ключові слова: онтологія, OWL, розподілена інтелектуальна система, теорія категорій, невизначеності, протиріччя.

AN ONTOLOGY REPRESENTATION MODEL WITH ACCOUNT UNCERTAINTIES AND CONTRADICTIONS

A.V. Prokhorov

The analysis of the characteristics of representation uncertainties in ontologies and works with contradictions describe in this article. The category theory is a universal ontological formalism which is independent of the ontology language choice and the implementation. For the working with uncertainties and contradictions in the process of merging ontologies and distributed intelligent system functioning used multi-valued presence calculation. For integration with system knowledge based proposed the OWL ontology converter to model multi-valued presence calculation, for which implemented a logical reasoning.

Keywords: ontology, OWL, distributed intelligent system, category theory, uncertainty and contradictions.