

РАЗРАБОТКА, ИСЛЕДОВАНИЕ И ПРЕДСТАВЛЕНИЕ ФУНКЦИЙ И ОПЕРАЦИЙ НА ОНТОЛОГИЯХ

Александр Палагин, Сергей Кривый, Николай Петренко

Аннотация: Определяется множество функций и операций на онтологиях в целях эффективного представления, анализа, преобразования и построения онтологий.

Ключевые слова: онтология, функции, операции, преобразование, построение.

ACM Classification Keywords: I.2 Artificial Intelligence – I.2.4 Knowledge Representation Formalisms and Methods.

Введение

Онтологический подход к выполнению научных исследований в настоящее время активно апробируется многими исследователями.

Научное исследование (НИ) – процесс выработки новых знаний, один из видов познавательной деятельности человека, характеризующийся объективностью, воспроизводимостью, доказательностью и точностью.

НИ имеют два уровня – *теоретический* и *эмпирический*, и делятся на *фундаментальные* и *прикладные*, *количественные* и *качественные*, *уникальные* и *комплексные*.

Исходя из этого общего определения, НИ – это некоторая последовательность необходимых действий, которые приводят к новому знанию или к пониманию невозможности его получения. Эта последовательность действий, как правило, носит недетерминированный характер, поскольку на каждом шаге выполнения того или иного действия требуется проверка нескольких гипотез или альтернатив. В связи с этим НИ имеет свою логику и систему методологических принципов, составляющих этапы жизненного цикла НИ.

Логика НИ должна четко (по возможности) определить:

- объект и предмет НИ;
- применяемые методы;
- последовательность применения методов;
- характер НИ (теоретический или эмпирический);

- имеющиеся факты, подтверждающие идею проведения НИ;
- имеющиеся средства для проведения НИ;
- предполагаемые результаты НИ.

Исходя из логики НИ *теоретический уровень* проведения НИ требует накопления необходимых знаний и связан с глубоким анализом фактов, с познанием имеющихся явлений и их толкованием. На этом этапе осуществляют прогнозирование возможных действий, событий, изменений в изучаемых явлениях с выработкой принципов и методов их практической реализации.

Эмпирический уровень, как правило, состоит из двух стадий:

- процесса получения (наблюдения) и фиксации фактов;
- первичной обработки фактов (их обобщения) и выявления их взаимосвязи.

На этих стадиях НИ также выполняется:

- критическая оценка и проверка каждого факта (на допустимом уровне абстракции);
- описание каждого факта в терминах языка НИ;
- отбор (классификация) фактов, характерных для данного НИ.

Характерной особенностью НИ последнего периода времени является их *трансдисциплинарный характер*. Примером трансдисциплинарности такого типа является информатика, которая тесно связана с математикой (в теоретическом смысле), с электроникой, с телекоммуникацией и другими областями научных знаний, а информационные технологии стали составляющими практически всех *Hi-Tech*. При этом очевидно, что одному специалисту или даже группе специалистов не под силу обладать таким широким спектром знаний из смежных областей и возникает проблема: *каким образом помочь исследователю решать возникающие перед ним задачи, выходящие за рамки его прямой компетенции?*

Сегодня в мире накоплено огромное количество информации по различным областям человеческой деятельности как теоретического, так и практического характера. Часто для решения задачи необходимо иметь быстрый доступ к такой информации, в частности, к той или иной программе или алгоритму либо к пакету прикладных программ для проверки возникшей гипотезы или альтернативы. Для того чтобы это осуществить, информация должна быть структурирована, классифицирована и специфицирована по разделам. Именно эти функции в настоящее время возлагаются на онтологические и компьютерные системы.

В данной работе затрагивается только небольшая часть вопросов организации таких систем, касающихся операций и функций на онтологиях.

Основная часть. Необходимые понятия

Когда речь идет об онтологиях, то предполагается, что их обработка выполняется в компьютере, а сами онтологии представляются в виде нагруженного ориентированного графа (который назовем онтографом) в памяти компьютера. Как правило, в таком онтографе нагружены вершины и, в зависимости от вида информации (декларативного или процедурного характера), нагружаемой вершины, зависит множество функций и операций на онтологиях. Далее в работе предлагается набор функций и операций, не зависящих от содержимого вершин, и предполагается, что в компьютере может быть представлена не одна, а совокупность онтологий, за содержимым которых необходимо следить и поддерживать их в рабочем состоянии. Исходя из этих предположений и рассматриваются ниже множества функций и операций на онтологиях, которые разделены на несколько групп:

- функции и операции на онтографах;
- функции модификации онтографа;
- функции модификации содержимого вершин онтографа;
- функции извлечения необходимой информации из вершин онтографа;
- функции сравнения онтографов или их подграфов;
- функции административного характера.

Ниже описываются функции и операции на онтологиях с учетом их представления и особенностей структур данных, с помощью которых представлены онтологии.

Формальное определение онтологии в виде упорядоченной четверки над некоторой областью D имеет вид:

$$O = (X, R, F, A),$$

где X – множество понятий (концептов);

- R – множество отношений на концептах, т. е. на множестве X ;
- F – функция (или функции) интерпретации концептов и отношений на области D ;
- A – множество ограничений (аксиом), которые используются для представления истинных высказываний (определений), рассматриваемых над областью D [1]. Область D называется областью интерпретации и может иметь достаточно сложную структуру.

Использование онтологий для представления и обработки знаний с помощью компьютера требует представления онтологии в виде структур данных, которые удобны при работе с таким

объектом. Такой структурой является графовая структура, т. е. онтология представляется в виде ориентированного графа $G = (V, E)$ или, в частном случае, ориентированного дерева, вершины которого представляют либо онтологии низшего (подчиненного) уровня, либо атомарные (неделимые) элементы онтологии. Если исходить из такого представления, то отношения между вершинами изображаются ориентированными ребрами. При этом отношения между вершинами могут быть не только концептуальными (т. е. отношениями из множества R , но и другими отношениями (подчиненности, доминирования и т. п.), которые будем называть **метаотношениями**.

Таким образом, первая совокупность функций и операций на онтологиях связана с операциями на графах.

Графовые операции на онтологиях

Этого типа функции и операции хорошо известны и сводятся к следующим:

- поиск (обход графа) в онтографе (вершины или подграфа с данным именем);
- операции построения и преобразования графа онтологии;
- операции сравнения онтографов или их подграфов;
- функции копирования и удаления вершин, ребер, подграфов онтографа;
- функции переподчинения одной вершины другой вершине онтографа.

Поиск в онтографе обычно реализуется с помощью двух функций: поиск в глубину и поиск в ширину. Обе функции реализуются алгоритмом поиска в глубину (*DFS*-алгоритм) и алгоритмом поиска в ширину (*BFS*-алгоритм). Оба алгоритма имеют временную сложность $O(m + n)$, где $m = |E|$, $n = |V|$, т. е. достаточно эффективны и ориентированы на представление графов списками смежности. Такое представление и алгоритмы подробно описаны в литературе (см., например, [2, 3]).

Операции построения и преобразования графов сводятся к таким пяти операциям:

- нульарная операция “абсолютно пустой граф” $\Lambda = (0, 0)$;
- бинарные операции;
- введения вершины u в граф $G = (V, E)$, что дает орграф $G + u = (V \cup \{u\}, E)$, где $u \notin V$;
- введения ребра e в граф $G = (V, E)$, что дает орграф $G + e = (V, E \cup \{e\})$, где $e \notin E$;

- удаления вершины u в графе $G = (V, E)$, что дает орграф $G - u = (V \setminus \{u\}, E \setminus \{e_1, e_2, \dots, e_k\})$, где $u \in V$, а $e_1, e_2, \dots, e_k \in E$ – ребра, начальной вершиной которых является вершина u ;
- удаления ребра e из графа $G = (V, E)$, что дает орграф $G - e = (V, E \setminus \{e\})$, где $e \in E$.

Реализация операции Δ означает заготовку места для некоторого орграфа.

Известно, что перечисленные пять операций составляют полное множество операций, с помощью которых может быть реализована любая операция на конечных графах [4]. Отсюда следует, что относительно этих операций и носителя, каким является множество конечных графов, имеем алгебру онтографов:

$$G = (U, \Omega),$$

где U – множество конечных графов над некоторым универсальным множеством (из которого берутся вершины). Обычно в качестве такого множества выступает множество натуральных чисел N или его декартово произведение N^k (см. приведенный ниже пример).

Функции и операции на вершинах онтографа

Кроме функций и операций на графах необходимо иметь также и средства работы с содержимым вершин онтографов, поскольку каждой вершине онтографа соответствует некоторая информация (текстовая, графическая и т. п.). В связи с этим рассмотрим следующие функции и операции:

- а) функция копирования содержимого одной вершины онтографа в содержимое другой вершины;
- б) функция редактирования (модернизации, пополнения) содержимого вершины онтографа;
- в) функция удаления содержимого вершины онтографа.

Основными операциями выступают следующие теоретико-множественные операции:

- 1) операция слияния двух онтографов (объединение онтографов);
- 2) операция выделения общего подграфа двух (и, следовательно, любого конечного числа) онтографов по именам и по содержимому вершин;
- 3) операция удаления общей части двух онтографов из одного из них (по именам и по содержимому вершин);
- 4) операция поиска информации, содержащейся в вершинах онтографа;

- 5) операция семантического сравнения информации, содержащейся в вершинах онтографа, и ее (тождественные) преобразования;
- 6) операция копирования содержимого вершины целиком или выделенной части;
- 7) функции и операции, связанные с доступом к программам и вызовом программ (это требует создания вычислительной среды для выполнения программы и получения результатов этого выполнения);
- 8) функции перехода из одной онтологии в другую по ссылке или по имени онтологии с сохранением связи перехода (необходимой для возврата).

Операции и функции административного характера

Эти функции и операции по отношению к компьютерной онтологии носят системный характер и предназначены для администратора онтологической базы знаний. К этим функциям и операциям относятся следующие:

- 1) функции создания запасных копий онтологий;
- 2) функции восстановления онтологии (онтологий) в результате аварийных ситуаций (внезапное отключение напряжения, атмосферные явления и т. п.);
- 3) функции перестройки структуры онтологии (очистка устаревших и утративших актуальность сведений в онтологии);
- 4) ведение журнала паролей, пользователей, статистики посещений онтологий, категории пользователей и т. п.

Пример представления онтографа и его наполнение

Рассмотрим пример представления онтографа, описывающего кластер одного из подразделов дискретной математики “Множества и отношения”. Этот онтограф, его уровни и связи (в том числе между листами) показаны ниже на рисунках 1,а – 1,г.

Предположим, что необходимо пополнить раздел “Виды отношений” новым видом бинарных отношений “Ассиметричное”. Тогда в онтографе осуществляется поиск вершин 2.2 “Виды отношений” 2.2.2 “Бинарное”. После этого добавляется к онтографу новая вершина 2.2.7 “Ассиметричное”, которая соединяется дугой с вершиной 2.2.2, и выполняется наполнение вершины информацией (определение ассиметричного отношения, его свойства, которые известны к данному моменту времени). Необходимость модификации содержимого вершин связана с тем, что с течением времени могут появиться новые свойства объекта, которому соответствует вершина.

Фрагмент онтографа “Множества и отношения”

Лист 1

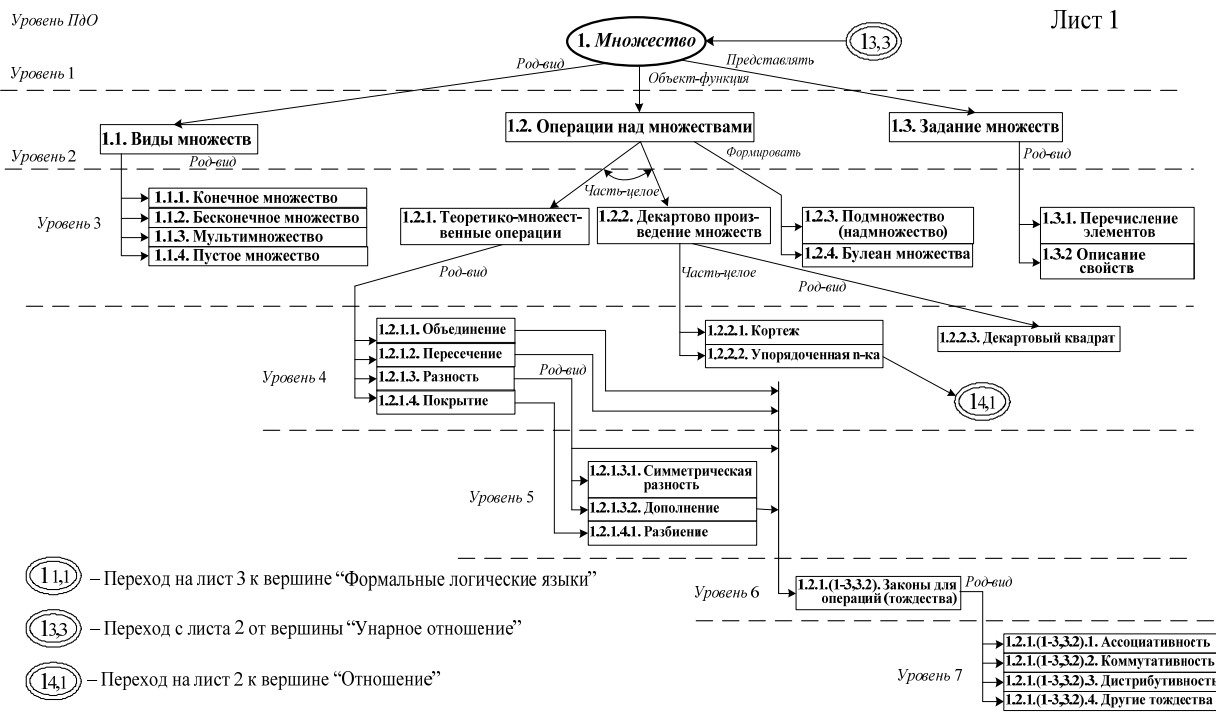


Рис.1, а. Онтограф кластера “Множества и отношения”

Лист 2

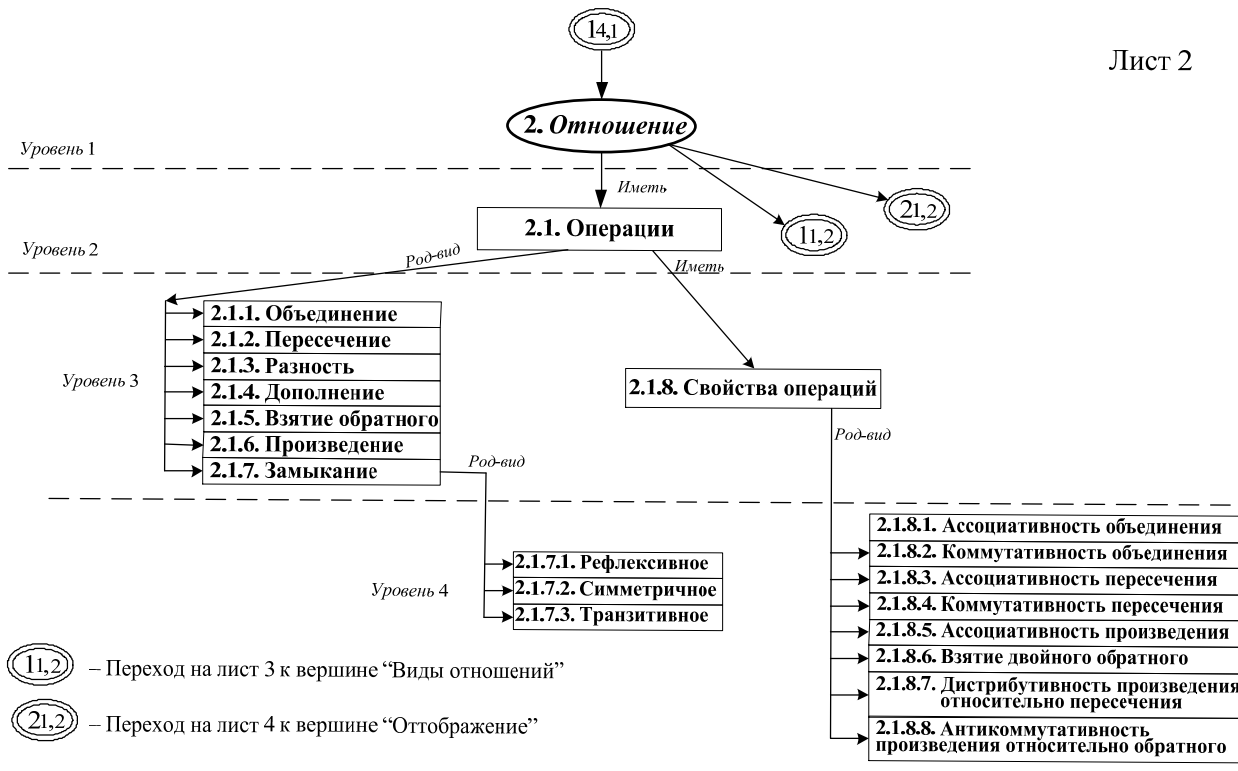


Рис.1, б. Онтограф кластера “Множества и отношения”

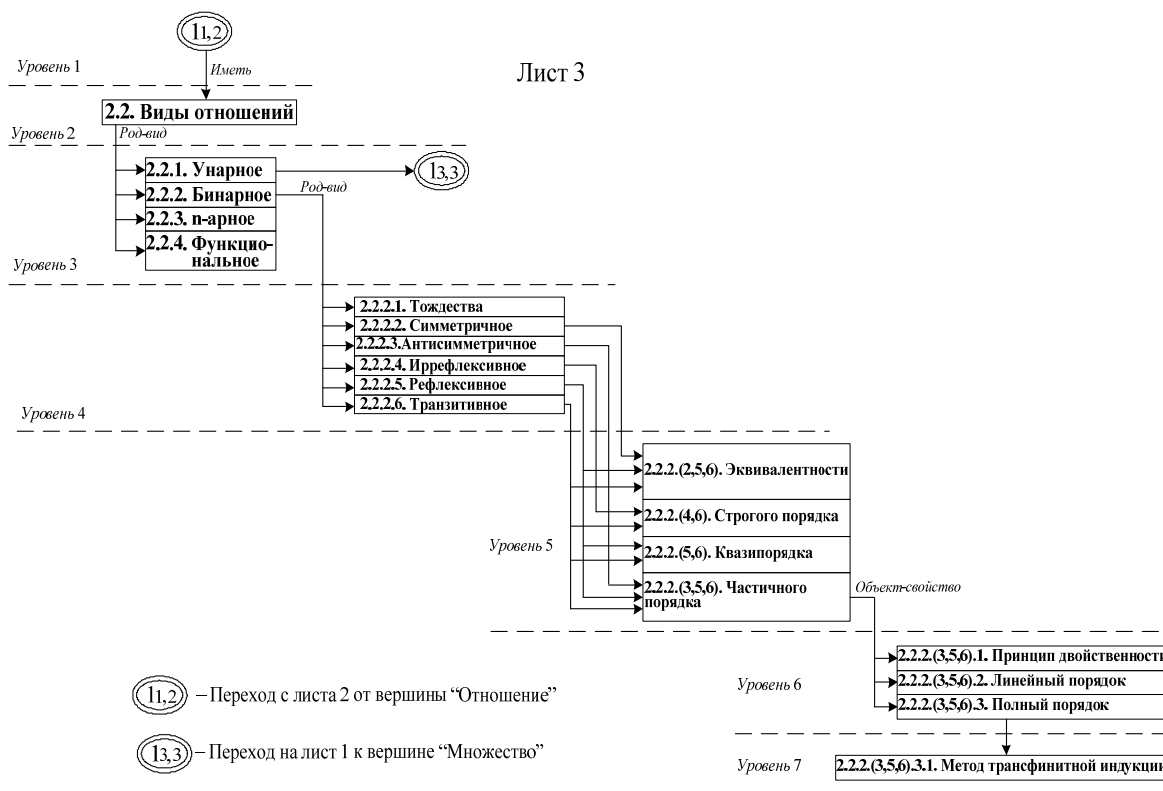


Рис.1, в. Онтограф кластера “Множества и отношения”

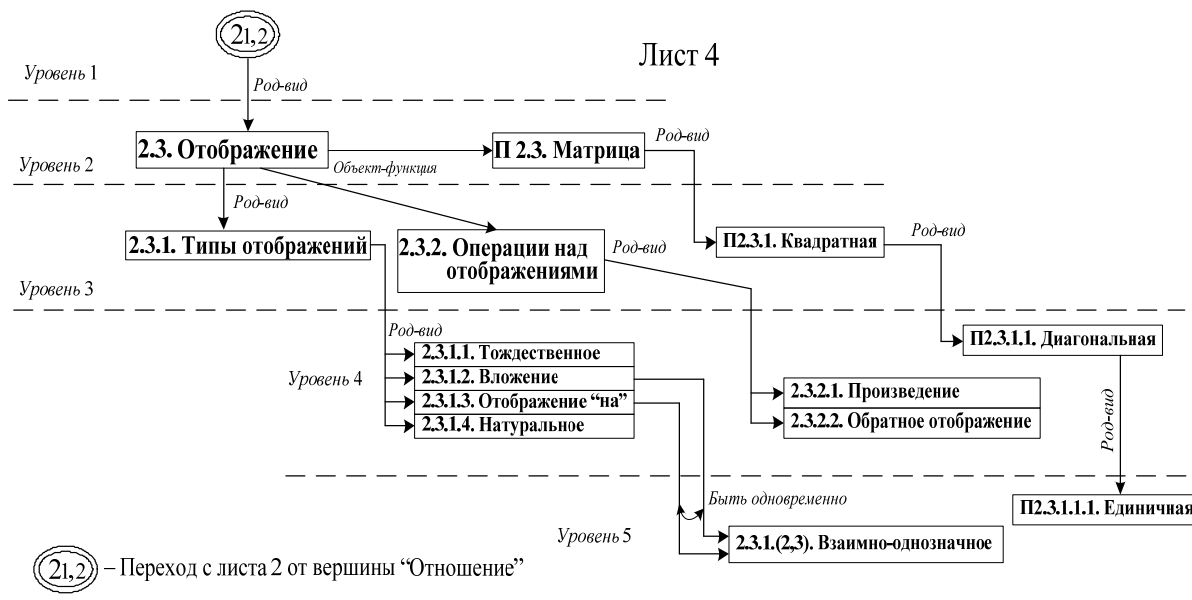


Рис.1, г. Онтограф кластера “Множества и отношения”

Содержимое вершин и выполнение операций

Для иллюстраций операций на онтографе рассмотрим задачу пополнения некоторого объекта в онтологии, представленной на рис. 1 онтографом “Множества и отношения”.

Рассмотрим некоторые способы описания (представления) содержимого вершин онтографа.

Первым из способов описания содержимого вершин онтографа является текстовое представление. Например, содержимым вершины 2.2.2.2 “Симметричное” есть текст следующего содержания: “Бинарное отношение R , заданное на множестве A ($R \subseteq A^2$), называется отношением, удовлетворяющее условию: для всех $a, b \in A$ из $(a, b) \in R$ следует $(b, a) \in R$.

Свойства $R : R = R^{-1}, R' = A^2 \setminus R$ тоже симметричное отношение”.

Вершина 2.2.2.2 “Симметричное” должна быть связана с вершинами, где определяется понятие обратного отношения (вершина 2.1.5) и дополнения (вершина 2.1.4). Кроме того, с понятием симметричного отношения связано понятие симметричного замыкания бинарного отношения и, поэтому, необходима связь с вершиной 2.1.7.2.

Другим возможным способом описания содержимого вершины может быть программно-алгоритмический способ. Например, если в процессе выполнения НИ возникнет необходимость вычисления транзитивного замыкания заданного бинарного отношения R на множестве A , то вершина 2.1.7.3 может содержать, кроме определений и свойства транзитивного замыкания (т. е. текстовой информации) ссылку на вершину 2.1.7.3. i, где расположены алгоритмы или программы вычисления этого замыкания с указанием языка программирования, ограничений на входные данные и другой необходимой информации.

Третьим возможным способом описания содержимого вершины может быть графическая информация (схемы, графики, фото-, рентгено снимки и другого типа информация), которая требует специальных способов обработки.

Из сказанного следует, что онтограф отображает только метаязык соответствующего раздела науки и его наполнение может быть достаточно сложным и аморфным. Каждая из вершин онтографа может быть, в свою очередь, начальной вершиной онтологии, описывающей соответствующий подраздел. При этом ранжирование по уровням вершин дает в некотором

смысле “строгую” классификацию информации, что необходимо для осуществления поиска и использования необходимой информации.

Операции логического вывода

Рассмотрим простой пример, связанный с проверкой правильности гипотез. Пусть в результате проведения НИ возникла необходимость проверки справедливости равенства $(R')^{-1} = (R^{-1})'$ для бинарного отношения R , заданного на множестве A , где R' – дополнение R в множестве A^2 , а R^{-1} – отношение, обратное к отношению R .

Для решения такого типа проверок необходимо иметь в качестве “подручного средства” систему логического вывода. Если считать, что такая система в нашем распоряжении имеется, то для ее работы необходимо найти всю требуемую информацию. В данном случае такая система должна “знать”, что это за объект “Бинарное отношение $R \subseteq A^2$, обратное отношение R^{-1} и его свойства, а также дополнение множества R' и его свойства”. Эти свойства необходимы для успешного решения поставленной задачи.

В данном случае требуемыми свойствами являются следующие свойства:

$$\text{а) } (a, b) \in R \Rightarrow (b, a) \in R^{-1};$$

$$\text{б) } (a, b) \notin R \Rightarrow (b, a) \notin R^{-1};$$

$$\text{в) } (a, b) \in R \Rightarrow (b, a) \notin R';$$

$$\text{г) } (a, b) \notin R \Rightarrow (b, a) \in R'.$$

Далее возможны, по крайней мере, два метода поиска решения задачи: аналитический метод, использующий тождества алгебры бинарных отношений (базирующийся на аксиоме экстенциональности теории множеств). Для второго метода можно применить свойства а) – г), приведенные выше.

В результате применения второго метода система (предположительно) сгенерирует следующий вывод:

$$(a, b) \in (R')^{-1} \Rightarrow (b, a) \in R' \Rightarrow (b, a) \notin R \Rightarrow (a, b) \notin (R)^{-1} \Rightarrow (a, b) \in (R^{-1})' \Rightarrow (R')^{-1} \subseteq (R^{-1})';$$

$$(a, b) \in (R^{-1})' \Rightarrow (a, b) \notin R^{-1} \Rightarrow (b, a) \notin R \Rightarrow (b, a) \in (R)' \Rightarrow (a, b) \in (R')^{-1} \Rightarrow (R^{-1})' \subseteq (R')^{-1}.$$

И выдаст вердикт: $(R')^{-1} = (R^{-1})'$ – правильное равенство.

Из этого простого примера следуют такие выводы.

1. Наполнение вершин онтографа должно быть как можно более “полным” в смысле описания свойств объекта, соответствующего данной вершине.
2. Пункт 1 предполагает выполнение операций пополнения онтологий и их модификации (в частности, при недостатке сведений об объекте).
3. Необходимо предусмотреть достаточный ресурс для хранения онтологий и выполнения задач вычислительного характера или работы систем логического вывода.
4. Необходимо иметь систему быстрого поиска (навигации по онтографам) необходимой информации (например, поиск по семантическим пространствам, ключевым словам, номерам или именам вершин онтографа).

Заключение

Из сказанного выше вытекает очень важный вопрос, связанный с реализацией упомянутых функций и операций на онтологиях: как должны быть организованы онтологии, чтобы эффективно можно было выполнять перечисленные функции и операции. Очевидно: первое, что требуется для успешной реализации функций и операций – это **введение некоторого стандарта на представление** онтографов и на представление информации в вершинах онтографов. Решение этой задачи, по-видимому, требует создания языка представления онтологий, который должен включать как элементы языка спецификаций, так и элементы языка программирования. Первое необходимо для спецификации объектов, с которыми нужно работать, а второе необходимо для того, чтобы можно было программировать логический вывод и выполнение программ (т.е. создавать среду вычисления). Такая стандартизация должна быть принята на международном уровне, что позволило бы навести определенный порядок в информационных системах, имеющих в глобальных сетях (например, в Интернете).

Литература

1. Палагин А.В., Крывый С.Л., Петренко Н.Г. Онтологические методы и средства обработки предметных знаний. – Луганск: Восточно-украинский нац. университет им. В. Даля. – 2012. – 324 с.
2. Кривий С.Л. Дискретна математика. Чернівці: Букрек. – 2014. – 567 с.
3. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. Второе издание. Изд дом “Вильямс”. – 2005. – 1296 с.

4. Сергієнко І.В., Кривий С.Л., Проватар О.І. Алгебраїчні аспекти інформаційних технологій. – Київ: Наукова думка. – 2011. – 400 с.

Информация об авторах



Александр Палагин – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: palagin_a@ukr.net

Область исследований: Общая теория знание-ориентированных информационных систем



Сергей Кривый – Киевский национальный университет им. Т. Шевченко, Украина Киев-187, ГСП, 03680, просп. акад. Глушкова, 40; email: krivoi@i.com.ua

Область исследований: Дискретная математика, теория автоматов, прикладная математическая логика, верификация программного обеспечения, программирование с ограничениями, онтологии.



Николай Петренко – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: petrng@ukr.net

Область исследований: Методология проектирования знание-ориентированных информационных систем с онтолого-управляемой архитектурой, инструментальных средств автоматизированного построения компьютерных онтологий предметных областей, толковых онтографических словарей

Development, Study and Presentation of Functions and Operations on Ontologies

Alexandr Palagin, Sergey Krivii, Nikolay Petrenko

Abstract: A set of functions and operations on ontologies for effective representation, analysis, transformation and construction of ontologies is defined.

Keywords: ontology, functions, operations, transformations, construction.