

УДК 004.048

С.Л. Крывый

## Формализованные онтологические модели в научных исследованиях

Рассмотрены проблемы, связанные с построением онтологий, тезаурусов и консенсуса в понимании интерпретации концептов предметной области, для которой строится такая онтология.

Розглянуто проблеми, пов'язані з побудовою онтологій, тезаурусів і консенсусу в розумінні інтерпретації концептів предметної області, для якої будується така онтологія.

**Введение.** Появление понятия *онтология* связывают с увеличением объемов информации и потребностями быстрой ориентации в ней. Огромные объемы разного рода информации невозможно использовать в полной мере без автоматизации процессов ее поиска и обработки. В частности, эта проблема возникает в процессе выполнения научных исследований (НИ), проведение которых немыслимо без широкого доступа к информации. Развитие научных исследований привело к тому, что они стали носить междисциплинарный характер (физхимия, биохимия, биофизика, биокибернетика, биомеханика и др.). Одни области научных знаний стали развиваться благодаря достижениям в других смежных областях, что особенно характерно для 20 ст.

Особенностью научных исследований конца 20 и начала 21 ст. является их трансдисциплинарность. Трансдисциплинарными стали не только исследования, проводимые на стыке многих наук, но и отдельно взятые науки стали трансдисциплинарными. Примером такой науки служит информатика, которая развивается благодаря достижениям математики, физики, электроники и других областей научных знаний, а информационные технологии стали составляющими практически всех *Hi-Tech*.

Очевидно, что при осуществлении НИ одному специалисту или даже группе специалистов не под силу овладеть таким широким спектром знаний из смежных областей одной и

той же науки, а тем более, разных наук. Часто исследователю нужны частные сведения, которые не требуют изучения некоторой области знаний. Например, если в процессе НИ возникает необходимость решения дифференциального уравнение определенного типа, то для этого нужна программа его решения, а не изучение теории дифференциальных уравнений. Другим частым примером в научных публикациях служит случай выяснения значения определенного обозначения или понятия, ссылка на которое дана, например, на двухтомник, каждый том которого содержит тысячу страниц. Естественно, в такой ситуации без автоматизации поиска нужной информации не обойтись. В связи с этим возникает проблема: *как помочь исследователю решать возникающие задачи?*

Для того чтобы осуществить такое содействие, информация должна быть структурирована. В настоящее время ведущей парадигмой структурирования информации являются *онтологии или иерархические концептуальные структуры*. Формирование таких структур в данное время возлагается на *аналитика* (инженера по знаниям).

### Онтологии

Если речь заходит об онтологиях, то формальное определение этого понятия дает искусственный интеллект (ИИ).

**Определение 1.** Онтологией предметной области  $D$  называется упорядоченная четверка  $O = (X, R, F, A)$ , где  $X$  – конечное множество,

элементы которого называются концептами (понятиями) из области  $D$ ,  $R \subseteq X^n$  – конечное множество, элементы которого называются отношениями между концептами,  $F$  – множество, элементы которого называются отношениями интерпретации элементов из  $X, R, A$  в области  $D$ , где  $A$  – множество аксиом (которое, в частности, может быть пустым).

Близким к понятию онтологии есть понятие *Тезауруса*. Чтобы понять общность и различие этих понятий, приведем определение Тезауруса, предлагаемое лингвистами [1, 2].

**Определение 2.** *Тезаурус* – это словарь, в котором слова и словоформы с близкими значениями сгруппированы в единицы, называемые *понятиями* (концептами, дескрипторами), где явно в виде отношений (отношений иерархии) указаны семантические отношения между этими понятиями.

Возникает вопрос: *какая разница между тезаурусом и онтологией?*

Действительно, из приведенных определений онтологии и тезауруса вытекает, что эти понятия почти совпадают. Это и не удивительно, так как до недавнего времени онтологии именовались тезаурусами.

Однако при внимательном рассмотрении этих понятий становится ясно, что Тезаурус описывает семантику на основе общих понятий, которая используется в качестве основы организации лексической информации, хотя при этом ничего не говорится о формализации самой семантики. А такая формализация необходима для обеспечения ее машинного применения в качестве моста между представлением знаний и естественно языковым текстом. Подобная организация Тезаурусов невозможна по ряду причин:

- семантика различных областей вряд ли имеет единый мета-каркас, который может служить основой организации любого узко ориентированного тезауруса;
- универсальные тезаурусы строятся на основе иерархической структуры, естественной для текстовых тезаурусов, но она не адекватна в реальных конкретных предметных областях.

Итак, Тезаурус представляет языковую организующую лингвистическую информацию, а Онтология – систему знаний об этой предметной области, т.е. ее формальную модель, использующую все доступные средства представления знаний, релевантных для данной области. Тезаурус представляет собой в определенном смысле онтологию лингвистических знаний о конкретной предметной области. Основную часть таких знаний составляет проблемно-ориентированная лексика, организованная с учетом морфологии, синтаксиса, словообразования, синонимии, омонимии и др. Отсюда следует, что основу интеллектуальной системы должна составлять пара «Тезаурус + Онтология».

Это разделение весьма условно, так как между Тезаурусом и Онтологией нет четко определенной границы, они слишком близки и взаимосвязаны и, возможно, в процессе эволюции со временем трансформируются в нечто новое.

Определение Онтологии подразумевает существование области интерпретации  $D$  множеств  $X, R, A$ , которой выступает данная предметная область  $D$ . В самом общем случае Онтология описывает иерархию концептов, связанных отношениями общего характера, а в более сложных случаях добавляются аксиомы, чтобы выразить более специфические отношения и ограничить их интерпретацию. Ограничение интерпретации позволяет сузить область возможных значений концептов, приближая отношение интерпретации к функциональному отношению. При этом к отношениям из множества  $R$  добавляются отношения из множества  $R_D$ , относящиеся к предметной области  $D$ . К отношениям из множества  $R$  относят отношения порядка (как правило, квазипорядка или частичного порядка) на множестве концептов и отношения подобия или, в лучшем случае, отношения эквивалентности. Для представления отношения порядка используется ориентированный граф, называемый *онтографом*, вершины которого формируются посредством отношения подобия. Таким образом, онтологическая модель предметной области представляется в виде ориентированного онтографа (на самом деле гиперграфа).

Отношения из множества  $R_D$  определяются той предметной областью  $D$ , на которой интерпретируются концепты. Из формального определения онтологии следует, что концептуализацией есть пара  $(D, R_D)$ , где  $D$  – область интерпретации данной предметной области (*домен*), а  $R_D$  – множество релевантных отношений конечной арности на  $D$ . Однако это определение не проясняет взаимосвязи между онтологией, возможными ее моделями и концептуализацией.

Один из способов представления содержания понятий состоит в том, чтобы определить их как функции от возможных миров и времени. Тогда пара  $(D, R_D)$  относится только к какому-то специфическому миру и это понятие необходимо расширить на множество миров.

Определим пару  $(D, W)$ , где  $D$  – домен, а  $W$  – множество возможных миров, как *доменное пространство*. В этом пространстве  $(D, W)$  определим отображение  $\rho: W \rightarrow B(D^n)$ , где  $B(D^n)$  – булеан множества  $D^n$ , как полностью определенную функцию из множества миров  $W$  в множество всех отношений конечной арности на  $D$ . Теперь концептуализация для области  $D$  определяется как упорядоченная тройка  $C = (D, W, R_D)$  где  $R_D$  – множество концептуальных отношений интерпретированных на доменном пространстве  $(D, W)$ . Для каждого возможного мира  $w \in W$  соответствующая ему структура согласно  $C$  имеет вид  $S_{wC} = (D, R_{wC})$ , где  $R_{wC} = \{\rho(w) \mid \rho \in R_D\}$  – множество расширений элементов из  $R_D$  относительно  $w$ .

Пусть имеется логический язык  $L$  в алфавите  $V$ , в котором записываются свойства предметной области. Определим модель языка как пару вида  $(S, I)$ , где  $S = (D, R_D)$  – структура мира, а  $I: V \rightarrow D \cup R_D$  – интерпретация элементов из  $V$  на заданных элементах из  $D$  и предикатов, соответствующих отношениям из множества  $R_D$ . Эта модель устанавливает *пространственную интерпретацию* логического языка  $L$ .

Исходя из сказанного, получается такая иерархия онтологий.

- **Онтологии верхнего уровня**, описывающие концепты общего характера, такие, как пространство, время, события и другие, не зависящие от отдельной предметной области или проблемы и отношений между этими концептами.

- **Онтологии предметных областей**, или онтологии заданий, описывающие актуальные задачи и результаты, накопленные в данной предметной области; эти онтологии используют терминологию онтологий верхнего уровня, пополняя ее своей терминологией, концептами и отношениями.

- **Онтологии прикладные**, описывающие конкретные задачи и методы их решения в конкретной предметной области и конкретные методы решения конкретных задач (программы, пакеты прикладных программ, описания технологий и пр.).

### **Знания и проблема совместности знаний**

Сложность построения Онтологии состоит в том, что два человека, смотрящие на один и тот же предмет, часто видят разные предметы. И поскольку построение Онтологии возлагается на инженера по знаниям, то построенная Онтология будет носить отпечаток индивидуальности ее создателя. Действительно, каждый отдельно взятый человек имеет свою интерпретацию увиденного, базирующуюся на его опыте восприятия мира и самого себя в этом мире. Эта ситуация – источник разного рода дискуссий, несогласия мнений, проявления вражды и др. Возникает вопрос: что же понимать под понятием *знание* и как определить достижение несколькими индивидами консенсуса в решении той или иной проблемы, что необходимо для построения адекватной модели?

Составной частью понятия *знание* является *суждение*, т.е. некоторое высказывание, где сформулирован какой-то фрагмент нашего понимания или представления о мире и объектах этого мира. Это высказывание может быть достаточно сложным, поскольку может включать в себя несколько предикатов с отрицанием и без такового. Посредством суждений выражаются как обычные предложения естественного языка, так и логические конструкции, такие как определения, толкование терминов, факты

из реальной жизни, выраженные посредством языка, математические теоремы, законы природы и пр. Каждый предикат суждения – это необходимый признак или условие существования объекта. Отсюда получаем, что суждение – это знания о субъекте ( $-ax$ ), выраженные через предикаты, которые могут быть ошибочными или вообще не иметь никакого отношения к реальности; основная проблема логического анализа суждений сводится не к выяснению истинности отдельных суждений, а к проверке их совместности [3].

Совокупность исходных суждений, подлежащих логическому анализу совместности, называется *рассуждением*. При таком анализе предполагается, что не все суждения истинны, а являются только некоторыми специфическими высказываниями, сомнительность которых может быть определена только в результате их совместного рассмотрения. Если такой анализ показывает *логическую некорректность*, то это служит основанием предположить, что наши аксиомы в совокупности некорректны. Таким образом, ограничение употребления терминов *истина* и *ложь* приводит нас к пониманию истины, которая *требует не декларирования знаний о мире, а обоснования этих знаний*.

Следовательно, **определение знания как совокупности предикатов, или отношений, представляющих суждения в рассуждениях**, не существенно сужает это понятие. Однако совокупность предикатов, описываемых средствами некоторого логического языка  $L$ , можно чисто формально проверить на противоречивость.

При этом возникают проблемы перевода с языка естественного на язык суждений, существования в естественной речи сведений, лежащих за пределами данного текста (т.е. знания вне контекста), и консенсуса понимания [4, 5].

*Первая проблема* связана с определенной языковой культурой, которая позволяет различать или отождествлять разнообразные субъекты и предикаты. Этой культурой пока не обладают современные компьютеры.

*Вторая проблема* является существенной преградой на пути автоматизации процесса из-

влечения знаний, если иметь в распоряжении только формальные правила преобразования наличного текста. Для этого необходимо помимо текста заложить в компьютер точную модель представленной в тексте ситуации. Получить такую модель на практике далеко не всегда возможно. В то же время для человека, когда речь идет о знакомых явлениях или событиях, восстановление недостающей информации – дело привычное, хотя и не всегда правильное.

*Третья проблема* связана с тем, что одни и те же факты разными людьми могут интерпретироваться не одинаково. Проблема консенсуса состоит в том, чтобы выработать общую (а лучше – единственную) интерпретацию полученных фактов.

**Знания и проблема совместности системы предикатов–отношений.** Основное понятие, необходимое для формулировки проблемы совместности системы предикатов–отношений, – понятие отношения, заданного на множестве  $D$ .

Пусть  $R_D$  означает множество всех отношений конечной арности, заданных на множестве  $D$ . Языком ограничений  $L_c$  на  $D$  называется некоторое непустое множество  $L_c \subseteq R_D$ .

**Определение 3.** Для произвольного множества концептов  $X$  онтологии  $(X, R, F, A)$  и любого языка ограничений  $L_c$  на  $D$  проблемой выполнимости ограничений  $CSP(L_c)$  ( $CSP$  – *constraint satisfaction problem*) есть решение такой комбинаторной задачи [6]:

Дано: тройка  $P = (X, D, C)$ , где

- $C$  – некоторое множество ограничений  $\{C_1, \dots, C_q\}$ ;
- каждое ограничение  $C_i \in C$  это пара  $(s_i, R_i)$ , где
  - $s_i = (x_{i1}, x_{i2}, \dots, x_{im})$ ,  $x_{ij} \in X, i = 1, 2, \dots, q$ , – кортеж концептов длины  $m$ ;
  - $R_i \in L_c$  –  $m$ -арное отношение на  $D$ , называемое отношением ограничения.

Выяснить: существует ли решение ограничения, т.е. существует ли такое отношение ин-

терпретации  $\varphi: X \rightarrow B(D)$ , где  $\varphi \in F$ ,  $B(D)$  – булеан множества  $D$ , что  $\forall (s_i, R_i) \in C$  кортеж  $(\varphi(x_{i1}), \dots, \varphi(x_{im})) \in R_i$  ?

Множество  $D$  в этом случае называется *областью проблемы*. Под понятием *знание* здесь понимаются отношения, извлеченные из текста или какого-нибудь другого источника информации об области  $D$  и составляющие множество отношений ограничения  $C$ . Тогда существование непустого отношения интерпретации  $\varphi$  означает *совместность знаний* (интерпретация  $\varphi$  – модель языка  $L_c$ ), а его отсутствие – *несовместность знаний*.

Если множество ограничений построено, то решение проблемы консенсуса состоит в том, чтобы выработать общее отношение интерпретации  $\varphi \in F$  (если оно существует) для данной совокупности людей. При решении задачи проверки совместности ограничений предполагается использование баз знаний, состоящее в том, чтобы найденное множество отношений ограничения с помощью базы знаний, снабженной процедурой логического вывода в подходящем логическом языке [10, 11], проверить на выполнимость (противоречивость) на области  $D$ .

Таким образом, проблема извлечения знаний из заданного источника информации, принадлежащего некоторой предметной области  $D$ , сводится к решению CSP, проблемы извлечения недостающей информации и проблемы консенсуса в понимании. *Консенсус в понимании* теперь можно сформулировать так. Пусть

$$\begin{aligned} \varphi_1(x_1) \subseteq D_1^1, \varphi_1(x_2) \subseteq D_2^1, \dots, \varphi_1(x_m) \subseteq D_m^1, \\ \varphi_2(x_1) \subseteq D_1^2, \varphi_2(x_2) \subseteq D_2^2, \dots, \varphi_2(x_m) \subseteq D_m^2, \\ \dots\dots\dots \\ \varphi_k(x_1) \subseteq D_1^k, \varphi_k(x_2) \subseteq D_2^k, \dots, \varphi_k(x_m) \subseteq D_m^k, \end{aligned}$$

– индивидуальные (субъективные) интерпретации  $\varphi_1, \varphi_2, \dots, \varphi_k$  концептов каждого из  $k$  участников решения данной CSP,  $D_1^i, D_2^i, \dots, D_m^i$  – подмножества из области  $D$  – множеств возможных значений концептов  $(x_1, \dots, x_m)$  соответственно. Построение отношения интерпре-

тации  $\varphi \in F$  выполняется исходя из интерпретаций  $\varphi_1, \varphi_2, \dots, \varphi_k$ :

$$\varphi(x_1) \subseteq \bigcap_{i=1}^k D_1^i, \varphi(x_2) \subseteq \bigcap_{i=1}^k D_2^i, \dots, \varphi(x_m) \subseteq \bigcap_{i=1}^k D_m^i.$$

Если  $\varphi(x_i) = \emptyset$  или  $(\varphi(x_1), \dots, \varphi(x_m)) \notin R_j$  хотя бы для одного  $j = 1, 2, \dots, q$ , то язык  $L$  считается противоречивым, а его система отношений  $C$  несовместной.

*Первый* путь преодоления противоречий в случае  $\varphi(x_i) = \emptyset$  состоит в расширении областей интерпретаций  $D_j^i$  с целью поиска возможного решения CSP( $L_c$ ), поскольку это может быть связано с проблемой извлечения недостающей информации.

*Второй* путь преодоления противоречий состоит в анализе причин их возникновения, т.е. выяснения того, какие отношения несовместны и почему. Это требует повторного анализа источника, из которого были получены отношения языка  $L_c$ , и анализа правильности определения этих отношений.

### Функции и операции на онтологиях

Рассмотрим совокупность функций и операций на онтологиях [7]. Поскольку онтологии представляются в виде онтографов, то первая совокупность функций и операций на онтологиях связана с операциями на графах.

**Графовые функции и операции на онтологиях.** Этого типа функции и операции известны и сводятся к следующим:

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

Поиск в онтографе обычно реализуется двумя функциями: поиском в глубину и поиском в ширину. Обе функции реализуются алгоритмами поиска в глубину (DFS-алгоритм) и в ширину (BFS-алгоритм). Оба алгоритма имеют вре-

менную сложность  $O(m + n)$ , где  $m = |E|$ ,  $n = |I|$ , т.е. достаточно эффективны и ориентированы на представление графов списками смежности. Эти алгоритмы и способы представления графов списками смежностей описаны в [8–9].

Построение и преобразование графов сводятся к таким пяти операциям:

- *нулевой операции*  $\Lambda = (\emptyset, \emptyset)$  – «абсолютно пустой граф»;

- *бинарным операциям* – введение вершины  $u$  в граф  $G = (V, E)$ :

результат – орграф  $G + u = (V \cup \{u\}, E)$ , где  $u \notin V$ ;

- введению ребра  $e$  в граф  $G = (V, E)$ :

результат – орграф  $G + e = (V, E \cup \{e\})$ , где  $e \notin E$ ;

- удалению вершины  $u$  в графе  $G = (V, E)$ :

результат – орграф  $G - u = (V \setminus \{u\}, E \setminus \{e_1, e_2, \dots, e_k\})$ , где  $u \in V$ , а  $e_1, e_2, \dots, e_k \in E$ , начальная вершина которых – вершина  $u$ ;

- удалению ребра  $e$  из графа  $G = (V, E)$ :

результат – орграф  $G - e = (V, E \setminus \{e\})$ , где  $e \in E$ .

Реализация операции  $\Lambda$  означает заготовку места для орграфа.

Известно, что перечисленные пять операций составляют *полное множество операций*, с помощью которых может быть реализована любая операция на конечных графах [8]. Отсюда следует, что относительно этих операций и носителя, каким есть множество конечных графов, имеем алгебру онтографов  $G = (U, \Omega)$ , где  $U$  – множество конечных графов над некоторым универсальным множеством (из которого берутся отметки вершин), а  $\Omega$  – сигнатура алгебры, состоящая из перечисленных пяти операций. Обычно в качестве множества  $U$  выступает множество натуральных чисел  $N$  или его декартово произведение.

Кроме приведенных функций и операций, изменяющих топологию онтографа, необходимо иметь также средства работы с содержимым вершин онтографов, поскольку каждой вершине

онтографа соответствует некоторая информация (текстовая, графическая и др.). В связи с этим рассмотрим следующие функции и операции:

- функцию копирования содержимого одной вершины онтографа в содержимое другой;

- функцию редактирования (модернизации, пополнения) содержимого вершины онтографа;

- функцию удаления содержимого вершины онтографа;

- функции, связанные с доступом к программам и вызовом подпрограмм (это предполагает наличие вычислительной среды для выполнения программ и получения результатов этого выполнения).

- функции перехода из одной онтологии в другую по ссылке или по имени онтологии с сохранением связи перехода (необходимой для возврата).

Основные операции таковы:

- слияние двух онтографов (объединение онтографов);

- выделение общего подграфа двух онтографов (и, следовательно, любого конечного числа) по именам и содержимому вершин;

- удаление общей части двух онтографов из одного из них (по именам и содержимому вершин);

- поиск информации, содержащейся в вершинах онтографа;

- семантическое сравнение информации, содержащейся в вершинах онтографа, и ее (тождественные) преобразования;

- копирование содержимого вершины в целом или выделенной части;

### **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 ним относятся:

- создание запасных копий онтологий;

- восстановление онтологии (онтологий) в результате аварийных ситуаций (внезапное отключение напряжения, атмосферные явления и пр.);

- перестройка структуры онтологии (очистка устаревших и утративших актуальность сведений в онтологии);

- ведение журнала паролей, пользователей, статистики посещений онтологий, категории пользователей и др.

### Пример представления онтографа и его наполнение

Рассмотрим пример представления онтографа, описывающего кластер одного из подразделов дискретной математики «Множества и отношения» [9]. Этот онтограф, его уровни и связи (в том числе между листами) показаны на рис. 1.

**Содержимое вершин и выполнение операций.** Для иллюстрации операций на онтографе рассмотрим задачу пополнения некоторого объекта в онтологии, представленной на рис. 1, (а–г) онтографом «Множества и отношения».

Предположим, что необходимо пополнить раздел *Виды отношений* новым видом бинарных отношений *Асимметричное*. Тогда в онтографе осуществляется поиск вершин 2.2 *Виды отношений* 2.2.2 *Бинарное*. После этого добавляется к онтографу новая вершина 2.2.2.7 *Асимметричное*, которая соединяется дугой с вершиной 2.2.2, и выполняется наполнение вершины информацией (определение асимметричного отношения, его свойства, которые известны к дан-

ному времени). Необходимость модификации содержимого вершин связана с тем, что с течением времени могут появиться новые свойства объекта, которому соответствует вершина.

Рассмотрим некоторые способы описания (представления) содержимого вершин онтографа.

*Первым* из способов описания содержимого вершин онтографа есть 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 *симметричным*, если оно удовлетворяет условию: для всех  $a, b \in A$  из  $(a, b) \in R$  следует  $(b, a) \in R$ .

Свойства  $R: R = R^{-1}$ ,  $R' = R^2 \setminus R$  – симметричные отношения, где  $R'$  означает дополнение  $R$  в множестве  $A^2$ ».

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.



Рис. 1, а

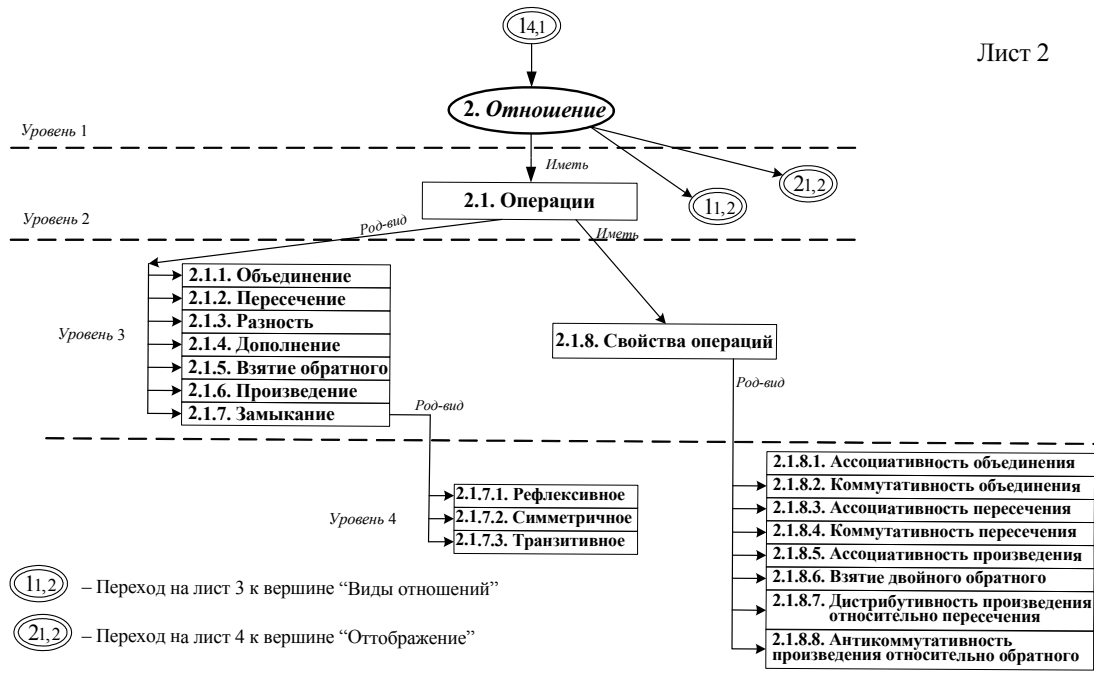


Рис. 1, б

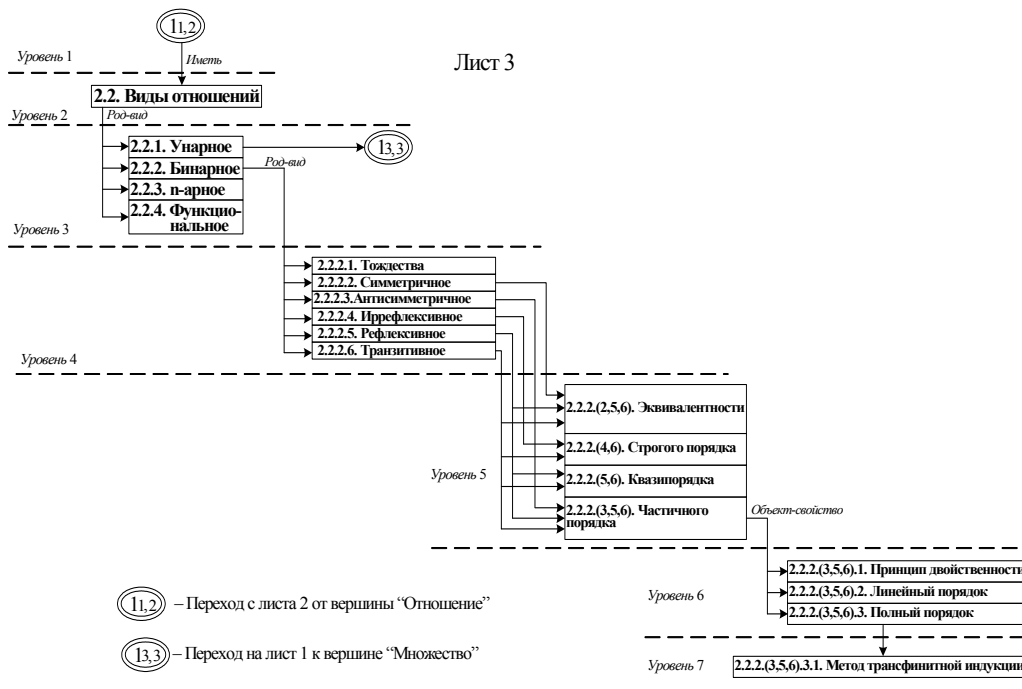
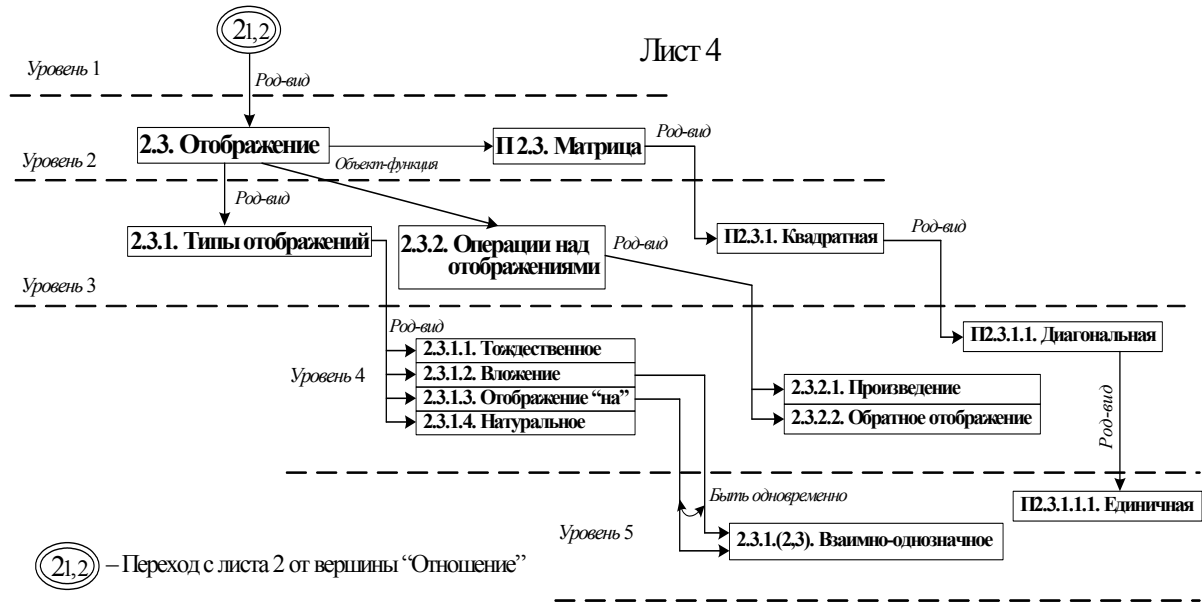


Рис. 1, в

Другим возможным способом описания содержимого вершины может быть программно-алгоритмический способ. Например, если в процессе выполнения НИ возникает необходимость вычисления транзитивного замыкания заданного бинарного отношения  $R$  на множе-

стве  $A$ , то вершина 2.1.7.3 может содержать, кроме определений и свойств 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 программирова-





(2,2) – Переход с листа 2 от вершины “Отношение”

Рис. 1, 2

ния, ограничений на входные данные и другой необходимой информации.

Третьим возможным способом описания содержимого вершины может быть графическая информация (схемы, графики, фото-, рентгено- снимки и другая информация), которая нуждается в специальных способах обработки. Из сказанного следует, что онтограф отображает только метауровень 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 эффективного поиска необходимой информации.

**Операции логического вывода.** Рассмотрим 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 принять, 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 требуемые свойства таковы:

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

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

$$в) (a, b) \in R \Leftrightarrow (b, a) \notin R';$$

$$г) (a, b) \notin R \Leftrightarrow (a, b) \in R'.$$

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

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

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

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

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

Из этого 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 необходимой информации (например, поиск по семантическим пространствам, ключевым словам, номерам или именам вершин онтографа).

Если речь идет об онтологиях и логическом выводе в формальных логических языках, то сегодня наиболее распространен язык *OWL*, [12], рекомендуемый консорциумом *W3 (World Wide Web)*. Этот язык имеет три диалекта: *OWL Lite*, *OWL DL*, *OWL Full*. Эти диалекты представления онтологий отличаются один от другого сложностью моделей, так как описываются разными логическими языками. Например, диалект *OWL DL* основывается на дескриптивной логике [13], основной принцип которой состоит в возможности автоматического доказательства непротиворечивости одного определения или отношения поглощения между разными определениями. Поглощение одного класса другим в дескриптивной логике означает, что для любых интерпретаций классов элементы поглощаемого класса есть также элементами и поглощающего класса. Это необходимо учитывать во многих задачах: проверки корректности онтологии, интеграции онтологий, обработки запросов в терминах он-

тологии. А диалект *OWL Full* позволяет определять неразрешимость задачи вывода.

В случае, когда автоматический вывод в онтологической модели невозможен ввиду ее сложности, доказательство возлагается на эксперта. Для этого служат системы интерактивного доказательства теорем, помогающие в доказательстве логических утверждений. Эти системы позволяют выполнять проверку корректности онтологий и их интеграцию, но плохо приспособлены к обработке запросов. Примером такой системы служит система *B-Toolkit* [14], основанная на логике первого порядка.

Еще один подход к решению такого типа задач базируется на *абстрактных типах данных (АТД)*. Если онтологические концепты выражены формальными спецификациями, то их можно рассматривать как АТД. В терминах АТД проверка совместности различных представлений о предметной области, описанных в терминах разных онтологий, сводится к проверке уточнения между спецификациями типов концептов. При этом спецификация *A* уточняет спецификацию *B*, если *A* можно использовать вместо *B*, не замечая факта подмены.

### Практическое использование онтологий

В последнее десятилетие онтологии нашли применение в мультиагентных технологиях в среде *семантической паутины (СП) – Semantic Web*. Последняя – это направление развития Всемирной паутины, основной целью которой есть представление информации в виде, удобном для машинной обработки на основе технологических стандартов (разработчик консорциум *W3*). СП предполагает запись информации в виде семантической сети с участием онтологий, что позволяет агентам непосредственно извлекать из паутины факты и генерировать логические следствия из этих фактов во взаимодействии с пользователем.

Разработка СП, как и онтологий, требует унификации представления данных и языка манипулирования этими данными. Эта унификация должна не только визуализировать данные, но и выполнять перенесение правил вывода в СП с целью превращения ее в систему семантического уровня.

Онтологический подход к разработке СП предусматривает развитие средств аннотирования документов, которыми могли бы воспользоваться *Web*-сервисы и специализированные программы-агенты в процессе обработки сложных запросов пользователей. В связи с этим консорциум *W3* решил, что для практического использования СП достаточно разработать универсальный язык представления знаний со ссылками на используемые онтологии (для чего служит *RDF* – язык низкого уровня представления метаданных), язык представления онтологий (*OWL* и *OWL2*), язык описания *Web*-сервисов (*WSDL*, *OWL-S*), язык запросов к базе знаний (*SPARQL*), семантические поисковые системы и программы-агенты.

Таким образом, СП представляет собой систему с элементами ИИ, а онтологии в такой системе служат вспомогательным средством концептуализации и унификации данных и доступа к ним. Для реализации доступа к данным, содержащимся в онтологиях, служит *RDF*-документ. В начале этого документа приведен список ссылок на объект из некоторой онтологии. Кроме того, узлы этого списка могут содержать дополнительную информацию, что позволяет строить на основе *RDF* языки более высокого уровня (*RDF Schema*, *OWL*). А это позволит создавать специализированные форматы для представления различных типов объектов. На пути решения этой проблемы консорциум *W3* разработал специализированный язык запросов к *RDF*-хранилищам – *SPARQL*, позволяющий осуществить сложные выборки из массивов метаданных.

На данном уровне развития СП основываются на трех базисных компонентах: программах-агентах, расширяемом языке разметки *XML* и *Web*-онтологиях.

Эти базисные принципы и компоненты использованы в различных отдельных проектах, среди которых выделяются два: *Dublin Core* (Дублинское ядро) и *BDpedia*. Первый из них реализован инициативной организацией *Dublin Core Metadata Initiative (DCMI)* и представляет открытый проект разработки метаданных, независимых от платформ и подходящих для использования в различных областях. В рамках

этого проекта осуществляется разработка словарей (тезаурусов) метаданных общего использования, стандартизирующих представление ресурсов в формате *RDF*.

Второй проект *BDpedia* реализует извлечение структурированной информации из проекта *Wikipedia*. При этом пользователь должен запрашивать информацию, базирующуюся на взаимосвязях и свойствах ресурсов *Wikipedia*, а также со ссылками на соответствующие базы данных. Этот проект использует *RDF* для представления извлеченной информации.

**Заключение.** Обозначенные проблемы, возникающие в процессе исследований с использованием онтологий, аккумулируют проблемы из области лингвистики, теории поиска доказательств, математической логики, баз знаний, алгебры, теории графов и психологии. Их решение встречает трудности, связанные прежде всего со сложностью таких объектов, как естественный язык и человеческая психология восприятия информации, заложенной в таком языке.

В перспективе онтологиям отводится специфическая роль в ИИ, компьютерной лингвистике, базах знаний, представлении знаний, качественном информационном моделировании, информационной интеграции знаний, поиске и извлечении информации, проектировании агентных систем и в создании онтолого-управляемых информационных систем.

1. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во. МГУ. – 2011. – 512 с.
2. Никитина С.Е. Семантический анализ языка науки. – М.: Наука, 1987. – 141 с.
3. Кулик Б.А. Логика естественных рассуждений. – С.-Пб.: Невский диалект, 2001. – 127 с.
4. Палагин А.В., Крытый С.Л., Петренко Н.Г. Знание-ориентированные информационные системы с обработкой естественно-языковых объектов: основы методологии и архитектурно-структурная организация. – УСиМ. – 2009. – № 3. – С. 42–55.
5. Апресян Ю.Д. Лингвистический процессор для сложных информационных систем. – М.: Наука, 1992. – 324 с.
6. Cohen D., Jeavons P. The Complexity of Constraint Languages. Handbook of Constraint Programming / Eds. by F. Rossi, P. van Beek, T. Walsh. – 2006. – P. 245–280.

7. Палагин А.В., Кривый С.Л., Петренко Н.Г. Разработка, исследование и представление функций и операций на онтологиях // Information theories and appl. – Boologaria. – ITHEA. – 2015. – 22, N 2. – P. 103–114.
8. Сергієнко І.В., Кривий С.Л., Проватар О.І. Алгебраїчні аспекти інформаційних технологій. – К: Наук. думка, 2011. – 399 с.
9. Кривий С.Л. Дискретна математика. – Чернівці: Букрек, 2014. – 567 с.
10. Клини С.К. Математическая логика. – М.: Мир, 1973. – 480 с.
11. Мендельсон Э. Введение в математическую логику. – М.: Наука, 1984. – 319 с.
12. Internet The Species of OWL. – <http://www.w3.org/TR/owlguide/#OwlVarieties>
13. The Description Logic Handbook. Theory, implementation and applications / F. Baader, D. Calvanese, D. McGuinness et al. – Cambridge Univ. Press, 2007. – 601 p.
14. B-Toolkit User's Manual – Release 3.4. – 1997. – Copyright B-Core (UK). – Ltd.

Поступила 18.02.2016

Тел. для справок: +38 044 521-3554 (Киев)

E-mail: [sl.krivoi@gmail.com](mailto:sl.krivoi@gmail.com)

© С.Л. Кривый, 2016

UDC 004.048

Кривий С.Л.

### Formal ontological models in scientific researchers

In the paper consider some problems connected to building ontologies, thesauruses and achieving of consensus in understanding of interpretation of concepts from ontology for a given domain.

In first part discusses of similarities and differences between ontologies and thesauruses. As a result of such discussion is the ontology and thesaurus must complements to each other

In second part of paper consider problems of knowledge consistency derived from natural language text or another sources of information. For knowledge presentation are used corresponding predicates or relations. The consistency problem of knowledge defines as constraint satisfaction problem over given domain D by using interpretation of concepts in this domain D. Consistence problem may be not solved over domain D, because some information is absences or set of predicates are inconsistency. This situations are discussed.

In third part of paper considered set of functions and operations over ontologies. For representation of hierarchy of ontology's concepts is used directed graph (more precisely hyper graph). These functions and operations are divided on set of functions and operations on graph and functions and operations of administrative character.

In four part of paper such functions and operations are demonstrated by examples.

In five part of paper is described same implementations of ontologies by using languages OWL, OWL2, RDF and so on.

In conclusion the future of applications of ontologies are described. In special case the main attention give a role of ontologies in transdisciplinary sciences researches.



### Внимание !

**Оформление подписки для желающих  
опубликовать статьи в нашем журнале обязательно.**

**В розничную продажу журнал не поступает.**

**Подписной индекс 71008**

### Внимание !

Авторы статей **обязательно** должны подать структурированную (*Introduction, Purpose, Methods, Results, Conclusion*) расширенную аннотацию на английском языке до одной страницы текста через два интервала, информацию об авторах на английском и, кроме пристатейного списка литературы (на языке статьи), список литературы в транслитерации (с указанием в скобках перевода на английский язык).