А.А. Кулинич

Концептуальные каркасы онтологий слабо структурированных предметных областей

Аннотация. Исследуются модели представления знаний о слабо структурированных предметных областях в виде концептуальных каркасов, построенных на основе знаний об одном из объектов этой предметной области. Рассмотрены вопросы экспертного построения онтологии предметной области на основе ее концептуального каркаса для случаев, когда признаки объекта принимают бинарные значения и значения из упорядоченного множества возможных значений.

Ключевые слова: онтология, концептуальный каркас, слабо структурированная предметная область, семантическое пространство, структуризация семантического пространства.

Введение

В системах поддержки принятия решений экспертные знания и опыт решения задач представляются в виде различных моделей представления знаний, в том числе и в виде онтологий, описывающих знания на разных уровнях и в разных аспектах. Обычно онтологии строятся экспертом, но если предметная область мало изучена, то и знания субъекта об этой предметной области носят фрагментарный и гипотетичный характер. В этом случае модель экспертных знаний – онтология – будет отражать фрагментарность и гипотетичность знаний субъекта о плохо определенной предметной области, и ее использование для решения практических задач в рамках системы поддержки принятия решений не даст желаемых результатов. В таких случаях эксперт должен пополнить свои знания о предметной области и только потом структурировать свои знания и представить их в виде онтологии. Таким образом, изучение сложных предметных областей – это длительных процесс, требующий значительного времени, которого часто при принятии решений у лица, принимающего решение, нет.

Один из возможных способов решения этой проблемы заключается в автоматическом построении онтологий на основе анализа текстов, описывающих предметную область, или автоматизированном построении онтологий на основе экспертно сформированном представительном множестве экземпляров предметной области.

В случае автоматического построения онтологии на первом этапе построения осуществляется подбор текстов, отражающих основные закономерности предметной области. На следующем этапе применяют один из подходов анализа текста:

1. На основе полного лингвистического анализа, предполагающего полный морфологический, синтаксический и семантический анализ текста. Результатом такого анализа является семантическая сеть концептов предметной области, связанных разного рода семантическими отношениями. Для построения онтологии выделяются отношения «класс-подкласс» («*Isa*»), которые образуют категориальную структуру онтологии предметной области.

2. На основе лексико-синтаксических шаблонов выделяются из текста различные отношения между понятиями предметной области. Например, выделению отношения «класс-подкласс» («*Isa*») посвящена работа [1]. В работе [2] разрабатывался язык для лексико-синтаксических шаблонов, которые применялись для анализа научно-технических документов.

3. Статистические методы исследовались в работе [3]. Суть этого подхода состоит в предварительной обработке корпуса текстов: определение всех слов и словосочетаний, определение их встречаемости в корпусе текстов. Далее осуществляется выделение отношений «класс-подкласс» на множестве всех слов и словосочетаний с использованием эвристик.

4. На основе продукций с применением генетического и автоматного программирования. Метод автоматического построения онтологий на основе продукционной модели исследовался в работе [4]. Здесь предлагается модель автоматического построения онтологий предметной области в виде системы продукций с применением генетического и автоматного программирования.

5. На основе анализа формальных понятий (FCA) [5,6]. Этот метод основан на лингвистическом анализе корпуса текстов с целью выявления в тексте экземпляров данной предметной области и представлении их в таблице формального контекста предметной области. В таблице формального контекста, в строках записаны названия экземпляров предметной области, а в столбцах названия их свойств и значения для каждого из объектов. Далее к формальному контексту применяются методы формального анализа понятий [7], позволяющие получить решетку понятий предметной области. Все понятия этой решетки связаны отношением «класс-подкласс».

Качество перечисленных подходов, опирающихся на автоматическую обработку текста, в значительной степени зависит от качества анализаторов текста. Известные теоретические трудности обработки естественного русского языка не позволяют говорить о высоком качестве построенных онтологий. Однако онтологии предметных областей, получаемые автоматическим способом, могут быть использованы в качестве основы для поддержки работы экспертов при построении онтологий предметных областей.

Интерес представляют работы по автоматизированному построению онтологий на основе экспертно - сформированного представительного множества объектов предметной области. Представительское множество объектов – это небольшое количество объектов, позволяющих построить онтологию всей предметной области. В работе [8] исследуются вопросы поддержки построения онтологии предметной области методом Bottom-Top. Здесь выделяется представительное множество объектов предметной области, описанных на языке дескриптивной логики (ДЛ). Для выделенного множества объектов строится формальный контекст, а затем - концептуальная решетка понятий (онтология) методами формального анализа понятий [7].

К сожалению, в работе [8] не исследованы вопросы формирования представительного множества объектов для построения качественной онтологии предметной области. Однако вопрос построения онтологий плохо определенных предметных областей на основе небольшого количества объектов предметной области актуален. Например, можно ли построить онтологию предметной области по одному ее объекту?

Рассмотрим одну из возможных теоретических моделей представления экспертных знаний о предметной области в условиях неопределенности [9]. Здесь все знания о предметной области представляются в виде множества сведений X. Любые сведения об объекте этой предметной области (информация о свойствах объекта, значениях этих свойств) называется элементарным сведением об объекте и определяется тройкой: $(p)\delta_i(x_0)\in X$, где $(p)\in[1,0]$ – степень истинности наличия свойства δ_i у объекта x_0. Если сведение истинно $(p=1)$, то ее опускают при описании объекта, т.е. $(1)\delta_i(x_0)=\delta_i(x_0)\in X$.

Информация о том, что у объекта x_0 нет свойства δ_i, $\neg\delta_i(x_0)$ также считается сведением об объекте. Если у объекта x_0 имеются два свойства $\delta_1(x_0)$ и $\delta_2(x_0)$, то конъюнкция и дизъюнкция этих свойств $\delta_1(x_0)\wedge\delta_2(x_0)$, $\delta_1(x_0)\vee\delta_2(x_0)$ также будут сведениями об объекте. Т.е., определены операции над элементарными сведениями $\delta_i(x_0)$ – $(\wedge, \vee, \neg)$, результаты которых также являются сведениями об объекте, $\delta_1(x_0)\wedge\delta_2(x_0)\in X$, $\delta_1(x_0)\vee\delta_2(x_0)\in X$. Утверждается, что все сведения об объекте образуют дистрибутивную решетку $(L(x_0), \wedge, \vee)$, которая называется решеткой понятий. Информацией об объекте считаются не просто элементарные сведения об объекте $\{\delta_i(x_0)\}$, но и все возможные их обобщения и сведения, полученные с помощью элементарных логических преобразований $(\wedge, \vee, \neg)$. Таким образом, объект реального мира, определенный множеством элементарных сведений $\{\delta_i(x_0)\}$ порождает структуру сведений – решетку понятий $(L(x_0), \wedge, \vee)$, которая несет информацию не только об объекте x_0, но и о предметной области, и структурной организации сведений об этой области.

В данной работе исследуется вопрос поддержки построения онтологии плохо определенной предметной области по одному ее объекту. Предлагается формальными методами строить концептуальный каркас онтологии предметной области, который затем применяется в экспертной процедуре построения ее онтологии.

1. Концептуальный каркас онтологии предметной области

Среди многих формальных определений онтологии предметной области выделим определение в виде кортежа:

$$\boldsymbol{O}^d=\langle C, A, R, D\rangle, \tag{1}$$

где C – множество классов предметной области, A– множество атрибутов классов, R – отношение частичного порядка на множестве классов, $R\subseteq C\times C$, D – множество доменов (экземпляры класса).

При таком определении отношения R в онтологии определен класс отношений «*Isa*» (отношение «класс-подкласс). Считается, что два класса c_i, $c_j\in C$ в онтологии $\boldsymbol{O}^d$ находятся в отношении - «*Isa*» если между атрибутами a_i, $a_j\in A$ и доменами d_i, $d_j\in D$ класса определены следующие зависимости: $a_i,\subset a_j$ & $d_i\supset d_j$. В этом случае класс c_i называется надклассом, а класс c_j – подклассом.

В онтологическом моделировании используют термины: класс (надкласс), атрибуты класса, экземпляр класса. Далее мы будем использовать синонимичную терминологию: понятие, обобщенное понятие – это класс, надкласс C; содержание понятия – это атрибуты класса A; объем понятия – это множество экземпляров класса или домен D.

В дескриптивной логике (ДЛ) [10], используемой в настоящее время для формального описания онтологий, на множестве экземпляров предметной области выделяется наиболее специфический объект – *msc (most specific concept),* имеющий более подробное описание свойств (большее число признаков), записанных на языке ДЛ, по сравнению с остальными объектами.

Допустим, что эксперт определил объект v^{msc} в некоторой предметной области, и будем считать, этот объект наиболее специфическим. Формально определим понятие (класс) этого специфического объекта тройкой: $\langle d,F(d),V(d)\rangle$, где d – имя понятия, $F(d)=\{f_j\}$- содержание понятия (множество признаков), $V(d)=\{v^{msc}\}$ – объем понятия, v^{msc} – наиболее специфический объект, имеющий признаки $F(d)$.

Пусть $B(F(d))=\{\varnothing, 2^{|F(d)|}\}$ – булеан содержания $F(d)$ (множества признаков) понятия d, где $2^{|F(d)|}$, множество всех подмножеств содержания $F(d)$. Известно, что элементы булеана образуют частично упорядоченное множество по включению его элементов, т.е. решетку $(B(F(d)), \wedge, \vee)$.

Как видим, элементы булеана образуют частично упорядоченное множество, так же как и классы онтологии в определении (1). Однако алгебраическую решетку, образованную элементами булеана можно считать прообразом онтологии предметной области, если сделать два следующих допущения:

1. Любой элемент $F(d^H)\in B(F(d))$, $H=1,\dots,|2^{|F(d)|}|$, полученной решетки формально будем считать содержанием понятия d^H обобщающего понятие d, если $F(d^H)\subseteq F(d)$. Это допущение означает, что в решетке $(B(F(d))\wedge, \vee)$ любое подмножество $F(d^H)\in 2^{|F(d)|}$ может быть интерпретировано как множество атрибутов $i-го$ класса A_i онтологии, т.е. $p(F(d^H)=A_i)$, где i–номер класса в онтологии, $p\in\{0,1\}$ - степень истинности того, что подмножество $F(d^H)$ содержания специфического понятия d в формальной алгебраической решетке $B(F(d))$ является множеством атрибутов класса онтологии предметной области.

2. Отношение включения содержаний элементов решетки $(F(d^H)\subseteq F(d))$ будем считать отношением «класс-подкласс» (*Isa*), при условии, что для объемов этих понятий выполняется достаточное условие $V(d)\subseteq V(d^H)$. Это допущение означает, что если определены (например, экспертным способом) элементы объема (экземпляры класса) удовлетворяющие условию $V(d)\subseteq V(d^H)$, то степень истинности p из первого допущения принимает значение «истина», т.е. $p=1$.

Определение 1. Решетку $K(d)=(B(F(d)), \wedge, \vee)$ всех подмножеств содержания начального понятия d будем называть концептуальным каркасом онтологии плохо определенной предметной области.

В концептуальном каркасе онтологии определены: множество признаков понятия $F(d^H)$ (атрибутов классов *A*); отношения частичного порядка *R* на множестве всех подмножеств содержания понятия *F(d),* но не определены; абстрактные имена понятий d^H (классов *C*) и их объемы $V(d^H)$ – экземпляры класса.

Концептуальный каркас считается необходимой структурой онтологии предметной области, к которой принадлежит понятие *d*. Это означает, что атрибуты $A=\{a_j\}$ классов *C* онтологии предметной области $\boldsymbol{O}^d$, построенной экспертом, будут принадлежать концептуальному каркасу *K(d)* объекта *d* этой предметной области, т.е. $a_j \in K(d),\ \forall j$.

Определение 2. Обобщенное понятие d^H в концептуальном каркасе онтологии *K(d)* будем называть реальным, если разности объемов обобщенного и необобщенного понятий непустое множество, $V(d^H)\backslash V(d) \neq \varnothing$, иначе это понятие будем называть виртуальным.

Определение 3. Онтологией предметной области $\boldsymbol{O}^d$ называется подмножество элементов концептуального каркаса *K(d),* в котором все содержания обобщенных понятий $F(d^H)$ реальны – $V(d^H)\backslash V(d)\neq\varnothing,\ \forall H$ и определены их имена d^H.

Пример концептуального каркаса и экспертно построенной онтологии предметной области геометрических фигур приводится в работе [11]. Здесь определен наиболее специфический объект – прямоугольный треугольник как плоская геометрическая фигура, ограниченная тремя сторонами, один угол прямой (Рис. 1). Его содержание определено двоичным вектором, включающим три единицы (1,1,1). Все собственные подмножества множества признаков этого понятия также обозначаем двоичным вектором, в котором отсутствующие признаки обозначаются нулем. Например, двоичный вектор (1,1,0) определяет содержание понятия с признаками: плоская геометрическая фигура и ограничена тремя сторонами – это треугольник.

Онтология геометрических фигур, построенная на основе концептуального каркаса с использованием экспертных процедур приведена на Рис.2. Суть экспертной процедуры заключается в следующем: эксперт по известному содержанию обобщенного понятия $F(d^H)$ (вершина концептуального каркаса, решетки) должен определить имя понятия (d^H) и факт включения в его объем обобщенного понятия $V(d)\subseteq V(d^H)$. Общее число обобщенных понятий, для которых

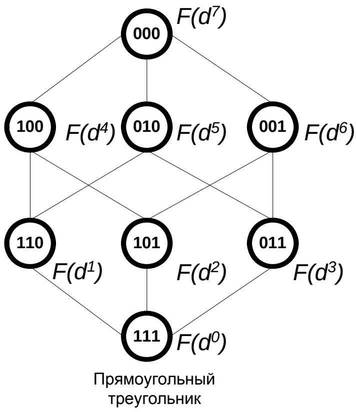

Рис. 1. Концептуальный каркас геометрических фигур

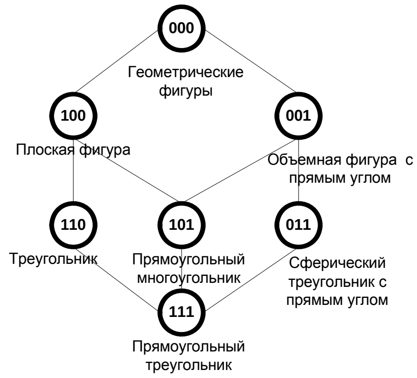

Рис. 2. Онтология геометрических фигур

эксперт должен установить имя и объем этого понятия, определиться по формуле (при условии, что число значений признаков одинаково):

$$N=n^k,$$

где *n* – число значений каждого признака; *k* – число признаков исходного понятия.

2. Качественные концептуальные каркасы в семантических пространствах

Ранее был рассмотрен концептуальный каркас объекта, признаки которого принимали бинарные значения, т.е. признак есть либо его нет. Рассмотрим случай, когда признак принимает значение из некоторого множества возможных значений.

Одной из моделей представления экспертных знаний, используемой в психологии для анализа их организации, является модель субъективного семантического пространства.

Субъективное семантическое пространство (от лат. Subjectum – подлежащее и греч. Semantikos – обозначающий) — система категорий индивидуального сознания, при помощи которых происходит оценка и классификация различных объектов, понятий [12].

Размерность семантического пространства определяется числом признаков понятия. Каждая ось семантического пространства соответствует одному из признаков понятия, а сами понятия представляются в виде точек в этом пространстве. Положение точки, характеризующее понятие, определяется значениями его признаков. Для оценки и классификации понятий в семантическом пространстве принимается гипотеза о том, что это пространство является метрическим, что подтверждают эксперименты [13].

В плохо определенной предметной области субъект наблюдает один объект, который представляется в семантическом пространстве, размерность которого определяется числом признаков понятия, характеризующего наблюдаемый объект.

Пусть объект v^0 имеет множество признаков $F=\{f_i\}$. Пусть известны множества возможных значений каждого признака наблюдаемого объекта – $Z_i=\{z_{ik}\}$.

Определение 4. Множество $Z_i=\{z_{ik}\}$ значений признака f_i будем называть качественным доменом, считая, что все элементы этого множества строго упорядочены, т.е. $z_{ik} > z_{ik-1}$, $\forall k$.

Определение 5. Семантическим пространством $SS(v^0)$ объекта v^0 будем называть пространство, определяемое прямым произведением качественных доменов всех его признаков, т.е. $SS(v^0)=\times_i Z_i$.

Определение 6. Объект v^0 в семантическом пространстве $SS(v^0)$ определяется вектором значений всех его признаков $v^0 = (z_1, \ldots, z_m)$, $v^0 \in SS(v^0)$.

В семантическом пространстве можно представить множество объектов предметной области в виде векторов значений их признаков $v_j^0 = (z_{j1}, \ldots, z_{jm})$, $v_j^0 \in SS(v^0)$, $\forall j$. Обычно считается, что в семантическом пространстве задана метрика $\delta(\cdot)$, а объекты, имеющие близкие по значению признаки $\delta\,(v_i; v_j)<\varepsilon$, принадлежат одному классу (понятию).

Рассмотрим окрестность точки семантического пространства, представляющей объект $v^0 = (z_1\pm\varepsilon_1, \ldots, z_m\pm\varepsilon_m)$, где $\pm\varepsilon_i \in Z_i$, $\forall i$.

Определение 7. Окрестность $z_i\pm\varepsilon_i$ значений i-го признака объекта $v^0 = (z_1, \ldots, z_m)$, $z_i \in \pm\varepsilon_i$, в котором не меняется имя класса, к которому он принадлежит, будем называть интервалом толерантности класса по этому признаку $\Delta_i = [z_{ik}+\varepsilon_{ik}, z_{ik} -\varepsilon_{ik}]$.

Определение 8. Подпространство семантического пространства $SS(d^0) \subseteq SS(v^0)$, полученное прямым произведением интервалов толерантности всех признаков объекта v^0, будем называть базовым понятием или классом d^0 объекта v^0:

$$SS(d^0)= \times_i [\, z_{ik}+\varepsilon_{ik}, z_{ik} -\varepsilon_{ik}]= \times_i \Delta_i\, , \; SS(d^0)\subseteq SS(v^0).$$

Определение 9. Содержанием $F(d^0)$ понятия d^0 заданного в семантическом пространстве $SS(v^0)$ будем называть подпространство $SS(d^0)$.

Определение 10. Объемом $V(d^0)$ понятия d^0 будем называть множество объектов $V(d^0)=\{v_j^0\}$, значения признаков которых попадают в область толерантности этого понятия $SS(d^0)$.

Для дальнейших рассуждений сделаем допущение, что любая точка семантического пространства $SS(v^0)$ является реальным объектом. Это допущение позволит нам в дальнейших рассуждениях опускать утверждение: «Если существует (объект) экземпляр». Т.е. мы считаем, что реальный объект определен для любой точки семантического пространства $SS(v^0)$.

Утверждение 1. Подпространство $SS(d^H) \subset SS(v^0)$, определяющее понятие d^H является обобщением базового понятия d^0, определенное подпространством $SS(d^0) \subset SS(v^0)$, в том и только в том случае если $SS(d^0)\subset SS(d^H)$.

Пусть имеется включение подпространств $SS(d^0)\subset SS(d^H)$. Определим объем понятия d^0: $V(d^0)=\{v_j^0\}$, где $v_j^0 \in SS(d^0)$; и понятия d^H: $V(d^H)=\{v_q^0\}$, где $v_q^0 \in SS(d^H)$, $\forall j,q$. Если учесть сделанное ранее допущение, о том, что в любой точке семантического пространства определен объект, то будет выполняться достаточное условие - включение объемов $V(d^0)\subset V(d^H)$, т.е., объем обобщенного понятия d^0 включен в объем обобщающего понятия d^H .

Определение 11. В семантическом пространстве $SS(v^0)$ обобщенное понятие (класс) d^H называется реальным обобщенным понятием, если разность объемов обобщенного понятия $V(d^H)$ и обобщаемого понятия $V(d^0)$ есть не пустое множество, т.е. $V(d^H)\backslash V(d^0)\neq\emptyset$. В противном случае это обобщенное понятие будем называть виртуальным.

Это означает, что если существует хотя бы один объект v_j^0, принадлежащий $v_j^0 \in SS(d^H) \backslash SS(d^0)$, то обобщенное понятие реальное и может быть представлено в концептуальной решетке вершиной.

Пусть задано базовое понятие в семантическом пространстве тройкой d^0, $SS(d^0)$, $V(d^0)$. С учетом определений 8-11 и утверждения 1 определим качественные обобщенные понятия d^H базового понятия d^0 в семантическом пространстве $SS(v^0)$.

Утверждение 2. Качественным положительным обобщением понятия d^0 по признаку q является понятие, имеющее области толерантности для признака q: $\Delta_q^h = [\, z_{qk} - \varepsilon_{qk}, supZ_q]$ и, соответственно, определенное подпространством: $SS(d^H) = \underset{i(i \neq q)}{\times} \Delta_i \times [\, z_{qk} - \varepsilon_{qk}, supZ_q]$.

Утверждение 3. Качественным отрицательным обобщением понятия d^0 по признаку q является понятие, имеющее области толерантности для признака q: $\Delta_q^h = [\, inf\, Z_q , z_{qk} + \varepsilon_{qk},]$ и, соответственно, определенное подпространством $SS(d^H) = \underset{i(i \neq q)}{\times} \Delta_i \times [\, inf\, Z_q, z_{qk} + \varepsilon_{qk},]$.

Утверждение 4. Качественным обобщением понятия d^0 по k-признакам является понятие, имеющее интервалы толерантности по необобщенным признакам равным интервалам толерантности базового понятия, а интервалы толерантности соответственно равны $\Delta_q^h = [\, inf\, Z_q , z_{qk} + \varepsilon_{qk},]$ или $[\, z_{qk} - \varepsilon_{qk}, supZ_q]$.

Отличительной особенностью качественного обобщения является, то, что при любом числе возможных значений признаков в семантическом пространстве рассматриваются только три интервала их значений:

- базовый интервал – $[z_{ik} - \varepsilon_{ik}, z_{ik} + \varepsilon_{ik}]$;
- больше базового интервала – $[z_{ik} - \varepsilon_{ik}, supZ_i]$ – положительное обобщение;
- меньше базового интервала – $[\, inf\, Z_i, z_{ik} + \varepsilon_{ik},]$ – отрицательное обобщение.

Сложность построения качественного концептуального каркаса равна $N = n^3$, т.е. $|\{SS(d^H)\}| = n^3$, $H = 0,\dots,(n^3 - 1)$, где n – число признаков.

Рассмотрим множество всех подпространств $\{SS(d^H)\}$, $\forall H$, полученных в результате качественного обобщения базового понятия по одному, двум и т.д. признакам в семантическом пространстве $SS(v^0)$, $SS(d^H) \subset SS(v^0)$. Для любой пары подпространств верны следующие свойства:

- рефлективности: $\forall a,\ SS(d^a) \subseteq SS(d^a)$;
- антисимметричности: $\forall i, q,\ SS(d^i) \subset SS(d^q) \wedge SS(d^q) \subset SS(d^i) \Rightarrow S(d^i) = SS(d^q)$;
- транзитивности: $\forall a, b, c,\ SS(d^a) \subset SS(d^b) \wedge SS(d^b) \subset SS(d^c) \Rightarrow SS(d^a) \subset SS(d^c)$.

Для любой пары подпространств из множества $\{SS(d^H)\}$ определены верхняя и нижняя границы. Справедливость этого утверждения основана на самом принципе структуризации семантического пространства на вложенные подпространства SS(d^H), включающие базовое понятие – подпространство SS(d^0).

Если множество подпространств $\{SS(d^H)\}$ удовлетворяет перечисленным свойствам и для любой пары подпространств определены верхняя и нижняя границы, то на этом множестве определена решетка $(\{SS(d^H)\}, \cap, \cup)$, структурирующая семантическое пространство.

Определение 12. Качественным концептуальным каркасом в семантическом пространстве будем называть решетку подпространств $(\{SS(d^H)\} \cap, \cup)$ этого семантического пространства.

Пример структуризации семантического пространства в виде качественного концептуального каркаса для объекта $v^0 = (Z_{1q}, Z_{2s})$ с двумя признаками, значения которых представлены упорядоченными множествами $Z_1 = (z_{11},\dots, z_{1n})$ и $Z_2 = (z_{21},\dots, z_{2m})$, показаны на Рис.3-Рис.6. Определим семантическое пространство SS(v^0)=$Z_1 \times Z_2$.

Базовое понятие (Рис.3) определено как подпространство:

$$SS(d^0) = [Z_{1q} + \varepsilon_1, Z_{1q} - \varepsilon_1] \times [Z_{2s} + \varepsilon_2, Z_{2s} - \varepsilon_2].$$

Графическое представление всех обобщений (d^1, d^2, d^3, d^4) базового понятия по одному признаку представлено на Рис.4. Подпространства обобщенных понятий определятся из соотношений:

$SS(d^1) = [Z_{1q} - \varepsilon_1, Z_{1n}] \times [Z_{2s} + \varepsilon_2, Z_{2s} - \varepsilon_2]$, $SS(d^0) \subset SS(d^1)$;
$SS(d^2) = [Z_{1q} + \varepsilon_1, Z_{1q} - \varepsilon_1] \times [Z_{2s} - \varepsilon_2, Z_{2m}]$, $SS(d^0) \subset SS(d^2)$;
$SS(d^3) = [Z_{1q} + \varepsilon_1, Z_{11}] \times [Z_{2s} + \varepsilon_2, Z_{2s} - \varepsilon_2]$, $SS(d^0) \subset SS(d^3)$;
$SS(d^4) = [Z_{1q} + \varepsilon_1, Z_{1q} - \varepsilon_1] \times [Z_{2s} + \varepsilon_2, Z_{21}]$, $SS(d^0) \subset SS(d^4)$.

На Рис.5 показаны обобщенные понятия (d^5, d^6, d^7, d^8), обобщающие базовое понятие по двух признакам.

Подпространства обобщенных по двум признакам понятий определятся из соотношений:

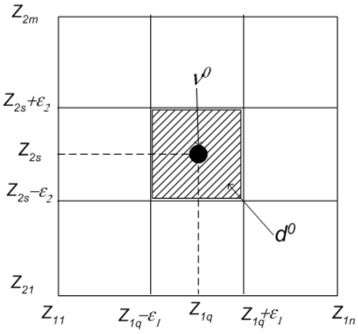

Рис. 3. Базовое понятие в семантическом пространстве

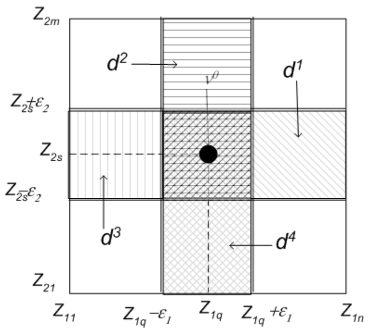

Рис. 4. Обобщения базового понятия по одному признаку

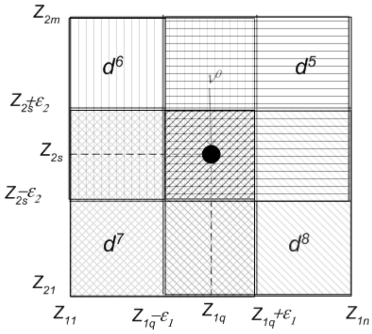

Рис. 5. Обобщения базового понятия по двум признакам

$$SS(d^5)= [Z_{1q} -\varepsilon_1, Z_{1n}]\times[Z_{2s} -\varepsilon_2, Z_{2m}],$$
$$SS(d^1)\subset SS(d^5),\ SS(d^2)\subset SS(d^5);$$
$$SS(d^6)= [Z_{1q} +\varepsilon_1, Z_{11}]\times[Z_{2s} -\varepsilon_2, Z_{2m}],$$
$$SS(d^2)\subset SS(d^6),\ SS(d^3)\subset SS(d^6);$$
$$SS(d^7)= [Z_{1q}+\varepsilon_1, Z_{11}]\times[Z_{2s}+\varepsilon_2, Z_{21}],$$
$$SS(d^3)\subset SS(d^7),\ SS(d^4)\subset SS(d^7);$$
$$SS(d^8)= [Z_{1q} -\varepsilon_1, Z_{1n}]\times[Z_{2s}+\varepsilon_2, Z_{21}],$$
$$SS(d^1)\subset SS(d^8),\ SS(d^4)\subset SS(d^8).$$

Качественный концептуальный каркас семантического пространства для абстрактного объекта v^0 с двумя признаками (Z_{1q}, Z_{2s}) показан на Рис.6.

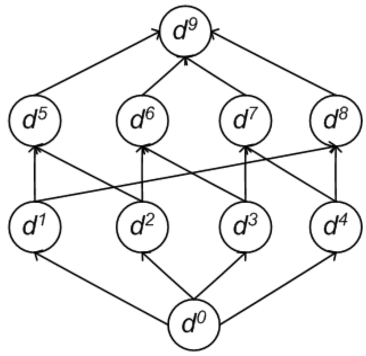

Рис. 6. Качественный концептуальный каркас в семантическом пространстве

3. Зависимости признаков и структура концептуального каркаса

Были рассмотрены концептуальные каркасы в предположении, что все признаки понятия независимы. Рассмотрим случай, когда для значений признаков определена функциональная зависимость. Функциональная зависимость имеет теоретико-множественное определение как *n-арное* отношение, заданное на декартовом произведении возможных значений всех признаков $\underset{i}{\times} Z_i$, т.е. отображение $\varphi: \underset{i}{\times} Z_i^X \rightarrow \underset{i}{\times} Z_i^Y$, где $\underset{i}{\times} Z_i^X$ - значения независимых признаков (аргументов) $f_i^X \in F$, $\underset{i}{\times} Z_i^Y$ - значения зависимых признаков $f_i^Y \in F$. Зависимость φ можно представить как множество упорядоченных *n-ок* $\{(z_{ie}^Y;\ z_{te}^X,\dots z_{je}^X)\}$, где z_{ie}^Y – значение зависимой переменной, $z_{te}^X,\dots z_{je}^X$ - значения независимых переменных. То есть для зависимых признаков заменен верхний индекс X f_i^X на Y, f_i^Y.

Рассмотрим пример задания функциональной зависимости на множестве признаков $F(d)$, имеющих бинарные значения (есть признак или нет признака). В этом случае множество $\varphi(F(d))$ будет подмножеством булеана $\boldsymbol{B}(F(d))$, т.е. $\varphi(F(d))\subseteq \boldsymbol{B}(F(d))$. Все подмножества содержания понятия с зависимостями его признаков $\varphi(F(d))$ образуют частично упорядоченное множество по включению его элементов, т.е. концептуальный каркас онтологии.

В качестве примера рассмотрим концептуальный каркас прямоугольного треугольника (Рис. 1). Однако второе свойство – геометрическая фигура «*ограничена тремя сторонами*» заменим другим определением – «*ограничена тремя прямыми*». Такое определение позволяет сформулировать зависимость, выраженную правилом: «*Если Геометрическая фигура ограничена тремя прямыми, То она Плоская*». Это правило порождает концептуальный каркас, показанный на Рис.7, в котором отсутствуют узлы $F(d^3)=(011)$ и $F(d^5)=(010)$, поскольку правило приводит содержание $F(d^3)=(011)$ к содержанию $F(d^0)=(111)$, и, соответственно, $F(d^5)=(010)$ к $F(d^1)=(110)$.

Рассмотрим теперь семантическое пространство $SS(v^0)$ объекта v^0. Пусть определен кон-

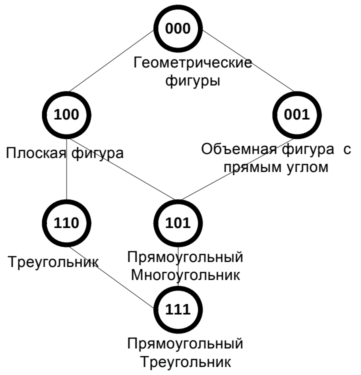

Рис. 7. Онтология геометрических фигур с учетом зависимостей признаков

цептуальный каркас в виде решетки подпространств семантического пространства $(\{SS(d^H)\}, \cap, \cup)$. Пусть определены зависимости между значениями признаков для наблюдаемого объекта $\varphi\colon \underset{i}{\times} Z_i^X \rightarrow \underset{i}{\times} Z_i^Y$. При этом зависимость φ между значениями признаков объекта ν^0, определенная для базового понятия $SS(d^0)= \underset{i}{\times}\Delta_i$, считается законом, действующим в предметной области, к которой этот объект принадлежит. В этом случае допущение о том, что в каждой точке семантического пространства определен объект уже не действует, но могут существовать объекты, значения признаков которых определены точками семантического пространства, связанные зависимостью φ. Из этого следует утверждение.

Утверждение 5. Если на множестве значений признаков объектов семантического пространства $SS(\nu^0)$ определена зависимость φ, то концептуальный каркас онтологии предметной области $(\{SS(d^H)\}, \cap, \cup)$ будет включать только те классы обобщенных понятий концептуального каркаса, для которых разности объемов обобщенного и необобщенного понятия непустое множество (утверждение 1), т.е. обобщенное понятие реально.

Поясним утверждение 5 на примере семантического пространства объекта, имеющего два признака (пример предыдущего раздела). Вначале определим все разности подпространств базового понятия d^0 и подпространств его обобщений по одному признаку (d^1, d^2, d^3, d^4):

$$SS(d^{1'})=SS(d^1)\backslash SS(d^0) = $$
$$= [Z_{1q}+\varepsilon_1, Z_{1n}]\times[Z_{2s}+\varepsilon_2, Z_{2s}-\varepsilon_2];$$
$$SS(d^{2'})=SS(d^2)\backslash SS(d^0) = $$
$$= [Z_{1q}-\varepsilon_1, Z_{11}]\times[Z_{2s}+\varepsilon_2, Z_{2s}-\varepsilon_2];$$
$$SS(d^{3'})=SS(d^3)\backslash SS(d^0) = $$
$$= [Z_{1q}-\varepsilon_1, Z_{1q}+\varepsilon_1]\times[Z_{2s}+\varepsilon_2, Z_{2m}];$$
$$SS(d^{4'})=SS(d^4)\backslash SS(d^0) = $$
$$= [Z_{1q}-\varepsilon_1, Z_{1q}+\varepsilon_1]\times[Z_{2s}-\varepsilon_2, Z_{21}].$$

На Рис.8 показаны все выделенные подпространства разностей обобщенных по одному признаку понятий и базового понятия (разности подпространств заштрихованы).

Далее рассмотрим разности подпространств понятий, обобщенных по двум признакам (d^5, d^6, d^7, d^8) и понятий, обобщенных по одному признаку (d^1, d^2, d^3, d^4). Выделим следующие подпространства (Рис.9):

$$SS(d^{5'})=SS(d^5)\backslash SS(d^1)\cup SS(d^3)=$$
$$= [Z_{1q}+\varepsilon_1, Z_{1n}]\times[Z_{2s}+\varepsilon_2, Z_{2m}];$$
$$SS(d^{6'})=SS(d^6)\backslash SS(d^2)\cup SS(d^3) = $$
$$= [Z_{11}, Z_{1q}-\varepsilon_1]\times[Z_{2s}+\varepsilon_2, Z_{2m}];$$
$$SS(d^{7'})=SS(d^7)\backslash SS(d^2)\cup SS(d^4) = $$
$$= [Z_{11}, Z_{1q}-\varepsilon_1]\times[Z_{2s}-\varepsilon_2, Z_{21}];$$
$$SS(d^{8'})=SS(d^8)\backslash SS(d^1)\cup SS(d^4) = $$
$$= [Z_{1q}+\varepsilon_1, Z_{1n}]\times[Z_{2s}-\varepsilon_2, Z_{21}].$$

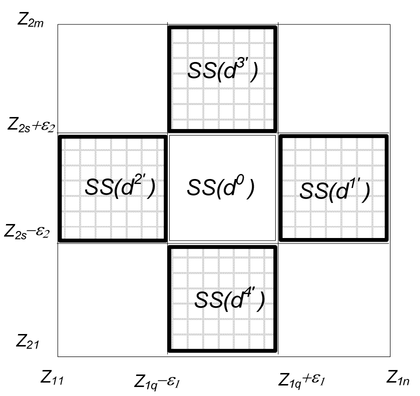

Рис.8. Разности подпространств понятий, обобщенных по одному признаку

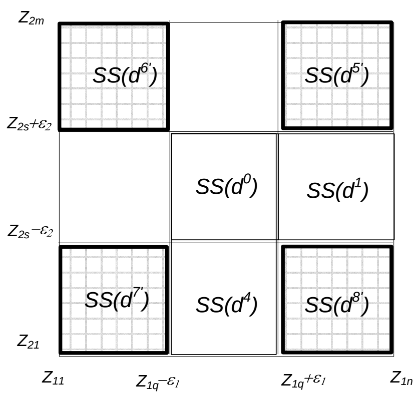

Рис. 9. Разности подпространств понятий, обобщенных по двум признакам

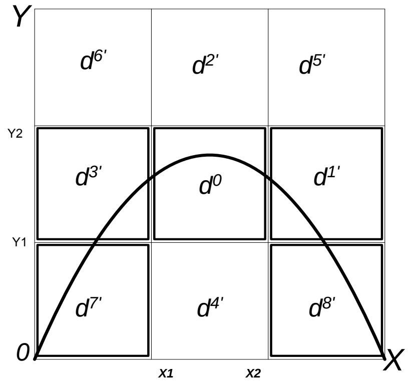

Рис. 10. Зависимость признаков $y=-kx^2+c$

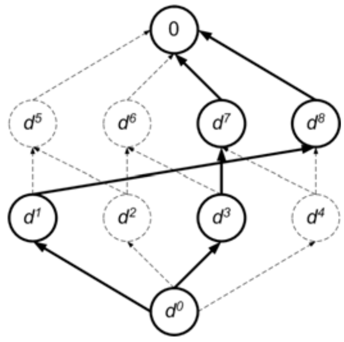

Рис. 11. Концептуальный каркас онтологии для зависимости признаков $y=-kx^2+c$

Согласно определению 11, для того чтобы обобщенное понятие было реальным, необходимо чтобы в подпространстве разности обобщенного и необобщенного понятий была точка, координаты которой соответствуют значениям признаков некоторого реального объекта. Т.е. если есть некоторый объект и известно, что значение его свойств попадают в подпространство $SS(d^{2'})$, то это означает, что обобщенное понятие d^2 реально. Или, например, если значения признаков объекта принадлежат подпространству $SS(d^{6'})$, то реальным будет обобщенное понятие d^6. Напомним, что реальные обобщенные понятия включаются в концептуальный каркас предметной области, которой принадлежит объект.

Рассмотрим еще несколько примеров концептуальных каркасов для случаев, когда зависимости между признаками объекта предметной области заданы графиками функций. Пусть дан гипотетический объект с двумя признаками, значения которых действительные числа, определенные на интервале числовой оси $x\in X$ и $y\in Y$, x, $y\in R^+$. Известна также зависимость значений признаков $y=-kx^2+c$, определенная для $\forall x\in X$ и $\forall\, y\in Y$. Пусть эксперт определил базовое понятие $SS(d^0)=[x1,x2]\times[y1,y2]$ в пространстве $SS(d^0)\subseteq X\times Y$. Определен также концептуальный каркас и разности всех подпространств обобщенных и необобщенных понятий концептуального каркаса. На Рис. 10 показан график функции $y=-kx^2+c$ в пространстве $X\times Y$, структурированном на подпространства разностей $SS(d^{1'}),\ldots,\ SS(d^{8'})$. Поскольку определенная зависимость считается законом для предметной области, то она верна для объекта с любыми значениями признаков $\forall x\in X$ и $\forall y\in Y$. Согласно утверждению 5 реальными обобщенными понятиями этой предметной области будут те, разности $SS(d^{1'}),\ldots,\ SS(d^{8'})$ которых пересекает график определенной ранее зависимости. В нашем случае – это понятия d^1 и d^3, обобщающие базовое понятие по одному признаку, и понятия d^7 и d^8, обобщающие по двум признакам. На Рис. 11 показан концептуальный каркас онтологии предметной области при заданной зависимости между значениями признаков. На Рис. 12 и Рис. 13 показан еще один пример построения концептуального каркаса для функциональной зависимости $y=x^k$ $(k<1)$.

Интерес представляет построение концептуальных каркасов процессов динамических систем. Под процессом будем понимать траекторию изменения состояния системы во времени в пространстве состояний динамической системы (фазовом пространстве).

Рассмотрим простой пример затухающего процесса колебания маятника. Его уравнение динамики зададим конечно-разностным уравнением $x(t+1)= -\,k\,x(t)$, $k<1$. Эта динамическая система характеризуется двумя признаками: отклонение x, $x\in[-1,1]$ и время t, $t=1,\ldots n$. Необходимо отметить, что приведенное уравнение качественно описывает изменение состояния маятника не по синусоидальному, а по линейному закону в виде пилообразных затухающих колебаний.

На Рис. 14 показано изменение состояния маятника (отклонения от состояния равновесия) в пространстве: {*Равновесие, Влево, Вправо*}. Дискретное время также представлено – {*Прошлое, Настоящее, Будущее*}. В пространстве состояний определено базовое понятие d^0 – состояние равновесия маятника, концепту-

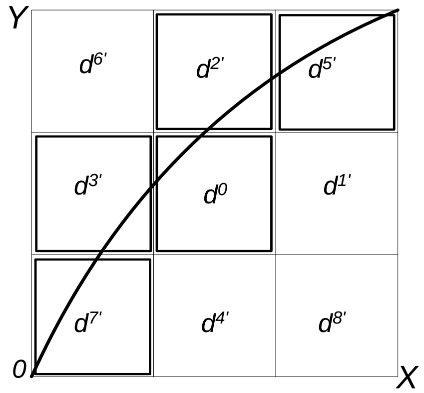

Рис. 12. Зависимость признаков $y=x^k$ $(k<1)$

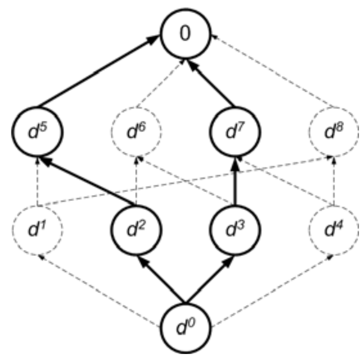

Рис. 13. Концептуальный каркас онтологии для зависимости признаков $y=x^k$ $(k<1)$

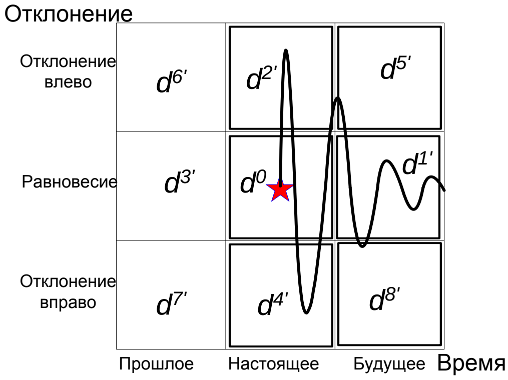

Рис. 14. График процесса «Затухающие колебания»

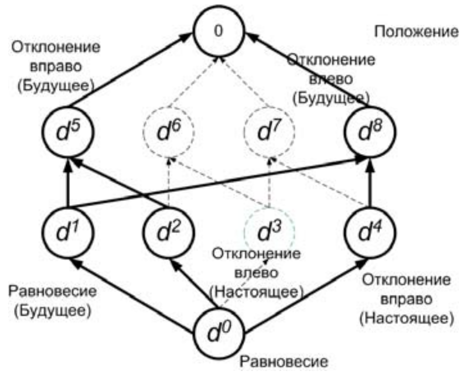

Рис. 15. Концептуальный каркас процесса «Затухающие колебания»

альный кластер, разности подпространств обобщенных понятий ($d^{1'}, d^{2'}, d^{4'}, d^{5'}, d^{8'}$). Изменение состояния процесса, описываемого уравнением динамики, представлено в виде затухающего процесса (линейный пилообразный затухающий процесс сглажен до синусоидального процесса). На Рис. 15 показан качественный концептуальный каркас этого процесса.

Заключение

Предложенные модель представления знаний и метод построения онтологий на основе концептуальных каркасов направлены на активизацию интеллектуальной деятельности и креативности эксперта на этапах построения онтологии плохо определенной предметной области. Концептуальные каркасы могут также применяться для поддержки интерпретации результатов моделирования в системах имитационного моделирования, для представления их в базах знаний интеллектуальных систем различного назначения.

Литература

1. Рабчевский Е.А. Автоматическое построение онтологий на основе лексико-синтаксических шаблонов для информационного поиска. Труды 11-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL’2009, Петрозаводск, Россия, 2009. с. 69-77.
2. Большакова Е.И., Васильева Н.Э., Морозов С.С. Лексико-синтаксические шаблоны для автоматического анализа научно-технических текстов // Десятая Национальная конференция по искусственному интеллекту с международным участием КИИ-2006. Труды конференции в 3-х томах. М.: Физматлит, 2006. Т. 2. С.506-524.
3. Мозжерина Е.С. Автоматическое построение онтологии по коллекции текстовых документов. Труды 13-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL’2011, Воронеж, Россия, 2011.
4. Найханова Л.В. Методы и модели автоматического построения онтологий на основе генетического и автоматного программирования: Автореф. дис. докт. тех. наук. — Красноярск, 2008. — 36 с.
5. Haav H. A semi-automatic method to ontology design by using FCA, in: V. Snásel, R. Belohlávek (Eds.), Proceed-

ings of Concept Lattices and their Applications (CLA), Ostrava, Czech Republic, 2004. pp. 13-24.

6. Cimiano P., Staab S., Tane J. Automatic acquisition of taxonomies from text: FCA meets NLP. In Proceedings of the PKDD/ECML'03 International Workshop on Adaptive Text Extraction and Mining (ATEM), 2003, pp.10-17.
7. Ganter B., Wille R. Formal Concept Analysis, Mathematical Foundations, Springer, 1999.
8. Baader F., Molitor R. Building and Structuring Description Logic Knowledge Bases Using Least Common Subsumers and Concept Analysis. In B. Ganter and G. W. Mineau, eds., Proceedings of the 8th International Conference on Conceptual Structures (ICCS 2000), volume 1867 of Lecture Notes in Computer Science, pp. 292-305. Springer-Verlag, 2000.
9. Чечкин А.В. Математическая информатика. – М.: Наука. Гл. ред. Физ-мат. лит., 1991. – 416 с.
10. Baader F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P.F. The Description Logic Handbook: Theory, Implementation, and Applications.
11. Кулинич А.А. Моделирование динамических процессов в понятийной системе субъекта для генерации креативных решений. Когнитивные исследования: Сборник научных трудов: Вып. 1/ Под редакцией В.Д. Соловьева. – 2006. с. 94-123.
12. Психологический словарь. - М.: Педагогика-Пресс, 1996.
13. Shepard R.N. Metric structures in ordinal data// J. of Math. Psy-chol. – 1966 -V.3. - №.2.

Кулинич Александр Алексеевич. Старший научный сотрудник института проблем управления им. В.А. Трапезникова РАН. В 1981г. окончил Государственный технический университет Армении (ЕрПИ). Кандидат технических наук. Автор более 60-ти печатных работ. Область научных интересов: искусственный интеллект, поддержка принятия решений, когнитивная психология. E-mail:kulinich@ipu.ru, alexkul@rambler.ru