OSTIS-2013
(Open Semantic Technologies for Intelligent Systems)

УДК 004.822

КОНЦЕПТУАЛЬНЫЕ «КАРКАСЫ» ПЛОХО ОПРЕДЕЛЕННЫХ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Кулинич А.А.

Федеральное Государственное учреждение науки Институт проблем управления им В.А. Трапезникова, Российской академии наук, г. Москва, Российская Федерация

kulinich@ipu.ru

Исследуются модели представления знаний о плохо определенных 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».

Ключевые слова: онтология; концептуальный «каркас»; плохо определенная предметная область.

Введение

В настоящее время в системах поддержки принятия решений (СППР) все более широкое применение находят формальные модели экспертных знаний, представленные онтологиями разных уровней. Онтологический подход в СППР расширяет возможности поддержки принятия решений при решении множества классов задач, множеством различных методов.

Важной особенностью онтологического подхода создания СППР является его открытость и возможность обработки разнородных данных, хранящихся различных базах данных или серверах сети Интернет. Открытость СППР, построенной на онтологических принципах выражается в возможности ее расширения, в том числе и за счет уже существующих онтологий предметных областей. Задачами поддержки принятия решений, решаемые в рамках онтологического подхода – это: поддержка выбора метода решения задач; классификация информации; задача интеграции разнородной информации; информационный поиск релевантной для поддержки принятия решения информации, интерпретация результатов решения задачи на естественном языке.

Узким местом при создании онтологической СППР в плохо изученной предметной области можно считать построение онтологий предметной области. Обычно онтологии строятся экспертом, но в последнее время появляются исследования автоматического или автоматизированного построения онтологий на основе анализа текстов, описывающих предметную область.

При автоматическом построении онтологии, на первом этапе построения осуществляется подбор текстов, отражающих закономерности предметной области. Далее применяют один из следующих подходов: полный лингвистический анализ, т.е. морфологический, синтаксический и семантический анализы текста; лексико-синтаксические шаблоны [Рабчевский, 2009], позволяющие выделить из текста основные понятия и отношения «Класс-Подкласс» («Isa») между ними; статистические методы, позволяющие выделить основные понятия и отношения «Класс-Подкласс» на множестве всех слов и словосочетаний с использованием эвристик [Мозжерина, 2011]; на основе продукций [Найханова, 2008] с применением генетического и автоматного программирования; на основе анализа формальных понятий (FCA) [Ganter, 1999].

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

Интересны работы по автоматизированному построению онтологий на основе экспертно сформированного представительного множества объектов предметной области. В работе [Baader, 2000] исследуются вопросы поддержки построения онтологии предметной области методом «Bottom-Top». Здесь выделяется представительное

Библиотека БГУИР

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 [Ganter, 1999].

К сожалению, в работе [Baader, 2000] не исследованы вопросы формирования представительного множества объектов для построения онтологии. Но этот вопрос интересен для построения онтологий плохо определенных предметных областей. Например, можно ли построить онтологию предметной области по одному объекту из этой области?

Рассмотрим одну из возможных теоретических моделей представления экспертных знаний о предметной области в условиях неопределенности [Чечкин, 1992]. 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: $(p)\delta_l(x_0)\in X$, где $(p)\in[1,0]$ – степень истинности наличия свойства δ_l у объекта x_0. Если сведение истинно $(p=1)$, то ее опускают при описании объекта, т.е. $(1)\delta_l(x_0)=\delta_l(x_0)\in X$.

Информация о том, что у объекта x_0 нет свойства δ_l, $\neg\delta_l(x_0)$ 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 $\delta_l(x_0)$ – $(\wedge, \vee, \neg)$, 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 считается не просто элементарные сведения об объекте $\{\delta_l(x_0)\}$, 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 $\{\delta_l(x_0)\}$ порождает структуру сведений – решетку понятий $(L(x_0), \wedge, \vee)$, которая несет информацию не только об объекте x_0, но и информацию о предметной области и ее структурной организации.

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

1. Концептуальный «каркас» онтологии предметной области

Среди многих определений онтологии выделим следующее: онтология предметной области – это кортеж:

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

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; содержание понятия – это атрибуты класса; объем понятия – это множество экземпляров класса.

В дескриптивной логике (ДЛ) [Baader, 2004], 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 специфический объект этой предметной области.

Формально определим понятие (класс) этого специфического объекта тройкой: $\langle d_i,F(d_i),V(d_i)\rangle$, где d_i – имя понятия, $F(d_i)=\{f_{ij}\}$- содержание понятия (множество признаков), $V(d_i)=\{v^{msc}\}$ – объем понятия, где v^{msc} – наиболее специфический объект, имеющие признаки $F(d_i)$.

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

Как видим, элементы булеана образуют частично упорядоченное множество, так же как и классы онтологии в определении (1).

Алгебраическую решетку, образованную элементами булеана можно считать прообразом онтологии предметной области, если сделать два

Библиотека БГУИР

следующих допущения:

1. Любой элемент $F(d_i^H)\in B(F(d_i))$, $H\in[1,2^{|F(d)|}]$, полученной решетки формально будем считать содержанием понятия d_i^H обобщающего понятие d_i, если $F(d_i^H)\subseteq F(d_i)$.

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

2. Отношение включения содержаний элементов решетки $(F(d_i^H)\subseteq F(d_i))$ будем считать отношением «Класс-Подкласс» (*Isa*), при условии, что для объемов этих понятий выполняется достаточное условие $V(d_i)\subseteq V(d_i^H)$.

Это допущение означает, что если определены (например, экспертным способом) элементы объема (экземпляры класса) удовлетворяющие условию $V(d_i)\subseteq V(d_i^H)$, то степень истинности p из первого допущения принимает значение «истина», т.е. p=1.

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

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

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

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

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

Пример концептуального «каркаса» и экспертно построенной онтологии предметной области геометрических фигур приводится в работе [Кулинич, 2006]. Здесь определен наиболее специфический объект – прямоугольный треугольник как: плоская геометрическая фигура; ограничена тремя сторонами; один угол прямой. 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.

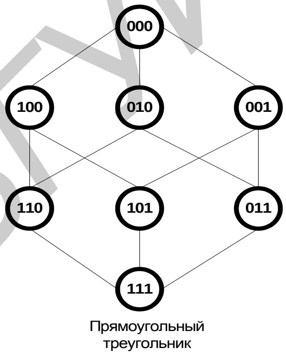

Рисунок 1 – Концептуальный «каркас» геометрических фигур

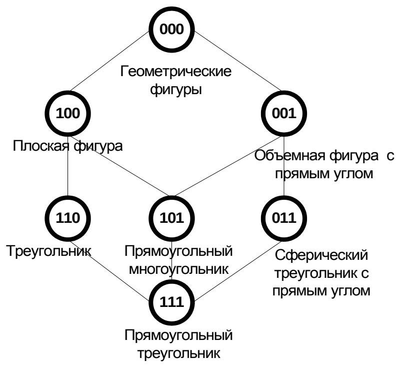

Рисунок 2 – Онтология геометрических фигур

Концептуальный «каркас» прямоугольного треугольника приведен на рисунке 1. Онтология геометрических фигур, построенная на основе этого концептуального «каркаса» с использованием экспертных процедур приведена на рисунке 2.

Общее число обобщенных понятий, для которых эксперт должен определить имя и объем этого понятия определиться по формуле (при условии, что число значений признаков одинаково):

$$N=n^k,$$

где, N – число обобщенных понятий в концептуальной решетке; n – число значений каждого признака; k – число признаков исходного понятия.

Библиотека БГУИР

2. Качественные концептуальные «каркасы» в семантических пространствах

Одной из моделей представления экспертных знаний для анализа их организации является модель субъективного семантического пространства, которое представляется как система категорий индивидуального сознания, при помощи которой происходит оценка и классификация объектов, понятий [Психологический словарь, 1996]. Размерность семантического пространства определяется числом признаков понятия. Каждая ось семантического пространства соответствует одному из признаков понятия, а сами понятия представляются в виде точек. 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 [Shepard, 1966]. В плохо определенной предметной области субъект наблюдает один объект, который представляется в семантическом пространстве, размерность которого определяется числом признаков понятия, характеризующего наблюдаемый объект.

Пусть есть 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 этого признака, 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. $SS(v^0)= \underset{i,j}{\times} Z_{ij}$.

Определение 6. Объект v_i^0 в семантическом пространстве $SS(v^0)$ определяется вектором значений всех его признаков $v_i^0 = (z_{ik}, \ \dots, \ z_{im})$, $v_i^0 \in SS(v^0)$.

Похожие объекты в семантических пространствах расположены на небольшом расстоянии (считаем, что в семантическом пространстве задана метрика) и принадлежат одному классу (понятию).

Рассмотрим окрестность точки семантического пространства, представляющей объект $v_i^0 = (z_{ik}\pm\varepsilon_{ik}, \dots, z_{im}\pm\varepsilon_{im})$, где $\pm\varepsilon_{ik}\in Z_i$.

Определение 7. Окрестность $z_{ik}\pm\varepsilon_{ik}$ значений ik-го признака объекта $v_i^0 = (z_{ik}, \ \dots, \ z_{im})$, $z_{ik}\in\pm\varepsilon_{ik}$, в котором не меняется имя класса, к которому этот объект 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 признаков объекта v^0 будем называть базовым понятием или классом объекта v^0:

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

Для дальнейших рассуждений сделаем допущение, что любая тока семантического пространства является экземпляром понятия. Т.е. $(z_{ik}, \ \dots, \ z_{im}) \in Z_i, \ \forall i,k,m, z_{ik}, \ \exists v_i \in SS(v^0)$.

Это допущение позволит нам в дальнейших рассуждениях опускать утверждение: «Если существует экземпляр». Т.е. мы считаем, что реальный объект определен в любой точке семантического пространства.

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

Пусть задано базовое понятие d^0, для множества признаков, которого определены интервалы толерантности каждого признака, $\Delta_i = [z_{ik}-\varepsilon_{ik}, z_{ik}+\varepsilon_{ik}]$ и подпространство $SS(d^0)$.

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:

$$\mathrm{SS}(d_i^H) = \underset{i(i\neq q)}{\times} \Delta_i \times [\, z_{qk} -\varepsilon_{qk}, supZ_q]$$

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:

$$\mathrm{SS}(d_i^H) = \underset{i(i\neq q)}{\times} \Delta_I \times [\, inf\, Z_q, z_{qk} +\varepsilon_{qk},]$$

Утверждение 4. Качественным обобщением понятия d^0 по k-признакам является понятие, имеющее интервалы толерантности по необобщенным признакам равными интервалам толерантности базового понятия, а интервалы толерантности по обобщенным признакам 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 следующих интервала 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.

Библиотека БГУИР

Сложность построения качественного концептуального «каркаса» равна – $N=n^3$, где n – число признаков.

Для всех элементов множества подпространств $\{SS(d_i^H)\}$, $\forall i,H$, определены следующие свойства:

1. Рефлективности: $\forall a$, $SS(d_a^H)\subset SS(d_a^H)$;
2. Антисимметричности: $\forall i, q$, $SS(d_i^H)\subset SS(d_q^H)\wedge SS(d_q^H)\subset SS(d_i^H)\Rightarrow S(d_i^H)=SS(d_q^H)$;
3. Транзитивности: $\forall a, b, c$, $SS(d_a^H)\subset SS(d_b^H)\wedge SS(d_b^H)\subset SS(d_c^H)\Rightarrow SS(d_a^H)\subset SS(d_c^H)$.

Наличие этих свойств означает, что на множестве подпространств $\{SS(d_i^H)\}$ определен частичный порядок всех подмножеств.

Утверждение 5. Для любой пары подпространств из множества $\{SS(d_i^H)\}$ определены верхняя и нижняя границы.

Если для пары подпространств $\{SS(d_i^H)\}$ определены верхняя и нижняя границы и они удовлетворяют свойствам 1-3, то на множестве подпространств определена решетка, которая является концептуальным «каркасом» предметной области.

Определение 9. Качественным концептуальным «каркасом» в семантическом пространстве будем называть решетку подпространств $(\{SS(d_i^H)\}\wedge,\vee)$ этого семантического пространства.

Пример структуризации семантического пространства для объекта с двумя признаками в виде качественного концептуального «каркаса» показан на рисунках 3-6.

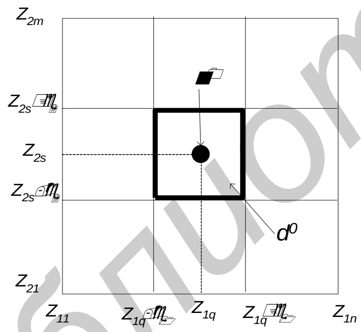

Рисунок 3 – Базовое понятие в семантическом пространстве

Базовое понятие (Рисунок 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 Рисунке 4. Подпространства обобщенных понятий определятся из следующих соотношений:

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

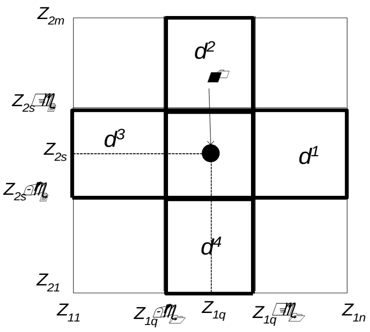

Рисунок 4 – Обобщения базового понятия

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

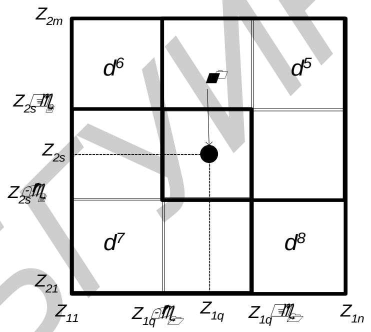

Рисунок 5 – Обобщения понятий (d^1, d^2, d^3, d^4).

Подпространства обобщенных по двум признакам понятий определятся их соотношений:

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

Качественный концептуальный «каркас» для абстрактного объекта с двумя признаками показан на рисунке 6.

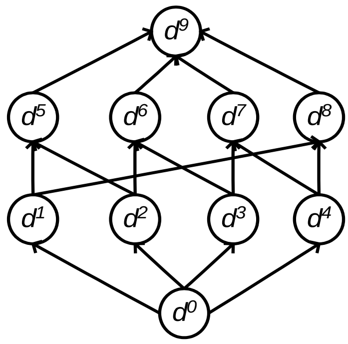

Рисунок 6 – Качественный концептуальный «каркас» в семантическом пространстве

3. Зависимости признаков и структура концептуального «каркаса»

Несколько снизить трудоемкость построения концептуального «каркаса» можно, если учитывать закономерности строения наиболее специфического объекта предметной области. Эти закономерности

Библиотека БГУИР

выражаются функциональными зависимостями между признаками объекта.

В этом случае понятие определяем четверкой:

$$\langle d_i, F(d_i), \varphi(F(d_i)), V(d_i) \rangle,$$

где $\varphi(F(d_i))$ – зависимость значений признаков понятия d_i .

Задание функциональной зависимости на множестве значений признаков означает, что множество всех подмножеств $\boldsymbol{B}(\varphi(F(d_i)))$ будет подмножеством булеана $\boldsymbol{B}(F(d_i))$, т.е. $\boldsymbol{B}(\varphi(F(d_i))) \subseteq \boldsymbol{B}(F(d_i))$.

Утверждение 6. Все подмножества содержания понятия с зависимостями его признаков $\boldsymbol{B}(\varphi(F(d_i)))$ образуют частично упорядоченное множество по включению его элементов, т.е. концептуальный «каркас» онтологии.

В качестве примера рассмотрим концептуальный «каркас» прямоугольного треугольника (Рисунок 1). Однако, второе свойство – геометрическая фигура «*ограничена тремя сторонами*» заменим другим определением – «*ограничена тремя прямыми*». Такое определение позволяет сформулировать зависимость, выраженную правилом: $\varphi(F(d_i))$=*(Если Геометрическая фигура ограничена тремя прямыми, То она Плоская)*. Это правило порождает концептуальный «каркас», показанный на рисунке 7 (отсутствуют узлы $F(d_3)=(011)$ и $F(d_5)=(010)$).

Рисунок 7 – Онтология геометрических фигур с учетом зависимостей признаков

Рассмотрим семантическое пространство $SS(v^0)$ некоторого объекта v^0. Пусть определено его базовое понятие ($SS(d^0)$) – класс. Пусть определен концептуальный «каркас» в виде решетки подпространств семантического пространства.

Пусть между значениями признаков разных объектов существует функциональная зависимость (определены отношения).

Сделаем ряд допущений. Считаем, что если определены зависимости между значениями признаков объекта, принадлежащему классу, то эти зависимости справедливы для любого объекта этого класса. Сделаем еще более сильное допущение: зависимость между признаками некоторого объекта считается законом, действующим в предметной области, к которой этот объект принадлежит.

Утверждение 7. Если для признаков объектов некоторого семантического пространства на множестве значений его признаков определены отношения $R \subseteq SS(v^0)$, то концептуальный «каркас» онтологии предметной с заданной функциональной зависимостью признаков будет включать только те классы обобщенных понятий концептуального «каркаса», подпространству которых принадлежат элементы этого отношения. Т.е. $K(d^0)=\{SS(d_i^H) \mid_{SS(d_i^H) \cap R \neq \varnothing}\}$.

В качестве примера рассмотрим многозначный формальный контекст представленный в таблице 1. Здесь перечислены пять сортов яблок, для каждого из которых определены значения двух признаков: *Цвет* из множества значений *{Зеленый, Желтый, Красный}* и *Вкус* из множества значений *{Кислый, Кисло-сладкий, Сладкий}*.

Таблица 1 – Многозначный формальный контекст

№	Сорт/ Признак	Цвет	Вкус
1	***Антоновка***	*Зеленый*	*Кислый*
2	***Фуджи***	*Красный*	*Сладкий*
3	***Белый налив***	*Желтый*	*Кисло-сладкий*
4	***Апорт***	*Красный*	*Кисло-сладкий*
5	***Ренет***	*Желтый*	*Кислый*

Для выявления зависимости между значениями признаков воспользуемся аппаратом анализа формальных понятий [Ganter, 1999], позволяющим по формальному контексту определить концептуальную решетку и выявить закономерности предметной области в виде правил, отражающих зависимости значений признаков перечисленных объектов. Были получены следующие правила:

1. Если **Вкус** *Кислый*, То **Цвет** *Зеленый* или *Желтый.*
2. Если **Вкус** *Кисло-сладкий*, То **Цвет** *Желтый* или *Красный*.
3. Если **Цвет** *Желтый*, То **Вкус** *Кислый* или *Кисло-сладкий*.
4. Если **Цвет** *Красный*, То **Вкус** *Сладкий* или *Кисло-сладкий*.

Правила, полученные на основе анализа, могут быть сформулированы экспертом, знакомым с этой предметной областью. Ниже (см. рисунки 8-11) рассмотрен пример построения качественного концептуального «каркаса» для желтого яблока, имеющего кисло-сладкий вкус на основе экспертных представлений о выше названных закономерностях предметной области (правил 1-4).

На рисунке 8 определено базовое понятие – «*Яблоко*» с признаками: Цвет – *Желтый*; Вкус – *Кисло-сладкий.*

Библиотека БГУИР

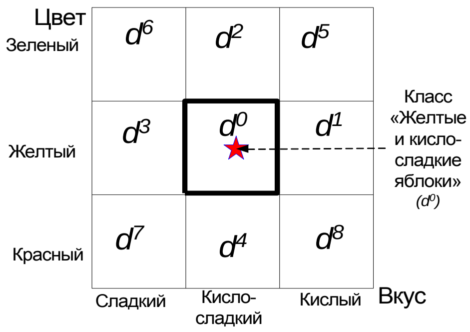

Рисунок 8 – Базовое понятие – *Желтое, кисло-сладкое яблоко*.

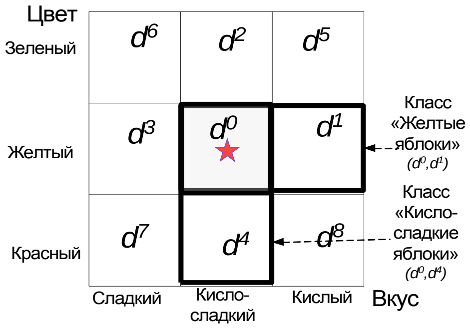

Рисунок 9 – Обобщения базового понятия по одному признаку

На рисунке 8 показано базовое понятие «*Желтое, кисло-сладкое яблоко*». На рисунке 9 показаны два обобщения первого уровня. По признаку *Вкус* обобщение осуществляется по правилу 3, т.е. ***Если*** цвет яблока *желтый*, ***То*** вкус - *кисло-сладкий* или *кислый*. Это - класс «*Желтые яблоки*». Обобщение по признаку *Цвет* строится по правилу 2, т.е. ***Если*** *Кисло-сладкие* яблоки, ***То*** цвет яблок - *желтый* или *красный*. Получили класс «*Кисло-сладкие яблоки*»

На рисунке 10 показаны два обобщения второго уровня - это классы: *Красные яблоки* имеют вкус *кисло-сладкий* или *сладкий* (используется правило 4) и *Кислые яблоки* имеют цвет *желтый* или *зеленый* (используется правило 1). На рисунке 11 показан концептуальный качественный «каркас» этой предметной области.

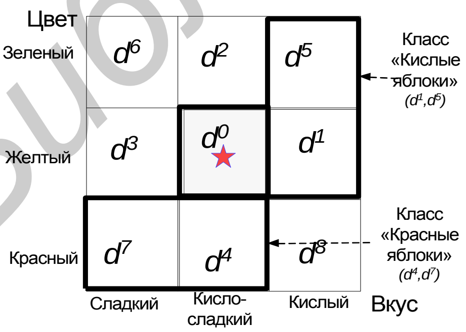

Рисунок 10 – Обобщения базового понятия по двум признакам.

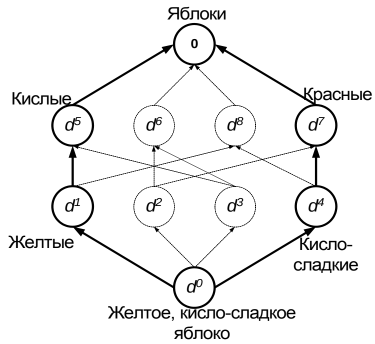

Рисунок 11 – Качественный концептуальный «каркас»

Зависимости между признаками объекта предметной области могут быть заданы графиками функций. В качестве примера рассмотрены два графика (Рисунки 12 и 14). В этом случае концептуальные «каркасы» будут включать только те классы обобщенных понятий, которые пересекает график этих функций (Рисунки 13 и 15).

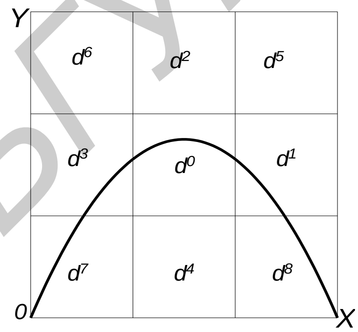

Рисунок 12 – Зависимость признаков $y=-kx^2+c$.

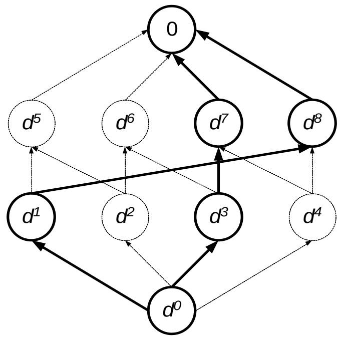

Рисунок 13 – Концептуальный «каркас» онтологии для зависимости признаков $y=-kx^2+c$.

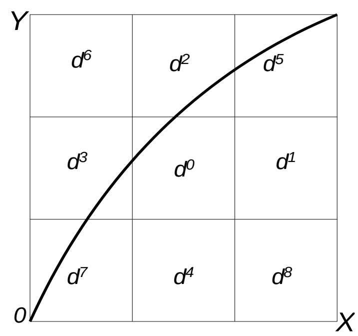

Рисунок 14 – Зависимость признаков $y=x^k$ $(k<1)$.

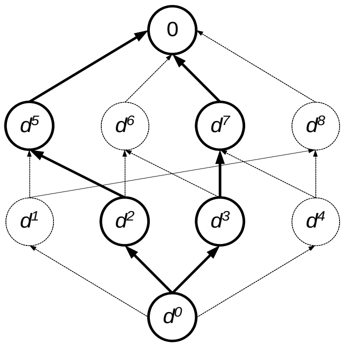

Рисунок 15 – Концептуальный «каркас» онтологии для зависимости признаков $y=x^k$ $(k<1)$.

Интерес представляет построение концептуальных «каркасов» процессов динамических систем. Под процессом будем понимать траекторию изменения состояния системы в фазовом пространстве. На рисунке 16 показано изменение состояния маятника. На рисунке 17 показан качественный концептуальный «каркас» этого процесса.

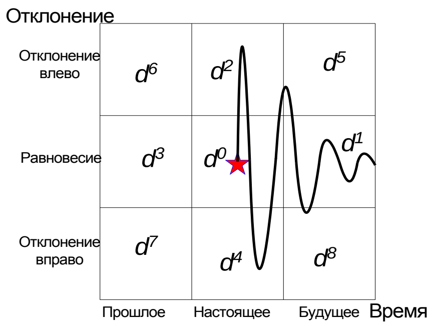

Рисунок 16 – График процесса «Затухающие колебания»

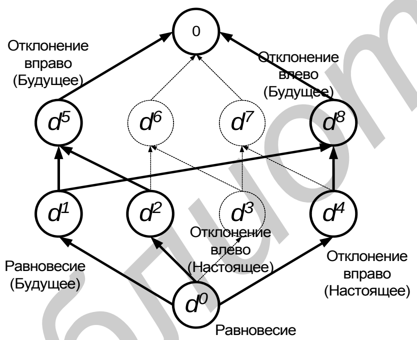

Рисунок 17 – Концептуальный «каркас» процесса «Затухающие колебания»

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.

Библиографический список

[Baader, 2000] F. Baader and R. Molitor. 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.

[Baader, 2004] Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. The Description Logic Handbook: Theory, Implementation, and Applications.

[Ganter, 1999] Ganter, B. and Wille, R.: Formal Concept Analysis, Mathematical Foundations, Springer, 1999.

[Shepard, 1966] Shepard R.N. Metric structures in ordinal data// J. of Math. Psy-chol. – 1966 -V.3. - №.2.

[Кулинич, 2006] 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.

[Мозжерина, 2011] Мозжерина Е. С. Автоматическое построение онтологии по коллекции текстовых документов. Труды 13й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2011, Воронеж, Россия, 2011.

[Найханова, 2008] Найханова Л. В. Методы и модели автоматического построения онтологий на основе генетического и автоматного программирования: Автореф. дис. докт. тех. наук. — Красноярск, 2008. — 36 с.

[Психологический словарь, 1996] Психологический словарь. - М.: Педагогика-Пресс, 1996.

[Рабчевский, 2009] 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. стр. 69-77.

[Чечкин, 1991] Чечкин А.В. Математическая информатика. – М.: Наука. Гл. ред. Физ-мат. лит., 1991. – 416 с. – ISBN 5-02-014136-4.

Conceptual "templates" of the uncertainty subject domains

A. A. Kulinich

V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Russia, Moscow.

kulinich@ipu.ru

The model of knowledge representation, and ontology construction method of the uncertain subject domain on the basis of conceptual "templates" is offered.

Introduction

Existing methods of the automated and automatic construction of subject domain ontology are considered. Their inefficiency at construction ontology in uncertain subject domains is shown.

Main Part

Questions of conceptual "template" construction of a subject domain using one object from this area are investigated. Dependence of conceptual "template" structure from dependences of characteristic values of object on which conceptual "template" has been constructed is shown.

Conclusion

The offered model of knowledge representations and ontology construction method on the basis of conceptual "templates" are directed on stimulation of intellectual activity and expert creativeness at stages uncertain subject domain ontology construction.

Библиотека БГУИР