



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

АЛГОРИТМИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ОНТОЛОГИЧЕСКОГО АНАЛИЗА ДАННЫХ

Зубцов Р.О. *, Семенова В.А. *, Смирнов С.В. **

*Самарский государственный аэрокосмический университет им. С.П. Королёва
(научно-исследовательский университет), г. Самара, Россия

zibtsov.r@gmail.com

queenbfjr@gmail.com

**Институт проблем управления сложными системами
Российской академии наук, г. Самара, Россия

smirnov@iccs.ru

В статье рассматриваются математические модели, методы и алгоритмы онтологического анализа данных, составляющие аналитическое ядро разрабатываемого инструментария семантического моделирования и проектирования, и предназначенные для вывода формальных онтологий на основе эмпирических объектно-признаковых данных, организованных в виде соответствия «объекты-свойства». Оцениваются перспективы развития соответствующего программного обеспечения.

Ключевые слова: онтологический анализ данных; анализ формальных понятий; многозначная логика; ограничение существования свойств; веб-сервис.

Введение

Изложенный в [Смирнов, 2015а] опыт создания средств семантического моделирования и проектирования указал в целом хорошо известные задачи семантических технологий в инструментальных системах, и среди этих задач была отмечена ключевая роль *онтологического анализа* для моделирования смысла предметной области (ПрО), составляющей интерес разработчика. Компьютерная поддержка этого анализа осуществляется либо путем предоставления эксперту ПрО инструментов «приобретения» знаний [Гаврилова, 2008] (т.е. средств *автоформализации*), либо благодаря когнитивному *у анализу данных* о ПрО [Загоруйко, 2013]. Поэтому аналитическим ядром развиваемого в течение ряда лет в Институте проблем управления сложными системами РАН инструментария семантического моделирования и проектирования стали методы и средства *онтологического анализа данных* (ОАД) [Семенова, 2015].

ОАД – методический комплекс для выявления *формальной онтологии* ПрО на основе эмпирических объектно-признаковых данных, структурированных согласно стандартизированному формату представления подобной информации,

известному как таблицы «объекты-свойства» (ТОС) [Барсегян, 2009].

В основу ОАД положен теоретически хорошо обоснованный метод *анализа формальных понятий* (АФП) [Ganter, 1999], в котором отражен классический подход к понятию как к фундаментальному элементу знания с объемом и содержанием. ОАД представляет собой прагматическую надстройку АФП, обладающую рядом новых способностей:

- поддерживается построение *произвольных отношений* между объектами (согласно концепции «отношение-проявление внутренних свойств объектов» [Смирнов, 2011]);
- допускается различного рода *неполнота* исходной эмпирической информации, и обрабатываются *противоречия* в извлекаемых из первичных данных *базовых семантических суждениях* (БСС) о ПрО вида $b_{xy} = \langle \text{объекту } x \text{ присуще свойство } y \rangle$;
- в анализ вовлекаются обычно имеющиеся у эксперта сведения о взаимосвязи измеряемых свойств объектов – так называемые «*ограничения существования*» свойств (ОСС) [Lammari, 2004].

Для реализации второй способности при оценке истинности БСС в отличие от *нечеткого АФП* [Yanget al., 2008] предложено использовать более

адекватную многозначную векторную логику [Смирнов, 2015b]. При этом и в том, и в другом варианте выявлена необходимость связывания модели неполноты исходных данных с моделью ОСС, и первоначально этот вопрос проработан применительно к концептуальной сопряженности свойств, возникающей в результате выполнения фундаментальной когнитивной процедуры АФП – концептуального шкалирования свойств [Офицеров, 2014].

В предлагаемой статье наряду с общим очерком операционной основы ОАД представлен метод и алгоритм работы с более общей моделью ОСС, а также проанализирована перспектива развития соответствующего программного обеспечения.

1. Математические модели и методы

В соответствии с общими положениями теории измерений ОАД основывается на том, что всякое измерение свойства объекта может дать специальный результат «None», свидетельствующий о нахождении значения измеряемого свойства за порогами чувствительности, вне динамического диапазона средства измерений. В АФП считается, что подобная фиксация семантического несоответствия исследуемого объекта и измерительной процедуры является, прежде всего, результатом фундаментального когнитивного приема – концептуального шкалирования [Ganter, 1999]. Т.е. исследователь сообразно субъективному видению априори «расщепляет» действительный диапазон процедуры измерения свойства, образуя набор новых свойств объектов ПрО, фактически измеряемых после этого в бинарной шкале наименований $\{X, \text{None}\}$, где X – лингвистическая константа, собирательно обозначающая любой символ шкалы динамического диапазона измерительной процедуры.

Как бы то ни было, «None-концепция» позволяет радикально изменить парадигму анализа данных и для начала естественным образом преобразовать ТОС в совокупность оценок истинности БСС $_{b_{xy}}$:

$$\|b_{xy}\| = \begin{cases} \text{Истина,} & \text{если результат измерения} \\ & \text{свойства у объекта } x \text{ есть } X; \\ \text{Ложь,} & \text{в противном случае.} \end{cases}$$

Именно на обработку таких данных ориентирован АФП, в котором используются следующие обозначения и модели:

- $K = (G^*, M, I)$ – формальный контекст (ФК), где $G^* = \{g_i\}_{i=1, \dots, r}$, $r = |G^*| \geq 1$ – набор объектов исследуемой ПрО, попавших в поле зрения исследователя (т.е. множество объектов обучающей выборки: $G^* \subseteq G$, где G – всё мыслимое множество объектов ПрО), $M = \{m_j\}_{j=1, \dots, s}$, $s = |M| \geq 1$ – множество измеряемых у объектов свойств, I – бинарное соответствие «объекты-свойства», т.е. совокупность оценок $\|b_{ij}\| \in \{\text{Истина}, \text{Ложь}\}$;

- операторы Галуа φ, ω (общая нотация «'») для контекста K :

$\varphi(X) = X' = \{m_j \mid m_j \in M, \forall g_i \in X: (g_i, m_j) \in I\}$ – общие свойства объектов, составляющих $X \subseteq G^*$;

$\omega(Y) = Y' = \{g_i \mid g_i \in G^*, \forall m_j \in Y: (g_i, m_j) \in I\}$ – объекты, которые обладают всеми свойствами из $Y \subseteq M$;

- (X, Y) – формальное понятие, у которого $X \subseteq G^*$ – объем, $Y \subseteq M$ – содержание, причем $X = Y', Y = X'$;

- $B(K)$ – множество формальных понятий контекста K ;

- $(B(K), \leq)$ – замкнутая решетка понятий, где $(X_1, Y_1) \leq (X_2, Y_2)$, если $X_1 \subseteq X_2$, или эквивалентно $Y_1 \supseteq Y_2$.

Естественную на практике многозначность оценок истинности БСС вызывает неполнота данных о ПрО (неточность, противоречивость, неопределенность и т.п.), которая вызывается реалиями накопления эмпирической информации: выполнением многократных независимых измерений свойства $m_j \in M$ у объекта $g_i \in G^*$, использованием для измерения одного и того же свойства m_j нескольких различных (но конгруэнтных) процедур, дифференциацией доверия к различным процедурам измерения. Поэтому в качестве адекватной модели исходных данных предлагается обобщенная ТОС, описываемая коротжем

$$(G^*, M, Se, Pr, A) \quad (1)$$

где:

- $Se = \bigcup_{i=1}^r Se_{(i)}$ – множество всех выполненных при зондировании ПрО серий измерений, $Se_{(i)} = \{se_{(i)k}\}_{k=1, \dots, q_{(i)}}$, $q_{(i)} \geq 1$, $i = 1, \dots, r$ – множество серий измерений, которым подвергнут объект $g_i \in G^*$;

- $Pr = \bigcup_{j=1}^s Pr_{(j)}$ – арсенал всех используемых при зондировании ПрО процедур измерения, $Pr_{(j)} = \{pr_{(j)k}\}_{k=1, \dots, p_{(j)}}$, $p_{(j)} \geq 1$, $j = 1, \dots, s$ – множество конгруэнтных процедур измерения свойства $m_j \in M$, причем всякая процедура $pr_{(j)k}$ характеризуется степенью доверия к ее результатам $t_{(j)k} \in (0, 1]$;

- $A = (a_{ij})_{i=1, \dots, m; j=1, \dots, n}$ – матрица результатов серий измерений Se свойств M у объектов из выборки G^* , выполненных с помощью процедур измерения Pr . Элементами этой матрицы могут быть константы X и None , а также еще две лингвистические константы. Константа **Failure** фиксирует отказ, сбой измерительного средства, воздержание при голосовании и т.п., т.е. тот нередко наблюдаемый на практике «результат» работы измерительной процедуры, который собирательно можно квалифицировать как «отказ от выполнения измерения». Константа **NM** (*not measured*) указывает, что в действительности в рассматриваемой серии измерений отдельная процедура измерения не использовалась (введение этого формального результата необходимо в том

числе и для сохранения двумерного характера обобщенной ТОС).

Задачей первичной обработки такой исходной эмпирической информации является извлечение оценок БСС $\|b_{ij}\|$. В [Смирнов, 2015b] на основе модели (1) и моделей векторной логики V^{TF} [Аршинский, 2007] дан метод построения формального контекста K с *нестрогим* соответствием I , которое образуют векторные оценки истинности $\|b_{ij}\| \in \langle b^+_{ij}, b^-_{ij} \rangle$, $b^+_{ij}, b^-_{ij} \in [0, 1]$, где компонент b^+_{ij} - *Истина* - формируется свидетельствами, подтверждающими БСС, а компонент b^-_{ij} - *Ложь* - отрицающими БСС.

Эффективные методы вывода понятийной структуры ПрО из «мягких» ФК основаны на их предварительной α -аппроксимации при задании экспертом порога доверия к исходным данным. Затем к полученным бинарным контекстам применяются (с различными дополнениями) апробированные АФП-методы вывода понятий. Однако в [Офицеров, 2014] показано, что в общем случае этот подход оказывается некорректным, поскольку стандартное α -сечение «мягкого» соответствия «объекты-свойства» «слепо» к *зависимостям* между измеряемыми свойствами объектов ПрО.

Общие модели подобных зависимостей предложены в [Lammari, 2004]. Здесь достаточно учесть, что пара свойств $m_j, m_k \in M$, $j \neq k$ для любого объекта ПрО (и, следовательно, для $\forall g_i \in G^*$) может быть:

- *несовместимой*, если, обладая свойством m_j , объект g_i заведомо не обладает свойством m_k , и наоборот, т.е.
 $E(m_j, m_k) \leftrightarrow \forall g_i \in G^*: m_j \in \{g_i\}' \rightarrow m_k \notin \{g_i\}'$;
- *обусловленной*, если, обладая свойством m_j , объект g_i непременно обладает свойством m_k (хотя обратное может быть неверно), т.е.
 $C(m_j, m_k) \leftrightarrow \forall g_i \in G^*: m_j \in \{g_i\}' \rightarrow m_k \in \{g_i\}'$.

Так, в результате концептуального шкалирования в описании задачи возникают непересекающиеся подмножества попарно несовместимых свойств - *группы концептуально сопряженных свойств* (ГКСС).

Важно подчеркнуть, что рассматриваемые зависимости между свойствами известны субъекту, строящему формальное описание ПрО, *априори*, т.к. он является «владельцем» арсенала процедур измерения свойств и, обладая об этих инструментах в идеале полной информацией, может знать об определенных связях между результатами измерения одного и того же объекта.

Поэтому в качестве интегрированной модели входных данных для ОАД целесообразно использовать кортеж

$$(G^*, M, Se, Pr, A, M_{\Pi}, E_{\Pi}, C_{\Pi}, V), \quad (2)$$

где дополнением к (1) являются:

- M_{Π} - множество актуальных для исследователя «*протосвойств*» объектов ПрО, $1 \leq |M_{\Pi}| \leq |M|$, $M_{\Pi} = M_{\Pi 1} \cup M_{\Pi 2}$, $M_{\Pi 1} \cap M_{\Pi 2} = \emptyset$; $M_{\Pi 1}$ - подмножество *единичных* протосвойств (т.е. актуальных свойств не подвергнутых концептуальному шкалированию, формально «*одинокных*» ГКСС); $M_{\Pi 2}$ - подмножество «*расщепленных*» протосвойств (на два и более измеряемых свойства), или совокупность «*множественных*» ГКСС, в каждой из которых составляющие её свойства несовместимы;
- E_{Π} - множество пар несовместимых протосвойств, $E_{\Pi} \subseteq M_{\Pi} \times M_{\Pi}$;
- C_{Π} - множество пар обусловленных протосвойств, $C_{\Pi} \subseteq M_{\Pi} \times M_{\Pi}$;
- V - пары сопряженных свойств-валентностей $V \subseteq M \times M$, определяемых предикаторами актуальных бинарных отношений между объектами ПрО [Смирнов, 2011].

Идея метода построения корректной α -аппроксимации *нестрогого* ФК (НФК) предусматривает замену работы с общим порогом доверия α выбором *индивидуальных* порогов доверия к *определённым* фрагментам нестрогого соответствия I , характеризующим *каждый* объект $g_i \in G^*$. При этом допустимую α -аппроксимацию каждой такой локальной зоны нестрогого соответствия I предлагается осуществлять *без участия* человека исходя из необходимого и достаточного набора условий, начиная с требования минимального *ужесточения* порога доверия к рассматриваемому фрагменту данных и заканчивая применением различных мер *ранжирования* векторов истинности в V^{TF} -логике [Офицеров, 2014].

2. Алгоритмический базис ОАД

Основу алгоритмического обеспечения ОАД составляют результативные реализации упоминавшихся в разделе 1 методов, в том числе:

- формирование НФК на основе обобщенной ТОС;
- построение бинарного ФК путем *рационального* α -сечения НФК;
- вывод формальных понятий с выявлением отношения *обобщения* на их множестве (другие отношения в получаемой *решетке* формальных понятий оказываются установленными благодаря картине распределения валентностей в содержаниях понятий);
- трансформация решетки формальных понятий в формальную *онтологию* (с выполнением ряда прагматических требований к подобным описаниям ПрО).

Изучение и эксперименты с различными алгоритмами вывода формальных понятий позволили выбрать и адаптировать для ОАД алгоритм, описанный в [Yang et al., 2008]. Его достоинствами являются простота и наглядность при приемлемой производительности.

Другие из перечисленных алгоритмов оригинальны. Последний из указанных достаточно подробно описан в [Семенова, 2015], а наиболее важным и интересным представляется алгоритм построения рационального α -сечения НФК. Схема этого алгоритма приведена на рисунке 1.

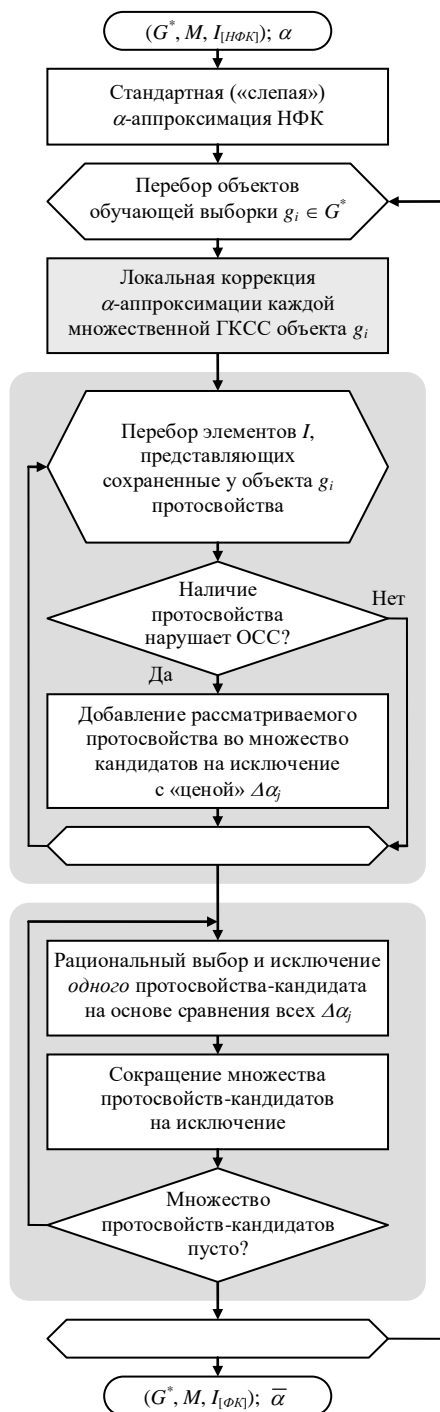


Рисунок 1 – Укрупнённая схема алгоритма построения рационального α -сечения нестрогого формального контекста

Рисунок 1 позволяет уяснить, что локальная коррекция стандартного α -сечения НФК производится на двух «уровнях»: сначала для каждой множественной ГКСС всех объектов $g_i \in G^*$ (см. выделенный заливкой блок), затем для множества протосвойств объектов обучающей выборки.

Такой подход целесообразен в силу двух обстоятельств. Во-первых, результат коррекции стандартной α -аппроксимации множественной ГКСС характеризуется существенной априорной определенностью: либо приходится признавать, что данный объект не обладает протосвойством, соответствующим рассматриваемой ГКСС, либо следует фиксировать « α -истинность» одного и только одного БСС, определяемого рассматриваемой ГКСС. А во-вторых, локальная коррекция α -аппроксимации множественной ГКСС не требует контроля соблюдения множества различных ОСС (оба вложенных цикла, выделенные на рисунке 1 фоном, включают эти акты) т.к. в такой группе эти ограничения «однородны» - всегда лишь попарно несовместимы.

Рисунки 2-4 иллюстрируют некоторые из задач, решаемых рассматриваемым алгоритмом на уровне множественных ГКСС объектов обучающей выборки.

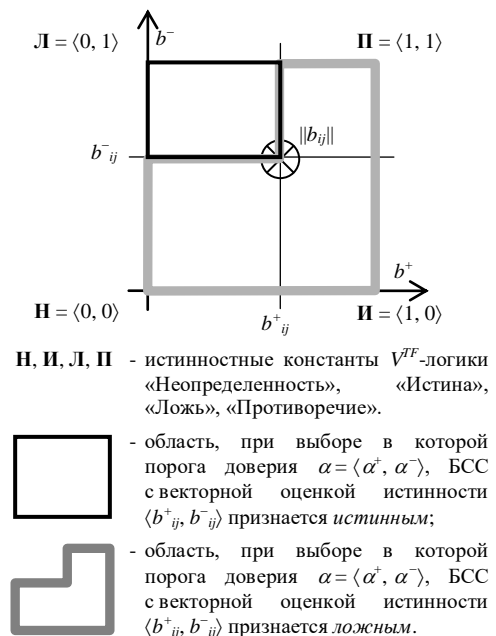


Рисунок 2 – Области выбора порога доверия к исходным данным, определяющие приближенную оценку истинности в шкале {Истина, Ложь} отдельно взятого базового семантического суждения о предметной области b_{ij}

Для каждой множественной ГКСС всякого объекта обучающей выборки алгоритм выполняет следующую последовательность действий:

- выявление *граничных элементов* множества оценок истинности *локальных* БСС (т.е. БСС соответствующих рассматриваемым объекту обучающей выборки и множественной ГКСС)- *множества недоминируемых альтернатив* выбора наиболее правдоподобных оценок (рисунок 3);
- установление для каждого граничного элемента *пределов* области выбора порога доверия, в которой истинно БСС, отвечающее этому элементу, а другие локальные БСС ложны, и вычисление значения меры *ужесточения* порога доверия, необходимого для такого выбора (рисунок 4);

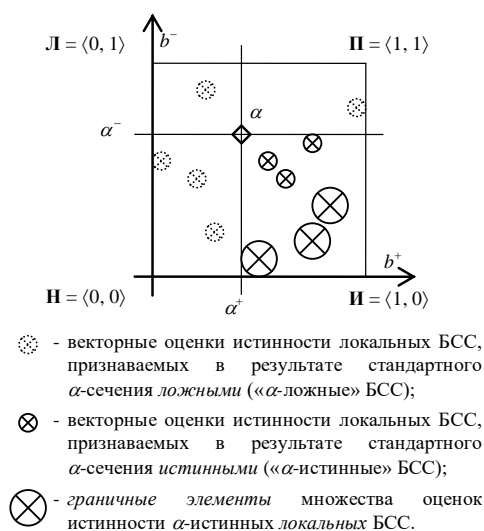


Рисунок 3 – Разделение оценок истинности локальных базовых семантических суждений

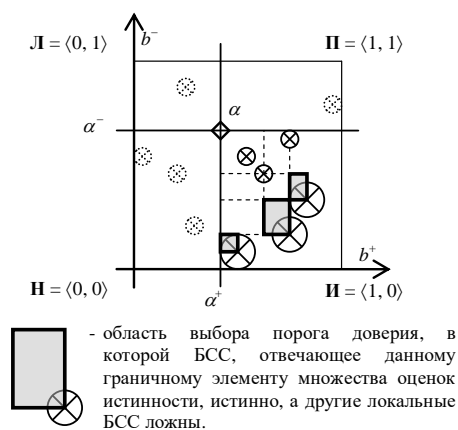


Рисунок 4 – Области корректного выбора порога доверия к данным, характеризующим множественную группу концептуально сопряженных свойств объекта обучающей выборки

- исключение из множества граничных элементов всех «дублированных» элементов (т.е. подмножеств эквивалентных оценок истинности), а также проверка существования неустрашимого противоречия локальных БСС – наличия среди них двух и более БСС, истинность которых оценивается вектором $\mathbf{I} = \langle 1, 0 \rangle$;

- если во множестве граничных элементов сохранилось два или более элементов (т.е. остаётся неопределённость выбора одного «лучшего» локального БСС), то производится *гарантированное* сокращение этого множества до *одноэлементного*. Этот результат достигается благодаря использованию трех скалярных мер, значения которых вычисляются по *текущему составу* множества граничных элементов. Сначала производится сохранение элемента(ов) с *наименьшим* значением меры *уменьшения порога доверия*. Если результат не достигнут, то выполняется сохранение элемента(ов) с *наибольшим* значением меры *достоверности* векторной оценки истинности [Аршинский, 2007]. Наконец (если выбор все ещё необходим), среди оставшихся граничных элементов указывается тот, который

характеризуется *наибольшим* значением меры *определённости* векторной оценки истинности [Аршинский, 2007].

3. Программное обеспечение и перспективы его развития

ОАД как аналитическое ядро *gB*-системы семантического моделирования и проектирования развивается на платформе MS Excel и языка программирования VBA [Смирнов, 2015a]. Однако очевидно, что оформленный в виде настольного приложения программный продукт удобней предоставлять конечным пользователям в качестве веб-сервиса. Для этого существует множество способов, и если приложение представляет собой отдельную самодостаточную программу, публикация его в интернете обычно не вызывает затруднений. Однако развертывание книги Excel, содержащей макросы, сопряжено с рядом проблем.

Стандартный экспорт в html позволяет сохранить определенную интерактивность приложения Excel, например сортировку данных или фильтры, однако не поддерживает макросы. Существуют специализированные сервисы, позволяющие публиковать книгу Excel как веб-страницу, например, Spreadsheet WEB или Excel Online, однако в настоящее время и они не осуществляют поддержку макросов^{1,2}.

Едва ли не единственным исключением здесь является продукт, разработанный компанией EASA Software³. Он позволяет экспортировать книгу Excel с макросами с минимальной затратой усилий – почти весь процесс преобразования происходит автоматически. Однако поддерживаются не все возможности VBA-макросов, например, работа контекстных меню. В целом данный программный продукт не пользуется особой популярностью и потому, что у него отсутствует достаточная документация.

Рассматривая перспективы развития программного обеспечения ОАД, следует отметить, что разработка *gB*-системы и все её составляющих, включая инструментарий ОАД, выполнена при последовательном соблюдении принципов объектно-ориентированного программирования и использовании соответствующих возможностей VBA. Поэтому для создания интерактивного веб-сервиса можно, переработать часть кода, использовать VBA в SharePoint⁴. Наконец, можно переписать весь функционал на Java с использованием, например, технологий JavaBeans, JSF и любого сервера приложений⁵.

¹ <https://blogs.office.com/2014/04/14/weve-updated-excel-online-whats-new-in-april-2014/>

² <http://wiki.pagos.com/display/SSWEB/Current+Restrictions>

³ <http://www.easasoftware.com/media-resources/easa-videos/>

⁴ [https://msdn.microsoft.com/en-us/library/office/aa159897\(v=office.11\).aspx](https://msdn.microsoft.com/en-us/library/office/aa159897(v=office.11).aspx)

⁵ <http://www.oracle.com/technetwork/java/javaee/javaserverfaces-139869.html>

Заключение

Наличие в арсенале систем семантического моделирования и проектирования средств онтологического анализа данных не просто расширяет их функциональный потенциал. Основу таких средств по необходимости составляют фундаментальные представления о когнитивных возможностях человека и соответствующие математические модели и методы работы со знаниями, которые могут быть использованы технологической системой моделирования и проектирования в целом (использование единой платформы представления знаний для анализа «внешней» действительности и «внутреннего» устройства системы, по меньшей мере, целесообразно).

Конкурентоспособность разрабатываемых авторами инструментальных средств семантического моделирования и проектирования во многом обеспечивает именно представленный в статье методический и алгоритмический комплекс интеллектуального онтологического анализа данных, а его программное обеспечение имеет достаточные резервы развития.

Работа выполнена при проведении исследований по теме «Модели и методы формирования согласованной системы понятий о предметной области управления в процессах коллективного принятия решений» в рамках государственного задания Института проблем управления сложными системами РАН на 2016 год, а также при государственной поддержке Министерства образования и науки РФ в рамках реализации мероприятий Программы повышения конкурентоспособности Самарского государственного аэрокосмического университета среди ведущих мировых научно-образовательных центров на 2013-2020 годы.

Библиографический список

- [Аршинский, 2007] Аршинский, Л.В. Векторные логики: основания, концепции, модели / Л.В. Аршинский. - Иркутск: Иркут. гос. ун-т, 2007. - 228 с.
- [Барсегян и др.]. Анализ данных и процессов / А.А. Барсегян и др. - СПб.: БХВ-Петербург, 2009. - 512 с.
- [Гаврилова, 2008] Гаврилова, Т.А. Интеллектуальные технологии в менеджменте: инструменты и системы / Т.А. Гаврилова, Д.И. Муромцев. - СПб.: Изд-во «Высшая школа менеджмента»; Изд. дом СПбГУ, 2008. - 488 с.
- [Загоруйко, 2013] Загоруйко, Н.Г. Когнитивный анализ данных / Н.Г. Загоруйко. - Новосибирск: Академическое изд-во «Гео», 2013. - 186 с.
- [Офицеров, 2014] Офицеров, В.П. Метод альфа-сечения нестрогих формальных контекстов в анализе формальных понятий / В.П. Офицеров, В.С. Смирнов, С.В. Смирнов // Проблемы управления и моделирования в сложных системах: Труды XVI международной конф. (30 июня - 03 июля 2014 г., Самара, Россия). - Самара: СамНЦ РАН, 2014. - С. 228-244.
- [Семенова, 2015] Семенова, В.А. OntoWorker: программная лаборатория для онтологического анализа данных / В.А. Семенова, В.С. Смирнов, С.В. Смирнов // Проблемы управления и моделирования в сложных системах: Труды XVII международной конф. (22-25 июня 2015 г., Самара, Россия). - Самара: СамНЦ РАН, 2015. - С. 382-393.
- [Смирнов, 2011] Смирнов, С.В. Построение онтологий предметных областей со структурными отношениями на основе

анализа формальных понятий / С.В. Смирнов // Знания – Онтология – Теория: Материалы Всероссийской конф. с международным участием (3-5 октября 2011 г., Новосибирск, Россия). Т. 2. - Новосибирск: Институт математики СО РАН, 2011. - С. 103-112.

[Смирнов, 2015a] Смирнов, С.В. Опыт создания средств семантического моделирования и проектирования на массовой программной платформе / С.В. Смирнов // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2015): Материалы V международ. науч.-тех. конф. (19-21 февраля 2015 г. Минск, Беларусь) / Редкол.: В.В. Голенков (отв. ред.) и др. - Минск: БГУИР, 2015. - С. 413-416.

[Смирнов, 2015b] Смирнов, С.В. Нестрогий анализ формальных понятий / С.В. Смирнов // Знания – Онтология – Теория: Материалы Всероссийской конф. с международным участием (6-8 октября 2015 г., Новосибирск, Россия). Т. 2. - Новосибирск: Институт математики СО РАН, 2015. - С. 142-150.

[Ganter, 1999] Ganter, B. Formal Concept Analysis. Mathematical foundations / B. Ganter, R. Wille. - Berlin-Heidelberg: Springer-Verlag, 1999. - 290 p.

[Lammari, 2004] Lammari, N. Building and maintaining ontologies: a set of algorithms / N. Lammari, E. Metais // Data & Knowledge Engineering. - 2004. - Vol. 48(2). - P. 155-176.

[Yang et al., 2008] Yang, K. M. Fuzzy Concept Mining based on Formal Concept Analysis / K.M. Yanget al. // Int. J. of Computers. - 2008. - Issue 3. Vol. 2. - P. 279-290.

ALGORITHMS AND SOFTWARE FOR ONTOLOGICAL DATA ANALYSIS

Zubtcov R.O., Semyonova V.A.,
Smirnov S.V.

*Samara State Aerospace University,
Samara, Russia*

*zubtcov.r@gmail.com
queenbfjr@gmail.com*

*Institute for the Control of Complex Systems of the
Russian Academy of Sciences, Samara, Russia*

smirnov@iccs.ru

The paper describes the analytical kernel of developing semantic modeling and designing tools. This kernel is composed of mathematical models, methods and algorithms of ontological data analysis, which are designed to bring the formal ontologies based on empirical data provided in the form of matching “objects-properties”. Ontological Data Analysis is based on theoretically well-established Formal Concept Analysis method and has a number of new abilities. To be able to handle incomplete and inconsistent data was suggested to use more adequate multi-valued vector logic. Dependencies called “existence constraints” are taken into account when constructing a correct context for the Formal Concept Analysis.

In addition, perspectives of extension and development of software are estimated. Possibility and ways to create a web service are considered.