

УДК 004.21:004.82

МЕХАНИЗМ НОРМАЛИЗАЦИИ ЭМПИРИЧЕСКОГО КОНТЕКСТА В ОНТОЛОГИЧЕСКОМ АНАЛИЗЕ ДАННЫХ

В. А. СЕМЕНОВА¹, С. В. СМИРНОВ²

¹queenbfjr@gmail.com, ²smirnov@iccs.ru

Самарский федеральный исследовательский центр РАН,
Институт проблем управления сложными системами РАН

Поступила в редакцию 30 июля 2021 г.

Аннотация. Областью исследований является «онтологический анализ данных», который понимается авторами как анализ, в общем случае, неполных и противоречивых эмпирических данных о неизученной, неструктурированной предметной области с целью построения ее формальной онтологии. Предметом исследования является обработка, в общем случае, логически многозначного формального контекста для вывода формальных понятий на основе объектно-признаковых данных и априорных ограничений существования свойств у объектов исследуемой предметной области. Задача исследования состоит в разработке эффективных алгоритмов управления «нормальностью» формального контекста, извлекаемого из эмпирических данных. В исследовании использованы методы теории множеств и бинарных отношений, модели и методы анализа формальных понятий, а также методология применения ограничений существования для построения формальных онтологий. Отличие и новизна предложенного решения заключается в замене «естественного» описания системы измеряемых у объектов свойств, как множества с заданными на нем экзистенциональными отношениями, спецификацией этой системы в виде совокупности субструктур, однородных по виду экзистенционального сопряжения свойств-членов. Значение полученных результатов состоит в алгоритмическом обеспечении ряда этапов онтологического анализа данных.

Ключевые слова: анализ формальных понятий; ограничения существования свойств; онтология; онтологический анализ данных; алгоритмизация.

ВВЕДЕНИЕ

Результативная разработка и эффективная эксплуатация информационных систем в науке, технике и производстве возможна лишь при адекватном компьютерном представлении их предметной области (ПрО). Систематизация имеющихся в этом смысле данных и знаний с построением соответствующих информационных моделей составляет современное содержание онтологического подхода в вычислительных науках [1–3], который развивается в русле

парадигмы представления и использования знаний в искусственном интеллекте [1, 4, 5].

Одним из путей построения формальных онтологий в информатике является их «вывод» из доступных эмпирических данных о ПрО, или онтологический анализ данных (ОАД). Данные рассматриваются при этом как результат измерений (в самом широком понимании) свойств объектов актуальной слабоструктурированной ПрО и сводятся в стандартизованные таблицы «объекты-свойства» (ТОС) [6, 7]. Наиболее результа-

тивные методы этого направления опираются на анализ формальных понятий (АФП) [8–11].

В АФП, благодаря концептуальному шкалированию [12] и/или учету информационной полноты процедур измерения (всякое измерение свойства объекта может дать особый результат «None», свидетельствующий о нахождении значения измеряемого свойства за порогами чувствительности, вне динамического диапазона средства измерений [3] или вследствие семантического несоответствия объекта ПрО и измерительной процедуры [13]) ТОС, трансформируется в так называемый формальный контекст ОАД-задачи, где результаты измерений заменяются совокупностью оценок истинности базовых семантических суждений (БСС) вида «объект g обладает свойством x ». На практике вследствие неполноты, неточности, неопределенности и противоречивости данных о ПрО такие оценки логически многозначны, и обеспечение вывода понятийной структуры ПрО в подобной ситуации – репертуар нечеткого (нестроого) АФП [2, 14–19].

Менее известна методика отологического инжиниринга, основанная на концепции ограничений существования свойств (ОСС) [20–22]. Однако нами показана общность этих формализованных подходов к построению онтологий (в том смысле, что когнитивный потенциал рассматриваемых методик возникает благодаря использованию одних и тех же гипотетико-дедуктивных приемов) и обоснована целесообразность их совместного использования [23, 24].

В [25] предложены алгоритмы формирования и трансформации «естественного» описания системы измеряемых свойств объектов ПрО (т.е. множества свойств с заданными на нем ОСС) в набор субструктур – групп сопряженных свойств. Цель данной статьи – представить механизм того, каким образом такая реструктуризация может быть использована для эффективной коррекции бинарной аппроксимации нестроого формального контекста ОАД-задачи, в результате чего он становится пригодным для применения апробирован-

ных АФП-методов вывода понятийной структуры актуальной ПрО.

СИСТЕМА ИЗМЕРЯЕМЫХ СВОЙСТВ И ПОСТАНОВКА ЗАДАЧИ

Постулатом при изучении актуальной ПрО является утверждение о существовании множества составляющих ее объектов G . Однако «в поле зрения» субъекта обычно оказывается лишь конечное подмножество $G^* \subseteq G$, и человек никогда не использует объекты как таковые, а эксплуатирует их свойства.

Согласно [23, 24], определение а priori состава конечного множества M , подлежащих измерению свойств объектов G^* в когнитивном плане есть не что иное, как выдвижение гипотез об искомой понятийной структуре актуальной ПрО. При этом кроме очевидного комбинаторного синтеза гипотетических понятий могут использоваться еще два (и согласно классической логике [26, 27] только два) приема их формирования: деление и ограничение гипотетических понятий. Эти способы и задают ОСС – бинарные экзистенциальные отношения на множестве M :

– обусловленность $C: M \times M \rightarrow \{\mathbf{True}, \mathbf{False}\}$, когда наперед устанавливается, что обладая свойством $x \in M$, всякий объект $g \in G$ непременно обладает свойством $y \in M$ (хотя обратное может быть неверно);

– несовместимость $E: M \times M \rightarrow \{\mathbf{True}, \mathbf{False}\}$, когда предопределяется, что обладая свойством x , всякий объект $g \in G$ заведомо не обладает свойством y , и наоборот.

Отношение C рефлексивно, несимметрично, транзитивно и порождает во множестве свойств рефлексивное, симметричное и транзитивное отношение взаимобусловленности $H: H(x, y) \leftrightarrow C(x, y) \wedge C(y, x)$. Отношение E антирефлексивно, симметрично и нетранзитивно, но характеризуется так называемой «транзитивностью относительно обусловленности» [22], что означает $\forall x, y, z \in M: C(x, y) \& E(y, z) \rightarrow E(x, z)$.

Таким образом, а priori субъект формирует (M, C, E) – систему измеряемых свойств, далее СИС, естественным описанием которой служит граф, в котором множе-

ство M определяет вершины, а отношения C и E – дуги и ребра соответственно [25].

Система измеряемых свойств (M, C, E) гипотетически предопределяет, что всякий объект $g \in G$ может обладать лишь «нормальным» подмножеством множества измеряемых свойств $N \subseteq M$ [22]: N содержит все свойства, обусловленные любым его элементом ($\forall x \in N: (\exists y \in M: C(x, y)) \rightarrow y \in N$), и любые два элемента N не связаны отношением несовместимости ($\forall x \in N: (\exists y \in M: E(x, y)) \rightarrow y \notin N$).

Если в АФП исходные данные для вывода понятийной структуры актуальной ПрО представлены формальным контекстом (G^*, M, I) , где I – совокупность оценок БСС базовых семантических а posteriori, то в ОАД к числу исходных данных добавляется система (M, C, E) , которая используется при нормализации формального контекста. Принципиальное отличие и главное содержание этой задачи – выяснение нормально или нет некоторое определенное подмножество свойств, а не выявление всех нормальных подмножеств СИС как в [20, 21]. Рассмотрим эффективное решение задачи нормализации эмпирического формального контекста за счет предварительной реструктуризации системы (M, C, E) .

РЕСТРУКТУРИЗАЦИЯ СИСТЕМЫ ИЗМЕРЯЕМЫХ СВОЙСТВ

Анализ формального определения нормального подмножества свойств позволяет указать в СИС субструктуры, которые однородны по характеру экзистенционального сопряжения свойств-членов. Для каждого вида таких субструктур, во-первых, легко формулируется и доказывается условие нормальности, а, во-вторых, на основании оценки нормальности отдельных субструктур, открывается возможность судить о нормальности отдельных подмножеств измеряемых свойств.

Первая из характерных субструктур СИС - «ВЗО-группа», - есть замкнутое подмножество взаимообусловленных (ВЗО) свойств (с учетом имманентной самообусловленности каждого измеряемого свойства). Нетрудно видеть, что ВЗО-группа

может принадлежать нормальному подмножеству измеряемых свойств только целиком.

В субструктуре СИС «О-группа» одна группа ВЗО-свойств обуславливает другую группу ВЗО-свойств (отношение обусловленности между ВЗО-группами обобщает это отношение между свойствами в том смысле, что каждое свойство-член одной группы обуславливает все свойства-члены другой; формальный анализ этого положения можно найти в [22]). Легко доказать, что О-группа либо входит в нормальное подмножество целиком, либо своей обуславливаемой частью.

Субструктуру СИС «Н-группа» образуют попарно несовместимые измеряемые свойства. Очевидно, что Н-группа может быть представлена в нормальном подмножестве только одним своим членом.

Ключевым для констатации «нормальности» любого подмножества СИС является утверждение, что подмножество СИС нормально тогда и только тогда, когда для него и пересекающихся с ним ВЗО-, О- и Н-групп справедливы вышеприведенные утверждения.

Таким образом, для нормализации эмпирического контекста актуальна реструктуризация СИС в набор субструктур – ВЗО-, О- и Н-групп.

Выявление групп взаимообусловленных свойств. Во-первых, каждое отдельное измеряемое свойство потенциально следует считать эмбрионом самостоятельной ВЗО-группы. Во-вторых, ясно, что признаком ВЗО-группы является не только взаимообусловленность входящих в нее свойств, но и отсутствие пересечения с другими ВЗО-группами (исключая, разумеется, некоторые ВЗО-группы, состоящие из одного свойства). Поэтому после выявления очередной ВЗО-группы ее свойства-члены следует исключить из дальнейшего анализа, что сокращает трудоемкость выявления каждой следующей такой группы. Детально алгоритм выявления ВЗО-групп представлен в [25].

Выявление групп с односторонней обусловленностью. Поскольку по определению О-группа – это две ВЗО-группы, одна из которых обуславливает другую, то алгоритм

выделения О-групп строится в предположении, что все ВЗО-группы в СИС уже выявлены. Детально алгоритм выявления О-групп представлен в [25].

Выявление групп попарно несовместимых свойств. Проблемой выявления Н-групп в СИС является не только их пересечения между собой, но, прежде всего, вложенности одних таких групп в другие. Как правило, подобные структуры эффективно обрабатываются лишь путем рекурсии. Далее приводится детальное описание порядка действий по выявлению Н-групп, уточняющее приведенные в [25] алгоритмы (в них $E(x)$ – подмножество измеряемых свойств, несовместимых со свойством x , $0 \leq |E(x)| \leq |M| - 1$).

Алгоритм «Выявление Н-групп»

Input: M – множество измеряемых свойств;
 E – отношение несовместимости на множестве M

Output: Es – множество Н-групп свойств в M

001 $Es := \emptyset$

002 **for each** x **in** M

003 $Grs :=$

Н-группы, включающие x и подмножества E_x

' x и E_x – аргументы функции, и
' $E_x = E(x)$; функция доставляет
'множество Н-групп, каждая из
'которых включает свойство x и
'определенное подмножество
'попарно несовместимых с ним
'свойств из $E(x)$

004 $Es := Es \cup Grs$

'бесповторное (по составу свойств)
'накапливание Н-групп

005 **next** x

Алгоритм рекурсивной функции «Н-группы, включающие x и подмножества E_x » выявляет как все пересекающиеся, так и вложенные одна в другую Н-группы, включающие в себя свойство x .

Алгоритм

«Н-группы, включающие x и подмножества E_x »

Input: M – множество измеряемых свойств;
 $x \in M$ – измеряемое свойство;
 E – отношение несовместимости

на множестве M ;

$E(x)$ – свойства, несовместимые с x

Output: Grs – искомое множество Н-групп

001 $Grs := \emptyset$

002 **for each** y **in** $E(x)$

003 $M^* := E(y) \cap E(x)$

'свойства, несовместимые как со
'свойством y , так и со свойством x

004 **if** $|M^*| \leq 1$ **then**

005 $Gr := \text{new } E$ 'E – класс «Н-группа»

006 $Gr := Gr \cup \{y\}$

007 **if** $|M^*| = 1$ **then** $Gr := Gr \cup M^*$

008 **else**

009 $Gr :=$

Н-группы, включающие x и подмножества E_x

' $x = y$, $E_x = M^*$; рекурсия для
'свойства y и M^*

010 **end if**

011 $Grs := Grs \cup Gr$

'бесповторное накопление
'искомых Н-групп

012 **next** y

013 **for each** Gr **in** Grs

014 $Gr := Gr \cup \{x\}$

015 **next** Gr

016 *Н-группы, включающие x и*
подмножества $E_x := Grs$

О МЕХАНИЗМЕ НОРМАЛИЗАЦИИ ЭМПИРИЧЕСКОГО ФОРМАЛЬНОГО КОНТЕКСТА

Если в эмпирическом формальном контексте (G^*, M, I) соответствие «объекты-свойства» I однозначно, т.е. $\forall g \in G^*, \forall x \in M: I(g, x) \in \{\text{True}, \text{False}\}$, то суть решения задачи нормализации – получение одного из следующих ответов:

– эмпирически установленные множества свойств объектов нормальны. Это означает, что результаты измерений не противоречат априорным гипотезам о понятийной структуре актуальной ПрО, и вывод формальных понятий из (G^*, M, I) есть выявление тех из этих гипотез, которые нашли подтверждение a posteriori;

– в формальном контексте имеются объекты, множества свойств которых не являются нормальными. Это означает, что эмпирический материал противоречит выдвинутым гипотезам о понятийной структуре актуальной ПрО, и субъект, выполняющий анализ, самостоятельно принимает решение о целесообразности вывода формальных понятий, выходящих за пределы его априорных представлений о ПрО.

Если эмпирическое соответствие в формальном контексте логически многозначно,

то возможен еще один вариант действий: принудительная нормализация эмпирически установленного множества свойств каждого объекта в формальном контексте за счет надлежущего исключения тех свойств-членов этого множества, БСС которых применительно к данному объекту наименее достоверны [28, 29]. Однако результат здесь не гарантирован, т.к. может обнаружиться неустранимое противоречие априорных гипотез и эмпирических данных, когда нормализация не может быть завершена без отказа от абсолютно достоверного эмпирического материала.

Заметим, что при работе с логически многозначным соответствием «объекты-свойства» оценки БСС, соответствующие неопределенности, лжи, противоречию, их промежуточным значениям (например, в нечеткой логике это лишь 0; в векторной V^{TF} -логике, являющейся простейшим обобщением нечеткой [30], это $\langle 0; 0 \rangle$ – «неопределенность», $\langle 0; 1 \rangle$ – «ложь», $\langle 1; 1 \rangle$ – «противоречие» и, если $b \in [0, 1)$, то $\langle 0; b \rangle$ – «неопределенность с долей лжи», $\langle b; 1 \rangle$ – «ложь с долей правды») интерпретируются как **False**, а остальные как **True**. Кроме того, субъект может установить собственный порог доверия, обращаясь в **False** незначимые по его мнению величины эмпирических оценок истинности БСС.

Опорный алгоритм нормализации устанавливает для каждого отдельно взятого объекта $g \in G^*$ статус любой субструктурной группы свойств в СИС. Группа получает статус «0» либо при отсутствии пересечения с множеством свойств $M^* \in M$ объекта g , либо при выполнении применительно к M^* условия утверждения 1, 2 или 3 (сообразно виду группы). Иначе группе присваивается статус «1» («проблемная группа»). Статус «1» переходит в «2» («критичная группа»), если группе невозможно вернуть статус «0» за счет надлежущего исключения свойств-членов этой группы, БСС которых применительно к объекту g наименее достоверны.

Алгоритм «Определение статуса st субструктурной группы свойств Gr для объекта g »

```

Input:  ( $G^*$ ,  $M$ ,  $I$ ) – формальный контекст
           $g \in G^*$  – объект измерения;
           $Gr$  – субструктурная группа свойств
          в системе ( $M$ ,  $C$ ,  $E$ );
           $MCs$ ,  $Cs$ ,  $Es$  – множество ВЗ0-, 0- и
          Н-групп свойств во множестве  $M$ 
Output:  $st$  – статус группы  $Gr$  для объекта
           $g$ ,  $st \in \{0, 1, 2\}$ 
001   $st := 0$  'исходное предположение
          'о статусе  $Gr$ 
002   $M^* := \{g\}$  'множество свойств объекта
          '  $g$  (« $\cdot$ » – оператор Галуа)
003  if  $Gr \in MCs$  And  $|Gr| > 1$  then
          ' $Gr$  – ВЗ0-группа с несколькими членами
004       $IS := |Gr \cap M^*|$  'множество свойств
          'объекта  $g$ , входящих
          'в ВЗ0-группу  $Gr$ 
005      if  $0 < |IS| < |Gr|$  then
          ' $Gr$  – проблемная ВЗ0-группа, т.к.
          'она пересекается с  $M^*$ , но не
          'вложена в него
006           $st :=$  Ненулевой_статус ( $g$ ,  $IS$ )
          ' $Gr$  – проблемная или критичная
          'группа
007      end if
008  elseif  $Gr \in Cs$  then ' $Gr$  – 0-группа
009       $IS := |Gr.C_0 \cap M^*|$  'множество
          'свойств объекта  $g$ , входящих в
          'обусловливаемую ВЗ0-группу
          '0-группы  $Gr$  ( $Gr.C_0$  суть эта
          'ВЗ0-группа)
010      if  $|IS| = 0$  then
011           $IS := |Gr.C_x \cap M^*|$  'множество
          'свойств объекта  $g$ , входящих в
          'обусловливающую ВЗ0-группу
          '0-группы  $Gr$  ( $Gr.C_x$  суть эта
          'ВЗ0-группа)
012          if  $|IS| = |Gr.C_x|$  then
          ' $Gr$  проблемная 0-группа, т.к. её
          'обусловливающая ВЗ0-группа
          'вложена в неё, а её
          'обусловливаемая ВЗ0-группа
          'с ней не пересекается
013               $st :=$  Ненулевой_статус ( $g$ ,  $IS$ )
              ' $Gr$  – проблемная или критичная
              'группа
014          end if
015      end if

```

```

016 elseif  $Gr \in Es$  then ' $Gr$  -  $H$ -группа
017    $IS := |Gr \cap M^*|$    'множество
      'свойств объекта  $g$ , входящих
      'в  $H$ -группу  $Gr$ 
018    $k := |IS|$ 
019   if  $k > 1$  then
      'мощность  $Gr \cap M$  больше 1
020      $st := 1$ 
      ' $Gr$  - проблемная  $H$ -группа
021     for each  $x$  in  $IS$ 
      'анализ вариантов исключения
      'свойств из  $IS$ 
022     if
Ненулевой_статус ( $g, IS \setminus \{x\}$ ) = 2 then
023        $k := k - 1$ 
024     end if
025     next  $x$ 
026   end if
027   if  $k = 0$  then  $st := 2$ 
      ' $Gr$  - критичная  $H$ -группа
028 end if

```

Алгоритм функции «Ненулевой_статус» (g, IS) выявляет возможность исключения множества свойств IS из M^* – множества свойств объекта g , за счет локального – для объекта g и его проблемной группы свойств – ужесточения порога доверия к эмпирическим данным. Очевидно, что для логически однозначного формального контекста функция всегда будет доставлять значение «2» (рассматриваемая проблемная группа свойств «критична» для объекта g). В случае логической многозначности эмпирического соответствия «объекты-свойства» такой результат будет получен лишь тогда, когда исключение IS из M^* потребует исключения из формального контекста абсолютно достоверных оценок БСС.

Выявление для $g \in G^*$ статуса каждой субструктурной группы свойств в СИС позволяет с учетом утверждения 4 дать следующее обоснованное заключение:

Утверждение 5. Множество свойств объекта g нормально тогда и только тогда, если по отношению к нему все субструктурные группы свойств имеют статус «0».

Что касается принудительной нормализации логически многозначного эмпирического контекста, то соответствующий алгоритм в цикле по всем объектам $g \in G^*$ и цикле по всем субструктурным группам в целом воспроизводит рассмотренный

опорный алгоритм нормализации и дополняется:

– в случае пороговых методов вывода четких и нечетких понятий [2, 16, 18] апробирован метод рационального выбора варианта исключения подмножества свойств объекта g до тех пор, пока множество свойств не станет нормальным, либо до обнаружения упоминавшегося выше неразрешимого противоречия между априорными гипотезами и эмпирическими данными о ПрО [2, 29];

– при использовании оператора нечеткого замыкания для вывода нечетких понятий [14, 16] может быть применен метод фильтрации декомпозиции нечеткого соответствия «объекты-свойства» с исключением четких слагаемых композиции, не являющимися нормальными множествами [28].

ЗАКЛЮЧЕНИЕ

– Трансформация исходного «естественного» представления СИС в набор субструктур, однородных по виду экзистенциального сопряжения свойств-членов, обеспечивает возможность эффективного практического использования априорных предположений и знаний субъекта онтологического анализа эмпирических данных. Основанием для такого вывода является возможность прямого соотнесения внутреннего устройства названных субструктур с прагматически важным и теоретически строгим понятием нормального подмножества измеряемых свойств. В целом открывается путь эффективного решения ранее не рассматривавшейся, важной в теоретическом и практическом плане проблемы нормализации эмпирических формальных контекстов.

– Очерченный в статье механизм нормализации формальных контекстов реализован в разрабатываемой в Институте проблем управления сложными системами РАН системе онтологического анализа данных на массовой программной платформе и апробирован при онтологическом анализе данных в различных прикладных областях.

СПИСОК ЛИТЕРАТУРЫ

1. Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И. Инженерия знаний. Модели и методы. СПб.: Лань, 2016.

324 с. [Т. А. Gavrilova, D. V. Kudryavtsev, D. I. Muromtsev, *Engineering knowledge. Models and Methods*, (in Russian). St. Petersburg: Lan, 2016.]

2. **Oficerov V. P., Smirnov S. V.** Fuzzy Formal Concept Analysis in the construction of ontologies // *Ontology of designing*. 2017. Vol. 7, no. 4. Pp. 487-495.

3. **Смирнов С. В.** Онтологический анализ предметных областей моделирования // *Известия Самарского научного центра РАН*. 2001. Т. 3, № 1. С. 62–70. [S. V. Smirnov, "Ontological analysis of modeling domains", (in Russian), in *Izvestiya Samarskogo nauchnogo centra RAN*, vol. 3, no. 1, pp. 62-70, 2001.]

4. **Искусственный интеллект**. В 3-х кн. Кн. 2. Модели и методы. Справочник / под ред. Д. А. Поспелова. М.: Радио и связь, 1990. 304 с. [D. A. Pospelova (ed.), *Artificial intelligence. In 3 books. Book. 2. Models and methods. Directory*, (in Russian). Moscow: Radio i svyaz, 1990.]

5. **Russel S. J., Norvig P.** Artificial Intelligence. A modern approach. New Jersey: Prentice Hall Press, 2003. 1152 p.

6. **Анализ данных и процессов** / А. А. Барсегян [и др.]. СПб.: БХВ-Петербург, 2009. 512 с. [A. A. Barseghyan, et al., *Data and process analysis*, (in Russian). St. Petersburg: BXV-Peterburg, 2009.]

7. **Загоруйко Н. Г.** Когнитивный анализ данных. Новосибирск: Академическое изд-во «Гео», 2013. 186 с. [N. G. Zagoruiko, *Cognitive data analysis*, (in Russian). Novosibirsk: Akademicheskoe izd-vo "Geo", 2013.]

8. **Formal Concept Analysis: From Knowledge Discovery to Knowledge Processing** / S. Ferré, et al. // *A Guided Tour of Artificial Intelligence Research*. Vol. II: AI Algorithms / P. Marquis, O. Papini, H. Prade (eds.). Springer Int. Publishing, 2020. Pp. 411-445.

9. **Obitko M., Snasel V., Smid J.** Ontology Design with Formal Concept Analysis / V. Snasel, R. Belohlavek (eds.) // *Proc. of the Int. Workshop on Concept Lattices and their Applications*. VŠB-Technical University of Ostrava, Dept. of Computer Science, 2004. Pp. 111-119.

10. **Priya M., Kumar Ch. A.** A Survey of State of the Art of Ontology Construction and Merging using Formal Concept Analysis // *Indian J. of Science and Technology*. 2015. Vol. 8, no. 24. Pp. 1-7.

11. **Sertkaya B.** A survey on how description logic ontologies benefit from FCA // *Proc. of the 7th Int. Conf. "Concept Lattices and Their Applications"*. University of Sevilla, 2010. Pp. 2-21.

12. **Ganter B., Wille R.** Conceptual scaling. Applications of Combinatorics and Graph Theory to the Biological and Social Sciences / F. Roberts (ed.). New York: Springer-Verlag, 1989. Pp. 139-167.

13. **Рубашкин В. Ш.** Представление и анализ смысла в интеллектуальных информационных системах. М.: Наука, 1989. 192 с. [V. Sh. Rubashkin, *Representation and analysis of meaning in intelligent information systems*, (in Russian). Moscow: Nauka, 1989.]

14. **Computing** the lattice of all fixpoints of a fuzzy closure operator / R. Belohlavek, et al. // *IEEE Trans. on Fuzzy systems*. 2010. Vol. 18, no. 3. Pp. 546-557.

15. **Unifying** fuzzy concept lattice construction methods / S. Boffa, et al. // *Proc. of IEEE Int. Conf. on Fuzzy Systems*. 2016. Pp. 209-216.

16. **Cross V., Kandasamy M.** Creating Fuzzy Concepts: The One-Sided Threshold, Fuzzy Closure and Factor Analysis

Methods // Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. 2011. Vol. 6743. Pp. 127-134.

17. **Fuzzy** and rough formal concept analysis: a survey / J. Poelmans, et al. // *Int. J. of General Systems*. 2014. Vol. 43, no. 2. Pp. 105-134.

18. **Fuzzy** Concept Mining based on Formal Concept Analysis / K. M. Yang, et al. // *Int. J. of Computers*. 2008. Vol. 2, no. 3. Pp. 279-290.

19. **Zhang Z.** Constructing L-fuzzy concept lattices without fuzzy Galois closure operation // *Fuzzy Sets and Systems*. 2018. Vol. 333. Pp. 71-86.

20. **Lammari N., Metais E.** Building and maintaining ontologies: a set of algorithms // *Data & Knowledge Engineering*. 2004. Vol. 48, no. 2. Pp. 155-176.

21. **Lammari N., du Mouza C., Metais E.** POEM: an Ontology Manager based on Existence Constraints / S. S. Bhowmick, J. Küng, R. Wagner (eds.) // *Proc. of the 19th Int. Conf. DEXA "Database and Expert Systems Applications"*. LNCS. 2008. Vol. 5181. Pp. 81-88.

22. **Пронина В. А., Шипилина Л. Б.** Использование отношений между атрибутами для построения онтологии предметной области // *Проблемы управления*. 2009. № 1. С. 27–32. [V. A. Pronina, L. B. Shipilina, "Use of relations between attributes for constructing a domain ontology", (in Russian), in *Problemy upravleniya*, no. 1, pp. 27-32, 2009.]

23. **Samoylov D. E., Semenova V. A., Smirnov S. V.** Multi-level recursive model of properties existence constraints in machine learning // *J. of Physics: Conf. Series*. 2018. Vol. 1096. Article number 012096.

24. **Смирнов С. В.** Две методологии вывода формальных понятий: когда и как они должны работать вместе // *Материалы VII международной конференции «Знания – Онтология – Теория» (ЗОНТ-2019)*. Новосибирск: Институт математики СО РАН, Новосибирский государственный ун-т, 2019. С. 355–363. [S. V. Smirnov, "Two methodologies for the derivation of formal concepts: when and how they should work together", (in Russian), in *Materials of the VII international conference "Knowledge - Ontology - Theories" (ZONT-2019)*, 2019, pp. 355-363.]

25. **Семенова В.А., Смирнов С.В.** Алгоритмизация формирования и прагматической трансформации ограничений существования свойств предметной области // *Онтология проектирования*. 2020. Т. 10, № 3 (37). С. 361–379. [V. A. Semenova, S. V. Smirnov, "Algorithmization of the formation and pragmatic transformation of restrictions on the existence of properties of the subject area", (in Russian), in *Ontologiya proektirovaniya*, vol. 10, no. 3 (37), pp. 361-379, 2020.]

26. **Гетманова А. Д.** Логика. Углубленный курс. М.: КНОРУС, 2016. 192 с. [A. D. Getmanova, *Logic. Advanced course*, (in Russian). Moscow: KNORUS, 2016.]

27. **Ивин А. А., Никифоров А. Л.** Словарь по логике. М.: Гуманит. изд. центр ВЛАДОС, 1997. 384 с. [A. A. Ivin, A. L. Nikiforov, *Dictionary of Logic*, (in Russian). Moscow: Gumanit. izd. centr VLADOS, 1997.]

28. **Samoilov A. E., Smirnov S. V.** Ways to accept the properties existence constraints in fuzzy formal concept analysis / V. Fursov, Y. Goshin, D. Kud-ryashov (eds.) // *CEUR Workshop Proceedings of the 6th Int. Conf. Information Technology and Nanotechnology. Session Data Science*. 2020. Vol. 2667. Pp. 269-272.

29. **Самойлов Д. Е., Семенова В. А., Смирнов С. В.** Анализ неполных данных в задачах построения формальных онтологий // *Онтология проектирования*. 2016. Т. 6, № 3 (21). С. 317–339. [D. E. Samoilov, V. A. Semenova, S. V. Smirnov, "Incomplete data analysis for building formal ontologies", (in Russian), in *Ontologiya proektirovaniya*, vol. 6, no. 3 (21), pp. 317-339, 2016.]

30. **Аршинский Л. В.** Векторные логики: основания, концепции, модели. Иркутск: Иркутский государственный ун-т, 2007. 228 с. [L. V. Arshinsky, *Vector Logics: Foundations, Concepts, Models*, (in Russian). Irkutsk: Irkutskij gosudarstvennyj un-t, 2007.]

ОБ АВТОРАХ

СЕМЕНОВА Валентина Андреевна, мл. науч. сотр., асп. каф. прикладной математики и информатики. М-р механики и мат. моделирования (Самарский ун-т, 2017). Готовит дисс. в обл. интеллектуального анализа данных.

СМИРНОВ Сергей Викторович, гл. науч. сотр, проф. каф. программного обеспечения и управления в технических системах. Дипл. инженер-математик (Куйбышевский авиационный институт, 1975). Д-р техн. наук по системному анализу, управлению и обработке информатики (защ. в СамГТУ, 2002). Иссл. в обл. прикладной математики, компьютерного моделирования сложных систем и создания интеллектуальных систем поддержки принятия решений.

METADATA

Title: Normalization of Empirical Context in Ontological Data Analysis.

Authors: V. A. Semenova¹, S. V. Smirnov²

Affiliation: Samara Federal Research Scientific Center RAS, Institute for the Control of Complex Systems RAS, Samara, Russia

Email: ¹ queenbfjr@gmail.com, ² smirnov@iccs.ru

Language: Russian.

Source: SIIT (scientific journal of Ufa State Aviation Technical University), vol. 3, no. 3 (7), pp. 45-52, 2021. ISSN 2686-7044 (Online), ISSN 2658-5014 (Print).

Abstract: The research field is "ontological data analysis", which is understood by the authors as an analysis of generally incomplete and inconsistent empirical data on an unexplored, unstructured subject domain in order to build its formal ontology. The subject of the research is the processing in the general case of a logically multivalued formal context for the derivation of formal concepts based on object-property data and a priori properties existence constraints at objects of the studied subject domain. Algorithms for controlling the "normality" of the formal context extracted from empirical data are being developed. The research uses the methods of the theory of sets and binary relations, models and methods of formal concept analysis, as well as the methodology for applying the existence constraints to construct formal ontologies. The difference and novelty of the proposed solution consists in replacing the "natural" description of the system of properties measured at objects as a set with existential relations given on it by specification of this system as a set of substructures, which are homogeneous in the form of existential relation of member properties. The significance of

the obtained results lies in the algorithmic support of a number of stages of ontological data analysis.

Key words: ontological data analysis; formal concept analysis; properties existence constraints; normal set of properties; algorithmization.

About authors:

SEMENOVA, Valentina Andreevna, researcher, Postgrad. Student.

SMIRNOV, Sergei Viktorovich, researcher, Prof., Dr. of Tech. Sci.