

## ФОРМИРОВАНИЕ КОНТЕКСТА ДЛЯ ВЫВОДА ФОРМАЛЬНЫХ ПОНЯТИЙ ИЗ НЕПОЛНЫХ И ПРОТИВОРЕЧИВЫХ ДАННЫХ

В. А. СЕМЕНОВА

**Аннотация.** В эпоху больших данных актуальность системного анализа данных возрастает, так как он помогает организациям эффективно обрабатывать и интерпретировать огромные объемы информации для достижения стратегических целей. Но зачастую данные об интересующей предметной области, требующие системного осмысления и извлечения из них полезной информации, характеризуются как неполные, расплывчатые и даже противоречивые. Для обработки подобного рода информации нередко применяется анализ формальных понятий, в частности, нечёткий. Однако иногда требуется учитывать и дополнительные априорные сведения исследователя об изучаемой предметной области, называемые ограничениями существования свойств. В данной статье ставится задача учёта ограничений существования свойств при выводе формальных понятий из неполных и противоречивых данных. Решить данную задачу помогает обращение к способам образования понятий в классической логике. В статье предлагается эвристический метод нормализации нестроого формального контекста с учётом ограничений существования свойств при выводе формальных понятий из неполных и противоречивых данных. В заключение приводится схема методики построения однозначного формального контекста из неполных и противоречивых данных с учетом ограничений существования свойств.

**Ключевые слова:** эмпирические данные; предметная область; формальный контекст; базовое семантическое суждение; ограничения существования свойств.

### ВВЕДЕНИЕ

Анализ формальных понятий [1, 2] является важным направлением в области интеллектуального анализа данных, применяемым в таких сферах, как медицина, психология, архивное дело, информатика, реинжиниринг программного обеспечения и других [3]. Этот анализ основывается на математической теории порядка и позволяет извлекать понятия, а также строить концептуальную иерархию из исходных данных, представленных в виде формального контекста.

При анализе формальных понятий отсутствуют определенные предположения о появлении объектов в рамках задачи. Тем не менее практика [4] демонстрирует, что такие предположения могут быть заложены в априорных знаниях исследователя о предметной области. Игнорирование этих знаний о взаимосвязях между измеряемыми свойствами объектов может привести к тому, что полученные формальные понятия не будут отражать действительность и окажутся некорректными. В статьях [5, 6] рассматриваются вопросы ограничений существования свойств и их важности для корректного вывода формальных понятий. Однако предлагается лишь рекомендация проверять свойства объектов формального контекста на соответствие этим ограничениям.

В данной статье предложен метод построения формального контекста, который будет соответствовать априорным представлениям исследователя и к которому можно будет применять известные методы анализа формальных понятий. Таким образом, задача получения однозначного формального контекста из данных измерения свойств объектов предметной области с учетом ограничений существования свойств становится самостоятельной проблемой.

## ТЕОРИЯ

В анализе формальных понятий используются следующие определения и модели:

- $K = (G, M, I)$  – формальный контекст, где  $G = \{g_i\}_{i=1, \dots, r}$ ;  $r = |G| \geq 1$  – набор объектов исследуемой предметной области, попавших в поле зрения исследователя;  $M = \{m_j\}_{j=1, \dots, s}$ ;  $s = |M| \geq 1$  – множество измеряемых у объектов свойств;  $I$  – бинарное соответствие «объекты–свойства», то есть совокупность оценок истинности базовых семантических суждений о предметной области  $b_{ij} =$  «объект  $g_i$  обладает свойством  $m_j$ »:  $\|b_{ij}\| \in \{\text{Истина, Ложь}\}$ ;

- операторы Галуа  $\varphi, \omega$  (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  $Y_1 \supseteq Y_2$ .

Данные для формального контекста основаны на результатах измерений, которые фиксируются в стандартной таблице «объекты–свойства» в виде базовых семантических суждений [7].

Мягкий контекст, используемый в рамках более распространенного нечеткого или нестроого анализа формальных понятий, формируется вследствие следующих реалий накопления эмпирических данных:

- повторяющиеся независимые измерения свойств у объекта из интересующей выборки. То есть усматривается, что каждый объект  $g_i \in G$  подвергается, вообще говоря, нескольким сериям измерений с помощью всего арсенала измерительных процедур, используемых для исследования предметной области, причем всякая серия имеет ограниченную достоверность;

- оценивание одного и того же свойства объекта с помощью нескольких независимых источников. То есть учитывается, что каждое актуальное для исследователя свойство  $m_j \in M$  измеряется, вообще говоря, с помощью множества измерительных процедур (конгруэнтных по назначению), причем всякая процедура характеризуется степенью доверия к ее результатам;

- безусловная результативность выполнения измерительной процедуры. При этом учитывается, что всякий акт измерения, кроме ожидаемого результата из динамического диапазона измерительной процедуры (собираательно всякий такой результат у любой процедуры измерения можно обозначить как «**X**»), может дать «**None**» (свидетельствует о нахождении значения измеряемого свойства за порогами чувствительности, вне динамического диапазона средства измерений, о «семантическом несоответствии» исследуемого объекта и измерительной процедуры и др. [8]) и «**Failure**» (фиксирует невыполнение задачи измерений: отказ, сбой измерительного средства, воздержание при голосовании и т. п.). Также, для сохранения двумерной структуры обобщенной таблицы «объекты–свойства» была введена лингвоконстанта «**NM**», означающая, что в данной серии данной процедурой измерение не проводилось [9].

Существуют различные способы обработки подобных эмпирических данных, как правило, в рамках нечеткого анализа формальных понятий, когда подходы различаются по структуре степеней истинности и сопряженным операциям, способам определения нечеткого формального контекста, обобщения соответствий Галуа и определения нечеткого формального понятия [10]. Но также обработка возможна путём применения метода классического анализа формальных понятий, когда контекст из нечеткого/нестроого предварительно приводится в бинарный вид.

Однако помимо неполных эмпирических данных, эксперт-исследователь может обладать также и априорными знаниями об экзистенциальных зависимостях между свойствами, измеряемыми у объектов исследуемой предметной области, которые именуются *ограничениями существования свойств*.

Общие модели характерных экзистенциальных зависимостей между свойствами предложены в [5] в форме бинарных отношений «ограничений существования». В частности, пара свойств  $m_j, m_k \in M$ ,  $j \neq k$  для любого объекта предметной области (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, то есть  $E(m_j, m_k) \leftrightarrow \forall g_i \in G^*: m_j \in \{g_i\}' \rightarrow m_k \notin \{g_i\}'$ ;
- обусловленной, если, обладая свойством  $m_j$ , объект  $g_i$  непременно обладает свойством  $m_k$  (хотя обратное может быть неверно), то есть  $C(m_j, m_k) \leftrightarrow \forall g_i \in G^*: m_j \in \{g_i\}' \rightarrow m_k \in \{g_i\}'$ .

### Цель

Необходимо на основе неполных и противоречивых данных построить однозначный формальный контекст, в котором будут учтены ограничения, связанные с существованием свойств. В работе [11] предложен метод, основанный на теореме декомпозиции нечётких множеств, но дающий в результате нечёткие формальные понятия, количество которых, как правило, является чрезмерным для извлечения полезной информации.

Для обработки подобной эмпирической информации требуется подобрать релевантную многозначную логику, которая обеспечит отсутствие отсеивания или искажения данных как при интерпретации, так и при консолидации [12]. В качестве подходящей логики была выбрана векторная логика  $V^{TF}$ , предложенная Аршинским [13]. В этой логике истинность базового семантического суждения представляется вектором  $\mathbf{b}_{ij} = \langle b_{ij}^+, b_{ij}^- \rangle$ ,  $b_{ij}^+, b_{ij}^- \in [0, 1]$ . Компонент  $b_{ij}^+$  формируется на основе показаний, подтверждающих суждение, а  $b_{ij}^-$  — на основе показаний, его опровергающих.

Для интерпретации первичных экспериментальных данных в выбранной векторной логике для константы **NM** установлена истинностная константа «*Неопределенность*»  $\{0, 0\}$  (так как при отсутствии измерения нет свидетельств ни в пользу Истины, ни в пользу Лжи), для константы **Failure** потребовалось введение значения истинности «*Амбивалентность*»  $\{0.5, 0.5\}$ , которая имеет аналоги в прагматически ориентированных многозначных логиках (см., например, значение истинности «круглый информационный ноль» в [14]), а для символов из динамических диапазонов измерительных процедур и константы **None** — истинностные константы «*Истина*»  $\{1, 0\}$  и «*Ложь*»  $\{0, 1\}$  соответственно.

Учет достоверности серии и степени доверия к процедуре измерения осуществляется через операцию «00 композиции»  $V^{TF}$ -логики на основе композиционного умножения по  $t$ -норме  $x \cdot y = xy$ . Дальнейшая консолидация всех базовых семантических суждений, истинность которых оценена в каждой выполненной для  $i$ -го объекта серии измерений каждой процедурой измерения  $j$ -го свойства, осуществляется по схеме «усиления–усреднения». Это частный случай «11 композиционного совмещения»  $V^{TF}$ -логики на основе композиционного сложения по  $s$ -норме  $x \oplus y = \min(1, x + y)$  с весами для совмещаемых оценок разных аспектов истинности, равными  $1/n_c$ , где  $n_c$  — число совмещаемых оценок. В результате получается исходный формальный контекст с нестрогим соответствием «объекты–свойства», который необходимо дефазифицировать.

Для дефазификации исходного нестрогого формального контекста предложена двухступенчатая схема: на первом этапе выявляются базовые семантические суждения, которые являются ложными при любом допустимом пороге доверия к эмпирической информации. Второй этап может быть повторен многократно с изменением субъективного порога доверия  $\alpha = \langle \alpha^+, \alpha^- \rangle$ .

## Методы

Природа ограничений существования свойств приводит к необходимости соответствия содержаний выводимых формальных понятий следующим условиям (составляющим определение нормального подмножества некоторого множества):

- интенционал понятия должен быть замкнут ( $Y \subseteq M$  замкнуто, если  $\forall m_j \in Y: (\exists m_k \in M, m_k \neq m_j: C(m_j, m_k)) \rightarrow m_k \in Y$ );
- интенционал понятия должен быть совместим ( $Y \subseteq M$  совместимо, если  $\forall m_j \in Y: (\exists m_k \in M, m_k \neq m_j: E(m_j, m_k)) \rightarrow m_k \notin Y$ ).

Таким образом, способом учета ограничений существования свойств является нормализация наборов свойств у каждого отдельно взятого объекта в формируемом однозначном формальном контексте. Предложено решение этой задачи на основе эвристики [15], согласно которой разумным подходом к нормализации эмпирически и субъективно выявленного набора свойств  $M_g^* \subseteq M$  объекта  $g \in G$  в рабочем формальном контексте будет денонсация базовых семантических суждений, которые описываются как истинные, но нарушают «нормальность» множества свойств объекта [16, 17]. При этом из нескольких таких вариантов денонсации базовых семантических суждений следует выбрать тот, что обеспечивает минимум агрегированного показателя достоверности денонсируемых базовых семантических суждений.

Для денонсации базового семантического суждения с оценкой истинности  $\mathbf{b}_{ij} = \langle b_{ij}^+, b_{ij}^- \rangle$ , которое в формальном контексте при субъективном пороге доверия  $\alpha = \langle \alpha^+, \alpha^- \rangle$  признано достоверным, достаточно установить порог доверия к эмпирическим данным, определяемый суммой вектора  $\alpha$  и вектора ужесточения порога  $\mathbf{T}_{TV}$  при условии

$$\alpha + \mathbf{T}_{TV} = \langle b_{ij}^+ + \varepsilon^+, b_{ij}^- - \varepsilon^- \rangle,$$

где сколь угодно малые  $\varepsilon^+ \geq 0$  и  $\varepsilon^- \geq 0$  одновременно не равны нулю. Полагая правую часть формулы практически равной  $b_{ij}$ , обнаруживаем, что длина вектора  $\mathbf{T}_{TV}$  в метрике L1 сколь угодно близка к достоверности денонсируемого базового семантического суждения, которая в логике  $V^{TF}$  определяется величиной  $(b_{ij}^+ - b_{ij}^-) \in [-1, 1]$  и устанавливает агрегированный показатель достоверности в случае денонсации отдельно взятого базового семантического суждения.

Исследована ситуация, когда для реализации сформулированной эвристики применительно к конкретному объекту  $g$  в рабочем формальном контексте необходимо денонсировать одновременно несколько базовых семантических суждений, которые признаны истинными при пороге  $\alpha$ , а в эмпирическом исходном формальном контексте оцениваются векторами истинности вида  $\mathbf{b}_{gx} = \langle b_{gx}^+, b_{gx}^- \rangle$ , где  $x \in X = M_g^\bullet$ ,  $M_g^\bullet \subseteq M_g^*$ ,  $|M_g^\bullet| \geq 1$ . Установлено, что из трех выявленных вариантов определения вектора ужесточения, обеспечивающего «отсечение» сразу всех векторов  $\mathbf{b}_{gx}$ , предпочтительным является тот, координаты стока которого определяет вектор

$$\langle \max_{x \in X} b_{gx}^+, \min_{x \in X} b_{gx}^- \rangle,$$

и его достоверность принимается в качестве агрегированного показателя достоверности денонсируемых базовых семантических суждений. Таким образом, агрегированный показатель достоверности является функцией достоверностей денонсируемых базовых семантических суждений, которая монотонно изменяется в зависимости от количества аргументов.

Метод нормализации набора свойств объекта  $g$  в формальном контексте, реализующий предложенную эвристику, циклически до достижения искомого результата выполняет следующие действия:

- безусловная денонсация базовых семантических суждений для свойств из  $M_g^*$ , которые обуславливают свойства, отсутствующие в  $M_g^*$ ;
- выявление в  $M_g^*$  групп попарно несовместимых свойств (Н-групп – нарушителей) и связанных подмножеств таких Н-групп. Если таких групп нет – завершение работы;

• *оптимальная редукция* каждого связного подмножества  $H$ -групп, нарушающего ограничения существования свойств.

Последний шаг цикла представляет собой NP-полную задачу, однако монотонность критерия агрегированного показателя достоверности даёт основания для реализации *лексикографического* метода с построением *рекурсивной процедуры* сокращенного обхода конечного *дерева решений*.

Предложенный метод был ускорен за счёт трансформации системы измеряемых свойств, являющейся совокупностью измеряемых свойств и отношений между ними. Отношение  $C(x, y)$  индуцирует на множестве  $M$  бинарное отношение взаимообусловленности  $MC$ :  $MC(x, y) \leftrightarrow C(x, y) \wedge C(y, x)$ , – которое рефлексивно, симметрично, транзитивно и, следовательно, разбивает множество  $M$  на классы эквивалентности – группы взаимообусловленных свойств (ВЗО-группы). При этом доказывается существование на множестве ВЗО-групп расширенных отношений обусловленности  $C^\circ$  и несовместимости  $E^\circ$ , тогда под  $H$ -группой будем понимать набор попарно несовместимых ВЗО-групп, а под  $O$ -группой – пару ВЗО-групп, связанных расширенным отношением обусловленности. Для каждой группы обосновывается связь с нормальностью содержащего их подмножества измеряемых свойств.

**Утверждение 1.** ВЗО-группа может принадлежать нормальному подмножеству измеряемых свойств только целиком.

*Доказательство.* Пусть  $X \subseteq M$  – ВЗО-группа,  $N \subseteq M$  – нормальное подмножество измеряемых свойств и  $X \cap N \neq \emptyset$ . Предположим, что  $x, y \in X$ ,  $x \neq y$  и  $x \in N$ ,  $y \notin N$ . Свойства в  $X$  взаимообусловлены, то есть  $(x, y) \in MC$  и, следовательно,  $(x, y) \in C$ , что, согласно условию замкнутости нормальных подмножеств свойств, требует  $y \in N$ . Полученное противоречие доказывает истинность утверждения 1.

**Утверждение 2.**  $O$ -группа представлена в нормальном подмножестве измеряемых свойств либо целиком, либо своей обусловливаемой частью.

*Доказательство.* Пусть  $X, Y \subseteq M$  – ВЗО-группы,  $(X, Y) \in C^\circ$  (то есть  $X$  – обусловливающая, а  $Y$  – обусловливаемая ВЗО-группа  $O$ -группы),  $N \subseteq M$  – нормальное подмножество измеряемых свойств, в котором представлена  $O$ -группа  $(X, Y)$ . Допустим, что вариантом такого представления может быть любое подмножество  $Z \subseteq X \cup Y$ , и покажем, что корректными из них являются только те, что указывает утверждение 3.2, то есть  $Z = X \cup Y$  и  $Z = Y$ . Действительно, согласно утверждению 1, возможны лишь варианты  $Z = X$ ,  $Z = Y$  и  $Z = X \cup Y$ , первый из которых исключается, так как при его реализации подмножество  $N$  оказалось бы незамкнутым:  $\forall y \in Y y \notin N$ .

**Утверждение 3.**  $H$ -группа может быть представлена в нормальном подмножестве не более чем одним своим членом.

*Доказательство.* Пусть  $X \subseteq M$  – ВЗО-группа,  $N \subseteq M$  – нормальное подмножество измеряемых свойств. Предположим, что существует  $H$ -группа, включающая  $X$  и некоторое множество  $X_Y$  других ВЗО-групп,  $|X_Y| \geq 1$ . Тогда  $\forall Y \in X_Y (X, Y) \in E^\circ$  и, следовательно,  $\forall (x \in X, y \in Y) E(x, y)$ . Поэтому условием совместимости  $N$  будет  $X \not\subseteq N \wedge Y \not\subseteq N \vee X \subseteq N \wedge Y \subseteq N \vee X \not\subseteq N \wedge Y \subseteq N$ , что и доказывает справедливость утверждения 3.

Ключевым для констатации «нормальности» подмножества множества свойств является

**Утверждение 4:** подмножество множества свойств объекта нормально тогда и только тогда, когда для него и пересекающихся с ним ВЗО-,  $O$ - и  $H$ -групп справедливы утверждения 1–3.

*Доказательство.* Действительно, пусть  $N \subseteq M$  – нормальное подмножество измеряемых свойств. Тогда для пересекающихся с  $N$  ВЗО- и  $O$ -групп нарушение утверждений 1 и 2 означает, что подмножество  $N$  незамкнуто, а для пересекающихся с  $N$   $H$ -групп нарушение утверждения 3 свидетельствует, что подмножество  $N$  несовместимо. Вместе и по отдельности и то,

и другое противоречит исходному допущению. С другой стороны, когда для некоторого подмножества  $N \subseteq M$  выполняются утверждения 1–3, оно будет замкнуто и совместимо, то есть нормально.

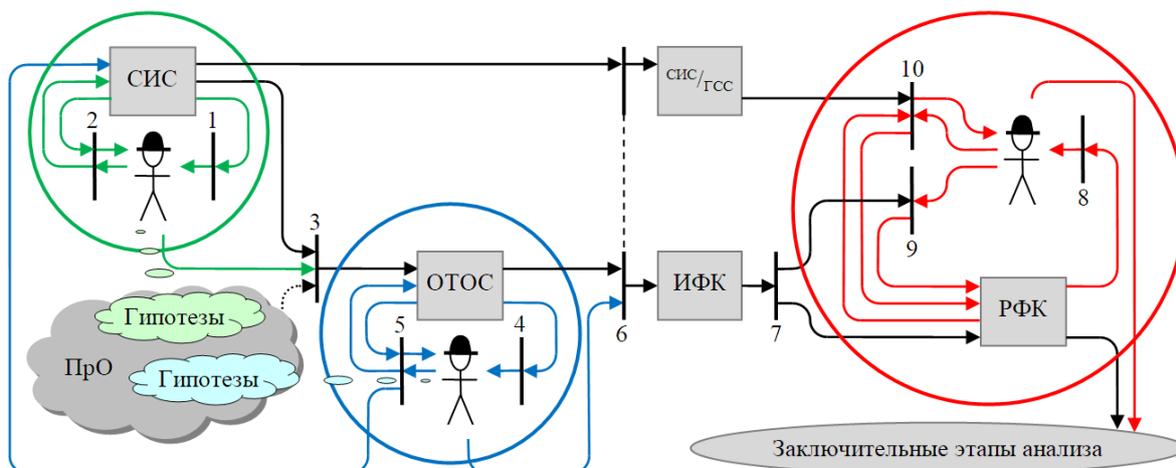
Теперь нормализацию признакового описания свойств можно существенно ускорить, осуществляя нормализацию признакового описания объекта в формальном контексте «крупными мазками», когда манипуляции со свойствами и связывающими их отношениями обусловленности и несовместимости заменяются оперированием с ВЗО-группами и отношениями  $S^\circ$  и  $E^\circ$ .

Предложенный метод нормализации свойств объекта в формальном контексте с учётом указанной трансформации системы измеряемых свойств необходимо лишь предварить денонсацией базовых семантических суждений для свойств, входящих в *некомплетные* ВЗО-группы, то есть такие, где хотя бы одно свойство не входит во множество свойств нормализуемого объекта. После выполнения данного заключительного этапа будет получен однозначный нормализованный формальный контекст, готовый к применению классического метода вывода формальных понятий.

### ЗАКЛЮЧЕНИЕ

В данной статье предложена методика построения однозначного формального контекста на основе неполных и противоречивых данных с учётом априорных знаний о зависимостях между свойствами объектов изучаемой предметной области.

В результате программной реализации предложенных моделей и методов сложилась схема построения формального контекста для вывода формальных понятий (рис.).



**Рис.** Схема формирования формального контекста для вывода формальных понятий и выполняемые при этом когнитивные акты.

Схема построения формального контекста состоит из следующих шагов:

- шаги 1, 4, 8, 17 – осмысление субъектом либо результата *предшествующего*, либо результата *любого из последующих* шагов анализа (фигуру субъекта, которая занимает в двудольной структуре схемы место «продукта», можно интерпретировать как измененное состояние «сознающего ума» [18] исследователя вследствие полученных результатов анализа).

Например, 4-й шаг может быть инициирован восприятием субъекта как обобщенной таблицы «объекты–свойства» (ОТОС), так и редуцированного представления системы измеряемых свойств (СИС/ГСС, где *группами сопряженных свойств* (ГСС) собирательно именуется ВЗО-, О- и Н-группы) или формальных понятий. Для исключения нагромождения связей такие переходы изображены лишь в пределах выделенных когнитивных актов;

- шаги 2, 5 – формирование и редактирование соответственно СИС и ОТОС;
- шаг 3 – измерение (в широком смысле) свойств объектов предметной области (ПрО);

- шаг 6 – построение СИС/ГСС и исходного формального контекста (ИФК);
- шаг 7 – возможное отождествление рабочего формального контекста (РФК) с ИФК;
- шаг 9 – дефазификация исходного формального контекста: установка субъективного порога доверия к эмпирическим семантическим суждениям о предметной области вида «объект  $x$  обладает свойством  $y$ », нестрогие оценки истинности которых составляют исходный формальный контекст, и обновление рабочего формального контекста путем построения его новой версии как стандартного сечения исходного формального контекста с выбранным субъективным порогом доверия;

- шаг 10 – нормализация рабочего формального контекста – преобразование рабочего формального контекста, в ходе которого снимаются противоречия между эмпирическими данными и априорными ограничениями свойств: множество свойств каждого отдельно взятого объекта измерения нормализуется путем отказа от части эмпирической информации, характеризующейся меньшим значением агрегированного показателя достоверности.

Обратное по отношению к субъекту воздействие связано с информированием о невозможности выполнить данное действие.

В заключительные этапы дальнейшего анализа входит извлечение из рабочего формального контекста формальных понятий, построение решётки формальных понятий и т. д. [19].

Все предложенные и разработанные модели и методы по подготовке контекста к выводу формальных понятий с учётом ограничений существования свойств были реализованы в программной лаборатории *OntoWorker*. Разработанные программные средства пригодны к применению в различных областях, например, при построении когнитивных карт коллективом экспертов [20, 21], и при решении задачи структурного анализа и синтеза технических решений [22]. Но целевым назначением предложенной методики является поддержка онтологического анализа данных, где формирование контекста является начальным этапом, первичной обработкой эмпирических данных.

### БЛАГОДАРНОСТИ И ПОДДЕРЖКА

Работа выполнена при финансовой поддержке Минобрнауки России, код научной темы FMRW-2022-0030. Автор выражает благодарность научному руководителю Сергею Викторовичу Смирнову за полезные замечания.

### СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- [1] Ganter B., Wille R. Formal Concept Analysis. Mathematical foundations. Berlin-Heidelberg: Springer-Verlag, 1999.
- [2] Ferré S., Huchard M., et al. Formal concept analysis: from knowledge discovery to knowledge processing // A Guided Tour of Artificial Intelligence Research. Vol. II: AI Algorithms. Springer Int. Publishing. 2020. Pp. 411–445.
- [3] Ignatov D. I. Introduction to formal concept analysis and its applications in information retrieval and related fields // Information Retrieval. Revised Selected Papers 8th Russian Summer School, RuSSIR. 2014 (August 18–22, 2014, Nizhny Novgorod, Russia). Springer Int. Publishing. 2015. Pp. 42–141.
- [4] Офицеров В. П., Смирнов В. С., Смирнов С. В. Метод альфа-сечения нестрогих формальных контекстов в анализе формальных понятий // Проблемы управления и моделирования в сложных системах: Тр. XVI междунар. конф. Самара: СамНЦ РАН, 2014. С. 228–244. EDN [VLBEIT](#). [[Ofitserov V. P., Smirnov V. S., Smirnov S. V. “Alpha-cutting method of non-strict formal contexts in the analysis of formal concepts” // Problems of Control and Modeling in Complex Systems: Proc. XVI Int. Conf. Samara, 2014, pp. 228–244. EDN [VLBEIT](#). (In Russian).]]
- [5] Lammari N., Metals E. Building and maintaining ontologies: a set of algorithms // Data & Knowledge Engineering. 2004. Vol. 48(2). Pp. 155–176.
- [6] Пронина В. А., Шипилина Л. Б. Использование отношений между атрибутами для построения онтологии предметной области // Проблемы управления. 2009. № 1. С. 27–32. EDN [JWKFLH](#). [[Pronina V. A., Shiplina L. B. “Using relationships between attributes to build an ontology of a subject area” // Problems of Management. 2009. No. 1, pp. 27–32. (In Russian).]]
- [7] Загоруйко Н. Г. Когнитивный анализ данных. Новосибирск: Гео, 2013. 186 с. [[Zagoruiko N. G. Cognitive Data Analysis. Novosibirsk: “Geo”, 2013. EDN [JWKFLH](#). (In Russian).]]
- [8] Рубашкин В. Ш. Представление и анализ смысла в интеллектуальных информационных системах. М.: Наука, 1989. 192 с. [[Rubashkin V. Sh. Representation and Analysis of Meaning in Intelligent Information Systems. Moscow, 1989. (In Russian).]]
- [9] Смирнов С. В. Двухсоставность феномена информации и анализ данных (с примерами из когнитивного анализа) // Информационные технологии и нанотехнологии (ИТНТ-2017). Самара: Новая техника, 2017. С. 1846–1849. EDN [YQLSRX](#).

- [[Smirnov S. V. "Two-component nature of the phenomenon of information and data analysis (with examples from cognitive analysis)" // Information Technologies and Nanotechnologies (ITNT-2017). Samara: New Technology, 2017, pp. 1846-1849. EDN [YQLSRX](#). (In Russian).]]
- [10] Голышев В. К. Нечёткий анализ формальных понятий: обзор подходов на примере // Математическое и программное обеспечение информационных, технических и экономических систем: Мат-лы VIII междунар. молодёжной науч. конф. Томск: Изд-во Томск. гос. ун-та, 2021. С. 28–32. [[(In Russian).]]
- [11] Самойлов А. Е., Смирнов С. В. Универсальный метод учета ограничений существования свойств в задачах вывода формальных понятий из нечетких объектно-признаковых данных // Сб. тр. VII междунар. конф. и молодеж. шк. (20-24 сентября 2021 г., Самара). Самара, 2021. С. 33862. EDN [DJDSRP](#). [[Samoilov A. E., Smirnov S. V. "Universal method for taking into account constraints on the existence of properties in problems of inferring formal concepts from fuzzy object-attribute data" // Proc. VII Int. Conf. and Youth Schools (September 20-24, 2021, Samara). Samara, 2021, pp. 33862. EDN [DJDSRP](#). (In Russian).]]
- [12] Семенова В. А. Выбор логической модели для представления эмпирической информации в онтологическом анализе данных // Информатика и вычислительная техника: XII Всеросс. науч.-тех. конф. аспирантов, студентов и молодых ученых ИВТ-2020 (15-16 июня 2020 г., Ульяновск): Сб. науч. тр. Ульяновск: УлГТУ, 2020. С. 205–210. EDN [EJYTKT](#). [[Semenova V. A. "Selection of a logical model for representing empirical information in ontological data analysis" // Informatics and Computing Engineering IVT-2020 (June 15-16, 2020, Ulyanovsk). Ulyanovsk: UISTU, 2020, pp. 205-210. EDN [EJYTKT](#). (In Russian).]]
- [13] Аршинский Л. В. Векторные логики: основания, концепции, модели. Иркутск, 2007. 228 с. EDN [QJSGNH](#). [[Arshinsky L. V. Vector Logics: Foundations, Concepts, Models. Irkutsk, 2007. EDN [QJSGNH](#). (In Russian).]]
- [14] Аншаков О. М., Виноградов Д. В., Финн В. К. Многозначные логики и их применения. Т. 1. Логические исчисления, алгебры и функциональные свойства. М.: URSS, 2020. 502 с. [[Anshakov O. M., Vinogradov D. V., Finn V. K. Many-Valued Logics and Their Applications. Vol. 1. Logical Calculi, Algebras and Functional Properties. Moscow: URSS, 2020 (In Russian).]]
- [15] Semenova V. A., Smirnov S. V. Extended methodology for deriving formal concepts // J. of Physics: Conf. Series. Int. Conf. «Marchuk Scientific Readings 2021» MSR-2021 (4–8 October 2021, Novosibirsk, Russia). 2021. Vol. 2099. 012026. Pp. 1–9. DOI: 10.1088/1742-6596/2099/1/012026. EDN [OVAZCP](#).
- [16] Семенова В. А. Эвристика и численный метод нормализации эмпирического  $V^{TF}$ -контекста в задаче онтологического анализа данных // Информационные и математические технологии в науке и управлении. 2021. № 2(22). С. 61–69. EDN [GRBPCE](#). [[Semenova V. A. "Heuristics and numerical method for normalizing the empirical  $V^{TF}$  context in the problem of ontological data analysis" // Information and Mathematical Technologies in Science and Management. 2021. No. 2(22), pp. 61–69. EDN [GRBPCE](#). (In Russian).]]
- [17] Семенова В. А., Смирнов С. В. Механизм нормализации эмпирического контекста в онтологическом анализе данных // СИИТ. 2021. Т. 3. № 3(7). С. 45–52. EDN [QXRTXB](#). [[Semenova V. A., Smirnov S. V. "Mechanism of normalization of empirical context in ontological data analysis" // СИИТ. 2021. Vol. 3, No. 3(7), pp. 45–52. EDN [QXRTXB](#). (In Russian).]]
- [18] Чалмерс Д. Сознательный ум: в поисках фундаментальной теории. М.: URSS, 2019. 512 с. [[Chalmers D. The Conscious Mind: In Search of a Fundamental Theory. Moscow: URSS, 2019. (In Russian).]]
- [19] Семенова В. А., Смирнов С. В. Функциональное наполнение и архитектура программной лаборатории для онтологического анализа данных // Вестник Самарского государственного технического университета. Серия Технические науки. 2023. Т. 31. № 2. С. 85–100. EDN [AFLYSK](#). [[Semenova V. A., Smirnov S. V. "Functional filling and architecture of a software laboratory for ontological data analysis" // Bulletin of Samara State Technical University. Series "Engineering Sciences". 2023. Vol. 31, No. 2, pp. 85–100. EDN [AFLYSK](#). (In Russian).]]
- [20] Смирнов С. В. Онтологический анализ экспертных данных в задаче формирования нечетких когнитивных карт // Труды ИСА РАН. 2019. Т. 69. Вып. 4. С. 79–86. DOI: 10.14357/20790279190410. EDN [DXIERS](#). [[Smirnov S. V. "Ontological analysis of expert data in the problem of forming fuzzy cognitive maps" // Proc. ISA RAS. 2019. Vol. 69. Issue. 4. Pp. 79–86. (In Russian). DOI 10.14357/20790279190410. EDN [DXIERS](#).]]
- [21] Андрейчиков А. В., Андрейчикова О. Н. Наука и искусство принятия решений. Кн. 3: Подходы к решению задач коллективного выбора. Методы генерации решений. М.: ЛЕНАНД, 2021. 248 с. [[Andreychikov A. V., Andreychikova O. N. Science and Art of Decision-Making. Book 3: Approaches to Solving Collective Choice Problems. Methods of Generating Solutions. Moscow: LENAND, 2021. (In Russian).]]
- [22] Семенова В. А., Смирнов С. В. Модели и методы онтологического анализа данных в задаче структурного анализа и синтеза технических решений // Онтология проектирования. 2023. Т. 13. № 4(50). С. 531–547. EDN [UADWCT](#). [[Semenova V. A., Smirnov S. V. "Models and methods of ontological data analysis in the problem of structural analysis and synthesis of technical solutions" // Ontology of Designing. 2023. Vol. 13, No. 4(50), pp. 531–547. EDN [UADWCT](#). (In Russian).]]

*Поступила в редакцию 28 октября 2024 г.*

#### МЕТАДАННЫЕ / METADATA

**Title:** Forming a context for deriving formal concepts from incomplete and inconsistent data.

**Abstract:** In the era of big data, the relevance of systems data analysis increases, as it helps organizations effectively process and interpret huge amounts of information to achieve strategic goals. However, data on the knowledge domain of interest that require systematic understanding and extraction of useful information from them are often characterized as incomplete, vague, and even inconsistent. To process this kind of information, formal concept analysis, in particular, fuzzy analysis, is often used. However, sometimes it is necessary to consider additional a priori information of the researcher on the knowledge domain under study, called

properties existence constraints. This article sets the task of considering the properties existence constraints when deriving formal concepts from incomplete and inconsistent data. This problem can be solved by turning to the methods of concept formation in classical logic. The article proposes a heuristic method for normalizing a non-strict formal context accounting properties existence constraint when deriving formal concepts from incomplete and inconsistent data. In conclusion, a scheme of the methodology for constructing an unambiguous formal context from incomplete and contradictory data is given, considering the constraints on the existence of properties.

**Key words:** empirical data; knowledge domain; formal context; basic semantic proposition; properties existence constraints.

**Язык статьи / Language:** Русский / Russian.

**Поддержка/Support:** Государственное задание Минобрнауки России, код темы FMRW-2022-0030.

#### **Об авторе / About the author:**

##### **СЕМЕНОВА Валентина Андреевна**

Институт проблем управления сложными системами РАН,  
Россия.

Мл. науч. сотр. лаб. анализа и моделирования сложных систем. Магистр механики и мат. моделирования (Самарск. нац. иссл. ун-т, 2017). Иссл. в обл. онтологического инжиниринга.

E-mail: azulave@rambler.ru

ORCID: <https://orcid.org/0000-0002-0557-3890>

##### **SEMENOVA Valentina Andreevna**

Institute of Complex Systems Control Problems of the RAS,  
Russia.

Jr. researcher, Lab. of Analysis and Modeling Complex Systems. Master of Mech and Math Modeling (Samara National Research Univ., 2004). Research in the field of ontological engineering.

E-mail: azulave@rambler.ru

ORCID: <https://orcid.org/0000-0002-0557-3890>