На правах рукописи

Семенова Валентина Андреевна

**МЕТОДИЧЕСКОЕ И АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ  
ПОДГОТОВКИ КОНТЕКСТА ДЛЯ ВЫВОДА ФОРМАЛЬНЫХ  
ПОНЯТИЙ В ОНТОЛОГИЧЕСКОМ АНАЛИЗЕ ДАННЫХ**

2.3.1 – Системный анализ, управление и обработка информации, статистика

АВТОРЕФЕРАТ  
диссертации на соискание ученой степени  
кандидата технических наук

Самара 2024

Работа выполнена на кафедре «Прикладная математика и информатика» ФГБОУ ВО «Самарский государственный технический университет» и  
в лаборатории анализа и моделирования сложных систем  
Института проблем управления сложными системами  
Российской академии наук – обособленном подразделении ФГБУН  
Самарского федерального исследовательского центра  
Российской академии наук (ИПУСС РАН – СамНЦ РАН).

|  |  |
| --- | --- |
| **Научный руководитель:** | **Смирнов Сергей Викторович**, д.т.н., доцент, главный научный сотрудник ИПУСС РАН – СамНЦ РАН |
| **Официальные оппоненты:** | **Иващенко Антон Владимирович**, д.т.н., профессор, директор НОЦ «Передовая медицинская инженерная школа», г. Самара  **Семенова Дарья Владиславовна**, к.ф.-м.н., доцент кафедры высшей и прикладной математики Сибирского федерального университета (СФУ), г. Красноярск |
| **Ведущая организация:** | ФГБУН Институт систем информатики им. А.П. Ершова СО РАН |

Защита состоится «хх» июня 2024 года в хх часов на заседании диссертационного совета 24.2.377.02 на базе ФГБОУ ВО  
«Самарский государственный технический университет»  
по адресу: г. Самара, ул. Молодогвардейская, 244, ауд. №200.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВО  
«Самарский государственный технический университет»  
по адресу: 443100, г. Самара, ул. Первомайская, 18.

Отзывы на автореферат (в двух экземплярах, заверенные печатью) просим направлять по адресу: 443100, г. Самара, ул. Молодогвардейская, 244,  
Главный корпус, ученому секретарю диссертационного совета 24.2.377.02.  
Тел. (846) 337-04-43, e-mail: *saushkin.mn@samgtu.ru*

Автореферат разослан «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2024 г.

Учёный секретарь  
диссертационного совета  
24.2.377.02 САУШКИН Михаил Николаевич

**ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

**Актуальность работы.** Практика показывает, чтосложные информационные системы результативны лишь при адекватном и согласованном представлении их *предметной области* (ПрО). Разработка и применение необходимых для этого информационных моделей составляют современное содержание *онтологического подхода* в вычислительных науках. Для построения подобных моделей, которые получили известность как [*формальные*] *онтологии*, широко используются основанные на экспертизе методы инженерии знаний, получившие развитие в работах Т.А. Гавриловой, В.В. Грибовой, Ю.А. Загорулько, А.В. Иващенко, Л.В. Массель, П.О. Скобелева, T. Gruber, N. Guarino, J. Sowa и др.

Одним из строго обоснованных методов построения «скелета» онтологии – множества *формальных понятий* (ФП) с заданным на нем отношением обобщения – является вывод ФП из *формального контекста* (ФК), создаваемого на основе объектно-признаковых эмпирических данных об интересующей субъекта ПрО. Наиболее результативные методики, реализующие это направление, опираются на *анализ формальных понятий* (АФП), предложенный R. Wille, и развитый в работах Е.Е. Витяева, С.О. Кузнецова, Д.В. Семёновой, R. Belohlávek, B. Ganter, S. Ferré и др. Менее известен существенно иной формальный метод, основанный на извлечении онтологии из *ограничений существования свойств* (ОСС) у объектов изучаемой ПрО (N. Lammari, E. Metais и др.). Предпринятая в работах В.А. Прониной и Л.Б. Шипилиной попытка совместить указанные подходы малопродуктивна, поскольку сведена лишь к рекомендации проверять ФК насчет соблюдения ОСС.

Актуальность проблемы гибридизации дедуктивных методов вывода ФП при учете реалий накопления эмпирической информации определила цель и задачи диссертации.

**Цель диссертационной работы** –разработка моделей, методов, алгоритмического обеспечения и программных средств подготовки ФК для вывода множества ФП, формирующих остов онтологического описания интересующей субъекта ПрО, на основе неполных и противоречивых эмпирических данных и учёта ОСС у объектов ПрО.

**Задачи, решаемые в диссертации**:

1. Анализ существующих методик вывода формальных понятий, необходимости и возможности их гибридизации.
2. Разработка моделей и методов структурирования гипотетических представлений и эмпирической информации о ПрО, включая метод интерпретации и консолидации объектно-признаковых данных с использованием многозначной логики для построения нестрогого *исходного* ФК (ИФК).
3. Разработка методики построения логически однозначного *рабочего* ФК (РФК), включающей метод учёта ОСС.
4. Разработка алгоритмического обеспечения подготовки ФК для вывода ФП и программных средств онтологического анализа данных.
5. Апробация разработанных методов и средств при решении технических и других задач.

**Объект исследования** – процесс первичной обработки объектно-признаковых данных, формирующий ФК для вывода ФП.

**Предмет исследования** – способы учёта при формировании ФК реалий накопления эмпирической информации и априорных гипотез субъекта о понятийном описании интересующей его ПрО.

**Методы исследования.** В диссертационной работе использовались методы теории множеств, бинарных отношений, классической и математической многозначной логики, имитационного статистического моделирования, многокритериальной оптимизации, объектно-ориентированного проектирования.

**Достоверность результатов** диссертации подтверждается корректным применением теоретического аппарата использованных методов исследования и апробацией предложенных в диссертации моделей, методов и алгоритмов для решения практически важных задач.

**Научная новизна** полученных в диссертации результатов:

1. Доказано единство логической природы существующих подходов к выводу формальных понятий, отличающееся обращением к их гипотетико-дедуктивным схемам, и предложен механизм их совмещения, отличающийся использованием нестрогих формальных контекстов, что в совокупности открыло возможность гибридизации дедуктивных методик вывода формальных понятий.

2. Предложены модели и методы структурирования эмпирической информации, отличающиеся использованием многозначной векторной логики, что дало возможность отразить реалии накопления эмпирических данных о предметной области.

3. Для нормализации множества свойств каждого объекта в контексте предложена эвристика, отличающаяся использованием сведений о достоверности базовых семантических суждений, образующих контекст, и последовательным приближением к нормальному множеству свойств объекта, что позволило разработать результативный метод нормализации контекста.

4. Предложен метод реструктуризации системы измеряемых свойств, отличающийся выявлением расширенных экзистенциональных отношений на классах эквивалентности множества измеряемых свойств, что обеспечило существенное сокращение его описания и, как следствие, ускорение нормализации контекста.

5. Разработано алгоритмическое обеспечение подготовки контекста для вывода формальных понятий, отличающееся использованием предложенных моделей и методов, что обеспечило корректную обработку неполных и противоречивых эмпирических данных многомерных измерений и наблюдений в предметной области с учётом ограничений существования свойств.

**Теоретическая значимость** работы заключается в обосновании единства гипотетико-дедуктивных схем сосуществующих подходов к выводу формальных понятий и разработке гибридной методологии подготовки контекста для вывода формальных понятий в онтологическом анализе данных.

**Практическая значимость** определяется реализацией программных инструментов первичной обработки эмпирических объектно-признаковых данных для построения ФК.

Результаты работы применены при выполнении следующих НИР по госзаданиям ИПУСС РАН – СамНЦ РАН:

* «Интерсубъективное управление инновационным развитием социотехнических объектов с применением онтологических моделей ситуаций» гос. рег. № АААА-А16-116040410061-0;
* «Разработка и исследование методов и средств аналитического конструирования, компьютерного представления знаний, вычислительных алгоритмов и мультиагентных технологий в задачах оптимизации процессов управления сложными системами» гос. рег. № АААА-А19-119030190053-2.

Результаты диссертационной работы использованы при выполнении научно-исследовательских и квалификационных работ по образовательным программам бакалавриата и магистратуры Самарского национального исследовательского университета имени академика С.П. Королёва на кафедре программных систем. Практическое использование подтверждено актом ООО «Открытый код» (г. Самара).

**Соответствие диссертации паспорту научной специальности 2.3.1:**

п. 1. Теоретические основы и методы системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта (пп. 1-3 новизны); п. 3. Разработка критериев и моделей описания и оценки эффективности решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта (пп. 2, 3 новизны); п. 4. Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта (пп. 2, 3 новизны); п. 5. Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта (пп. 4 новизны); п. 13. Методы получения, анализа и обработки экспертной информации, в том числе на основе статистических показателей (пп. 2, 3 новизны).

**Личный вклад.** Постановка и обсуждение возможных путей решения задач исследования осуществлялась совместно с научным руководителем. Результаты, составляющие научную новизну и выносимые на защиту, получены лично автором, либо в неделимом соавторстве. Автором выполнено проектирование архитектуры и функциональности программной лаборатории для онтологического анализа данных, проведена апробация программных средств при решении задач в различных научных и технических областях.

**Апробация работы.** Основные результаты работы докладывались на международных конференциях «Информационные технологии и нанотехнологии» (Самара, 2015, 2018, 2019, 2022 гг.), «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (Санкт-Петербург, 2022 г.), «Марчуковские научные чтения» (Новосибирск, 2021 г.), «Перспективные информационные технологии» (Самара, 2016, 2017, 2018, 2020 гг.), «Проблемы управления и моделирования в сложных системах» (Самара, 2017, 2018, 2019 гг.), «Открытые семантические технологии проектирования интеллектуальных систем» (Минск, 2016 г.); на всероссийских конференциях с международным участием «Информационные и математические технологии в науке и управлении» (Иркутск, 2021, 2023 гг.), «Информационные технологии интеллектуальной поддержки принятия решений» (Уфа, 2021 г.); на всероссийской конференции аспирантов, студентов и молодых ученых «Информатика и вычислительная техника» (Ульяновск, 2020 г.).

**Положения, выносимые на защиту:**

1. Существующие дедуктивные методики вывода формальных понятий имеют единую логическую природу, их гибридизация основана на отказе от части эмпирической информации.

2. Прагматические модели системы измеряемых свойств и результатов многомерных измерений предметной области. Методика использования многозначной векторной логики для интерпретации и консолидации эмпирической информации.

3. Эвристический метод нормализации нестрогого формального контекста.

4. Метод реструктуризации системы измеряемых свойств.

5. Архитектурные решения и алгоритмическое обеспечение программной лаборатории для онтологического анализа данных.

**Публикации.** Результаты исследования опубликованы в 8 статьях в журналах, рекомендованных ВАК. 5 статей опубликованы в изданиях, индексируемых в зарубежной базе Scopus, 12 в других изданиях. Получено свидетельство о государственной регистрации программы для ЭВМ.

**Структура и объём работы.** Диссертационная работа состоит из введения, пяти разделов, заключения, списка сокращений и обозначений, списка литературы и приложений. Объём работы: 178 страниц текста, включая 57 рисунков и 6 таблиц; 26 страниц занимают 6 приложений.

**ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** показана актуальность рассматриваемой проблемы, сформулирована цель исследования и определены задачи, необходимые для ее достижения. Сформулированы положения, составляющие новизну полученных результатов и выносимые на защиту. Дан общий обзор работы.

**В первом разделе** проведен анализ когнитивной деятельности субъекта при применении двух существующих методик вывода ФП. Обосновано единство логической природы этих методик, что подтверждает насущность и открывает возможность их гибридизации.

АФП направлен на обработку эмпирических объектно-признаковых данных, универсальной и стандартизированной формой структуризации которых является таблица «*объекты-свойства*», а моделью – кортеж

(*G*, *M*, *D*, ***I***), (1)

где *G* = {*gi*}*i*= 1,…,*r*, *r* = |*G*| ≥ 1 – конечное множество наблюдавшихся объектов ПрО, ставших *объектами измерения*; *M* = {*mj*}*j*= 1,…,*s*, *s* = |*M*| ≥ 1 – конечное множество свойств, измеряемых у каждого объекта; *D* – совокупное множество зафиксированных значений разных свойств: *D* ⊆ *D*\* = ∪*j*= 1,…,*sDj*, *Dj* – домен свойства *mj* (в технике ему сопоставляется динамический диапазон *процедуры измерения* *j*-го свойства); ***I*** – тернарное отношение между *G*, *M* и *D*, определённое для всех пар из *G* × *M*. Для вывода ФП модель (1) путем *концептуального шкалирования* и/или учета ограниченности динамических диапазонов процедур измерения преобразуется в *логически однозначный* ФК:

(*G*, *M*, *I*), (2)

где *I*: *G* × *M*→ {**True**, **False**}, *I*⊆ *G* × *M* – *бинарное соответствие*, интерпретируемое как совокупность оценок истинности *bij* *базовых семантических суждений* (БСС) о ПрО: «объекту *gi* присуще свойство *mj*».

Метод, основанный на извлечении онтологии из ОСС, исходит из знания *экзистенциональных* отношений между принимаемыми во внимание измеряемыми свойствами объектов ПрО:

* *обусловленности* *C*: *M*×*M*→ {**True**, **False**}, *C*⊆ *M* × *M*, *С*(*mj*, *mk*) ↔ ∀*g*∈*G*: *mj*∈{*g*}' → *mk*∈{*g*}', «'» - оператор Галуа. *С* рефлексивно, несимметрично и транзитивно (рисунок 1a);
* *несовместимости* *E*: *M*×*M*→ {**True**, **False**}, *E*⊆ *M* × *M*, *E*(*mj*, *mk*) ↔ ∀*g*∈*G*: *mj*∈{*g*}' → *mk*∉{*g*}'. *E* антирефлексивно, симметрично и нетранзитивно, но «*транзитивно относительно обусловленности*» (рисунок 1б): ∀ *x*, *y*, z ∈ *M*, *x* ≠ z, *y* ≠ z: *С*(*x*, *y*) ∧ *E*(*y*, *z*) → *E*(*x*, *z*).

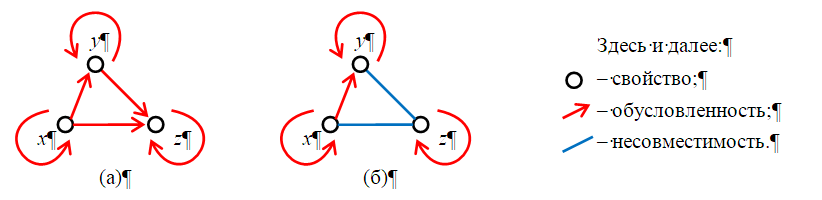


Рисунок 1 – Примеры экзистенциональных зависимостей между свойствами

Содержания ФП, выводимых из *системы измеряемых свойств* (СИС; см. пример на рисунке 2), моделируемой кортежем

(*M*, *C*, *E*), (3)

есть *нормальные* подмножества *M*, которые *замкнуты* и *совместимы*:

* *Y* ⊆ *M* замкнуто, если ∀*mj* ∈ *Y*: (∃*mk* ∈ *M*, *mk* ≠ *mj*: *С*(*mj*, *mk*)) → *mk*∈*Y*;
* *Y* ⊆ *M* совместимо, если ∀*mj* ∈ *Y*: (∃*mk* ∈ *M*, *mk* ≠ *mj*: *Е*(*mj*, *mk*)) → *mk*∉*Y*.

В диссертации показано, что априорная когнитивная деятельность субъекта в обеих методиках вывода ФП связана с выдвижением *гипотез* о понятийном описании исследуемой ПрО *путем формирования* СИС (3):

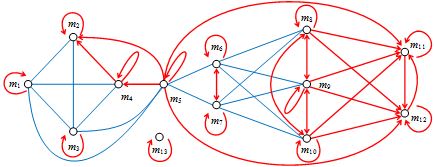


Рисунок 2 – Пример системы измеряемых свойств

* в АФП гипотетические понятия формируются простейшим способом на основе всех подмножеств *M* в качестве содержаний понятий (при этом *C* = ∅, *E* = ∅, а на объемы не накладываются какие-либо ограничения);
* использование ОСС де-факто реализует два (и только два) другие известные в классической логике способы формирования понятий – *ограничение* и *деление*, порождающих на множестве свойств *C* и *E* соответственно.

В качестве иллюстрации тезиса о необходимости гибридизации сосуществующих методик при *развитии* онтологического анализа данных (ОАД) в работе приводится пример использования *нечеткого* АФП, когда в (2) соответствие *I* отражает «размытые», «мягкие» оценки истинности БСС. Показано, что обычная в подобных случаях установка *субъективного порога доверия* (СПД) для дефаззификации *I* с целью формирования однозначного ФК приводит без учета ОСС к выводу ложных понятий.

Вместе с тем, использование практически адекватных «мягких» ИФК открывает ранее не рассматривавшуюся возможность:

* выделить в ограниченно достоверной эмпирической информации о ПрО определенную часть, отказ от которой приведет к однозначному ФК, удовлетворяющему ОСС;
* среди всех таких частей для конкретного ФК, выбрать и исключить из дальнейшего анализа характеризующуюся меньшей достоверностью.

**Во втором разделе** разработаны прагматические модели СИС и таблицы «объекты-свойства», методы их реструктуризации с помощью концептуального шкалирования, метод интерпретации и консолидации эмпирических данных для формирования «мягкого» ИФК.

Модификация модели (3) *–* кортеж (*M*, *CNR*, *E*, *H*) *–* определяет практически ориентированное представление СИС (рисунок 3), где:

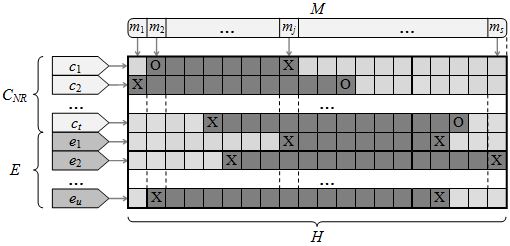


Рисунок 3 – Схема модифицированной модели системы измеряемых свойств  
(«X» и «О» – указатели инцидентности, причем второй – для обусловленного свойства)

* отношение *C* замещено соответствующей нерефлексивной обусловленностью *CNR* = {*ck*}*k*= 1,…,*t*, 0 ≤ *t* = |*CNR*|, описывающей множество пар разных свойств, находящихся в отношении обусловленности;
* *E*= {*el*}*l*= 1,…,*u*, 0 ≤ *u* = |*E*| – множество пар несовместимых свойств;
* *H* = (*hkl*)*k*= 1,…,*t*;*l*= 1,…,*u* – матрица инцидентности «*сопряженные пары разных свойств – свойства*».

Для обобщения модели (1) с целью учета реалий накопления эмпирических данных (многократные независимые измерения свойств у одного и того же объекта измерения, с ограниченной *достоверностью*; использование нескольких независимых процедур измерения одного и того же свойства, с ограниченной *степенью доверия*; неполнота серий измерений, выражающаяся в пропуске выполнения некоторых процедур измерения (фиксируется как получение результата **NM** «измерение не производилось»); наличие у всякой процедуры измерения результатов **None**, или «нет информации», и **Failure**, означающего сбой процедуры измерения, или отказ выполнить измерение) использована модель *обобщенной таблицы* «*объект-свойство*» (ОТОС; см. рисунок 4)

(*G*, *M*, *Se*, *Pr*, *A*): (4)

* *Sе* = {*Sе*(*i*)}*i*= 1,…,*r*, |*Sе*| = Σ*ri*= 1|*Sе*(*i*)| = *m* – множество выполненных для изучаемой ПрО серий измерений, *Sе*(*i*) = {*sе*(*i*)*k*}*k*= 1,…,*q*(*i*), *q*(*i*) ≥ 1, *i* = 1,…, *r* – множество серий измерений, выполненных для объекта *gi* ∈ *G*, каждая серия *se*(*i*)*k* характеризуется степенью достоверности ее результатов *st*(*i*)*k*;
* *Pr* = {*Pr*(*j*)}*j*= 1,…,*s*, |*Pr*| = Σ*sj*= 1|*Pr*(*j*)| = *n* – арсенал используемых процедур измерения, *Pr*(*j*) = {*pr*(*j*)*k*}*k*= 1,…,*p*(*j*), *p*(*j*) ≥ 1, *j* = 1,…, *s* – множество процедур измерения свойства *mj* ∈ *M*, всякая процедура *pr*(*j*)*k* характеризуется степенью доверия к ее результатам *pt*(*j*)*k*;
* *A* = (*aij*)*i*= 1,…,*m*;*j*= 1,…,*n* – матрица результатов серий измерений свойств *M* у объектов из *G*, выполненных с помощью процедур измерения. В качестве элементов этой матрицы выступают значения, принадлежащие динамическим диапазонам процедур измерения, *лингвоконстанта* **X**, свидетельствующая о получении такого результата, а также *лингвоконстанты* **NM**, **None**, **Failure**.

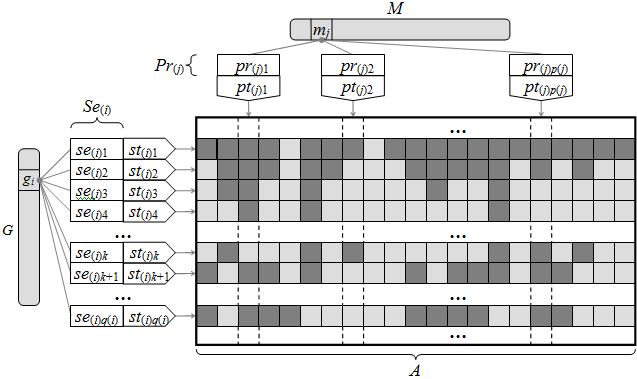


Рисунок 4 – Схема обобщенной таблицы «объекты-свойства»  
(светло-серые ячейки соответствуют результату измерения **NM**)

Показано, что адекватное оценивание истинности БСС на основе ОТОС, предоставляющей в общем случае неполную и противоречивую информацию о ПрО, возможно лишь в рамках *многозначной логики*. Проведенный анализ выявил в качестве наиболее предпочтительной *векторную* VTF-логику, в которой истинность суждения оценивается вектором***b****ij*=〈*bij*+, *bij*–〉, *bij*+,*bij*–∈[0, 1], где компонент *bij*+ формируется показаниями, подтверждающими суждение, а *bij*– – показаниями его отрицающими.

Смысловая интерпретация эмпирических данных в *A* начинается с построения *нестрогого* соответствия «*серии измерений* – *процедуры измерений*» *IA* = (***b****ij*)*i*=1,…,*m*;*j*=1,…,*n*:

⎧ 〈1, 0〉, если *aij* ∈ *D*\* или *aij* = **X**; ⎫

***b****ij* = ⎨ 〈0, 1〉, если *aij* = **None**; ⎬.

⎪ 〈0.5, 0.5〉, если *aij* = **Failure**; ⎪

⎩ 〈0, 0〉, если *aij* = **NM**. ⎭

Учёт достоверности серии и степени доверия к процедуре измерения осуществляется операцией «00‑композиции» VTF-логики на основе композиционного умножения согласно *t*-норме *x*• *y = xy*. *Совмещение* всех БСС, истинность которых оценена в каждой выполненной для *i*-го объекта серии измерений каждой процедурой измерения *j*-го свойства, реализуется по схеме «усиления-усреднения» (частный случай «11‑композиционного совмещения» VTF-логики на основе композиционного сложения согласно *s*-норме *x*⊕ *y =***min**(1*, x* *+* *y*)с весами для совмещаемых оценок разных аспектов истинности, равными1/*nc*, где *nc* – число совмещаемых оценок).

В результате из *IA* получаем нестрогое соответствие *I* = (***b****ij*)*i*=1,…,*r*;*j*=1,…,*s*, определяющее *нестрогий* ИФК вида (2).

**Третий раздел** посвящен разработке методического обеспечения апостериорных когнитивных актов субъекта, направленных на формирование РФК.

Предложена двухэтапная схема пороговой дефаззификации ИФК: на первом единожды выполняемом этапе выявляются БСС, ложные при любом допустимом пороге доверия к эмпирической информации, а ко второму этапу субъект может возвращаться неоднократно, варьируя СПД ***α*** = 〈*α*+, *α*−〉.

Когнитивное содержание субъективного решения об учете ОСС состоит в требовании соответствия выводимого понятийного описания изучаемой ПрО и априорных гипотез об этой понятийной структуре, индуцируемых СИС, а способом учета ОСС является нормализация наборов свойств *у каждого отдельно взятого* объекта в формируемом РФК.

Проанализированы существующие подходы к нормализации «мягких» ФК и констатированы их недостатки в разрабатываемой парадигме онтологического анализа данных.

Предложено новое решение этой задачи на основе эвристики, согласно которой разумным подходом к нормализации эмпирически и субъективно выявленного набора свойств *Mg*\* ⊆ *M* объекта *g* ∈ *G* в РФК будет *денонсация* некоторых БСС, которые описываются как истинные, но *нарушают* «нормальность» множества свойств объекта. При этом из нескольких таких вариантов денонсации БСС следует выбрать тот, что обеспечивает минимум *агрегированного показателя достоверности* (АПД) денонсируемых БСС.

Для денонсации БСС с оценкой истинности ***b****ij* = 〈*bij*+, *bij*−〉, которое в РФК при СПД ***α*** = 〈*α*+, *α*−〉 признано достоверным, достаточно установить порог доверия к эмпирическим данным, определяемый суммой вектора ***α*** и *вектора ужесточения порога* ***T****TV* при условии

***α*** + ***T****TV* = 〈*bij*+ + *ε*+, *bij*− – *ε*−〉, (5)

где сколь угодно малые *ε*+ ≥ 0 и *ε*− ≥ 0 одновременно не равны нулю. Полагая правую часть (5) практически равной ***b****ij*, обнаруживаем, что длина вектора ***T****TV* в метрике L1 сколь угодно близка к *достоверности* денонсируемого БСС, которая в логике VTF определяется величиной (*bij*+ – *bij*−) ∈ [-1, 1] и устанавливает АПД в случае денонсации отдельно взятого БСС.

Исследована ситуация, когда для реализации сформулированной эвристики применительно к конкретному объекту *g* в РФК необходимо денонсировать *одновременно* *несколько* БСС, которые признаны истинными при пороге ***α***, а в эмпирическом ИФК оцениваются векторами истинности вида ***b****gx* = 〈*bgx*+, *bgx*−〉, где *x* ∈ *X* = *Mg*•, *Mg*• ⊆ *Mg*\*, |*Mg*•| ≥ 1. Установлено, что из трех выявленных вариантов определения вектора ужесточения, обеспечивающего «отсечение» сразу всех векторов ***b****gx*, предпочтительным является тот, координаты стока которого определяет вектор

〈**max***x*∈*X* *bgx*+, **min***x*∈*X* *bgm*−〉,

и его достоверность принимается в качестве АПД денонсируемых БСС. Таким образом, АПД является функцией достоверностей денонсируемых БСС, которая *монотонно* изменяется в зависимости от *количества* аргументов.

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

* безусловная денонсация БСС для свойств из *Mg*\*, которые *обусловливают* свойства, отсутствующие в *Mg*\*;
* выявление в *Mg*\* групп *попарно несовместимых* свойств (Н-групп – нарушителей) и *связных* подмножеств таких Н-групп. Если таких групп нет – завершение работы;
* *оптимальная редукция* каждого связного подмножества Н-групп, нарушающего ОСС (монотонность критерия АПД даёт основания для реализации *лексикографического* метода с построением *рекурсивной процедуры* сокращенного обхода конечного *дерева решений*).

Дополнительное ускорение этого метода осуществлено за счёт трансформации модели СИС. Отношение *C*(*x*, *y*) индуцирует на множестве *M* бинарное отношение взаимообусловленности *MC*: *MC*(*x*, *y*) ↔ *С*(*x*, *y*) ∧ *С*(*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*° и несовместимости *E*°, что в итоге определяет новую редуцированную модель СИС – граф, приведённый на рисунке 2, трансформируется в граф на рисунке 5.

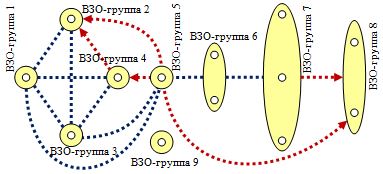


Рисунок 5 – Пример реструктурированной системы измеряемых свойств

С учётом существования отношений *E*° и *C*° под Н-группой будем понимать набор попарно несовместимых ВЗО-групп, а под О-группой – пару ВЗО-групп, связанных расширенным отношением обусловленности. Для каждой группы обосновывается связь с нормальностью содержащего их подмножества измеряемых свойств. Тогда предложенный метод нормализации с минимальной коррекцией применим для новой модели СИС.

Эффект сжатия объема обрабатываемых данных, обусловленный предложенной трансформацией модели СИС и вызывающий сокращение вычислительных затрат метода нормализации, исследован и подтвержден в имитационных статистических экспериментах, где вопрос о надлежащем количестве статистических испытаний решался на основе неравенства Хёвдинга. Рисунок 6 демонстрирует сокращение количества сущностей в редуцированной модели СИС (т.е. соотношение количества ВЗО-групп и измеряемых свойств; для получаемого выигрыша ЭС присущ затухающий экспоненциальный рост). Аналогично исследованы средние относительных сокращений количества обусловленностей ЭО и несовместимостей ЭН.

**В четвертом разделе** описано архитектурное проектирование программной лаборатории на основе известных диаграммных техник UML и смысловых карт. Приведены ключевые разработанные алгоритмы, поддерживающие формирование, редактирование и реструктуризацию СИС, работу с ОТОС и построение ИФК, а также нормализацию рабочего ФК (РФК). Продемонстрирована работа OntoWorker, начиная с построения СИС и ОТОС, заканчивая получением нормализованного ФК вида (2), пригодного для обработки методами АФП.

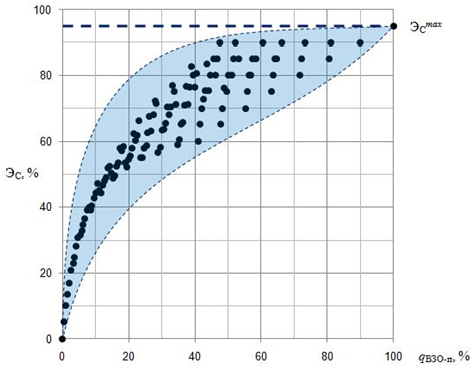


Рисунок 6 – Значения показателя эффективности сокращения количества сущностей Эс в обобщенном представлении  
системы |*M*| = 20 измеряемых свойств в зависимости от доли  
ВЗО-пар *q*ВЗО-п в её исходном представлении

Приведена схема формирования РФК для вывода ФП (рисунок 7), сложившаяся при программной реализации разработанных моделей и методов:

* 1, 4, 8, 17 — осмысление субъектом, находящимся на определенном этапе решения ОАД-задачи, либо результата *предшествующего*, либо результата *любого из последующих* шагов анализа;
* 2, 5 — формирование и редактирование соответственно СИС и ОТОС, включая концептуальное шкалирование измеряемых свойств;

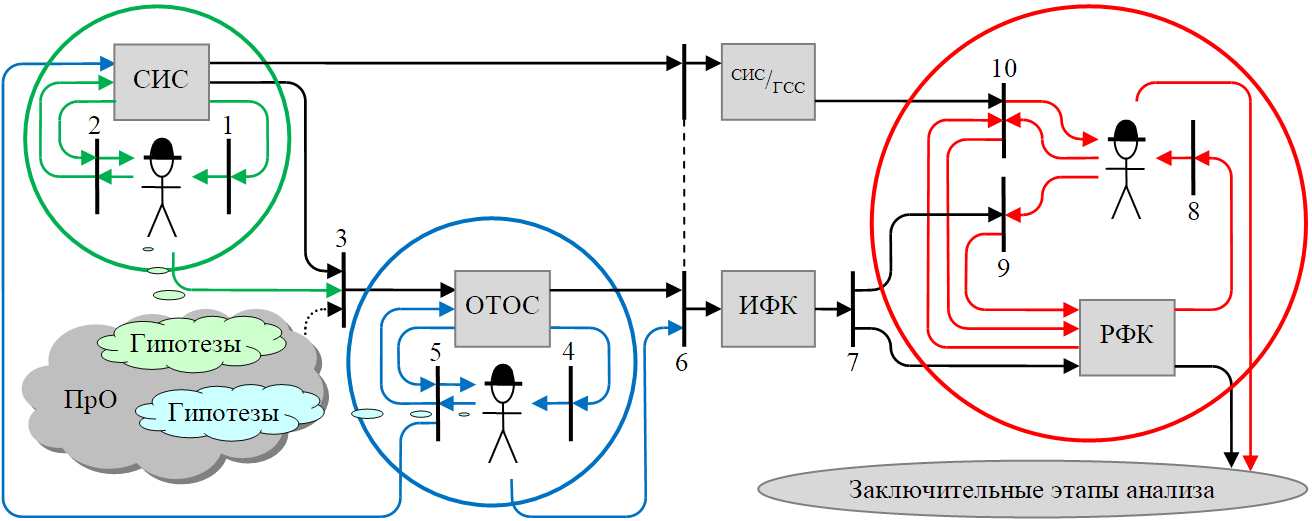


Рисунок 7 – Схема формирования рабочего формального контекста для вывода формальных понятий и выполняемые при этом когнитивные акты

* 3 — измерение (в широком смысле) свойств объектов ПрО;
* 6 — построение СИС/ГСС и ИФК, где СИС/ГСС – редуцированное представление СИС, а *группами сопряженных свойств* (ГСС) собирательно именуются ВЗО-, О- и Н-группы;
* 7 — исходное отождествление РФК с ИФК;
* 9 — дефаззификация ИФК;
* 10 — нормализация РФК.

**В пятом разделе** приведены результаты апробации разработанных методов и программных средств. Их целевое назначение - поддержка ОАД, где формирование контекста является начальным этапом, первичной обработкой эмпирических данных. Вместе с тем задача формирования контекста имеет самостоятельное значение, и в данном разделе рассматриваются примеры соответствующих приложений.

Вполне использование предложенных моделей и методов иллюстрирует поддержка *построения когнитивных карт коллективом экспертов*. Эксперты выступают в качестве «процедур измерения», объектами служат пары факторов, которые по мнению экспертов отражают динамику моделируемой системы, а нормализованный однозначный ФК, выводимый из совокупности многопланово противоречивых данных, фиксирующих взгляд разных экспертов, является матрицей смежности искомого знакового графа.

Использование разработанной методики в случае одного объекта в ОТОС имеет характер *семантической идентификации объекта*, которая являет самостоятельную ценность в системном анализе. Это демонстрируется при решении задачи *структурного анализа и синтеза технических решений* (ТР).

Содержание структурного анализа и синтеза ТР определяется тем, что на практике функциональный потенциал и структура создаваемого артефакта объективно тесно взаимосвязаны, а реализация структурных элементов многовариантна. Для поддержки построения соответствующей функционально-структурной модели артефакта существует формализованный подход – *морфологический анализ* ТР.

Морфологический анализ области, где предстоит выбрать решение, состоит в формировании таблицы (пример – таблица 1), описывающей «*морфологическое пространство*» выбора ТР как множество дискретных точек, каждая из которых определяет комбинацию-*трансверсаль* *вариантов реализаций* (ВР) *всех* *функционально-структурных подсистем* (ФСП) проектируемого артефакта. Синтез ТР сводится к реализации некоторой *процедуры рационального выбора* определённой точки морфологического пространства.

В работе предложено следующее формальное описание морфологического пространства:

* множество ФСП *W* = {*wi*}*i*= 1,…,*n*, где *n* > 0;
* множество ВР ФСП *V* такое, что *V* = ∪*i*= 1,…,*n**Vi*, **, *ki* > 0, и ∀ *i*, *j*, *i* ≠ *j*: *Vi* ∩ *Vj* = ∅;
* множество точек-трансверсалей *T*, у которого |*T*| = Π*i*= 1, 2,…,*n**ki*;
* бинарное отношение *взаимообусловленности*, которое полно на множестве *W* – *MC*: *W* × *W* → **True** (в морфологическом анализе и синтезе это отношение явно не фиксируется, но фактически ТР состоятельно тогда и только тогда, когда реализованы все ФСП проектируемого артефакта);
* бинарное отношение *несовместимости* (попарной несовместимости) *E*: *V* × *V* → {**True**, **False**}, – которое здесь конфигурируется так, что

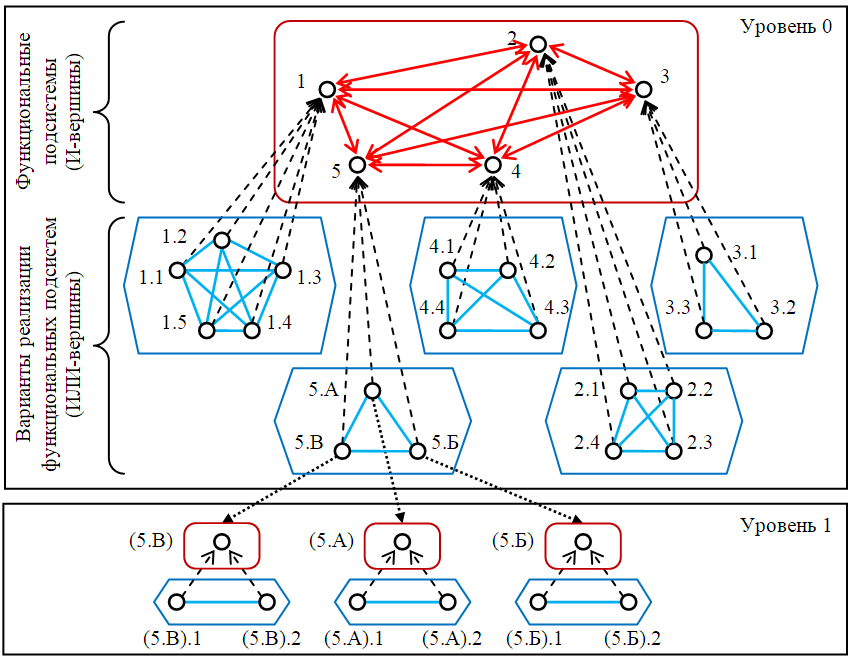
∀ *i*∀ *x* ∈ *Vi*: *E*(*x*) = *Vi* \ {*x*} (3)

* бинарное отношение «*является вариантом* *реализации*» *R*: *V* × *W* → {**True**, **False**}, – характерная особенностью которого состоит в том, что ∀ *i*∀ *x* ∈ *Vi*: *R*(*x*) = *wi* (т.е. *R* – отношение между каждым ВР некоторой ФСП и этой ФСП);
* бинарное отношение «*ТР включает ВР ФСП*» *S*: *T* ×*V* → {**True**, **False**}, – причем ∀ *x* ∈ *T*, ∀ *i*: (|*S*(*x*)| = *n*) ∧(|*S*(*x*) ∩ *Vi*|) = 1.

Таблица 1 – Морфологическое пространство компоновок холодильной машины

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ФСП | Возможные варианты реализации функционально-структурной подсистемы | | | | | | | | | | | | | | | | |
| 1*Хладагент* | 1.1 Аммиак | | 1.2 Пропан | | | | | 1.3 Пропилен | | | 1.4 Изобутан | | | | | 1.5 Тетрафторэтан | |
| 2 *Компрессор* | 2.1 Поршневой | | | 2.2 Ротационный | | | | | 2.3 Центробежный | | | | | | 2.4 Тепловой (абсорбер) | | |
| 3 *Конденсатор* | 3.1 Жидкостного испарения | | | | | 3.2 Воздушного испарения | | | | | | 3.3 Испарительного охлаждения | | | | | |
| 4*Дроссель* | 4.1 Капиллярная трубка | | | | 4.2 Автомати-ческий клапан | | | | 4.3 Терморегули-руемый вентиль | | | | | 4.4 Электронно-управляемый вентиль | | | |
| 5 *Испаритель* | 5.1  Трубчатый медный | 5.2  Трубчатый стальной | | | | | 5.3 Плас-тинчатый медный | | | 5.4 Плас- тинчатый стальной | | | 5.5  Ребристый медный | | | | 5.6  Ребристый стальной |

Развитие морфологического анализа, осуществляющее идею многоуровневых морфологических представлений достигается путём введения бинарного отношения между множеством ВР ФСП, описанными на уровне *q* и множеством ФСП на уровне *q* + 1, *q* ≥ 0. Это отношение «*содержит подсистему*» *P*: *V*(*q*) × *W*(*q*+1) → {**True**, **False**}, – отличающееся тем, что ∀ *x*, *y* ∈ *V*: *P*(*x*) ≠ *P*(*y*), и ∀ *x* ∈ *W*(*q*+1): |*P*(*x*)| = 1. В совокупности с ранее введёнными формализмами получаем представление морфологического пространства в виде широко известной математической модели И/ИЛИ-дерева (графа). Соответствующая многоуровневая модель для холодильной машины приведена на рисунке 8.



|  |
| --- |
| - оси и точки на осях морфологического пространства;  - дуга «взаимообусловленность»; - ребро «несовместимость»;  - дуга «*является вариантом реализации*»; - дуга «*содержит подсистему*» |

Рисунок 8 – Многоуровневая морфологическая модель холодильной машины в форме  
И/ИЛИ графа с детализацией отношений между элементами морфологического пространства  
(здесь: 5.А, 5.Б, 5.В – трубчатая, пластинчатая и ребристая конструкции испарителя;  
(5.\*) – материал, пригодный для изготовления испарителя «\*»; (5.\*).1 – медь; (5.\*).2 – сталь)

Обычно на ВР ФСП накладываются различные – физические, патентные, технические, технологические,гуманитарные и др. –ограничения. Подобные обстоятельства в морфологическом плане могут быть интерпретированы как *ограниченная сочетаемость* ВР *различных* ФСП в *одном* ТР, которая проявляется либо как предусловие, либо как постусловие выбора ВР одной ФСП – *x* ∈ *Vi*, – на выбор ВР другой ФСП – *y* ∈ *Vj*, *j* ≠ *i*.

Для отражения возможной несовместимости ВР различных ФСП, во-первых, уточнено условие (3):

∀ *i*∀ *x* ∈ *Vi*: *E*(*x*) = *Vi* \ {*x*} ∪ *Y*, (4)

где *Y* ⊆ *V* \ *Vi*, т.е. в морфологическом пространстве ВР некоторой ФСП несовместим не только с другими ВР этой ФСП, но, возможно, и с некоторыми ВР других ФСП. Во-вторых, для описания потенциальной обусловленности ВР различных ФСП в модель введено соответствующее отношение – *C*: *V* × *V* → {**True**, **False**}. Таким образом, обнаружилось, что в общем случае на множестве ВР заданы *ограничения существования* аналогичные ОСС в ОАД.

Задача структурного синтеза допустимых ТР, представленных многоуровневыми морфологическими моделями, с позиций ОАД трактуется как задача семантической идентификации объекта на каждом ИЛИ-уровне и сводится при этом к выявлению «нормальных» подмножеств ВР каждой ФСП при условии реализации всех ФСП.

Установлено, что при наличии информации об издержках и/или выгодах ВР ФСП для рационального структурного синтеза рационального ТР пригоден разработанный метод нормализации ФК. Для иллюстрации этого положения рассмотрен пример синтеза холодильной машины, когда, во-первых, её морфологическая модель (таблица 1) дополнена ограничениями существования, учитывающими разрушительное действие аммиака на медную арматуру, а из хладагентов, лишь аммиак пригоден для постройки абсорбционного холодильника, а, во-вторых, известны нормированные значения выгод и издержек ВР рассматриваемых ФСП: 〈*b+*, *b–*〉, *b+*, *b–*∈ [0, 1].

Рисунок 9 иллюстрирует найденное предложенным методом оптимальное решение (по критерию превышения выгод над издержками) задачи синтеза ТР на примере холодильной машины. Слева представлены ВР ФСП и ограничения существования: морфологическая таблица 1 устанавливает несовместимость ВР каждой ФСП, а дуга и ребра описывают ограниченную сочетаемость ВР разных ФСП. В средней части рисунка 9 приведены нормированные значения выгод и издержек, связанных с включением в ТР отдельных ВР ФСП (в процентах), справа демонстрируются компоненты найденного оптимального ТР, найденного методом семантической идентификации объекта.

**В заключении** диссертации констатируется достижение цели диссертационного исследования и формулируются основные результаты работы.

**Основные выводы и результаты работы**

1. Обоснована необходимость комбинирования существующих методик вывода формальных понятий, выявлено единство их логической природы и выдвинута идея осуществления гибридизации методик за счет отказа отчасти наименее достоверной эмпирической информации.
2. Разработаны прагматические модели системы измеряемых свойств и протокола измерений «объект-свойство», структурирующие соответственно гипотетические представления и эмпирическую информацию о ПрО, а также метод интерпретации и консолидации эмпирических данных с использованием многозначной векторной логики VTF с целью построения нестрогого ФК.

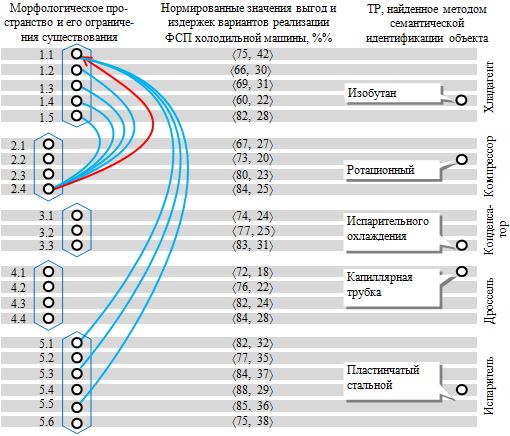


Рисунок 9 – Пример решения задачи синтеза облика артефакта, полученного методом  
семантической идентификации объекта на основе учета ограниченной сочетаемости вариантов реализации функционально-структурных подсистем в форме ограничений существования

1. Разработана методика построения рабочего однозначного ФК, включающая дефаззификацию нестрогого ФК и эвристический метод его нормализации, учитывающий ограничения существования свойств.
2. Разработано алгоритмическое обеспечение подготовки ФК для вывода формальных понятий и программные средства ОАД.
3. Разработанные методы и средства апробированы при решении различных научно-технических системологических задач и, в частности, задачи структурного анализа и синтеза ТР.

**СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ РАБОТЫ**

**Публикации в изданиях, рекомендованных ВАК РФ**

1. **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.
2. **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.
3. Смирнов С.В., **Семенова В.А.** Эффективность редукции ограничений существования свойств в задаче идентификации признаков объекта // Информационные и математические технологии в науке и управлении. 2023. №2(30). С. 5-13.
4. **Semenova V.A.**, Smirnov S.V. Determination and Processing of Properties Existence Constraints in Ontological Data Analysis // Информационные технологии. 2022. Т. 28, №12. С. 636-643.
5. **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.
6. **Семенова В.А.** Опыт бикластеризации данных о сортах сельскохозяйственных культур // Известия Самарского научного центра Российской академии наук, Т. 22, №1, 2020. С. 86-92.
7. Самойлов Д.Е., **Семенова В.А.**, Смирнов С.В. Структура системы измеряемых свойств у объектов многомерного наблюдения и экспериментов // Вестник Самарского государственного технического университета. Серия «Технические науки». 2018. №3(59). С. 56-71.
8. 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.

**Публикации в изданиях, индексируемых в международной базе данных Scopus**

1. **Semenova V**, Smirnov S. Revealing attributes of an object based on incomplete and inconsistent empirical data // Proc. VIII Int. Conf. on Information Technology and Nanotechnology. IEEE Xplore. 2022. P. 1-5.
2. **Semenova V.A.**, Smirnov S.V. Extended methodology for deriving formal concepts // Proc. Int. Conf. «Marchuk Scientific Readings 2021». Journal of Physics: Conf. Series. 2021. Vol. 2099, 012026. P. 1-9.
3. Samoylov, D.E., **Semenova V.A.**, Smirnov S.V. Defuzzification of the initial context in Formal Concept Analysis // Proc. of the Data Science Session at the V Int. Conf. on Information Technology and Nanotechnology. CEUR Workshop Proceedings. 2019. Vol. 2416. P. 1-9.
4. Samoylov, D.E., **Semenova V.A.**, Smirnov S.V. Multilevel recursive model of properties existence constraints in machine learning // Proc. IV Int. conf. Information Technology and Nanotechnology. Journal of Physics: Conf. Series. 2018. Vol. 1096, 012096. P. 1-11.
5. **Semenova V.A.**, Smirnov S.V. Intelligent analysis of incomplete data to building formal ontologies // Proc. Int. conf. Information Technology and Nanotechnology. CEUR Workshop Proceedings. 2016. Vol. 1638. P. 796-805.

**Публикации в других изданиях**

1. **Семенова В.А.** Опыт использования программной лаборатории для онтологического анализа данных // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2022): сб. науч. статей XI междунар. н.-т. и н.-метод. конф. Т. 2. – СПб: СПбГУТ. 2022. С. 485-489.
2. **Семенова В.А.**, Смирнов С.В. Механизм нормализации эмпирического контекста в онтологическом анализе данных // Системная инженерия и информационные технологии. 2021. Т. 3, №3(7). С. 45-52.
3. **Семенова В.А.** Эвристика и численный метод нормализации эмпирического VTF-контекста в онтологическом анализе данных // Информационные и математические технологии в науке и управлении. 2021. №2(22). С. 61-69.
4. **Семенова В.А.** Выбор логической модели для представления эмпирической информации в онтологическом анализе данных // Тр. XII Всеросс. н.-т. конф. аспирантов, студентов и молодых ученых «Информатика и вычислительная техника». - Ульяновск: УлГТУ, 2020. С. 205-210.
5. Лещева Д.В., **Семенова В.А.** Управление данными при кластеризации объектов многомерных наблюдений и экспериментов // ПИТ 2020: тр. междунар. н.-т. конф. - Самара: СамНЦ РАН, 2020. С. 127-132.
6. Лещева Д.В., **Семенова В.А.**, Смирнов С.В. О поддержке когнитивных актов в задачах онтологического анализа данных // ПУМСС: тр. XXI междунар. конф. - Самара: «Офорт», 2019. Т. 2. С. 260-263.
7. Самойлов Д.Е., **Семенова В.А.**, Смирнов С.В. Паттерны структурной организации системы измеряемых свойств в онтологическом анализе данных // Проблемы управления и моделирования в сложных системах: тр. XX междунар. конф. - Самара: «Офорт», 2018. С. 358-366.
8. Лещева Д.В., **Семенова В.А.**, Смирнов С.В. Пользовательский интерфейс программной лаборатории для онтологического анализа данных // ПИТ 2018: тр. междунар. н.-т. конф. - Самара: СамНЦ РАН, 2018. С. 479-482.
9. Самойлов Д.Е., **Семенова В.А.**, Смирнов С.В. Модель ограничений существования свойств в онтологическом моделировании // ПУМСС: тр. XIX междунар. конф. - Самара: «Офорт», 2017. С. 468-473.
10. **Семенова В.А.** Построение формальных онтологий на основе противоречивых эмпирических данных // ПИТ 2017: тр. междунар. н.-т. конф. - Самара: СамНЦ РАН, 2017. С. 164-168.
11. **Семенова В.А.**, Смирнов С.В. Семантическая идентификация объекта в задачах концептуального моделирования // ПИТ 2016: тр. междунар. н.-т. конф. - Самара: СамНЦ РАН, 2016. С. 330-333.
12. Зубцов Р.О., **Семенова В.А.**, Смирнов С.В. Алгоритмическое и программное обеспечение онтологического анализа данных // Материалы VI междунар. н.-т. конф. «Открытые семантические технологии проектирования интеллектуальных систем = OSTIS». Минск: БГУИР, 2016. С. 83-88.

**Свидетельства о регистрации программ для ЭВМ**

1. Свидетельство о государственной регистрации программы для ЭВМ № 2022669423. Программная лаборатория для онтологического анализа данных OntoWorker / **В.А. Семенова**, С.В. Смирнов. Зарегистрировано 19.10.2022. – М.: Роспатент, 2022.