

На правах рукописи



**Туральчук Константин Анатольевич**

**Метод естественной системной классификации**

Специальность 05.13.01 — «Системный анализ, управление и обработка информации (технические системы)»

Автореферат  
диссертации на соискание учёной степени  
кандидата технических наук

Санкт-Петербург — 2020

Работа выполнена:

на кафедре автоматики и процессов управления ФГАОУ ВО «Санкт-Петербургский Государственный Электротехнический Университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» и в Высшей школе интеллектуальных систем и суперкомпьютерных технологий ФГАОУ ВО «Санкт-Петербургский Политехнический Университет Петра Великого».

Научный руководитель: **Фомин Борис Федорович**  
доктор технических наук, профессор

Официальные оппоненты: **Соколов Борис Владимирович**,  
доктор технических наук, профессор,  
руководитель лаборатории информационных технологий в системном анализе и моделировании, ФГБУН «Санкт-Петербургский институт информатики и автоматизации» Российской академии наук, г. Санкт-Петербург

**Коршунов Игорь Львович**,  
кандидат технических наук, доцент,  
заведующий кафедрой информационных систем и технологий, ФГБОУ ВО «Санкт-Петербургский государственный экономический университет», г. Санкт-Петербург

Ведущая организация: ВУНЦ ВМФ «Военно-морская академия», г. Санкт-Петербург

Защита состоится 17 февраля 2020 г. на заседании диссертационного совета Д 212.238.07 при ФГАОУ ВО «Санкт-Петербургский Государственный Электротехнический Университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» по адресу: 197376, Санкт-Петербург, ул. Проф. Попова, д. 5.

С диссертацией можно ознакомиться в библиотеке ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» и на сайте университета [www.etu.ru](http://www.etu.ru).

Автореферат разослан 16 декабря 2019 года.

Ученый секретарь  
диссертационного совета  
Д 212.238.07,  
канд. техн. наук



Цехановский В. В.

## Общая характеристика работы

### Актуальность темы.

Основным направлением развития информационных технологий становится наука о данных. В настоящее время наука о данных развивается преимущественно за счет методов машинного обучения и Data Mining. Применительно к классификационным задачам такие методы в основном направлены на решение прогностических задач - настраиваются на заданные классы в обучающей выборке для распознавания классов новых объектов. Иной подход заключается в построении естественных классификаций, направленных на решение объяснительных задач.

В противоположность искусственному (инструменталистскому) подходу к классификации подход естественной классификации, применяемый в фундаментальных дисциплинах, состоит в открытии (реконструкции) естественной системы. Идея естественной классификации уже несколько столетий направляет действия классификаторов.

Большое внимание проблематике естественной классификации уделяли исследователи, работающие в разных областях знания: Мейен С. В., Шрейдер Ю. А., Кожара В. Л., Розова С. С., Субботин А. Л., Любищев А. А., Любарский Г. Ю., Покровский М. П., Павлинов И. Я., Витяев Е. Е., Соукал Р., Снит П., Парочиа Д., Невил П.

Наибольший опыт научных исследований по проблеме естественной классификации накоплен биологической наукой. Проблема естественной классификации возникла в биологии как проблема реальности классов и реальности признаков классов. Биология предложила и развивает подход к решению данной проблемы, в основе которого лежат понятия онтологии предметной области, интенционала и экстенционала классов, мерономии и таксономии. В рамках этого подхода пока остаются без ответа вопросы определения меронов (естественных признаков класса) и построения воспроизводимого метода познания онтологии предметной области.

Ключевой проблемой естественной классификации стало познание онтологии предметной области исследования. Эта проблема была решена в рамках системологии. Физика открытых систем (ФОС), разработанная Фоминым Б. Ф. и Качановой Т.Л., обеспечила познание общей онтологии открытых природных, общественных, антропогенных, кибер-физических и сложных технических систем с сотнями и тысячами показателей на основе больших массивов полимодальных гетерогенных эмпирических данных о системах. С появлением и развитием ФОС актуальная проблема естественного классифицирования получила реальную перспективу решения.

Следуя этой перспективе, надо было в рамках ФОС на базе знания общей онтологии систем предложить новую концепцию, и создать на ее основе новый метод рационального естественного классифицирования в

сложных предметных областях. При таком подходе знание об общей онтологии системы, полученное ФОС, должно порождать знание об онтологии предметной области, необходимое для решения задачи естественного классифицирования.

**Целью** диссертации стала разработка научного метода естественной системной классификации (ЕСК) на основе знания онтологии системы.

Для достижения поставленной цели в диссертации решены следующие **задачи**:

1. Обосновать актуальность разработки метода ЕСК для сложных предметных областей на основе знания онтологии систем.
2. Выполнить анализ ценности знания онтологии системы, полученного ФОС, и оценить перспективы его использования для решения задачи ЕСК.
3. Разработать концепцию и модель естественного классифицирования на основе знания онтологии систем.
4. Разработать метод ЕСК, способный выявлять и оформлять системно-обусловленные естественные признаки классов в условиях гетерогенности классифицируемых объектов, осуществлять логическую реконструкцию интенционалов классов, создавать когнитивные репрезентации классов.
5. Разработать программный комплекс, реализующий метод ЕСК, как элемент конструктивного компонента технологической платформы ФОС.
6. Провести апробацию метода ЕСК с целью аттестации его воспроизводимости и возможностей эффективного использования в разных предметных областях.

**Объект исследования.** Естественное классифицирование в сложных предметных областях.

**Предмет исследования.** Выявление естественных признаков и естественных классов на базе онтологического знания о системах.

**Научная новизна:**

1. Впервые установлена возможность применить эталоны состояний собственных качеств системы, открытые ФОС, как системные гомологи естественной классификации, способные определять и различать классы в сложных предметных областях.
2. Впервые предложена концепция ЕСК, основанная на экспликации знания общей онтологии системы в интенционалы классов, обеспечившая разработку единой системы базовых понятий и отношений между понятиями модели метода ЕСК.
3. Впервые создан метод ЕСК, отличающийся от существующих методов классификации: способностью выявлять естественные системно-обусловленные признаки классов; способностью строить на основе этих признаков иерархические структуры моделей классов

(интенционалы классов); способностью получать репрезентанты (прототипы) классов и раскрывать морфологию каждого класса.

### **Практическая значимость:**

1. Метод ЕСК создан для применения к природным, антропогенным, кибер-физическим и техническим системам, исходно заданным большими массивами многосортных, полимодальных, неоднородных эмпирических данных.
2. Разработаны алгоритмы вычисления формальных объектов, атрибутов объектов и отношений между объектами метода ЕСК. Сложность построенных алгоритмов зависит от степени гетерогенности классификационного поля (КП) и от правильности поставленных экспертами классов. Рост сложности ограничен гиперболическим законом Ципфа-Мандельброта для вычисляемых коррелированных объектов.
3. Программный комплекс, реализующий метод ЕСК, включен в состав конструктивного компонента технологической платформы ФОС, готов к практическому применению для решения задачи естественной системной классификации.

**Методология и методы исследования.** Методологические основания, метод и информационные технологии ФОС, таксономия, меронимия, теория множеств, логика, теория вероятностей и математическая статистика, лексикология, теория графов, квалиметрия.

### **Основные положения, выносимые на защиту:**

1. Онтологическое знание о системах, представляющих классификационный универсум, имплицитно содержит знание, на основе которого выявляются системные гомологи естественных признаков и естественные классы онтологически единых объектов.
2. Концепция и модель системы классифицирования, основанные на интенциональном и экстенциональном подходах, обеспечивающие экспликацию знания системной онтологии классификационного поля (далее – КП) в архетипические и таксономические компоненты классов.
3. Метод ЕСК как конструктивное воплощение модели системы классифицирования, обеспечивший преобразование понятий модели в формальные конструкты, а отношения между понятиями в вычислительные процедуры построения семантических и денотативных компонентов классов.

### **Достоверность:**

- достоверность научных оснований концепции и модели системы естественного классифицирования определяют: научный метод ФОС, обеспечивающий получение знания об общей онтологии системы; семиотический треугольник Г. Фреге, задающий связь классов КП с их сигнификатом и денотатом; меронимический и

- таксономический подходы, положенные в основу разработанного формального аппарата классифицирования;
- достоверность результатов предложенного метода определяют: качественные оценки ценности онтологического знания о КП как системе; критерии оценивания качества полученных интенционалов и экстенционалов классов (ранги признаков, ранги формальных дескрипторов, компактность архетипа и др.);
  - достоверность вычислительных процедур метода ЕСК определяет корректное применение методов математической статистики, теории множеств, теории графов;
  - достоверность положений и выводов диссертации определяют успешная верификация результатов применения метода к реальным данным в практических решениях классификационных задач из разных сложных предметных областей, а также сравнения с результатами, полученными методами искусственной классификации.

**Апробация работы.** Основные положения и результаты диссертации апробированы на научных симпозиумах, семинарах и конференциях: на всероссийской научной конференции по проблемам управления в технических системах (ПУТС-2015); на IX конференции «Информационные технологии в управлении (ИТУ–2016)»; на научном семинаре «ИКТ», СПИИРАН, (11 ноября 2016 г.); на XXI международной конференции по мягким вычислениям и измерениям (SCM-2018); на XIII международном симпозиуме "Intelligent Systems – 2018"(INTELS-18).

**Личный вклад.** На постановочном этапе создания метода ЕСК: постановка и проведение вычислительных экспериментов, в результате которых получено обоснование имплицитной связи онтологического знания о системе и онтологии классов. В экспериментальных исследованиях: анализ достоинств и недостатков методов искусственной классификации при их применении в сложных предметных областях, обоснование актуальности разработки воспроизводимого метода ЕСК. На этапе разработки модели классифицирования: выявление системных гомологов и их экспликация в элементы онтологического знания о классах. В разработке метода ЕСК: построение конструктивных форм понятий модели классифицирования и алгоритмов их вычислений. На этапе верификации метода ЕСК: подбор задач классификации для наилучшей аттестации метода, выполнение вычислений, оценивание результатов.

**Публикации.** Основные результаты по теме диссертации изложены в 9 печатных изданиях, 4 из которых изданы в журналах, рекомендованных ВАК, 2 — в научно-технической базе SCOPUS, 3 — в материалах конференций.

## Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, формулируется цель, ставятся задачи работы, излагается научная новизна и практическая значимость представляемой работы.

**В первой главе** рассматривается проблематика естественной классификации и ограничения искусственной классификации. Наибольшее распространение в области обработки данных получили методы машинного обучения и Data Mining (далее ML & DM). Такие методы являются аппроксимационными - рассматривают исследуемую систему как черный ящик с неизвестной внутренней природой и выполняют настройку той или иной модели таким образом, чтобы обеспечить наилучшее приближение поведения модели к известной обучающей выборке. Методы ML & DM хорошо зарекомендовали себя в таких областях, где объекты разных классов достаточно хорошо различимы (распознавание устойчивых шаблонов). В сложных предметных областях (биосистемы, системы техносферы, киберфизические системы, технические системы) искусственные классификации малопродуктивны ввиду гетерогенности объектов, обусловленной большим разнообразием элементов, структур и внутренних процессов.

В искусственных классификациях заданные классы рассматриваются как правильные, не подвергаются сомнению и критике. Во многих областях знания при исследовании малоизученных явлений классы могут содержать неточности, ошибки. Между классами возможны пограничные случаи, когда объект обладает свойствами разных классов. Класс может быть неоднороден, состоять из подклассов. Вопрос о реальности класса требует системного исследования.

Сформулированы основные вызовы естественной классификации: существенность онтологического основания классификации; реальность применяемых признаков и реальность выделенных классов; когнитивно-содержательный потенциал; систематизация.

Ответом на вызовы естественной классификации в диссертационной работе является следующий подход: ФОС извлекает знание об общей онтологии «системы КП» из ее эмпирического описания. При определении «системы КП» признак класса во внимание не принимается, поэтому состояния системы не имеют прямой информации о классах. Задача ЕСК решается в опоре на знание общей онтологии «системы КП». На этой базе ЕСК выявляет все системно-обусловленные естественные признаки (формализованные дескрипторы) классов и создает на их основе интенционалы и экстенционалы классов.

**Вторая глава** посвящена обзору методологических оснований ФОС (аксиом, принципов, симметрий).

Научный метод ФОС получает полное завершённое знание об общей онтологии системы из больших массивов гетерогенных эмпирических данных. Знание онтологии системы имплицитно содержит знание о классах. Для естественного классифицирования в сложных предметных областях метод ФОС предоставляет: полное множество моделей собственных качеств системы, модели эталонных состояний ситемы в каждом ее собственном качестве; модели реконструкции всех актуальных состояний системы. На базе этого знания об онтологии «системы КП» метод ЕСК создает системные модели референтов системы, системно-обусловленные естественные признаки классов, интенционалы классов, экстенционалы классов.

В диссертации представляется аналитическое ядро технологий ФОС, обеспечивающих автоматическую генерацию полного завершённого онтологического знания об онтологии систем.

В главе приводятся описания форм представления «системы КП», применяемых на разных этапах жизненного цикла «производства» онтологического знания: «система в данных», «система в отношениях», «система в эталонах состояния собственных качеств», «система в эталонах», «система в формах воплощения эталонов», «система в состояниях».

Система исходно представлена как «система в данных»:  $D = \langle P, O, I_v \rangle$ , где  $P = \{p_i\}$  - множество показателей (атрибутов), характеризующих состояния «системы КП»,  $O = \{o_j\}$  - множество объектов (референтов), представляющих состояния «системы КП», функция  $I_v : (p_i, o_j) \in Values_i$ , где  $Values_i$  - множество допустимых значений показателя  $p_i$ .

Каждый объект является представителем (референтом) определенно-го класса:  $\forall o_j \in O : Class(o_j) \in C$ . Информация о классах не используется для получения общего онтологического знания о системе, а привлекается на этапе экспликации - переходе к онтологическому знанию о классах.

«Система в эталонах состояния собственных качеств» (далее кратко - «система в эталонах») – основа для получения полных наборов системных гомологов в методе ЕСК. Каждый эталон состояния представлен уникальной атрибутированной структурой показателей с определенной морфологией. Он выражает состояние одного уникального собственного качества системы как части единого целого и всего целого в условиях этой части.

Каждый эталон  $m_j \in M$  характеризуется множеством показателей  $Params(m_j) \subseteq P$ , множеством объектов  $Objects(m_j) \subseteq O$ . Каждый показатель эталона  $p_i \in Params(m_j)$  характеризуется атрибутами, раскрывающими его системную роль, смысловую активность, системный и предметный веса. В методе ЕСК используются следующие атрибуты показателя: системная роль  $Role(p_i, m_j) \in \{Special, Kernel, Locality\}$ ; уровень значений  $Level(p_i, m_j) \in \{Low, High\}$ ; область допустимых значений показателя  $Value(p_i, m_j) \subset Values_i$ .



«Система в формах воплощения эталонов» представляет отображение моделей эталонов состояний на референты классов.

«Система в состояниях» (полное множество моделей реконструкций эмпирических описаний референтов классов) представляет каждое актуальное состояние системы как единое целое, порожденное уникальной «сборкой» эталонов состояний, объясняющей внутрисистемные механизмы, детерминирующие данное состояние.

В каждой реконструкции состояния значение каждого показателя моделируется определенным набором воплощенных эталонов состояния, формирующих изменчивость этого показателя. Значения показателей кодируются уровнями величин на специальной шкале уровней:  $\forall p_i \in P, \forall o_j \in O : RLevel(p_i, o_j) \in LevelScale$ , где шкала *LevelScale* содержит 17 уровней значений от низкого до высокого. В методе ЕСК используется шкала с двумя уровнями значений  $RLevel^*(p_i, o_j) \in \{Low, High\}$ .

Полное множество реконструкций актуальных состояний служит базой решения задачи естественного классифицирования.

**В третьей главе** рассматриваются концепция метода ЕСК, модель метода ЕСК, основные объекты метода ЕСК и функциональность метода ЕСК.

Онтологическое знание о системе, применяемое методом ЕСК, включает знание собственных качеств системы, знание эталонов состояний собственных качеств, знание реконструкций актуальных состояний системы. Никакие из этих элементов знания не содержат в явном виде информацию о классах.

Главную идею метода ЕСК определяют следующие положения:

1. КП – система. Онтологическое знание о «системе КП» образует ресурс знания для решения проблемы ЕСК;
2. Эталоны состояний собственных качеств, способные различать классы, становятся системными гомологами классов;
3. На базе системных гомологов определяются системно-обусловленные существенные признаки каждого класса в отдельности, эти признаки различают классы и объединяют референты одного и того же класса;
4. Существенные признаки каждого класса связаны системой отношений архетипического характера (общее-частное, род-вид, часть-целое);
5. Интенционал класса порождает экстенционал класса и его морфологию.

Метод ЕСК формируется на основе когнитивной модели классификации – системы базовых понятий, воплощающих идею ЕСК, рисунок 1.

*Понятия модели метода ЕСК* развернуты по двум категориям. Категория «Представление классов» (градации «Целое», «Частное», «Конкретное») задает три формы выражения смыслов классов.

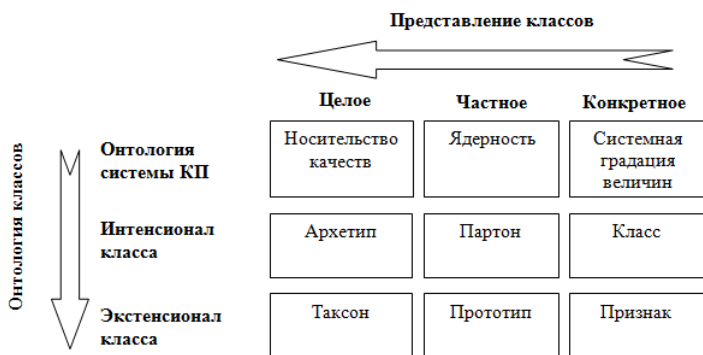


Рис. 1 — Модель ЕСК

Категория «Онтология классов» (градации «Онтология Системы КП», «Интенционал», «Экстенционал») представляет этапы смыслового анализа классов. Смысловым центром модели является понятие «Партон» - системно-обусловленный естественный признак класса. «Архетип» выражает иерархическую структуру партонов, возникающую на базе отношений между партонами. «Прототип» выражает идеальный объект класса, обладающий всеми существенными признаками класса.

*Объекты метода ЕСК.* Ключевые положения метода воплощены в его конструктивных объектах и отношениях между этими объектами. Каждому понятию модели соответствует вычислимый объект метода: партону – формализованный дескриптор, архетипу – структурно-смысловая модель (ССМ) класса, прототипу – репрезентант.

Метод осуществляет естественную системную классификацию последовательно раскрывая КП в четырех его представлениях: «КП в референтах классов», «КП как система», «КП в проблемном пространстве», «КП в пространстве решения», рисунок 2.

«КП в референтах классов» и «КП как система» определяются на основе научного метода ФОС и служат онтологической базой решения задачи ЕСК. «КП в пространстве решения» включает семантический (интенционал КП) и денотативный (экстенционал КП) компоненты. «КП в проблемном пространстве» отвечает на вопрос о возможности решения задачи ЕСК на основе знания онтологии системы и характеризует качество полученного решения.

Качество решений классификационных задач определяют: полнота и представительность исходного представления КП в референтах, ценность знания онтологии КП как единой системы, правильное и полное представление интенционалов классов КП и интенционала КП в целом, совершенство процедур построения экстенционалов классов и экстенционала КП в целом.



Рис. 2 — Формы представления КП

Метод проникает в онтологическую проблему предметной области каждой классификационной задачи. Он воспринимает исходно заданный состав классов и их обусловленность исключительно как условие предметно-содержательного исследования, нацеленного на познание, понимание и рациональное объяснение научно-достоверных онтологических оснований, выражающих основные закономерности и существенные свойства классифицируемых объектов. Решение классификационной задачи раскрывает смысловое содержание классов и устройство денотата КП, дает каждому классу КП научное подтверждение или обоснованное опровержение. В первом случае решение гарантирует успешную идентификацию классов, во втором случае – служит основанием для улучшения обусловленности классов, дает полезные ориентиры для когнитивной и аналитической работы в этом направлении.

Операциональные возможности метода ЕСК характеризуют отношения между объектами метода:

*Дистрибуция ядерностей и показателей* заключается в выявлении элементов онтологического знания, способных различать классы: множество доминантных ядерностей (ядерных компонент эталонов)  $M^* \subset M$  и множество доминантных показателей  $P^* \subset P$ . Используются статистические критерии различимости и ранги характерности.

*Синтагматический анализ* направлен на генерацию формализованных дескрипторов на базе ядерностей. Формализованный дескриптор несет в себе знание о каком-то одном определенном характерном и значимом свойстве класса.

Формализованный дескриптор характеризуется следующими элементами:  $D_i = \langle M_i, P_i, O_i \rangle$ , где  $M_i$  - множество ядерностей, составляющих дескриптор,  $P_i$  - множество показателей с уровнями значений;  $O_i$  - множество референтов класса, соответствующих дескриптору (уровням значений

показателей). Формирование дескрипторов выполняется итеративно, начиная с отдельных ядерностей  $m_j \in M$ .

Ядерность или комбинация ядерностей является формализованным дескриптором, если отвечает условиям: референтной соотнесенности:  $|O_i| \geq s^*$ , где  $s^*$  - минимально-допустимая поддержка; сопряженности с классом:  $\forall o_j \in O_i : Class(o_j) = c^*$ , где  $c^* \in C$ ; доминантности:  $\exists m_j \in M_i : m_j \in M^*$ ; совместимости показателей по уровням значений:  $\forall p_j \in P_i, \forall m_k \in M_i : p_j \in Params(m_k) \implies Level(p_j, m_k) = l^*$  и  $\forall o_y \in O_i : RLevel^*(p_j, o_y) = l^*$ , где  $l^* \in \{Low, High\}$ .

Сборка формализованных дескрипторов из нескольких ядерностей обусловлена неоднородностью класса. Чем больше ранг дескриптора  $Rank(D_i) = |M_i|$ , тем менее существенным он является.

*Парадигматический анализ* обеспечивает получение конструктивного определения интенционалов классов через построение многоуровневой иерархической структуры для каждого класса. Такие структуры для диспозиции понятий парадигматического отношения «Общее-частное» получены через бинарные отношения (тождества, включения, толерантности, наследования, подобия) на множестве формализованных дескрипторов классов. Например, отношение включения для дескрипторов  $D_i$  и  $D_j$  определяется следующим образом:  $O_i \subset O_j$ .

Схематическое представление ССМ класса представлено на рисунке 3.

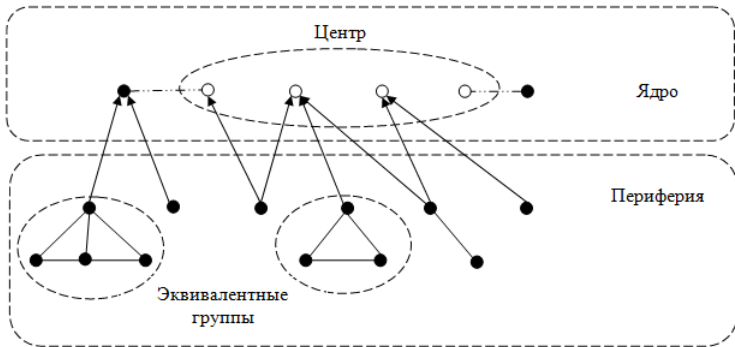


Рис. 3 — Схематическое представление ССМ

Каждая вершина соответствует одному формализованному дескриптору. Ребра соответствуют отношениям тождественности, дуги – отношениям включения, пунктирные ребра – отношениям толерантности. ССМ имеет ядро, центр и периферию. Ядро содержит формализованные дескрипторы, способные идентифицировать класс в целом. Центр определен на базе ядра через отношения толерантности между элементами ядра.

Периферия включает дескрипторы, характеризующие отдельные группы референтов класса, но не класс в целом.

*Прототипический подход* обеспечивает построение идеализированных образцов референтов–репрезентантов классов. Для каждого класса такой образец задается набором наиболее характерных системных признаков данного класса. Для получения таких признаков используется множество формализованных дескрипторов, образующих центр структурно-смысловой модели класса.

*Таксономический подход* обеспечивает конструктивное определение денотата (экстенционала) класса. Каждый репрезентант класса выделяет в КП область денотата класса. В этой области репрезентант является центром референции смысла (опорной точкой) класса. Мера близости к опорной точке задает устройство денотата. Референты всех классов получают признаки принадлежности к одной из шести областей денотата.

*Оценка качества классификации.* В проблемном пространстве метода ЕСК введены оценки качества классификации на разных уровнях раскрытия онтологии классов: уровень системной модели референтов (доля доминантных показателей; доля доминантных моделей); уровень формализованных дескрипторов (доля дескрипторов 1-ранга; доля референтов класса, покрытых формализованными дескрипторами); уровень ССМ (компактность ядра как доля дескрипторов ядра от общего числа дескрипторов класса, компактность центра как доля дескрипторов, не связанных отношением толерантности); уровень структуры классов (доля референтов ядерной и околядерной областей).

Метод ЕСК воплощен в программно-аналитический комплекс (решатель ЕСК) и является частью конструктивного компонента единой информационной технологии ФОС.

В четвертой главе приведены результаты экспериментальной апробации метода ЕСК и разработанного комплекса программ на четырех конкретных задачах из разных предметных областей.

*Задача «ToxEffects».* Естественная классификация по эмпирическим данным о токсических эффектах острых отравлений людей фосфорорганическими веществами (ФОВ). КП задачи включает референты трех классов: класс веществ, содержащих атомы фтора (V-газы, зарин, зоман) – 82 референта; класс веществ, содержащих атомы серы (карбофос, метафос, тиофос) – 118 референтов; класс веществ, содержащих атомы хлора (дихлорфос, хлорофос) – 200 референтов. Классы определены экспертами в зависимости от типов отравляющих веществ. Каждый референт классификационного поля описан набором значений 123 показателей.

*Задача «Numbers».* Естественная классификация рукописных цифр 0 - 9 по эмпирическим данным о рукописных графических образах этих цифр. КП задачи: каждая цифра (класс) представлена 200 референтами; каждый референт КП описан набором значений 396 показателей.

*Задача «Leuco».* Естественная классификация двух типов лейкемии по геномными данным, полученным от микрочипов. КП задачи включает два класса: острый миелоидный лейкоз (AML, 25 референтов), острый лимфобластный лейкоз (ALL, 47 референтов). Каждый референт КП описан набором значений 320 показателей.

*Задача «Secom».* Естественная классификация годных и забракованных полупроводниковых изделий по эмпирическим данным мониторинга производственного процесса. КП задачи включает два класса: годные изделия (1463 референта); забракованные изделия (104 референта). Каждый референт КП описан набором значений 425 показателей.

Задачи «Numbers» и «Leuco» получили удовлетворительное решение методами искусственной классификации. Задачи «ToxEffects» и «Secom» до настоящего времени не получили удовлетворительного решения.

Решение каждой из этих задач получено единым методом ЕСК и развернуто в логической последовательности этапов естественного классифицирования.

Для каждой задачи получены формализованные дескрипторы, выражающие характерные свойства классов. В таблице 1 приведены примеры формализованных дескрипторов, идентифицирующих классы КП.

Таблица 1 — Примеры формализованных дескрипторов

Система	Предикатная форма	Расширенная предикатная форма	Поддержка
ToxEffects	<b>Tachypn LR</b>	Tachypn↓ & Lowhem↑ & HEMOGLOB↓ & Distvis↑ & G5↓	38
SECOM	<b>P28 LL &amp; P455 HL</b>	P28↓ & P106↓ & P33↑ & P455↑ & P298↓ & & P439↓ & P163↓	60
Leuco	<b>P2714 LL</b>	P2714↓ & P5062↑ & P5188↑ & P1087↑	18
Numbers	<b>Pix175 LL &amp; Kar57 LR</b>	Pix175↓ & Pix162↓ & Pix163↓ & Pix147↓ & Pix149↓ & Kar57↓ & Pix105↓ & Pix239↑	90

Для каждого дескриптора указаны: предикатная форма (имя эталона или комбинация имен эталонов); расширенная предикатная форма, состоящая из показателей ядерностей дескриптора и уровня значения для каждого показателя (символ ↓ (↑) или L(H) – низкий (высокий) уровень); референтная поддержка. Распределение дескрипторов по рангам проявляет сложность различения классов в каждой задаче, 2.

Таблица 2 — Обзор данных о формализованных дескрипторах

Оценки	ToxEffects	Secom	Numbers	Leuco
Общее количество формализованных дескрипторов	4315	2171	12168	250
Распределение дескрипторов по рангам				
1-ранг	6	4	1	250
2-ранг	448	1162	12095	0
3-ранг	3852	1005	72	0
4-ранг	9	0	0	0
Референтное покрытие	98 %	96 %	100 %	100 %

Для каждого класса получена его иерархическая ССМ, в которой выделены: центр (набор конститuentов денотативного компонента КП); ядро (набор системно-обусловленных естественных признаков класса как целого); периферия (набор менее значимых системно-обусловленных признаков, характерных для отдельных групп референтов класса). Морфологические характеристики ССМ проявляют гетерогенность или гомогенность классов КП, таблица 3.

Таблица 3 — Оценки ССМ

Оценки	ToxEffects	Secom	Numbers	Leuco
Общее количество формализованных дескрипторов	4315	2171	12168	250
Референтное покрытие	98%	96%	100%	100%
Дескрипторы ядра	1317	1666	6642	194
Референтное покрытие дескрипторами ядра	98%	96%	100%	100%
Дескрипторы центра	102	226	326	10
Референтное покрытие дескрипторами ядра	97%	96%	99.9%	100%

Неполный охват референтов дескрипторами в «ToxEffects» и «Secom» свидетельствует о гетерогенности классов КП в этих задачах. Неохваченные референты не типичны, обладают выраженной индивидуальностью и не отвечают идее частотности. Переход от ядра к центру обеспечивает минимизацию числа дескрипторов, характеризующих классы КП.

В каждой задаче для каждого класса КП построены репрезентанты (идеальные образцы классов), содержащие наиболее существенные признаки класса. Репрезентанты всех классов попарно различны по составу и уровням (L - Low или H - High) значений показателей, таблица 4.

Каждый референт попадает в «сферу притяжения» репрезентанта того или иного класса и принадлежит к определенной области денотата этого класса.

Таблица 4 — Фрагмент описания репрезентантов классов в системе данных «Numbers»

Показатель	Класс									
	0	1	2	3	4	5	6	7	8	9
Pix2	L	L	L	H	L	L	L	H	L	L
Pix215	H	L	H	H	L	H	H	L	H	L
Zern18	H	L	L	L	L	H	H	L	H	H
Pix26	H	L	–	H	L	H	L	H	L	H
Pix67	L	H	L	H	H	H	L	L	-	L
Fr73	L	H	H	L	H	L	H	H	L	L
Zern8	L	H	H	H	H	H	L	H	L	-

Таблица 5 содержит распределения референтов классов по областям денотата КП.

Таблица 5 — Распределение референтов КП по областям структур классов КП

Области	ToxEffects	Secom	Numbers	Leuco
Ядерная область	16%	0%	2%	29%
Околоядерная область	34%	2%	35%	27%
Ближняя периферия	15%	12%	41%	10%
Дальняя периферия	20%	39%	21%	25%
Смежная область	15%	47%	1%	9%

Наибольшая неоднородность выявлена для «Secom». Единственный репрезентант класса не охватывает все референты класса. Это связано с изначальной нецельностью классов – среди референтов как 1-го, так и 2-го класса присутствуют разнотипные изделия.

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. *Обоснована идея решения задачи ЕСК* на базе метода ФОС, обеспечивающего производство завершеного научно-достоверного онтологического знания об открытых природных, общественных, антропогенных, технических и киберфизических системах из больших многомерных массивов гетерогенных эмпирических данных. Благодаря ФОС возникла реальная перспектива решения актуальной задачи рационального естественного классифицирования на основе знания онтологии системы. До настоящего времени эта задача не получила воспроизводимого решения методами искусственной классификации, не располагающими знанием онтологии предметной области.
2. *Обоснована применимость* глубинного научно понятого, рационально объясненного знания об онтологии систем, полученного



- ФОС, в качестве решения проблемы естественного классифицирования.
3. *Предложена концепция ЕСК*, на основе которой разработана модель метода ЕСК, в результате чего решение проблемы естественного классифицирования получило свое полное завершённое оформление. Разработана система ключевых понятий метода ЕСК, порожденных знанием онтологии, отвечающих принципам сходства и различия, общности и существенности признаков классов.
  4. *Разработан научный метод ЕСК*, в котором ключевые понятия и отношения понятий модели метода ЕСК получили конструктивное оформление. Получены формализованные представления системно-обусловленных естественных признаков классов, архетипов классов, репрезентантов классов, устройства таксонов. Определены состав и содержание вычислительных процедур метода ЕСК.
  5. *Разработан программный комплекс*, реализующий метод ЕСК в рамках технологической платформы ФОС.
  6. *Верифицирован и аттестован* полный технологический цикл решения задачи ЕСК в разных предметных областях, характеризующихся разным числом и составом классов, многосортностью шкал данных, размерностью описания объектов, в результате чего продемонстрировано аналитическое и вычислительное превосходство метода ЕСК по отношению к методам искусственной классификации.
  7. *Определены перспективы*: совершенствования метода ЕСК путем включения партитивных и родовидовых отношений естественных признаков классов; создание отчуждаемого программного продукта – решателя задачи диагностики (распознавания) в разных предметных областях; оснащение метода ЕСК когнитивными процедурами понимания и объяснения предметного содержания естественных признаков и классов.

## Публикации автора по теме диссертации

### В изданиях из списка ВАК РФ

1. *Качанова, Т. Л.* Естественная классификация острых отравлений фосфорорганическими веществами / Т. Л. Качанова, К. А. Туральчук, Б. Ф. Фомин и др. // Известия СПбГЭТУ "ЛЭТИ". — 2015. — Т. 8. — С. 8–17.
2. *Туральчук, К. А.* Метод отбраковки негодных полупроводниковых изделий на базе технологий физики систем / К. А. Туральчук, Т. Л. Качанова // Известия СПбГЭТУ "ЛЭТИ". — 2013. — Т. 7. — С. 46–53.

3. *Агеев, В. О.* Генерация системного знания по проблемам социальной напряженности в регионах России / В. О. Агеев, Т. Л. Качанова, К. А. Туральчук и др. // Научно-технические ведомости СПбГПУ. — 2012. — Т. 147. — С. 300—308.
4. *Качанова, Т. Л.* Анализ аварийности сетей городского водоснабжения на базе системного знания / Т. Л. Качанова, К. А. Туральчук, Б. Ф. Фомин и др. // Научно-технические ведомости СПбГПУ. — 2015. — Т. 121, № 3. — С. 297—303.

#### **В научно-технической базе SCOPUS**

5. *Kachanova, T.* Class Reconstruction in the Space of Natural System Classification / T. Kachanova, K. Turalchuk, B. Fomin // Procedia Computer Science. — 2019. — Vol. 150. — P.140–146.
6. *Kachanova, T.* Scientific understanding of ontological knowledge about open systems that is automatically mined from big data / T. Kachanova, K. Turalchuk, B. Fomin // Proceedings of the 33rd International Business Information Management Association Conference. — 2019.

#### **В прочих изданиях**

7. *Качанова, Т. Л.* Системная онтология классов в задаче обнаружения негодных изделий в полупроводниковом производстве / Т. Л. Качанова, Б. Ф. Фомин, К. А. Туральчук // Международная конференция по мягким вычислениям и измерениям. Т. 2. — 2018. — С. 87—90.
8. *Качанова, Т. Л.* Метод естественной классификации на основе информационных технологий физики открытых систем / Т. Л. Качанова, К. А. Туральчук, Б. Ф. Фомин и др. // Информационные технологии в управлении (ИТУ-2016) Материалы 9-й конференции по проблемам управления. — 2016. — С. 15—24.
9. *Качанова, Т. Л.* Системная онтология классов и естественная классификация / Т. Л. Качанова, К. А. Туральчук, Б. Ф. Фомин и др. // Всероссийская научная конференция по проблемам управления в технических системах. — 2015. — № 1. — С. 50—53.

*Туралъчук Константин Анатольевич*

Метод естественной системной классификации

Автореф. дис. на соискание ученой степени канд. техн. наук

Подписано в печать \_\_\_\_\_.\_\_\_\_.\_\_\_\_\_. Заказ № \_\_\_\_\_

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.

Типография \_\_\_\_\_

