

Федеральное агентство научных организаций

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ИНСТИТУТ ДИНАМИКИ СИСТЕМ И ТЕОРИИ УПРАВЛЕНИЯ
ИМЕНИ В.М. МАТРОСОВА
СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК

УДК 004.5 + 528.9

№ госрегистрации АААА-А17-117032210079-1

Инв. № 2017-10

УТВЕРЖДАЮ

Директор ИДСТУ СО РАН

академик



И.В. Бычков

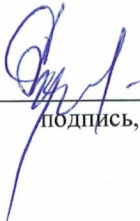
«23» января 2018 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

то теме:

МЕТОДЫ И ТЕХНОЛОГИИ СОЗДАНИЯ РАСПРЕДЕЛЁННОЙ СЕРВИСНО-ОРИЕНТИРОВАННОЙ СРЕДЫ СБОРА, ХРАНЕНИЯ, ОБРАБОТКИ БОЛЬШИХ ОБЪЁМОВ РАЗНОФОРМАТНЫХ МЕЖДИСЦИПЛИНАРНЫХ НАУЧНЫХ ДАННЫХ И ЗНАНИЙ, ОСНОВАННЫЕ НА КОНСТРУКТИВНЫХ СРЕДСТВАХ СПЕЦИФИКАЦИИ, ПОРОЖДАЮЩЕМ ПРОГРАММИРОВАНИИ И ИНТЕЛЛЕКТУАЛИЗАЦИИ
(промежуточный)

Научный руководитель проекта

 Д.Т.Н. Г.М. Ружников
подпись, дата 23.01.2018

Иркутск 2017

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель темы:
Зав. отделения
д-р технич. наук

 28.12.2017
подпись, дата

Г.М. Ружников (введение,
раздел 1, 2, заключение)

Исполнители темы:

Зав. лабораторией
канд. технич. наук

 28.12.2017
подпись, дата

А.Е. Хмельнов (раздел 2)

Вед. науч. сотр.
канд. технич. наук

 28.12.2018
подпись, дата

Р.К. Фёдоров (раздел 1, 2)

Зав. лабораторией
канд. технич. наук

 28.12.2018
подпись, дата

А.Ю. Юрин (раздел 2)

Ст. науч. сотр.
д-р. технич. наук

 28.12.2017
подпись, дата

О.А. Николайчук (раздел 1,
2)

Гл. науч. сотр.
д-р. технич. наук

 28.12.2017
подпись, дата

А.Ф. Берман (раздел 2)

Ст. науч. сотр.
канд. технич. наук

 28.12.2017
подпись, дата

А.И. Павлов (раздел 1, 2)

Ст. науч. сотр.
канд. технич. наук

 28.12.17
подпись, дата

А.О. Шигаров (раздел 1, 2)

Ст. науч. сотр.
канд. технич. наук

 28.12.2017
подпись, дата

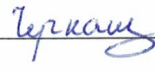
А.С. Гаченко (раздел 2)

Ст. науч. сотр.
канд. технич. наук

 28.12.2017
подпись, дата

В.В. Парамонов (раздел 1, 2)

Ст. науч. сотр.
канд. технич. наук

 28.12.2017
подпись, дата

Е.А. Черкашин (раздел 2)

Науч. сотр.
канд. технич. наук

 28.12.2017
подпись, дата

Е.С. Фереферов (раздел 2)

Науч. сотр.
канд. технич. наук

 28.12.2017
подпись, дата

А.П. Попова (раздел 2)

Мл. науч. сотр.

 28.12.2017
подпись, дата

А.А. Михайлов (раздел 2)

Мл. науч. сотр.

 28.12.2017
подпись, дата

Н.О. Дородных (раздел 2)

Программист

 28.12.2017
подпись, дата

А.А. Ветров (раздел 2)

Программист	<u>Хмельная 28.12.2017</u> подпись, дата	Т.Ф. Хмельнова (раздел 2)
Программист	<u>Шумилов 28.12.2017</u> подпись, дата	А.С. Шумилов (раздел 1, 2)
Программист	<u>Авраменко 28.12.2017</u> подпись, дата	Ю.В. Авраменко (раздел 1, 2)
Программист	<u>Малтугуева 28.12.2017</u> подпись, дата	Г.С. Малтугуева (раздел 2)
Нормоконтролер канд. технич. наук	<u>Фереферов 23.01.2018</u> подпись, дата	Е.С. Фереферов

РЕФЕРАТ

Отчет 60 стр., 13 рис., 4 табл., 96 источников и публикаций.

ГЕОПОРТАЛ, ПРОСТРАНСТВЕННЫЕ ДАННЫЕ, ДЕКЛАРАТИВНЫЕ СПЕЦИФИКАЦИИ, НЕСТРУКТУРИРОВАННЫЕ ДАННЫЕ, СЕРВИСЫ

Объектом исследования являются пространственные и тематические данные, знания и технологии (сервисы) их обработки, конструктивные средства спецификации.

Цель проекта – разработать эффективные методы и технологии создания распределенной сервисно-ориентированной среды сбора, хранения, обработки больших объемов разноформатных междисциплинарных научных данных и знаний, основанные на конструктивных средствах спецификации, порождающем программировании и интеллектуализации.

Разработка технологий и сервисов распределённой обработки пространственных и тематических данных, проводится с использованием современных методов основанных на декларативных спецификациях и знаниях.

В ходе выполнения НИР получены следующие научные результаты:

1. Реализованы иерархии классов для абстрагирования технологий доступа к данным в настраиваемых при помощи спецификаций приложениях БД АРМ-ов.
2. Разработана технология создания информационных систем в облачной среде для поддержки междисциплинарных научных исследований, использующая геопортальный подход.
3. Разработаны новые возможности трансформации данных из произвольных электронных таблиц в реляционную форму на основе исполнения правил их анализа и интерпретации.
4. Разработаны модели информационных процессов создания и функционирования интеллектуальных систем поддержки принятия решений, основанные на реализации модельных трансформаций для систем различного типа. Разработана модель информационного процесса создания и функционирования системы агентного имитационного моделирования, обеспечивающая возможность автоматизированного создания прикладных агентных имитационных моделей.

Проект соответствует:

- приоритетным направлениям развития науки, технологий и техники в РФ: Информационно-телекоммуникационные системы (3);
- критическим технологиям РФ: Технологии и программное обеспечение распределенных и высокопроизводительных вычислительных систем (18), Технологии информационных, управляющих, навигационных систем (13), Технологии предупреждения и ликвидации чрезвычайных ситуаций природного и техногенного характера (21);
- технологической платформе – Национальная программная платформа.

СОДЕРЖАНИЕ

РЕФЕРАТ	4
ОПРЕДЕЛЕНИЯ.....	6
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	7
ВВЕДЕНИЕ	8
1 Исследование методов создания технологий обработки и анализа больших объёмов разноформатных данных и программного кода, основанные на использовании конструктивных средств спецификации	13
2 Создание методов формирования распределенной сервисно- ориентированной среды, а также технологий и программных систем формирования композиции сервисов в виде Javascript сценариев на основе анализа спецификаций существующих сервисов и данных	19
3 Создание методологического и инструментального обеспечения поддержки процессов трансформации неструктурированных данных из произвольных таблиц	23
4 Исследование методов и технологий создания проблемно-ориентированных и интеллектуальных систем поддержки принятия решений, основанные на использовании модельно-управляемого подхода, порождающего программирования и декларативных описаний, в контексте решения задач управления риском техногенных катастроф	32
ЗАКЛЮЧЕНИЕ	46
СПИСОК ПУБЛИКАЦИЙ	50

ОПРЕДЕЛЕНИЯ

Автоматизированное рабочее место – индивидуальный комплекс аппаратно-программных средств.

База данных – совокупность данных, организованных по определённым правилам, устанавливающим общие принципы описания, хранения, манипулирования данными.

Базовые пространственные данные – это описания базовых пространственных объектов в заданной системе координат, которые следует использовать для определения координат любых близлежащих пространственных объектов и явлений.

Геопортал – Web-портал, отображающий и предоставляющий доступ к географической информации посредством Web-сервисов.

Геоинформационная система – информационная система, обеспечивающая сбор, хранение, обработку, доступ, отображение и распространение пространственных данных.

Декларативное программирование – это парадигма программирования, в которой задаётся спецификация решения задачи, то есть описывается, что представляет собой проблема и ожидаемый результат

Инфраструктура пространственных данных – это иерархически упорядоченная система (информационная среда), построенная с использованием информационных и геоинформационных технологий, основанная на общих стандартах пространственных данных и метаданных, а также сети географических информационных узлов – геопорталов и каталогов метаданных.

Модель пространственных данных – это способ цифрового представления пространственных объектов, тип структуры пространственных данных.

Метаданные – это описательная информация о структуре и смысле данных, а также приложений и процессов, которые манипулируют данными.

Пространственные данные – данные о географических объектах, которые являются формализованными цифровыми моделями материальных или абстрактных объектов реального или виртуального мира

Хранилище данных – предметно-ориентированный, интегрированный, неизменчивый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

АРМ	–	автоматизированное рабочее место
БД	–	база данных
БПД	–	база пространственных данных
ГИС	–	геоинформационная система
ИС	–	информационная система
ИПД	–	инфраструктура пространственных данных
ПД	–	пространственные данные
СХД	–	система хранения данных
СУБД	–	система управления БД
ХД	–	хранилище данных
API	–	Application Programming Interface - интерфейс прикладного программирования
DNS	–	Domain Name System — система доменных имён, компьютерная распределённая система для получения информации о доменах.
IDE	–	Integrated development environment – интегрированная среда разработки
MFC	–	Microsoft Foundation Classes - библиотека на языке C++, призванная облегчить разработку GUI-приложений для Microsoft Windows путём использования богатого набора библиотечных классов.
MDA	–	Model Driven Architecture - архитектура, управляемая моделью
OGC	–	Open Geospatial Consortium - консорциум открытых ГИС
SOA	–	Service-oriented architecture - сервис-ориентированная архитектура
UML	–	Unified Modeling Language - унифицированный язык моделирования — язык графического описания для объектного моделирования в области разработки программного обеспечения, моделирования бизнес-процессов, системного проектирования и отображения организационных структур.
VCL	–	Visual Component Library - объектно-ориентированная библиотека для разработки программного обеспечения.
WMS	–	Web Map Service стандарт Web-служб OGC (Web-сервис), предоставляющий интерфейс http-запросов для получения браузером или настольным приложением в растровом виде геопривязанных изображений карт в форматах PNG, TIFF, JPEG и т.п.
WPS	–	Web Processing Service стандарт Web-служб OGC, (WPS-сервис), описывает правила стандартизации входящих и исходящих данных (запросов и ответов на них), сервисов обработки.
WWW	–	World Wide Web - всемирная информационная паутина

ВВЕДЕНИЕ

Большие объемы накопленных междисциплинарных научных данных и знаний в России и за рубежом, а также распределённость и развитость информационно-телекоммуникационной инфраструктуры обосновывают актуальность исследования технологий создания распределенной сервисно-ориентированной информационной среды сбора, хранения, обработки больших объёмов разноформатных данных и знаний.

Технологии и инструментальные средства автоматизации создания информационных систем (ИС), в том числе интеллектуальных, позволяют сократить сроки и, как следствие, затраты на создание и модернизацию систем. К наиболее перспективным направлениям автоматизации разработки ИС следует отнести такие подходы, как MDA (Model Driven Architecture), порождающее программирование, использование онтологических моделей и декларативных спецификаций. Следует отметить, что декларативные спецификации удобны своей компактностью, предметной ориентированностью и выразительностью, а также возможностью интерпретации различными трансформационными и другими процедурами.

Широкое внедрение сервисно-ориентированного подхода при создании информационно-вычислительных сред обусловлено распределённостью их информационных ресурсов, простотой использования сервисов, легким тестированием и отладкой, масштабируемостью, возможностью повторного использования. Для решения комплексных задач сервисы могут интегрироваться с использованием сложных логических конструкций для обработки передаваемых данных, управления потоком выполнения и т.д. Большое разнообразие создаваемых сервисов обуславливает актуальность их комплексного использования, передачи данных между сервисами, поиску сервисов, согласованию форматов данных, организации асинхронного вычислительного процесса, где операторами являются сервисы и т.д. Одним из наиболее популярных решений для интеграции информационно-вычислительных сервисов является Oracle BPEL Process Manager (Web Services Business Process Execution Language Version 2.0.-корпорация Oracle), строящий цепочки выполнения сервисов с помощью графического интерфейса. Другим средством интеграции сервисов является программная система Taverna (J. de Jesus, P. Walker, M. Grant, S. Groom и др.).

В свою очередь для эффективной обработки распределенной сервисно-ориентированной средой больших объёмов данных необходимо создание методологической основы извлечения и трансформации данных из таблиц и табличных форм, технологий их очистки.

Сравнительная оценка уровня проделанной работы в этом направлении в СО РАН, РАН и в мире. Большинство современных ИС для хранения и обработки данных используют СУБД. Поэтому практически всегда при разработке ИС тратится много усилий на реализацию пользовательского интерфейса для ввода, редактирования и поиска данных в БД. Для решения этой задачи в основном используются библиотеки классов визуальных элементов управления и невизуальных объектов, описывающих данные (например, VCL, MFC, FCL). Также применяется объектно-реляционное отображение (ORM – object-relational mapping), например Hibernate/NHibernate, Entity Framework, позволяющее ускорить разработку ИС за счёт автоматизации построения объектной модели, обеспечивающей взаимодействие с сущностями БД и являющейся описанием структуры БД для ИС. При этом часть программного кода, в частности классы соответствующие сущностям БД, и классы, обеспечивающие преобразование объектных запросов в SQL-запросы, генерируется автоматически. В настоящее время развиваются подходы позволяющие ускорить как разработку пользовательских интерфейсов (например, Model-Based User Interface Development(Paulo Pinheiro da Silva)), так и ИС в целом (например, MDA, порождающее программирование). Основной тенденцией в этих исследованиях является разработка современных методов структурирования данных о структуре ИС в виде моделей системы различными средствами (например, надстройки над моделями классов UML), применение модельных каркасов (Черткова Е.А.), или построения онтологий предметной области (Грибова В.В., Клещёв А.С.). Проблемой данных подходов является то, что сгенерированный по моделям программный код является грубым каркасом ИС и практически всегда требует доработки вручную, при этом полученные изменения не отражаются в исходных абстрактных моделях системы.

В ИДСТУ СО РАН развивается подход к разработке ИС на основе декларативных спецификаций и универсальных алгоритмов для работы с БД, управляемых такими спецификациями. Реализованы алгоритмы решения задач: построение объектов, описывающих структуру БД в приложении, генерация пользовательского интерфейса, интерактивное построение запросов, формирование отчётов, отображение результатов запросов на карте и т.д. Реализовано ПО для создания клиент-серверных приложений БД под ОС Windows и информационно-справочных Web-приложений по спецификациям приложений БД. В рамках проекта должны быть расширены возможности имеющихся универсальных алгоритмов, разработаны новые, и реализованы аналогичные приложения для других платформ (Linux, Android).

Следует отметить, что в ИДСТУ СО РАН также развиваются технологии создания информационных систем (в том числе интеллектуальных) на основе порождающего

программирования, онтологических моделей, а также построения каркасов ИС на основе UML моделей с применением правил трансформации, учитывающих, в частности, свойства среды реализации. Можно отметить общую тенденцию к использованию концептуальных моделей при создании баз знаний и ориентацию на непрограммирующих пользователей (Аликин С.С., Вяткин В.В., Гаврилова Т.А., Дубинин В.Н., Ермаков А.Е., Жидаков В.П., Найденова К.А., Стенников В.А., Fouad K.M., Garcia-Crespo A., Nofal M., Ruiz-Mezcua B.), активно развиваются подходы к созданию интеллектуальных систем на основе онтологий и семантических технологий (Голенков В.В., Грибова В.В., Гулякина Н.А., Клещев А.С., Загорулько Ю.А.). При этом перспективным является использование подходов, основанных на порождающем программировании (Айзенекер У., Чернецки К.), в частности, модельно-управляемом подходе – Model-Driven Engineering (MDE) или Model Driven Development (MDD) (Book M., Gruhn V., Sami B.) и его разновидностях, например, Model Driven Architecture (MDA) (Грибачев К.Г., Frankel D.) – концепции реализации MDD от Object Management Group (OMG). В ИСЭМ СО РАН развивается модельно-управляемый подход создания специализированных информационных систем. Можно выделить работы Голенкова В.В. (БГУИР, Белоруссия) по технологии компонентного проектирования интеллектуальных систем; Canadas J., Canfora G., Distantе D., Palma J., Pedone P., Rossi G., Tunes S. (University of Sannio, Sapienza University, Италия; AGH University of Science and Technology, Польша) связанные с разработкой и встраиваемых систем (программных компонентов) для веб-приложений на основе порождающего программирования. Проблемой данных подходов является то, что сгенерированный по моделям программный код является грубым каркасом ИС и практически всегда требует доработки вручную, при этом полученные изменения не отражаются в исходных абстрактных моделях системы

В ИДСТУ СО РАН ведутся исследования по созданию основных компонентов сервисно-ориентированной среды хранения пользовательских данных и выполнения WPS-сервисов над ними. Для организации выполнения WPS-сервисов разработаны модули инвентаризации сервисов, запуск, обмена данными, контроля выполнения WPS-сервисов. В СО РАН можно выделить работы ИВТ СО РАН (Пестунов И.А.), направленные на разработку сервисов, поддерживающих стандарт WPS. В Южном научном центре РАН разрабатывается (Каляев И.А.) мультиагентный подход к построению адаптивной распределенной вычислительной среды. В ИАПУ ДВО РАН проводятся исследования (Грибова В.В., Гулякина Н.А., Клещев А.С.) по разработке сервисно-ориентированной системы создания виртуальных лабораторий. В Научно-исследовательском институте наукоемких компьютерных технологий (Бухановский А.В.), разрабатывается среда

вычислительных сервисов. Среди зарубежных исследователей отметим X. Yang and H. Zhang, реализующих сервисно-ориентированную архитектуру на основе облачных вычислений, а также Zhao Wu, Naixue Xiong, Yannong Huang, Qiong Gu, Chunyang Hu, Zhongbo Wu, and Bo Hang, решающих проблему обеспечения надежности выполнения композиции облачных сервисов.

Ведутся исследования по созданию методологической основы извлечения и трансформации данных из таблиц и табличных форм. В частности, предлагаются модели таблиц (Hurst, 2006; Lautert, 2013), методы обнаружения и распознавания таблиц (Ramel, 2003; Amano, 2006; Hassan, 2007; Oro, 2009), методы анализа и интерпретации таблиц (Wang, 2012; Embley 2005, 2014; Nagy, 2014), методы анализа и трансформации электронных таблиц (Hung, 2011; Chen, 2013, 2014). В последние годы получены результаты по извлечению таблиц из неформатированного текста и документов печати в формате Enhanced Metafile, анализа и интерпретации произвольных размеченных таблиц (Shigarov, 2009-15). Работы по созданию инструментальных средств разработки программного обеспечения для трансформации данных из произвольных таблиц только начинают развиваться, например, созданы специализированные формальные языки TranSheet (Университет Нового Южного Уэльса, Австралия) и CRL (ИДСТУ СО РАН, Россия). На сегодняшний день отсутствуют проблемно-ориентированные библиотеки модели и алгоритмы трансформации табличных данных. В проекте впервые планируется исследовать алгоритмы поддержки процессов трансформации неструктурированных данных из произвольных таблиц, включая манипулирование, анализ, интерпретацию, очистку и отслеживание происхождения.

Цель проекта, предполагаемые результаты исследований и возможные области их применения: создать эффективные методы и технологии формирования, основанные на конструктивных средствах спецификации, порождающем программировании и интеллектуализации, формирования распределенной сервисно-ориентированной среды сбора, хранения, обработки больших объемов разноформатных междисциплинарных научных данных и знаний.

Задачи 2017 года:

1. Реализовать иерархии классов для абстрагирования технологий доступа к данным в настраиваемых при помощи спецификаций приложений БД АРМ-ов.
2. Исследовать существующие стандарты спецификаций сервисов и данных.
3. Провести аналитический обзор современного состояния в области понимания таблиц, систематизированный по моделям, методам, оценкам и приложениям.
4. Исследовать методы и разработать принципы оригинальных технологий

создания проблемно-ориентированных систем и сервисов интеллектуальной поддержки принятия решений и агентного имитационного моделирования.

Основанием для проведения научно-исследовательских работ по теме «Методы и технологии создания распределенной сервисно-ориентированной среды сбора, хранения, обработки больших объёмов разноформатных междисциплинарных научных данных и знаний, основанные на конструктивных средствах спецификации, порождающем программировании и интеллектуализации» является Государственное задание Федерального государственного бюджетного учреждения науки Института динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук на плановый период 2017 - 2020 гг.

1 Исследование методов создания технологий обработки и анализа больших объёмов разноформатных данных и программного кода, основанные на использовании конструктивных средств спецификации

1.1 Иерархии классов для абстрагирования технологий доступа к данным при помощи спецификаций приложений баз данных

Программный код для автоматизации разработки приложений баз данных (БД) с использованием спецификаций был существенно переработан. Ранее для поддержки работы с использованием различных технологий доступа к данным (ADO и BDE) использовалась условная компиляция, что существенно затрудняло переработку кода. Основные фрагменты кода, различающиеся в зависимости от выбранной технологии доступа к данным, были систематизированы и реализованы в виде виртуальных методов классов – потомков абстрактного базового класса TBaseDAT (DAT, Data Access Technology – технология доступа к данным, далее ТДД).

Для того, чтобы не приходилось собирать весь код, зависящий от ТДД, в один модуль, предложен способ организации кода с использованием *реестров зависимых классов* и *вариаторов* – классов, позволяющих варьировать код, используемый в определённом модуле, в зависимости от применяемой ТДД. Реестр зависимых классов используется для выбора типа при создании экземпляра вспомогательного класса в зависимости от применяемой ТДД. При этом в модуле, определяющем вспомогательный базовый класс, объявляется реестр его классов-потомков, в котором эти классы-потомки регистрируются при подключении в программу модулей, определяющих такие классы. Реестр зависимых классов используется для выбора типа при создании экземпляра вспомогательного класса в зависимости от применяемой ТДД.

В некоторых случаях неудобно создавать классы-потомки в зависимости от ТДД, в основном это касается классов визуальных форм. Так, если в приложении используется одна общая форма для отображения содержимого БД при разработке спецификации, то хочется иметь возможность изменения её поведения в зависимости от выбора ТДД. Для этих целей используются специальные классы – вариаторы поведения. При этом таблица вариаторов создаётся у модифицируемого объекта и, в зависимости от выбранной ТДД, используется один из этих вариаторов.

Всё это технически позволяет разрабатывать приложения, использующие сразу несколько ТДД, но эта возможность ещё не была опробована на практике.

С использованием разработанных иерархий классов была поддержана работа с данными с использованием библиотеки доступа к данным FireDAC (Рисунок 1).

Библиотека FireDAC является основной для разработки мультиплатформенных приложений в последних версиях Delphi, её использование позволяет в перспективе разработать версии программ для мобильных устройств.

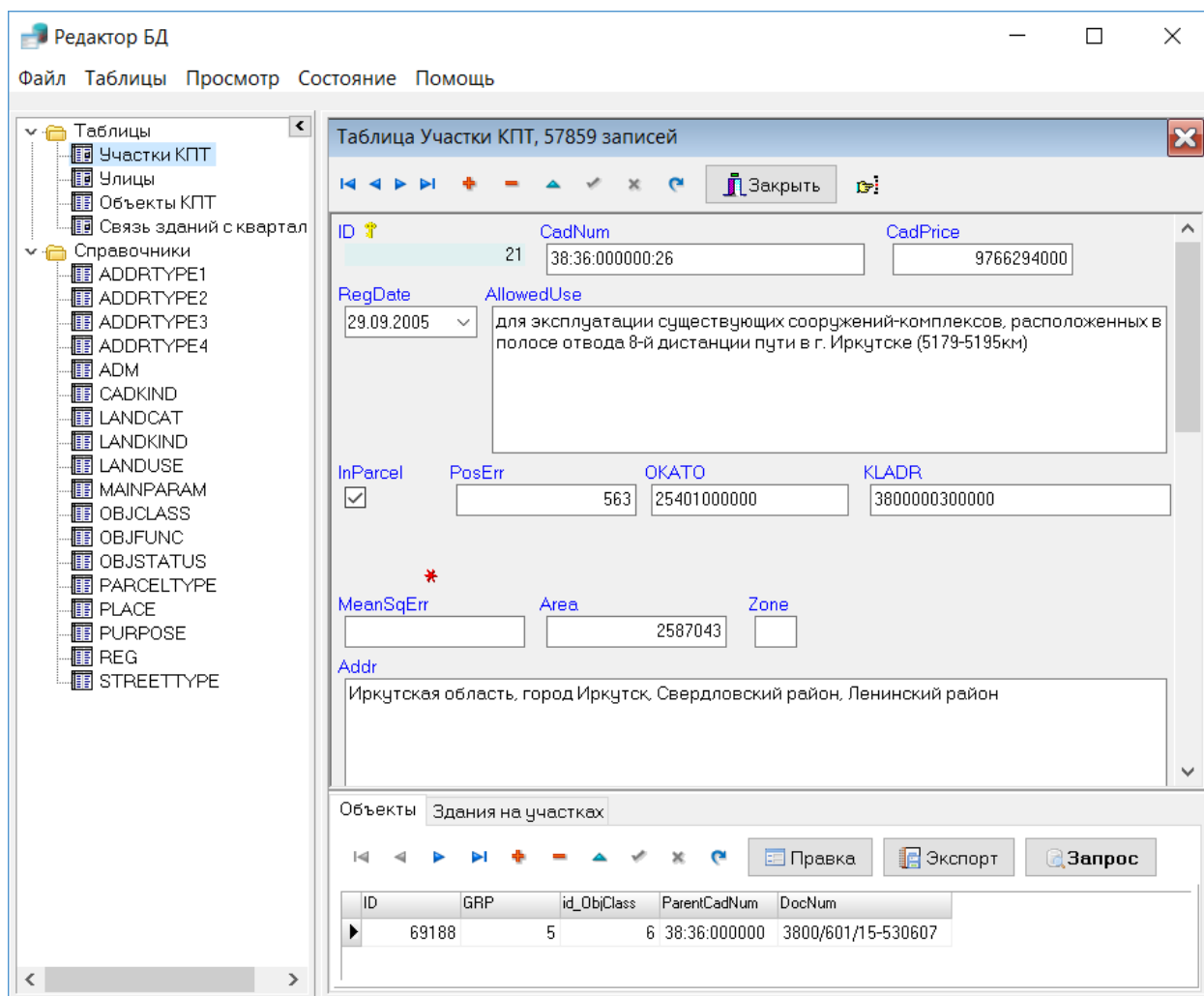


Рисунок 1. Приложение БД, работающее с использованием спецификаций с использованием ТДД FireDAC

1.2 Создание алгоритмов анализа потоков данных в дизассемблере программного кода из файлов DCU

Для отображения дополнительной информации об операндах инструкций и ячейках памяти в программе DCU32INT реализован итерационный алгоритм анализа потоков данных подпрограмм. Для сбора информации об операндах инструкций в потоке данных используется метод достигающих определений. Компилятор Delphi генерирует код таким образом, что все обращения к памяти происходят в пределах адресного пространства данной процедуры, что позволяет производить анализ в рамках процедуры. Анализ потоков данных проводится на графе потока управления для подпрограммы, состоящем из

базовых блоков. Базовый блок представляет собой последовательность инструкций с одним входом и одним выходом.

Промежуточный код для базовых блоков строится путём символьной интерпретации каждой команды СП и сопоставления ей выражения (экземпляра класса), реализующего семантику. Начальное состояние каждого линейного участка кода характеризуется значениями параметров подпрограммы и локальных переменных, а также состоянием стека. Конечное состояние определяется в результате символьной интерпретации.

Далее применяется итерационный алгоритм анализа потоков данных для достигающих определений. Состояние локальных переменных и аргументов подпрограммы являются общим контекстом для всех базовых блоков, и используется только для хранения результатов, а не вычисления выражений. Поэтому в качестве входного и выходного множества для передаточной функции рассматривается только состояние стека.

1.3 Разработка методов геокодирования в условиях неточной информации

Разработан прототип веб-сервиса, посредством которого объединены в единый ресурс оригинальная библиотека по разбору почтовых адресов и отображение найденного адреса на картографической основе.

Модуль по разбору адресов производит поиск в иерархических адресных справочниках (статичных) наилучших вариантов сопоставления для строки почтового адреса, которая может содержать ошибки. При построении дерева альтернативных вариантов разбора почтовых адресов, возникает проблема ранжирования результатов и выбора одного или нескольких предпочтительных вариантов сопоставления. Для решения этой задачи разработан алгоритм информированного поиска, позволяющий на каждом этапе иметь дело с наиболее перспективным к текущему моменту вариантом разбора. Алгоритм поиска вариантов сопоставления получает на вход последовательность лексем, являющуюся результатом работы лексического анализатора. В своей работе алгоритм использует машину состояний, предназначенную для хранения и контроля состояний разбора. Основными структурами данных, используемыми при работе алгоритма, являются: очередь с приоритетами, предоставляющая на каждом этапе разбора наиболее перспективный вариант перехода между состояниями; и хэш-таблица состояний, позволяющая избежать повторной обработки состояний разбора. Каждое состояние

разбора идентифицируется парой (узел справочника адресов, позиция в списке лексем) и содержит информацию о наилучшем известном к текущему моменту способе сопоставления названий из данного узла и его предков с соответствующим позиции начальным фрагментом списка лексем. Таким образом, в предложенном алгоритме реализуется схема динамического программирования. Для сравнения найденных вариантов разбора предложена система оценок, которая позволяет учитывать расстояние между сопоставляемыми словами, перестановки слов в словосочетании и наличие элементов адреса в строке.

Разработанная технология может быть использована в задачах геокодирования информации базы данных, нормализации почтовых адресов, в обработке массивов данных, содержащих адресную информацию, а также в информационных поисковых системах для поиска корректных почтовых адресов и отображения их на топооснове.

1.4 Разработка методики приобретения (acquisition) структурного описания документа на основе анализа его изменений

Исследован подход к автоматизации разметки документов по аналогии с уже размеченным примером на основе компьютерного обучения. Процесс состоит из трех шагов: 1). разметка одного документа метаданными в виде набора атрибутов абзацев; 2). машинное обучение, ориентированное на сопоставление значений атрибутов роли абзаца в документе; 3). классификация абзацев в аналогичных документах при помощи обученного алгоритма. На данной задаче протестированы стандартные для библиотеки `scipy` (Python) алгоритмы компьютерного обучения: построение дерева решений, сеть Байеса и нейронная сеть. При помощи дерева решений выделен набор наиболее информативных атрибутов, на основе которых выполняется распознавание классов абзацев в тексте документа: отсутствие или присутствие в абзаце глагола, наличие выраженного абзацного отступа, наличие отступа красной строки, наличие номера раздела, наличие слова-разделителя структуры документа, наличие заголовочного абзаца перед текущим и т.п.

В результате применения методики новые документы получают дополнительную общую структурную и семантическую разметку. На примере рабочих программ вуза удалось выделить с достоверностью 0.7 основные разделы рабочей программы, разделы формального описания курса, разобрать структуру титульного листа.

Для поддержки решения данной задачи произведен анализ существующих стандартизованных онтологий (концептуальных моделей) и реализован сервис хранения документов и их размеченной логической структуры.

Логическая структура документа представляется с использованием дескриптивных возможностей технологий Семантического веба (Semantic Web), применяемых в формате HTML5, а также технологии открытых связанных данных (Linked Open Data, LOD). Логическая структура документа представляется при помощи онтологий Open Annotation (oa) — представление содержимого (аннотации), описывающего другое содержимое; Friend-of-a-friend (foaf), которая позволяет представлять информацию об агентах: физических, юридических лицах и программных агентах; Provenance (prov) — основа описания информационных потоков в документах и их взаимосвязи; Dublin Core (dc) представляет в аннотации метаданную о творческом произведении: авторов, формат содержимого, его описание и др.; DBpedia resource (dbr) — пространство имен конкретных объектов (ресурсов-экземпляров) Wikipedia; Schema.org (schema) представляет объекты, распознаваемые поисковыми агентами Google, Yandex, Yahoo и др.; Bibliographic Ontology (bibo) — представление списков литературы; онтологии из проекта NEPOMUK (<https://userbase.kde.org/Nepomuk>), предназначенные для описания объектов, хранимых в полнотекстовых индексных цифровых архивах. Пример описания части документа при помощи выбранных онтологий изображён на Рисунок 2.

Разработаны программные средства (сервер) хранения документов на основе открытых связанных данных. Архитектура созданного сервера изображена на Рисунок 3.

Основные функции сервера заключаются в предоставлении доступа к данным по адресу ресурса и различным типам контекста: по содержимому документа (полнотекстовый индекс), по запросу SPARQL, а также Datalog-запросу (при помощи механизма Penguins). В реализации функций использованы современные технологии: Elasticsearch — полнотекстовый поиск по содержимому документов и данным, хранимым в графах онтологий; интерпретатор SWI-Prolog — реализация Datalog-запросов; экспериментальный сервер онтологий ClioPatria — система хранения графов онтологий и интерпретации основных RDF- и RDFs-отношений. Все основные функциональные блоки представлены в виде компонент Zope, что дает возможность их использования по отдельности, так и в рамках одного конфигурируемого декларативно приложения. Средствами системы программирования Prolog реализована подсистема обработки данных логического слоя на основе формализованных знаний. Знания хранятся как на сервере, так и могут быть подружены с клиентского веб-браузера. В качестве приложения

разрабатываемых технологий реализовано новое программное средство для автоматизации верстки (authoring) документов.

Важным преимуществом использования LOD при создании информационных сред является ослабление требований к хранилищам публикуемой информации: сам документ является хранилищем данных в формализованном виде. В некоторой степени это позволяет время, затрачиваемое на проектирование структуры базы данных для хранилищ частичноструктурированных документов, перенаправить на процесс решения предметной задачи.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html lang="ru" xmlns=http://www.w3.org/1999/xhtml
xmlns:taa=http://irnok.net/engine/rdfa-manipulation
xml:lang="ru" metal:define-macro="page">
<head> <!-- Подключение стилевых таблиц и модулей -->
</head>
<body prefix="rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns# ... foaf:
http://xmlns.com/foaf/0.1/ imei: imei.html#
course: https://irnok.net/college/plan/01.03.02-16-1234-2461\_1%D0%BA\_PB-SM.plm.xml.xlsx-
%D0%911.%D0%92.%D0%94%D0%92.3.1.html#"
resource="#post" typeof="schema:CreativeWork sioc:Post prov:Entity">
<!-- Панель управления приложением -->
<main lang="ru" resource="#annotation" typeof="oa:Annotation"
id="main-document-container"><div property="oa:hasTarget"
resource="#course-work-program"></div>
<article property="oa:hasBody"
typeof="schema:Article foaf:Document curr:WorkingProgram"
resource="#course-work-program" id="main-document">
<div taa:content="imei:title-page"></div> <!--Титульный лист.. -->
<div taa:content="imei:neg-UMK"></div> <!--Лист согласования..-->
<section id="contents" class="break-after"> <h2 class="nocount c">Содержание</h2>
<div id="tableOfContents"></div>
</section>
<section id="course-description" resource="#description"
property="schema:hasPart" typeof="schema:CreativeWork">
<div property="schema:hasPart" resource="#purpose"
typeof="dc:Text cnt:ContentAsText" >
<div property="cnt:chars" datatype="xsd:string">
<h2 property="dc:title" datatype="xsd:string">Цели и задачи дисциплины
(модуля)</h2>
<p>Целью преподавания дисциплины "Технологии программирования" является
освоение ...</p>
</div>
</div>
. . . . .
<div property="schema:hasPart" typeof="dc:Text
cnt:ContentAsText" resource="#volume">
<div property="cnt:chars" datatype="xsd:string">
<h2 property="dc:title" datatype="xsd:string"> Объем дисциплины (модуля) и виды
учебной работы (разделяется по формам обучения)</h2>
<div taa:content="course:time-distrib"></div>
</div>
</div>
```

Рисунок 2. Пример логического представления документа (рабочей программы вуза)

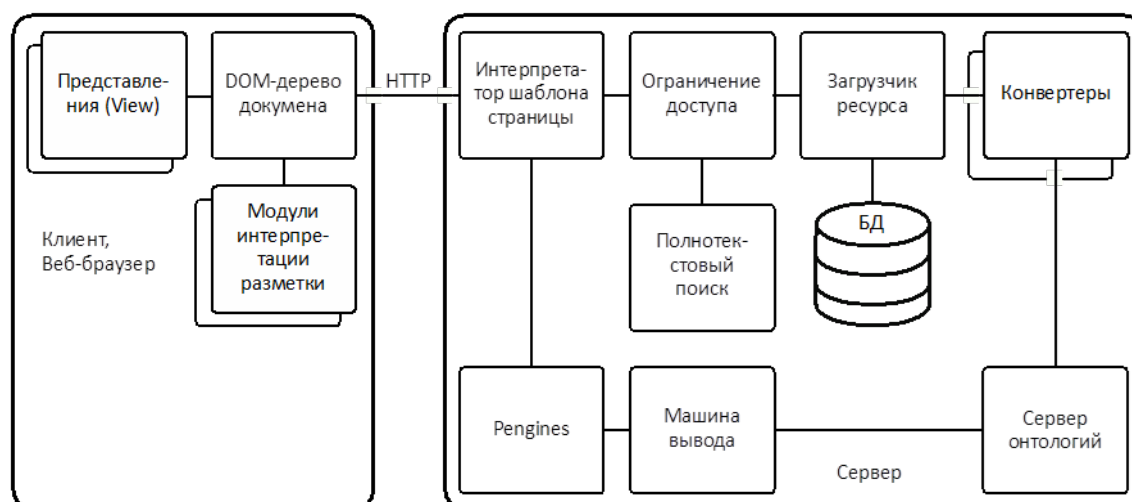


Рисунок 3. Архитектура сервера хранения документов и их логической разметки

2 Создание методов формирования распределенной сервисно-ориентированной среды, а также технологий и программных систем формирования композиции сервисов в виде Javascript сценариев на основе анализа спецификаций существующих сервисов и данных

2.1 Исследование стандартов спецификаций сервисов пространственных данных и формирования инфраструктуры пространственных данных

Как правило, существующие программные системы анализа данных требуют представления данных в реляционном (табличном) виде с определенным набором полей, типами полей, форматами данных, используемыми справочниками, единицами измерения и т.д. Соответственно при применении некоторой программной системы анализа от исследователя требуется унификация реляционных данных под требования, что может быть не тривиальным. Например, методы создания цифровой модели рельефа требуют наличия атрибута, содержащего положение объекта (точка или полилиния) и атрибута со значением высоты в метрах. Данные исследователя могут быть в других единицах измерения, что потребует от него поиска и применения специализированных средств или ручной переработки. Решением проблемы унификации данных и методов является публикация спецификаций, описывающих требуемую структуру реляционных данных. Предлагается с помощью спецификаций фиксировать структуру реляционных таблиц, требуемых для определенного сервиса анализа или обработки данных. Соответственно по этим спецификациям предусматривается возможность ввода данных для сервисов и

создание самих таблиц. Метаданные о структурах таблиц необходимо упорядочить в виде некоторых иерархий и реализовать механизмы наследования и полиморфизма в терминах объектно-ориентированного подхода. Система может обеспечить интеграцию отличающихся по составу данных для совместного анализа.

Хранение спецификаций о структуре таблиц дает возможность сбора данных, ориентированных на использование в конкретных сервисах анализа, а также создавать сервисы обработки, ориентированные на данные, соответствующие опубликованным метаданным.

В рамках геопортала разработан каталог спецификаций для хранения спецификаций о структуре таблиц и их предоставления по запросу. Спецификации хранятся в формате JSON. В рамках каталога спецификаций разработаны сервисы предоставления и обмена спецификаций, который предоставляет программный REST интерфейс:

- 1) /dataset/list – получение списка структур таблиц;
- 2) /dataset/list?f_id=<Идентификатор спецификаций> – получение спецификаций определенной таблицы.

Представленный интерфейс позволяет автоматически обмениваться спецификациями между геопорталами. Добавление новых спецификаций в каталог производится с помощью специального инструмента создания таблиц (см. Рисунок 4). Данный инструмент позволяет определить структуру таблиц и ее название, определить типы полей и элементы управления. Кроме того в рамках этого инструмента можно выбрать спецификации из каталога и создать на их основе таблицу.

Рисунок 4. Инструмент для создания таблиц

Разработаны спецификации наиболее часто используемых данных в области геообработки:

- административные границы;
- объекты транспортной сети;
- населенные пункты;
- здания и сооружения;
- гидрография;
- историко-культурные достопримечательности;
- образцы видов растительности (гербарии) и т.д.

На основе созданных спецификаций создан ряд таблиц, которые в свою очередь зарегистрированы в каталоге.

В проекте проведено развитие формирования инфраструктуры пространственных данных в облачной среде. Технология базируется на оригинальном типовом геопортале, реализующим общие для систем функции. Некоторые из функций вынесены на отдельные узлы облачной среды. Пользователь работает с помощью браузера.

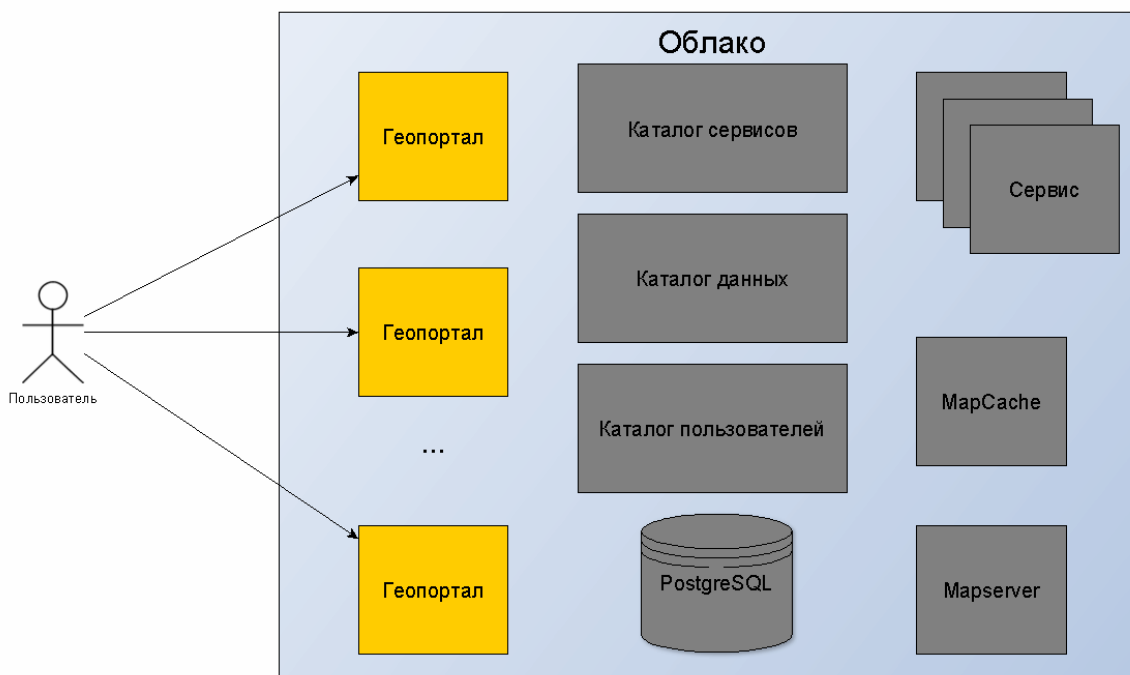


Рисунок 5. Архитектура облачной среды

Типовой геопортал разработан на основе программной платформы Node.js и СМС Calipso.

«Каталог данных» — предназначен для хранения спецификаций таблиц и их структурных спецификаций. Каталог обеспечивает поиск данных по всем геопорталам.

«Сервисы» — вычислительные узлы виртуальной среды, выполняющие методы обработки данных. Методы должны быть реализованы в соответствии со стандартом WPS.

«Каталог сервисов» — предназначен для хранения метаданных сервисов. Каталог обеспечивает поиск сервисов и их выполнение.

«Mapserver» — проводит на стороне сервера генерацию изображений слоев карт в соответствии со стандартом WMS. Выделен на отдельный вычислительный узел из-за необходимости использования 80 порта. Многие организации блокируют доступ к остальным портам.

«MapCache» — проводит кэширование карт для получения скоростного доступа.

«PostgreSQL» — СУБД. Для обработки пространственных данных в СУБД установлено расширение PostGIS. Каждому пользователю системы предоставляется схема базы данных, в которой он может создавать таблицы. СУБД может функционировать на каждой виртуальной машине геопорталов или на общей СУБД.

Все рассмотренные компоненты архитектуры функционируют в рамках облачной среды. Разработанная технология базируется на рассмотренной архитектуре и состоит из следующих этапов:

- 1) развертывание нового узла облачной среды на основе типового геопортала и присвоение DNS имени, что позволяет получить общедоступный сайт в сети Интернет;
- 2) разработка таблиц для хранения данных пользователей на основе структурных спецификаций;
- 3) модификация пользовательского интерфейса.

Виртуальные машины создаются на основе шаблонов (в терминологии VMware это "template", в Openstack это "image"), где заранее установлен и сконфигурирован типовой геопортал. Типовой геопортал содержит программную систему ввода и редактирования данных, которая обеспечивает: многопользовательскую работу через Интернет, ввод и редактирование данных разных по структуре реляционных таблиц, и т.д.

3 Создание методологического и инструментального обеспечения поддержки процессов трансформации неструктурированных данных из произвольных таблиц

3.1 Исследования в области понимания таблиц и очистки данных. Систематизация методов очистки данных

Проведен обзор методов сопоставления текстовых данных, представленных на естественных языках. Для очистки такой информации обычно применяются методы нечеткого сравнения строк и фонетического сопоставления. При этом, в случае использования значений, отличных от канонических, представленных в словарях требуется привлечение методов словообразования. Анализ методов фонетического сопоставления показал, что их алгоритмы ориентированы на использование фонетических правил английского языка. Реализованы некоторые модификации алгоритмов, например, для французского, испанского языков [1]. Существуют и другие адаптации фонетических алгоритмов для языков, отличных от английского. Как правило, в этом случае используется транслитерация, а в качестве алгоритма используются вариации Soundex [1,2]. Транслитерация в большинстве случаев не позволяет учесть особенности фонетики искомого языка. Был продолжен и апробирован алгоритм Polyphon, использующий фонетические особенности русского языка.

В качестве развития методов очистки данных предлагается использовать синтез, основанный на применении методов нечеткого сравнения строк и методов фонетического сходства, в том числе посредством разработанного алгоритма. Методы очистки данных,

представленных на естественном языке также использованы для идентификации вычисляемых значений, содержащихся в слабоструктурированных табличных документах [3]. В частности предложенные методы использовались для корректной идентификации слов, описывающих табличные данные.

В настоящее время отмечается востребованность инструментальных и технологических решений, позволяющих обеспечить интеграцию и генерализацию данных. Применение средств данного класса представляет возможность аккумуляции информации полученных из различных источников с учетом её разноформатности. Так, например, в научных, государственных учреждениях существует множество открытых тематических данных, представленных в виде таблиц, в форматах Excel, CSV (Comma Separated Values), MDB (Microsoft Data Base). Как правило, при сборе и хранении данных учитываются лишь требования, обеспечивающие удобство работы для их владельца. В связи с этим актуальна задача интеграции различных тематических данных, представленных в разных форматах и имеющих различную структуру. Для обеспечения интеграции такого рода данных требуется решить ряд задач, одна из которых – очистка данных. Для принятия обоснованных решений необходима надежная система данных. В связи с этим возрастает потребность в новых методах и технологиях для организации процессов интеграции данных. Интегрированные данные имеют некий единый интерфейс, что упрощает доступ к ним. Благодаря большому объему и наличию расширенного числа показателей представляется возможность проведения их анализа с учетом множества критериев, что повышает ценность данных. Также интеграция предоставляет более широкие возможности для дальнейшего проведения совместных научных исследований. Ввиду того, что информация может быть получена из различных источников, данные могут значительно отличаться по формату своей организации и представлению. В первичных документах, являющихся основой для формирования массива данных, встречаются вычисляемые данные. Как правило, они представляют собой различные агрегированные значения, полученные в результате математических операций (сумма, среднее значение, максимальное, минимальное значения, нарастающий итог и пр.) над каким-либо набором данных, т.е. использованы функции обрабатывающие набор значений для подсчета и возвращающих одно значение.

В случае слияния данных, подобные значения являются избыточными, т.к. их всегда можно получить по имеющейся информации. К тому же вычисляемые значения, представленные в электронных таблицах, часто содержат ошибки, что в итоге отражается на качестве интегрированной информации. Таким образом, в рамках вопросов очистки интегрируемой информации представляется важным проведение работ по идентификация

вычисляемых значений и их коррекции или экстрагирования для повышения качества данных.

Выделяют несколько проблем, связанных с необходимостью очистки данных перед проведением процедур консолидации [4]. Это, в частности, корректность форматов и представлений данных; уникальность первичных ключей в таблицах базы данных; полнота и целостность данных; полнота связей; соответствие некоторым аналитическим ограничениям и т.д. Используя методы нечеткого фонетического сравнения можно повысить эффективность решения таких проблем как отсутствие уникальности и целостности данных. Процедура очистки данных позволяет выявить и, в ряде случаев, исправить ошибки и несоответствия в сырых (пользовательских) данных. Такими ошибками являются как дублирование, опечатки, ненужные префиксы, противоречия, пропуски, и т.п.

Целью очистки является улучшение качества данных. В исследовании проведена адаптация фонетического алгоритма для обработки слов на русском языке. При этом в отличие от большинства используемых фонетических алгоритмов, не строится транслитерация, что позволяет учитывать особенности произношения русских слов. Так, фонетическое сопоставление слов может быть использовано для эффективного сопоставления пользовательского текста информации, содержащейся в различных классификаторах. В совокупности с применением методов нечеткого сравнения строк, данный подход может использоваться в алгоритмах очистки данных и представляет интерес для создания интегрированной информационно-аналитической среды, создаваемой «добровольцами» по принципу краудсорсинга. Предложенный алгоритм был адаптирован для монгольского языка. Для монгольского языка можно выделить, что также как и русский язык он использует кириллические буквы. В обоих языках существуют звуки похожие друг на друга, например гласные [а], [о] и согласные [м], [с]. Языки имеют некоторые общие фонетические законы, хотя они не всегда проявляются в каждом из них одинаково. Таким образом, учитывая похожесть алфавита, некоторых фонетических правил – представляется возможной адаптация Polyphon для использования в монгольском языке, не смотря на то, что он не относится к восточнославянской языковой группе. Перед тем как провести кодировку слов проводится модификация строки. Все буквы слова переводятся в нижний регистр; удаляются все символы, не принадлежащие алфавиту; удаляются буквы «ъ», «ь»; проводится замена парных букв. В случае, если на вход подается множество слов, то между ними убираются пробелы и множество слов рассматривается как одно. Некоторые буквы монгольского алфавита имеют сходное написание с буквами латинского алфавита. Это такие буквы как: а, е, к, м, н, о, у, х,

заглавная буква В. Во избежание случайных или преднамеренных замен монгольских букв на латинские, проводится их анализ. Все подобные латинские буквы заменяются соответствующими монгольскими. Все остальные, не принадлежащие языку, символы – удаляются.

Следующим шагом проводится замена букв, которые могут быть созвучны. Дифтонги преобразуются в одну букву: ай, ой, эй -> э; уй, үй -> у. Результаты представлены в таблице №1. При этом оценка похожести звуков, получаемых от гласных букв, делается достаточно грубо.

Таблица 1. Замена некоторых букв

Буквы	а, е, ё, и, о, Ө, ы, э, я	Б	В	г	д	з	щ	ж	м	у,ю
Результат замены	Э	п	Ф	к	т	с	ш	ш	н	у

В современном монгольском языке краткость и долгота гласных имеет смысловоразличительную функцию. В связи с этим осуществляется преобразование более двух идущих подряд букв к двум: ааа - аа, ооо - оо, ууу - уу и т.п. Так, как можно считать, что в данном случае имеет место опечатка. При этом преобразование двух букв к одной не приводится. Так как это может сказаться на смысловой интерпретации слова: цас – цаас, ул – уул и т.п. Замена относится как к гласным, так и к согласным буквам. При этом особенности силлабической структуры не учитываются.

Далее, проводится анализ слова с учетом поведенных ранее возможных замен. Соответствие кодов и букв приведено в таблице № 2.

Таблица 2. Кодирование букв

Буква	К	Л	Н	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Э
Код	2	3	5	7	11	13	17	19	23	29	31	37	41	43

Результат преобразования представляется в виде последовательности простых чисел: «багшийн» -> «пэкшэн» -> **74324143435**.

Для нечеткого фонетического сравнения, по аналогии с русской версией, полученные числовые фонетические характеристики суммируются: «багшийн» -> **7 + 43 + 2 + 41 + 43 + 43 + 5 = 184**. В данном случае не учитывается порядок букв в слове. Применение нечеткого фонетического сопоставления кодов строк позволяет сравнивать слова более гибко.

Для апробации был использовано множество из 11601 различных слов, взятых из словаря Ожегова куда были внесены следующие виды ошибок:

- позиционные изменения;
- ассимиляционные и диссимиляционные процессы.

В результате было проведено сопоставление на слова, с целью идентифицировать их и привести к каноническому виду. Результаты разработанного алгоритма сравнивались с результатами других известных алгоритмов. При этом использовалась транслитерация кириллических символов по ГОСТ Р 52535.1-2006. Результаты представлены в таблице №3.

Таблица 3. Результаты тестирования алгоритма

Алгоритм	Совпадение фонетических кодов (in %)	Время работы (в миллисекундах)
Polyphon	95.12	2003
<i>Polyphon</i> (используется нечеткое сопоставление кодов)	98.8	1623
Soundex	90.24	1096
Metaphone	90.29	870
Double Metaphone	96.15	1451
Caverphone	90.41	9770
NYSIIS	75.97	1517
DaitchMokotoffSoundex	96.84%	1763

3.2 Методологическое, инструментальное обеспечение поддержки процессов концептуализации произвольных таблиц на основе глобальных таксономий понятий, связанных данных и предметных онтологий и очистки табличных данных.

Направление исследования основывается на том, что в мире циркулирует большое количество произвольных электронных таблиц в форматах HTML, Excel, CSV. Современные оценки, сделанные на основе изучения экспериментальных срезов содержания сети Интернет (Common Crawl), показывают, что их количество исчисляется сотнями миллионов [5, 6]. Предположительно они содержат миллиарды фактов [7]. Такие таблицы характеризуются большим разнообразием и разнородностью компоновок, стилей и содержания [8]. С учетом постоянного и быстрого роста объема такой информации её относят к Большим Данным [9]. Большой объем и свойства структуры таких таблиц делают их ценным источником в приложениях науки о данных и бизнес-аналитики. Однако, как правило, они не сопровождаются явной семантикой необходимой для машинной интерпретации своего содержания так, как задумано их автором.

Накапливаемая в них информация часто является неструктурированной и не стандартизированной. Анализ этих данных нуждается в их предварительном извлечении и трансформации к структурированному представлению с заданной формальной моделью. Анализ современного состояния исследований данной области [10] показывает большой интерес к рассматриваемым вопросам со стороны исследователей и разработчиков систем обработки документов, информационного поиска и управления данными. Сегодня продолжают активно развиваться специализированные инструменты для извлечения и трансформации данных из произвольных электронных таблиц в структурированную форму, в том числе, системы трансформации табличных данных, системы извлечения связанных данных [7, 11], система извлечения реляционных данных из таблиц с иерархиями заголовков SENBAZURU [12], система трансформации произвольных таблиц к реляционной форме на основе поиска критических ячеек MIPS [9]. По результатам проведенного исследования предлагается свободная система извлечения и трансформации данных из произвольных электронных таблиц, называемая TabbyXL.

Система извлечения и трансформации данных из произвольных электронных таблиц охватывает задачи автоматического восстановления семантической разметки таблиц, очистки и отслеживания происхождения табличных данных, генерации реляционных данных. Рассматриваемые процессы трансформации табличных данных служат для восстановления недостающих метаданных (семантики) о структуре и содержании произвольной таблицы, позволяя переходить к более структурированному представлению. В отличие от перечисленных решений система ТАВВУХЛ использует объектную модель таблицы, не ограничивающую структуру обрабатываемых таблиц функциональными регионами. Ограничения, которые в существующих решениях встроены в их алгоритмы, мы предлагаем представлять в виде наборов правил анализа и интерпретации таблиц. При этом трансформация данных обеспечивается исполнением этих правил.

Как правило, произвольная электронная таблица не включает метаданные, описывающие её отношения. Неизвестно то, какие роли играют её ячейки (например, содержат ли они данные или атрибуты), как они связаны между собой (например, заголовки и значения данных), какими предметными понятиями (категориями) описывается её содержание. Для того чтобы перейти к структурированному представлению, необходимо, в том числе, выделить внутри содержания произвольной таблицы функциональные единицы (значения, метки, категории, агрегированные значения, сноски, название и др.), восстановить не представленные явным образом отношения между значениями данных (вхождениями) и описывающими их ключами

(метками), соотнести её содержание с понятиями (категориями) внешнего словаря (концептуальной онтологии или тезауруса).

В основе представления исходных и восстанавливаемых фактов о таблице в рабочей памяти системы исполнения правил лежит двухуровневая объектная модель (Рисунок 6). Её физический (синтаксический) уровень описывает компоновочные (координаты в пространстве строк и столбцов), стилевые (шрифтовое форматирование, выравнивание, цветовое выделение и др.) и содержательные (текст, отступы) свойства ячеек. Логический (семантический) уровень определяет связанные функциональные единицы данных: вхождения, метки и категории. Данная модель позволяет представлять таблицы с произвольным расположением функциональных единиц внутри содержания ячеек. Она также предусматривает взаимные ссылки между двумя уровнями. С одной стороны, это обеспечивает отслеживание происхождения данных, с другой позволяет организовать удобный доступ к её объектам.

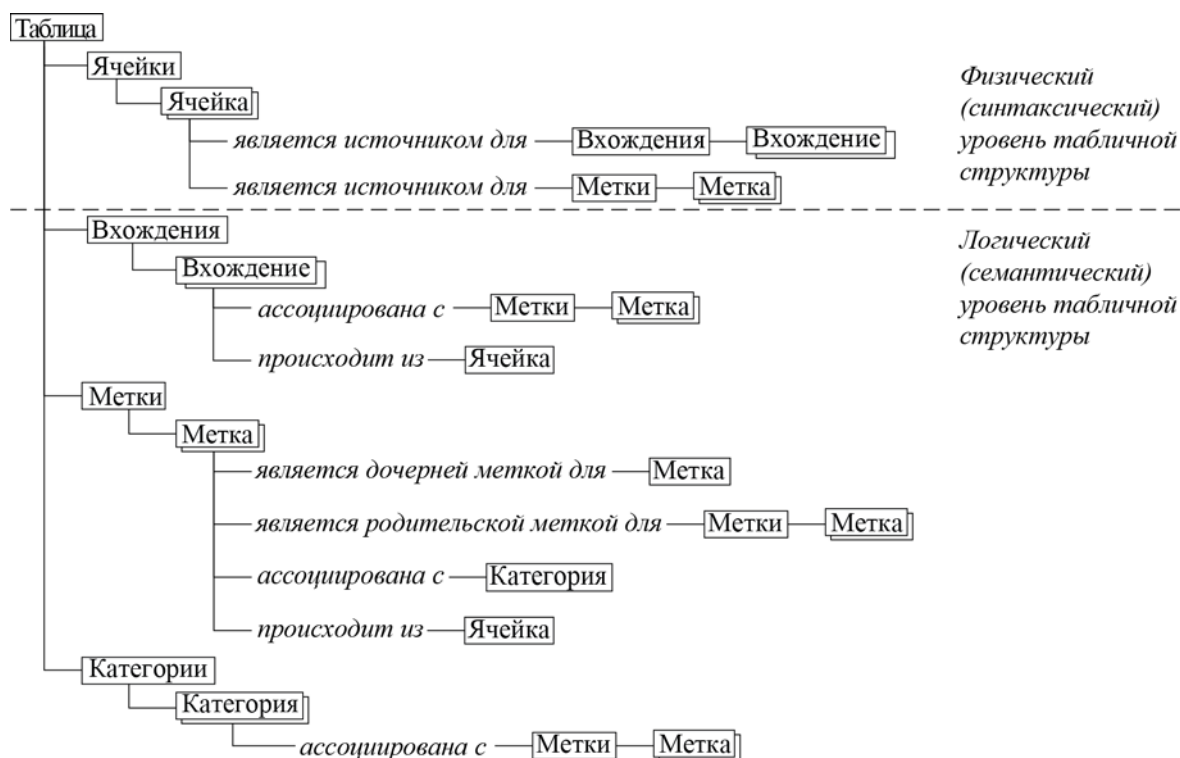


Рисунок 6. Двухуровневая объектная модель произвольной таблицы.

В системе ТАВВУХЛ процесс трансформации табличных данных строится как последовательность следующих этапов (шагов) анализа и интерпретации таблиц.

1. *Реформатирование и очистка табличных данных* необходима для изменения синтаксической структуры и содержания ячеек, в тех случаях, они представлены неаккуратно, содержат ошибки, несогласованности и неточности. При этом для очистки данных могут привлекаться различные готовые алгоритмы обработки строк.

2. *Ролевой анализ таблицы* состоит в том, чтобы восстановить отдельные функциональные единицы: вхождения (значения данных) и метки (ключи, описывающие значения данных) из естественно-языкового содержания ячеек.
3. *Структурный анализ таблицы* позволяет ассоциировать (связать) между собой вхождения и метки, т.е. восстановить пары «вхождения-метка» и «метка-метка».
4. *Интерпретация таблицы* служит для того, чтобы определить категории меток, т.е. восстановить пары «метка-категория».

В системе TABBYXL каждый шаг может быть выражен одним или несколькими продукционными правилами, отображающими известные факты (объекты) обоих уровней в неизвестные изначально факты. Такие правила могут быть выражены на предметно-ориентированном языке правил CRL (Cells Rule Language), который ограничен рассматриваемым предметом — пониманием таблиц.

Один набор CRL правил может охватывать широкий диапазон таблиц, разделяющих общий набор свойств, например, статистических таблиц или паспортов безопасности продуктов. Предполагается, что такие таблицы имеют различную на просвет структуру ячеек, однако разделяют общие неизменяемые компоновочные, стилевые или содержательные свойства, например, наличие критической ячейки, разделяющей таблицу на функциональные области, постоянное количество отступов, идентифицирующих некоторый уровень в иерархии заголовков, или числовой литеральный тип для всех значения данных.

В процессе исполнения CRL правила отображают имеющиеся изначально факты о компоновке, стилях и содержании ячеек таблицы в отсутствующие факты о её семантике (вхождениях, метках, категориях и отношениях между ними). Все факты рабочей памяти системы исполнения правил представляются с помощью объектной модели таблиц. Реализация процессов анализа и интерпретации данных не обеспечивает абсолютной точности и полноты. Трансформация произвольных таблиц может приводить к накоплению ошибок данных. Поэтому TABBYXL позволяет отслеживать происхождение данных. Эта информация накапливается в представленной объектной модели таблицы.

Архитектура TABBYXL представлена на Рисунке 7. Трансляция и исполнение CRL правил реализовано в свободной системе Drools Expert (<http://drools.jboss.org/drools-expert>). В результате трансформации из восстановленной семантической информации генерируются электронные таблицы в реляционной (канонической) форме, которая включает поле DATA — данные (вхождения) и поля CAT_1,..., CAT_n меток для каждой восстановленной категории. Исходные коды проекта разработки системы TABBYXL опубликованы под свободной лицензией по адресу: <https://github.com/cellsrg/tabbyxl>.

Оценка производительности TABBYXL выполнена на известном тестовом наборе TROY200 (http://tc11.cvc.uab.es/datasets/Troy_200_1) из 200 произвольных таблиц, собранных с 10 различных сайтов государственной статистики. Разработан набор из 16 CRL правил. Для оценки восстановления вхождений, меток, пар «вхождения-метка» и «метка-метка» использовались стандартные метрики, используемые в информационном поиске: полнота, точность и F-мера.

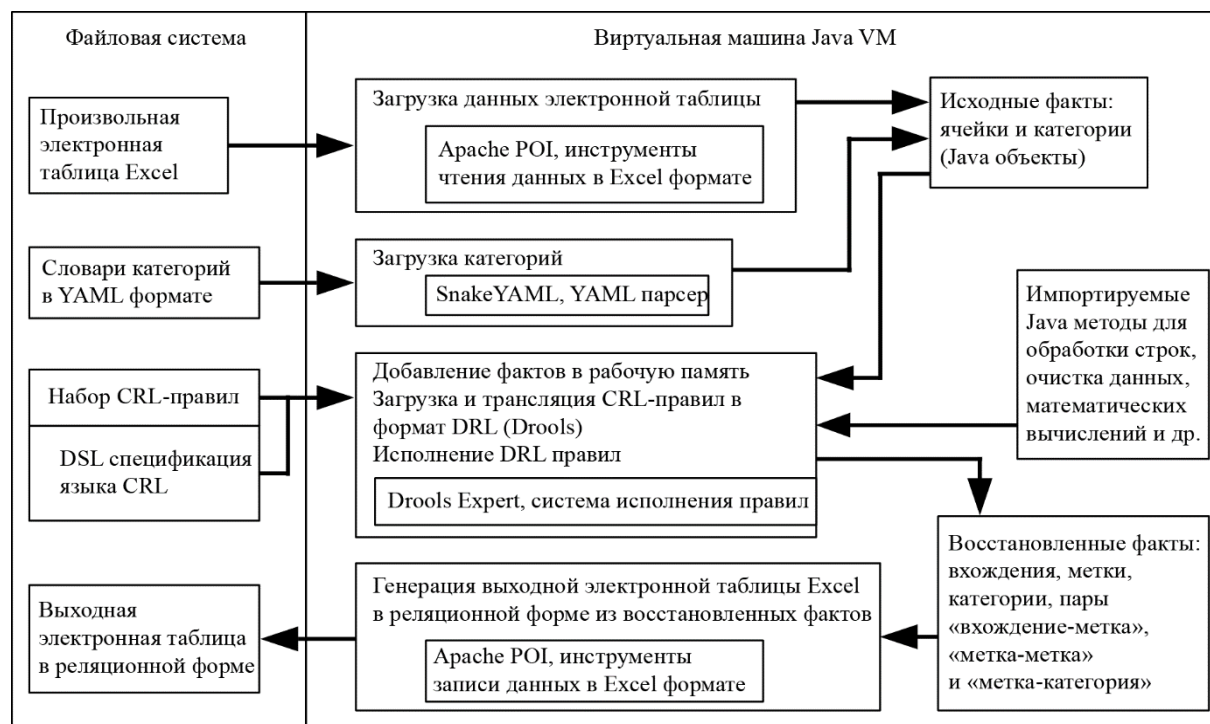


Рисунок 7. Архитектура системы трансформации данных произвольных электронных таблиц TABBYXL.

Полученные результаты приводятся в таблице 4. Исходные произвольные и соответствующие эталонные реляционные таблицы, шаги воспроизведения, CRL правила и результаты эксперимента опубликованы в виде набора данных (<https://data.mendeley.com/datasets/448jdx7gcr/1>).

Таблица 4. Экспериментальные результаты.

	Вхождения	Метки	Пары «вхождения-метка»	Пары «метка-метка»
Полнота	0,9813	0,9979	0,9773	0,9389
Точность	0,1000	0,9365	0,9966	0,9784
F-мера	0,9906	0,9662	0,9869	0,9582

Экспериментальные результаты показывают, что TABBYXL может применяться для массовой обработки произвольных таблиц. При этом один набор правил может разрабатываться для широкого диапазона таблиц, разделяющих общие неизменяемые компоновочные, стилевые и содержательные свойства.

Таким образом, в работе продемонстрированы полученные новые возможности в трансформации данных из произвольных электронных таблиц в реляционную форму на основе исполнения правил их анализа и интерпретации. Представленная двухуровневая

объектная модель таблицы описывает, как физическую (синтаксическую), так и логическую (семантическую) табличную структуру. В отличие от известных подходов она не опирается на функциональные регионы (например, боковик, шапка и тело), а напротив предполагает произвольное расположение функциональных единиц содержания внутри ячеек.

CRL, предметно-ориентированный язык, служит для выражения правил анализа и интерпретации таблиц. По сравнению с известными языками трансформации данных он выстраивает этот процесс как последовательность шагов автоматического понимания таблиц: их ролевого анализа, структурного анализа и интерпретации. В отличие от языков правил общего назначения CRL позволяет фокусироваться на предмете (понимании таблиц), скрывая не существенные детали.

TabbyXL, система трансформации произвольных электронных таблиц, реализует предложенную модель и язык правил. В будущем TabbyXL может интегрироваться с инструментами извлечения таблиц, например, TabbyPDF, для выстраивания сквозного процесса восстановления семантических данных в системах анализа и распознавания документов. Перспективным направлением дальнейшего развития является создание на её базе инструментальной платформы извлечения и трансформации семантических данных из произвольных таблиц.

4 Исследование методов и технологий создания проблемно-ориентированных и интеллектуальных систем поддержки принятия решений, основанные на использовании модельно-управляемого подхода, порождающего программирования и декларативных описаний, в контексте решения задач управления риском техногенных катастроф

4.1 Модели информационных процессов создания и функционирования продукционных экспертных систем, систем синтеза баз знаний, агентных имитационных моделей и сервиса визуализации результатов имитационного моделирования, учитывающие специфику модельно-управляемого подхода.

В ходе выполнения проекта были разработаны модели информационных процессов создания и функционирования интеллектуальных систем поддержки принятия решений,

основанные на реализации модельных трансформаций для систем различного типа. Трансформация моделей является одной из основных составляющих модельно-ориентированного (-управляемого) подхода к разработке программного обеспечения (Model-Driven Engineering, MDE). В рамках данного подхода, процесс разработки программного обеспечения представляет собой последовательное преобразование моделей с разным уровнем детализации, где на завершающем этапе генерируется исходный код программы или ее спецификации.

Использование модельных трансформаций и принципов модельно-управляемого подхода позволяет максимально вовлечь непрограммирующих специалистов в процесс создания программного обеспечения с минимизацией ошибок программирования за счет использования когнитивной графики и элементов визуального программирования, а также автоматической генерации кодов и спецификаций.

4.2 Модель информационного процесса создания продукционных баз знаний на основе концептуальных моделей.

С целью создания продукционных баз знаний с использованием модельных трансформаций разработана модель информационного процесса, особенностью которой является комплексное использование: XML подобных форматов для представления концептуальных моделей; онтологии в качестве универсального промежуточного хранения знаний; авторской нотации RVML (Rule Visual Modeling Language) – для описания продукций. Предлагаемая модель позволяет учесть в цепочке модельных трансформаций специфику процесса разработки интеллектуальных систем, такую как наличие этапов концептуализации и формализации.

Формально разработанная модель представлена следующим образом:

$$T : CM \rightarrow KB$$

где T - оператор преобразования, CM – исходная концептуальная модель; KB – целевая база знаний.

При этом множество информационных моделей ограничено XML-подобными форматами, в частности, для представления: UML моделей - XMI (XML Metadata Interchange), концепт-карт - CXL (Concept Mapping Extensible Language), деревьев событий - ETXL (Event Tree Mapping Extensible Language), таким образом: $CM = \langle CM^{XMI}, CM^{CXL}, CM^{EXTL} \rangle$. В качестве целевого языка программирования баз знаний выбран CLIPS (C Language Integrated Production System), тогда $KB = \langle Code^{CLIPS} \rangle$.

Согласно MDE процесс создания баз знаний на основе концептуальных моделей может быть представлен последовательностью трансформаций: $CM \rightarrow Ont \rightarrow RM \rightarrow KB$ где CIM - вычислительно-независимая модель, RM - модель продукций.

Таким образом, оператор преобразования концептуальной модели:

$$T = \langle T_{CM-Ont}, T_{Ont-RM}, T_{RM-KB} \rangle.$$

В качестве программных средств реализации модели информационного процесса предполагается использовать авторские инструментальные средства Knowledge Base Development System (KBDS) и Personal Knowledge Base Designer (PKBD).

Предлагаемый информационный процесс создания баз знаний на основе концептуальных моделей рассмотрен на Рисунке 8.

Предлагаемая модель информационного процесса создания продукционных баз знаний является одним из элементов модель информационного процесса создания экспертных систем.

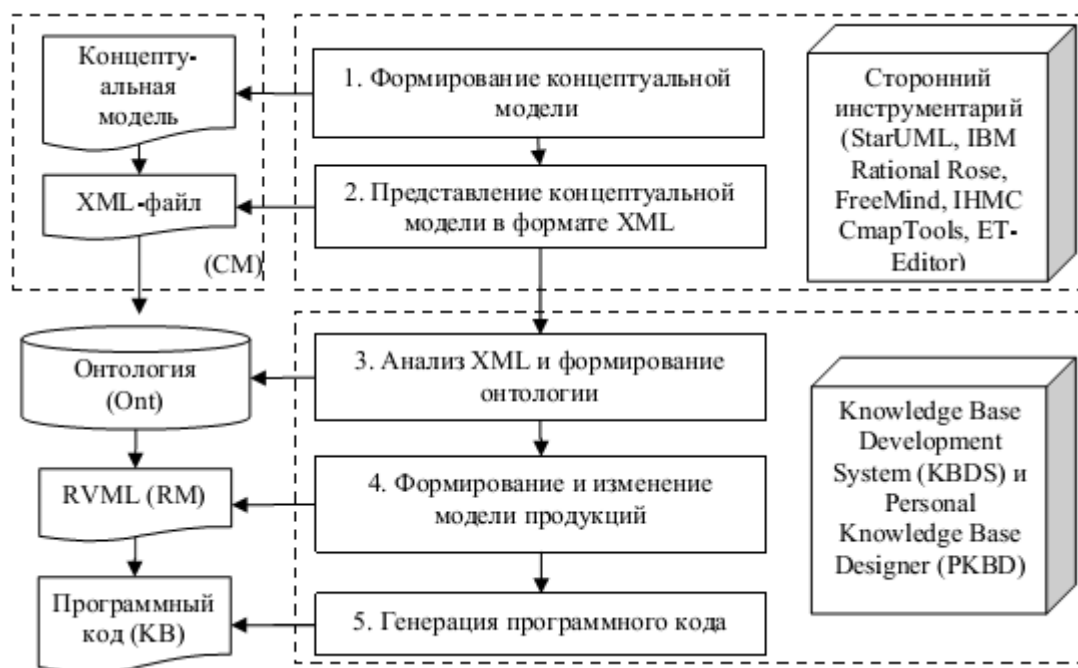


Рисунок 8. Модель информационного процесса создания баз знаний на основе концептуальных моделей

4.3 Модель информационного процесса создания и функционирования системы визуализации результатов имитационного моделирования

Информационный процесс создания и функционирования системы визуализации результатов имитационного моделирования (Рисунок 9) состоит из следующих основных этапов:

- Определение понятий исследуемой области путем создания онтологии предметной области, т.е. описание тех понятий (а также их атрибутов и связей), которые далее будут участвовать в визуализации.

- Создание трехмерных объектов, визуализирующих объекты предметной области, с использованием редактора визуальных объектов, включающего библиотеку стандарта WebGL и библиотеку dat.gui. Визуальное представление большинства реальных объектов условно можно разложить на множество более мелких элементов стандартных форм (куб, цилиндр, сфера, плоскости различной формы и т.д.), поэтому редактор визуальных объектов представляет собой достаточно простой конструктор формирования объектов из геометрических примитивов. В тех случаях, когда одних примитивов недостаточно для создания объекта с необходимой детализацией, объекты могут создаваться на основе полигональной сетки. Их использование позволяет более точно настроить форму объекта, а также дополнять простые объекты через добавление мелких деталей. Описание всех созданных визуальных объектов сохраняется в онтологии визуальных объектов и онтологии визуализации приложения.

- Компоновка созданного набора визуальных объектов в единую сцену. Для этого с помощью редактора визуальных сцен необходимо описать их расположение и поведение на сцене в виде правил. Правила определяют то, как будут расположены все объекты на сцене и их дальнейшее поведение в зависимости от различных параметров (входных данных, полученных от пользователя, или параметров других объектов). Данные правила служат основой для последующей трансляции и генерации кода визуальной сцены.

- Генерация программного кода визуальной сцены на основе полученной структуры всей визуальной сцены. Структура сцены поступает в модуль генерации кода, где на ее основе формируется программный код на языке JavaScript.

- Отрисовка визуальной сцены в браузере (Рисунок 10). Код всех основных функций встраивается в формируемый модулем визуализации код пустой визуальной сцены, тем самым наполняя ее объектами и событиями. Полученная визуальная сцена отображается при помощи браузера и принимает входные данные, определяющие ее поведение, образуя тем самым подсистему визуализации (1).

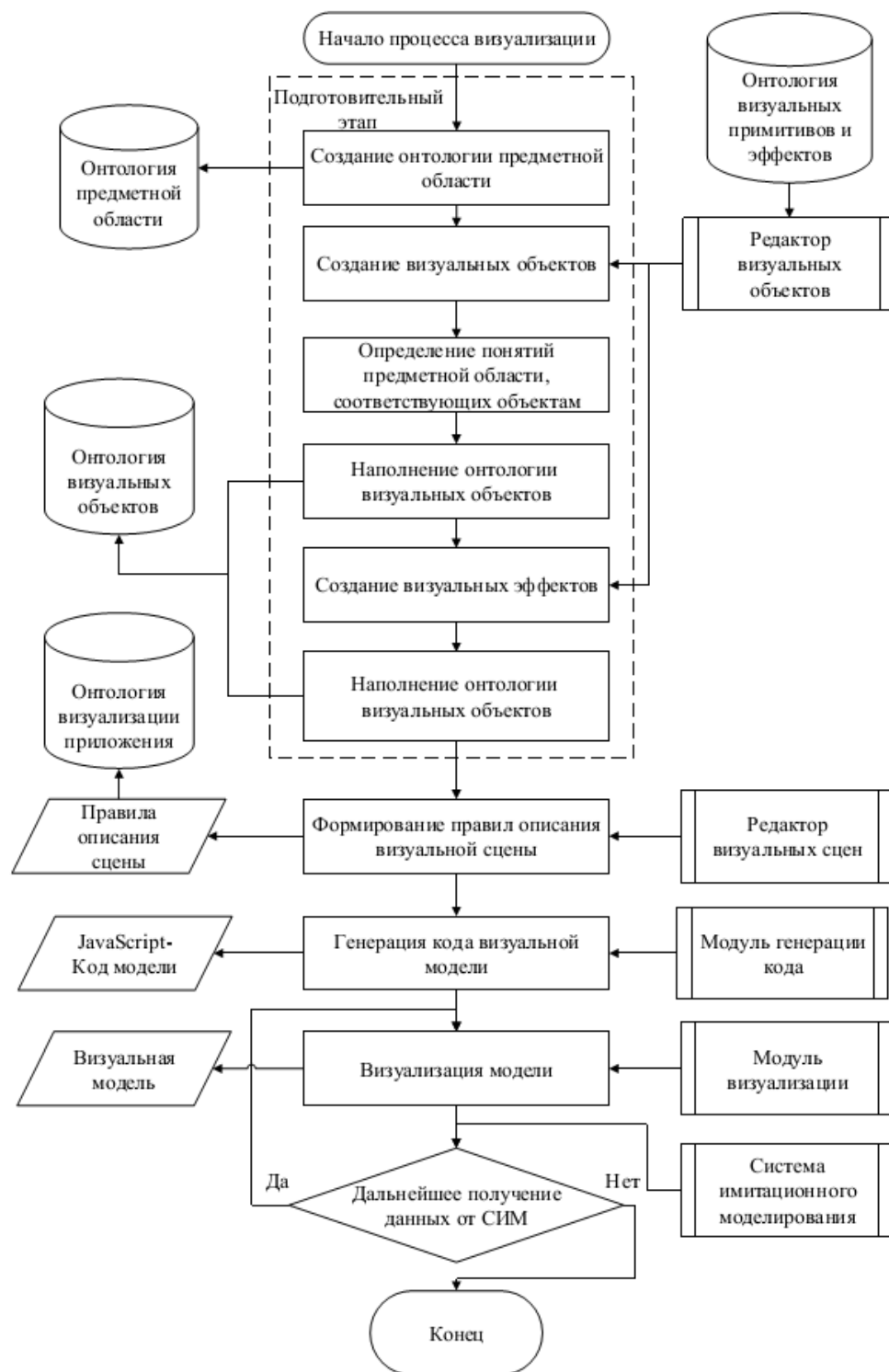


Рисунок 9. Модель информационного процесса создания и функционирования системы визуализации результатов имитационного моделирования

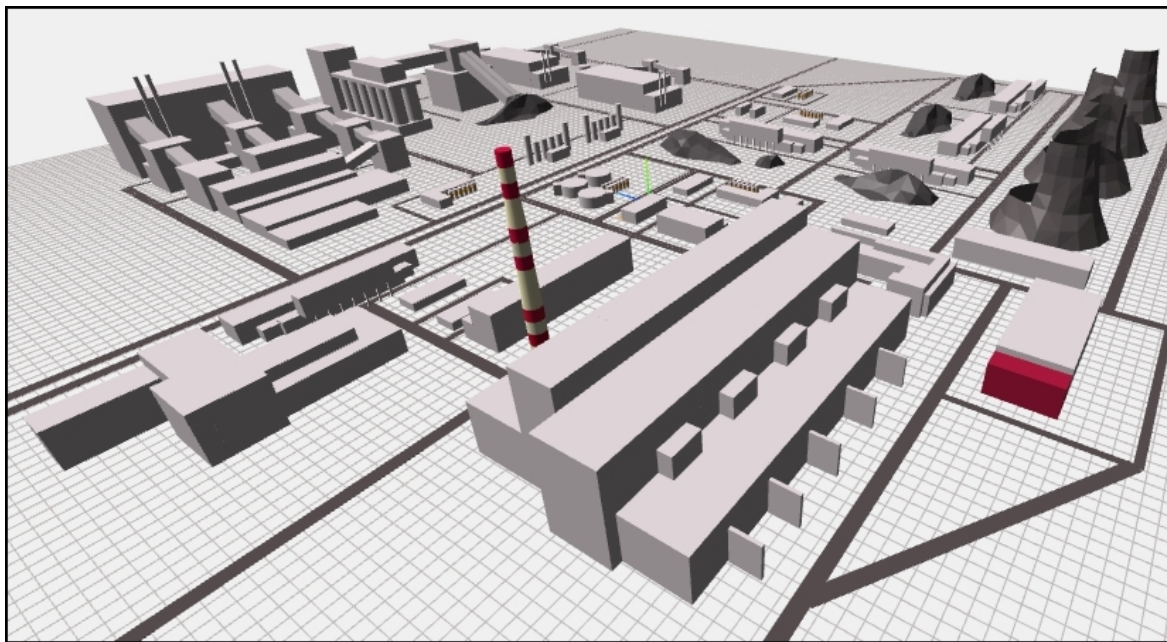


Рисунок 10. Визуальная сцена промышленной зоны

4.4 Модель информационного процесса создания и функционирования системы агентного имитационного моделирования.

Разработана модель информационного процесса создания и функционирования системы агентного имитационного моделирования, обеспечивающая, в соответствии с принципами MDE, возможность автоматизированного создания прикладных агентных имитационных моделей (АИМ). В рамках предложенного информационного процесса используется следующая иерархия метамodelей:

- M3 – определяет наиболее абстрактные элементы для спецификации M2:
 - M^{Ont} – метамодель онтологии,
 - M^{KB} – метамодель продукционной базы знаний,
 - M^{Com} – метамодель операций;
- M2 – спецификация элементов M1 согласно выбранной методологии разработки АИМ:
 - M^{St-A} – метамодель элементов структуры агентной модели,
 - M^{Com-A} – метамодель поведения агентной модели,
 - M^A – предметно-независимая агентно-ориентированная метамодель,
- M1 – спецификация АИМ:
 - M^{Ont-D} – онтология предметной области,
 - M^{KB-D} – база знаний, описывающая поведение объектов предметной области,

- o M^{St-A-D} – структура агентной модели конкретной предметной области,
- o M^{A-D} – агентная модель конкретной предметной области,
- M_0 – имитационная модель во время выполнения:
 - o $codI$ – выполнение спецификации программой-интерпретатором,
 - o $codG$ – сгенерированный исходный код имитационной модели.

Для перечисленных моделей определено следующее множество преобразований:

- o $M^{St-A} \times M^{Com} \rightarrow M^{Com-A}$,
- o $M^{Com-A} \times M^{St-A} \rightarrow M^A$,
- o $M^{KB} \times M^{Ont-D} \rightarrow M^{KB-D}$,
- o $M^{Ont-D} \times M^{St-A} \rightarrow M^{St-A-D}$,
- o $M^A \times M^{St-A-D} \times M^{KB-D} \rightarrow M^{A-D}$,
- o $M^{A-D} \rightarrow codI|codG$.

Информационный процесс включает в себя следующие этапы (Рисунок 11):

- **Онтологическое моделирование структуры агентной модели.** На данном этапе с помощью редактора онтологий необходимо описать понятия, которых будут использованы в качестве элементов (строительных блоков) для описания архитектуры агента, среды и других объектов АИМ.
- **Онтологическое моделирование поведения агентной модели.** На данном этапе с помощью редактора операций необходимо описать поведение активных элементов АИМ (агент, среда), используя в качестве элементов (строительных блоков) методы функциональных компонентов системы агентного имитационного моделирования (компонент работы с базой данных, коммуникационный компонент, компонент рассуждений на основе правил и т.п.). При этом каждый метод функционального компонента является базовой операцией, которые можно объединять в последовательности для реализации более сложного поведения. Операции, состоящие из последовательности базовых, называются производными.
- **Формирование агентной модели.** На данном этапе с помощью редактора агентных моделей осуществляется объединение описания возможных элементов структуры и поведения АИМ в единую метамодель АИМ, которая в дальнейшем будет использована в процессе разработки прикладных АИМ.
- **Онтологическое моделирование предметной области.** На данном этапе с помощью редактора онтологий описываются понятия заданной предметной области, их свойства и связи.

- Формирование структуры агентной модели для заданной предметной области. На данном этапе с помощью редактора агентных моделей осуществляется установка соответствия между понятиями заданной предметной области и понятиями, описывающими элементы структуры АИМ. Например, «Специфицированное изделие» – «Агент», «Датчик давления» – «Сенсор», «Увеличить давление» – «Действие».
- Формирование поведения агентной модели для заданной предметной области. На данном этапе с помощью редактора продукционных баз знаний осуществляется разработка баз знаний, описывающих поведение объектов предметной области. При этом допускается использование уже существующих вычислительных модулей (web-сервисов), вызов которых осуществляется через коммуникационный компонент.
- Формирование агентной модели для заданной предметной области. На данном этапе с помощью редактора агентных моделей происходит конкретизация метамодели АИМ описанием заданной предметной области. При этом происходит формирование структуры прикладной АИМ, путем объединения агентно-ориентированных и предметных свойств, а также выполняется конкретизация поведения, при которой происходит распределение ранее созданных баз знаний и обращений к вычислительным модулям по соответствующим параметрам операций активных элементов прикладной АИМ.
- Агентное имитационное моделирование. На данном этапе с помощью модуля интерпретации моделей осуществляется апробация и тестирование полученной прикладной АИМ.
- Генерация кода агентной модели для заданной предметной области. На данном этапе с помощью модуля трансформации моделей осуществляется генерация отчуждаемого кода прикладной АИМ.

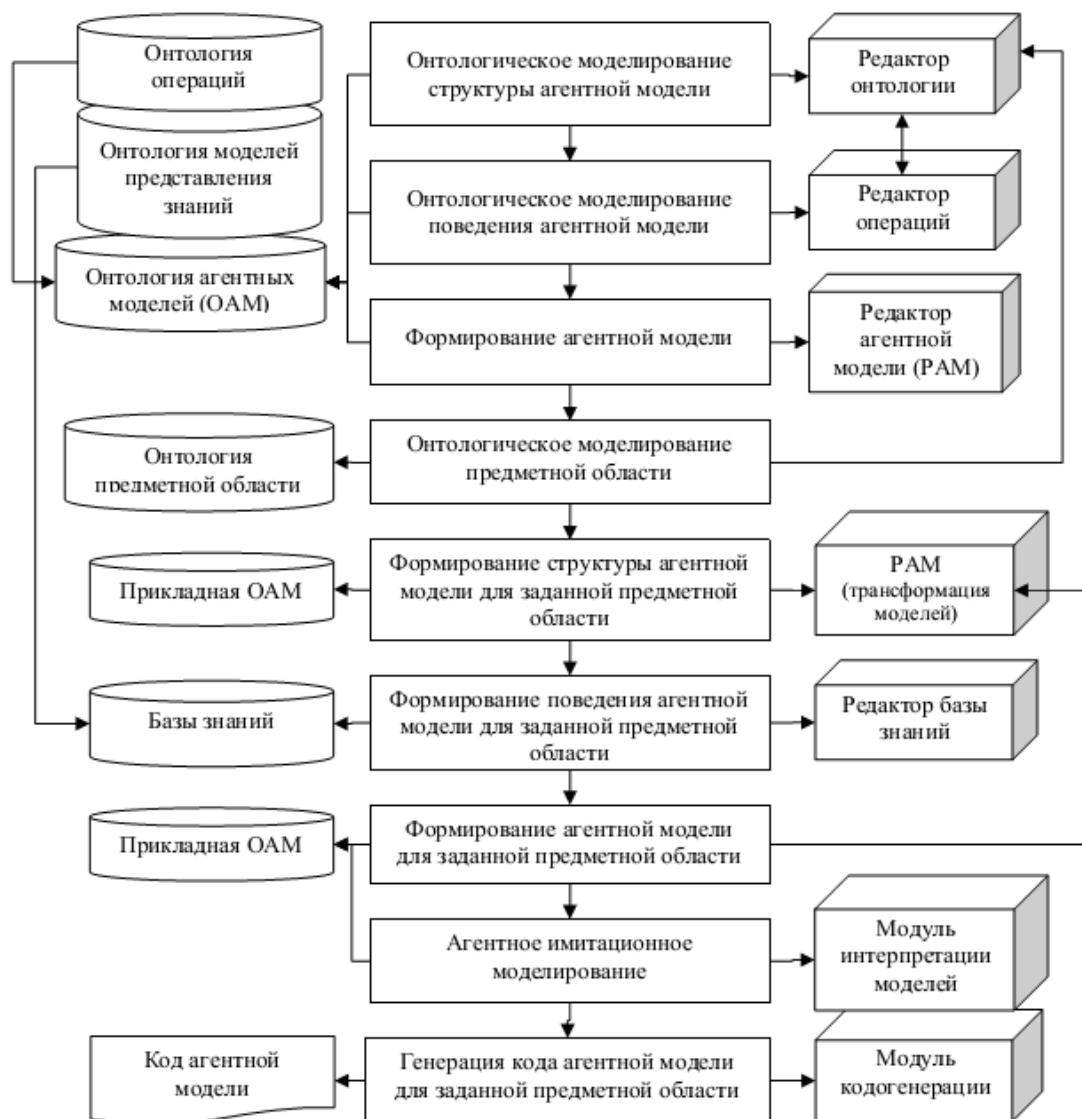


Рисунок 11. Модель информационного процесса создания и функционирования системы агентного имитационного моделирования

4.5 Анализ методов и средств принятия решений, использующих аппарат нечетких мультимножеств в метрическом пространстве А.Б. Петровского.

Проведен анализ методов и средств принятия решений, базирующихся на использовании аппарата мультимножеств, в частности уточнена формулировка понятия мультимножества. Согласно работам Петровского А.Б., Ягера (Yager), Миямото (Miyamoto), Ли (Li), Славина О.А., Тарасова В.Б., Литвиновой А.В., Фуремс.Е.М., Гнеденко Л.С., Ройзензона Г.В., Заболевской-Зотовой А.В., Чеснокова А.М., Воробьева В.В. Паршиковой Е.А., Гусевой М.В., Редько В.И. мультимножество – это множество в повторяющихся элементами. Другими словами, обобщение понятия множества, допускающее включение одного и того же элемента по несколько раз. Впервые описано

мультимножество в работе 1970-х годов Николаса де Брейна. В настоящее время понятие мультимножества активно используется в таких научных направлениях как теория принятия решений, искусственный интеллект. В нашей стране автором первой систематической работы по теории мультимножеств является Петровский А.Б., им же впервые введены операция над мультимножествами, метрики. В области искусственного интеллекта Ягером введено понятие нечеткого мультимножества, которое сейчас успешно применяется для выполнения мягких вычислений.

Задачу принятия решений можно представить в виде кортежа $T = \langle A, C, E, IP \rangle$, где:

- T – задача принятия решений;
- $A = \{A_1, A_2, \dots, A_n\}$, $n \geq 2$ – конечное, заранее заданное множество вариантов, которое не изменяется в процессе решения задачи;
- $C = \{C_1, C_2, \dots, C_m\}$, $m \geq 2$ – конечное, согласованное со всеми заинтересованными сторонами множество оцениваемых характеристик вариантов - критериев;
- $E = \{E_1, E_2, \dots, E_k\}$, $k \geq 1$ – лицо, принимающее решение, или группа заинтересованных лиц, участвующих в процессе принятия решения, – экспертов;
- $IP = \{IP_1, IP_2, \dots, IP_k\}$, $k \geq 1$ – профиль индивидуальных предпочтений, которые могут быть представлены в любой из форм: оценки (количественные, качественные), упорядочения вариантов, парные сравнения (числовые, вербальные значения);

На основе мультимножеств разработаны следующие методы решения задач принятия решения:

- процедуры иерархической и неиерархической кластеризации – позволяют разбивать множество вариантов на произвольное число кластеров,
- процедура построения обобщенного решающего правила для классификации многопризнаковых объектов – предназначена для распределения вариантов по заранее определенным классам,
- метод ранжирования многопризнаковых объектов (АРАМИС) – применяется для одновременной обработки вербальных и/или числовых оценок экспертов,
- технологии ПАКС (ПАКС-М) – обеспечивают снижение размерности признакового пространства, построение нескольких иерархических систем составных критериев и интегрального показателя качества, которые агрегируют исходные признаки, классификацию и/или упорядочение многопризнаковых объектов, используя единственный метод принятия решений (разные комбинации нескольких методов принятия решений).

Все перечисленных методы позволяют решать основные задачи в области принятия решений: выбор лучшего варианта, упорядочение вариантов, классификация. Основными достоинствами методов, базирующихся на аппарате мультимножеств, являются:

- отсутствие необоснованных преобразований,
- их применение не приводит к искажению или потере исходной информации,
- могут применять как при индивидуальном, так и при групповом принятии решений,
- позволяют формировать интерпретацию (обоснование) окончательного решения.

На практике наиболее часто возникают задачи выбора и/или упорядочения вариантов, в которых всю информацию о вариантах (профиль индивидуальных предпочтений), в которых информацию об оценках вариантов можно представить в виде мультимножеств следующего вида: $A_i = \{k_{Ai}(x_1^1) \circ x_1^1, \dots, k_{Ai}(x_1^{h1}) \circ x_1^{h1}, \dots, k_{Ai}(x_m^1) \circ x_m^1, \dots, k_{Ai}(x_m^{hm}) \circ x_m^{hm}\}$, где $k_{Ai}(x_s^{es})$ – число экспертов, давших объекту A_i оценку $x_s^{es} \in X_s, s=1, \dots, m$, $X=\{x_1, x_2, \dots\}$ – шкала оценивания. Такое преобразование мнений экспертов не только не приводит к потере или искажению информации, но и позволяет построить методы обработки с единой методологической основой [Малтугуева Г.С., Юрин А.Ю. Многометодная процедура принятия управленческих решений // Труды XI международной школы-симпозиума «Анализ, моделирование, управление развитие социально-экономических систем», г. Судак, Россия, 14-27 сентября 2017 г. – Симферополь: ИП Корниенко, 2017. – С. 254 - 258]. Отметим, что в литературе описаны разнообразные примеры использования методов, базирующихся на теории мультимножеств, а также использующих данную форму представления разнообразной информации (тексы, базы данных).

С целью повышения эффективности принятия решений разработаны разнообразные программные продукты, например: ГАС «Выборы», «СВИРЬ», «Выбор», «Общий мозг», «ПОИСК-ИТ», «ДУМА», система стратегического планирования предвыборной кампании, «ОЦЕНКА и ВЫБОР», системы электронного голосования и др. Однако в литературе отсутствует описание информационной системы, реализующей методы с использованием мультимножеств. Для повышения эффективности применения таких методов используется табличный редактор Microsoft Excel. В связи с этим задача разработки полноценного программного продукта, реализующего методы с использованием аппарата мультимножеств, является актуальной, которая позволит:

- оказывать полноценную поддержку лицу, принимающему решение,
- решать разнообразные задачи принятия решений.

4.6 Структура системы управления исследованиями монодисциплинарных задач обеспечения приемлемого техногенного риска.

В рамках решения поставленной задачи определена модель трансдисциплинарной задачи (ТДЗ) согласно модели динамики технического состояния сложного опасного объекта (Рисунок 12). Элементами декомпозиции являются дисциплинарные (монодисциплинарные) задачи (ДЗ) и междисциплинарные задачи (МДЗ).

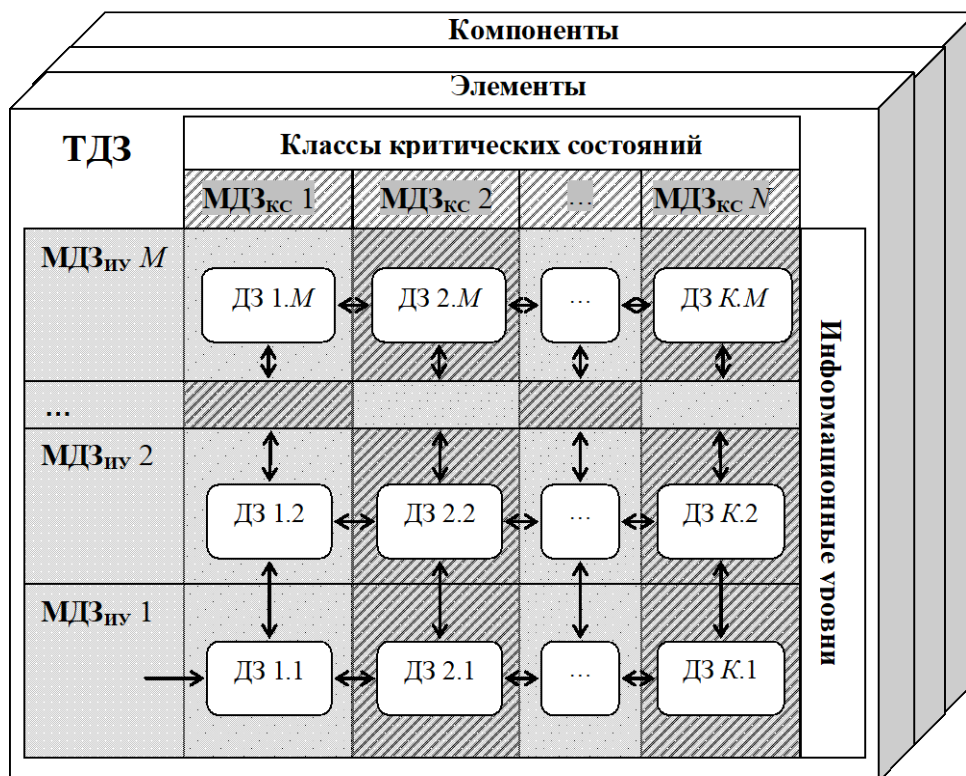


Рисунок 12. Трансдисциплинарная задача (ТДЗ) обоснования допустимого значения риска сложного технического объекта: (МДЗИУ 1 – МДЗИУ М) – междисциплинарные задачи информационного уровня; (МДЗКС 1 – МДЗКС М) – междисциплинарные задачи классов критических состояний; (ДЗ 1.1 – ДЗ К.М) – дисциплинарные (монодисциплинарные) задачи.

Монодисциплинарные задачи предназначены для обеспечения приемлемого риска Классов Критических Состояний (ДЗ 1.1 – ДЗ К.М) (ККС) (рис.1) и требуют для своего решения условно одну-пять дисциплин. Например, для решения задачи исследования и обеспечения требуемых параметров класса «Исходное состояние» достаточно знаний, входящих в специальности «Детали машин и основы конструирования», «Материаловедение», «Сопротивление материалов». Количество параметров, при решении монодисциплинарных задач, колеблется в пределах от 100 до 300 и зависит от уровня сложности СОО. Монодисциплинарные задачи одновременно принадлежат как к

МДЗ соответствующих информационных уровней (горизонтالي), так и к МДЗ уровней критичности состояний (вертикали).

Предложенная модель ТДЗ положена в основу создаваемой системы научных исследований. Для ее управления разработана структура системы управления (СУ) процессом исследования для решения монодисциплинарных задач обеспечения приемлемого техногенного риска рассматриваемого Класа Критического Состояния (ККС) сложного опасного объекта.

Структура системы управления монодисциплинарной задачи определяется информационным, математическим и интеллектуальным обеспечением, а также алгоритмом управления (Рисунок 13). Первое базируется на онтологическом представлении информации, в частности, на разработанных онтологических моделях дисциплин, задач, методов и средств их реализации. Второе – на алгоритме самоорганизации решателей монодисциплинарных задач.

Целью управления является обеспечение обобщенного параметра класса (параметра порядка), характеризующего техногенный риск – R .

Предлагается обобщенный параметр риска рассматриваемого класса критического состояния (ККС) вычислять по формуле:

$$\mathfrak{R} = \prod_{i=1}^n r_i k_i \leq [R],$$

где $\mathfrak{R}$ – обобщенный параметр риска ККС; $[R]$ – допустимое значение обобщенного параметра риска ККС; r_i – риск параметра состояния; k_i – коэффициент весомости параметра; n – количество параметров класса.

Инициализация системы управления осуществляется в случае превышения параметром R допустимого значения для рассматриваемого класса. В этом случае, система управления на основе самоорганизующегося алгоритма формирует вычислительную структуру, которая реализуются решателем дисциплинарных задач.

Структуру системы управления исследованиями представим в следующем виде (рис.8):

$$CS_DSS = (Dt, Knl, Ont, MS, P, \hat{P}, Slv, Crd, R_{DSS}, Ind, Pln)$$

где Dt – базы данных, Knl – базы знаний, Ont – онтология предметной и проблемной областей, $Ont \rightarrow \{P, \hat{P}, Dt, Knl, R_{DSS}\}$, MS – системы измерений, P – задачи, $\hat{P}$ – иерархия задач, Slv – «решатели», Crd – «координаторы» задач, R_{DSS} – отношения

между компонентами CS_DSS , $R_{DSS} = \{R^{PP}, R^{PCrd}, R^{PSlv}\}$, R^{PP} – отношения между задачами, R^{PCrd} – отношения между задачами и «координаторами», R^{PSlv} – отношения между задачами и «решателями», Ind – индикаторы состояния, $Ind = \{I^{DefP}, I^{SP}, I^{Un}, I^{Risk}\}$ – множество индикаторов формулировки, состояния, неопределенности задачи и риска, Pln – планировщик, реализующий самоорганизующийся алгоритм $SAlg$ решения задачи на основе локальных правил $LRule$, $Pln = \langle LRule, SAlg \rangle$.

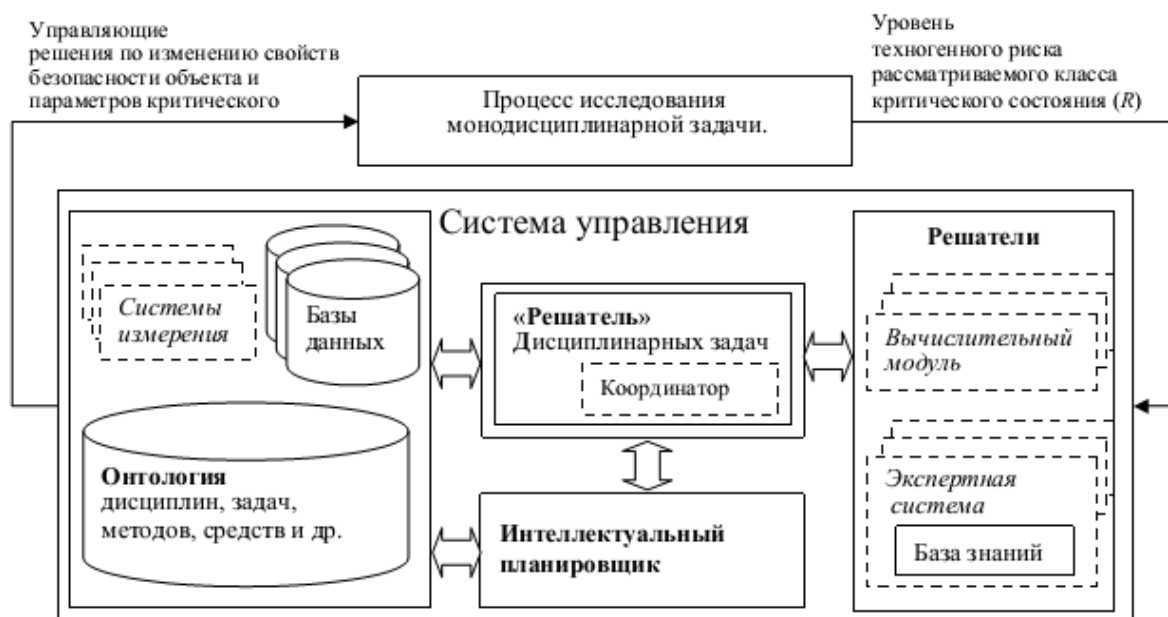


Рисунок 13. Структура системы управления исследованиями монодисциплинарных задач

ЗАКЛЮЧЕНИЕ

Все исследования в рамках этапа НИР 2017 г. выполнены в соответствии с государственным заданием ИДСТУ СО РАН на 2017-2020 гг. по теме IV.38.1.2 «Методы и технологии создания распределенной сервисно-ориентированной среды сбора, хранения, обработки больших объемов разноформатных междисциплинарных научных данных и знаний, основанные на конструктивных средствах спецификации, порождающем программировании и интеллектуализации» (№ госрегистрации АААА-А17-117032210079-1). Содержание НИР раскрыто в Планах научно-исследовательских работ ИДСТУ СО РАН на 2017-2020 гг.

Запланированные на 2017 год задачи выполнены полностью и создают основу фундаментальных исследований методов и технологий автоматизации разработки программного обеспечения, которые основаны на развитии и применении декларативных спецификаций, моделей программных систем и логических методов. Эффективность использования спецификаций обусловлена тем, что они позволяют отделить информацию о специфике создаваемых систем от программного кода, реализующего конкретные алгоритмы обработки информации. В проекте исследовались технологии применения управляемых декларативными спецификациями для создания сервисов обработки пространственных и тематических данных, алгоритмов очистки данных, работы с неструктурированными данными, анализа, обработки, хранения информации, представленной, как в СУБД, так и в бинарных файлах.

В ходе выполнения НИР по Блоку 1 «Исследование методов создания технологий обработки и анализа больших объемов разноформатных данных и программного кода, основанных на использовании конструктивных средств спецификации» разработаны:

- Оригинальные иерархии классов для абстрагирования технологий доступа к данным в настраиваемых при помощи спецификаций приложений БД АРМах;
- Эффективные алгоритмы анализа потоков данных в дизассемблере программного кода из файлов DCU;
- Новые алгоритмы геокодирования в условиях неточной информации;
- Оригинальная методика приобретения (acquisition) структурного описания документа на основе анализа его изменений.

По Блоку 2 «Создание методов формирования распределенной сервисно-ориентированной среды, а также технологий и программных систем формирования композиции сервисов в виде Javascript сценариев на основе анализа спецификаций существующих сервисов и данных» на этапе 2017 года проведено исследование существующих стандартов спецификаций сервисов и данных и формирования

инфраструктуры пространственных данных и предложена технология создания информационных систем (ИС) в облачной среде, использующая геопортальный подход.

В рамках исследований по Блоку 3 «Создание методологического и инструментального обеспечения поддержки процессов трансформации неструктурированных данных из произвольных таблиц»:

- проведен аналитический обзор современного состояния в области понимания таблиц, систематизированный по моделям, методам, оценкам и приложениям.
- проведен обзор и систематизация методов для эффективной и аккуратной очистки данных представленных на различных естественных языках.
- предложено оригинальное методологическое, инструментальное обеспечение поддержки процессов концептуализации произвольных таблиц на основе глобальных таксономий понятий, связанных данных и предметных онтологий и очистки табличных данных.

Результатами исследований по Блоку 4 «Исследование методов и технологий создания проблемно-ориентированных и интеллектуальных систем поддержки принятия решений, основанные на использовании модельно-управляемого подхода, порождающего программирования и декларативных описаний, в контексте решения задач управления риском техногенных катастроф» стали:

- Модели информационных процессов создания и функционирования продукционных экспертных систем, систем синтеза баз знаний, агентных имитационных моделей и сервиса визуализации результатов имитационного моделирования, учитывающие специфику модельно-управляемого подхода.
- Модель информационного процесса создания продукционных баз знаний на основе концептуальных моделей.
- Модель информационного процесса создания и функционирования системы визуализации результатов имитационного моделирования
- Модель информационного процесса создания и функционирования системы агентного имитационного моделирования.
- Анализ методов и средств принятия решений, использующих аппарат нечетких мультимножеств в метрическом пространстве А.Б. Петровского.
- Структура системы управления исследованиями монодисциплинарных задач обеспечения приемлемого техногенного риска.

По результатам этапа НИР 2017 года опубликовано 83 работы, из них **11** статей в журналах, включенных в международные базы цитирования Web of Science и Scopus, **29** статей в изданиях, включенных в базу данных РИНЦ, а также получено **6** свидетельств о

государственной регистрации программ для ЭВМ в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Alotaibi Yousef, Meftah Ali Review of distinctive phonetic features and the Arabic share in related modern research // Turkish Journal of Electrical Engineering & Computer Sciences 2013, Vol. 21 Issue 5, pp.1426-1439.
2. Howida A. Shedeed A New Intelligent Methodology for Computer based Assessment of Short Answer Question based on a new Enhanced Soundex phonetic Algorithm for Arabic Language // International Journal of Computer Applications (0975 – 8887) Volume 34– No. 10, November 2011. pp. 40-47.
3. Парамонов В. В., Шигаров А. О. Идентификация вычисляемых значений в слабоструктурированных табличных документах // Материалы международной научной конференции "Информационные технологии и системы-2017" (ИТС-2017) Международная научная конференция "Информационные технологии и системы-2017" — Минск: БГУИР, 2017. — С. 256 – 257.
4. David Loshin. Master Data Management. Morgan Kaufmann, 2010. 304 p.
5. Eberius J., Braunschweig K., Hentsch M., Thiele M., Ahmadov A., Lehner W. Building the dresden web table corpus: A classification approach // Proc. IEEE/ACM 2nd Int. Symposium on Big Data Computing, 2015, pp. 41-50. DOI:10.1109/BDC.2015.30.
6. Lehmberg O., Ritze D., Meusel R., Bizer C. A large public corpus of web tables containing time and context metadata // Proc. 25th Int. Conf. Companion on World Wide Web, 2016, pp. 75-76. DOI:10.1145/2872518.2889386.
7. Galkin M., Mouromtsev D., Auer S. Identifying web tables: Supporting a neglected type of content on the web // Proc. 6th Int. Conf. Knowledge Engineering and Semantic Web, 2015, pp. 48-62. DOI: 10.1007/978-3-319-24543-0_4.
8. Braunschweig K. Recovering the Semantics of Tabular Web Data, Ph.D. thesis, Technischen Universitt Dresden, Dresden, Germany, 2015.
9. Embley D.W., Krishnamoorthy M.S., Nagy G., Seth S. Converting heterogeneous statistical tables on the web to searchable databases // Int. J. Document Analysis and Recognition. 2016. vol. 19, pp. 119-138. DOI: 10.1007/s10032-016-0259-1.
10. Shigarov A., Mikhailov A. Rule-Based Spreadsheet Data Transformation from Arbitrary to Relational Tables // Information Systems. Elsevier. 2017. Vol. 71, pp. 123-136. DOI: 10.1016/j.is.2017.08.004.
11. Fiorelli M., Lorenzetti T., Pazienza M.T., Stellato A., Turbati A., Sheet2RDF: a flexible and dynamic spreadsheet import&lifting framework for RDF // Proc. 28th Int. Conf. Industrial, Engineering and Other Applications of Applied Intelligent Systems, 2015, pp. 131-140. DOI: 10.1007/978-3-319-19066-2_13.
12. Chen Z. Information Extraction on Para-Relational Data, Ph.D. thesis, Univ. of Michigan, US, 2016.

СПИСОК ПУБЛИКАЦИЙ

1. Фёдоров Р. К., Шумилов А. С., Авраменко Ю. В. Обработка векторных данных с помощью спецификаций в соответствии с моделью MapReduce // Вестник Бурятского государственного университета. – 2017. – Т. 2. – С. 12 – 19.
2. Фёдоров Р. К., Шумилов А. С. Сценарий расчета временной доступности объектов образования // Вестник Бурятского государственного университета. – 2017. – Т. 2. – С. 20 – 32.
3. Фёдоров Р. К., Шумилов А. С. Расчет временной доступности для географических объектов с помощью системы распределенных сервисо-ориентированных вычислений // Информационные и математические технологии в науке и управлении. – 2017. – Т. 4. – С. 94-104.
4. Шумилов А. С., Авраменко Ю. В. Спецификация распараллеливания обработки векторных данных в модели MapReduce // Информационные и математические технологии в науке и управлении. – 2017. – Т. 4. – С. 85- 93.
5. Черкашин Е.А., Орлова И.В. Инструментарий создания цифровых архивов документов на основе связанных данных //Современные технологии. Системный анализ. Моделирование. – 2017. – С. 100-107.
6. Берман А.Ф., Николайчук О.А., Павлов А.И., Юрин А.Ю. Обеспечение надежности и безопасности химических и нефтехимических производств методами искусственного интеллекта. Часть 2 // Прикладная информатика. – 2017. – Т. 12, №1 (67). – С. 26-38.
7. Дородных Н.О., Николайчук О.А., Юрин А.Ю. Автоматизированное создание продукционных баз знаний на основе деревьев событий // Информационные и математические технологии в науке и управлении. – 2017, №2 (6). – С. 30-41.
8. Дородных Н.О., Юрин А.Ю. Использование концепт-карт для автоматизированного создания продукционных баз знаний // Программные продукты и системы. – 2017. – Т. 30. № 4. – С. 658-662.
9. Коршунов С.А., Николайчук О.А., Павлов А.И. Программно средство визуализации трехмерных сцен на основе онтологий // Научная визуализация. – 2007. – №2. – С. 102–119.
10. Николайчук О.А., Берман А.Ф., Павлов А.И. Прогнозирование технического состояния опасных объектов методом имитационного моделирования // Проблемы машиностроения и надежности машин. – 2017. – №2. – С. 131–142.

11. Феоктистов А.Г., Башарина О.Ю., Дядькин Ю.А., Фереферов Е.С. Автоматизация распределенного имитационного моделирования систем массового обслуживания // Вестник ИрГТУ. – 2017. – № 12. – С. 105-113
12. Михайлов А. А., Хмельнов А. Е. Декомпиляция объектных файлов *.dcuil // Труды Института системного программирования РАН. – 2017. – Т.29, вып.6. – С. 124-136
13. Mihailov A., Shigarov A. Rule-based spreadsheet data transformation from arbitrary to relational tables // Information Systems. – 2017. – Т. 71. – С. 123 – 136.
14. Voronov V. A., Vetrov A. A., Banshikov A. V., Rusanov V. A., Daneev A. V. A posteriori simulation of dynamic model of the elastic element of satellite-gyrostet // Far East Journal of Mathematical Sciences. – 2017. – Т. 101, № 9. – С. 2079 – 2094.
15. Бычков И.В., Ружников Г.М., Фереферов Е.С., Хмельнов А.Е., Будням С. Технология создания информационных систем на основе декларативных спецификации // Бизнес и Инновации, Улан-Батор. – 2017. – №3. – С. 23 – 41.
16. Будням С., Батбилэг С., Баянжаргал Д., Ружников Г.М., Федоров Р.К. Макроэконометрическое моделирование экономики Монголии // Бизнес и Инновации, Улан-Батор. – 2017. – №3. – С. 32 – 47.
17. Будням С., Батбилэг С., Баянжаргал Д., Ружников Г.М., Федоров Р.К. Математическая модель производственного потенциала и оценки технологической эффективности регионального производства// Бизнес и Инновации, Улан-Батор. – 2017. – № 6(08). – С. 21-32 (на монгольском языке).
18. Bychkov I., Gachenko A., Ruzhnikov G., Hmelnov A. 3-D modeling of Angara river bed // Proceedings of the International Conference Mathematical and Information Technologies, MIT-2016. Vrnjacka Banja, Serbia – Budva, Montenegro, August 28 – September 5, 2016, [Электронный ресурс] <http://ceur-ws.org/Vol-1839/MIT2016-p03.pdf> (дата обращения: 04.11.2017).
19. Nikolaychuk O.A., Berman A.F., Pavlov A.I. Predicting the Technical State of Hazardous Objects via Simulation Modeling // Journal of Machinery Manufacture and Reliability. – 2017. – V. 46, No. 2. – pp. 209–218. © Allerton Press, Inc., 2017.
20. Берман А.Ф., Грищенко М.А., Дородных Н.О., Николайчук О.А., Юрин А.Ю. A model-driven approach and a tool to support creation of rule-based expert systems for industrial safety expertise // Proceedings IFKAD2017 (12th International Forum on Knowledge Asset Dynamics), St. Petersburg, Russia 7-9 June 2017. pp.2034-2050.
21. Михайлов А. А., Хмельнов А. Е. Анализ программного кода в объектных файлах Delphi, скомпилированных под платформу .NET // Труды конференции Языки программирования и компиляторы — 2017.— С. 202 – 204.

22. Бычков И. В., Ружников Г. М., Парамонов В. В., Шумилов А. С., Фёдоров Р. К., Будням С. Инфраструктурный подход обработки пространственных данных в задачах управления территорией // Сборник трудов всероссийской конференции "Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов" — Новосибирск: 2017, ИВТ СО РАН, — С. 7 – 9.
23. Гаченко А. С., Хмельнов А. Е., Осипчук Е. Н., Абасов Н. В. Технология моделирования зон затопления а нижнем бьефе Иркутской ГЭС при высоких расходах через её гидростворы // Сборник трудов всероссийской конференции "Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов" — Новосибирск: 2017, ИВТ СО РАН, — С. 252 – 256.
24. Фёдоров Р. К., Шумилов А. С., Ружников Г. М. Облако геопорталов // Сборник трудов всероссийской конференции "Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов" — Новосибирск: , 2017, ИВТ СО РАН — С. 305 – 308.
25. Шигаров А.О., Бычков И.В., Ружников Г.М., Парамонов В.В., Михайлов А.А. TABBYXL: система трансформации данных из произвольных электронных таблиц в реляционную форму / XVI Российская конференция «Распределенные информационно - вычислительные ресурсы. Наука – цифровой экономике» (DICR-2017): Труды XVI Всероссийской конференции (4-7 декабря 2017 г.). Новосибирск: ИВТ СО РАН. 2017. – С.150-156. - ISBN: 978-5-905569-10-4
26. Бычков И.В., Ружников Г.М., Фёдоров Р.К., Шигаров А.О., Парамонов В.В. Инфраструктура цифровой экономики Иркутской области / XVI Российская конференция «Распределенные информационно - вычислительные ресурсы. Наука – цифровой экономике» (DICR-2017): Труды XVI Всероссийской конференции (4-7 декабря 2017 г.). Новосибирск: ИВТ СО РАН. 2017 – С.38-48. - ISBN: 978-5-905569-10-4.
27. Будням С., Батбилэг С., Баянжаргал Д., Ружников Г.М., Федоров Р.К. Математическая модель производственного потенциала и оценки технологической эффективности регионального производства // Сборник докладов 10-ой научной конференции Финансово-Экономического Университета Монголии. Улан-Батор. – 2017.– С. 94 – 100.
28. Бычков И.В., Хмельнов А.Е., Ружников Г.М., Фёдоров Р.К., Маджара Т.И., Верховина А.В., Сороковой А.А. Информационно-аналитическое и телекоммуникационное обеспечение междисциплинарных научных исследований территориального развития // Материалы XII международной конференции «Окружающая среда и устойчивое

- развитие монгольского плато и сопредельных территорий», Улан-Удэ, БНЦ СО РАН. – 2017. – С. 230 – 233.
29. Бычков И. В., Хмельнов А. Е., Гаченко А. С., Никитин В. М., Абасов Н. В., Осипчук Е. Н. Подход к 3D моделированию русла реки Ангара // Материалы XII международной научной конференции "Окружающая среда и устойчивое развитие Монгольского плато и сопредельных территорий" — Улан-Удэ: БНЦ СО РАН, 2017. — С. 233 – 235.
30. Парамонов В. В. Поиск вычислимых значений в табличных документах // XVIII Всероссийская конференция молодых учёных по математическому моделированию и информационным технологиям — Новосибирск: ИВТ СО РАН, 2017. — С. 85 – 86.
31. Фёдоров Р. К., Бычков И. В., Ружников Г. М., Авраменко Ю. В. Реализация функций принадлежности в логическом методе идентификации объектов // САИТ-2017 Труды конференции Седьмая международная конференция "Системный анализ и информационные технологии" — Калининград: , 2017. — С. 592 – 596.
32. Шумилов А. С., Авраменко Ю. В. Сервис идентификации объектов на спутниковых снимках // XVIII Всероссийская конференция молодых учёных по математическому моделированию и информационным технологиям — Новосибирск: ИВТ СО РАН, 2017. — С. 64 – 65.
33. Черкашин Е. А., Шигаров А. О., Михайлов И. С., Орлова И. В. Использование технологий Linked Open Data при подготовке и публикации текстовых документов // Материалы Всероссийской конференции с международным участием "Знания-Онтологии-Теории" (ЗОНТ-2017) — Новосибирск: , 2017. — С. 138 – 147.
34. Фёдоров Р. К., Бычков И. В., Шумилов А. С., Ружников Г. М., Верхозина А. В. Облачная среда для научных исследований // Материалы Международной научно-практической конференции "Использование современных информационных технологий в ботанических исследованиях" — Апатиты: , 2017. — С. 123 – 126.
35. Фёдоров Р. К., Верхозина А. В. Развитие информационно-аналитической системы по фиторазнообразию Байкальской Сибири // Материалы Международной научно-практической конференции "Использование современных информационных технологий в ботанических исследованиях" — Апатиты: , 2017. — С. 26 – 28.
36. Фёдоров Р. К., Ветров А. А. Сервис многомерного анализа данных с использованием подготовленных таблиц измерений // Материалы XVIII Всероссийской конференции молодых учёных по математическому моделированию — Новосибирск: ИВТ СО РАН, 2017. — С. 68 – 69.

37. Шигаров А. О., Парамонов В. В. Идентификация вычисляемых значений в слабоструктурированных табличных документах // Материалы международной научной конференции "Информационные технологии и системы-2017" (ИТС-2017) Международная научная конференция "Информационные технологии и системы-2017" — Минск: БГУИР, 2017. — С. 256 – 257.
38. Черкашин Е. А., Шигаров А. О., Михайлов И. С., Орлова И. В. Использование технологий Linked Open Data при подготовке и публикации текстовых документов // Материалы Всероссийской конференции с международным участием "Знания-Онтологии-Теории" (ЗОНТ-2017) — Новосибирск: , 2017. — С. 138 – 147.
39. Фёдоров Р. К., Бычков И. В., Гаченко А. С., Ружников Г. М., Хмельнов А. Е. ГИС «Адресный план» г. Иркутска // Материалы IV научной конференции «Региональные проблемы дистанционного зондирования земли». — 2017. — С. 70 – 73
40. Cherkashin E. A., Shigarov A., Morozov A., Pascal K., Malkov F. AnEnvironment for Metagenomic Analysis // Procs. of International Conference on Applied Internet and Information Technologies (ICAИТ) 2017 — Zrenjanin, Serbia: 2017. — С. 110 – 117.
41. Cherkashin E. A., Shigarov A., Mikhailov I., Orlova I. Authoring and Publishing Text Documents by means of Linked Open Data Technologies // Procs. of International Conference on Applied Internet and Information Technologies (ICAИТ) 2017 — Zrenjanin, Serbia: 2017. — С. 98 – 109.
42. Fedorov R., Voskoboinikov M. L. Method of Collecting, Processing and Storing Geolocational Data Received From Mobile Devices // Procs. of International Conference on Applied Internet and Information Technologies (ICAИТ) 2017 — Zrenjanin, Serbia: 2017. — С. 45 – 52.
43. Шевченко Б.С., Черкашин Е.А. Математическое обеспечение рекомендательных информационных систем и его реализации // Молодежный вестник ИРГТУ. 2017. № 2(26). <http://mvestnik.istu.irk.ru/journals/2017/02/articles/32>
44. Берман А.Ф., Николайчук О.А., Павлов А.И. Метод приобретения мультидисциплинарных знаний на основе онтологии // Седьмая международная конференция «Системный анализ и информационные технологии», САИТ – 2017 (13-18 июня 2017 г., г. Светлогорск, Россия): Труды конференции. В 2-х т.. – М : ИСА РАН, 2017. – Т1. – С. 295–302.
45. Берман А.Ф., Николайчук О.А., Павлов А.И. Самоорганизующийся алгоритм формирования решений для обеспечения требуемого технического состояния сложных опасных объектов // Седьмая международная конференция «Системный анализ и информационные технологии», САИТ – 2017 (13-18 июня 2017 г., г.

- Светлогорск, Россия): Труды конференции. В 2-х т.. – М : ИСА РАН, 2017. – Т1. – С. 377-384.
46. Берман А.Ф., Николайчук О.А., Тюленев Р.П. Трансдисциплинарная модель задач для обеспечения безопасности технических объектов // Международная научно-практическая конференция «Логистика и экономика ресурсосбережения и энергосбережения в промышленности» (МНПК «ЛЭРЭП-11-2017»). Россия, г. Тула. 15-17 ноября 2017 г.
47. Dorodnykh N.O. Web-based software for automating development of knowledge bases on the basis of transformation of conceptual models// Открытые семантические технологии проектирования интеллектуальных систем. 2017, № 7. С. 145-150.
48. Дородных Н.О. Программная система автоматизации разработки web-сервисов для генерации баз знаний // Труды конференции «Системный анализ и информационные технологии. САИТ-2017», САИТ – 2017 (13-18 июня 2017 г., г. Светлогорск, Россия): Труды конференции. В 2-х т.. – М : ИСА РАН, 2017. – Т1. – С. 222-229.
49. Дородных Н.О., Малтугуева Г.С., Николайчук О.А. Метод представления и обработки экспертных знаний, извлеченных из концептуальных моделей // Седьмая международная конференция «Системный анализ и информационные технологии», САИТ – 2017 (13-18 июня 2017 г., г. Светлогорск, Россия): Труды конференции. В 2-х т.. – М : ИСА РАН, 2017. – Т1. – С. 363-368.
50. Dorodnykh N.O., Yurin A.Yu. About the specialization of model-driven approach for creation of case-based intelligence decision support systems // Открытые семантические технологии проектирования интеллектуальных систем. 2017, № 7. С. 151-154.
51. Дородных Н.О., Юрин А.Ю. Автоматизированное создание продукционных баз знаний на основе коцепт-карт // Труды конференции «Системный анализ и информационные технологии. САИТ-2017», САИТ – 2017 (13-18 июня 2017 г., г. Светлогорск, Россия): Труды конференции. В 2-х т.. – М : ИСА РАН, 2017. – Т1. – С.337-341.
52. Дородных Н.О., Юрин А.Ю. Использование онтологий при разработке продукционных экспертных систем // Материалы Всероссийской конференции с международным участием «Знания-Онтологии-Теория» (ЗОНТ-2017), г. Новосибирск, 2-6 октября 2017 г. – Новосибирск, Институт математики им. С.Л. Соболева СО РАН – Т.1. – С.129-138.
53. Малтугуева Г.С., Юрин А.Ю. Многометодная процедура принятия управленческих решений // Анализ, моделирование, управление, развитие социально-экономических систем: сборник научных трудов XI Международной школы-симпозиума АМУР-

- 2017, Симферополь-Судак, 14-27 сентября 2017 / Под общей редакцией А.В. Сигала. – Симферополь : ИП Корниенко А.А., 2017. – С.254-258.
54. Малтугуева Г.С., Юрин А.Ю., Дородных Н.О. Система поддержки принятия решений «Выбор» // Труды Шестой Международной науч. конф. Информационные технологии и системы (Банное, Россия, 1–5 марта 2017 г. (ИТиС — 2017): науч. электрон. изд. (1 файл 28,0 Мб). Челябинск: Изд-во Челяб. гос. ун-та, 2017. С. 163 – 166.
55. Махутов Н.А., Берман А.Ф., Николайчук О.А. Алгоритм самоорганизующегося процесса обоснования свойств технического состояния сложных опасных объектов // Материалы десятой международной конференции «Управление развитием крупномасштабных систем» (MLSD'2017), 2-4 октября 2017 г., Москва, том 1, с 98-101.
56. Николайчук О.А., Павлов А.И. Web-ориентированная программная система визуального проектирования продукционных баз знаний на основе онтологии // Материалы всероссийской конференции с международным участием «Знания-Онтологии-Теории (ЗОНТ-17)» (2-6 октября 2017 г., Новосибирск) в 2-х т.: Труды конференции. – Новосибирск: ИМ СО РАН, 2017. –Т.2. – С. 105-110.
57. Николайчук О.А., Павлов А.И., Столбов А.Б. Особенности разработки агентных имитационных моделей на основе модельно-управляемого подхода // Восьмая всероссийская научно-практическая конференция по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование. Теория и практика» (ИММОД-2017) (18-20 октября 2017 г., С-Петербург). – С-Петербург, 2017. – С. 288-293.
58. Павлов А.И., Столбов А.Б. Система разработки агентных имитационных моделей // Международная научно-практическая конференция «Логистика и экономика ресурсосбережения и энергосбережения в промышленности» (МНПК «ЛЭРЭП-11-2017»). Россия, г. Тула. – С.290-292. 15-17 ноября 2017 г.
59. Авраменко Ю. В. Логические описания деформируемых моделей // Материалы XIII Всероссийской конференции молодых ученых "Моделирование, оптимизация и информационные технологии" памяти профессора В.И. Гурмана — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 4.
60. Новицкий Ю. А., Гаченко А. С., Хмельнов А. Е. ГИС «Адресный план г. Иркутска» в задачах развития муниципалитета // Тезисы докладов XIII Всероссийской конференции молодых ученых «Моделирование, оптимизация и информационные технологии» XIII Всероссийская конференция молодых ученых «Моделирование,

оптимизация и информационные технологии» памяти профессора В.И. Гурмана — Иркутск: , 2017. — С. 20.

61. Воронин В.И., Леви К.Г., Ружников Г.М. Потенциал древесно-кольцевых хронологий для палеореконструкций в Сибири / Фундаментальные проблемы экологии России / Тезисы докладов Всероссийской научной конференции, г. Иркутск – пос. Листвянка (оз. Байкал). – Иркутск: Изд-во Института географии им. В. Б. Сочавы СО РАН, 2017. – С. 51
62. Ружников Г.М., Федоров Р.К., Шигаров А.О., Парамонов В.В., Хмельнов А.Е. Перспективы развития инфраструктуры цифровой экономики Иркутской области // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 43-44.
63. Хмельнов А.Е. Алгоритмы улучшения цифровых моделей рельефа, представленных триангуляциями // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 59.
64. Гаченко А.С., Хмельнов А.Е. Механизм построения 3-D моделей рельефа водных объектов // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 11.
65. Алтаев А.А., Михайлов А.А., Шигаров А.О. TABBYPDF: ВЕБ-сервис извлечения таблиц из PDF документов // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 3.
66. Гаченко А.С., Хмельнов А.Е., Фереферов Е.С., Федоров Р.К. WEB-система мониторинга и оценки антропогенного воздействия на озеро Байкал // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 12.
67. Берман А.Ф., Николайчук О.А. Трансдисциплинарный подход для решения комплексных мультидисциплинарных проблем// Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 7.
68. Малков Ф.С., Черкашин Е.А., Шигаров А.О. Перспективы применения системы управления рабочим процессом TAVERNA в задачах обработки метагеномных данных при исследовании микробиома озера Байкал // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 32.
69. Малтугуева Г.С. Критерии оценки методов решения задач принятия решений // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 12.

70. Малтугуева Г.С., Николайчук О.А., Метод коллективной обработки нечёткой и неполной информации, извлечённой из концептуальных моделей // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 34.
71. Николайчук О.А., Павлов А.И., Столбов А.Б. Шаблоны информационных процессов интеллектуальных мультиагентных систем // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 38.
72. Николайчук О.А., Павлов А.И., Столбов А.Б. Об одном подходе к разработке агентных имитационных моделей на основе модельно-управляемого подхода // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 39.
73. Черкашин Е.А., Лобачёва И.Ю. Математическое моделирование пространственного взаимодействия природных явлений на глобальном уровне // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 61.
74. Ветров А.А., Фереферов Е.С. Развитие информационной системы учета научной деятельности // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 9.
75. Малков Ф.С., Черкашин Е.А., Шигаров А.О. Перспективы применения системы управления рабочим процессом TAVERNA в задачах обработки метагеномных данных при исследовании микробиома озера Байкал // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 32.
76. Федоров Р.К., Шумилов А.С. Организация работы с устройствами и датчиками с помощью системы выполнения композиций сервисов // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 54.
77. Воскобойников М.Л., Федоров Р.К. Разработка методов анализа применения пользователем WEB-сервисов // Материалы конференции «Ляпуновские чтения» — 2017 — Иркутск: РИО ИДСТУ СО РАН, 2017. — С. 10.
78. Дородных Н.О., Малтугуева Г.С., Юрин А.Ю. Автоматизация подбора методов многокритериального принятия решений // Материалы XVIII Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям, г. Иркутск, Россия, 21-25 августа 2017 г. — Новосибирск ИВТ СО РАН, 2017. — С.82. (РИНЦ)
79. Дородных Н.О., Николайчук О.А., Юрин А.Ю. Создание продукционных баз знаний на основе диаграмм Исикавы // Материалы XVIII Всероссийской конференции молодых ученых по математическому моделированию и информационным

технологиям, (г. Иркутск, Россия, 21-25 августа 2017 г.) – Новосибирск: ИВТ СО РАН, 2017. – С.74.

80. Малтугуева Г.С. Метод обработки индивидуальных предпочтений с помощью мультимножеств // Тезисы докладов XIII Всероссийской конференции молодых ученых «Моделирование, оптимизация и информационные технологии» (Иркутск – Старая Ангасолка (оз. Байкал, 13-18 марта 2017 г.). – Иркутск: ИДСТУ СО РАН, 2017. С. 45.
81. Малтугуева Г.С., Николайчук О.А. Нечеткие мультимножества и прецедентный подход // Тезисы докладов XIII Всероссийской конференции молодых ученых «Моделирование, оптимизация и информационные технологии» (Иркутск – Старая Ангасолка (оз. Байкал, 13-18 марта 2017 г.). – Иркутск: ИДСТУ СО РАН, 2017. С. 46.
82. Дородных Н.О., Юрин А.Ю. Создание прецедентных экспертных систем на основе трансформации моделей // Тезисы докладов XIII Всероссийской конференции молодых ученых «Моделирование, оптимизация и информационные технологии» (Иркутск – Старая Ангасолка (оз.Байкал), 13 - 18 марта 2017 г. – Иркутск: ИДСТУ СО РАН, 2017. – С.31.

Свидетельства о государственной регистрации программ

1. Фёдоров Р. К., Бычков И. В., Шумилов А. С., Ружников Г. М. Среда выполнения сервисов и их сценариев Свидетельство о государственной регистрации программ для ЭВМ № 2017617913 М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.

2. Фёдоров Р. К., Парамонов В. В., Ружников Г. М., Данчинова Г. А., Хасантинов М. А., Ляпунов А. В. WEB-сервис импорта данных из реляционных таблиц в CSV формате Свидетельство о государственной регистрации программ для ЭВМ № 2017615230 М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.

3. Данчинова Г. А., Ляпунов А. В., Хасантинов М. А., Манзарова Э. Л., Болотова Н. А., Соловаров И. С., Петрова И. В., Шулепова С. Ю., Лазарева Е. Л., Фёдоров Р. К., Парамонов В. В. Информационно-справочная система «Обращаемость людей, пострадавших от присасываний иксодовых клещей на территории Бурятии» (ИСС «Бурятия-клещи») Свидетельство о государственной регистрации программ для ЭВМ № 2017620505 М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.

4. Хмельницкий И., Олещенко И.Г., Хмельнов А.Е. Способ прогнозирования уровня седации во время анестезиологического пособия. Свидетельство о государственной регистрации № 2017127808, М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.

5. Дородных Н.О. Web-ориентированный редактор моделей трансформаций (Web Transformation Model Editor): Свидетельство о государственной регистрации программ для ЭВМ № 2017618430 от 01.08.2017. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.

6. Дородных Н.О. RVML editor (Web Knowledge Base Designer): Свидетельство о государственной регистрации программ для ЭВМ № 2017618446 от 01.08.2017. М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2017.