

ОГЛАВЛЕНИЕ

Часть 1. Тематический выпуск

«ЦИФРОВЫЕ ТЕХНОЛОГИИ БУДУЩЕГО – СОВРЕМЕННЫЕ РЕШЕНИЯ В НАУКАХ О ЗЕМЛЕ. ITES-2025»

В. В. Наумова

VII ВСЕРОССИЙСКАЯ КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

ЦИФРОВЫЕ ТЕХНОЛОГИИ БУДУЩЕГО – СОВРЕМЕННЫЕ РЕШЕНИЯ

В НАУКАХ О ЗЕМЛЕ. ITES-2025

721–726

Л. П. Брагинская, А. П. Григорюк, В. В. Ковалевский,

А. А. Добрынина, М. С. Ким

ВАРИАЦИИ СПЕКТРАЛЬНОГО СОСТАВА МИКРОСЕЙСМ

КАК ПРОГНОЗНЫЙ ПАРАМЕТР ЗЕМЛЕТРЯСЕНИЙ

В БАЙКАЛЬСКОЙ РИФТОВОЙ СИСТЕМЕ

727–739

И. В. Бычков, Е. А. Черкашин, Цзинь Чжан, Т. Ю. Черкашина,

В. А. Попова, О. А. Мазаева, О. В. Лунина

ТЕХНОЛОГИИ СЕМАНТИЧЕСКОГО ВЕБА ДЛЯ ПОДДЕРЖКИ

ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

740–780

В. С. Ерёменко, В. В. Наумова

ЦИФРОВОЙ ПОМОЩНИК ГЕОЛОГА-ИССЛЕДОВАТЕЛЯ

781–787

В. В. Наумова, М. И. Патук, А. С. Еременко, А. А. Загуменнов,

В. С. Ерёменко

ЦИФРОВЫЕ ТЕХНОЛОГИИ БУДУЩЕГО

ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

788–805

М. И. Патук, В. В. Наумова

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТЕКСТОВ ГЕОЛОГИЧЕСКИХ ПУБЛИКАЦИЙ

С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

806–821

Л. Е. Чесалов

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ГЕОЛОГИИ: СОСТОЯНИЕ И ПЕРСПЕКТИВЫ

822–828

Часть 2. Оригинальные статьи

- М. В. Бобырь, Н. А. Милостная, Б. А. Бондаренко, М. М. Бобырь**
ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ПОРОГОВОГО МЕТОДА HSV
И НЕЙРОННОЙ СЕТИ U-NET В ЗАДАЧЕ РАСПОЗНАВАНИЯ ПОЖАРОВ
С ПОМОЩЬЮ НЕЙРО-НЕЧЕТКОГО ПОДХОДА 829–851
- Я. А. Вахтерова, Д. А. Леонтьева**
ОБРАТНАЯ ЗАДАЧА ИДЕНТИФИКАЦИИ
ТЕРМОФИЗИЧЕСКИХ ПАРАМЕТРОВ МОДЕЛИ ГРИНА – НАГДИ III ТИПА
ДЛЯ УПРУГОГО СТЕРЖНЯ НА ОСНОВЕ
ФИЗИЧЕСКИ ИНФОРМИРОВАННОЙ НЕЙРОННОЙ СЕТИ 852–869
- Э. А. Заарур, С. В. Симушкин**
ЭМПИРИЧЕСКИЕ АНАЛОГИ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ
С ГАРАНТИРОВАННЫМ ВЫВОДОМ 870–883
- Р. Н. Минниханов, Т. Р. Баторшин, Р. М. Габбазов, Р. И. Фахразиев,**
А. С. Катасёв, М. В. Дагаева, И. Р. Бадрутдинов
ЦИФРОВОЙ ДВОЙНИК ПАРКОВОЧНОГО ПРОСТРАНСТВА 884–902
- В. Р. Раҳманкулова, В. В. Кугуракова**
ОНЛАЙН-ИНСТРУМЕНТ TULA ДЛЯ БАЛАНСИРОВКИ ВИДЕОИГР 903–930
- Х. Салем, А. С. Тощев**
АЛГОРИТМИЧЕСКИЙ ФРЕЙМВОРК ДЛЯ ИЗВЛЕЧЕНИЯ
ИНФОРМАЦИОННОГО ЯДРА ВЕБ-СТРАНИЦЫ 931–942
- А. Е. Семичастнов, Д. А. Балакин**
МОДЕЛИРОВАНИЕ СЦЕНАРИЕВ РАБОТЫ
РАДИОЛОКАЦИОННОЙ СИСТЕМЫ ДЛЯ КЛАССИФИКАЦИИ
БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ И ПТИЦ
НА ОСНОВЕ МИКРОДОПЛЕРОВСКИХ СИГНАТУР В СРЕДЕ ENGEE 943–952

В. Т. Трофимчук, В. В. Кугуракова

**А РАЗРАБОТКА АДАПТИВНОЙ СИСТЕМЫ ГЕНЕРАЦИИ
ИГРОВЫХ КВЕСТОВ И ДИАЛОГОВ
НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ**

953–993

УДК 004.85

VII ВСЕРОССИЙСКАЯ КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ «ЦИФРОВЫЕ ТЕХНОЛОГИИ БУДУЩЕГО – СОВРЕМЕННЫЕ РЕШЕНИЯ В НАУКАХ О ЗЕМЛЕ. ITES- 2025»

В. В. Наумова [0000-0002-3001-1638]

*Государственный геологический музей им. В. И. Вернадского РАН, г. Москва,
Россия
naumova_new@mail.ru*

Аннотация

Представлена краткая информация о VII Всероссийской конференции с международным участием «Цифровые технологии будущего – современные решения в науках о Земле. ITES-2025» которая прошла 22–26 сентября 2025 г. во Владивостоке.

Ключевые слова: конференция ITES-2025, цифровые технологии будущего, науки о Земле.

22–26 сентября 2025 г. в Дальневосточном федеральном университете (ДВФУ) (г. Владивосток) прошла VII Всероссийская конференция с международным участием «Цифровые технологии будущего – современные решения в науках о Земле. ITES-2025».

Организаторы конференции: Государственный геологический музей им. В. И. Вернадского РАН (г. Москва) и Институт математики и компьютерных технологий ДВФУ (г. Владивосток).

Ранее конференции ITES проводились в городах Магадане (2008), Владивостоке (2010), Петропавловске-Камчатском (2014), Южно-Сахалинске (2016), Москве (2019), Владивостоке (2022). Интеграция академической и вузовской науки Москвы и Владивостока стала определяющим фактором при выборе региона для проведения конференции 2025.

Организаторы надеются, что эти конференции служат стимулом для развития информационных технологий в научных исследованиях и образовании на Дальнем Востоке РФ и способствуют привлечению молодежи в науку.

Организаторы предполагают, что конференция 2025 г. будет также способствовать повышению эффективности использования результатов научной деятельности органами исполнительной власти и организациями, связанными с контролем и управлением природными ресурсами, предупреждением и ликвидацией последствий чрезвычайных природных и антропогенных ситуаций, проектно-изыскательскими работами.

Цели и задачи Конференции: представление ведущими специалистами докладов по актуальным направлениям исследований в области цифровых технологий будущего для наук о Земле: больших данных, облачных технологий, искусственного интеллекта, цифровых двойников, беспилотных летательных аппаратов и др. Важными также являются осуждения вопросов цифровой популяризации научной деятельности, в также вопросов вовлечения в науку молодежи и студентов и организация взаимодействия между учеными, представляющими различные области знаний.

Направления работы Конференции

В ходе конференции были рассмотрены современные достижения, полученные в рамках четырех главных направлений исследований, в том числе и их практические приложения.

Первое направление объединяет разработки, связанные с особенностями данных (в том числе Big Data) в науках о Земле: новые концепции и методы, инструменты их сбора, интеграции и обработки в различных информационных системах, в том числе и в системах с интенсивным использованием данных.

Второе направление связано с разработкой и адаптацией методов и технологий искусственного интеллекта для решения научных, в частности, геологических и геофизических задач.

Третье – с цифровыми инструментами анализа и обработки информации и виртуальными ассистентами.

Четвертое – с популяризацией научных знаний.

В конференции приняли участие очно ученые из трех стран мира (РФ, Белоруссия, ЮАР). РФ была представлена 20 городами. Были представлены научные материалы исследователей из 23 институтов РАН и 12 российских университетов. Кроме того, приняли участие в обсуждениях важных тем конференции 3 федеральные организации (МЧС, СБЕР, Министерство природных ресурсов РФ) и 6 крупных горнорудных компаний и ассоциаций РФ (Алроса, Нордголд, Полиметалл и др.).

Основные темы конференции:

- унифицированные платформы геологических данных;
- облачные технологии для интеграции данных, их обработки и анализа, визуализации в режиме совместной работы территориально распределенных геологов-пользователей;
- разработка и адаптация методов и технологий искусственного интеллекта для решения геологических задач. Технологии искусственного интеллекта и обработки естественного языка для научных исследований в геологии. Опыт применения технологий машинного обучения в задачах прогнозирования и диагностирования геологических процессов;
- дистанционное зондирование Земли из космоса. Беспилотные летательные аппараты в геологоразведке: создание карт местности, ортофотопланов, трехмерных моделей рельефа и объектов на нем, а также обнаружение различных аномалий геологического характера и др.;
- цифровые двойники геологических объектов и процессов;
- современные цифровые системы популяризации геологических знаний.

На пленарных сессиях были представлены доклады ведущих специалистов, профессоров и приглашенных докладчиков. Доклад академика РАН И. В. Бычкова (Институт динамики систем и теории управления СО РАН)

был посвящен технологиям семантического веба для поддержки фундаментальных исследований в геологии; доклад Л. М. Чесалова (ФГБУ «Гидроспецгеология») – цифровым технологиям в геологии и недропользовании: оценке их состояния и перспективам развития; доклад В. В. Наумовой (Государственный геологический музей им. В. И. Вернадского РАН) – цифровым технологиям будущего и их современным решениям в науках о Земле; доклад А. С. Еременко (Государственный геологический музей им. В.И. Вернадского РАН) – Метавселенной «История Земли» в качестве инструмента революции в образовании и др.

В ходе работы конференции на трех секциях были озвучены и обсуждены более 50 устных докладов по темам:

- Цифровые технологии в геологии, геофизике, геохимии;
- Цифровые технологии в океанологии, физике атмосферы и географии;
- Современные цифровые инструменты сбора, интеграции и обработки данных для наук о Земле;
- Современные цифровые системы популяризации наук о Земле.

Помимо применения традиционных методов сбора, анализа и обработки данных с использованием информационных технологий в области наук о Земле ряд докладов был посвящен применению технологий будущего в области наук о Земле: искусственного интеллекта (разные направления), цифровых двойников, БЛА, виртуальной и дополненной реальности, беспроводных технологий и др.

Краткие тезисы докладов опубликованы в Материалах конференции [1]. Презентации докладов размещены на сайте конференции [2].

Конференция организована и проведена в рамках Государственного задания ГГМ РАН по теме № 1021061009468-8-1.5.1 «Цифровая платформа интеграции и анализа геологических и музеиных данных» и при финансовой поддержке Гранта Государственной Программы стратегического академического лидерства «Приоритет-2030».

СПИСОК ЛИТЕРАТУРЫ

1. Цифровые технологии будущего — современные решения в науках о Земле ITES-2025: Материалы VII Всероссийской конференции с международным участием, Владивосток, 22–26 сентября 2025 г. / сост. А.А. Стрельцова. Владивосток: Издательство Дальневосточного федерального университета, 2025. 82 с.

2. Презентации докладов ITES-2025.

<https://ites2025.sgm.ru/images/data/ITES-2025-materialy.pdf>

VII ALL-RUSSIAN CONFERENCE WITH INTERNATIONAL PARTICIPATION «DIGITAL TECHNOLOGIES OF THE FUTURE — MODERN SOLUTIONS IN EARTH SCIENCES. ITES-2025»

Vera V. Naumova^[0000-0002-3001-1638]

*Vernadsky State Geological Museum of the Russian Academy of Sciences,
Moscow, Russia*

naumova_new@mail.ru

Abstract

The publication provides brief information about the VII All-Russian Conference with international participation "VII All-Russian Conference with International Participation «Digital technologies of the future — modern solutions in earth sciences. ITES-2025»", which took place on September 22–26, 2025 in Vladivostok.

Keywords: conference *ITES-2025*, digital technologies of the future, earth sciences.

REFERENCES

1. Digital technologies of the future — modern solutions in Earth sciences ITES-2025: Proceedings of the VII All-Russian Conference with International Participation, Vladivostok, September 22–26, 2025 / comp. A.A. Streltsova. Vladivostok: Publishing House of the Far Eastern Federal University, 2025. 82 p.

2. Presentations of the ITES-2025 reports.

<https://ites2025.sgm.ru/images/data/ITES-2025-materialy.pdf>

СВЕДЕНИЯ ОБ АВТОРЕ



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Vera V. NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: naumova_new@mail.ru

ORCID: 0000-0002-3001-1638

Материал поступил в редакцию 25 октября 2025 года

УДК 004.5+004.415

ВАРИАЦИИ СПЕКТРАЛЬНОГО СОСТАВА МИКРОСЕЙСМ КАК ПРОГНОЗНЫЙ ПАРАМЕТР ЗЕМЛЕТРЯСЕНИЙ В БАЙКАЛЬСКОЙ РИФТОВОЙ СИСТЕМЕ

Л. П. Брагинская¹ [0000-0002-3469-3373], А. П. Григорюк² [0000-0003-0388-2259],
В. В. Ковалевский³ [0000-0001-7143-3354], А. А. Добрынина⁴ [0000-0002-2087-4202],
М. С. Ким⁵ [0009-0001-7339-291X]

^{1, 2, 3, 5}Институт вычислительной математики и математической геофизики
СО РАН, г. Новосибирск, Россия

⁴Институт земной коры СО РАН, г. Иркутск, Россия

⁵Новосибирский государственный университет, г. Новосибирск, Россия

¹ludmila@opg.sscc.ru, ²and@opg.sscc.ru, ³kovalevsky@sscc.ru,

⁴dobrynina@crust.irk.ru, ⁵matvey3693@gmail.com

Аннотация

Исследован спектральный состав микросейсмического шума за несколько часов до умеренных и сильных сейсмических событий. Рассмотрены 40 землетрясений с энергетическим классом $K = 9.5\text{--}14.5$ на эпицентральных расстояниях от 10 до 120 км. Установлено статистически значимое повышение спектральной плотности мощности (СПМ) в диапазоне 0.8–2.4 Гц. Методами машинного обучения построена модель бинарной классификации, позволяющая по значениям СПМ микросейсм в указанном диапазоне частот определить процессы подготовки землетрясений за несколько часов до толчка.

Ключевые слова: геофизический мониторинг, машинное обучение, цифровая платформа, предвестники, сейсмический прогноз, землетрясения.

ВВЕДЕНИЕ

Проблема краткосрочного прогноза землетрясений до сих пор остается в области дискуссий и неопределенности. В настоящее время приоритетным является статистический подход к прогнозированию землетрясений. Однако в последнее десятилетие появилось значительное количество публикаций, в которых сообщается о необычных вариациях и аномальном поведении параметров различных геофизических полей в областях подготовки сильных землетрясений [1-3], а также о разработке различных алгоритмов, включая технологии когнитивного распознавания предвестников [4, 5].

Анализ данных, полученных сейсмологическими методами мониторинга, включает измерение спектрального состава колебаний, типичность или аномальность первых вступлений поперечных и продольных волн, выявление тенденции к группированию (это называют роем сейсмических событий), оценку вероятности активизации тех или иных тектонически активных структур и др.

Мониторинг микросейсмического режима вызывает особый интерес, поскольку микросейсмические колебания, возникающие в результате внутренних процессов, содержат информацию о деформациях в земной коре. Эти процессы могут варьироваться от движения тектонических плит и крупных землетрясений до лунно-солнечных приливных деформаций и микроземлетрясений.

Примером признака подготовки землетрясений являются обнаруженные закономерности в вариациях микросейсмического шума на примере близких умеренных и сильных сейсмических событий в Байкальской рифтовой системе (БРС) [6, 7].

Целью настоящей работы было развитие методики обнаружения признаков подготовки близких умеренных и сильных землетрясений по данным мониторинга микросейсмического шума БРС.

ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ ДАННЫХ

Институтом земной коры СО РАН создана pilotная сеть комплексного геофизического мониторинга, состоящая из полигонов «Приольхонье», «Бугульдейка», «Листвянка» и «Зун-Мурино». Основу наблюдательной сети в пределах каждого полигона составляет пункт комплексного мониторинга [8]. Данные

со всех пунктов мониторинга поступают на цифровую платформу (ЦП) комплексного мониторинга, разработанную в Институте вычислительной математики и математической геофизики (ИВМиМГ) СО РАН [9, 10]. ЦП позволяет анализировать и визуализировать на одной временной сетке данные любой станции и любого из методов, предоставляя исследователям доступ к широкому набору вычислительных, картографических сервисов, данным каталогов землетрясений, подробной информации о пунктах мониторинга и активных разломах (см. пример на рис. 1).

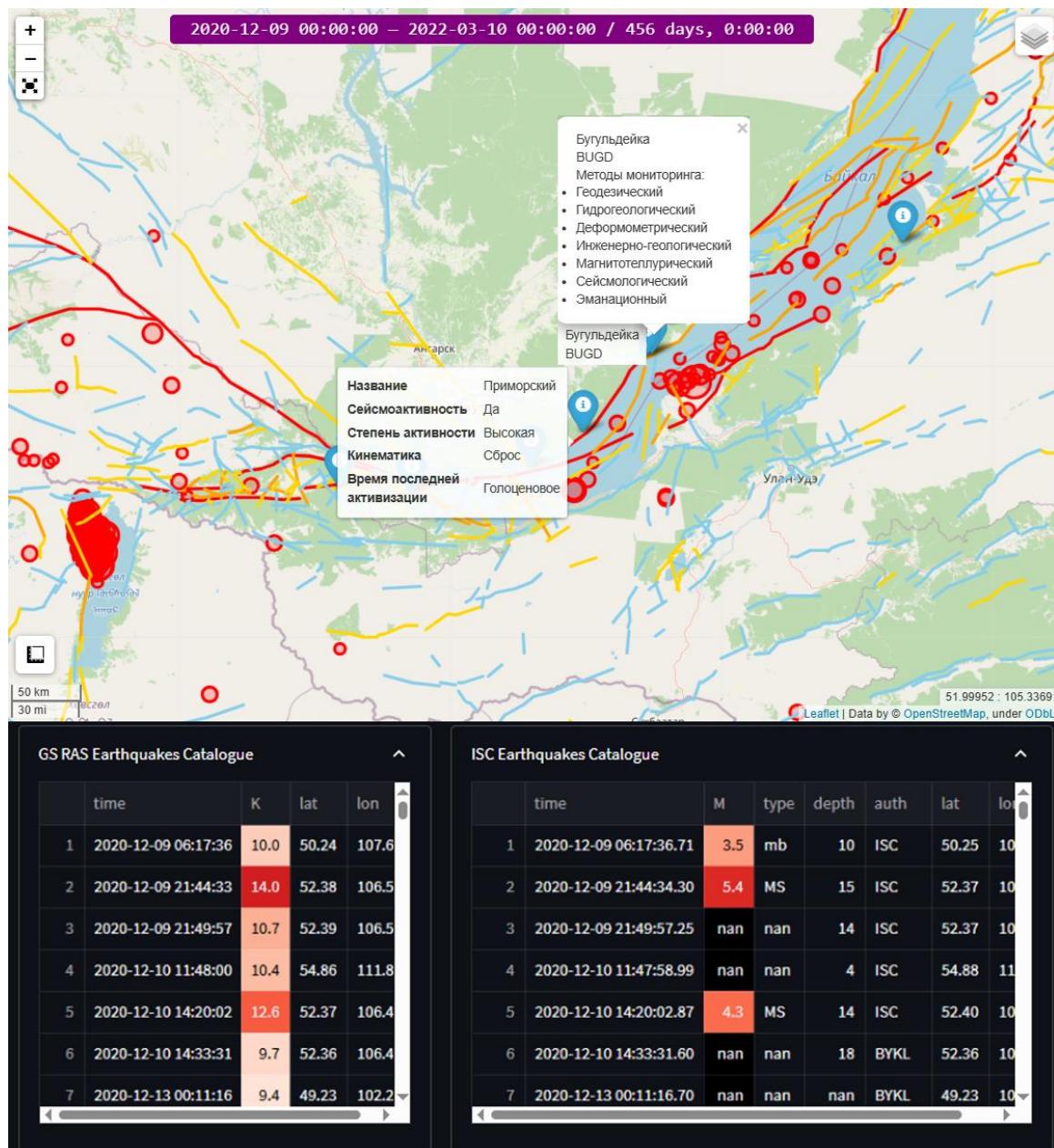


Рис. 1. Пример страницы ЦП

В настоящей работе анализировались данные микросейсмического мониторинга, зарегистрированные на пункте комплексного мониторинга опасных геологических ситуаций «Бугульдейка», расположенным вблизи зоны влияния крупнейших разломов центральной части БРС – Приморского и Морского. Этот пункт оснащен современным геофизическим оборудованием, в том числе широкополосной сейсмической станцией. Параметры аппаратуры позволяют не только регистрировать локальные, региональные и удаленные землетрясения, но и фиксировать и анализировать вариации микросейсмического поля.

Для поиска краткосрочных предвестников землетрясений проводилась оценка вариаций спектральной плотности мощности (СПМ) участков записи микросейсмических полей, зарегистрированных накануне землетрясения и в спокойные периоды.

Для анализа были выбраны 40 землетрясений различного энергетического диапазона ($K = 9.5\text{--}14.9$, магнитуда $M = 3.1\text{--}5.9$) за период 2021–2023 гг., которые в основном локализованы в зоне влияния Обручевского, Морского и Приморского разломов. Расстояние от эпицентров землетрясений до пункта регистрации составляло от 10 до 120 км. Выбранная область характеризуется как область чистого растяжения и охватывает сейсмические события, происходящие в зонах влияния этих разломов, предоставляя полный набор данных для анализа закономерностей и поиска предвестников будущих землетрясений, которые могут повлиять на регион.

СПМ количественно описывает распределение энергии по частотам и активно применяется при анализе землетрясений благодаря своей способности выявлять ключевые характеристики сейсмических процессов.

Для оценки СПМ использовался метод периодограмм Уэлча на 10-минутных участках записи микросейсмического шума (МСШ). Параметры периодограмм: длительность сегментов 10 сек с перекрытием 50%, значения СПМ медианные. Это позволило получить оценки СПМ с шагом 0.1 Гц. Анализировались записи МСШ за 250, 200, 150, 100 и 50 мин до землетрясения. Полученные оценки СПМ сравнивались с таковыми, полученными в спокойные периоды,

не менее чем через 12 ч после землетрясения. Таким образом, для каждого землетрясения анализировались десять 10-минутных участков записей, а всего было получено 400 образцов СПМ.

Усредненные спектры МСШ для спокойных периодов времени и активированных периодов (за несколько часов до землетрясений) приведены на диаграмме рис. 2а.

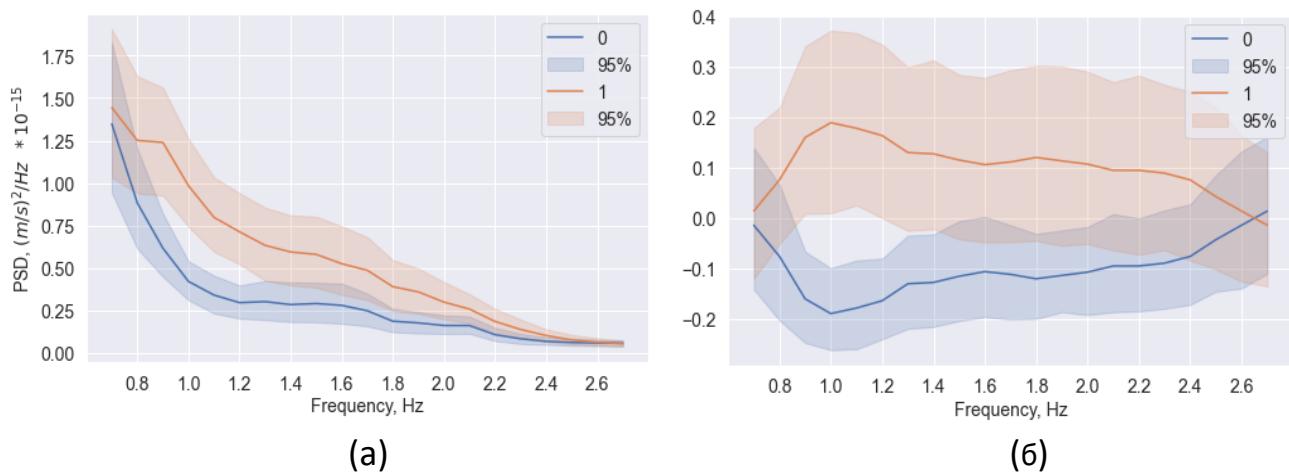


Рис. 2. Зависимости усредненных значений СПМ от частоты:

(а) исходные зависимости, (б) нормализованные зависимости.

Обозначения: «0» (синим цветом) – спокойные периоды времени, «1» (красным цветом) – активированные периоды времени.

Для большей наглядности на рис. 2б представлены нормализованные спектры. Нормализация осуществлялась путем удаления среднего значения и масштабирования до единичного стандартного отклонения. На диаграммах также показаны доверительные интервалы 95%. Как можно видеть из диаграмм, в диапазоне 0.8–2.4 Гц наблюдается статистически значимое повышение СПМ для активированных периодов времени, при этом также увеличивается стандартное отклонение. Эти образцы использовались для обучения модели бинарной классификации записей МСШ.

Первоначально анализ и сравнение различных участков записей МСШ выполнялись с помощь ЦП в интерактивном режиме, позднее были написаны программы на языке Python для автоматического отбора и анализа данных. Построение интерактивных карт с эпицентрами землетрясений обеспечивает входящая в состав платформы ГИС.

ПОСТРОЕНИЕ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

Для моделирования связи между вариациями СПМ и признаком активации участка записи микросейсм использовались алгоритмы бинарной классификации трех моделей контролируемого машинного обучения – моделей KNN, Gradient Busting, Random Forest.

В машинном обучении использовался датасет, состоящий из 400 образцов СПМ, который был разделен на 300 образцов для обучения 100 для тестирования. В качестве независимых переменных (признаков) для обучения и тестирования моделей брались значения СПМ в диапазоне 0.8–2.4 Гц с шагом 0.1 Гц, всего 17 значений.

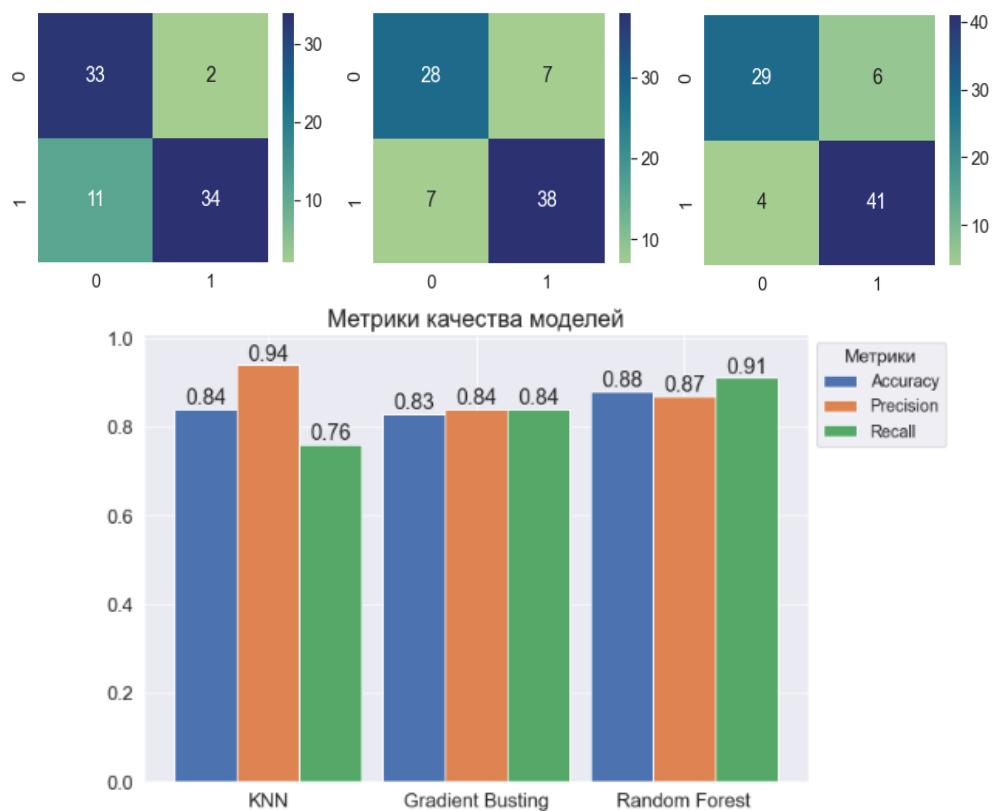


Рис. 3. Матрицы ошибок и метрики качества моделей KNN, Gradient Busting, Random Forest

Метрика *Accuracy* равна доле правильно классифицированных образцов. Значение полноты (*Recall*) отражает долю предсказанных положительных результатов среди фактических положительных результатов. Более высокие значения

полноты (*Recall*) указывают на меньшее количество пропущенных прогнозов. Значение точности (*Precision*) отражает долю фактических положительных результатов среди предсказанных положительных результатов. Более высокие значения точности (*Precision*) указывают на меньшее количество неверных прогнозов.

Методом кросс-валидации установлено, что наилучшие результаты показала модель «Случайный лес» (Random Forest) из библиотеки Sklearn. «Случайный лес» — это ансамблевый метод обучения, объединяющий несколько деревьев решений для создания более точной и надежной модели.

Основная идея модели «Случайный лес» заключается в создании нескольких деревьев решений, каждое из которых обучено на подмножестве данных и случайном подмножестве признаков [11]. Каждое дерево формирует прогноз, а окончательный прогноз представляет собой среднее значение (для регрессии) или моду (для классификации) прогнозов отдельных деревьев. Создавая множество деревьев и вычисляя их среднее значение, «Случайный лес» может снизить влияние переобучения и повысить точность и устойчивость модели. Наилучшее найденное разделение на классы может варьироваться даже при одинаковых обучающих данных. Поиск наилучшего качества модели проводится путем подбора гиперпараметров модели, максимизирующего ее метрики на этапе тестирования. На рис. 4 представлены основные метрики и матрица ошибок после подбора гиперпараметров для разработанной модели машинного обучения.

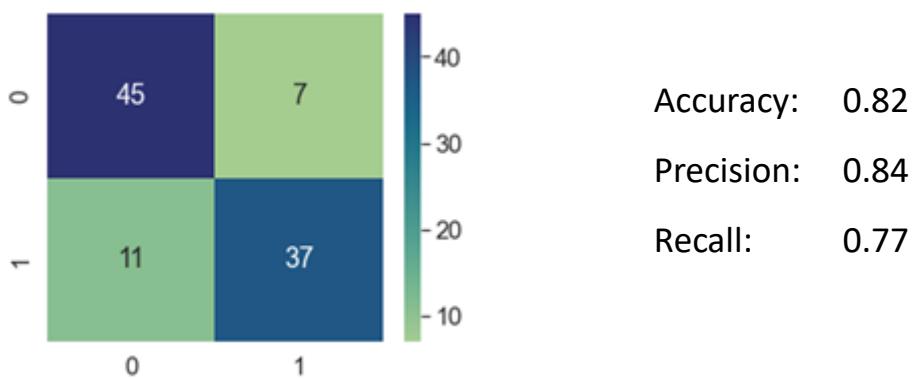


Рис. 4. Матрица ошибок и метрики для модели «Случайный лес»

Тестирование производилось на 100 образцах МСШ, не участвовавших в обучении.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований получила развитие методика обнаружения признаков подготовки землетрясений на примере близких умеренных и сильных сейсмических событий по данным о вариациях спектрального состава МСШ. На примере 40 землетрясений БРС с энергетическим классом К = 9.5–14.5 на эпицентralьных расстояниях от 10 до 120 км установлено статистически значимое повышение СПМ микросейсмического шума в диапазоне 0.8–2.4 Гц за несколько часов перед толчком. На основе полученных данных методом машинного обучения создана модель для бинарной классификации записей МСШ. Зафиксированное повышение спектральной плотности мощности МСШ может классифицироваться как краткосрочный предвестник, связанный с консолидацией среды за 3–4 ч до землетрясения.

Благодарности

Работы выполнены в рамках Госзадания ИВМиМГ СО РАН, тема FWNM-2025-0004, и гранта Министерства науки и высшего образования РФ № 075-15-2024-533.

СПИСОК ЛИТЕРАТУРЫ

1. *Pulinets S., Herrera V.M.V. Earthquake Precursors: The Physics, Identification, and Application // Geosciences.* 2024. Vol. 14 (209), P. 2–33.
<https://doi.org/10.3390/geosciences14080209>
2. *Bogdanov V., Gavrilov V., Pulinets S., Ouzounov D. Responses to the preparation of strong Kamchatka earthquakes in the lithosphere–atmosphere–ionosphere system, based on new data from integrated ground and ionospheric monitoring // E3S Web Conf.* 2020. Vol. 196 (03005) P. 1–14.
<https://doi.org/10.1051/e3sconf/202019603005>
3. *Saltykov V.A. On the Possibility of Using the Tidal Modulation of Seismic Waves for Forecasting Earthquakes // Izvestiya, Physics of the Solid Earth.* 2017. Vol. 53 (2), P. 250–261. <https://doi.org/10.1134/S1069351317010128>
4. *Li J., Zhai H., Jiang C. et al. Application of artificial intelligence technology in the study of anthropogenic earthquakes // Artif Intell.* 2025. Vol. 58 (155).
<https://doi.org/10.1007/s10462-025-11157-2>

5. Kubo H., Nao M., and Kano M. Recent advances in earthquake seismology using machine learning // Earth Planets. 2024. Vol 76 (36).
<https://doi.org/10.1186/s40623-024-01982-0>
 6. Korol S.A., Sankov A.V., Dobrynina A.A., Sankov V.A. Ambient Seismic Noise Variations before Earthquakes in the Baikal Rift System // Geodynamics & Tectonophysics. 2022. Vol. 13 (2):0632 (In Russ.).
<https://doi.org/10.5800/GT-2022-13-2s-0632>
 7. Sobolev G.A., Lyubushin A.A., Zakrzhevskaya N.A. Asymmetrical Pulses, the Periodicity and Synchronization of Low Frequency Microseisms // Journal of Volcanology and Seismology. 2008. Vol. 2, No. 2. P. 118–134.
<https://doi.org/10.1134/S074204630802005X>.
 8. Семинский К.Ж., Добрынина А.А., Борняков С.А., Саньков В.А., Постпееев А.В., Рассказов С.В., Перевалова Н.П., Семинский И.К., Лухнев А.В., Бобров А.А., Чебыкин Е.П., Едемский И.К., Ильясова А.М., Салко Д.В., Саньков А.В., Король С.А. Комплексный мониторинг опасных геологических процессов в Прибайкалье: организация пилотной сети и первые результаты // Геодинамика и тектонофизика. 2022. Т. 13(5), 0677. <https://doi.org/10.5800/GT-2022-13-5-0677>
 9. Григорюк А.П., Брагинская Л.П., Семинский И.К., Семинский К.Ж., Ковалевский В.В. Цифровая платформа для интеграции и анализа данных геофизического мониторинга байкальской природной территории // Электронные библиотеки. 2022. Т. 25. № 4. С. 303–316.
<https://doi.org/10.26907/1562-5419-2022-25-4-303-316>
 10. Брагинская Л.П., Григорюк А.П., Ковалевский В.В., Добрынина А.А. Цифровая платформа для комплексных геофизических исследований в Байкальском регионе // Сейсмические приборы. 2023. Т. 59, № 4. С. 36–49.
<https://doi.org/10.21455/si2023.4-3>
 11. Earthquake-prediction-using-Machine-learning-models // A project done for the course CSE3505 — Essentials of Data Analytics under ELANGO N M. URL: <https://github.com/akash-r34/Earthquake-prediction-using-Machine-learning-models?tab=readme-ov-file#earthquake-prediction-using-machine-learning-models>
-

VARIATIONS IN MICROSEISMIC NOISE SPECTRA AS A FORECAST PARAMETER OF EARTHQUAKES IN THE BAIKAL RIFT SYSTEM

Ludmila P. Braginskaya¹ [0000-0002-3469-3373], Andrey P. Grigoryuk² [0000-0003-0388-2259], Valery V. Kovalevsky³ [0000-0001-7143-3354], Anna A. Dobrynina⁴ [0000-0002-2087-4202] Matvey S. Kim⁵ [0009-0001-7339-291X]

^{1, 2, 3, 5}*Institute of Computational Mathematics and Mathematical Geophysics
Siberian Branch of the Russian Academy of Sciences, Novosibirsk*

⁴*Institute of the Earth's Crust Siberian Branch of the Russian Academy of Sciences, Irkutsk*

⁵*Novosibirsk State University, Novosibirsk*

¹ludmila@opg.sscc.ru, ²and@opg.sscc.ru, ³kovalevsky@sscc.ru,
⁴dobrynina@crust.irk.ru, ⁵matvey3693@gmail.com

Abstract

This paper examines the microseismic noise spectra a few hours before moderate and strong seismic events. Forty earthquakes with an energy class of K=9.5–14.5 at epicentral distances of 10 to 120 km were considered. A statistically significant increase in the spectral power density (SPD) was detected in the 0.8–2.4 Hz range. Machine learning methods were used to construct a binary classification model that allows detection of earthquake preparations a few hours before an event based on microseismic SPD values in the specified frequency range.

Keywords: geophysical monitoring, machine learning, digital platform, precursors, seismic forecast, earthquakes.

REFERENCES

1. Pulinets S., Herrera V.M.V. Earthquake Precursors: The Physics, Identification, and Application // Geosciences. 2024. Vol. 14 (209), P. 2–33.
<https://doi.org/10.3390/geosciences14080209>
2. Bogdanov V., Gavrilov V., Pulinets S., Ouzounov D. Responses to the preparation of strong Kamchatka earthquakes in the lithosphere–atmosphere–ionosphere system, based on new data from integrated ground and ionospheric monitoring // E3S Web Conf. 2020. Vol. 196 (03005) P. 1–14.

<https://doi.org/10.1051/e3sconf/202019603005>

3. Saltykov V.A. On the Possibility of Using the Tidal Modulation of Seismic Waves for Forecasting Earthquakes // Izvestiya, Physics of the Solid Earth. 2017. Vol. 53 (2), P. 250–261. <https://doi.org/10.1134/S1069351317010128>

4. Li J., Zhai H., Jiang C. et al. Application of artificial intelligence technology in the study of anthropogenic earthquakes // Artif Intell. 2025. Vol. 58 (155). <https://doi.org/10.1007/s10462-025-11157-2>

5. Kubo H., Nao M., Kano M. Recent advances in earthquake seismology using machine learning // Earth Planets. 2024. Vol. 76 (36).

<https://doi.org/10.1186/s40623-024-01982-0>

6. Korol S.A., Sankov A.V., Dobrynina A.A., Sankov V.A. Ambient Seismic Noise Variations before Earthquakes in the Baikal Rift System // Geodynamics & Tectonophysics. 2022. Vol. 13 (2), 0632 (In Russ.).

<https://doi.org/10.5800/GT-2022-13-2s-0632>

7. Sobolev G.A., Lyubushin A.A., Zakrzhevskaya N.A. Asymmetrical Pulses, the Periodicity and Synchronization of Low Frequency Microseisms // Journal of Volcanology and Seismology. 2008. Vol. 2, No. 2. P. 118–134.

<https://doi.org/10.1134/S074204630802005X>

8. Seminsky K.Zh., Dobrynina A.A., Bornyakov S.A., Sankov V.A., Pospeev A.V., Rasskazov S.V., Perevalova N.P., Seminskiy I.K., Lukhnev A.V., Bobrov A.A., Chebykin E.P., Edemskiy I.K., Ilyasova A.M., Salko D.V., Sankov A.V., Korol S.A. Integrated monitoring of hazardous geological processes in Pribaikalye: pilot network and first results // Geodynamics & Tectonophysics. 2022. Vol. 13 (5), 0677 (In Russ.). <https://doi.org/10.5800/GT-2022-13-5-0677>

9. Grigoryuk A.P., Braginskaya L.P., Seminsky I.K., Seminsky K.Zh., Kovalevsky V.V. A Digital Platform for Integration and Analysis of Geophysical Monitoring Data from the Baikal Natural Zone // Russian Digital Libraries Journal. 2022. Vol. 25, No. 4. P. 303–316. <https://doi.org/10.26907/1562-5419-2022-25-4-303-316>.

10. Braginskaya L., Grigoryuk A., Kovalevsky V., Dobrynina A. Digital platform for integrated geophysical investigations in the Baikal region // Seismic Instruments. 2023. Vol. 59 (4), P. 36–49. <https://doi.org/10.21455/si2023.4-3>

11. Earthquake-prediction-using-Machine-learning-models // A project done for the course CSE3505 — Essentials of Data Analytics under ELANGO N M.

URL: <https://github.com/akash-r34/Earthquake-prediction-using-Machine-learning-models?tab=readme-ov-file#earthquake-prediction-using-machine-learning-models>

СВЕДЕНИЯ ОБ АВТОРАХ



БРАГИНСКАЯ Людмила Петровна, ведущий программист Института вычислительной математики и математической геофизики СО РАН, г. Новосибирск.

Lyudmila BRAGINSKAYA, Lead Programmer at the Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk.

email: ludmila@opg.ssc.ru

ORCID: 0000-0002-3469-3373



ГРИГОРЮК Андрей Павлович, научный сотрудник Института вычислительной математики и математической геофизики СО РАН, г. Новосибирск.

Andrey GRIGORYUK, Researcher at the Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk.

email: and@opg.ssc.ru

ORCID: 0000-0003-0388-2259



КОВАЛЕВСКИЙ Валерий Викторович, доктор технических наук, заместитель директора по научной работе Института вычислительной математики и математической геофизики СО РАН, г. Новосибирск.

Valeriy KOVALEVSKIY, Doctor of Technical Sciences, Deputy Director for Science, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk.

email: kovalevsky@sscc.ru

ORCID: 0000-0001-7143-3354



Добрынина Анна Александровна, ведущий научный сотрудник Института земной коры СО РАН, г. Иркутск

Anna Dobrynina, lead researcher at the Institute of the Earth's Crust SB RAS, Irkutsk

email: dobrynina@crust.irk.ru

ORCID: 0000-0002-2087-4202



Ким Матвей Сергеевич, инженер Института вычислительной математики и математической геофизики СО РАН, г. Новосибирск.

Matvey KIM, Engineer at the Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk.

email: matvey3693@gmail.com

ORCID: 0009-0001-7339-291X

Материал поступил в редакцию 10 октября 2025 года

УДК 004.652.8+004.057.8+0004.75

ТЕХНОЛОГИИ СЕМАНТИЧЕСКОГО ВЕБА ДЛЯ ПОДДЕРЖКИ ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

И. В. Бычков¹ [0000-0002-1765-0769], Е. А. Черкашин² [0000-0003-2428-2471],
Цзинь Чжан³ [0000-0003-0009-824X], Т. Ю. Черкашина⁴ [0000-0002-1443-6839],
В. А. Попова⁵ [0000-0002-7764-1995], О. А. Мазаева⁶ [0000-0002-3053-5511],
О. В. Лунина⁷ [0000-0001-7743-8877]

^{1, 2, 5} Институт динамики систем и теории управления им. В. М. Матросова
Сибирского отделения РАН, г. Иркутск, Россия

^{1, 2, 4} Институт математики и информационных наук, Яньтайский
университет, г. Яньтай, Китай

³ Институт гражданского строительства, Яньтайский университет,
г. Яньтай, Китай

^{4, 6, 7} Институт земной коры Сибирского отделения РАН, г. Иркутск, Россия

^{1, 5} Институт математики и информатики, Иркутский государственный
университет, г. Иркутск, Россия

¹bychkov@icc.ru, ²eugeneai@icc.ru, ³zhangjin513@outlook.com,

⁴tcherk@crust.irk.ru, ⁵victorypopova1@gmail.com, ⁶moks@crust.irk.ru,

⁷lounina@crust.irk.ru

Аннотация

Представлена инновационная методология применения технологий семантического веба для поддержки фундаментальных геологических исследований. Рассмотрена проблема семантической интеграции разнородных геологических данных, характеризующихся масштабом разного уровня и междисциплинарностью. Разработана пятиэтапная методология, включающая анализ предметной области, онтологическое концептуальное моделирование, трансформацию данных в граф знаний, развертывание инфраструктуры распределенного доступа к данным на основе концептуальной модели, а также интеграцию с процедурами обработки и анализа. Практическая апробация проведена на трех кейсах: анализе геохимических данных для оценки уровня загрязнения территории, создании информационной системы о разломах и исследовании динамики береговой зоны

водохранилищ. Предложенный онтологический подход обеспечивает соответствие FAIR-принципам и преодоление «семантического барьера» в геологических исследованиях. Показано, что технологии семантического веба позволяют перейти от фрагментированных информационных массивов к целостному семантическому пространству геологических знаний, что открывает новые возможности для генерации комплексных научных гипотез и кросс-дисциплинарных исследований.

Ключевые слова: семантический веб, графы знаний, онтологическое моделирование, семантическое пространство научных знаний, фундаментальные геологические исследования.

ВВЕДЕНИЕ

Современная геология как фундаментальная наука переживает этап цифровой трансформации, характеризующийся большим ростом объемов и разнообразия данных. Этот процесс сопровождается фундаментальным парадоксом: несмотря на увеличение объема доступной информации, ее разрозненность и семантическая несовместимость существенно ограничивают возможности генерации новых комплексных гипотез для исследования. Существует проблема поиска и обработки полученных количественных данных в различных областях знаний. Ученый-исследователь вынужден тратить до 70–80% рабочего времени на поиск информации, приведение ее к единому формату и согласование данных, а не на их анализ и научную интерпретацию.

Проблема интеграции разнородных, разного масштаба и междисциплинарных данных является одной из ключевых в области наук о Земле. Ее актуальность и сложность подчеркиваются в фундаментальных работах, охватывающих последнее десятилетие. Так, еще в 2013 г. Крис Мэттманн в своей программной статье в *Nature* [1] обозначил необходимость создания автоматизированных методов для работы с тысячами форматов файлов и бесшовной интеграции алгоритмов. Развитие семантических технологий, подробно рассмотренное Паскалем Хитцлером [2, 3], предлагает мощный аппарат для решения этих задач через онтологии и графы знаний. Наконец, практическая реализация этих подходов на

национальном уровне продемонстрирована в таких проектах, как Геохронологический атлас-справочник (ГАС) ВСЕГЕИ [4], в котором решается задача в рамках централизованной системы с жесткой экспертной оценкой. В настоящей работе предложен подход, основанный на технологиях семантического веба (СВ), для решения обозначенных задач.

Технологии СВ предлагают принципиально новый подход к решению этой проблемы, позволяя преодолеть «семантический барьер» в геологических исследованиях. Особую актуальность данное направление приобретает в контексте глобальной научной инициативы FAIR-принципов [5] (Findable, Accessible, Interoperable, Reusable), направленной на обеспечение обнаружения, доступности, совместимости и многократного использования научных данных.

Фундаментальные геологические исследования характеризуются разнообразием используемых данных, таких как литостратиграфические колонки и палеонтологические определения; результаты петрографических и минералогических исследований; геохимические и изотопные анализы горных пород; геофизические данные различных методов (сейсмологии, гравиметрии, магнитотеллурического зондирования); структурные наблюдения и тектонические построения; данные дистанционного зондирования и цифровые модели рельефа; результаты экспериментальных исследований и математического моделирования.

Проблематика больших данных (Big Data) в геологии имеет выраженную предметную специфику. Наиболее ярко характеристики больших данных проявляются в области геохимии, где можно выделить следующие аспекты концепции Five-V [6, 7].

Variety (разнообразие) – геохимические данные существуют в принципиально различных форматах: табличные результаты физико-химических методов исследования, структурные данные, изотопные соотношения, пространственные распределения химических элементов, что создает значительную сложность в их интеграции и совместного анализа.

Value (ценность) - скрытые закономерности в геохимических данных имеют фундаментальное значение для понимания процессов петрогенезиса и рудообразования, геодинамических процессов, однако их изучение требует комплексного анализа разнородных данных.

Veracity (достоверность) – большое разнообразие аналитических методов, имеющих отличные друг от друга стандарты методических измерений, создают проблемы при интеграции полученных из разных лабораторий количественных данных, что требует учета их метрологических характеристик.

Velocity (скорость) – современные аналитические комплексы (ICP-MS, LA-ICP-MS, электронные сканирующие микроскопы) генерируют непрерывные потоки данных, требующих оперативной обработки и интерпретации прежде всего для публикации результатов в виде научных статей.

Volume (объем) – накопленные массивы аналитической информации достигают значительных объемов, создавая вызовы для хранения, индексации и эффективного поиска релевантных данных.

Актуальным становится требование сохранения научной автономии исследовательских коллективов в условиях необходимости интеграции данных. Каждое учреждение (геологический институт, университет, производственная организация) должно сохранять возможность управления собственными данными и знаниями, при этом обеспечивая их семантическую интероперабельность для совместных исследовательских проектов. Это требование приобретает особую значимость в контексте распределенного характера геологических исследований, когда данные собираются различными организациями в разных регионах, в разное время и для разных предметных областей.

Отметим, что существующие системы управления геологическими данными имеют фундаментальные ограничения, не позволяющие в полной мере решать перечисленные проблемы:

- жесткость схем реляционных баз данных не соответствует эволюционирующему характеру геологических знаний и постоянно уточняющимся классификациям;
- отсутствие семантической интеграции между различными геологическими дисциплинами приводит к созданию независимых, трудно интегрируемых «информационных силосов»;
- неспособность поддерживать сложные виды обработки информации для выявления скрытых взаимосвязей и генерации нового знания;

- ограниченность средств формализации сложных геологических концепций и теорий, имеющих множество исключений и частных случаев.

Для решения этих задач в настоящей работе предложено ориентироваться на технологический стек СВ, который используется при реализации трех ключевых направлений:

- 1) представление разнородных геологических данных в виде единого графа знаний [8] на основе разработанных онтологий, что обеспечивает их семантическую согласованность;
- 2) обеспечение федеративного доступа к распределенным графикам знаний через стандартизованные SPARQL-эндпоинты, позволяющие интегрировать данные без их централизации;
- 3) реализация целевых приложений, использующих семантически интегрированные данные для решения конкретных научных задач, относящихся к многомерному статистическому анализу геохимических данных (пример 1), визуализации структурно-тектонической информации (пример 2) и идентификации моделей динамики береговой зоны (пример 3).

Технологии СВ обеспечивают уникальные возможности, принципиально отличающиеся от традиционных подходов, а именно:

- онтологическое моделирование позволяет формализовать сложные геологические концепции (стратиграфические подразделения, тектонические элементы), создавая единую семантическую основу для интеграции знаний из различных дисциплин;
- гибкость RDF-формата обеспечивает адаптацию к развивающимся научным парадигмам и методикам исследований без необходимости перманентного перепроектирования всей системы хранения данных;
- гибкое управление данными в соответствии с принципами графов знаний [8] – возможность откладывать строгие формализации на более поздние этапы исследования. Эмерджентность знаний – это проявление новых закономерностей по мере интеграции данных из различных источников. Преобразование форматов данных как преобразование графов через SPARQL-запросы.

Технологии СВ совместимы не только с процедурами многомерного статистического анализа данных и большими языковыми моделями, но и с механизмами логического вывода, позволяющими не просто находить различного рода корреляции, а генерировать новое знание – автоматически идентифицировать геодинамическую обстановку формирования горной породы на основе комплекса геохимических и минералогических признаков, formalизованных в онтологии или базе знаний. Архитектура открытых связанных данных (Linked Data), специальный вариант СВ, поддерживает создание распределенной сети геологических знаний с сохранением автономии участников, что соответствует организационной структуре современной науки.

Целью настоящего исследования была разработка методологии применения технологий СВ для поддержки фундаментальных геологических исследований, обеспечивающей семантическую интеграцию распределенных геологических знаний и переход от фрагментированных информационных массивов к целостному семантическому пространству научных знаний.

В задачи исследования входило:

- 1) разработка онтологических моделей для представления фундаментальных геологических знаний, охватывающих как основные дисциплины, так и конкретные научные задачи, и их взаимосвязи;
- 2) создание механизмов семантической интеграции распределенных данных исследовательских организаций с сохранением их автономии;
- 3) разработка методов интеграции с механизмами логического вывода и системами, основанными на formalизованных знаниях, для решения конкретных научных задач в различных областях фундаментальной геологии;
- 4) апробация разработанного подхода на практических примерах (интерпретация геохимических данных, естественно-языковая система доступа к данным о геологических разломах, идентификация модели изменения береговой зоны водохранилищ);
- 5) оценка эффективности предложенных решений и их соответствия FAIR-принципам.

Ниже рассмотрены три примера решения задач с помощью предложенного онтологического подхода.

Пример 1: Анализ геохимических данных для оценки уровня загрязнения территории (геохимия, геоэкология)

Исследование эколого-геохимического состояния различных природных территорий требует интеграции данных об измеренных концентрациях химических элементов, результатах мониторинга загрязнений, различных нормативных показателей, а также информации о наличии природных или техногенных источников поступления загрязняющих компонентов. Отсутствие единой семантической модели затрудняет комплексную оценку экологического состояния территории и выявление возможных источников загрязнения, особенно когда данные поступают из различных ведомств и исследовательских групп.

Пример 2: Информационная система о геологических разломах (сейсмология, структурная геология)

Необходима интеграция разрозненных данных о разрывных нарушениях: кинематические характеристики, возраст образования, современная активность, связь с сейсмичностью, геофизические аномалии. Отсутствие единой семантической модели препятствует созданию целостной картины тектонического строения регионов и адекватной оценке сейсмической опасности.

Пример 3: Исследование береговой зоны (инженерная геология)

Требуется интеграция данных инженерно-геологических изысканий, гидрологических наблюдений, мониторинга береговых процессов. Разрозненность этих данных, хранящихся в различных форматах и у разных организаций, затрудняет создание комплексных моделей и прогнозирование развития береговой зоны, что критически важно для обеспечения безопасности прибрежной инфраструктуры.

Таким образом, предлагаемый онтологический подход и архитектура семантической интеграции позволяют преодолеть традиционные ограничения в работе с геологическими данными. Научная новизна исследования заключается в разработке комплексной онтологической модели, охватывающей междисциплинарные связи между геохимией, геоэкологией, инженерной геологией и тектоникой, что отсутствует в существующих отраслевых решениях. В отличие от подходов, ориентированных на отдельные типы данных, предлагаемая методология

обеспечивает сквозную семантическую интеграцию, позволяя выявлять кроссдисциплинарные закономерности.

Онтологический подход и графы знаний не просто решают проблему интеграции данных, но и становятся новым инструментом научного познания, позволяя формулировать и проверять гипотезы, которые были недоступны при работе с разрозненными информационными массивами. Это создает новые перспективы для фундаментальных открытий через выявление ранее скрытых взаимосвязей и закономерностей.

Структура статьи отражает логику исследования: последующие разделы представляют разработанную онтологическую модель, архитектуру системы, примеры реализации для различных геологических задач, оценку эффективности и анализ перспектив развития подхода.

МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ СЕМАНТИЧЕСКОГО ВЕБА

Разработанная методология представляет собой структурированный подход к семантической интеграции геологических данных, состоящий из пяти последовательных этапов, обеспечивающих переход от разрозненных информационных массивов к целостному семантическому пространству научных знаний.

Этап 1. Доменный анализ и идентификация семантических барьеров

Начальный этап предполагает комплексный анализ предметной области с выявлением ключевых семантических разрывов. Проводится инвентаризация источников данных, анализируются используемые классификации и таксономии, идентифицируются типичные сценарии исследовательских задач. Особое внимание уделяется выявлению неявных знаний и контекстных зависимостей, характерных для геологических дисциплин.

Этап 2. Онтологическое моделирование

На основании результатов доменного анализа разрабатывается модульная онтологическая модель, сочетающая общие геологические концепции со специализированными доменными расширениями. Модель строится по принципу се-

мантической совместимости с существующими стандартами и онтологиями. Формализуются классы, свойства, отношения и ограничения, обеспечивающие адекватное представление предметной области.

Этап 3. Трансформация данных в граф знаний

На данном этапе происходит преобразование разнородных геологических данных в RDF-формат с сохранением семантической целостности. Разрабатываются правила картирования для различных форматов исходных данных, обеспечивающие корректное отображение в термины онтологии. Реализуются механизмы контроля качества и верификации преобразованных данных.

Этап 4. Развёртывание семантической инфраструктуры

В рамках этого этапа создается архитектура доступа к семантически интегрированным данным, которая включает SPARQL-эндпоинты, специализированные API и пользовательские интерфейсы. Обеспечивается поддержка федеративных запросов к распределенным источникам данных. Разрабатываются механизмы кэширования и оптимизации производительности для работы с крупномасштабными графиками знаний.

Этап 5. Интеграция с исследовательскими процедурами

Завершающий этап направлен на внедрение семантических технологий в существующие исследовательские процедуры. Разрабатываются интерфейсы для интеграции с аналитическими инструментами, системами визуализации и средствами статистического анализа. Создаются механизмы логического вывода и семантического обогащения данных.

Сквозные принципы методологии:

- итеративный характер разработки с непрерывной валидацией. Ориентация на соответствие FAIR-принципам;
- поддержка эволюции онтологических моделей;
- обеспечение семантической интероперабельности;
- Сохранение автономии источников данных.

Данная методология обеспечивает системный подход к преодолению семантических барьеров в геологических исследованиях и созданию инфраструктуры для генерации нового знания на основе интегрированных разнородных данных.

Рассмотрим три примера решения геологических задач с использованием предложенного онтологического подхода.

ПРИМЕР 1: АНАЛИЗ ГЕОХИМИЧЕСКИХ ДАННЫХ ДЛЯ ОЦЕНКИ УРОВНЯ ЗАГРЯЗНЕНИЯ ПРИРОДНЫХ ТЕРРИТОРИЙ

Рассмотрим три примера решения геологических задач при помощи предложенного онтологического подхода.

Исследование эколого-геохимического состояния территорий требует комплексной интеграции данных об измеренных концентрациях химических элементов, методах анализа вещества, точек отбора образцов и их геологическом контексте [9–13]. Технологии СВ предоставляют принципиально новый подход к решению этой задачи через создание единого графа знаний на основе разработанной онтологии.

Разработанная онтологическая модель охватывает все аспекты геохимических исследований: от базовых классов образцов и измерений до сложных систем контроля качества и пространственного контекста. Ключевые классы включают pt:Sample, pt:GeochemicalSample, pt:Measurement со специализированными подклассами для элементных и суммарных измерений, что обеспечивает семантически богатое представление данных.

Приведенный ниже пример конкретного образца почвы в формате RDF/Turtle демонстрирует полноту представления информации:

```
pt:AGS-0085 a pt:GeochemicalSample ;
    rdfs:label "AGS-0085" ;
    wgs:lat 53.21610 ;
    wgs:long 107.4481 ;
    pt:sampleDepth 0.5 ;
    pt:sampleDate "2023-07-15"^^xsd:date ;
    pt:location pt:Ольхон_Хужир ;
    pt:formation pt:Палеозойские_граниты ;
```

```
pt:hasMeasurement [  
    a pt:ElementMeasurement ;  
    pt:measuredElement mt:As ;  
    pt:measurementValue 3.4 ;  
    pt:measurementUncertainty 0.2 ;  
    pt:detectionLimit 0.1 ;  
    pt:analyticalMethod "ICP-MS" ;  
    pt:qualityFlag pt:Good  
].
```

Для извлечения и анализа данных разработан специализированный модуль kgquery.ext, который предоставляет исследователям удобный интерфейс доступа к графу знаний. Модуль абстрагирует сложность SPARQL-запросов, позволяя формулировать запросы на концептуальном уровне, близком к предметной области геологии.

Типичный SPARQL-запрос для анализа распределения мышьяка в почвах на территории острова Ольхон демонстрирует высокую эффективность семантического подхода:

```
PREFIX pt: <http://example.org/ontology/project#>  
PREFIX mt: <http://example.org/ontology/mineral#>  
PREFIX wgs: <http://www.w3.org/2003/01/geo/wgs84_pos#>  
  
SELECT ?sample ?lat ?long ?value ?method ?quality ?formation  
WHERE {  
    ?sample a pt:GeochemicalSample ;  
    wgs:lat ?lat ;  
    wgs:long ?long ;  
    pt:location/pt:formation ?formation ;  
    pt:hasMeasurement ?measurement .  
  
    ?measurement pt:measuredElement mt:As ;  
        pt:measurementValue ?value ;  
        pt:analyticalMethod ?method ;  
        pt:qualityFlag ?quality .
```

```
FILTER(?value > 2.0 && ?quality IN (pt:Excellent, pt:Good))
}
ORDER BY DESC(?value)
```

Особенностью такого подхода является возможность семантического обогащения данных через механизмы логического вывода. Система автоматически классифицирует пробы как «аномальные» на основе формализованных правил, учитывающих абсолютные концентрации элементов и их соотношения, а также геологический контекст.

Полученные количественные данные обрабатываются с применением методов статистического анализа, включая метод главных компонент (PCA), для выявления скрытых закономерностей в распределении химических элементов и межэлементных взаимосвязей. Пространственная интерполяция на основе регрессионных моделей позволяет построить карты поверхностного распределения токсичных элементов, отражающие различные уровни загрязнений окружающей среды с учетом физических параметров территории.

Система контроля качества, реализованная в онтологии через классы pt:DataQuality и pt:QualityLevel, обеспечивает достоверность аналитических результатов. Это особенно важно при интеграции данных из различных лабораторий, использующих для измерения разные аналитические методы с отличными друг от друга метрологическими характеристиками.

Интеграция с внешними источниками данных, такими как цифровые модели рельефа SRTM, позволяет учитывать влияние геоморфологических факторов на распределение загрязняющих веществ. Автоматическое вычисление производных параметров (уклонов, экспозиций) расширяет аналитические возможности системы.

Предложенный подход успешно апробирован на примере количественной оценки концентраций мышьяка в почвах острова Ольхон, показав возможность выявления пространственных закономерностей и аномалий в распределении этого элемента. Семантическая интеграция данных позволила учсть не только химические концентрации, но и геоморфологические характеристики территории (см. рис. 1).

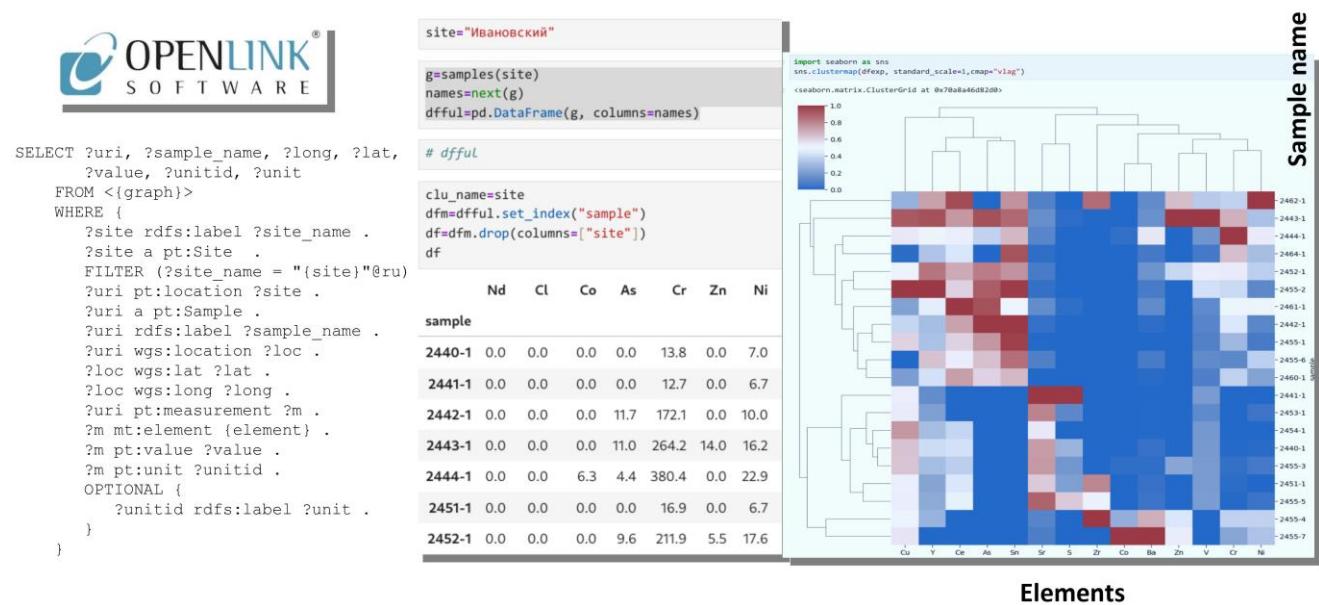


Рис. 1. Корреляционный анализ геохимических данных для почв острова Ольхон

Разработанный алгоритм демонстрирует переход от фрагментированных массивов геохимических данных к целостному семантическому пространству, где каждый новый факт автоматически обогащает общее знание и становится доступным для последующего анализа и открытия новых закономерностей в распределении загрязняющих веществ.

ПРИМЕР 2: ИНФОРМАЦИОННАЯ СИСТЕМА СЕМАНТИЧЕСКИХ ДАННЫХ О ГЕОЛОГИЧЕСКИХ РАЗЛОМАХ

Исходные данные о геологических разломах хранились в реляционной базе DBF-формата, содержащей 72 поля [14]. Эта структура включала идентификатор разлома, географическое название, различные характеристики с соответствующими уточнениями, уровень сейсмической активности, имя исследователя-компилятора и дату уточнения. Номера записей связывали записи базы данных с пространственными объектами слоя разломов. Важным ограничением формата DBF являлось то, что имена полей не могли быть длиннее десяти символов, должны были быть заглавными, а количество полей не могло превышать 255.

Идентификаторы атрибутов с общими префиксами определяли одно значение с уточнением. Например, атрибуты “Depth” и “Length” разлома, измеряемые в километрах, уточнялись атрибутом “quality” с суффиксом “_Q” в конце. Содер-

жение таблицы базы данных разломов было заполнено разреженно: многие атрибуты содержали нулевые значения. Поля “Geomorphol...” и некоторые другие заполнялись значениями из предопределенных наборов. Согласованность данных контролировалась алгоритмически с помощью модулей расширения QGIS. Эта плоская структура и формат представления были предназначены для легкого доступа и реализации с помощью стандартных инструментов реляционных баз данных, но общие нарушения стандартных нормальных форм вынуждали разработчика реализовывать подпрограммы контроля содержания записи в дополнение к команде UPDATE DML.

Процесс преобразования табличных данных в граф знаний начался с определения T-Box, включающего базовый класс geo:Fault и специализированные классы измерений: geo:Strike, geo:Slip, geo:Rate, geo:Feature, geo:Activity, geo:Event, которые являются подклассами geo:Measurement. Все записи о разломах последовательно конвертировали в RDF-триплеты с присвоением соответствующих классов, формируя таким образом A-Box (ящик экземпляров) – семантически структурированный граф знаний экземпляров разломов [15].

Универсальные свойства онтологии организованы вокруг объектного свойства geo:hasMeasurement, связывающего разломы с их измерениями, и набора дата-свойств для описания характеристик: geo:value (значение), geo:type (тип), geo:quality (качество), geo:unit (единица измерения) и geo:grade (балльная оценка). Дополнительные свойства geo:id, geo:name и geo:location описывают базовые атрибуты разломов, а свойства geo:compiledBy, geo:compilerName и geo:date фиксируют метаданные о компиляторах информации, обеспечивая полноту и прослеживаемость данных [15].

Для организации пространства имен использовали префикс geo: <<http://crust.irk.ru/ontologies/fault/terms/1.0/>>. Техническая реализация процесса осуществлена с помощью специализированной Python-программы, которая загружала исходные DBF-файлы и генерировала OWL2 XML-документы, после чего полученные онтологии T-Box и A-Box проходили визуальную верификацию в среде Protégé с последующей конвертацией в формат Turtle и загрузкой в семантические серверы GraphDB и Jena для последующего использования.

Ниже представлен пример преобразованного объекта – разлом «Главный Саянский» (RUAF_1), который демонстрирует насыщенность семантической модели:

```
<http://.../RUAF_1> a geo:Fault ;
    geo:id "RUAF_1" ;
    geo:name "Главный Саянский" ;
    geo:location "Вдоль юго-западной границы Сибирской платформы" ;

    # Strike measurements
    geo:hasMeasurement [
        a geo:Strike ;
        geo:value "301" ;
        geo:quality "ЛС"
    ] ;

    # .....
    # Activity
    geo:hasMeasurement [
        a geo:Activity ;
        geo:value "аномально высокая" ;
        geo:grade 32
    ], [
        a geo:Activity ;
        geo:type "LastActivationAge" ;
        geo:value "Голоценовое" ;
        geo:grade 2
    ] ;

    # Events
    geo:hasMeasurement [
        a geo:Event ;
        geo:type "PotentialMagnitude" ;
        geo:value "8" ;
        geo:quality "ЛС"
    ] ;
```

```
# Compiler
geo:compiledBy [
    a geo:Compiler ;
    geo:compilerName "Лунина О.В." ;
    geo:date "2011-02-23"^^xsd:date
].
```

АРХИТЕКТУРА ПРИЛОЖЕНИЯ WEB-GIS

Программа приложения Web-GIS реализована как одностраничное Web 2.0 приложение, построенное на основе React-компонентов с управлением состоянием через Redux. Данная архитектура обеспечивает эффективное управление интерфейсом и состоянием приложения, позволяя динамически обновлять содержимое виджетов при изменении набора визуализируемых геологических разломов.

Для визуализации картографических данных (рис. 2.) используется React-обертка библиотеки leaflet.js, которая предоставляет возможности отображения интерактивных графических объектов на топологической основе. Система поддерживает различные источники картографических подложек, включая OpenStreetMap в качестве базового варианта. Контуры разломов были преобразованы из формата KML в JSON-структуры, совместимые с leaflet, и могут быть загружены приложением по мере необходимости с веб-сервера. Важной особенностью этой системы является сопоставление контуров через International Resource Identifiers (IRI) разломов, что обеспечивает семантическую связь между графическими представлениями и данными в графе знаний.

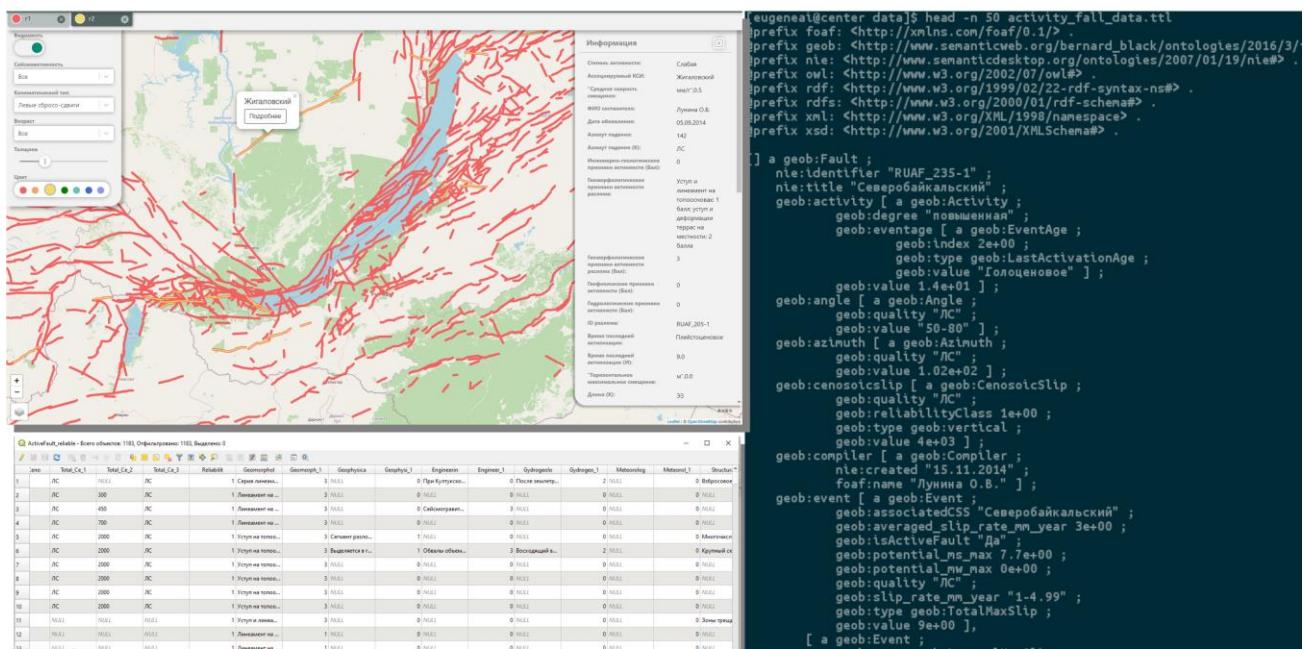


Рис. 2. Программа просмотра данных о геологических разломах

Доступ к семантическим данным организован через интерпретацию результатов выполнения SPARQL-запросов к конечной точке СВ-сервера. Пользовательский интерфейс включает форму для задания условий фильтрации, которая автоматически модифицирует стандартные SPARQL-запросы, добавляя соответствующие ограничения. Для целей отладки и разработки предусмотрена возможность выполнения свободных SPARQL-запросов с визуализацией результатов на карте. Отображение атрибутов разломов реализовано через формы с RDFa-разметкой, что обеспечивает семантически обогащенное представление данных.

Пример простого SPARQL-запроса для получения всех активных разломов, имеющих высокий потенциал вызвать землетрясение с магнитудой выше 7 баллов:

```
PREFIX geo: <http://crust.irk.ru/ontologies/fault/terms/1.0/>
SELECT ?fault ?name ?magnitude
WHERE {
  ?fault a geo:Fault ;
  geo:name ?name ;
  geo:hasMeasurement [
    a geo:Event ;
    geo:type "PotentialMagnitude" ;
```

```
geo:value ?magnitude
].
FILTER(xsd:decimal(?magnitude) > 7.0)
}
```

ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Система предоставляет естественно-языковой интерфейс доступа к данным графа знаний, автоматически преобразуя запросы на естественном языке в формализованные SPARQL-запросы. Это позволяет исследователям, не владеющим семантическими технологиями, формулировать сложные запросы, аналогичные приведенным в примере 1, а также запросы о разломах с голоценовой активизацией и идентификации левосдвиговых разломов длиной более 500 км с аномально высокой активностью.

По умолчанию система отображает все разломы из графа знаний, что соответствует результату SPARQL-запроса, выбирающего все объекты класса geo:Fault. Пользовательский интерфейс организован таким образом, чтобы обеспечить интуитивно понятное взаимодействие с данными. Он включает топооснову с наложенными контурами разломов, панель фильтрации для задания критериев отбора, всплывающие подсказки с названиями разломов при наведении курсора, а также детализированную форму отображения атрибутов выбранного разлома.

Система успешно преодолевает «семантический барьер» в работе с данными о разломах, обеспечивая гибкое представление неполных и эволюционирующих данных, соответствующих принципам Linked Open Data.

Ключевым достижением является создание естественно-языкового интерфейса, доступного для пользователей без специализированной подготовки в области семантических технологий. Система обеспечивает глубокую интеграцию пространственных и атрибутивных данных через семантические связи, что открывает новые возможности для комплексного анализа геологической информации.

Полученный опыт подтвердил практическую возможность построения Web-GIS-систем для научных сообществ, поддерживающих постоянное накопление, агрегацию и анализ данных благодаря фундаментальным свойствам графов зна-

ний. Такие системы способны адаптироваться к изменяющимся требованиям исследований и обеспечивать долгосрочную поддержку научной деятельности в области геологии и сейсмологии.

ПРИМЕР 3: СЕМАНТИЧЕСКАЯ ИНТЕГРАЦИЯ ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ БЕРЕГОВОЙ ЗОНЫ ВОДОХРАНИЛИЩ

Исследование динамики береговой зоны крупных равнинных водохранилищ характеризуется необходимостью интеграции разнородных данных инженерно-геологических изысканий, гидрологических наблюдений и мониторинга береговых процессов [16, 17]. Как отмечено выше, неупорядоченность этих данных затрудняет создание комплексных прогнозных моделей и оценку рисков для прибрежной инфраструктуры. Традиционные подходы, основанные на реляционных базах данных [18, 19] с жесткой схемой (в основном таблицы Excel), неспособны эффективно аккумулировать эволюционирующие классификации типов берегов и многообразие экзогенных геологических процессов, что ограничивает возможности комплексного анализа, мониторинга и прогнозирования.

Для решения обозначенных проблем была разработана онтологическая модель, представляющая береговую зону как единое семантическое пространство, в котором взаимосвязаны природные объекты, геологические процессы и условия их развития (см. рис. 3). Ключевой архитектурной идеей является процессоцентричная модель, где экзогенный процесс выступает активным агентом, меняющим состояние берега.

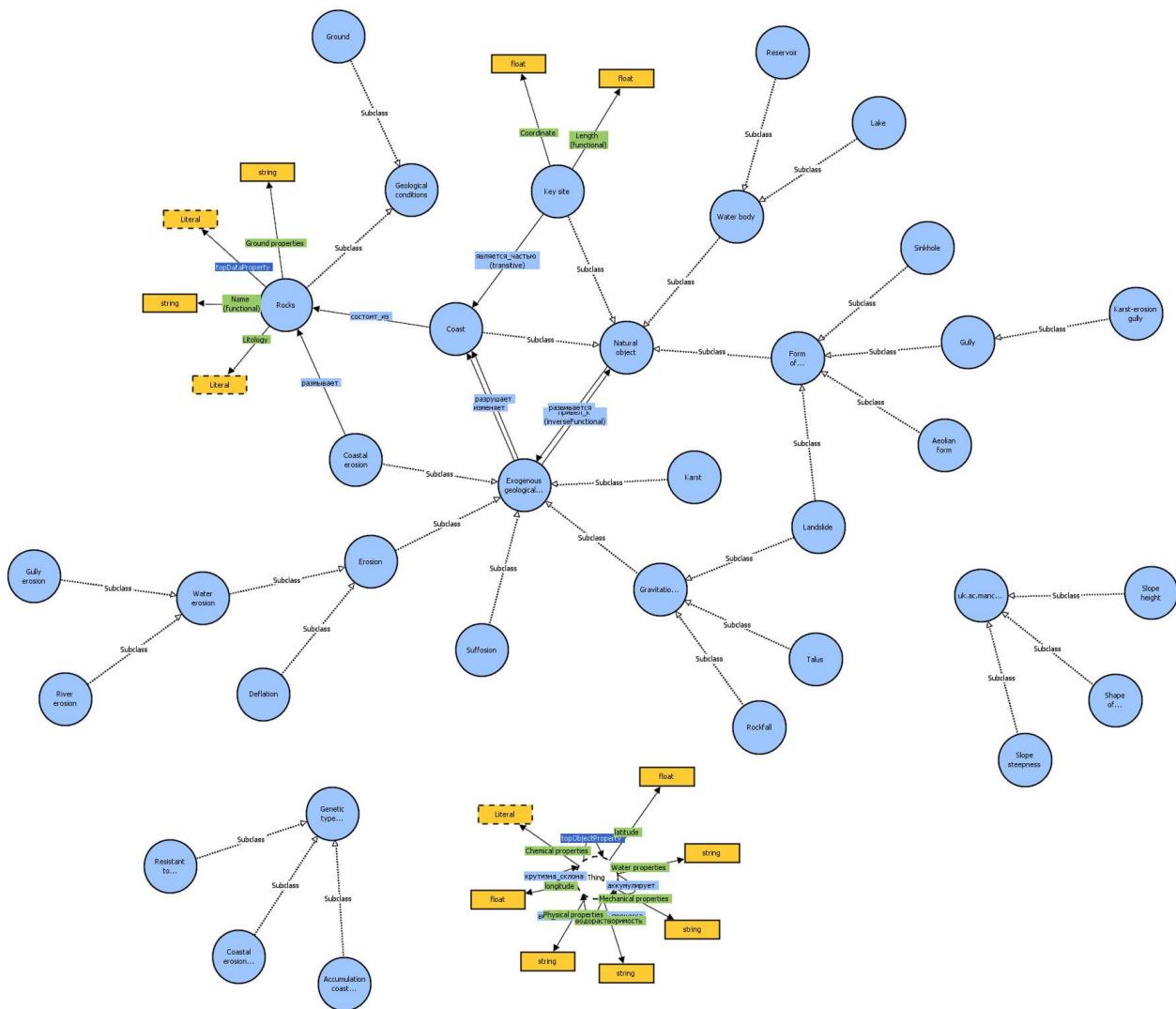


Рис. 3. Структура онтологии деградационных процессов береговой зоны водохранилищ

T-Box-онтологии (фрагмент). На концептуальном уровне онтология задает систему классов, свойств и их отношений, формализующих предметную область. Ниже представлен пример ключевого фрагмента T-Box.

```
# Ключевые классы (Class Hierarchy)
:Природный_объект rdf:type owl:Class .
:Берег rdfs:subClassOf :Природный_OBJECT .
:Водоем rdfs:subClassOf :Природный_объект .
```

:Водохранилище rdfs:subClassOf :Водоем .
:Участок rdfs:subClassOf :Природный_объект .

:Геологический_процесс rdf:type owl:Class .
:Абраузия rdfs:subClassOf :Геологический_процесс .
:Оползень rdfs:subClassOf :Геологический_процесс .
:Карстовый rdfs:subClassOf :Геологический_процесс .

:Тип_берега rdf:type owl:Class .
:Абразионный rdfs:subClassOf :Тип_берега .
:Аккумулятивный rdfs:subClassOf :Тип_берега .

Свойства отношений (Object Properties)
:изменяет rdf:type owl:ObjectProperty ;
 rdfs:domain :Геологический_процесс ;
 rdfs:range :Берег .

:привел_к rdf:type owl:ObjectProperty ;
 rdfs:domain :Геологический_процесс ;
 rdfs:range :Природный_объект .

:является_частью rdf:type owl:ObjectProperty, owl:TransitiveProperty ;
 rdfs:domain :Участок ;
 rdfs:range :Берег .

Предложенный подход приобретает важную практическую значимость в контексте решения задачи идентификации доминирующей модели динамики береговой зоны, эта задача требует комплексного анализа следующих факторов:

- литолого-петрографического состава пород берегового уступа;
- гидродинамического режима водоема;
- морфометрических характеристик берегового склона;
- наличия и степени активности экзогенных геологических процессов.

Для формализации этой задачи в онтологии были введены дополнительные аксиомы и правила логического вывода. В частности, свойство 'доминирую-

щая_модель' связывает участок берега с конкретным типом динамики (абразионная, аккумулятивная, абразионно-оползневая и др.) на основе комплекса диагностических признаков.

Пример SPARQL-запроса для идентификации модели динамики:

```
PREFIX : <http://www.semanticweb.org/moks/ontologies/2023/1/geolprocesses#>
SELECT ?Участок ?МодельДинамики ?Обоснование
WHERE {
    ?Участок :доминирующая_модель ?МодельДинамики .
    ?Участок :имеет_обоснование ?Обоснование .
    FILTER(?МодельДинамики IN (:Абразионная, :Абразионно-Оползневая, :Аккумулятивная))
}
```

А-Вох-онтологии. Наполнение онтологии конкретными данными и их интеграция в граф знаний позволили реализовать сценарий семантического запроса для решения конкретной научно-прикладной задачи. В качестве примера рассмотрим задачу выявления всех участков Братского водохранилища, на которых развиваются оползневые процессы, с указанием типа берега, SPARQL-запрос:

```
PREFIX : <http://www.semanticweb.org/moks/ontologies/2023/1/geolprocesses#>
SELECT ?Участок ?ТипБерега ?Процесс
WHERE {
    ?Участок :является_частью/:является_частью? :Береговая_зона_Братского_водохр .
    ?Участок rdf:type ?ТипБерега .
    FILTER(?ТипБерега IN (:Абразионный, :Аккумулятивный))
    ?Процесс rdf:type :Оползень ;
        :изменяет ?Участок .
}
```

Результат выполнения запроса: Система возвращает конкретные участки (например, Хадахан), для которых установлена связь между типом берега (Абразионный) и активным экзогенным процессом (Оползень_Хадахан). Это позволяет исследователю не просто видеть разрозненные данные, а идентифицировать зоны повышенного риска.

Таким образом, применение онтологического подхода к задаче интеграции данных о береговой зоне позволило преодолеть традиционную разрозненность информации. Построенный граф знаний обеспечил семантическую связность данных, что предоставило возможность формулировать комплексные запросы, расширяющие традиционные дисциплинарные границы между инженерной геологией и геоморфологией. Важным результатом является возможность идентификации доминирующей модели динамики береговой зоны, что критически важно для прогнозирования развития береговых процессов и обеспечения безопасности прибрежной инфраструктуры. Данный пример демонстрирует, каким образом технологии СВ превращают разрозненные массивы данных в инструмент для поддержки принятия решений при оценке рисков и планировании хозяйственной деятельности в береговой зоне.

ТЕХНИЧЕСКАЯ РЕАЛИЗАЦИЯ МЕТОДОЛОГИИ

Реализация предложенной методологии основана на современном технологическом стеке, обеспечивающем полный цикл работы с семантическими данными. Архитектура системы построена по многоуровневому принципу, где каждый уровень решает определенный класс задач. На уровне хранения и обработки данных используются графовые базы данных GraphDB, Apache Jena Fuseki, Openlink Virtuoso, которые обеспечивают надежное хранение RDF-данных и предоставление SPARQL-эндпоинтов для доступа к семантически структурированной информации.

Для представления данных применяются стандартизованные форматы: RDF/Turtle для онтологий, OWL2 XML для обмена метаданными между системами, JSON-LD для веб-интеграции. Серверная часть реализована на Python 3.x с использованием специализированных библиотек RDLib и SPARQLWrapper, которые обеспечивают выполнение ETL-процессов по преобразованию геологических данных в семантические форматы.

Разработка онтологических моделей осуществляется с помощью инструментария Protégé, который служит основной средой для визуального проектирования и валидации Т-Бокс-онтологий. Для выполнения запросов к данным используется язык SPARQL 1.1, стандартно поддерживающий федеративные запросы к распределенным источникам информации.

Пользовательские интерфейсы веб-приложений построены на основе React.js и React-оберток библиотек, обеспечивающих отображение табличных, а также пространственных данных на топологических основах. Семантическое обогащение веб-интерфейсов достигается за счет применения RDFa-разметки, а прямой доступ к графикам знаний предоставляется через специализированные SPARQL-интерфейсы.

Обеспечение соответствия FAIR-принципам [5] является ключевым аспектом технической реализации. Для реализации принципа находимости (Findable) всем объектам присваиваются устойчивые URI через зарегистрированные пространства имен, а онтологии регистрируются в специализированных репозиториях. Доступность (Accessible) обеспечивается через открытые SPARQL-эндпоинты с современными механизмами аутентификации и поддержкой стандартных форматов данных.

Совместимость (Interoperable) достигается за счет использования стандартных онтологий в качестве основы и поддержки кросс-доменных запросов. Многоразовость (Reusable) обеспечивается модульной архитектурой онтологий с четким разделением доменов и детальными метаданными о происхождении и качестве данных. Все компоненты системы лицензируются в соответствии с открытыми стандартами, что способствует их широкому распространению в научном сообществе.

СРАВНЕНИЕ С АНАЛОГАМИ

Проведенный анализ современных подходов к интеграции геологических данных выявил несколько ключевых направлений, каждое из которых обладает определенными преимуществами и ограничениями в контексте решения задач семантической интеграции разнородных геологических данных.

В области стандартизации геологических данных сформировались два принципиально различных подхода. GeoSciML [20] представляет собой практико-ориентированный стандарт обмена геологической информацией, основанный на языке GML (Geography Markup Language). Этот стандарт, разработанный Комиссией по управлению и применению геонаучной информации (CGI), обеспечивает

синтаксическую интероперабельность данных через интеграцию с OGC-сервисами, что делает его удобным инструментом для организаций, работающих в рамках существующей геоинформационной инфраструктуры. Однако GeoSciML не решает задач глубокой семантической интеграции и ограничивается стандартизацией форматов обмена.

Альтернативный подход представлен онтологией SWEET (Semantic Web for Earth and Environmental Terminology) [21], разработанной NASA для семантической разметки данных о Земле и окружающей среде. SWEET обеспечивает формальное описание понятий и их взаимосвязей через ортогональные концепции (пространство, время, физические свойства) и синтетические знания (явления, события). Это позволяет автоматически расширять поисковые запросы синонимами и связанными понятиями, значительно повышая релевантность результатов. Однако SWEET остается преимущественно теоретическим конструктом, слабо адаптированным для работы с реальными полевыми данными и решением конкретных прикладных задач.

Международные инициативы в области интеграции геологических данных демонстрируют различные стратегические подходы. OneGeology [22] представляет собой глобальную распределенную систему создания цифровой геологической карты планеты с участием 117 стран. Проект успешно реализует модель распределенной архитектуры, где данные сохраняются у организаций-владельцев, которые обеспечивают их актуальность и контроль. Однако фокус OneGeology ограничен преимущественно картографическими данными, что не позволяет решать задачи глубокой семантической интеграции разнородных геологических данных из различных дисциплин.

Более комплексные подходы демонстрируют проекты EarthCube (США) [23] и EPOS (Европейский союз). EarthCube ориентирован на создание киберинфраструктуры в области «Науки о Земле» через построение «системы систем» с распределенным управлением по аналогии с интернетом. Проект сталкивается с комплексными вызовами, включая концептуальные (время, масштаб), социально-культурные (мотивация, сотрудничество) и технические (интероперабельность) аспекты. European Plate Observing System (EPOS) [24] представляет собой панъевропейскую инициативу по интеграции исследовательских инфраструктур в

области наук о Земле, объединяя более 250 инфраструктур из 18 стран. Особое внимание в EPOS уделяется мониторингу тектонических процессов и связанных с ними природных опасностей, включая специализированные решения для работы в арктическом регионе.

В российской практике преобладает парадигма централизованных систем, характерным примером которой является ГАС ВСЕГЕИ [4]. Данный проект демонстрирует высокую эффективность в решении узкоспециализированной задачи унификации геохронологической информации для территории России. Ключевыми преимуществами системы являются строгий контроль качества через экспертуру валидацию (отбраковка ~20% исходных определений) и интеграция с серийными легендами Государственных геологических карт. Централизованная архитектура через аккредитованный Центр изотопных исследований ВСЕГЕИ обеспечивает единые стандарты качества и методологии. Однако такая модель создает существенные ограничения для интеграции разнородных данных из смежных дисциплин и обладает высоким «порогом входа» для независимых исследователей.

Проведенный анализ выявил устойчивый разрыв между теоретическими возможностями семантических технологий (SWEET) и их практической реализацией в существующих системах (GeoSciML, OneGeology, EarthCube, EPOS, ГАС). Собственно, этот разрыв и был преодолен в данной работе через разработку методологии, которая синтезирует семантическую строгость онтологических подходов с практической ориентированностью существующих стандартов. Предложенное решение обеспечивает гибкость распределенных архитектур и позволяет интегрировать разнородные алгоритмы и данные для решения конкретных междисциплинарных научных задач в геологии.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Представленный подход доказал свою практическую ценность в различных геологических дисциплинах, что подтверждается успешным применением в геохимии, тектонике и инженерной геологии. Возможность кросс-дисциплинарного использования свидетельствует о продуманной архитектуре, способной работать с разнородными геологическими данными.

Использование стандартизованных технологий СВ (RML, SPARQL) обеспечивает совместимость систем и снижает порог внедрения за счет повторно используемых компонентов.

Однако следует учитывать существенные ограничения. Высокие требования к экспертизе в области СВ создают серьезный барьер для внедрения, особенно в геологических коллективах без соответствующей подготовки. Разработанная методология проверена на данных умеренного объема, тогда как ее масштабируемость для работы с петабайтными массивами геологической информации требует дополнительного изучения.

Эффективность предложенного подхода сильно зависит от исходного качества данных. Как показала практика, неоднородность единиц измерений, методов анализа и таксономий требует значительных усилий по их нормализации. Особенно критичен этап онтологического моделирования: ошибки проектирования на этой стадии негативно влияют на все последующие этапы.

Сложности интеграции с унаследованным геологическим ПО и специализированными инструментами также заслуживают внимания. Отсутствие готовых решений для подключения требует разработки собственных адаптеров, что увеличивает сложность и стоимость сопровождения.

Несмотря на отмеченные ограничения, методология успешно решает ключевые задачи семантической интеграции и открывает новые возможности для исследований благодаря федеративным запросам и логическому выводу. Итеративный процесс разработки с непрерывной валидацией обеспечивает необходимую гибкость для адаптации к конкретным предметным областям.

Перспективные направления развития включают создание масштабируемых архитектур для работы с большими данными, разработку учебных материалов для геологов и стандартизацию интеграции с распространенным геологическим ПО.

Предложенная методология применима в исследовательских проектах средней сложности, требующих интеграции разнородных данных. Для промышленного использования понадобятся дополнительные проверки и оптимизация производительности.

ЗАКЛЮЧЕНИЕ

Разработана и апробирована методология применения технологий СВ для поддержки фундаментальных геологических исследований. Основным результатом работы стало создание части комплексной онтологической модели, обеспечивающей семантическую интеграцию разнородных геологических данных из различных дисциплин: геохимии, тектоники и инженерной геологии. Практическая реализация подхода продемонстрирована на трех кейсах: анализе геохимических данных для оценки уровня загрязнения территории, создании информационной системы о разломах и изменении береговой зоны. Представленная архитектура на основе графов знаний и SPARQL-эндпоинтов подтвердила свою эффективность в преодолении «семантического барьера» и обеспечении соответствия FAIR-принципам.

Полученный опыт эксплуатации приложений, разработанных на основе предложенного подхода, выявил как значительные преимущества, так и определенные ограничения. К достоинствам относятся гибкость онтологической модели, способность адаптироваться к эволюционирующими научным парадигмам, поддержка распределенной архитектуры с сохранением автономии исследовательских коллективов, а также возможность создания подсистем естественно-языковых запросов на основе интеграции онтологических моделей и больших языковых моделей (LLM). Однако отмечена высокая сложность начального этапа разработки онтологий, необходимость обучения исследователей работе с семантическими технологиями, а также вычислительная нагрузка при обработке крупных графов знаний. Эти факторы могут ограничивать широкое внедрение подхода в геологическом сообществе.

Перспективы развития исследования связаны с несколькими направлениями. Во-первых, это расширение онтологической модели для охвата дополнительных геологических дисциплин и интеграции с международными стандартами типа GeoSciML. Во-вторых, разработка специализированных инструментов автоматического преобразования традиционных геологических данных в RDF-формат. Особый интерес представляет создание когнитивных систем поддержки принятия решений на базе разработанной платформы, которые будут объединять се-

мантически интегрированные данные, механизмы логического вывода и предиктивную аналитику для решения таких сложных задач, как прогнозирование развития деградационных процессов на определенных участках береговой зоны и экспертная оценка хозяйственного использования этой зоны, оценка сейсмической опасности и оптимизация работ по экологическому мониторингу. Таким образом, дальнейшая работа будет направлена на создание открытой платформы для семантической интеграции геологических знаний с вовлечением широкого круга заинтересованных лиц.

Благодарности

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках гранта № 075-15-2024-533 на реализацию крупных научных проектов по приоритетным направлениям научно-технологического развития, проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы для экологического мониторинга окружающей среды».

СПИСОК ЛИТЕРАТУРЫ

- 1) *Mattmann C.* A vision for data science // Nature. 2013. Vol. 493. P. 473–475. <https://doi.org/10.1038/493473a>
- 2) *Hitzler P., Janowicz K.* Semantic Web for Earth and Environmental Science: Current state and future directions // Semantic Web. 2010. Vol. 1, No. 1–2. P. 85–98. <https://doi.org/10.3233/SW-2010-0012>
- 3) *Hitzler P.* A review of the semantic web field // Communications of the ACM. 2021. Vol. 64, No. 2. P. 76–83. <https://doi.org/10.1145/3397512>
- 4) Синькова Е.А., Петров О.В., Ханчук А.И. Геохронологический атлас-справочник основных структурно-вещественных комплексов России – базовый информационный ресурс для геологической отрасли страны // Разведка и охрана недр. 2022. № 90. С. 5–14. https://doi.org/10.52349/0869-7892_2022_90_5-14
- 5) Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al. The FAIR Guiding Principles for scientific data management and stewardship // Scientific Data. 2016. Vol. 3. P. 160018. <https://doi.org/10.1038/sdata.2016.18>

- 6) *Laney D.* 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group Research Note. 2001. Vol. 6, No. 70.
URL: <https://diegonogare.net/wp-content/uploads/2020/08/3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- 7) *Gandomi A., Haider M.* Beyond the hype: Big data concepts, methods, and analytics // International Journal of Information Management. 2015. Vol. 35, No. 2. P. 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- 8) *Hogan A., Blomqvist E., Cochez M., D'Amato C. et al.* Knowledge Graphs // ACM Computing Surveys. 2021. Vol. 54, No. 4. <https://doi.org/10.1145/3447772>
- 9) *Pellinen V.A., Cherkashina T.Y., Gustaitis M.A.* Assessment of metal pollution and subsequent ecological risk in the coastal zone of Olkhon Island, Lake Baikal, Russia // Science of the Total Environment. 2021. Vol. 786. P. 147441.
<https://doi.org/10.1016/j.scitotenv.2021.147441>
- 10) *Zerizghi T., Yang Y., Wang W., Zhou Y., Zhang J., Yi Y.* Ecological risk assessment of heavy metal concentrations in sediment and fish of a shallow lake: a case study of Baiyangdian Lake, North China // Environmental Monitoring and Assessment. 2020. Vol. 192. P. 154. <https://doi.org/10.1007/s10661-020-8078-8>
- 11) *Zhang J., Wang K., Yi Q., Zhang T., Shi W., Zhou X.* Transport and partitioning of metals in river networks of a plain area with sedimentary resuspension and implications for downstream lakes // Environmental Pollution. 2022. Vol. 294. P. 118668.
<https://doi.org/10.1016/j.envpol.2021.118668>
- 12) *Gopal V., Krishnamurthy R.R., Vignesh R., Nathan C.S. et al.* Assessment of heavy metal contamination in the surface sediments of the Vedaranyam coast, Southern India // Regional Studies in Marine Science. 2023. Vol. 65. P. 103081.
<https://doi.org/10.1016/j.rsma.2023.103081>
- 13) *Chubarov V., Cherkashina T., Maltsev A., Chuparina E., Amosova A., Prosekkin S.* Investigation of Soils and Pine Needles Using WDXRF and TXRF Techniques for Assessment of the Environmental Pollution of Shelekhov District, Eastern Siberia, by the Aluminum Industry and Heat Power Engineering // Agronomy. 2022. Vol. 12. P. 454.
<https://doi.org/10.3390/agronomy12020454>

- 14) *Lunina O.V.* The digital map of the Pliocene–Quaternary crustal faults in the Southern East Siberia and the adjacent Northern Mongolia // Geodynamics and Tectonophysics. 2016. Vol. 7, No. 3. P. 407–434.
<https://doi.org/110.5800/GT-2016-7-3-0215>
- 15) *Cherkashin E.A., Lunina O.V., Demyanov L.O., Tsygankov A.V.* Web-GIS viewer for active faults data represented as a knowledge graph // The 4th Scientific-practical Workshop Information Technologies: Algorithms Models, Systems. September 14, 2021, Irkutsk, Russia / CEUR Workshop Proceedings. 2021. Vol. 2984. P. 55–65. URL: <https://ceur-ws.org/Vol-2984/paper8.pdf>
- 16) *Mazaeva O., Babicheva V., Kozyreva E.* Geomorphological process development under the impact of man-made reservoir operation, a case study: Bratsk reservoir, Baikal-Angara hydroengineering system, Russia // Bulletin of Engineering Geology and the Environment. 2019. Vol. 78. P. 4659–4672.
<https://doi.org/110.1007/s10064-018-1428-x>
- 17) *Овчинников Г.И., Павлов С.Х., Тржцинский Ю.Б.* Изменения геологической среды в зонах влияния ангаро-енисейских водохранилищ. Новосибирск: Наука, 1999. 254 с.
- 18) *Ranatunga S., Ødegård R.S., Jetlund K., Onstein E.* Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geo-spatial Data: A Framework Based on Ontology-Based Data Access // ISPRS International Journal of Geo-Information. 2025. Vol. 14, No. 2. P. 52.
<https://doi.org/110.3390/ijgi14020052>
- 19) *Moura A.-M., Porto F., Vidal V., Magalhães R.P. et al.* A semantic integration approach to publish and retrieve ecological data // International Journal of Web Information Systems. 2015. Vol. 11, No. 1. P. 87–119.
<https://doi.org/110.1108/IJWIS-08-2014-0028>
- 20) *Marcus S., Tim D.* GeoSciML: Development of a generic GeoScience Markup Language // Computers & Geosciences. 2005. Vol. 31, No. 9. P. 1095–1103.
<https://doi.org/10.1016/j.cageo.2004.12.003>
- 21) *Raskin R., Pan M.* Knowledge representation in the semantic web for Earth and environmental terminology (SWEET) // Computers & Geosciences. 2005. Vol. 31, No. 9. P. 1119–1125. <https://doi.org/110.1016/j.cageo.2004.12.004>
-

- 22) *Simons B.A., Raymond O., Jackson I., Lee K.* OneGeology – Improving global access to geoscience // The 5th Global Workshop on Digital Soil Mapping. April 10–13, 2012, Sydney, Australia / Digital Soil Assessments and Beyond: Proceedings of CRC Press, 2012. P. 265–275.
- 23) *Allison M.L., Ahern T., Arctur D. et al.* EarthCube Governance Framework: A Proposal to the Community (Version 1.0) // EarthCube Governance Working Group Technical Report. 2012. 237 p.
- 24) *Atakan K., Bjerrum L.W., Bungum H. et al.* The European Plate Observing System and the Arctic // Arctic. 2015. Vol. 68, Suppl. 1.
<https://doi.org/10.14430/arctic4446>

SEMANTIC WEB TECHNOLOGIES FOR SUPPORTING FUNDAMENTAL RESEARCH IN GEOLOGY

E. A. Cherkashin¹ [0000-0003-2428-2471], I. V. Bychkov² [0000-0002-1765-0769],
Jin Zhang³ [0000-0003-0009-824X], T. Yu. Cherkashina⁴ [0000-0002-1443-6839],
V. A. Popova⁵ [0000-0002-7764-1995], O. A. Mazaeva⁶ [0000-0002-3053-5511],
O. V. Lunina⁷ [0000-0001-7743-8877]

^{1, 2, 5} V. M. Matrosov Institute of System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, 664033, Russia

^{1, 2, 4} School of Mathematics and Information Science, Yantai University, Yantai, 264005, China

³School of Civil Engineering, Yantai University, Yantai, 264005, China

^{4, 6, 7} Institute of the Earth's Crust of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, 664033, Russia

^{1, 5}Institute of Mathematics and Informatics, Irkutsk State University, Irkutsk, 664003, Russia

¹eugeneai@icc.ru, ²bychkov@icc.ru, ³zhangjin513@outlook.com, tcherk@crust.irk.ru,
⁵victorypopova1@gmail.com, ⁶moks@crust.irk.ru, ⁷lounina@crust.irk.ru

Abstract

The article presents an innovative methodology for applying Semantic Web technologies to support fundamental geological research. The problem of semantic integration of heterogeneous geological data, characterized by different scales and interdisciplinarity, is considered. A five-stage methodology is developed, including domain analysis, ontological conceptual modeling, data transformation into a knowledge graph, deployment of a distributed data access infrastructure based on the conceptual model, and integration with processing and analysis procedures. Practical testing was conducted on three case studies: analysis of geochemical data for assessing territory pollution levels, creation of an information system about faults, and research on reservoir shoreline dynamics. The proposed ontological approach ensures compliance with FAIR principles and overcoming the "semantic barrier" in geological research. It is shown that Semantic Web technologies enable a transition from fragmented infor-

mation arrays to a holistic semantic space of geological knowledge, opening new opportunities for generating comprehensive scientific hypotheses and cross-disciplinary research.

Keywords: Semantic Web, knowledge graphs, ontological modeling, semantic environment of scientific knowledge, fundamental geological research.

REFERENCES

1. *Mattmann C.* A vision for data science // Nature. 2013. Vol. 493. P. 473–475. <https://doi.org/10.1038/493473a>
2. *Hitzler P., Janowicz K.* Semantic Web for Earth and Environmental Science: Current state and future directions // Semantic Web. 2010. Vol. 1, No. 1–2. P. 85–98. <https://doi.org/10.3233/SW-2010-0012>
3. *Hitzler P.* A review of the semantic web field // Communications of the ACM. 2021. Vol. 64, No. 2. P. 76–83. <https://doi.org/10.1145/3397512>
4. *Sinkova E.A., Petrov O.V., Khancuk A.I.* Geochronological Atlas and Reference Book of the Main Structural-Material Complexes of Russia – a Basic Information Resource for the Country's Geological Industry // Exploration and Protection of Mineral Resources. 2022. No. 90. P. 5–14.
https://doi.org/10.52349/0869-7892_2022_90_5-14
5. *Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al.* The FAIR Guiding Principles for scientific data management and stewardship // Scientific Data. 2016. Vol. 3. P. 160018. <https://doi.org/10.1038/sdata.2016.18>
6. *Laney D.* 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group Research Note. 2001. Vol. 6, No. 70.
URL: <https://diegonogare.net/wp-content/uploads/2020/08/3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
7. *Gandomi A., Haider M.* Beyond the hype: Big data concepts, methods, and analytics // International Journal of Information Management. 2015. Vol. 35, No. 2. P. 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
8. *Hogan A., Blomqvist E., Cochez M., D'Amato C. et al.* Knowledge Graphs // ACM Computing Surveys. 2021. Vol. 54, No. 4. <https://doi.org/10.1145/3447772>

9. *Pellinen V.A., Cherkashina T.Y., Gustaitis M.A.* Assessment of metal pollution and subsequent ecological risk in the coastal zone of Olkhon Island, Lake Baikal, Russia // *Science of the Total Environment*. 2021. Vol. 786. P. 147441.
<https://doi.org/110.1016/j.scitotenv.2021.147441>
10. *Zerizghi T., Yang Y., Wang W., Zhou Y., Zhang J., Yi Y.* Ecological risk assessment of heavy metal concentrations in sediment and fish of a shallow lake: a case study of Baiyangdian Lake, North China // *Environmental Monitoring and Assessment*. 2020. Vol. 192. P. 154. <https://doi.org/110.1007/s10661-020-8078-8>
11. *Zhang J., Wang K., Yi Q., Zhang T., Shi W., Zhou X.* Transport and partitioning of metals in river networks of a plain area with sedimentary resuspension and implications for downstream lakes // *Environmental Pollution*. 2022. Vol. 294. P. 118668.
<https://doi.org/110.1016/j.envpol.2021.118668>
12. *Gopal V., Krishnamurthy R.R., Vignesh R., Nathan C.S. et al.* Assessment of heavy metal contamination in the surface sediments of the Vedaranyam coast, Southern India // *Regional Studies in Marine Science*. 2023. Vol. 65. P. 103081.
<https://doi.org/110.1016/j.rsma.2023.103081>
13. *Chubarov V., Cherkashina T., Maltsev A., Chuparina E., Amosova A., Prosekin S.* Investigation of Soils and Pine Needles Using WDXRF and TXRF Techniques for Assessment of the Environmental Pollution of Shelekhov District, Eastern Siberia, by the Aluminum Industry and Heat Power Engineering // *Agronomy*. 2022. Vol. 12. P. 454.
<https://doi.org/110.3390/agronomy12020454>
14. *Lunina O.V.* The digital map of the Pliocene–Quaternary crustal faults in the Southern East Siberia and the adjacent Northern Mongolia // *Geodynamics and Tectonophysics*. 2016. Vol. 7, No. 3. P. 407–434.
<https://doi.org/110.5800/GT-2016-7-3-0215>
15. *Cherkashin E.A., Lunina O.V., Demyanov L.O., Tsygankov A.V.* Web-GIS viewer for active faults data represented as a knowledge graph // The 4th Scientific-practical Workshop Information Technologies: Algorithms Models, Systems. September 14, 2021, Irkutsk, Russia / CEUR Workshop Proceedings. 2021. Vol. 2984. P. 55–65. URL: <https://ceur-ws.org/Vol-2984/paper8.pdf>

16. *Mazaeva O., Babicheva V., Kozyreva E.* Geomorphological process development under the impact of man-made reservoir operation, a case study: Bratsk reservoir, Baikal-Angara hydroengineering system, Russia // Bulletin of Engineering Geology and the Environment. 2019. Vol. 78. P. 4659–4672.
<https://doi.org/110.1007/s10064-018-1428-x>
17. *Ovchinnikov G.I., Pavlov S.Kh., Trzhitsinsky Yu B.* Changes in the Geological Environment in the Zones of Influence of the Angara–Yenisei Reservoirs. Novosibirsk: Nauka, 1999. 254 p.
18. *Ranatunga S., Ødegård R.S., Jetlund K., Onstein E.* Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geo-spatial Data: A Framework Based on Ontology-Based Data Access // ISPRS International Journal of Geo-Information. 2025. Vol. 14, No. 2. P. 52.
<https://doi.org/110.3390/ijgi14020052>
19. *Moura A.-M., Porto F., Vidal V., Magalhães R.P. et al.* A semantic integration approach to publish and retrieve ecological data // International Journal of Web Information Systems. 2015. Vol. 11, No. 1. P. 87–119.
<https://doi.org/110.1108/IJWIS-08-2014-0028>
20. *Marcus S., Tim D.* GeoSciML: Development of a generic GeoScience Markup Language // Computers & Geosciences. 2005. Vol. 31, No. 9. P. 1095–1103.
<https://doi.org/10.1016/j.cageo.2004.12.003>.
21. *Raskin R., Pan M.* Knowledge representation in the semantic web for Earth and environmental terminology (SWEET) // Computers & Geosciences. 2005. Vol. 31, No. 9. P. 1119–1125. <https://doi.org/110.1016/j.cageo.2004.12.004>
22. *Simons B.A., Raymond O., Jackson I., Lee K.* OneGeology – Improving global access to geoscience // The 5th Global Workshop on Digital Soil Mapping. April 10–13, 2012, Sydney, Australia / Digital Soil Assessments and Beyond: Proceedings of CRC Press, 2012. P. 265–275.
23. *Allison M.L., Ahern T., Arctur D. et al.* EarthCube Governance Framework: A Proposal to the Community (Version 1.0) // EarthCube Governance Working Group Technical Report. 2012. 237 p.

24. Atakan K., Bjerrum L.W., Bungum H. et al. The European Plate Observing System and the Arctic // Arctic. 2015. Vol. 68, Suppl. 1. <https://doi.org/110.14430/arctic4446>
-

СВЕДЕНИЯ ОБ АВТОРАХ



БЫЧКОВ Игорь Вячеславович – доктор технических наук, профессор, академик РАН, директор Института динамики систем и теории управления им. В.М. Матросова Сибирского отделения РАН, член Президиума РАН, член бюро Отделения нанотехнологий и информационных технологий РАН, почетный профессор Института математики и компьютерных наук Янтайского университета (Китай).

Научные интересы: искусственный интеллект, геоинформационные системы, web-технологии, системы интеллектуального анализа данных, математическое моделирование, облачные вычисления, большие данные, многоагентные системы, информационно-вычислительные и телекоммуникационные технологии.

Igor Vyacheslavovich BYCHKOV – Doctor of Technical Sciences, Professor, Member of the Russian Academy of Sciences, Director of the V.M. Matrosov Institute of System Dynamics and Control Theory of Siberian Branch of the Russian Academy of Sciences, Member of the Presidium of RAS, Member of the Bureau of the Department of Nanotechnology and Information Technologies of RAS, honorary professor of Institute of Mathematics and Computer Science at Yantai University (China).

Research interests: artificial intelligence, geoinformation systems, web technologies, data mining systems, mathematical modeling, cloud computing, big data, multi-agent systems, information-computing and telecommunication technologies.

email: bychkov@icc.ru

ORCID: 0000-0002-1765-0769



ЧЕРКАШИН Евгений Александрович – кандидат технических наук, доцент, заведующий кафедрой информационных технологий Института математики и информатики Иркутского государственного университета, старший научный сотрудник Института динамики систем и теории управления им. В.М. Матросова Сибирского отделения РАН, профессор Института математики и компьютерных наук Янтайского университета (Китай).

Научные интересы: синтез программных объектов на основе абстрактных моделей (Model driven architecture); логическое программирование; системы, основанные на формализованных знаниях; обработка естественного языка; системы поддержки принятия решений на основе математического моделирования; психология педагогического процесса.

Evgeny Alexandrovich CHERKASHIN – Candidate of Technical Sciences, Associate Professor, Head of the Department of Information Technologies of Institute of Mathematics and Computer Science at Irkutsk State University, Senior Researcher at the V.M. Matrosov Institute of System Dynamics and Control Theory of Siberian Branch of the Russian Academy of Sciences.

Research interests: software objects synthesis based on abstract models (Model driven architecture), logical programming, systems based on formalized knowledge, natural language processing, decision support systems based on mathematical modeling, psychology of pedagogical process, honorary professor of Institute of Mathematics and Computer Science at Yantai University (China).

email: eugeneai@icc.ru

ORCID: 0000-0003-2428-2471



ЧЖАН Цзинь – профессор, заместитель декана Института гражданского строительства Янтайского университета.

Научные интересы: экологическая инженерия, фосфорный цикл в водных системах, мембранные загрязнения, очистка сточных вод с использованием микроводорослей.

Jin ZHANG – Professor, Vice Dean of the School of Civil Engineering, Yantai University.

Research interests: environmental engineering, phosphorus cycling in aquatic systems, membrane fouling, microalgae-based wastewater treatment.

email: zhangjin513@outlook.com

ORCID: 0000-0003-0009-824X



ЧЕРКАШИНА Татьяна Юрьевна – кандидат геолого-минералогических наук, старший научный сотрудник Центра коллективного пользования «Геодинамика и геохронология» Института земной коры Сибирского отделения РАН, профессор Института математики и компьютерных наук Янтайского университета (Китай).

Научные интересы: рентгенофлуоресцентный анализ горных пород, геохимия осадочных бассейнов юга Сибирской платформы, оценка экологического состояния почвенного покрова и береговой зоны Байкальской природной территории.

Tatiana Yurievna CHERKASHINA – Candidate of Geological and Mineralogical Sciences, Senior Researcher at the Center for Geodynamics and Geochronology of the Institute of the Earth's Crust, Siberian Branch of the Russian Academy of Sciences, professor of Institute of Mathematics and Computer Science at Yantai University (China).

Research interests: X-ray fluorescence analysis of rocks, geochemistry of sedimentary basins of the Siberian Platform, assessment of the ecological and geochemical state of the soil cover and the coastal zone of the Baikal natural territory.

email: therk@crust.irk.ru

ORCID: 0000-0002-1443-6839



ПОПОВА Виктория Алексеевна – ассистент кафедры алгебраических и информационных систем Института математики и информационных технологий Иркутского государственного университета, инженер-программист.

Научные интересы: разработка программного обеспечения, криптографические методы защиты информации, статистические анализаторы программного кода, разработка web-приложений.

Victoria Alexeevna POPOVA – Assistant of the Department of Algebraic and Information Systems, Institute of Mathematics and Information Technology of Irkutsk State University, Software Engineer.

Research interests: software development, cryptographic methods of information protection, static code analyzers, web-application development.

email: victorypopova1@gmail.com

ORCID: 0000-0002-7764-1995



МАЗАЕВА Оксана Анатольевна – кандидат геолого-минералогических наук, научный сотрудник лаборатории инженерной геологии и геоэкологии Института земной коры Сибирского отделения РАН.

Научные интересы: исследование эрозионных и других экзогенных геологических процессов в природных и природно-технических геосистемах юга Сибирской платформы, изучение состава и свойств рыхлых отложений.

Oksana Anatolievna MAZAEVA – Candidate of Geological and Mineralogical Sciences, Researcher at the Laboratory of Engineering Geology and Geoecology of the Institute of the Earth's Crust, Siberian Branch of the Russian Academy of Sciences.

Research interests: study of erosion and other exogenous geological processes in natural and natural-technical geosystems in the south of the Siberian Platform, investigation of the composition and properties of loose sediments.

email: moks@crust.irk.ru

ORCID: 0000-0002-3053-5511



ЛУНИНА Оксана Викторовна – доктор геолого-минералогических наук, главный научный сотрудник лаборатории тектонофизики Института земной коры Сибирского отделения РАН.

Научные интересы: тектоника и геодинамика, структурная геология, тектонофизика, сейсмотектоника, изучение строения земной коры и опасных природных процессов.

Oksana Viktorovna LUNINA – Doctor of Geological and Mineralogical Sciences, Chief Researcher at the Laboratory of Tectonophysics of the Institute of the Earth's Crust, Siberian Branch of the Russian Academy of Sciences.

Research interests: tectonics and geodynamics, structural geology, tectonophysics, seismotectonics, study of the Earth's crust structure and hazardous natural processes.

email: lounina@crust.irk.ru

ORCID: 0000-0001-7743-8877

Материал поступил в редакцию 14 октября 2025 года

УДК 004.85

ЦИФРОВОЙ ПОМОЩНИК ГЕОЛОГА-ИССЛЕДОВАТЕЛЯ

В. С. Ерёменко¹ [0000-0002-5250-5743], В. В. Наумова² [0000-0002-3001-1638]

^{1, 2}Государственный геологический музей им. В.И. Вернадского РАН,
г. Москва, Россия

¹vitaer@gmail.com, ²naumova_new@mail.ru

Аннотация

Представлены концепция и архитектура мультиагентной системы, пред назначенной для функционирования в роли цифрового ассистента геолога-исследователя. Система нацелена на автоматизацию ключевых этапов научного исследования: от формулировки темы и анализа литературы до выдвижения гипотез и оформления результатов. Описана интеграция системы с платформой GeologyScience.ru, обеспечивающей доступ к разнородным геологическим данным и инструментам анализа, а также подходы к адаптации больших языковых моделей (LLM) для решения специализированных научных задач.

Ключевые слова: цифровой помощник геолога-исследователя, мультиагентная система, искусственный интеллект, LLM.

ВВЕДЕНИЕ

Искусственный интеллект (ИИ) используется в геологии для автоматизации анализа данных, моделирования процессов и предсказания природных явлений. Это позволяет ускорить исследовательские процессы и повысить точность результатов, особенно в областях, где классические методы теряют эффективность.

Примеры использования мультиагентных систем для решения научных задач встречаются в различных научных областях [1]. Эффективность мультиагентных систем для решения сложных геологических задач подтверждается рядом успешных работ. Среди них можно выделить обнаружение месторождений железной руды в Западной Австралии [2], выявление дополнительных запасов нефти в уже разрабатываемых пластах [3]. Эти примеры демонстрируют потенциал ИИ систем для интеллектуального анализа данных и поддержки принятия

решений в геологии. Имеются и другие примеры использования ИИ в рассматриваемой области [3].

ЦИФРОВОЙ АССИСТЕНТ И ЦИФРОВЫЕ ИИ-АГЕНТЫ ДЛЯ ПОДДЕРЖКИ И СОПРОВОЖДЕНИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

Цифровой помощник геолога-исследователя определяется как программа, взаимодействующая с пользователем в диалоговом режиме на естественном языке. Его ключевая задача – это сопровождение пользователя через протокол научного исследования.

Базовой основой цифрового ассистента геолога-исследователя на Портале <https://geologyscience.ru> является чат-бот по формализации и сопровождению научного исследования на всех его этапах: это помочь в формулировке темы, поиск возможных путей исследования, поиск и анализ литературных источников, различных видов данных.

Использование современных больших языковых моделей (LLM) общего назначения без дополнительной адаптации не обеспечивает необходимого качества для постановки научных задач. Для повышения качества получаемых результатов предлагается использовать комбинацию следующих подходов:

- составление специализированных промптов для управления диалогом и фокусировки на научной проблематике;
- генерация, дополненная поиском, (Retrieval-Augmented Generation, RAG) для доступа к актуальным и релевантным данным из научных публикаций и базам знаний платформы;
- тонкая настройка (Fine-tuning) для повышения компетентности модели в предметной области геологии.

Предлагаемая нами в качестве сопровождения мультиагентная система представляет собой комбинацию небольших узкоспециализированных программных ИИ-агентов, контролируемых общим управляемым агентом (рис. 1).



Рис. 1. Концептуальная схема работы мультиагентной системы.

Мультиагентная система функционирует как управляющий узел, взаимодействующий с ресурсами и сервисами платформы GeologyScience.ru [4]. Платформа предоставляет доступ к геологическим данным, таким как пространственные данные, количественные данные, спутниковые снимки, научные публикации, тематические и государственные геологические базы данных; а также к инструментам обработки и анализа, таким как ИИ-анализ геологических текстов, обработка геохимических данных, статистический анализ, визуализация [5].

Пользовательский интерфейс для взаимодействия с сервисами реализован как виртуальный рабочий кабинет с возможностью совместной работы, доступный через веб-браузер (рис. 2). Функциональность системы расширяется за счет подключения плагинов [6].

The screenshot shows the 'GeologyScience.ru' website's user interface. At the top, there is a navigation bar with links for 'ДАННЫЕ', 'СЕРВИСЫ', 'ПОИСК', 'НОВОСТИ', 'ГЕОЛОГИ', and 'О ПРОЕКТЕ'. Below the navigation bar, a header reads 'Рабочий кабинет пользователя: pkvkc'. On the left, a sidebar titled 'Данные' contains buttons for 'Поиск данных' and 'Загрузить'. The main workspace is divided into several windows: one showing a PDF document titled 'Золото-сульфидная минерализация в кварц-сульфидных проприях Рудоразведочного участка «Курбай»' (Gold-sulfide mineralization in quartz-sulfide veins of the Kurbay Prospecting Area); another showing an Excel spreadsheet titled 'Пример файла.xlsx'; and a third showing a PDF document titled 'Step_16.pdf'. To the right, a sidebar titled 'Сервисы' lists various services: 'Раздел "Сервисы" в разработке. Работает в тестовом режиме.', 'Последние новости', 'Поиск данных', 'Анализ данных', 'Визуализация данных', 'Сервисы с использованием ИИ', 'Чат бот по месторождениям', 'Сервисы нейросетевого анализа геологических текстов', and 'Создание программного кода'.

Рис. 2. Интерфейс рабочего кабинета пользователя

Взаимодействие с системой происходит через чат-интерфейс на естественном языке (рис. 3).

The screenshot shows a conversational interface with a sidebar of tasks and a main chat area.

Чат-бот

- План работы 21.08.2025
- Тестирование работы 21.08.2025
- Проверка гипотезы** 07.07.2025
- Анализ 21.08.2025

Проверка гипотезы

Создан: 07.07.2025, 17:34:12

Сихотэ-Алинский пояс:
Khanchuk A.I. et al. (2016) *Geodynamics of the Sikhote-Alin Orogen*. Работы ГИН РАН по Дальнему Востоку.

Геохимия базитов:
Pearce, J.A. (2008) *Geochemical fingerprinting of oceanic basalts*.

5. Ожидаемая научная новизна
Уточнение геодинамической позиции Калиновского комплекса.
Выявление специфики его магматизма (плаунте-мантийное взаимодействие, метасоматоз).
Сравнение с другими оphiолитами Восточной Азии (например, Ханкайский массив).

6. Практическая значимость
Понимание эволюции Сихотэ-Алиня как аккреционного орогена.
Возможные металлогенические аспекты (Cr, Ni, PGE минерализация).

7. План работы
Введение (актуальность, цели).
Литературный обзор (офиолиты, геология Сихотэ-Алиня).
Методы (полевые, лабораторные).
Результаты (описание комплекса, геохимия).
Обсуждение (сравнение с океаническими аналогами, геодинамика).
Заключение (выводы, перспективы).
Если у тебя есть конкретные данные (геохимия, карты), можно углубиться в детали. Также полезно посмотреть **цифровые геологические карты** (USGS, ГеоЛКарта России) и базы данных (EarthChem, PetDB).
Нужна помощь с анализом конкретных данных или подбором литературы?

17:39:58
Можешь ли ты порекомендовать литературу из репозитория научных публикаций

Новый чат Введите сообщение... Отправить

Рис. 3. Пользовательский интерфейс для взаимодействия с прототипом цифрового ассистента

В процессе диалога помощник предлагает релевантные научные публикации с возможностью их скачивания, выделяет и формулирует гипотезы на основе проведенного анализа данных, а также обеспечивает доступ к инструментам обработки и визуализации в рамках виртуального рабочего кабинета.

Для реализации системы используются ИИ-модель DeepSeek R1 [7], Python фреймворки Ollama и LangChain. Пользовательский интерфейс реализован в виде веб-приложения с использованием фреймворка Vue 3, а также RESTful API для интеграции с инструментами и базами данных GeologyScience.ru.

ЗАКЛЮЧЕНИЕ

Представлен прототип цифрового ассистента геолога-исследователя, разработанный для поддержки научных исследований в геологии. Система позволяет структурировать процесс исследования, обеспечивая пользователя интеллектуальным интерфейсом взаимодействия с нейронной сетью, а также доступом к

научной информации и геологическим данным и цифровым инструментам обработки и анализа. Дальнейшее развитие проекта связано с расширением протоколов взаимодействия между агентскими системами, что позволит увеличить функциональность и автономность цифрового ассистента, а также возможность непрерывной адаптации используемых ИИ-моделей к актуальным научным задачам.

Благодарности

Работа выполнена в рамках Государственной темы Государственного геологического музея им. В.И. Вернадского РАН (FMMR-2025-0001) Рег. №1022061000107-4-1.5.1 «Естественнонаучные музеи в едином цифровом пространстве геологических знаний».

СПИСОК ЛИТЕРАТУРЫ

1. KoBold used AI to find copper — now investors are piling in to the tune of \$537M. URL: <https://techcrunch.com/2025/01/02/kobold-used-ai-to-find-copper-now-investors-are-piling-in-to-the-tune-of-537m/>
 2. Цифровая система «Газпром нефти» нашла новые пласты нефти в ЯНАО. URL: <https://www.gazprom-neft.ru/press-center/news/tsifrovaya-sistema-gazprom-nefti-nashla-novye-plasty-nefti-v-yanao/>
 3. Gottweis J., Wei-Hung Weng, Daryin A. et al. Towards an AI co-scientist // arxiv.org. 2025. <https://doi.org/10.48550/arXiv.2502.18864>
 4. Наумова В.В., Еременко В.С., Загуменнов А.А., Патук М.И. Научный портал Geologyscience.ru: текущее состояние и перспективы развития// Геоинформатика. 2023. № 3. С. 33–43.
 5. Наумова В.В., Патук М.И., Еременко А.С., Загуменнов А.А., Еременко В.С. Цифровые технологии будущего – современные решения в науках о Земле // Геоинформатика. 2024. № 3. С. 53–63.
 6. Еременко В.С., Наумова В.В., Загуменнов А.А. Разработка виртуальных пространств с интеллектуальным адаптивным интерфейсом для цифровых геологических систем // Информационные и математические технологии в науке и управлении. 2024. № 1 (33). С. 146–153.
 7. DeepSeek 1R. URL: <https://www.deepseek.com>
-
-

DIGITAL ASSISTANT FOR GEOLOGIST-RESEARCHERS

V. S. Eremenko¹ [0000-0002-5250-5743], V. V. Naumova² [0000-0002-3001-1638]

Vernadsky State Geological Museum of Russian Academy of Sciences, Moscow;

²vitaer@gmail.com, ²naumova_new@mail.ru

Abstract

This article presents the concept and architecture of a multi-agent system designed to function as a digital assistant for geologist researchers. The system aims to automate key stages of scientific research: from topic formulation and literature review to hypothesis generation and presentation of results. The article describes the system's integration with the GeologyScience.ru platform, which provides access to diverse geological data and analysis tools, as well as approaches to adapting large-scale language models (LLM) to solve specialized scientific problems.

Keywords: *multi-agent system, artificial intelligence, digital assistant, LLM.*

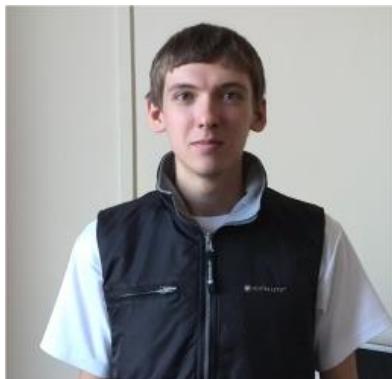
REFERENCES

1. KoBold used AI to find copper — now investors are piling in to the tune of \$537M. URL: <https://techcrunch.com/2025/01/02/kobold-used-ai-to-find-copper-now-investors-are-piling-in-to-the-tune-of-537m/>
2. Cifrovaja sistema «Gazprom nefti» nashla novye plasty nefti v JaNAO. URL: <https://www.gazprom-neft.ru/press-center/news/tsifrovaya-sistema-gazprom-nefti-nashla-novye-plasty-nefti-v-yanao/>
3. Gottweis J., Wei-Hung Weng, Daryin A. et al. Towards an AI co-scientist // arxiv.org. 2025. <https://doi.org/10.48550/arXiv.2502.18864>
4. Naumova V.V., Eremenko V.S., Zagumennov A.A., Patuk M.I. Nauchnyj portal Geologyscience.ru: tekushhee sostojanie i perspektivy razvitiya// Geoinformatika. 2023. № 3. S. 33–43.
5. Naumova V.V., Patuk M.I., Eremenko A.S., Zagumennov A.A., Eremenko V.S. Cifrovye tehnologii budushhego — sovremennoye reshenija v naukah o Zemle // Geoinformatika. 2024. № 3. S. 53–63.
6. Eremenko V.S., Naumova V.V., Zagumennov A.A. Razrabotka virtual'nyh prostranstv s intellektual'nym adaptivnym interfejsom dlja cifrovyh geologicheskikh sistem

// Informacionnye i matematicheskie tehnologii v nauke i upravlenii. 2024. № 1 (33).
P. 146–153.

7. DeepSeek 1R. URL: <https://www.deepseek.com>

СВЕДЕНИЯ ОБ АВТОРАХ



ЕРЕМЕНКО Виталий Сергеевич – научный сотрудник, Государственный геологический музей им. В.И. Вернадского РАН, Москва.

Vitaliy S. EREMENKO – researcher of the SGM scientific department, Vernadsky State Geological Museum RAS, Moscow.

email: vitaer@gmail.com

ORCID: 0000-0002-5250-5743



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва.

Vera V. NAUMOVA – Prof., head of the SGM scientific department, Vernadsky State Geological Museum RAS, Moscow.

email: naumova_new@mail.ru

ORCID: 0000-0002-3001-1638

Материал поступил в редакцию 31 августа 2025 года

УДК 004.85

ЦИФРОВЫЕ ТЕХНОЛОГИИ БУДУЩЕГО ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

В. В. Наумова¹ [0000-0002-3001-1638], М. И. Патук² [0000-0003-3036-2275],

А. С. Еременко³ [0000-0003-1923-8417], А. А. Загуменнов⁴ [0000-0002-0501-5362],

В. С. Ерёменко⁵ [0000-0002-5250-5743]

¹⁻⁵Государственный геологический музей им. В. И. Вернадского РАН,
г. Москва, Россия

^{3, 4}Институт автоматики и процессов управления Дальневосточного
отделения РАН, г. Владивосток, Россия

^{3, 4}Дальневосточный федеральный университет, г. Владивосток, Россия

¹naumova_new@mail.ru, ²patuk@mail.ru, ³academy21@gmail.com,

⁴truepikvic@gmail.com, ⁵vitaer@gmail.com

Аннотация

Рассмотрены технологии, способные радикально изменить развитие сразу многих областей: искусственный интеллект, квантовые технологии, большие данные, технологии беспроводной связи, системы распределенного реестра и др. Представлен ряд перспективных технологий ближайшего будущего, которые в настоящее время имеют перспективы применяться в науках о Земле. Дан обзор применения этих технологий для решения различных геологических задач, в том числе и результатов, полученных авторами.

Ключевые слова: искусственный интеллект, числовые инструменты, виртуальные ассистенты, интеллектуальные агенты, числовые технологии будущего в геологии.

ВВЕДЕНИЕ

Всемирный экономический форум (ВЭФ) с 2011 г. публикует ежегодный отчет с описанием десяти лучших инновационных технологий, которые потенциально могут стать революционными в течение нескольких лет. В июне 2024 г. ВЭФ опубликовал новый отчет, который основан на идеях более 300 ведущих

щих мировых ученых и исследователей – участников Сети советов по вопросам глобального будущего, а также сети Frontiers, с более чем двумя тысячами главных редакторов ведущих научных изданий.

По версии ВЭФ в 2024 г. выделены 10 следующих перспективных технологий [1].

1. **Искусственный интеллект (ИИ) для научных открытий.** Ожидается, что достижения в области глубокого обучения, генеративного ИИ и фундаментальных моделей революционизируют процесс научных открытий.

2. **Технологии повышения конфиденциальности.** Синтетические данные защищают персональную информацию и предоставляют новые возможности для глобального сотрудничества и обмена данными.

3. **Реконфигурируемые интеллектуальные поверхности.** Превращение обычных стен и других поверхностей в интеллектуальные компоненты для беспроводной связи повысит энергоэффективность сетей и откроет новые перспективы в таких областях, как интеллектуальные фабрики или транспорт.

4. **Высотные платформы** (High-Altitude Platform Station — HAPS, геостационарные атмосферные спутники). Применение самолетов, дирижаблей и воздушных шаров может расширить доступ к сетям мобильной связи в отдаленных регионах, помогая преодолеть цифровой разрыв.

5. **Интегрированное обнаружение и связь.** Сети 6G облегчат одновременный сбор данных (зондирование) и их передачу (связь), что позволит создавать системы мониторинга пространства, применимые в сельском хозяйстве, экологии, городском планировании.

6. **Иммерсивная технология для построенного мира.** Метавселенная и цифровые двойники с вычислительными мощностями, виртуальной и дополненной реальностью открывают новые перспективы для строительной отрасли.

7. **Эластокалорики.**

8. **Микроны, улавливающие углерод.**

9. **Альтернативные корма для животных.**

10. **Геномные технологии для трансплантаций.**

Рассмотрим построенный на этой основе ряд перспективных цифровых технологий, способных радикально изменить развитие геологических исследований, которые в настоящее время уже применяются или имеют перспективы применяться:

1. Искусственный интеллект;
2. Технологии беспроводной связи;
3. Системы распределенного реестра;
4. Большие данные и квантовые вычисления;
5. Виртуальная и дополненная реальность (VR и AR);
6. Цифровые инструменты и виртуальные ассистенты.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Выделяют семь направлений в области ИИ:

- компьютерное зрение;
- обработка естественного языка;
- распознавание и синтез речи;
- рекомендательные системы и интеллектуальные системы поддержки принятия решений;
- перспективные методы и технологии в ИИ;
- нейропротезирование;
- нейроинтерфейсы, нейростимуляция и нейросенсинг.

Хотя ИИ уже используется в научных исследованиях, достижения в области глубокого обучения, генеративного ИИ и базовых моделей коренным образом изменяют процесс научных открытий. Искусственный интеллект позволяет исследователям устанавливать беспрецедентные связи и добиваться прогресса в изучении геологических объектов и процессов.

Развитие методов ИИ, связанное с обработкой и генерацией текстов, открыло серьезные возможности по извлечению новых знаний из потока научной геологической информации, которую очень трудно, а зачастую и невозможно переработать традиционным методом чтения. С помощью методов обработки естественного языка решаются следующие задачи: выделение геоло-

гических и географических именованных сущностей, извлечение пространственных и временных взаимосвязей, классификация, кластеризация, реферирование геологических отчетов и публикаций.

Разработан широкий спектр алгоритмов машинного обучения, подходящих для выполнения специализированного геологического анализа. Машинное обучение можно использовать с целью выявления геологоразведочных объектов в условиях избытка данных, автоматического выявления геологических зон залегания полезных ископаемых, оценки твердости руды на основе результатов анализа и др.

Разработка методов ИИ, связанных с анализом изображений, в том числе и 3D-изображений, открывает новые возможности, связанные с анализом изображений пород и минералов.

Достаточно подробный анализ применения ИИ для решения геологических задач дан в статье [2].

ТЕХНОЛОГИИ БЕСПРОВОДНОЙ СВЯЗИ

Технология беспроводных сенсорных сетей позволяет совместить сенсоры, микропроцессор для сбора и первичной обработки данных, системы питания и радиосвязи в небольшом защищенном корпусе и объединить много таких устройств в беспроводную сеть сбора данных, которая способна к самонастройке и автономной работе на открытой местности в течение нескольких лет.

Сенсорные сети используются для мониторинга вулканов — объектов, требующих постоянного наблюдения и изучения. Такие сети — это распределенные, самоорганизующиеся сети множества датчиков (сенсоров), объединенных между собой посредством радиоканала. Цель — фиксировать физические параметры, связанные с геологическими процессами под поверхностью, и передавать данные в центры мониторинга.

В наше время цифровизация имеет тенденцию к подключению устройств беспроводных сенсорных сетей к так называемому интернету вещей (IoT — системе взаимосвязанных устройств, которые могут собирать данные, обрабатывать их и обмениваться ими между собой, с человеком и серверами)

для реального мониторинга процессов. Системы WSN и IoT основаны на новейших лазерных или ультразвуковых технологиях и могут проводить измерения с точностью до нескольких миллиметров или выше.

СИСТЕМЫ РАСПРЕДЕЛЕННОГО РЕЕСТРА

Системы распределенного реестра применяются в геологии для обеспечения безопасной, неизменяемой и прозрачной передачи данных между участниками геологоразведочных работ, добычи полезных ископаемых и мониторинга геотехнических объектов. Такие системы позволяют:

- хранить информацию в электронной системе БД, распределенной между несколькими сетевыми узлами или устройствами, без единого управляющего узла;
- синхронизировать данные между узлами с помощью автоматического алгоритма консенсуса: каждый узел самостоятельно обновляется, после чего все «голосуют» за обновление реестра и достигают согласия в отношении одной из его копий;
- защищать данные уникальной электронной подписью, что обеспечивает защищенность БД, а также ее прозрачность.

Отметим системы распределенного реестра, которые используются в геологии в РФ:

- В ФГИС «Единый фонд геологической информации о недрах» (ЕФГИ) — федеральная государственная информационная система, которая содержит реестр геологической информации, имеющейся в федеральном и территориальных фондах геологической информации, а также в организациях, находящихся в ведении органов государственной власти. Цель — информационное обеспечение управления государственным фондом недр, ведения государственного кадастра месторождений и проявлений полезных ископаемых, государственного баланса запасов полезных ископаемых, реестра работ по геологическому изучению недр.
- В Информационное пространство «Единый банк геолого-геофизической информации ВНИГНИ» — веб-ориентированная программная среда для организации распределенного хранения и доступа к геолого-геофизической информации, накопленной во ФГБУ «ВНИГНИ» [3].

КВАНТОВЫЕ ВЫЧИСЛЕНИЯ

Квантовые компьютеры смогут решать задачи, непосильные для классических систем, что открывает новые горизонты в науке, криптографии и оптимизации. Благодаря квантовым явлениям, квантовые компьютеры будут способны выполнять вычисления с беспрецедентной скоростью. Квантовая криптография обеспечит абсолютно безопасную передачу данных, защищенную от взлома.

Квантовые вычисления (квантовые алгоритмы) применяются в геологии для решения задач, которые требуют больших вычислительных ресурсов и занимают много времени на классических компьютерах. Это связано с использованием квантовых вычислителей («отжигателей») и квантовых сенсоров — устройств, которые фиксируют мельчайшие изменения в окружающей среде.

В 2024 г. Геологическая служба США (USGS) и компания Q-Ctrl объявили о заключении договора о сотрудничестве, цель которого состоит в использовании квантовых технологий для поиска полезных ископаемых [4]. Кроме того, стороны должны освоить новые методы прогнозирования стихийных бедствий. Одним из направлений сотрудничества станет применение технологий квантового зондирования для точного определения запасов минералов и воды. Стороны также намерены использовать другие сложные квантовые методы, такие как квантовая гравиметрия, квантовая магнитометрия и квантовая оптимизация. Ожидается, что это не только повысит эффективность разведывательных работ, но и снизит затраты на добычу ресурсов благодаря высокоточному планированию.

ВИРТУАЛЬНАЯ И ДОПОЛНЕННАЯ РЕАЛЬНОСТЬ

Технологии виртуальной (VR) и дополненной (AR) реальности применяются в геологии для визуализации геологических данных, моделирования и анализа структур. Эти технологии позволяют специалистам погружаться в геологические структуры, изучать их изнутри и получать более глубокое понимание строения недр.

Одним из применений виртуальной реальности в геологии являются визуализация и моделирование геологических данных. С помощью VR специалисты могут создавать трехмерные модели горных пород, разломов, пластов

и других геологических структур, что позволяет им лучше понять их структуру и взаимодействие.

Виртуальная реальность также используется для проведения геологических исследований в недоступных или опасных районах. Специалисты могут использовать VR, чтобы исследовать вулканы, карстовые ямы, пещеры и другие опасные места без риска для своей жизни.

Кроме того, VR позволяет создавать виртуальные пробы и моделировать геологические процессы, что помогает специалистам прогнозировать и анализировать различные явления, такие как землетрясения, извержения вулканов, наводнения и другие природные катастрофы.

Другим важным применением VR в геологии являются обучение и тренировка специалистов. Студенты и геологи смогут погрузиться в виртуальные геологические местности и изучать их без фактического присутствия на месте, что позволит им приобрести ценный опыт и знания, не выходя из аудитории или кабинета.

Таким образом, применение виртуальной реальности в геологии и исследованиях является важным инструментом, который помогает ученым и специалистам проводить более точные, эффективные и безопасные исследования, а также обучать будущих геологов.

ЦИФРОВЫЕ ИНСТРУМЕНТЫ И ВИРТУАЛЬНЫЕ АССИСТЕНТЫ

В геологии используются цифровые инструменты и виртуальные ассистенты на основе ИИ и виртуальной реальности. Эти технологии помогают изучать недра, анализировать данные и принимать решения на их основе. Среди них можно отметить:

- **Обнаружение месторождения железной руды в Западной Австралии.** Специалисты стартапа Earth AI с помощью алгоритмов искусственного интеллекта проанализировали архивные спутниковые снимки за последние 20 лет, данные геофизических исследований и результаты предыдущих разведочных работ. В результате затраты на поиск сократились на 30%, а время разведки — с 3 лет до 8 месяцев.

- **Выявление дополнительных запасов нефти в уже разрабатываемых пластах.** Компания «Газпром нефть» применила ИИ-алгоритмы, которые проанализировали исторические данные бурения и сейсмики за 15 лет. По оценкам экспертов, такие «цифровые месторождения» могут увеличить добычу на 5–15% без дополнительных капитальных вложений.

- **Обнаружение никелевых месторождений в Индонезии.** Спутник Phisat-2 с ИИ-алгоритмами обнаружил никелевые месторождения по повышенной температуре поверхности в районах древних вулканов.

- **Обнаружение крупных залежей фосфатов в Западной Африке.** Система на основе ИИ от EOS Data Analytics выявила участки с аномальным содержанием фосфора в почве по изменению ее отражательной способности.

- **Поиск кимберлитовых трубок — главных источников алмазов.** В 2023–2024 годах специалисты Института геологии алмаза (Якутск) успешно внедрили систему на основе искусственного интеллекта, которая анализирует комплекс данных: от спектральных характеристик растительности до магнитных аномалий.

В работе [5] проведен анализ приложений в геологии цифровых инструментов и виртуальных ассистентов.

Технологии будущего в геологии связаны с развитием цифровых решений, которые повышают эффективность и точность геологических исследований, открывают новые возможности для поиска и разведки полезных ископаемых. Главные тренды состоят в интеграции ИИ, IoT и Big Data, а также в развитии квантовых вычислений для сверхточного моделирования.

ЦИФРОВАЯ ПЛАТФОРМА ИНТЕГРАЦИИ И АНАЛИЗА ГЕОЛОГИЧЕСКИХ ДАННЫХ

Основной целью работ, которые выполняются нами в Государственном геологическом музее им. В. И. Вернадского РАН (ГГМ РАН), является разработка Единого цифрового пространства геологических знаний, которое рассматривается как некоторая перспектива создания, развития и интеграции разно типовых информационных ресурсов РФ и сервисов, в том числе осно-

ванных на моделях и технологиях ИИ, в современной цифровой интеллектуальной среде. Результаты этих работ доступны пользователям на Портале <https://geologyscience.ru> (рис. 1) [6].



Рис. 1. Портал GeologyScience.ru

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТЕКСТОВ ГЕОЛОГИЧЕСКИХ ПУБЛИКАЦИЙ С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Стремительный рост объема публикаций во всех областях геологических наук делает критически важным внедрение методов автоматизированной обработки научных текстов. Одним из наиболее перспективных инструментов для решения этой задачи выступают большие языковые модели на основе нейронных сетей. Наши работы по семантическому поиску публикаций с использованием дополнительно тренированных языковых моделей и расчету меры близости геологических текстов показали хорошие результаты. Но используемые модели оказались неспособны выполнить глубокий анализ текстов. Сравнительный анализ современных архитектур позволил нам выделить модель DeepSeek R1, относящуюся к классу систем с расширенными возможностями логического вывода. Проведено тестирование предложенного подхода для сравнительного анализа геологических публикаций. Тестирование показало результаты, вызывающие доверие.

Нами разработан ряд сервисов анализа геологических текстов с помощью больших нейросетевых языковых моделей (рис. 2) [7]. Они позволяют выполнять поиск на основе введенного запроса или анализировать предоставленные тексты, включая сравнение нескольких публикаций, том числе поиск семантически близких публикаций, определение косинусной близости двух текстов, сравнительного анализа текстов геологических публикаций.

The screenshot shows a web browser window with the URL <http://83.149.241.22/analyze/>. The main title is "Сервис анализа PDF файлов геологических публикаций с помощью нейросетевой языковой модели DeepSeek R1". Below it is a sub-section titled "Патук М.И. ГГМ РАН". A text box states: "Сервис предназначен для анализа научных публикаций геологической направленности на русском и английском языках." To the right is a button labeled "К сервису". Another text box says: "Для начала работы по анализу PDF файла щелкните по кнопке "Обзор..." и выберите файл, который вы хотите проанализировать. Если вам надо проанализировать только один файл, то слева, где написано "Введите свой вопрос..." напишите примерно следующее - "Проанализируй этот файл" и нажмите кнопку "Отправить" или клавишу Enter." Below this are three buttons: "Проанализируй этот файл", "Обзор...", and "Отправить". A large blue-bordered box contains the text "Ответ ИИ" and a snippet of analysis results: "Анализ статьи "Месторождение алмазов кимберлитовой трубы Мир: основные этапы изучения, особенности и результаты разведки" Цель исследования".

Рис. 2. Сервис анализа pdf-файлов геологических публикаций на портале GeologyScience.ru

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ ПЛАТФОРМЫ ПО РАБОТЕ СО СПУТНИКОВЫМИ И ПРОСТРАНСТВЕННЫМИ ДАННЫМИ В ВИДЕ МУЛЬТИАГЕНТНОЙ СИСТЕМЫ НА БАЗЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В ГГМ РАН осуществляется разработка информационно-вычислительного узла для поиска и обработки спутниковых и пространственных данных для решения задач геологии на основе программной платформы Open Data Cube AI [8], при этом реализуется интеграция технологий DataCube AI, GeoAI и LLM для поиска и обработки спутниковых данных.

В ходе работ были исследованы и апробированы технологии Spatio-Temporal Asset Catalog (STAC) и Cloud-Optimized GeoTIFF (COG) для доступа

к спутниковым данным посредством запросов на естественном языке. Для обработки спутниковых данных применялась технология Open Data Cube, позволяющая получить доступ к каждому пикселю спутниковых данных в потоковом режиме, что совместно с технологией COG позволило обрабатывать спутниковые данные без необходимости их предварительной загрузки целиком, а работать только с нужным географическим регионом, что существенно снизило нагрузку и повысило производительность, а также исключило промежуточное звено в виде хранения данных, что позволяет говорить о распределенной системе обработки. Большие языковые модели также показали состоятельность при формировании визуализации спутниковых данных в соответствии с решаемой задачей геолога-исследователя, что позволит динамически адаптировать интерфейс пользователя, составляя его из типовых элементов.

Полученные разработки в ближайшее время планируется развернуть в виде тестового прототипа на портале <https://geologyscience.ru>

ВИРТУАЛЬНЫЙ АССИСТЕНТ ГЕОЛОГА-ИССЛЕДОВАТЕЛЯ

Виртуальный ассистент геолога-исследователя разрабатывается нами для того, чтобы выступать в качестве полезного помощника геолога, способствуя ускорению процесса научных открытий [9]. Предполагается, что система будет представлять собой сложную мультиагентную систему ИИ, разработанную для отражения процесса рассуждений, лежащего в основе научного метода. Ассистент по искусенному интеллекту призван помочь раскрыть новые, оригинальные знания и наглядно сформулировать новые исследовательские гипотезы и предложения, основанные на предыдущих доказательствах и согласованные с предоставленными научными целями исследования и руководящими указаниями.

A	B	C	D
1	Depth	Quartz	K-Feldspar
2	3173.97	27.56	0.4
3	3183.11	42.92	0.64
4	3192.26	44.65	0.56
5	3201.4	42.65	0.38
6	3210.54	42.81	1.26
7	3219.69	40.13	1.1
8	3228.83	40.32	5.2
9	3237.98	36.02	0.08
10	3247.12	36.48	0.29
11	3256.26	32.51	0.41
12	3265.41	36.07	0
13	3274.55	38.02	0.79
14	3283.7	36.21	0.23
15	3292.84	39.91	1.22
16	3301.98	38.23	0.42
17	3311.13	39.69	0.51
18	3338.56	39.49	1.15
19	3347.7	35.87	0
20	3356.85	32.91	0.18
21	3365.99	30.02	0.29
22	3375.14	32.87	0
23	3384.28	37.41	0
24	3393.42	39.32	0
25	3402.57	34.75	1.38
26	3411.71	34.41	0.75
27	3420.86	31.81	0.77
28	3430	31.78	8.39
		1.74	8.21

Рис. 3. Рабочий кабинет пользователя на портале <https://geologyscience.ru>

МЕТАВСЕЛЕННАЯ «ИСТОРИЯ ЗЕМЛИ»

Популяризацию геологических знаний мы осуществляляем с помощью виртуальных пространств. Это цифровые среды, в которых пользователи могут взаимодействовать как по образовательным и научным мотивам, так и просто находить собеседников по интересам и общаться. Такие среды называют метавселеными. Чтобы реализовать возможность полного виртуального погружения в метавселенную, для создания трехмерного мира используются такие передовые технологии, как виртуальная реальность, 3D-реконструкция, искусственный интеллект и Интернет вещей.

Нами рассмотрена новая концепция популяризации научного знания и получения образовательного опыта через взаимодействие с виртуальным окружающим миром с привлечением ИИ-агентов, подключенных к накопленной базе научно-популярного знания. Таким образом, метавселенная — это постоянно действующее цифровое пространство, в котором пользователи могут взаимодействовать с виртуальными мирами и другими пользователями в реальном времени.

В статье [10] описан первый этап проектирования Метавселенной «История Земли».

ЗАКЛЮЧЕНИЕ

Предложены методы и разработаны цифровые технологии для аккумуляции больших объемов территориально распределенной геологической и музейной информации (Big Data), а также для веб-сервисов для обработки и анализа всех типов геологической информации, в том числе и интеллектуальных для анализа геологических публикаций с использованием больших языковых моделей ИИ. Предложена первая версия интеграции геологических данных и сервисов на базе ИИ. В основе интеграции лежит цифровое решение постановки геологических задач. На этом решении разработан прототип цифрового ассистента геолога-исследователя для геологических информационных систем. Результат имеет принципиальную новизну. Возможные сферы практического применения – это поддержка научных геологических исследований, решение задач производственной геологии, получение аналитических материалов органами государственной власти и др через разработанный нами портал ГГМ РАН <https://geologyscience.ru>.

БЛАГОДАРНОСТИ

Работа выполнена в рамках Государственной темы Государственного геологического музея им. В. И. Вернадского РАН (FMMR-2025-0001) Рег. №1022061000107-4-1.5.1 «Естественнонаучные музеи в едином цифровом пространстве геологических знаний».

СПИСОК ЛИТЕРАТУРЫ

1. Top 10 Emerging Technologies of 2024,
URL: <https://www.weforum.org/publications/top-10-emerging-technologies-2024/>
2. Наумова В.В., Патук М.И., Еременко А.С., Загумённов А.А., Ерёменко В.С. Цифровые технологии будущего — современные решения в науках о Земле // Геоинформатика. 2024. № 3. С. 53–63.
3. Марков К.Н., Жуков К.А., Конева А.А., Костылева Т.В. Распределенный банк геолого-геофизической информации ВНИГНИ как составная часть Единого фонда геологической информации отрасли// Геология нефти и газа. 2021. №3., С.67-76.

4. Q-CTRL Partners with USGS to Pioneer Quantum Sensing and Computing Applications. URL: <https://q-ctrl.com/blog/q-ctrl-partners-with-usgs-to-pioneer-quantum-sensing-and-computing-applications>
5. *Naumova V.V., Zagumennov A.A., Eremenko V.S.* Digital Tools and Virtual Assistants to Support Scientific Research in Geology // Automatic Documentation and Mathematical Linguistics, 2024, Vol. 58, Suppl. 5, P. S287–S291.
6. *Наумова В.В., Еременко В.С., Загуменнов А.А., Патук М.И.* Научный портал Geologyscience.ru: текущее состояние и перспективы развития// Геоинформатика. 2023. № 3. С. 33–43.
7. *Патук М.И., Наумова В.В.* Сравнительный анализ текстов геологических публикаций с использованием больших языковых моделей. // Цифровые технологии будущего — современные решения в науках о Земле ITES-2025: Материалы VII Всероссийской конференции с международным участием, Владивосток, 22–26 сентября 2025 г. / сост. А.А. Стрельцова. Владивосток: Издательство Дальневосточного федерального университета, 2025. С. 23
8. *Загуменнов А.А., Наумова В.В.* Сервисы на основе методов искусственного интеллекта для поддержки научных исследований в геологии // Материалы Международной конференции «Марчуковские научные чтения 2024» (МНЧ-2024), 7–11 октября 2024 г., Академгородок, Новосибирск, Россия. С. 125.
9. *Еременко В.С., Наумова В.В.* Мультиагентная система – AI виртуальный соавтор геолога-исследователя // Цифровые технологии будущего — современные решения в науках о Земле ITES-2025: Материалы VII Всероссийской конференции с международным участием, Владивосток, 22–26 сентября 2025 г. / сост. А.А. Стрельцова. Владивосток: Издательство Дальневосточного федерального университета, 2025. С. 22.
10. *Еременко А.С.* Метавселенная «История Земли» - новый взгляд на популяризацию геологических знаний // Электронные библиотеки. 2025. Т. 28, № 1. С. 1–21. <https://doi.org/10.26907/1562-5419-2025-28-1-2-21>

DIGITAL TECHNOLOGIES OF THE FUTURE FOR SCIENTIFIC RESEARCH IN GEOLOGY

V. V. Naumova¹ [0000-0002-3001-1638], M. A. Patuk² [0000-0003-3036-2275],
A. S. Eremenko³ [0000-0003-1923-84171], A. A. Zagumennov⁴ [0000-0002-0501-5362],
V. S. Eremenko⁵ [0000-0002-5250-5743]

¹⁻⁵*Vernadsky State Geological Museum of the Russian Academy of Sciences,
Moscow, Russia*

^{3, 4}*Institute of automatic and control processes of FEB RAS, Vladivostok, Russia*
^{3, 4}*Far East Federal University, Vladivostok, Russia*

¹naumova_new@mail.ru, ²patuk@mail.ru, ³academy21@gmail.com, ⁴truepik-vic@gmail.com, ⁵vitaer@gmail.com

Abstract

The article discusses technologies that can radically change the development of many areas at once: artificial intelligence, quantum technologies, big data, wireless communication technologies, distributed registry systems. The authors consider a number of promising technologies of the near future that currently have prospects for application in Earth sciences. The review of the application of these technologies to solve various geological problems, including the results obtained by the authors, is carried out.

Keywords: Artificial intelligence, numerical tools, virtual assistants, intelligent agents, numerical technologies of the future in geology.

REFERENCES

1. Top 10 Emerging Technologies of 2024.
URL: <https://www.weforum.org/publications/top-10-emerging-technologies-2024/>
2. Naumova V.V., Patuk M.I., Yeryomenko A.S., Zagumyonnov A.A., Yeryomenko V.S. Tsifrovie tekhnologii budushchego — sovremennie resheniya v naukakh o Zemle // Geoinformatika. 2024. № 3. S. 53–63.
3. Markov K.N., Zhukov K.A., Koneva A.A., Kostileva T.V. Raspredelennii bank geologo-geofizicheskoi informatsii VNIGNI kak sostavnaya chast Yedinogo

fonda geologicheskoi informatsii otrasi// Geologiya nefti i gaza. 2021. №3. S. 67-76.

4. Q-CTRL Partners with USGS to Pioneer Quantum Sensing and Computing Applications, URL: <https://q-ctrl.com/blog/q-ctrl-partners-with-usgs-to-pioneer-quantum-sensing-and-computing-applications>

5. Naumova V.V., Zagumennov A.A., Eremenko V.S. Digital Tools and Virtual Assistants to Support Scientific Research in Geology // Automatic Documentation and Mathematical Linguistics. 2024. Vol. 58, Suppl. 5. P. S287–S291.

6. Naumova V.V., Yeremenko V.S., Zagumennov A.A., Patuk M.I. Nauchnii portal Geologyscience.ru: tekushchee sostoyanie i perspektivi razvitiya // Geoinformatika. 2023. № 3. S. 33–43.

7. Patuk M.I., Naumova V.V. Sravnitelni analiz tekstov geologicheskikh publikatsii s is-polzovaniem bolshikh yazikovikh modelei // Tsifrovie tekhnologii budushchego — sovremennie resheniya v naukakh o Zemle ITES-2025: Materiali VII Vserossiiskoi konferentsii s mezhdunarodnim uchastiem, Vladivostok, 22–26 sentyabrya 2025 g. / sost. A.A. Streltsova. Vladivostok: Izdatelstvo Dalnevostochnogo federalnogo universiteta, 2025. S. 23

8. Zagumennov A.A., Naumova V.V. Servisi na osnove metodov iskusstvennogo intellekta dlya podderzhki nauchnikh issledovanii v geologii // Materiali Mezhdunarodnoi konferentsii «Marchukovskie nauchnie chteniya 2024» (MNCh-2024), 7–11 oktyabrya 2024 g., Akademgorodok, Novosibirsk, Rossiya. S. 125.

9. Eremenko V.S., Naumova V.V. Multiagentnaya sistema – AI virtualnii soavtor geologa-issledovatelya // Tsifrovie tekhnologii budushchego — sovremennie resheniya v naukakh o Zemle ITES-2025: Materiali VII Vserossiiskoi konferentsii s mezhdunarodnim uchastiem, Vladivostok, 22–26 sentyabrya 2025 g. / sost. A.A. Streltsova. Vladivostok: Izdatelstvo Dalnevostochnogo federalnogo universiteta, 2025. S. 22.

10. Eremenko A.S. Metavselennaya «Istoriya Zemli» – novii vzglyad na popularizatsiyu geologicheskikh znanii // Russian Digital Libraries Journal. 2025. Tom 28, № 1, S. 1–21. <https://doi.org/10.26907/1562-5419-2025-28-1-2-21>

СВЕДЕНИЯ ОБ АВТОРАХ



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва.

Vera V. NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow.

email: naumova_new@mail.ru,

ORCID: 0000-0002-3001-1638



ПАТУК Михаил Иванович – к. г.-м. н., н. с., научный отдел Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Michail I. PATUK – PhD, scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: patuk@mail.ru

ORCID: 0000-0003-3036-2275



ЕРЕМЕНКО Александр Сергеевич – программист (внештатный сотрудник), Государственный Геологический музей им. В.И. Вернадского РАН; старший научный сотрудник, кандидат технических наук, Институт автоматики и процессов управления ДВО РАН, Владивосток, Россия.

Alexander EREMENKO – contract programmer at the SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia); Ph.D, Senior researcher of Institute of automatic and control processes of FEB RAS, Vladivostok (Russia).

email: academy21@gmail.com

ORCID: 0000-0003-1923-8417

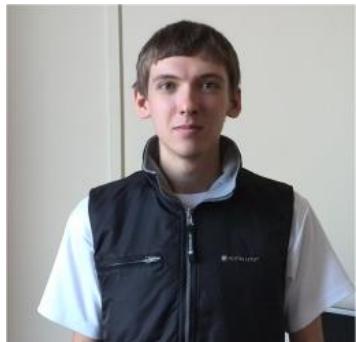


ЗАГУМЁННОВ Алексей Андреевич – программист (внештатный сотрудник), Государственный Геологический музей им. В.И. Вернадского РАН, Москва; младший научный сотрудник, Институт автоматики и процессов управления ДВО РАН, Владивосток.

Aleksey A. ZAGUMENNOV – contract programmer at the SGM scientific department, Vernadsky State Geological Museum RAS, Moscow; Junior researcher of Institute of automatic and control processes of FEB RAS, Vladivostok.

email: truepikvic@gmail.com

ORCID: 0000-0002-0501-5362



ЕРЕМЕНКО Виталий Сергеевич – научный сотрудник, Государственный геологический музей им. В.И. Вернадского РАН, Москва.

Vitaliy S. EREMENKO – researcher at the SGM scientific department, Vernadsky State Geological Museum RAS, Moscow.

email: vitaer@gmail.com

ORCID: 0000-0002-5250-5743

Материал поступил в редакцию 20 октября 2025 года

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТЕКСТОВ ГЕОЛОГИЧЕСКИХ ПУБЛИКАЦИЙ С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

М. И. Патук¹ [0000-0003-3036-2275], В. В. Наумова² [0000-0002-3001-1638]

^{1, 2}Государственный геологический музей им. В.И. Вернадского РАН, г. Москва, Россия

¹patuk@mail.ru, ²naumova_new@mail.ru

Аннотация

Стремительный рост объема публикаций во всех областях геологических наук делает критически важным внедрение методов автоматизированной обработки научных текстов. Одним из наиболее перспективных инструментов для решения этой задачи выступают большие языковые модели на основе нейронных сетей. Огромный прорыв в области искусственного интеллекта за последние годы превратил такие модели в незаменимых помощников для исследователей. Наши работы по семантическому поиску публикаций с использованием дополнительно тренированных языковых моделей и нахождения меры близости геологических текстов показали хорошие результаты. Но используемые модели оказались неспособны выполнить глубокий анализ текстов. Сравнительный анализ современных архитектур позволил нам выделить модель DeepSeek R1, относящуюся к классу систем с расширенными возможностями логического вывода. Данный тип моделей демонстрирует принципиально новый уровень качества генерации. На базе выбранной модели разработан веб-сервис, предоставляющий уникальный функционал, осуществляющий сравнительный анализ до 5 научных статей стандартного объема; поддержку мультиязычных источников (ввод текстов на английском, китайском, русском и др. языках); формирование структурированных отчетов на русском языке с выделением ключевых тезисов, противоречий и паттернов. Проведено тестирование предложенного подхода для сравнительного анализа геологических публикаций. Тестирование показало результаты, вызывающие доверие.

Ключевые слова: искусственный интеллект, большие языковые модели, обработка естественного языка, анализ текстов, геология.

ВВЕДЕНИЕ

Стремительный рост объема публикаций во всех областях геологических наук является серьезным вызовом для исследователей, что заставляет уделять все больше времени изучению литературы. Проблема дополнительного инструмента, позволяющего автоматизировать процесс анализа периодических научных изданий, давно назрела. По нашему мнению, таким инструментом могут являться языковые модели и, особенно, большие языковые модели [1].

Наши работы в области анализа научных публикаций начинались с самых простых задач – бинарной классификации и реферирования текстов [2]. Были решены задачи классификации месторождений золота и железа по аннотациям статей и выделения ключевых слов из них. Эти задачи решались с помощью дополнительно тренированных языковых моделей относительно небольшого объема.

Следующей задачей, которая являлась более практически значимой, была задача семантического поиска [3]. Для ее решения была выполнена дополнительная тренировка другой языковой модели, которая была предназначена для нахождения семантической близости документов.

На основе введенного поискового запроса происходит поиск наиболее семантически близких публикаций в репозитории Государственного геологического музея (ГГМ) им. В. И. Вернадского РАН [4]. Поиск близости ведется по названиям и аннотациям статей.

С помощью этой же языковой модели была решена еще одна задача – поиск близости аннотаций двух статей геологической направленности. В качестве параметра близости рассчитывалась косинусная близость двух текстов. Но числовой параметр близости мало информативен, поэтому было решено найти более развернутую оценку схожести и различия двух текстов. Данная задача также была реализована, и мы получили текстовое описание сходства и различия двух текстов (рис. 1).

На рис. приведен фрагмент описания различий двух месторождений алмазов – трубы «Мир» [5] и трубы «Аргайл» [6] (Австралия). Обращают на себя внимание некорректные утверждения типа: «тип месторождения не указан в тексте», «текст не содержит деталей геологии», «в тексте не описываются конкретные характеристики». Данный недостаток является очевидным, т. к. мы использовали

для анализа аннотации статей, а не полные тексты. С другой стороны, мы были ограничены параметрами используемых языковых моделей, контекстное окно (количество текста для анализа) которых ограничено 512 токенами (около 200–250 слов русского языка, в английском немного более).

Различия:

1. Тип месторождения и геология:

- **Мир:** Четко указан тип - **кимберлитовая трубка**. Текст фокусируется на **изменениях геологического строения с глубиной** (увеличение объема рудного тела >1500 м, стабильность состава).
- **Аргайл:** **Тип месторождения не указан в тексте** (известно, что это **лампроитовая трубка**, что и было вызовом традициям). Текст **не содержит деталей геологии**, акцент на историю добычи и характеристики продукции.

2. Характер продукции (Алмазы):

- **Мир:** **В тексте не описываются конкретные характеристики** добываемых алмазов (размер, цвет, качество). Упор на объем руды и запасы алмазов в каратах.

Рис. 1. Текстовое описание различий аннотаций двух статей.

Стало очевидно, что для анализа полного текста статей нужно переходить на языковые модели другого класса – большие языковые модели. Модели предыдущего поколения, которые мы использовали, могли обрабатывать небольшие объемы текста и были ориентированы на узкий круг задач обработки естественного языка. Это было плюсом с точки зрения адаптации их под конкретные задачи пользователя. Можно было дообучать их на своих наборах данных, размещать на локальных серверах. Но, как мы показали выше, для них есть свои естественные ограничения.

АНАЛИЗ ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ

Большие языковые модели (LLM) стали очень бурно развиваться с начала 2020-х годов. Особенно большой резонанс вызвал выход в 2022 г. первой версии модели ChatGPT [7]. Данной модели и ее использованию в научной деятельности посвящено огромное количество публикаций. Большая часть исследователей в академических кругах обратила внимание в первую очередь на генеративные свойства этой модели.

ChatGPT может создавать тексты на заданные темы и неплохо владеет научной терминологией. Исследование 45000 статей, вышедших за 3 месяца после выхода ChatGPT, показало, что около 10% статей использовали его в той или иной степени [8]. Большинство авторов сходится во мнении, что ChatGPT может быть полезен в автоматизации рутинных задач, таких как: анализ текстов, перевод, классификация, обзор литературы [9]. Хотя по поводу обзоров литературы не все авторы разделяют оптимизм [10, 11]. Сравнение результатов ChatGPT с реальными исследованиями показало существенную неполноту искусственного интеллекта (ИИ) и ограниченность выборки материалов датой подготовки материалов для обучения модели. Особенно неприятным явлением стали так называемые «галлюцинации» [12], вымышленные ответы системы, не имеющие отношения к действительности.

Со временем появилось понимание того что эти модели, по крайней мере в их текущем состоянии, надо использовать как помощников в исследованиях [13]; выполнять с их помощью трудоемкие, но рутинные задачи. При этом от пользователя таких инструментов требуется глубокое понимание предметной области, чтобы направлять ИИ и проверять его выводы.

Вместе с тем не все исследователи признают первенство ChatGPT над другими моделями. Там, где требуется глубокое знание предметной области, дополнительно обученные модели предыдущего поколения, типа BERT, способны показывать лучшие результаты [14]. Авторы решали традиционную для обработки естественного языка задачу классификации кернов скважин по типам осадочных пород на основании текстового описания этих кернов. При этом была поставлена довольно сложная задача – кроме основного литологического типа выделять еще

два дополнительных. Их дополнительно обученная модель показала результаты классификации в среднем на 11% лучше, чем GPT-4.

МЕТОДИКА РАБОТЫ

В настоящее время существует большое количество различных больших языковых моделей [15]. Но, к сожалению, многие из них нам недоступны в силу различных причин. Мы выбрали для анализа три модели: T-lite-it-1.0-Q8_0-GGUF [16] от Т-банка, GigaChat [17] от Сбера и китайскую модель DeepSeek [18]. Первую модель можно было запускать локально, две другие вызывались по API.

С учетом нашей задачи – сравнительного анализа геологических текстов – мы построили методику анализа работы указанных моделей.

Все современные большие языковые модели являются многоязычными и поддерживают перевод текстов. Были использованы несколько пар текстов описания различных месторождений на русском и английском языках. Указанные тексты подавались на вход моделям в режиме чат-бота с основным вопросом: описать сходство и различие полученных текстов. После получения ответа он анализировался на соответствие действительности. Кроме того, задавались различные вопросы с целью проверки качества запоминания каждого отдельного текста. Например, если в качестве сравниваемых текстов выступали описания месторождений золота и железа, мы просили модель указать, в каком из текстов содержится описание минерала «гематит» (по факту в описании месторождения железа) или описание минерала «арсенопирит» (по факту в описании месторождения золота).

По результатам этих тестов модель от Т-банка показала худшие результаты. Она плохо запоминала каждый отдельный текст и показывала большое количество «галлюцинаций», т. е. выдуманных фактов. Модель от Сбера оказалась немного лучше. Но безусловным лидером оказалась модель DeepSeek R1. Она относится к классу систем с расширенными возможностями логического вывода. Перед каждым выводом модель демонстрирует цепочку своих рассуждений по теме поставленного вопроса. Она всегда четко запоминала каждый отдельный текст, не путалась в содержании текстов и выдавала четкий структурированный ответ. Данная модель была использована для создания веб-сервиса по сравнению текстов геологических статей.

ВЕБ-СЕРВИС

Указанный веб-сервис вместе с другими сервисами анализа текстов расположен на портале GeologyScience.ru Государственного геологического музея им. В. И. Вернадского РАН. Он доступен через пункты меню «Сервисы» – «ИИ-анализ геологических текстов» или поциальному адресу <http://83.149.241.22/> (рис. 2).

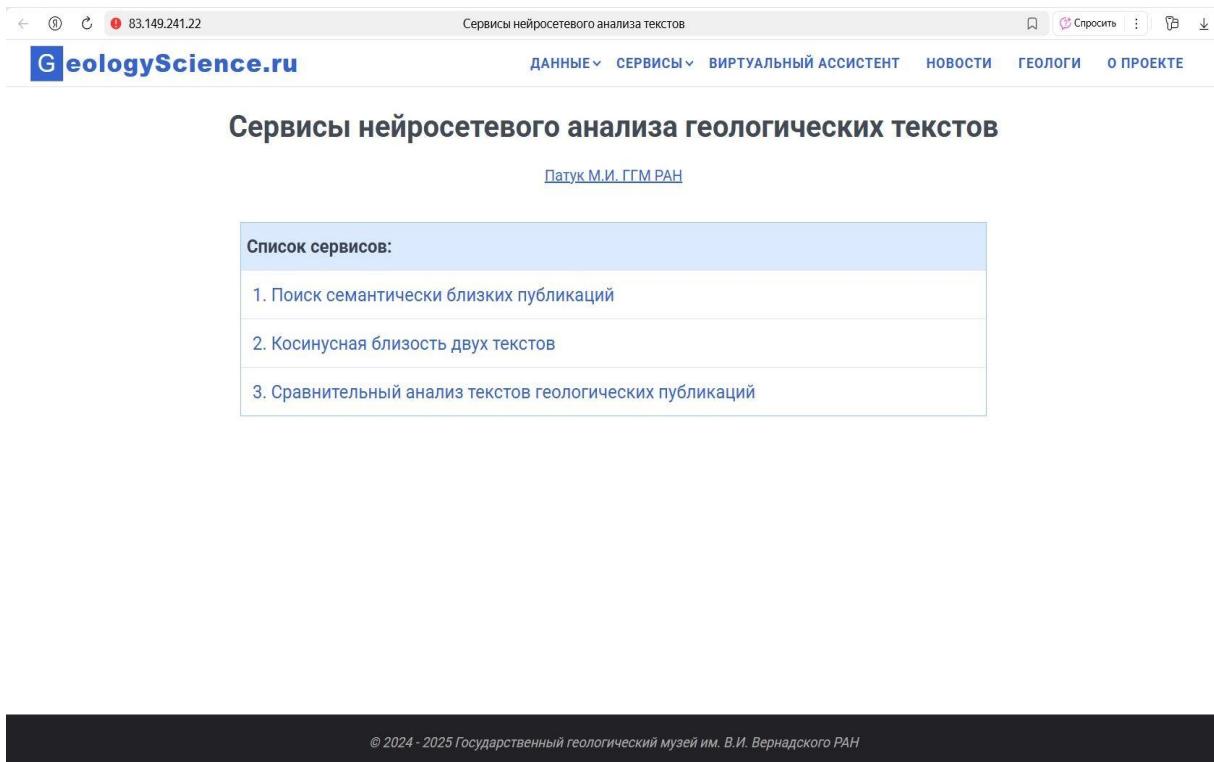


Рис. 2. Сервисы нейросетевого анализа текстов.

Сервис «Сравнительный анализ текстов геологических публикаций» принимает на вход статьи в формате PDF, которые должны сопровождаться основным вопросом типа: «Сравни эти два текста». Кроме того, можно задавать любое количество дополнительных вопросов: «В чем различие геологических описаний двух приведенных месторождений», «Какой возраст вмещающих пород первого месторождения» и т. п. Полученный PDF-файл преобразуется в текст и вместе с вопросом пользователя по API отправляется на сервер DeepSeek. Полученный ответ демонстрируется пользователю в окне сервиса.

Нами были повторно проанализированы две статьи по месторождениям алмазов – трубка «Мир» [5] и трубка «Аргайл» [6] (Австралия), которые не совсем удачно мы пытались анализировать с помощью своего предыдущего сервиса.

При этом вместо аннотаций анализу были подвергнуты полные тексты статей. Результат представлен ниже.

**Сравнение двух геологических текстов с помощью нейросетевой модели
DeepSeek R1**

**Сравнительный анализ алмазных месторождений: кимберлитовая трубка
«Мир» (Россия) и месторождение Аргайл (Австралия)**

1. Геологические особенности

Критерий	Трубка «Мир»	Аргайл
Тип породы	Кимберлитовая трубка	Лампроитовая трубка (первое крупное лампроитовое месторождение)
Возраст	Возраст кимберлитов: ~360 млн лет	Возраст лампроита: ~1.2 млрд лет; алмазы: ~1.58 млрд лет
Морфология	Удлиненно-овальная форма на глубине, расширение объема рудного тела с глубиной	Трубка в форме «головастика», наклонена на 30°, длина 2000 м
Геодинамическая обстановка	В пределах Якутской алмазонной провинции, на краю Сибирского кратона	В зоне Холлс-Крик (окраина Кимберлийского кратона), тектонически активная зона

Ключевое отличие: Аргайл изменил парадигму геологии алмазов, доказав, что они могут формироваться в лампроитах, а не только в кимберлитах.

2. Объемы добычи и запасы

Критерий	Трубка «Мир»	Аргайл
Пиковая добыча	Нет точных данных, но за 40 лет карьером извлечено 350 млн м ³ породы	42.8 млн карат в 1994 г. (40% мировой добычи)
Средняя годовая добыча	На подземном руднике (после 2009 г.) — 1 млн т руды в год	~25–30 млн карат/год в 1990-х
Запасы	+60 млн карат после разведки глубоких горизонтов (общие запасы >60 млн карат)	61 млн т руды с содержанием 2.9 карат/т (2000 г.)
Содержание алмазов	~3–6 карат/т	3 карата/т (в 3–10 раз выше типичных кимберлитов)

3. Технологии добычи и переработки

Критерий	Трубка «Мир»	Аргайл
Способ добычи	Карьер (до 525 м) → подземный рудник «Мир-Глубокий» после 2017 г.	Открытая добыча (планировался переход на подземную к 2006 г.)
Методы разведки	Наклонно-направленное бурение, керновое опробование, ГИС	Бурение сетки скважин, тяжело средная сепарация, рентгеновская сортировка
Переработка	Механическое дробление, гравитационное обогащение	Многоступенчатое дробление, тяжело средная сепарация, рентгеновские сортеры

Инновации: Аргайл разработал уникальные рентгеновские сортеры для обработки мелких алмазов (до 200 камней/сек). В «Мире» акцент на разведку глубоких горизонтов с использованием телеметрии.

4. Маркетинг и продукция

Критерий	Трубка «Мир»	Аргайл
Основная продукция	Бесцветные и слабо-окрашенные алмазы	Мелкие коричневые/желтые (72%), редкие розовые (<1%)
Маркетинговые стратегии	Фокус на промышленное значение, увеличение запасов	Продвижение «шампанских» и «коночных» алмазов (шкала C1–C7), аукционы розовых алмазов
Рынки сбыта	Внутренний (Россия), экспорт через АЛРОСА	Индия (гранка), глобальные аукционы (розовые алмазы)

Особенность: Аргайл создал новый сегмент рынка для цветных алмазов, тогда как «Мир» ориентирован на массовую добычу.

5. Экологические и социальные аспекты

Критерий	Трубка «Мир»	Аргайл
Экологические меры	Управление водными ресурсами после затопления 2017 г., «мокрая консервация»	Сертификация ISO 14001, рекультивация земель, создание озера из карьера
Социальное влияние	Основа алмазной промышленности России, создание рабочих мест в Якутии	Занятость 800+ человек, взаимодействие с аборигенами, вахтовый метод работы
Проблемы	Затопление шахты в 2017 г., необходимость перехода на подземную добычу	Высокая себестоимость при углублении карьера, удаленность месторождения

6. Проблемы и решения

- «Мир»:
 - **Проблема:** Катастрофическое затопление в 2017 г. остановило добычу.
 - **Решение:** Проект «Мир-Глубокий» для подземной разработки, интенсификация геологоразведки.

- **Аргайл:**
 - *Проблема:* Низкая рентабельность мелких алмазов.
 - *Решение:* Партнерство с Индией для огранки, создание премиального сегмента (розовые алмазы).
-

Выводы

- **Геология:** Аргайл расширил понимание алмазоносных пород, доказав роль лампроитов. «Мир» остается эталоном кимберлитовых месторождений.
- **Экономика:** Аргайл доминировал по объемам в 1990-х, но «Мир» обеспечивает долгосрочные запасы для России.
- **Инновации:** Оба месторождения используют современные технологии, но Аргайл преуспел в маркетинге, а «Мир» – в решении инженерных вызовов.
- **Устойчивость:** Аргайл лидирует в экологических практиках, «Мир» фокусируется на восстановлении после аварий.

Оба месторождения стали символами национальной гордости и драйверами развития своих регионов, демонстрируя разные пути адаптации к геологическим и рыночным вызовам.

ЗАКЛЮЧЕНИЕ

Как можно видеть, языковая модель DeepSeek R1 подробно по 6 основным пунктам выполнила анализ двух статей. Поскольку одна из статей была на английском языке, моделью был выполнен корректный перевод, что наглядно видно в приведенном анализе. Хотя следует отметить, что в процессе наших экспериментов модель иногда оставляла некоторые термины на английском языке или даже, очень редко, на китайском. Отметим, что мы не указывали, по каким пунктам следует выполнять сравнение. Они были установлены моделью на основе анализа текстов. Кроме того, были проверены все числовые данные: даты, геологический возраст, размеры карьеров, объемы добычи. Все цифры строго соответствуют тексту. Галлюцинации отсутствуют.

Размер контекстного окна модели DeepSeek равен 128 k токенов (60–65 тыс. слов русского языка). Это позволяет выполнять совместный анализ до 5 файлов

среднего размера (~10 страниц текста). Это будет предметом наших дальнейших исследований.

Применение модели DeepSeek для решения задач, с которых мы начинали свои работы по анализу текстов, бинарной классификации и извлечения ключевых слов, показало хорошие результаты, сопоставимые или превосходящие наши первые опыты. При этом эта модель не требовала дополнительно тренировки в предметной области для решения таких задач, поскольку перечисленные задачи были достаточно простыми.

Но обратим еще раз внимание на работу [14], в которой авторы показали превосходство дополнительно тренированной специализированной модели в задаче классификации осадочных пород над ChatGPT. И вслед за авторами этой работы повторим их вывод: если вы незнакомы с машинным обучением или у вас нет доступа к соответствующим техническим средствам, то, конечно, следует пользоваться такими инструментами, как ChatGPT или DeepSeek. Но специализированные модели в современных условиях могут дать гораздо лучший результат.

Кроме того, отметим специфику нашей задачи сравнительного анализа текстов для языковых моделей. Мы ограничиваем свободу генеративных возможностей модели текстами, предоставленными для анализа. Полученные выводы легко проверяются и в случае неверных ответов, как это имело место на этапе тестирования, легко заметить ошибки. Разумеется, для более глубокого анализа результатов работы больших языковых моделей по анализу геологических текстов требуется проверка ответов моделей со стороны специалистов в предметной области. Мы надеемся, что наша работа послужит импульсом для специалистов геологического профиля активнее применять подобные инструменты.

Наши тестовые результаты показывают, что современные большие языковые модели являются хорошим и доступным инструментом для анализа текстов научных публикаций. Но, как и любой инструмент, он требует дальнейшей проверки от заинтересованных специалистов.

Благодарности

Работа выполнена в рамках Государственной темы Государственного геологического музея им. В. И. Вернадского РАН (FMMR-2025-0001) Рег.

№1022061000107-4-1.5.1 «Естественно научные музеи в едином цифровом пространстве геологических знаний».

СПИСОК ЛИТЕРАТУРЫ

1. Большая языковая модель

URL: https://ru.wikipedia.org/wiki/Большая_языковая_модель (дата обращения 01.10.2025)

2. Патук М.И., Наумова В.В. Методы искусственного интеллекта для научных исследований в геологии // Электронные библиотеки. 2023. Т. 26, №5. С. 673–696. <https://doi.org/10.26907/1562-5419-2023-26-5-673-696>

3. Патук М.И., Наумова В.В. Использование семантического поиска для выбора и ранжирования научных геологических публикаций // Электронные библиотеки. 2024. Т. 27, №5. С. 758–773.

<https://doi.org/10.26907/1562-5419-2024-27-5-758-773>

4. Патук М.И., Наумова В.В., Еременко В.С. Цифровой репозиторий "geologyscience.ru": открытый доступ к научным публикациям по геологии России // Электронные библиотеки. 2020. Т. 23, № 6. С. 1324–1338.

5. Килижеков О.К., Толстов А.В., Яхин Ш.М., Зырянов И.В. Месторождение алмазов кимберлитовой трубки Мир: основные этапы изучения, особенности и результаты разведки. Горная промышленность. 2025. № 1. С. 49–56.

<https://doi.org/10.30686/1609-9192-2025-1-49-56>

6. Shigley J., Chapman J., Ellison R. Discovery and Mining of the Argyle Diamond Deposit, Australia // Gems and Gemology. 2001. Vol. 37. P. 26–41. <https://doi.org/10.5741/GEMS.37.1.26>

7. ChatGPT. <https://ru.wikipedia.org/wiki/ChatGPT> (дата обращения 01.10.2025)

8. Picazo-Sanchez P., Ortiz-Martin L. Analysing the impact of ChatGPT in research // Applied Intelligence. 2024. Vol. 54. P. 4172–4188.

<https://doi.org/10.1007/s10489-024-05298-0>

9. Islam I., Islam M.N. Exploring the opportunities and challenges of ChatGPT in academia // Discover Education. 2024. Vol. 3. Article no. 31. <https://doi.org/10.1007/s44217-024-00114-w>

10. *Faiza Farhat F., Sohail Sh. S., Dag Øivind Madsen D.Ø.* How trustworthy is ChatGPT? The case of bibliometric analyses // Cogent Engineering. 2023. Vol. 10. Article no. 2222988. <https://doi.org/10.1080/23311916.2023.2222988>
 11. *Захарина И.М.* Подготовка научной статьи: справится ли ChatGPT? // Высшее образование в России. 2023. Т. 32. № 8-9. С. 24–47.
<https://doi.org/10.31992/0869-3617-2023-32-8-9-24-47>
 12. Галлюцинация (искусственный интеллект),
URL: https://ru.wikipedia.org/wiki/%D0%93%D0%BB%D0%BE%D1%84%D0%BE%D0%BD%D0%BA%D0%B0%D0%BD%D1%8F%D1%82%D0%BD%D1%8B%D0%8E_%D0%80%D0%84%D0%80%D0%8F%D0%82%D0%8C%D0%BD%D0%BE%D0%BC
(дата обращения 01.10.2025)
 13. *Salvagno M., Taccone F.S., Gerli A.G.* Can artificial intelligence help for scientific writing? // Critical Care. 2023. Vol. 27. Article no. 75.
<https://doi.org/10.1186/s13054-023-04380-2>
 14. *Ghorbanfekr H., Kerstens P.J., Dirix K.* Classification of geological borehole descriptions using a domain adapted large language model // Applied Computing and Geosciences. 2025. Vol. 25. Article no. 100229.
 15. LLM Leaderboard, <https://artificialanalysis.ai/leaderboards/models> (дата обращения 01.10.2025)
 16. T-lite, https://huggingface.co/t-tech/T-lite-it-1.0-Q8_0-GGUF (дата обращения 01.10.2025)
 17. GigaChat. <https://giga.chat/> (дата обращения 01.10.2025)
 18. DeepSeek. <https://www.deepseek.com/en> (дата обращения 01.10.2025)

COMPARATIVE ANALYSIS OF GEOLOGICAL TEXTS USING LARGE LANGUAGE MODELS

Mikhail I. Patuk¹ [0000-0003-3036-2275], Vera V. Naumova² [0000-0002-3001-1638]

^{1, 2} Vernadsky State Geological Museum of the Russian Academy of Sciences, Moscow, Russia

¹patuk@mail.ru; ²Naumova_new@mail.ru

Abstract

The rapid increase in the volume of publications in various fields of geology makes it crucial to introduce methods for automated processing of scientific texts. Large language models based on neural networks represent one of the most promising approaches to solving this challenge. The recent breakthroughs in artificial intelligence have made such models indispensable tools for researchers. Our work on semantic search for publications using additionally trained language models and measuring the similarity between geological texts yielded good results. However, the models we used were unable to perform in-depth text analysis. A comparative analysis of modern architectures identified the DeepSeek R1 model as belonging to a class of systems with advanced logical inference abilities. This type of model represents a fundamentally new level of quality in text generation. Based on the chosen model, we have developed a web service that provides unique functionality for comparative analysis of up to 5 scientific articles. The service supports multilingual sources, allowing users to input text in English, Chinese, Russian, etc. It generates structured reports in Russian, highlighting key theses, contradictions, and patterns. The proposed approach has been tested on geological publications, and the results have been promising.

Keywords: artificial intelligence, large language models, natural language processing, texts analysis, geology.

REFERENCES

1. Large language model.

https://en.wikipedia.org/wiki/Large_language_model?ysclid=mg7ip9ev9d289421479
(date of access 01.10.2025)

2. *Patuk M.I., Naumova V.V.* Artificial Intelligence Methods for Scientific Research in Geology // *Russian Digital Libraries Journal*. 2023. Vol. 26, No. 5. P. 673–696. (In Russ.). <https://doi.org/10.26907/1562-5419-2023-26-5-673-696>
3. *Patuk M.I., Naumova V.V.* Using Semantic Search to Select and Rank Geological Publications // *Automatic Documentation and Mathematical Linguistics*. 2024. Vol. 58, Suppl. 5. P. S294–S298. <https://doi.org/10.3103/S0005105525700372>
4. *Patuk M.I., Naumova V.V., Eryomenko V.S.* Digital repository "geology-science.ru": open access to scientific publications on russian geology // *Russian Digital Library Journal*. 2020. Vol. 23, No. 6. P. 1324–1338 (in Russian).
5. *Kilizhekov O.K., Tolstov A.V., Yakhin Sh.M., Zyryanov I.V.* Diamond deposit of the Mir kimberlite pipe: main research stages, specific features and results of exploration // *Russian Mining Industry*. 2025. No. 1. P. 49–56 (In Russ.).
<https://doi.org/10.30686/1609-9192-2025-1-49-56>
6. *Shigley J., Chapman J., Ellison R.* Discovery and Mining of the Argyle Diamond Deposit, Australia // *Gems and Gemology*. 2001. Vol. 37. P. 26–41.
<https://doi.org/10.5741/GEMS.37.1.26>
7. ChatGPT.
URL: <https://en.wikipedia.org/wiki/ChatGPT> (date of access 01.10.2025)
8. *Picazo-Sanchez P., Ortiz-Martin L.* Analysing the impact of ChatGPT in research // *Applied Intelligence*. 2024. Vol. 54. P. 4172–4188.
<https://doi.org/10.1007/s10489-024-05298-0>
9. *Islam I., Islam M.N.* Exploring the opportunities and challenges of ChatGPT in academia // *Discover Education*. 2024. Vol. 3. Article no. 31.
<https://doi.org/10.1007/s44217-024-00114-w>
10. *Faiza Farhat F., Sohail Sh. S., Dag Øivind Madsen D.Ø.* How trustworthy is ChatGPT? The case of bibliometric analyses // *Cogent Engineering*. 2023. Vol. 10. Article no. 2222988. <https://doi.org/10.1080/23311916.2023.2222988>
11. *Zashikhina I.M.* Scientific Article Writing: Will ChatGPT Help? Vysshee obrazovanie v Rossii // *Higher Education in Russia*. 2023. Vol. 32, no. 8. P. 24–47.
<https://doi.org/10.31992/0869-3617-2023-32-8-9-24-47> (In Russ., abstract in Eng.)
12. Hallucination (artificial intelligence). URL: [https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)) (date of access 01.10.2025)

13. *Salvagno M., Taccone F.S., Gerli A.G.* Can artificial intelligence help for scientific writing? // Critical Care. 2023. Vol. 27. Article no. 75.
<https://doi.org/10.1186/s13054-023-04380-2>
 14. *Hossein Ghorbanfekr, Pieter Jan Kerstens, Katrijn Dirix.* Classification of geological borehole descriptions using a domain adapted large language model // Applied Computing and Geosciences. 2025. Vol. 25. Article no. 100229.
 15. LLM Leaderboard.
<https://artificialanalysis.ai/leaderboards/models> (date of access 01.10.2025)
 16. T-lite. https://huggingface.co/t-tech/T-lite-it-1.0-Q8_0-GGUF (date of access 01.10.2025)
 17. GigaChat. <https://giga.chat/> (date of access 01.10.2025)
 18. DeepSeek. <https://www.deepseek.com/en> (date of access 01.10.2025)
-

СВЕДЕНИЯ ОБ АВТОРАХ



ПАТУК Михаил Иванович – к. г.-м. н., н. с., научный отдел Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Michail I. PATUK – PhD, scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: patuk@mail.ru

ORCID: 0000-0003-3036-2275



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Vera V. NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: naumova_new@mail.ru

ORCID: 0000-0002-3001-1638

Материал поступил в редакцию 30 августа 2025 года

УДК 550.8.05

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ГЕОЛОГИИ: СОСТОЯНИЕ И ПЕРСПЕКТИВЫ

Л. Е. Чесалов

ФГБУ «Гидроспецгеология», г. Москва, Россия

chesalov@specgeo.ru

Аннотация

Рассмотрен опыт развития цифровых технологий в геологии и добывающих отраслях, включая государственные и частные структуры, промышленные, научные и образовательные учреждения. Приведены результаты использования цифровых технологий для широкого спектра исследований в области геологоразведки и добычи, рассмотрены преимущества и недостатки. Предложены подходы, способные обеспечить качественный прирост знаний и информации в области геологии.

Ключевые слова: цифровые технологии, геология, добывающие отрасли, информационные системы.

Цифровые технологии в геологии развиваются более 60 лет. Использование цифровых технологий в геологии и недропользовании за последние годы находится под влиянием двух разнонаправленных факторов – взрывной интерес к технологиям искусственного интеллекта (ИИ) и снижение конъюнктуры минерального сырья и энергетических ресурсов. Следует отметить, что снижение конъюнктуры сырья ведет к поиску более дешевых методов геологического изучения, что также усиливает интерес к цифровым технологиям прогнозирования, поисков разведки и разработки месторождений.

Внедрение цифровых технологий на раннем этапе (70–80-е годы XX в.), как правило, не давало прямого экономического эффекта: затраты на информационную инфраструктуру не окупались. Однако появлялись новые возможности, особенно в области сбора, хранения и обработки геологической информации, что стимулировало разработку алгоритмов и программ [1]. Современный этап начался в конце 80-х – начале 90-х XX в. и связан с резким удешевлением компо-

нентов информационной инфраструктуры (IBM-совместимые персональные компьютеры). За прошедшие 35 лет практически исчезли из массового сегмента разнообразные RISC-процессоры и архитектуры, несовместимые с Intel. Развитие открытого программного обеспечения обеспечило доступность алгоритмов обработки и интерпретации, а относительная низкая эффективность их реализации компенсировалась вычислительной мощностью. Рост возможностей и конкуренция технологий ИИ (OpenAI, DeepSeek и др.) за последнее время также создают уникальные возможности для развития цифровых технологий в прикладных областях.

Использование цифровых технологий в геологии и недропользовании можно условно разделить на следующие три направления.

1. *Традиционное (условно-пассивное)*, связанное со сбором и хранением геологической информации. В его рамках развиваются автоматизация сбора, повышение эффективности хранения и доступа, стандартизация и технологии обмена данными. Именно это направление создает основу (в т. ч. пресловутые «датасеты») для дальнейшего использования и развития. Резкое удешевление твердотельных накопителей существенно расширило возможности по хранению данных.

2. *Традиционное (условно-активное)*, связанное с промышленными аналитическими решениями по обработке геологической, геофизической и геохимической информации. В рамках этого направления решаются задачи информационной поддержки разработки месторождений и переработки сырья. Многие реализуемые алгоритмы являются вычислительно-емкими. К сожалению, иллюзия бесконечного роста вычислительной мощности (эмпирический «закон Мура», действовавший более 50 лет) отодвинула на второй план поиск более эффективных алгоритмов обработки.

3. *Инновационные решения*, направленные как на решение новых задач, так и на повышение эффективности традиционных методов обработки. Эти решения активно развиваются во всех областях геологии, однако есть области, вызывающие больший интерес. Всеобщий бум машинного обучения ведет к пере-

оценке качества получаемых решений, спорным подходам к формированию датасетов для обучения и часто к игнорированию физических основ методов и измерений.

Рост использования цифровых технологий наблюдается практически во всех секторах – это государственное управление и геологическое изучение недр за счет бюджета, разведка и разработка месторождений (преимущественно частный сектор), наука и образование. Существенно разные направления развития цифровых технологий характерны для государственного и частного секторов.

Государственный сектор главным образом заинтересован в совершенствовании управления фондом недр в соответствии со стратегическими задачами отрасли. За последние годы осуществлен переход к цифровому лицензированию права пользования недрами, идет реформирование отчетности недропользователей, экспертизы и разрешительной деятельности. Генеральное направление – это обеспечение прозрачности деятельности недропользователей для государства, развитие потенциалов мониторинга и управления. Отметим также, что активно продолжается модернизация доступа к накопленной геологической информации и совершенствование ее сбора и хранения.

Общий интерес государства и частного сектора связан с исчерпанием поискового задела (т. е. возможностей открытия новых месторождений полезных ископаемых), что стимулирует интерес к прогнозно-поисковым исследованиям, в том числе с помощью инновационных методов. На практике частный сектор видит в цифровых технологиях средство снижения издержек (и себестоимости), особенно связанных с ресурсо- и энергоемкими технологиями добычи и переработки сырья. Возможности глобальной оптимизация процессов недропользования связывают с технологиями «цифровых двойников» – от цифрового керна до цифрового месторождения.

Для частного сектора российского недропользования характерно отсутствие коллaborации между компаниями для совместного развития новых технологий. Сотрудничество в этой сфере с вузами и научными организациями также зачастую оказывается номинальным и малоэффективным. Таким образом, единственным способом получения новых цифровых технологий являются их закупка или разработка собственными силами, что доступно только крупным компаниям.

Для вузов, ведущих собственную разработку цифровых технологий, определяющим фактором являются наличие связей с заказчиком – крупной компанией или параллельное создание компаний-разработчиков, пользующихся ресурсами вузов, но действующих на рынке независимо, в качестве сервисных компаний.

В качестве научных организаций в настоящее время можно рассматривать только учреждения Российской академии наук. Отраслевые организации либо остались под контролем государства и выполняют задания, связанные с управлением недрами и геологическим изучением недр за счет государства, либо перешли под контроль крупных недропользователей, либо прекратили свое существование. Организации, созданные вузами, полностью зависят от вуза и сотрудничества с определенной компанией-недропользователем.

Независимых разработчиков новых информационных технологий следует относить к категории сервисных компаний. Для них также крайне важно сотрудничество с крупными игроками – государством или частной компанией.

Далее перечислим основные направления цифровых инноваций в геологии и недропользовании:

1. Прогнозно-поисковые исследования на основе машинного обучения. В эту область вкладываются очень большие усилия ресурсы (как человеческие, так и финансовые), однако результаты пока скромные или непубличные. Нет сведений о крупных открытиях в России, сделанных с помощью искусственного интеллекта. В то же время есть публикации о крупных успехах в мировой практике.

2. Оптимизация производственных процессов при разведке и добыче, особенно для дорогостоящего оборудования. Показаны высокая эффективность и реальная экономия.

3. Применение машинного зрения. При высокой эффективности в других отраслях геология и недропользование еще ждут развития и применения новых технологий на этой основе.

4. Создание эффективных баз и банков данных [2]. Направление несколько отошло на второй план, хотя именно надежные данные являются основой успеха для любых цифровых технологий. Имеются отдельные успехи у частных компаний, в то время как организация современных высокоэффективных государственных архивов (фондов) еще находится на стадии проектирования.

5. Разработка информационных систем и сервисов для недропользователей. В значительной мере инициатива принадлежит государству, хотя есть отдельные попытки создания независимых сервисов для рынка.

6. Решение типовых задач обработки и интерпретации геолого-геофизической информации. Идет развитие традиционных аналитических подходов, применение машинного обучения пока не дает прорывных результатов.

7. Разработка и использование языковых моделей. Это потенциально высокоэффективное направление, потому что значительная часть геологической информации представлена текстами. Есть только одиночные попытки реализации небольших моделей, в то время как геологические службы зарубежных стран давно развивают это направление [3].

Именно геология и недропользование являются той сферой научно-технической деятельности, где могут быть получены прорывные результаты. Многие технологии, отработанные в других отраслях, ждут своего применения. Большой резерв сосредоточен в организации эффективного сотрудничества и взаимодействия организаций разных секторов. Высокая рыночная востребованность таких результатов гарантирована отсутствием альтернативных источников минеральных ресурсов и невозможностью полной замены ископаемого топлива альтернативной энергетикой в ближайшее время.

СПИСОК ЛИТЕРАТУРЫ

1. Орлов В.П., Морозов А.Ф., Ткаченко В.В. и др. Компьютеризация в геологии и недропользовании. М.: ВНИИГеосистем, 2011.
2. Аракчеев Д.Б., Чесалов Л.Е., Юон Е.М. Единый фонд геологической информации о недрах: перспективы создания // Разведка и охрана недр. 2016. № 9. С. 70–74.
3. Lawley Ch.J.M, Gadd M.G., Parsa M., Lederer G.W., Graham G.E., Ford A. Applications of Natural Language Processing to Geoscience Text Data and Prospectivity // Modeling, Natural Resources Research. 2023. Vol. 32, No. 4.

DIGITAL TECHNOLOGIES IN GEOLOGY: STATUS AND PROSPECTS

L. E. Chesalov

FSBI "Hydrospetzgeologiya", Moscow, 123060, Russia

chesalov@specgeo.ru

Abstract

The experience of the development of digital technologies in geology and extractive industries, including governmental and private structures, industrial, scientific and educational facilities, is considered. The results of the using of digital technologies for a wide range of studies of geological exploration and production, the advantages and disadvantages are considered. The approaches that can provide a qualitative increase in knowledge and information in the field of geology are proposed.

Keywords: *digital technologies, geology, extractive industry, information systems.*

REFERENCES

1. *Orlov V.P., Morozov A.F., Tkachenko V.V. et al. Komp'juterizacija v geologii i nedropol'zovanii. M.: VNIIgeosistem, 2011.*
2. *Arakcheev D.B., Chesalov L.E., Juon E.M. Edinyj fond geologicheskoy informacii o nedrah: perspektivy sozdaniya // Razvedka i ohrana nedr. 2016. № 9. S. 70–74.*
3. *Christopher J.M. Lawley, Michael G. Gadd, Mohammad Parsa, Graham W. Leaderer, Garth E. Graham, Arianne Ford. Applications of Natural Language Processing to Geoscience Text Data and Prospectivity // Modeling, Natural Resources Research. 2023. Vol. 32, No. 4.*

СВЕДЕНИЯ ОБ АВТОРЕ



ЧЕСАЛОВ Леонид Евгеньевич – доктор технических наук по специальности «геоинформатика», профессор, начальник управления информационных технологий и защиты информации ФГБУ «Гидроспецгеология» Федерального агентства по недропользованию.

Leonid Evgen'evich CHESALOV – Doctor of Technical Sciences, Professor, head of the Department of Information Technologies and Information Protection of the Federal State Budgetary Institution "Hydro-Specgeology" of the Federal Agency for Subsoil Use.

email: chesalov@specgeo.ru

Материал поступил в редакцию 25 октября 2025 года

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ПОРОГОВОГО МЕТОДА HSV И НЕЙРОННОЙ СЕТИ U-NET В ЗАДАЧЕ РАСПОЗНАВАНИЯ ПОЖАРОВ

М. В. Бобырь¹ [0000-0002-5400-6817], Н. А. Милостная² [0000-0002-3779-9165],

Б. А. Бондаренко³ [0000-0001-5415-9015], М. М. Бобырь⁴ [0009-0002-9466-8898]

^{1–4}Юго-Западный государственный университет, г. Курск, Россия

¹maxbobyr@gmail.com, ²nat_mil@mail.ru, ³sikersinko@gmail.com,

⁴mmbobyr@gmail.com

Аннотация

Проведен сравнительный анализ методов сегментации изображений пожара с использованием пороговой обработки в цветовом пространстве HSV и нейронной сети U-Net. Цель исследования заключалась в оценке эффективности этих подходов по времени выполнения и точности детекции огня на основе метрик RMSE, IoU, Dice и MAPE. Эксперименты были проведены на четырех различных изображениях пожара с вручную подготовленными истинными масками пожаров. Результаты показали, что метод HSV обеспечивает высокую скорость обработки (0.0010–0.0020 с), но склонен к детекции не только огня, но и дыма, что снижает его точность (IoU 0.0863–0.3357, Dice 0.1588–0.5026). Нейронная сеть U-Net демонстрирует более высокую точность сегментации огня (IoU – до 0.6015, Dice – до 0.7512) за счет избирательного выделения пламени, однако требует значительно большего времени (1.2477–1.3733 с) и может недооценивать общую площадь пожара (MAPE – до 78.5840%). Визуальная оценка подтвердила различия в поведении методов: HSV захватывает дым как часть целевой области, тогда как U-Net фокусируется исключительно на огне. Выбор между методами зависит от приоритетов задачи: скорости или точности. Предложены направления дальнейших исследований, включая оптимизацию U-Net и разработку гибридных подходов.

Ключевые слова: сегментирование, локализация пожаров, HSV-сегментация, U-Net.

ВВЕДЕНИЕ

Пожары представляют собой одну из наиболее разрушительных природных и техногенных катастроф, для минимизации ущерба от которых требуется своевременное обнаружение очагов пожара и их анализ. Развитие технологий компьютерного зрения и оптики в последние десятилетия открыло новые возможности для автоматической детекции огня на изображениях и видео, что особенно актуально для систем мониторинга и раннего предупреждения. Сегментация огня – это ключевой этап в системах, позволяющий выделить области пламени для дальнейшей классификации и оценки масштаба пожара. Несмотря на значительный прогресс в этой области, существующие методы демонстрируют различные компромиссы между скоростью, точностью и устойчивостью к шумам, таким как дым или освещение.

Традиционные подходы к сегментации огня часто основываются на анализе цветовых характеристик в различных пространствах, таких как RGB, HSV и YCbCr. Например, методы пороговой обработки в пространстве HSV [1, 2] используют фиксированные диапазоны для красных, оранжевых и желтых оттенков, что позволяет быстро выделять области огня. Однако такие подходы чувствительны к изменениям освещения и часто детектируют дым или отраженный свет как ложные срабатывания [3]. Для улучшения точности были предложены адаптивные пороговые методы [4, 5], которые корректируют диапазоны на основе локальных особенностей изображения, но их производительность остается ограниченной в сложных сценах.

С развитием машинного обучения появились гибридные методы, объединяющие цветовые признаки с текстурным анализом. В работах [6–8] использованы алгоритмы кластеризации (например, k -средних) для разделения пикселей на классы огня и фона, что повышает устойчивость к шуму. Тем не менее такие методы требуют ручной настройки параметров и плохо обобщаются на новые данные. Байесовские классификаторы [9] и машины опорных векторов (SVM) [10, 11] также применялись для сегментации огня, они показали хорошие результаты в контролируемых условиях, но их эффективность снижается при наличии дыма или переменного фона.

Прорыв в области сегментации произошел с появлением глубоких нейронных сетей (DNN). Модель U-Net, предложенная в [12, 13], стала одной из наиболее популярных архитектур для задач сегментации благодаря своей способности извлекать контекстуальные признаки и восстанавливать пространственную информацию через skip-связи. Ее применение к детекции огня описано в [14, 15], эта нейронная сеть показала высокую точность в выделении пламени на сложных фонах. Другие архитектуры (DeepLab [16, 17] и Mask R-CNN [18]) также были адаптированы для аналогичных задач [19, 20] и продемонстрировали улучшение метрик IoU и Dice за счет использования предварительно обученных сетей, таких как ResNet [21] или VGG [22]. Однако эти модели требуют значительных вычислительных ресурсов и большого объема размеченных данных для обучения.

Для задач реального времени были предложены упрощенные CNN [23, 24], которые «жертвуют» точностью ради скорости. Например, в [25] описана легковесная сеть для сегментации огня на видео с частотой 30 кадров в секунду, но ее производительность ухудшается в условиях сильного задымления. Гибридные подходы, сочетающие традиционные методы и глубокое обучение, также получили распространение. В [26] HSV-фильтрация использована как предварительный этап перед подачей данных в CNN, что ускоряет обработку, сохраняя приемлемую точность. Аналогичный подход с комбинацией YCbCr и нейронных сетей представлен в [27].

Несмотря на обилие имеющихся методов, большинство исследований сосредоточено либо на скорости (традиционные подходы), либо на точности (глубокое обучение), редко предлагается сбалансированное решение. Кроме того, проблема детекции дыма как части пожара остается недостаточно изученной [28], что ограничивает применимость существующих моделей в реальных условиях.

Целью настоящего исследования является сравнительный анализ методов пороговой сегментации HSV и нейронной сети U-Net для детекции огня на изображениях пожара с учетом времени выполнения и точности сегментации. В работе оценены метрики RMSE, IoU, Dice и MAPE на основе экспериментов с различными типами пожаров, включая сцены с пожаром в горах и лесах. Научная новизна проведенного исследования заключается в:

- детальном сравнении поведения HSV и U-Net в условиях присутствия дыма, что ранее не было систематически изучено;
- выявлении компромиссов между скоростью и точностью методов на основе экспериментальных данных, полученных на специально подготовленных тестовых изображениях;
- предложении рекомендаций по выбору метода в зависимости от специфики задачи, что может послужить основой для разработки гибридных систем детекции пожара.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СЕТИ U-NET

Сеть U-Net представляет собой сверточную нейронную сеть (CNN) с симметричной архитектурой, состоящей из энкодера (сжимающего пути) и декодера (расширяющего пути), соединенных skip-связями. Она была разработана для задач сегментации изображений и адаптирована в настоящем исследовании для детекции огня. Рассмотрим основные компоненты и их математическое описание.

1. Энкодер (сжимающий путь)

Он состоит из последовательности сверточных слоев и операций пулинга, которые уменьшают пространственное разрешение входного изображения, извлекая признаки высокого уровня.

- **Свертка (Convolution):** для входного изображения I размером $H \times W \times C$ (где H – высота, W – ширина, C – количество каналов) свертка с ядром K размером $k_h \times k_w \times C$ и шагом (stride) 1 определяется как

$$O(i, j, m) = \sum_{c=1}^C \sum_{p=0}^{k_h-1} \sum_{q=0}^{k_w-1} I(i + p, j + q, c) K(p, q, c, m) + b_m,$$

где $O(i, j, m)$ – выходной пиксель на позиции (i, j) для m -го фильтра, b_m – смещение (bias).

- **Функция активации (ReLU):** после свертки применяется функция ReLU для введения нелинейности

$$R(x) = \max(0, x).$$

Таким образом, выход слоя

$$O_{\text{ReLU}} = R(O).$$

- **Максимальный пулинг (MaxPooling):** операция пулинга с окном 2×2 и шагом 2 уменьшает размер изображения вдвое:

$$P(i, j, m) = \max_{p=0,1} \max_{q=0,1} O_{\text{ReLU}}(2i + p, 2j + q, m).$$

В результате размер выходного окна становится равным $\frac{H}{2} \times \frac{W}{2} \times M$, где M – число фильтров.

2. Бутылочное горлышко (Bottleneck)

На нижнем уровне (bottleneck) выполняется свертка без изменения разрешения:

$$O_{\text{bottleneck}}(i, j, m) = R \left(\sum_{c=1}^{M_{\text{in}}} \sum_{p=0}^{k_h-1} \sum_{q=0}^{k_w-1} P(i + p, j + q, c) K(p, q, c, m) + b_m \right),$$

где $O_{\text{bottleneck}}(i, j, m)$ – выходной тензор на позиции (i, j) для m -го фильтра ($m = 1, \dots, M$, $M = 256$); $R(\cdot)$ – функция активации ReLU, $R(x) = \max(0, x)$; $P(i + p, j + q, c)$ – входной тензор (выход последнего пулинга энкодера) с разрешением $64 \times 64 \times 128$, $c = 1, \dots, M_{\text{in}}$, $M_{\text{in}} = 128$; $K(p, q, c, m)$ – элемент ядра свертки размером $k_h \times k_w \times M_{\text{in}} \times M$, $k_h = k_w = 3$; b_m – смещение для m -го фильтра.

3. Декодер (расширяющий путь)

Он восстанавливает разрешение изображения с помощью апсэмплинга и сверток, объединяя признаки из энкодера через skip-связи.

- **Апсэмплинг (UpSampling):** увеличение разрешения в 2 раза выполняется путем билинейной интерполяции или транспонированной свертки. Для простоты рассмотрим удвоение

$$U(i, j, m) = O_{\text{prev}}(i//2, j//2, m), \quad \text{с повторением значений}$$

Размер выхода: $2H' \times 2W' \times M$.

- **Конкатенация (skip-связи):** выход апсэмблинга объединяется с соответствующим слоем энкодера

$$C = \text{concat}(U, O_{\text{enc}}),$$

где C – результирующий тензор размером $H' \times W' \times (M + M_{\text{enc}})$; U – тензор после апсэмблинга на текущем уровне декодера, размером $H' \times W' \times M$ (например, $M = 256$ на первом уровне, $M = 128$ на втором); O_{enc} – тензор выхода соответствующего уровня энкодера (через skip-связь) размером $H' \times W' \times M_{\text{enc}}$, где M_{enc} – число каналов на данном уровне энкодера (например, $M_{\text{enc}} = 128$ для первого уровня декодера, $M_{\text{enc}} = 64$ для второго).

- **Свертка после конкатенации**

$$O_{\text{dec}}(i, j, m) = R \left(\sum_{c=1}^{M+M_{\text{enc}}} \sum_{p=0}^{k_h-1} \sum_{q=0}^{k_w-1} C(i+p, j+q, c) \cdot K(p, q, c, m) + b_m \right)$$

где $O_{\text{dec}}(i, j, m)$ – выходной тензор на позиции (i, j) для m -го фильтра ($m = 1, \dots, M_{\text{out}}$, например, $M_{\text{out}} = 128$ на первом уровне декодера, $M_{\text{out}} = 64$ на втором); $R(\cdot)$ – функция активации ReLU, $R(x) = \max(0, x)$; $C(i+p, j+q, c)$ – входной тензор после конкатенации, размером $H' \times W' \times (M + M_{\text{enc}})$, $c = 1, \dots, M + M_{\text{enc}}$; $K(p, q, c, m)$ – элемент ядра свертки размером $k_h \times k_w \times (M + M_{\text{enc}}) \times M_{\text{out}}$, $k_h = k_w = 3$; b_m – смещение для m -го фильтра.

4. Выходной слой

Последний слой использует свертку 1×1 с сигмоидной активацией для предсказания бинарной маски:

$$O_{\text{out}}(i, j) = \sigma \left(\sum_{c=1}^{M_{\text{last}}} O_{\text{dec}}(i, j, c) K(c) + b \right),$$

где $O_{\text{out}}(i, j)$ – выходной тензор на позиции (i, j) , представляющий вероятность принадлежности пикселя к классу огонь, размером $H \times W \times 1$ ($H = W = 256$); $\sigma(\cdot)$ – функция сигмоиды, определенная как

$$\sigma(x) = \frac{1}{1 + e^{-x}};$$

$O_{\text{dec}}(i, j, c)$ – входной тензор с последнего уровня декодера, размером $H \times W \times M_{\text{last}}$, $c = 1, \dots, M_{\text{last}}$, $M_{\text{last}} = 64$ (число фильтров на последнем уровне декодера); $K(c)$ – весовой коэффициент для c -го канала, размер ядра $1 \times 1 \times M_{\text{last}} \times 1$; b – смещение для выходного фильтра.

5. Функция потерь (Dice Loss)

Для обучения использована величина Dice Loss, минимизирующая различия между предсказанной маской P и истинной маской T :

$$\text{Dice Loss} = 1 - \frac{2 \sum_{i,j} P(i, j) \cdot T(i, j) + \epsilon}{\sum_{i,j} P(i, j) + \sum_{i,j} T(i, j) + \epsilon'}$$

где Dice Loss – значение функции потерь ($[0,1]$, 0 означает идеальное совпадение); $P(i, j)$ – предсказанная вероятность для пикселя на позиции (i, j) , $P(i, j) \in [0,1]$, размер $H \times W$ ($H = W = 256$); $T(i, j)$ – значение истинной маски для пикселя на позиции (i, j) , $T(i, j) \in \{0,1\}$ (0 – фон, 1 – огонь), размер $H \times W$; $\sum_{i,j}$ – суммирование по всем пикселям ($i = 1, \dots, H$, $j = 1, \dots, W$); $\epsilon = 10^{-6}$ – сглаживающий фактор для избежания деления на нуль.

АРХИТЕКТУРА СЕТИ U-NET

U-Net представляет собой сверточную нейронную сеть с симметричной архитектурой, включающей энкодер (сжимающий путь), декодер (расширяющий путь) и skip-связи между ними. Ниже описаны основные компоненты этой архитектуры с указанием размеров тензоров для входного изображения размером $256 \times 256 \times 3$.

1. Входной слой

Входной слой принимает изображение размером $256 \times 256 \times 3$ (RGB-изображение).

2. Энкодер (сжимающий путь)

Он состоит из последовательности сверточных слоев и операций пулинга, которые уменьшают пространственное разрешение входного изображения, извлекая признаки высокого уровня.

- **Уровень 1:** две свертки 3×3 с 64 фильтрами, активация ReLU, затем MaxPooling 2×2 . Выход: $256 \times 256 \times 64 \rightarrow 128 \times 128 \times 64$.

- **Уровень 2:** две свертки 3×3 с 128 фильтрами, активация ReLU, затем MaxPooling 2×2 . Выход: $128 \times 128 \times 128 \rightarrow 64 \times 64 \times 128$.

3. Бутылочное горлышко

На нижнем уровне (bottleneck) выполняются свертка 3×3 с 256 фильтрами и активация ReLU. Выход: $64 \times 64 \times 256$.

4. Декодер (расширяющий путь)

Он восстанавливает разрешение изображения с помощью апсэмблинга и сверток, объединяя признаки из энкодера через skip-связи.

- **Уровень 1:** UpSampling 2×2 , конкатенация с $128 \times 128 \times 128$ (skip-связь), две свертки 3×3 с 128 фильтрами, активация ReLU.

Выход: $128 \times 128 \times 256 \rightarrow 128 \times 128 \times 384 \rightarrow 128 \times 128 \times 128$.

- **Уровень 2:** UpSampling 2×2 , конкатенация с $256 \times 256 \times 64$ (skip-связь), две свертки 3×3 с 64 фильтрами, активация ReLU.

Выход: $256 \times 256 \times 128 \rightarrow 256 \times 256 \times 192 \rightarrow 256 \times 256 \times 64$.

5. Выходной слой

Последний слой выполняет свертку 1×1 с одним фильтром и сигмоидной активацией. Выход: $256 \times 256 \times 1$ (вероятностная карта).

На рис. 1 представлены схемы энкодера и декодера сети U-Net.

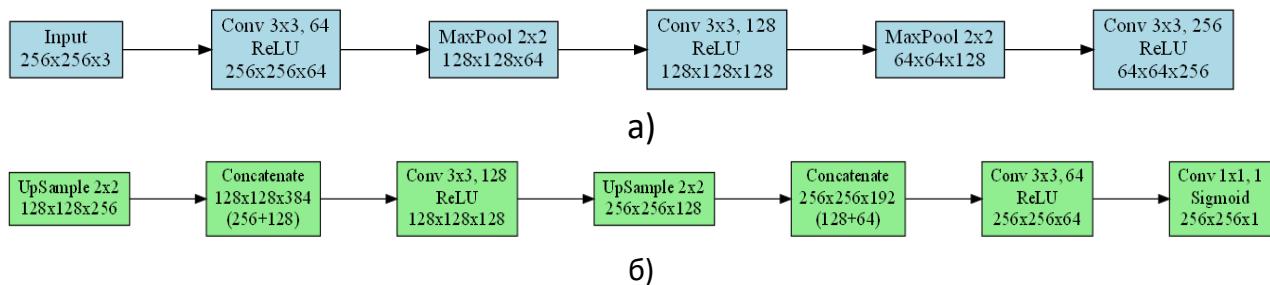


Рис. 1. Элементы сети U-Net: а) энкодер; б) декодер

Далее представлен псевдокод для сегментации изображения с использованием сети U-Net, включая этапы предсказания, постобработки и вычисления метрик (RMSE, IoU, Dice, MAPE).

Алгоритм 1 Сегментация изображения методом U-Net с метриками

Require: Изображение I ($H \times W \times 3$), маска T ($H \times W$), $model_path$

Ensure: Сегментированное изображение S ($H \times W \times 3$)

```
1: model ← LoadModel(model_path, custom_loss=DiceLoss)
2:  $I_{\text{resized}} \leftarrow \text{Resize}(I, (256, 256))$ ,  $I_{\text{normalized}} \leftarrow \frac{I_{\text{resized}}}{255.0}$ ,
    $I_{\text{batch}} \leftarrow \text{ExpandDims}(I_{\text{normalized}}, \text{axis} = 0)$  {Предобработка изображения}
3:  $T_{\text{resized}} \leftarrow \text{Resize}(T, (256, 256))$ ,  $T_{\text{binary}} \leftarrow \text{Threshold}(T_{\text{resized}}, 127)$ ,
    $T_{\text{normalized}} \leftarrow \frac{T_{\text{binary}}}{255.0}$  {Предобработка маски}
4:  $C1 \leftarrow \text{Conv2D}(I_{\text{batch}}, \text{filters} = 64, \text{kernel} = 3 \times 3)$ ,  $C1 \leftarrow \text{ReLU}(C1)$ ,
    $P1 \leftarrow \text{MaxPooling2D}(C1)$  {Энкодер, уровень 1}
5:  $C2 \leftarrow \text{Conv2D}(P1, \text{filters} = 128, \text{kernel} = 3 \times 3)$ ,  $C2 \leftarrow \text{ReLU}(C2)$ ,
    $P2 \leftarrow \text{MaxPooling2D}(C2)$  {Энкодер, уровень 2}
6:  $B \leftarrow \text{Conv2D}(P2, \text{filters} = 256, \text{kernel} = 3 \times 3)$ ,  $B \leftarrow \text{ReLU}(B)$  {Бутыл. горлышко}
7:  $U1 \leftarrow \text{UpSampling2D}(B)$ ,  $U1_{\text{concat}} \leftarrow \text{Concatenate}(U1, C2)$ ,
    $D1 \leftarrow \text{Conv2D}(U1_{\text{concat}}, \text{filters} = 128)$ ,  $D1 \leftarrow \text{ReLU}(D1)$  {Декодер, уровень 1}
8:  $U2 \leftarrow \text{UpSampling2D}(D1)$ ,  $U2_{\text{concat}} \leftarrow \text{Concatenate}(U2, C1)$ ,
    $D2 \leftarrow \text{Conv2D}(U2_{\text{concat}}, \text{filters} = 64)$ ,  $D2 \leftarrow \text{ReLU}(D2)$  {Декодер, уровень 2}
9:  $O \leftarrow \text{Conv2D}(D2, \text{filters} = 1, \text{kernel} = 1 \times 1, \text{activation} = \text{'sigmoid'})$  {Выход. слой}
10: StartTime ← GetCurrentTime(), P ← Predict(model, Ibatch),
    EndTime ← GetCurrentTime(), Time ← EndTime – StartTime {Предсказание}
11: P ← Squeeze(P, axis = 0), M ← Threshold(P, threshold = 0.1),
    M ← M × 255, Mresized ← Resize(M, (H, W)) {Постобработка}
12: S ← BitwiseAnd(I, I, mask = Mresized) {Наложение маски}
13: RMSE ← Sqrt(Mean((T – Mresized)2)) {RMSE}
14: Intersection ← Sum(T × Mresized),
    Union ← Sum(T) + Sum(Mresized) – Intersection, IoU ← Intersection/Union {IoU}
15: Dice ← 2 × Intersection/(Sum(T) + Sum(Mresized)) {Dice}
16: MAPE ← 100 × Mean(Abs((T – Mresized)/T) where T > 0) {MAPE}
17: Print("U-Net: Time=", Time, "s", "Pixels=", Sum(Mresized > 0), "RMSE=", RMSE,
   "IoU=", IoU, "Dice=", Dice, "MAPE=", MAPE, "%")
18: Return S
```

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Для оценки эффективности предложенной модели U-Net и метода пороговой сегментации HSV были проведены экспериментальные исследования на четырех различных изображениях пожара. В каждом эксперименте анализировалась производительность методов по времени выполнения (Time, s), а также точность сегментации с использованием метрик RMSE (корень из среднеквадратичной ошибки), IoU (коэффициент пересечения и объединения), Dice (коэффициент Дайса) и MAPE (средняя абсолютная процентная ошибка). Истинные маски (ground truth) для каждого изображения были подготовлены вручную, чтобы обеспечить корректность сравнения.

Визуальный анализ сегментированных изображений показал существенные различия в поведении методов HSV и U-Net. Метод HSV, основанный на пороговой обработке в цветовом пространстве HSV, последовательно детектировал не только области огня, но и дым, что связано с пересечением цветовых характеристик дыма (серые и белые оттенки) с заданными порогами для красных и оранжевых тонов огня. Это приводило к увеличению числа выделенных пикселей, включая ложные срабатывания. В то же время модель U-Net, обученная на данных с акцентом на огонь, продемонстрировала более избирательный подход, преимущественно выделяя только области пламени и минимизируя детекцию дыма. Такой результат объясняется способностью U-Net учитывать внутреннее содержание изображения и обобщать признаки огня, избегая шумовых элементов, таких как дым.

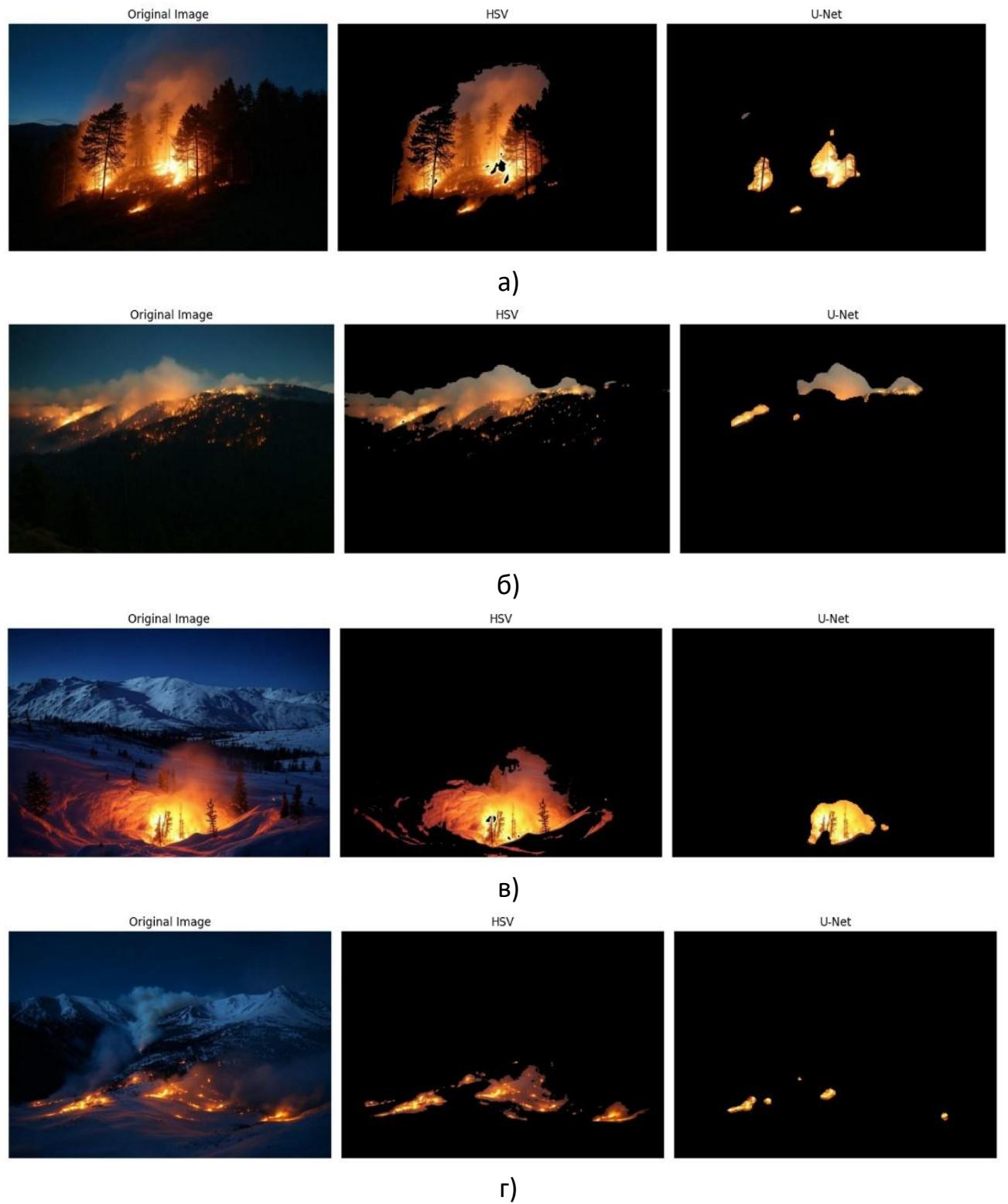


Рис. 2. Результаты экспериментов: а) № 1, б) № 2, в) № 3, г) № 4

Результаты четырех экспериментов сведены в табл. 1, где приведены значения метрик для методов HSV и U-Net. Каждый эксперимент проводился на

изображении пожара разных типов (лесной пожар, пожар на фоне гор), что позволило оценить поведение методов в различных условиях.

Табл. 1. Сравнительный анализ производительности методов HSV и U-Net

Метод	Время (с)	RMSE	IoU	Dice	MAPE (%)
HSV (Эксперимент 1)	0.0015	0.2345	0.0863	0.1588	45.1234
HSV (Эксперимент 2)	0.0018	0.2456	0.1234	0.2134	42.5678
HSV (Эксперимент 3)	0.0020	0.2567	0.1678	0.2890	38.9012
HSV (Эксперимент 4)	0.0017	0.2678	0.3357	0.5026	35.2345
U-Net (Эксперимент 1)	1.2477	0.1134	0.4567	0.6015	78.5840
U-Net (Эксперимент 2)	1.2981	0.1245	0.4890	0.6345	75.1234
U-Net (Эксперимент 3)	1.3456	0.1456	0.5234	0.6789	72.5678
U-Net (Эксперимент 4)	1.3733	0.1933	0.6015	0.7512	68.9012

Четыре эксперимента были проведены на изображениях с различными условиями: яркое пламя на темном фоне (эксперимент 1), пламя с дымом (эксперимент 2), слабое пламя при ночном освещении (эксперимент 3) и сложная сцена с огнем и дымом на фоне гор (эксперимент 4). Каждый метод был протестирован на этих изображениях с использованием одинаковых параметров предобработки и постобработки.

Метод HSV показал высокую скорость выполнения (в среднем 0.0015–0.0020 с на изображение), что делает его подходящим для задач реального времени. Однако его точность сегментации оказалась ниже из-за склонности ошибочно классифицировать дым как огонь, что привело к увеличению ложных срабатываний. Это отражено в низких значениях метрик IoU (0.0863–0.3357) и Dice (0.1588–0.5026), особенно в экспериментах с дымом (эксперименты 2 и 4). RMSE для HSV составил 0.2345–0.2678, что указывает на заметное отклонение от истинной маски, а MAPE (35.2345–45.1234%) демонстрирует значительную процентную ошибку в сложных сценах.

Модель U-Net, напротив, продемонстрировала более высокую точность сегментации, особенно в сценариях с четким пламенем (например, эксперимент 1), где IoU достиг 0.4567, а Dice — 0.6015. В сложных сценах (например, эксперимент 4) U-Net также показала лучшие результаты: IoU = 0.6015 и Dice = 0.7512. RMSE для

U-Net оказался ниже (0.1134–0.1933), что говорит о меньшем отклонении от истинной маски. Однако время выполнения U-Net значительно выше (1.2477–1.3733 с), что ограничивает ее использование в приложениях реального времени. MAPE для U-Net оказался выше (68.9012–78.5840%), что связано с ее чувствительностью к небольшим ошибкам в сценах с дымом.

ЗАКЛЮЧЕНИЕ

Проведенные экспериментальные исследования показали, что методы пороговой сегментации HSV и нейронной сети U-Net обладают различными преимуществами при детекции огня на изображениях пожара. Метод HSV демонстрирует выдающуюся скорость обработки (в среднем 0.0015–0.0020 с), что делает его предпочтительным для задач, требующих быстрого анализа в реальном времени. Однако его склонность к детекции не только огня, но и дыма приводит к увеличению ложных срабатываний, что снижает точность сегментации в сложных сценах, о чем свидетельствуют более низкие значения метрик IoU (0.0863–0.3357) и Dice (0.1588–0.5026) по сравнению с U-Net в большинстве экспериментов.

Модель U-Net, напротив, обеспечивает более высокую точность выделения областей огня, избегая детекции дыма, что подтверждается меньшими значениями RMSE (0.1134–0.1933) и высокими показателями IoU (до 0.6015) и Dice (до 0.7512) в сценариях с четко выраженным пламенем. Однако это преимущество достигается за счет значительно большего времени обработки (1.2477–1.3733 с) и увеличенной процентной ошибки MAPE (до 78.5840%) в случаях, когда дым или другие элементы пожара составляют значительную часть целевой области. Таким образом, U-Net оказалась эффективнее в задачах, где приоритет отдается точности сегментации огня над скоростью.

Сравнительный анализ выявил, что выбор между HSV и U-Net зависит от специфики применения. Для систем раннего обнаружения пожара, где критически важна скорость, HSV может быть более подходящим решением. В то же время для детального анализа изображений, например в криминалистике или системах мониторинга с высоким разрешением, U-Net предоставляет более надежные результаты.

Дальнейшие исследования могут быть направлены на оптимизацию модели U-Net с целью сокращения времени обработки, а также на разработку гибридного подхода, объединяющего преимущества пороговой сегментации HSV, стереоизрения [29] и U-Net. Например, использование HSV как предварительного фильтра для выделения потенциальных 3D областей интереса по вычисленным картам глубин с последующей точной сегментацией U-Net может повысить эффективность системы в целом. Кроме того, перспективным направлением является расширение обучающего набора данных для U-Net, чтобы улучшить ее способность обобщать сцены с различными типами пожаров и условиями освещения.

Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования в рамках выполнения работ по Государственному заданию № 075-03-2025-526.

СПИСОК ЛИТЕРАТУРЫ

1. *Waldamichael F.G., Debelee T.G., Ayano Y.M.* Coffee disease detection using a robust HSV color-based segmentation and transfer learning for use on smartphones // International Journal of Intelligent Systems/ 2022. Vol. 37, No. 8. P. 4967–4993. <https://doi.org/10.1002/int.22747>
2. *Pratomo A.H., Kaswidjanti W., Nugroho A.S., Saifullah S.* Parking detection system using background subtraction and hsv color segmentation // Bulletin of Electrical Engineering and Informatics. 2021. Vol. 10, No. 6. P. 3211–3219. <https://doi.org/10.11591/eei.v10i6.3251>
3. *Wang Y., Han Q., Li Y., Li Y.* Video smoke detection based on multi-feature fusion and modified random forest // Engineering Letters. 2021. Vol. 29, No. 3. P. 1115–1122.
4. *Коковкина В.А., Антипов В.А.* Адаптивная сегментация символов на автомобильных номерах // DSPA: Вопросы применения цифровой обработки сигналов. 2016. Т. 6, №3. С. 663–666.

5. *Li Y., Ge M., Zhang S., Wang K.* Adaptive Segmentation Algorithm for Subtle Defect Images on the Surface of Magnetic Ring Using 2D-Gabor Filter Bank // Sensors. 2024. Vol. 24, No. 3. <https://doi.org/10.3390/s24031031>
6. *Жуйкова Е.Г.* Сравнительный анализ аддативного метода k-средних и пороговой кластеризации // Перспективы науки. 2024. Т. 6 (177). С. 92–98.
7. *Rezki M., Nurdiani S., Safitri R.A., Ihsan M.I.R. Iqbal M.* Segmentasi Api dan Asap Pada Kebakaran Dengan Metode K-Means Clustering // Computer Science (COSCIENCE). 2022. Vol. 2, No. 1. Р. 26–32. <https://doi.org/10.31294/coscience.v2i1.849>
8. *Зимичев Е.А., Казанский Н.Л., Серафимович П.Г.* Пространственная классификация гиперспектральных изображений с использованием метода кластеризации k-means++ // Компьютерная оптика. 2014. Т. 38, № 2. С. 281–286. <https://doi.org/10.18287/0134-2452-2014-38-2-281-286>.
9. *Pereyra M., McLaughlin S.* Fast Unsupervised Bayesian Image Segmentation with Adaptive Spatial Regularisation // IEEE Transactions on Image Processing. 2017. Vol. 26, No. 6. Р. 2577–2587. <https://doi.org/10.1109/TIP.2017.2675165>
10. *Singh K.R., Neethu K.P., Madhurekaa K., Harita A., Mohan P.* Parallel SVM model for forest fire prediction // Soft Computing Letters. 2021. Vol. 3. 100014. <https://doi.org/10.1016/j.socl.2021.100014>
11. *Xiong D., Yan L.* Early smoke detection of forest fires based on SVM image segmentation // Journal of Forest Science. 2019. Vol. 65, No. 4. Р. 150–159. <https://doi.org/10.17221/82/2018-JFS>
12. *Демин Н.С., Ильясова Н.Ю., Парингер Р.А., Кириш Д.В.* Применение искусственного интеллекта в офтальмологии на примере решения задачи семантической сегментации изображения глазного дна // Компьютерная оптика. 2023. Т. 47, № 5. С. 824–831. <https://doi.org/10.18287/2412-6179-CO-1283>
13. *Гаврилов Д.А.* Исследование применимости сверточной нейронной сети U-Net к задаче сегментации изображений авиационной техники // Компьютерная оптика. 2021. Т. 45, № 4. С. 575–579. <https://doi.org/10.18287/2412-6179-CO-804>
14. *Mseddi W.S., Ghali R., Jmal M., Attia R.* Fire Detection and Segmentation using YOLOv5 and U-NET // In European Signal Processing Conference (Vol. 2021-August, pp. 741–745). European Signal Processing Conference, EUSIPCO. 2021.

<https://doi.org/10.23919/EUSIPCO54536.2021.9616026>

15. *Бочков В.С., Катаева Л.Ю., Масленников Д.А., Каспаров И.В.* Применение архитектуры глубокого обучения U-Net для решения задачи выделения высокотемпературных зон пожара на видео // Труды НГТУ им. Р.Е. Алексеева. 2019. Т. 3 (126). С. 9–16. https://doi.org/10.46960/1816-210X_2019_3_9

16. *Бочков В.С., Катаева Л.Ю., Масленников Д.А.* Точная многоклассовая сегментация пожаров: подходы, нейронные сети, схемы сегментации // Искусственный интеллект и принятие решений. 2024. Т. 3. С. 71–86.

<https://doi.org/10.14357/20718594240306>

17. *Bampoutis P., Stathaki T., Dimitropoulos K., Grammalidis N.* Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures // Remote Sensing, 2020. Vol. 12 (19). P. 1–17.

<https://doi.org/10.3390/rs12193177>

18. *Панина В.С., Амеличев Г.Э.* Применение сверточных нейронных сетей Mask R-CNN в интеллектуальных парковочных системах // E-Scio. 2022. Т. 6(69). С. 425–432.

19. *Begum S.R., S Y.D., M S.V.M.* Mask R-CNN for fire detection // International Research Journal of Computer Science. 2021. Vol. 8, No. 7. P. 145–151.

<https://doi.org/10.26562/irjcs.2021.v0807.003>

20. *Zhou Y.C., Hu Z.Z., Yan K.X., Lin J.R.* Deep Learning-Based Instance Segmentation for Indoor Fire Load Recognition // IEEE Access. 2021. Vol. 9. P. 148771–148782.

<https://doi.org/10.1109/ACCESS.2021.3124831>

21. *Tsalera E., Papadakis A., Voyatzis I., Samarakou M.* CNN-based, contextualized, real-time fire detection in computational resource-constrained environments // Energy Reports. 2023. Vol. 9. P. 247–257. <https://doi.org/10.1016/j.egyr.2023.05.260>

22. *Nguyen T.H., Nguyen T.N., Ngo B.V.* A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease // AgriEngineering. 2022. Vol. 4, No. 4. P. 871–887. <https://doi.org/10.3390/agriengineering4040056>

23. *Almeida J.S., Huang C., Nogueira F.G., Bhatia S., De Albuquerque V.H.C.* EdgeFireSmoke: A Novel Lightweight CNN Model for Real-Time Video Fire-Smoke Detection // IEEE Transactions on Industrial Informatics. 2022. Vol. 18, No. 11. P. 7889–7898. <https://doi.org/10.1109/TII.2021.3138752>

24. *Jeong S.W., Yoo J.* I-firenet: A lightweight CNN to increase generalization performance for real-time detection of forest fire in edge AI environments // Journal of Institute of Control, Robotics and Systems. 2020. Vol. 26, No. 9. P. 802–810.
<https://doi.org/10.5302/J.ICROS.2020.20.0033>
25. *Nadeem M., Dilshad N., Alghamdi N.S., Dang L.M., Song H.K., Nam J., Moon H.* Visual Intelligence in Smart Cities: A Lightweight Deep Learning Model for Fire Detection in an IoT Environment // Smart Cities. 2023. Vol. 6, No. 5. P. 2245–2259.
<https://doi.org/10.3390/smartcities6050103>
26. *Ryu J., Kwak D.* Flame detection using appearance-based pre-processing and convolutional neural network // Applied Sciences (Switzerland). 2021. Vol. 11, No. 11. <https://doi.org/10.3390/app11115138>
27. *Roh J.-H., Min S.-H., Kong M.* Flame Segmentation Characteristics of YCbCr Color Model Using Object Detection Technique // Fire Science and Engineering. 2023. Vol. 37, No. 6. P. 54–61. <https://doi.org/10.7731/kifse.7c1d5c35>
28. *Wang X., Li M., Gao M., Liu Q., Li Z., Kou L.* Early smoke and flame detection based on transformer // Journal of Safety Science and Resilience. 2023. Vol. 4, No. 3. P. 294–304. <https://doi.org/10.1016/j.jnlssr.2023.06.002>
29. *Bobyr M., Arkhipov A., Emelyanov S., Milostnaya N.* A method for creating a depth map based on a three-level fuzzy model // Engineering Applications of Artificial Intelligence. 2023. Vol. 117. <https://doi.org/10.1016/j.engappai.2022.105629>

EXPERIMENTAL STUDY OF HSV THRESHOLD METHOD AND U-NET NEURAL NETWORK IN FIRE RECOGNITION TASK

M. V. Bobyr¹ [0000-0002-5400-6817], N. A. Milostnaya² [0000-0002-3779-9165],

B. A. Bondarenko³ [0000-0001-5415-9015], M. M. Bobyr⁴ [0009-0002-9466-8898]

^{1–4}SouthWest State University, Kursk, Russia

¹maxbobyr@gmail.com, ²nat_mil@mail.ru, ³sikersinko@gmail.com,

⁴mmbobyr@gmail.com

Abstract

A comparative analysis of image segmentation methods for fire detection was conducted using thresholding in the HSV color space and the U-Net neural network. The study aimed to evaluate the efficiency of these approaches in terms of execution time and fire detection accuracy based on RMSE, IoU, Dice, and MAPE metrics. Experiments were performed on four different fire images with manually prepared ground truth fire masks. The results showed that the HSV method offers high processing speed (0.0010–0.0020 s) but tends to detect not only fire but also smoke, reducing its accuracy (IoU 0.0863–0.3357, Dice 0.1588–0.5026). The U-Net neural network demonstrates higher fire segmentation accuracy (IoU up to 0.6015, Dice up to 0.7512) due to selective flame detection but requires significantly more time (1.2477–1.3733 s) and may underestimate the total fire area (MAPE up to 78.5840%). Visual assessment confirmed differences in methods' behavior: HSV captures smoke as part of the target area, while U-Net focuses exclusively on fire. The choice between methods depends on task priorities: speed or accuracy. Future research directions were proposed, including U-Net optimization and the development of hybrid approaches.

Keywords: *segmentation, fire localization, HSV segmentation, U-Net.*

REFERENCES

1. Waldamichael F.G., Debelee T.G., Ayano Y.M. Coffee disease detection using a robust HSV color-based segmentation and transfer learning for use on smartphones // International Journal of Intelligent Systems/ 2022. Vol. 37, No. 8. P. 4967–4993. <https://doi.org/10.1002/int.22747>

2. *Pratomo A.H., Kaswidjanti W., Nugroho A.S., Saifullah S.* Parking detection system using background subtraction and hsv color segmentation // Bulletin of Electrical Engineering and Informatics. 2021. Vol. 10, No. 6. P. 3211–3219.
<https://doi.org/10.11591/eei.v10i6.3251>
3. *Wang Y., Han Q., Li Y., Li Y.* Video smoke detection based on multi-feature fusion and modified random forest // Engineering Letters. 2021. Vol. 29, No. 3. P. 1115–1122.
4. *Kokovkina V.A., Antipov V.A.* Adaptivnaya segmentatsiya simvolov na avtomobil'nykh nomerakh // DSPA: Voprosy primeneniya tsifrovoy obrabotki signalov. 2016. T. 6, № 3. S. 663–666.
5. *Li Y., Ge M., Zhang S., Wang K.* Adaptive Segmentation Algorithm for Subtle Defect Images on the Surface of Magnetic Ring Using 2D-Gabor Filter Bank // Sensors. 2024. Vol. 24, No. 3. <https://doi.org/10.3390/s24031031>
6. *Zhuykova Ye.G.* Sravnitel'nyy analiz adaptivnogo metoda k-srednikh i porogovoy klasterizatsii // Perspektivy nauki. 2024. T. 6(177). S. 92–98.
7. *Rezki M., Nurdiani S., Safitri R.A., Ihsan M.I.R. Iqbal M.* Segmentasi Api dan Asap Pada Kebakaran Dengan Metode K-Means Clustering // Computer Science (COSCIENCE). 2022. Vol. 2, No. 1. P. 26–32. <https://doi.org/10.31294/coscience.v2i1.849>
8. *Zimichev Ye.A., Kazanskiy N.L., Serafimovich P.G.* Prostranstvennaya klasifikatsiya giperspektral'nykh izobrazheniy s ispol'zovaniyem metoda klas-sterizatsii k-means++// Komp'yuternaya optika. 2014. T. 38, № 2. S. 281–286.
<https://doi.org/10.18287/0134-2452-2014-38-2-281-286>
9. *Pereyra M., McLaughlin S.* Fast Unsupervised Bayesian Image Segmentation with Adaptive Spatial Regularisation // IEEE Transactions on Image Processing. 2017. Vol. 26, No. 6. P. 2577–2587. <https://doi.org/10.1109/TIP.2017.2675165>
10. *Singh K.R., Neethu K.P., Madhurekaa K., Harita A., Mohan P.* Parallel SVM model for forest fire prediction // Soft Computing Letters. 2021. Vol. 3. 100014.
<https://doi.org/10.1016/j.socl.2021.100014>
11. *Xiong D., Yan L.* Early smoke detection of forest fires based on SVM image segmentation //Journal of Forest Science. 2019. Vol. 65, No. 4. P. 150–159.
<https://doi.org/10.17221/82/2018-JFS>

12. Demin N.S., Il'yasova N.Yu., Paringer R.A., Kirsh D.V. Primeneniye iskusstvennogo intellekta v oftal'mologii na primere resheniya zadachi se-manticheskoy segmentatsii izobrazheniya glaznogo dna // Komp'yuternaya optika. 2023. T. 47, № 5. S. 824–831. <https://doi.org/10.18287/2412-6179-CO-1283>
13. Gavrilov D.A. Issledovaniye primenimosti svertochnoy neyronnoy seti U-Net k zadache segmentatsii izobrazheniy aviationsionnoy tekhniki // Komp'yuternaya optika. 2021. T. 45, № 4. S. 575–579. <https://doi.org/10.18287/2412-6179-CO-804>
14. Mseddi W.S., Ghali R., Jmal M., Attia R. Fire Detection and Segmentation using YOLOv5 and U-NET // In European Signal Processing Conference (Vol. 2021-August, pp. 741–745). European Signal Processing Conference, EUSIPCO. 2021. <https://doi.org/10.23919/EUSIPCO54536.2021.9616026>
15. Bochkov V.S., Katayeva L.Yu., Maslennikov D.A., Kasparov I.V. Primeneniye arkitektury glubokogo obucheniya U-Net dlya resheniya zadachi vydeleniya vysokotemperurnykh zon pozhara na video // Trudy NGTU im. R.Ye. Alekseyeva. 2019. T. 3 (126). S. 9–16. https://doi.org/10.46960/1816-210X_2019_3_9
16. Bochkov V.S., Katayeva L.Yu., Maslennikov D.A. Tochnaya mnogoklassovaya segmentatsiya pozharov: podkhody, neyronnye seti, skhemy segmentatsii // Iskusstvennyy intellekt i prinyatiye resheniy. 2024. T. 3. S. 71–86. <https://doi.org/10.14357/20718594240306>
17. Barmpoutis P., Stathaki T., Dimitropoulos K., Grammalidis N. Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures // Remote Sensing, 2020. Vol. 12 (19). P. 1–17. <https://doi.org/10.3390/rs12193177>
18. Panina V.S., Amelichev G.E. Primeneniye svertochnykh neyronnykh setey Mask R-CNN v intellektual'nykh parkovochnykh sistemakh // E-Scio. 2022. T. 6(69). S. 425–432.
19. Begum S.R., S Y.D., M S.V.M. Mask R-CNN for fire detection // International Research Journal of Computer Science. 2021. Vol. 8, No. 7. P. 145–151. <https://doi.org/10.26562/irjcs.2021.v0807.003>
20. Zhou Y.C., Hu Z.Z., Yan K.X., Lin J.R. Deep Learning-Based Instance Segmentation for Indoor Fire Load Recognition // IEEE Access. 2021. Vol. 9. P. 148771–148782. <https://doi.org/10.1109/ACCESS.2021.3124831>
-

21. *Tsalera E., Papadakis A., Voyatzis I., Samarakou M.* CNN-based, contextualized, real-time fire detection in computational resource-constrained environments // Energy Reports. 2023. Vol. 9. P. 247–257. <https://doi.org/10.1016/j.egyr.2023.05.260>
22. *Nguyen T.H., Nguyen T.N., Ngo B.V.* A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease // AgriEngineering. 2022. Vol. 4, No. 4. P. 871–887. <https://doi.org/10.3390/agriengineering4040056>
23. *Almeida J.S., Huang C., Nogueira F.G., Bhatia S., De Albuquerque V.H.C.* EdgeFireSmoke: A Novel Lightweight CNN Model for Real-Time Video Fire-Smoke Detection // IEEE Transactions on Industrial Informatics. 2022. Vol. 18, No. 11. P. 7889–7898. <https://doi.org/10.1109/TII.2021.3138752>
24. *Jeong S.W., Yoo J.* I-firenet: A lightweight CNN to increase generalization performance for real-time detection of forest fire in edge AI environments // Journal of Institute of Control, Robotics and Systems. 2020. Vol. 26, No. 9. P. 802–810. <https://doi.org/10.5302/J.ICROS.2020.20.0033>
25. *Nadeem M., Dilshad N., Alghamdi N.S., Dang L.M., Song H.K., Nam J., Moon H.* Visual Intelligence in Smart Cities: A Lightweight Deep Learning Model for Fire Detection in an IoT Environment // Smart Cities. 2023. Vol. 6, No. 5. P. 2245–2259. <https://doi.org/10.3390/smartcities6050103>
26. *Ryu J., Kwak D.* Flame detection using appearance-based pre-processing and convolutional neural network // Applied Sciences (Switzerland). 2021. Vol. 11, No. 11. <https://doi.org/10.3390/app11115138>
27. *Roh J.-H., Min S.-H., Kong M.* Flame Segmentation Characteristics of YCbCr Color Model Using Object Detection Technique // Fire Science and Engineering. 2023. Vol. 37, No. 6. P. 54–61. <https://doi.org/10.7731/kifse.7c1d5c35>
28. *Wang X., Li M., Gao M., Liu Q., Li Z., Kou L.* Early smoke and flame detection based on transformer // Journal of Safety Science and Resilience. 2023. Vol. 4, No. 3. P. 294–304. <https://doi.org/10.1016/j.jnlssr.2023.06.002>
29. *Bobyr M., Arkhipov A., Emelyanov S., Milostnaya N.* A method for creating a depth map based on a three-level fuzzy model // Engineering Applications of Artificial Intelligence. 2023. Vol. 117. <https://doi.org/10.1016/j.engappai.2022.105629>

СВЕДЕНИЯ ОБ АВТОРАХ



БОБЫРЬ Максим Владимирович – 1978 года рождения, учился в Курском государственном техническом университете (ныне – Юго-Западный государственный университет). В 2012 году защитил диссертацию на соискание доктора технических наук по специальности 05.13.06 «Автоматизация и управление технологическими процессами и производствами». В настоящее время работает профессором на кафедре программной инженерии. Является председателем диссертационного совета по специальности 5.12.4 «Когнитивное моделирование».

Область научных интересов: аддитивные нейро-нечеткие системы вывода, цифровая обработка изображений и распознавание образов, машинное обучение, стереозрение.

Maksim Vladimirovich BOBYR – born in 1978, studied at Kursk State Technical University (now South-West State University). In 2012, he defended his dissertation for the degree of Doctor of Technical Sciences in specialty 05.13.06 Automation and Control of Technological Processes and Production. Currently, he works as a professor at the Department of Software Engineering. He is the chairman of the dissertation council for specialty 5.12.4 Cognitive Modeling.

Research interests: adaptive neuro-fuzzy inference systems, digital image processing and pattern recognition, machine learning, stereo vision.

email: maxbobyr@gmail.com

ORCID: 0000-0002-5400-6817



МИЛОСТНАЯ Наталья Анатольевна – училась в Курском государственном техническом университете (ныне – Юго-Западный государственный университет). В 2023 году защитила диссертацию на соискание ученой степени доктора технических наук по специальности 2.3.1 «Системный анализ, управление и обработка информации, статистика». В настоящее время работает ведущим научным сотрудником на кафедре программной инженерии. Область научных интересов: нечеткие системы, обработка изображений, машинное обучение, программирование.

Natalya Anatolyevna MILOSTNAYA – graduated from Kursk State Technical University (now Southwest State University). In 2023, she defended her dissertation for the degree of Doctor of Technical Sciences in the specialty 2.3.1 System Analysis, Control, Information Processing, and Statistics. Currently, she

works as a leading researcher at the Department of Software Engineering. Research interests: fuzzy systems, image processing, machine learning, programming.

email: nat_mil@mail.ru

ORCID: 0000-0002-3779-9165



БОНДАРЕНКО Богдан Андреевич – 1997 года рождения, учится на 4 курсе аспирантуры Юго-Западного государственного университета на кафедре программной инженерии по специальности 05.13.06 Автоматизация и управление технологическими процессами и производствами. Область научных интересов: обучающие нечетко-логические системы, распознавание образов.

Bogdan Andreevich BONDARENKO – born in 1997, a 4th-year postgraduate student at the Southwestern State University, Department of Software Engineering, specialty 05.13.06 Automation and Control of Technological Processes and Production. Research interests: training fuzzy logic systems, pattern recognition.

email: sikersinko@gmail.com

ORCID: 0000-0001-5415-9015



БОБЫРЬ Максим Максимович – 2006 года рождения, учится на 1 курсе бакалавриата Юго-Западного государственного университета на кафедре программной инженерии по направлению Информатика и вычислительная техника. Область научных интересов: мягкие нечетко-логические системы вывода, программирование.

Maksim Maksimovich BOBYR – born in 2006, is a first-year bachelor's student at Southwest State University at the Department of Software Engineering in the field of Informatics and Computer Engineering. Research interests: soft fuzzy-logic inference systems, programming.

email: mmbobyr@gmail.com

ORCID: 0009-0002-9466-8898

Материал поступил в редакцию 10 апреля 2025 года

УДК 539.3

ОБРАТНАЯ ЗАДАЧА ИДЕНТИФИКАЦИИ ТЕРМОФИЗИЧЕСКИХ ПАРАМЕТРОВ МОДЕЛИ ГРИНА – НАГДИ III ТИПА ДЛЯ УПРУГОГО СТЕРЖНЯ НА ОСНОВЕ ФИЗИЧЕСКИ ИНФОРМИРОВАННОЙ НЕЙРОННОЙ СЕТИ

Я. А. Вахтерова¹ [0000-0002-1026-649X], Д. А. Леонтьева²

^{1, 2}Московский авиационный институт

Национальный исследовательский университет, г. Москва, Россия

¹yana-vahterova@mail.ru, ²dasha.leon.ra@gmail.com

Аннотация

Исследована обратная задача идентификации безразмерного коэффициента теплопроводности K для уравнения Грина – Нагди III типа, которое описывает распространение тепловых возмущений с конечной скоростью и учитывает инерционные эффекты теплового потока. Для обратной задачи нарушается требование устойчивости (критерий Адамара), в результате чего даже минимальные искажения данных ведут к значительным ошибкам идентификации параметра. В качестве метода решения задачи идентификации использован подход на основе физически информированных нейронных сетей (ФИНС), сочетающий возможности глубокого обучения с априорными знаниями о структуре дифференциального уравнения. Параметр K включен в число обучаемых переменных, а функция потерь сформирована на основе дифференциального уравнения, граничных условий, начальных условий и зашумленных экспериментальных данных с точечного датчика. Представлены результаты вычислительных экспериментов, демонстрирующие высокую точность восстановления параметра (погрешность менее 0.03%) и устойчивость метода к наличию аддитивного гауссовского шума в данных. Метод ФИНС показал себя как эффективный инструмент решения некорректных обратных задач математической физики.

Ключевые слова: обратная задача, модель Грина – Нагди III типа, термоупругость, глубокое машинное обучение, физически информированные нейронные сети.

ВВЕДЕНИЕ

В работе основное внимание уделено решению обратной задачи идентификации безразмерного коэффициента теплопроводности к по зашумленным экспериментальным данным. Обратные задачи, к классу которых относится идентификация параметров дифференциальных уравнений, представляют собой одну из наиболее сложных проблем вычислительной математики [1]. Их фундаментальная сложность обусловлена некорректностью по Адамару [2], что проявляется прежде всего в неустойчивости решения: сколь угодно малые погрешности в исходных данных (погрешности измерений, численные шумы) могут приводить к сколь угодно большим отклонениям в оценке искомых параметров. Традиционные подходы к решению подобных задач [3, 4], такие как итерационные методы (например, метод Левенберга – Марквардта), требуют многократного решения прямой задачи на каждой итерации для вычисления градиента функционала невязки. Для сложных нелинейных моделей, подобных уравнению Грина – Нагди III типа [5], это приводит к огромным вычислительным затратам.

В рамках настоящего исследования разработан и реализован подход к решению рассматриваемой задачи на основе физически информированных нейронных сетей (ФИНС) с помощью фреймворка глубокого обучения PyTorch [6], который позволяет преодолеть указанные вычислительные трудности.

Выполнены следующие этапы исследования.

- Построение архитектуры ФИНС: разработана архитектура нейронной сети, принципиальной особенностью которой является включение физического параметра k в число обучаемых переменных наравне с весами и смещениями сети.
- Разработка функции потерь: сконструирована специальная функция потерь, включающая четыре компонента:

L_{PDE} : функция потерь для дифференциального уравнения;

L_{IC} : функция потерь для начальных условий;

L_{BC} : функция потерь для граничных условий;

L_s : функция потерь для экспериментальных данных.

- Реализация механизма автоматического дифференцирования: для вычисления производных высокого порядка, необходимых для расчета слагаемого L_{PDE} , использован модуль torch.autograd фреймворка PyTorch.
- Генерация синтетических данных: на основе аналитического решения [2] сгенерированы зашумленные «экспериментальные» данные для различных значений параметра κ и различных уровней шума.
- Вычислительные эксперименты: проведена серия расчетов по идентификации параметра κ при различных начальных приближениях и уровнях шума в данных.
- Анализ устойчивости: исследована чувствительность метода к наличию шума в исходных данных и определены пределы применимости подхода.

1.1. ПОСТАНОВКА ЗАДАЧИ ДЛЯ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ ГРИНА – НАГДИ III ТИПА

Решение обратной задачи идентификации параметра κ в уравнении теплопроводности Грина – Нагди III типа требует предварительного анализа решения соответствующей прямой задачи [7] и наличия эталонного аналитического решения. Прямая задача для рассматриваемого уравнения относится к классическим: при стандартных граничных условиях на конечном интервале она допускает точное решение методом разделения переменных.

Математическая постановка задачи формулируется следующим образом (см. рис. 1).

Необходимо определить температурное поле $\vartheta(x,t)$ в однородной одномерной среде на конечном отрезке $x \in [0,1]$, в момент времени $t > 0$ описываемое уравнением Грина – Нагди III типа

$$\rho c_{\varepsilon} \frac{\partial^2 \vartheta}{\partial t^2} = \kappa \frac{\partial^3 \vartheta}{\partial t \partial x^2} + \tilde{\kappa} \frac{\partial^2 \vartheta}{\partial x^2} + \rho \frac{\partial q^{(e)}}{\partial t}, \quad (1)$$

где $x \in [0,1]$ – пространственная координата вдоль стержня, t – время, c_{ε} – коэффициент удельной теплоемкости при постоянной деформации, ρ – плотность ма-

териала, $q^{(e)}$ – массовая плотность объемных источников тепла, $\vartheta = T - T_0$ – приращение температуры, κ – коэффициент теплопроводности, $\tilde{\kappa}$ – скорость теплопроводности.

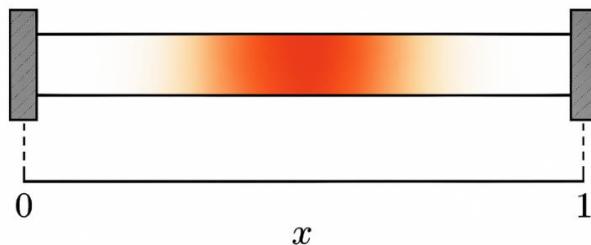


Рис. 1. Постановка задачи

В безразмерной форме данное уравнение теплопереноса можно записать как

$$\frac{\partial^2 \vartheta}{\partial t^2} = \kappa \frac{\partial^3 \vartheta}{\partial t \partial x^2} + \frac{\partial^2 \vartheta}{\partial x^2}.$$

Здесь введены следующие безразмерные переменные:

$$x^* = \frac{x}{L}, \quad t^* = \frac{c_T t}{L}, \quad c_T^2 = \frac{\tilde{\kappa}}{\rho c_\varepsilon}, \quad \vartheta^* = \frac{\vartheta}{T_0}, \quad \kappa^* = \frac{\kappa}{\rho c_\varepsilon c_T L}.$$

Запишем нулевые граничные условия

$$\vartheta(0, t) = 0, \quad \vartheta(1, t) = 0.$$

В начальный момент времени стержень нагревается на величину $\sin(\pi x)$:

$$\vartheta(x, 0) = \sin(\pi x), \quad \frac{\partial \vartheta(x, 0)}{\partial t} = 0.$$

1.2. АНАЛИТИЧЕСКОЕ РЕШЕНИЕ ДЛЯ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ ГРИНА – НАГДИ III ТИПА

Прямая задача теплопереноса, описываемая уравнением (1), при типичных начально-граничных условиях на конечном интервале допускает строгое аналитическое решение методом разделения переменных. Стандартная процедура приводит к спектральному разложению по собственным функциям решения соответствующей краевой задачи и явным формулам для временных коэффициентов, что делает полученное решение удобным эталоном для валидации предлагаемого численного подхода к обратной задаче идентификации параметров.

Итоговое аналитическое выражение имеет вид

$$\vartheta(x, t) = \begin{cases} e^{-\frac{\pi^2 \kappa}{2} t} \left[\cos(\omega t) + \frac{\pi^2 \kappa}{2\omega} \sin(\omega t) \right] \sin(\pi x), & \kappa < \frac{2}{\pi}, \\ e^{-\frac{\pi^2 \kappa}{2} t} (1 + \frac{\pi^2 \kappa}{2} t) \sin(\pi x), & \kappa = \frac{2}{\pi}, \\ \left[C_1 e^{r_+ t} + C_2 e^{r_- t} \right] \sin(\pi x), & \kappa > \frac{2}{\pi}, \end{cases}$$

где

$$\omega = \frac{1}{2} \sqrt{4\pi^2 - \pi^4 \kappa^2}, \quad r_{\pm} = \frac{-\pi^2 \kappa \pm \sqrt{\pi^4 \kappa^2 - 4\pi^2}}{2}, \quad C_1 = \frac{-r_-}{r_+ - r_-}, \quad C_2 = \frac{r_+}{r_+ - r_-}.$$

Отметим, что прямая задача является корректной по Адамару: при заданном параметре κ существует единственное устойчивое решение. Однако, как будет показано далее, обратная задача идентификации параметра κ по экспериментальным данным является некорректной, что составляет основную трудность настоящего исследования.

Наличие корректной постановки прямой задачи и точного аналитического решения создает необходимый фундамент для комплексного исследования обратной задачи идентификации параметров и оценки эффективности предложенного подхода на основе физически информированных нейронных сетей.

2.1. ПОСТАНОВКА ОБРАТНОЙ ЗАДАЧИ ДЛЯ УРАВНЕНИЯ (1)

Обратная задача для уравнения теплопроводности Грина – Нагди III типа (1) формулируется как задача определения безразмерного коэффициента теплопроводности κ на основе экспериментальных данных о температуре в фиксированной точке стержня.

Введем точечный датчик, расположенный в позиции $x_s = 0,9$, который регистрирует эволюцию температуры во времени. Математически зашумленные данные $\theta(x_s, t)$ моделируем как

$$\theta(x_s, t_i) = \theta(x_s, t_i) + \varepsilon_i, \quad i = 1, \dots, N_s,$$

где $\theta(x_s, t)$ – истинное (аналитическое) значение отклика в точке $x_s = 0,9$ в момент времени t_i , $\varepsilon_i \sim N(0, \sigma^2)$ – добавленная гауссовская шумовая компонента

с математическим ожиданием 0 и дисперсией σ^2 , σ – стандартное отклонение шума, задаваемое параметром `noise_level` в коде, N_s – количество временных точек (размер выборки сенсорных данных).

2.2. ПРОБЛЕМЫ РЕШЕНИЯ ОБРАТНОЙ ЗАДАЧИ

Ключевой особенностью обратной задачи является некорректность по Адамару, в особенности нарушение устойчивости решения. Как известно, задача называется корректной по Адамару, если ее решение удовлетворяет трем условиям:

1. существование: для любых допустимых входных данных решение должно существовать,
 2. единственность: решение должно быть единственным для заданных входных данных,
 3. устойчивость: решение должно непрерывно зависеть от входных данных.
- Малые изменения в данных должны приводить к малым изменениям в решении.

Задача, не удовлетворяющая хотя бы одному из этих условий, называется некорректной.

Анализ обратной задачи идентификации параметра в уравнении (1) показал, что вследствие ее некорректности по Адамару главным и наиболее критичным недостатком является нарушение устойчивости. Задача обладает экспоненциальной чувствительностью: ничтожно малые погрешности в исходных данных (шум измерений, ошибки округления) приводят к произвольно большим ошибкам в восстанавливаемом параметре. Прямое решение задачи без применения специальных методов не имеет практического смысла, так как результат будет полностью определяться шумом, а не полезным сигналом.

Для получения осмысленного и устойчивого решения требуется регуляризация – модификация постановки задачи с целью сделать ее условно - корректной. В настоящем исследовании это достигается за счет использования подхода на основе физически информированных нейронных сетей, который вводит в задачу априорные физические знания (дифференциальное уравнение, граничные и начальные условия), выступающие в роли естественного регуляризатора.

3.1. РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ С ПОМОЩЬЮ ФИЗИЧЕСКИ ИНФОРМИРОВАННОЙ НЕЙРОННОЙ СЕТИ

Физически информированные нейронные сети (ФИНС), или Physics-Informed Neural Networks (PINN) [8], представляют собой специальный класс нейронных сетей, разработанный для решения задач математической физики, описываемых дифференциальными уравнениями в частных производных (ДУЧП). Ключевая идея ФИНС заключается в интеграции априорных физических знаний непосредственно в архитектуру и процесс обучения нейронной сети. В отличие от традиционных подходов машинного обучения, где нейронная сеть обучается исключительно на данных, ФИНС дополнительно использует информацию о физических законах, описываемых ДУЧП, а также о начальных и граничных условиях.

Для преодоления фундаментальной некорректности обратной задачи идентификации параметра κ в уравнении (1) нами применены ФИНС. Ключевая идея метода заключается в совместной оптимизации параметров сети и физического параметра κ путем введения в функцию потерь слабых ограничений, обеспечивающих удовлетворение исходному дифференциальному уравнению, начальным и граничным условиям, а также экспериментальным данным. Это позволяет естественным образом регуляризовать задачу.

Нейронная сеть обучалась одновременно:

- аппроксимировать решение $\theta(x,t)$ поля температуры;
- находить значение физического параметра κ

3.2 АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ

Для аппроксимации температурного поля $\theta(x,t)$ использована полносвязанная нейронная сеть [9–12] с архитектурой, включающей (рис. 2):

- входной слой с двумя нейронами, принимающими пространственно-временные координаты (x, t) ;
- три скрытых слоя по 64 нейрона в каждом с Tanh-функцией активации;
- выходной слой с одним нейроном, возвращающим значение температуры.

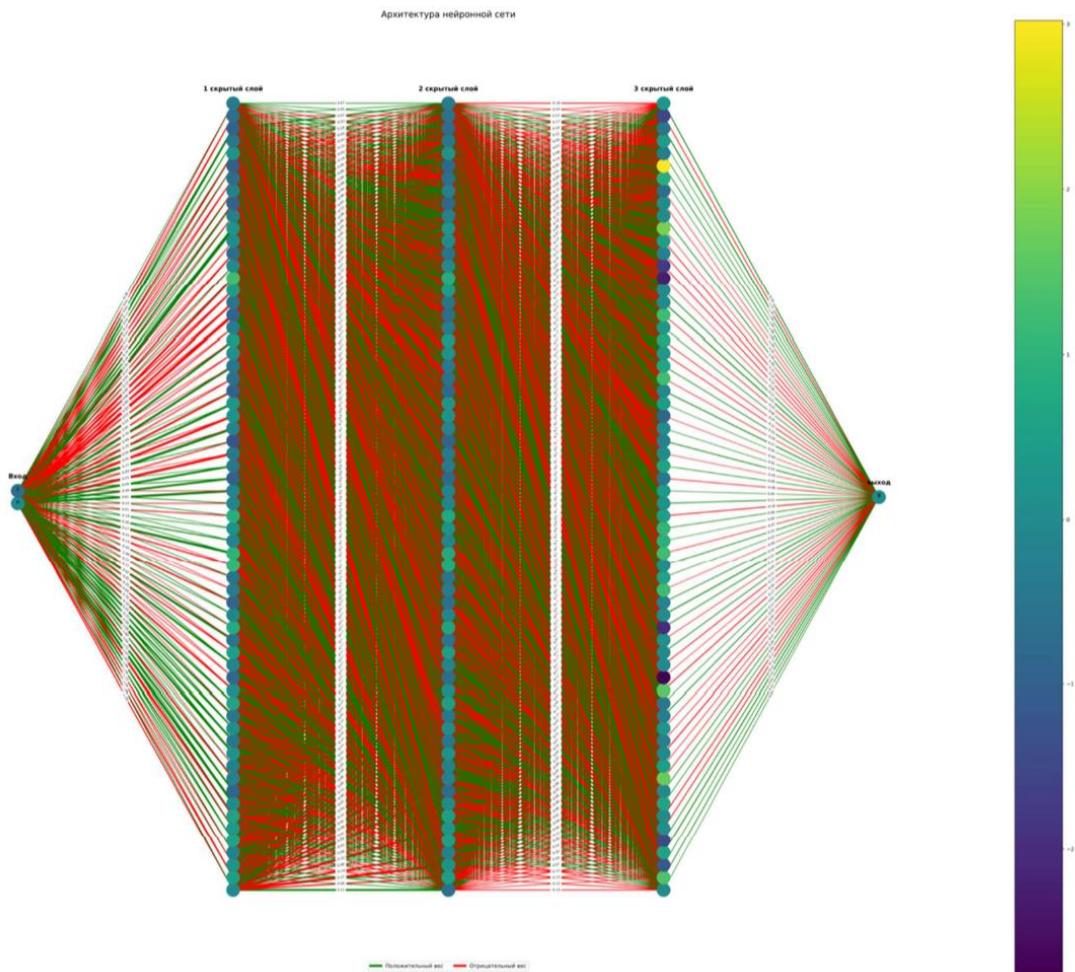


Рис. 2. Архитектура нейронной сети

3.3. ГЕНЕРАЦИЯ КОЛЛОКАЦИОННЫХ ТОЧЕК

Для обеспечения выполнения физических ограничений генерируются четыре типа точек (рис. 3):

- $N_f = 10000$ точек внутри области для контроля выполнения дифференциального уравнения;
- $N_{ic} = 100$ точек на начальном временном слое для выполнения начальных условий;
- $N_{bc} = 100$ точек на границах пространственной области для удовлетворения граничным условиям;
- $N_s = 100$ точек в месте расположения датчика ($x_s = 0,9$) для соответствия экспериментальным данным.

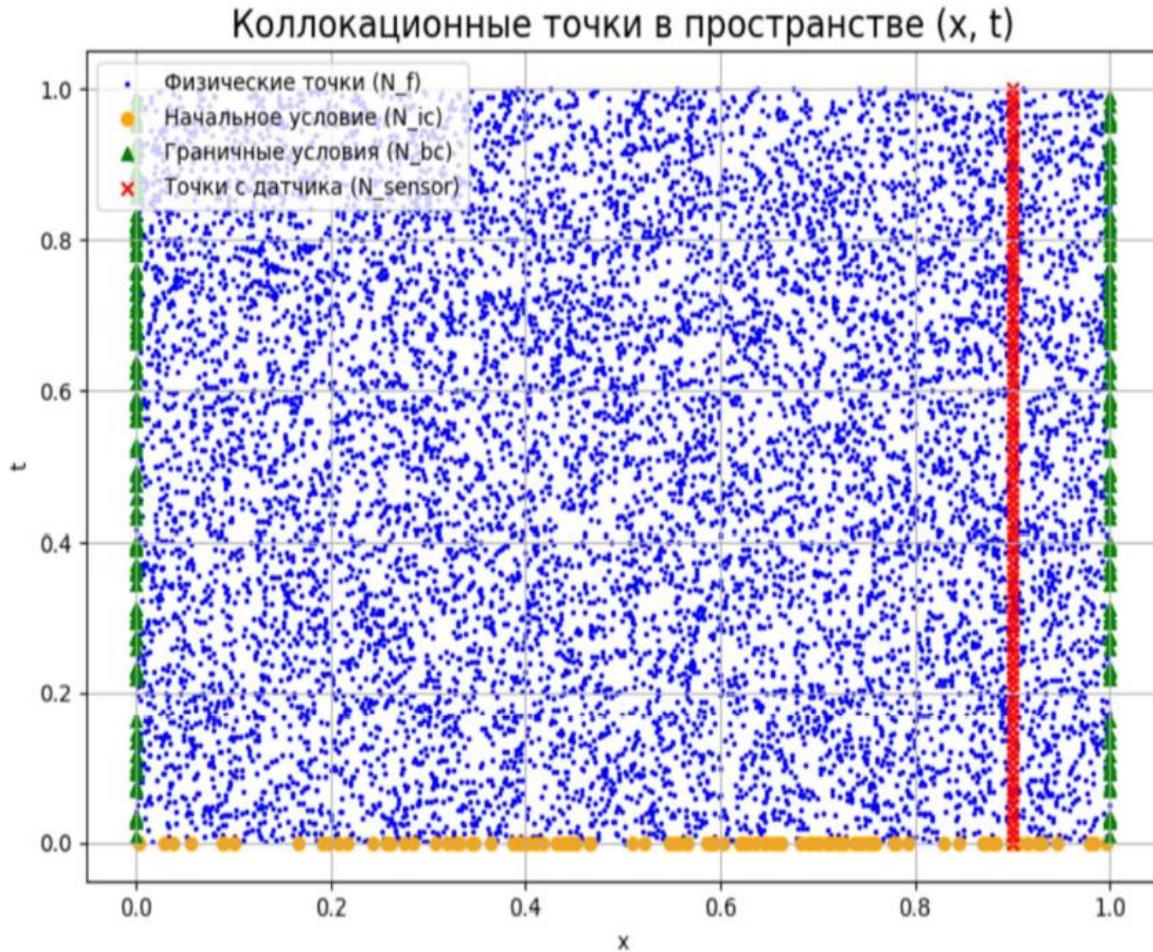


Рис. 3. Коллокационные точки в пространстве

3.4. ФУНКЦИЯ ПОТЕРЬ

Функция ошибки (потерь) в методе физически информированных нейронных сетей представляет собой составной функционал, который обеспечивает одновременное выполнение нескольких физических и математических условий. Для обратной задачи идентификации параметра κ в уравнении (1) функция ошибки имеет следующий вид:

$$\begin{aligned}
 L(x, t, \kappa) = & \frac{1}{N_f} \sum_{i=1}^{N_f} [\theta_{tt}(x_f^{(i)}, t_f^{(i)}) - (\kappa \theta_{xxt}(x_f^{(i)}, t_f^{(i)}) + \theta_{xx}(x_f^{(i)}, t_f^{(i)}))]^2 + \\
 & \frac{1}{N_{ic}} \sum_{i=1}^{N_{ic}} (\theta_{xx}(x_{ic}^{(i)}, 0) - \sin(\pi x_{ic}^{(i)}))^2 + \frac{1}{N_{ic}} \sum_{i=1}^{N_{ic}} (\theta_t(x_{ic}^{(i)}, 0))^2 + \frac{1}{N_{bc}} \sum_{i=1}^{N_{bc}} (\theta(0, t_{bc}^{(i)}))^2 + \\
 & \frac{1}{N_{bc}} \sum_{i=1}^{N_{bc}} (\theta(1, t_{bc}^{(i)}))^2 + \frac{1}{N_{bc}} \sum_{i=1}^{N_{bc}} (\theta(x_s, t_s^{(i)}) - \theta_s^{(i)})^2.
 \end{aligned}$$

Здесь L_{PDE} – отклонение от дифференциального уравнения; L_{IC} – отклонение от начальных условий; L_{BC} – отклонение от граничных условий; L_s – отклонение от экспериментальных данных.

3.5. ОБУЧЕНИЕ

Процесс обучения физически информированной нейронной сети для решения обратной задачи представляет собой итерационную процедуру совместной оптимизации параметров сети и физического коэффициента κ с использованием алгоритма Adam (Adaptive Moment Estimation).

Процесс начинается с задания начальных значений: скорости градиентного спуска η , которая определяет величину шага на каждой итерации; параметров алгоритма (смещения) $\beta_1, \beta_2 \in [0, 1]$, которые отвечают за степень влияния предыдущих значений градиентов на текущие расчеты; вектора θ_0 обучаемых переменных θ ; первого и второго моментов градиента функции $L(\theta)$: $m_0 = 0, v_0 = 0$; задается также малая величина $\varepsilon \ll 1$; для обеспечения численной стабильности.

На каждой итерации алгоритма с номером $k = 1, 2, 3, \dots$ выполняются следующие действия.

Сначала вычисляем градиент функции по обучаемым переменным – вектор, указывающий направление ее наискорейшего роста: $g_k = \nabla_{\theta} L(\theta_{k-1})$.

Затем обновляем оценки первых и вторых моментов градиента. Первый момент можно рассматривать как сглаженную траекторию градиента, а второй – как накопленную информацию о квадратах градиентов:

$$m_k = \beta_1 m_{k-1} + (1 - \beta_1) g_k, \quad v_k = \beta_2 v_{k-1} + (1 - \beta_2) g_k^2.$$

Поскольку начальные значения моментов равны нулю, на начальных этапах обучения может возникать смещение. Для его компенсации вычисляем оценки первого и второго моментов с поправкой на смещения. Эти поправки особенно важны в начале процесса обучения:

$$m_k = \frac{m_k}{(1 - \beta_1^k)}, \quad \hat{v}_k = \frac{v_k}{(1 - \beta_2^k)}.$$

Далее выполняем корректировку обучаемых переменных. Размер шага для каждого параметра адаптивно регулируется: он уменьшается для параметров

с большими градиентами и увеличивается для тех, градиенты которых были малыми:

$$\theta_k = \theta_{k-1} - \eta \frac{m_k}{\sqrt{\hat{v}_k} + \epsilon}.$$

Это позволяет эффективно работать с разреженными градиентами и ускоряет сходимость. Благодаря этому алгоритм Adam демонстрирует высокую эффективность при обучении сложных моделей, включая физически - информированные нейронные сети.

4.1. РЕЗУЛЬТАТЫ РАБОТЫ

В ходе численных экспериментов была проведена успешная идентификация безразмерного коэффициента теплопроводности κ в уравнении (1). Обучение нейронной сети проводилось в течение 25000 эпох (рис. 4). Идентифицированное значение коэффициента теплопроводности κ равно 0,44186335802078247..

Epoch	50		Loss:	2.157e-01		kappa:	0.54532
Epoch	100		Loss:	1.888e-01		kappa:	0.57604
Epoch	150		Loss:	1.238e-01		kappa:	0.64980
Epoch	200		Loss:	3.749e-02		kappa:	0.72703
Epoch	250		Loss:	1.473e-02		kappa:	0.77307
Epoch	300		Loss:	8.393e-03		kappa:	0.78232
Epoch	350		Loss:	5.443e-03		kappa:	0.77596
Epoch	400		Loss:	4.038e-03		kappa:	0.76689
Epoch	450		Loss:	3.221e-03		kappa:	0.75813
Epoch	500		Loss:	2.559e-03		kappa:	0.75054
Epoch	550		Loss:	2.447e-03		kappa:	0.74359
Epoch	600		Loss:	1.996e-03		kappa:	0.73719
Epoch	650		Loss:	1.797e-03		kappa:	0.73166
Epoch	700		Loss:	1.631e-03		kappa:	0.72660
Epoch	750		Loss:	5.789e-03		kappa:	0.72170
Epoch	800		Loss:	1.447e-03		kappa:	0.71665
Epoch	850		Loss:	1.316e-03		kappa:	0.71259
Epoch	900		Loss:	1.219e-03		kappa:	0.70863
Epoch	950		Loss:	1.130e-03		kappa:	0.70471
Epoch	1000		Loss:	2.107e-03		kappa:	0.70091
Epoch	1050		Loss:	1.026e-03		kappa:	0.69698
Epoch	1100		Loss:	9.609e-04		kappa:	0.69399
Epoch	1150		Loss:	9.045e-04		kappa:	0.69097
Epoch	1200		Loss:	8.541e-04		kappa:	0.68792
Epoch	1250		Loss:	8.090e-04		kappa:	0.68481
...							
Epoch	24900		Loss:	9.626e-06		kappa:	0.44191
Epoch	24950		Loss:	1.163e-04		kappa:	0.44197
Epoch	25000		Loss:	1.106e-05		kappa:	0.44186
Final recovered kappa: 0.44186335802078247							

Рис. 4. Обучение нейросети

Наблюдалась устойчивая сходимость метода без признаков переобучения или расходимости. В процессе обучения параметр κ демонстрировал монотонное приближение к эталонному значению.

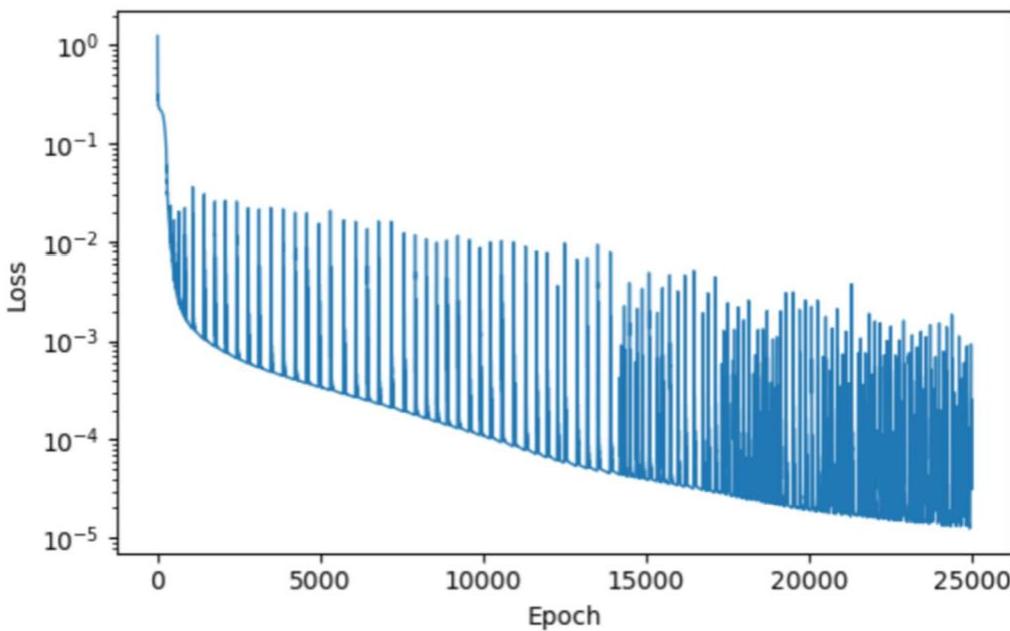


Рис. 5. График функции потерь

Представленный график (рис. 5) является наглядным доказательством эффективности подхода PINN для решения рассматриваемой обратной задачи. Модель продемонстрировала быструю сходимость и достигла высокой точности, что позволяет быть уверенным в корректности полученного решения.

Для оценки точности метода проведено сравнение с эталонным значением $\kappa_{\text{true}} = 0.442$:

$$\delta = \frac{|\kappa - \kappa_{\text{true}}|}{\kappa} * 100\% = 0.03\% .$$

ЗАКЛЮЧЕНИЕ

Успешно решена обратная задача идентификации безразмерного коэффициента теплопроводности κ в уравнении теплопроводности Грина – Нагди III типа с использованием метода ФИНС. Предложенная реализация сочетает вычислительную эффективность методов глубокого обучения с физической обоснованностью традиционных подходов к решению задач математической физики, обеспечивая устойчивое решение обратной задачи идентификации параметра.

Благодарности. Исследование выполнено за счет гранта Российского научного фонда № 23-19-00680, <https://rscf.ru/project/23-19-00680/.>)

СПИСОК ЛИТЕРАТУРЫ

1. *Smirnova V., Semenova E., Prunov V., Zamaliev R.; Sachenkov O.* Topological Approach for Material Structure Analyses in Terms of R2 Orientation Distribution Function // Mathematics. 2023. Vol. 11, No. 12. 2639.
<https://doi.org/10.3390/math11122639>
2. *Hadamard J.* Le probleme de Cauchy et les equations aux derivers partielles lineaires hyperbolique. Paris: Hermann, 1932. 542 p.
3. *Lokteva N.A., Serdyuk D.O., Skopintsev P.D.* Non-stationary influence function for an unbounded anisotropic Kirchoff–Love shell // Journal of Applied Engineering Science, 2020. Vol. 18, No. 4. P. 737–744. <https://doi.org/10.5937/jaes0-28205>
4. *Serdyuk A.O., Fedotenko G.V.* Unsteady bending function for an unlimited anisotropic plate // Vestnik Samarskogo Gosudarstvennogo Tekhnicheskogo Universiteta, Seriya Fiziko-Matematicheskie Nauki, 2021. Vol. 25, No. 1. P. 111–126.
<https://doi.org/10.14498/vsgtu1793>
5. *Орехов А.А., Рабинский Л.Н., Федотенков Г.В.* Фундаментальные решения уравнений классической и обобщенной моделей теплопроводности // Ученые записки Казанского университета. Серия Физико-математические науки. 2023. Т. 165(4). С. 404–414. <https://doi.org/10.26907/2541-7746.2023.4.404-414>
6. *Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., Killeen T., Lin Z., Gimelshein N., Antiga L., Desmaison A., Köpf A., Yang E., DeVito Z., Raison M., Tejani A., Chilamkurthy S., Steiner B., Fang L., Bai J., Chintala S.* PyTorch: An Imperative Style, High-Performance Deep Learning Library // NeurIPS. 2019.
<https://doi.org/10.48550/arXiv.1912.01703>
7. *Вахтерова Я.А., Рабинский Л.Н.* Физически информированная нейронная сеть для решения уравнения теплопроводности Грина–Нагди III типа // СТИН. 2025. №9. С. 28–32.
8. *Raissi M., Perdikaris P., Karniadakis G.E.* Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations // arXiv:1711.10561, 2017, URL: <https://arxiv.org/abs/1711.10561v1>.
<https://doi.org/10.48550/arXiv.1711.10561>
9. *Федотенков Г.В., Киреенков А.А.* Алгоритм решения контактных задач с использованием технологий глубокого машинного обучения // СТИН. 2024. № 12. С. 24–27. <https://www.elibrary.ru/joqxsg>.

10. Гончаренко В.И., Олешко В.С. Использование искусственных нейронных сетей в неразрушающем контроле деталей авиационной техники // Известия высших учебных заведений. Авиационная техника. 2024. № 3. С. 30–35.
 11. Ivanova A., Kharin N., Baltina T., and Sachenkov O., Muscle tone control system based on LIF model neural network, VIII International Conference on Information Technology and Nanotechnology (ITNT), Samara, Russian Federation, 2022, pp. 1–4. <https://doi.org/10.1109/ITNT55410.2022.9848650>
 12. Ivanova A., Kharin N., Smirnova V., Tufanova E., Sachenkov O. Stabilization of a pendulum on an elastic foundation using a multilayer perceptron// Journal of Physics: Conference Series. 2022. Vol. 2308. 012005.
<https://doi.org/10.1088/1742-6596/2308/1/012005>.
-

INVERSE PROBLEM OF IDENTIFICATION OF THERMOPHYSICAL PARAMETERS OF THE GREEN-NAGDI TYPE III MODEL FOR AN ELASTIC ROD BASED ON A PHYSICALLY INFORMED NEURAL NETWORK

Y. A. Vakhterova¹ [0000-0002-1026-649X], D. A. Leonteva²

^{1,2}Moscow Aviation Institute (National Research University), Moscow, Russia

¹yana-vahterova@mail.ru, ²dasha.leon.ra@gmail.com

Abstract

In this paper, we study the inverse problem of identifying the dimensionless thermal conductivity coefficient for the Green–Naghdi equation of type III, which describes the propagation of thermal disturbances with a finite velocity and takes into account the inertial effects of heat flux. For the inverse problem, the stability requirement (Hadamard criteria) is violated, as a result of which even minimal data distortions lead to significant errors in parameter identification. As a solution method, we use an approach based on physically informed neural networks (PINN), which combines the capabilities of deep learning with a priori knowledge of the structure of the differential equation. The parameter is included among the trained variables, and the loss function is formed based on the deviation from the differential equation, boundary conditions, initial conditions, and noisy experimental data from a point sensor. The results of computational experiments are presented, demonstrating high accuracy of parameter re-

covery (error less than 0.03%) and the stability of the method with respect to the presence of additive Gaussian noise in the data. The PINN method has proven itself to be an effective tool for solving ill-posed inverse problems of mathematical physics.

Keywords: *inverse problem, Green-Naghdi model type III, thermoelasticity, deep machine learning, physics informed neural networks.*

REFERENCES

1. Smirnova V., Semenova E., Prunov V., Zamaliev R.; Sachenkov O. Topological Approach for Material Structure Analyses in Terms of R2 Orientation Distribution Function // Mathematics. 2023. Vol. 11, No. 12. 2639.
<https://doi.org/10.3390/math11122639>
2. Hadamard J. Le probleme de Cauchy et les equations aux derivees partielles lineaires hyperbolique. Paris: Hermann, 1932. 542 p.
3. Lokteva N.A., Serdyuk D.O., Skopintsev P.D. Non-stationary influence function for an unbounded anisotropic Kirchoff–Love shell // Journal of Applied Engineering Science, 2020. Vol. 18, No. 4. P. 737–744. <https://doi.org/10.5937/jaes0-28205>
4. Serdyuk A.O., Fedotenkov G.V. Unsteady bending function for an unlimited anisotropic plate // Vestnik Samarskogo Gosudarstvennogo Tekhnicheskogo Universiteta, Seriya Fiziko-Matematicheskie Nauki, 2021. Vol. 25, No. 1. P. 111–126.
<https://doi.org/10.14498/vsgtu1793>
5. Orekhov A.A., Rabinskij L.N., Fedotenkov G.V. Fundamental'nye re-sheniya uravnenij klassicheskoy i obobshchennoj modeley teploprovodnosti // Uchenye zapiski Kazanskogo universiteta. Seriya Fiziko-matematicheskie nauki. 2023. T. 165(4). S. 404–414. <https://doi.org/10.26907/2541-7746.2023.4.404-414>
6. Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., Killeen T., Lin Z., Gimelshein N., Antiga L., Desmaison A., Köpf A., Yang E., DeVito Z., Raison M., Tejani A., Chilamkurthy S., Steiner B., Fang L., Bai J., Chintala S. PyTorch: An Imperative Style, High-Performance Deep Learning Library // NeurIPS. 2019.
<https://doi.org/10.48550/arXiv.1912.01703>
7. Vahterova YA.A., Rabinskij L.N. Fizicheski informirovannaya nejronnaya set' dlya resheniya uravneniya teploprovodnosti Grina-Nagdi III tipa // STIN. 2025. №9. S. 28–32.

8. *Raissi M., Perdikaris P., Karniadakis G.E.* Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations // arXiv:1711.10561, 2017, URL: <https://arxiv.org/abs/1711.10561v1>.
<https://doi.org/10.48550/arXiv.1711.10561>
 9. *Fedotenkov G.V., Kireenkov A.A.* Algoritm resheniya kontaktnyh za-dach s ispol'zovaniem tekhnologij glubokogo mashinnogo obucheniya // STIN. 2024. № 12. S. 24–27. <https://www.elibrary.ru/joqxsg>.
 10. *Goncharenko V.I., Oleshko V.S.* Ispol'zovanie iskusstvennyh nejron-nyh setej v nerazrushayushchem kontrole detalej aviacionnoj tekhniki // Izve-stiya vysshih uchebnyh zavedenij. Aviacionnaya tekhnika. 2024. № 3. S. 30–35.
 11. *Ivanova A., Kharin N., Baltina T., Sachenkov O.* Muscle tone control system based on LIF model neural network // VIII International Conference on Information Technology and Nanotechnology (ITNT), Samara, Russian Federation, 2022. P. 1–4. <https://doi.org/10.1109/ITNT55410.2022.9848650>
 12. *Ivanova A., Kharin N., Smirnova V., Tufanova E., Sachenkov O.* Stabilization of a pendulum on an elastic foundation using a multilayer perceptron// Journal of Physics: Conference Series. 2022. Vol. 2308. 012005.
<https://doi.org/10.1088/1742-6596/2308/1/012005>.
-

СВЕДЕНИЯ ОБ АВТОРАХ



ВАХТЕРОВА Яна Андреевна – к. ф.-м. н., доцент кафедры 902 «Сопротивление материалов, динамика и прочность машин» Московского авиационного института. Научные интересы: механика деформируемого твердого тела (динамика, термоупругость, волновые задачи, упругое полупространство); обратные нестационарные задачи идентификации параметров, нагрузок и источников по частичным измерениям; физически информированные нейронные сети для решения прямых и обратных нестационарных задач механики

Yana Andreevna VAKTEROVA – Candidate of Physical and Mathematical Sciences, Associate Professor of Department 902 "Strength of Materials, Dynamics and Strength of Machines" of the Moscow Aviation Institute. Research interests: mechanics of deformable solids (dynamics, thermoelasticity, wave problems, elastic half-space); inverse non-stationary problems of identifying parameters, loads and sources from partial measurements; physically informed neural networks for solving direct and inverse non-stationary problems of mechanics

email: yana-vahterova@mail.ru

ORCID: 0000-0002-1026-649X



ЛЕОНТЬЕВА Дарья Андреевна – студентка 4 курса бакалавриата Института №6 «Аэрокосмический», профиль обучения «Прикладная механика» Московского авиационного института, техник кафедры 902 «Сопротивление материалов, динамика и прочность машин» Московского авиационного института

Darya Andreevna LEONTYEVA – 4th year student of the bachelor's degree program at the Institute No. 6 "Aerospace", majoring in "Applied Mechanics" at the Moscow Aviation Institute, technician at the Department 902 "Strength of Materials, Dynamics and Strength of Machines" at the Moscow Aviation Institute

email: dasha.leon.ra@gmail.com

Материал поступил в редакцию 8 сентября 2025 года

ЭМПИРИЧЕСКИЕ АНАЛОГИ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ С ГАРАНТИРОВАННЫМ ВЫВОДОМ

Э. А. Заарур¹ [0009-0008-8585-4174], С. В. Симушкин² [0000-0003-2544-0267]

^{1, 2}Казанский федеральный университет, г. Казань, Россия

¹zrwrz05@gmail.com, ²smshkn@gmail.com

Аннотация

Для построения гарантийных процедур различия двух односторонних гипотез применены методы ядерного оценивания априорной плотности в задаче деконволюции. Рассмотрена ситуация, когда наблюдаемая случайная величина представляет собой сумму неизвестного параметра и центрированной нормальной ошибки с известной дисперсией. Построены состоятельные эмпирические оценки для функции d-апостериорного риска. Установлена сходимость соответствующей критической константы к оптимальному значению. Точность процедур проиллюстрирована численно на трех вариантах априорного распределения.

Ключевые слова: эмпирический байесовский подход, проблема деконволюции, гарантированный статистический вывод, d-апостериорный подход.

ВВЕДЕНИЕ

Пусть в эксперименте наблюдается выборка $X^{(n)} = (X_1, \dots, X_n)$, $n \geq 1$, независимых нормальных $\mathcal{N}(\theta, \sigma^2)$ случайных величин. По результатам наблюдений $X^{(n)} = x^{(n)}$ требуется проверить гипотезу $H_0: \theta \leq \theta_0$ против альтернативы $H_1: \theta > \theta_0$. В ситуации, когда задача проверки гипотез возникает многократно, можно предположить, что значение параметра θ есть реализация случайной величины ϑ с некоторым априорным распределением. Здесь мы рассматриваем случай, когда ϑ имеет абсолютно непрерывное распределение относительно меры Лебега с плотностью $g(\theta)$, $\theta \in R^1$. Качество решений будет оценено посредством так называемого d-апостериорного риска.

Пусть $\delta = \delta(x^{(n)})$, $x^{(n)} \in R^n$, есть решающее правило, принимающее решение d_0 в пользу гипотезы H_0 или решение d_1 в пользу гипотезы H_1 . d-апостериорный риск первого рода решающего правила δ определяется как

средняя доля ошибочных решений среди экспериментов, закончившихся принятием нулевой гипотезы:

$$\mathcal{R}_0(c; g) = \mathbb{P}(\vartheta > \theta_0 \mid \delta(X^{(n)}) = d_0).$$

Решающее правило должно быть гарантийным, т. е. должно удовлетворять заданному ограничению $\mathcal{R}_0(c; g) \leq \beta$ на величину d-апостериорного риска первого рода (см. [1, 2]). Как всегда, величина β выбирается исследователем.

Поскольку нормальная вероятностная модель имеет монотонное отношение правдоподобия, оптимальное гарантийное решающее правило, которое минимизирует d-апостериорный риск второго рода

$$\mathcal{R}_1(c; g) = \mathbb{P}(\vartheta \leq \theta_0 \mid \delta(X^{(n)}) = d_1),$$

принимает решение в пользу H_0 , когда выборочное среднее $\bar{x} \leq c$, где критическая константа находится из условия $\mathbb{P}(\vartheta > \theta_0 \mid \bar{X} \leq c) = \beta$ (см. [1, 2]). Эта константа зависит от априорной плотности. Чтобы подчеркнуть этот факт, будем обозначать решение предыдущего уравнения через $c(g)$.

Обозначим через Φ функцию распределения стандартного нормального $\mathcal{N}(0,1)$ закона, а через φ – ее функцию плотности. Известно, что при фиксированном θ выборочное среднее \bar{X} имеет нормальное $\mathcal{N}(\theta, \sigma^2/n)$ распределение. В настоящей статье мы не касаемся вопросов планирования эксперимента, поэтому в дальнейшем будем считать, что в эксперименте наблюдается случайная величина $X \sim \mathcal{N}(\theta, \sigma^2)$.

Кроме того, мы рассматриваем только такие априорные плотности, которые сосредоточены на ограниченном открытом интервале (A, B) числовой прямой: $g(\theta) = 0, \theta \notin (A, B)$.

Таким образом, d-риск первого рода оптимального решения можно записать как $\mathcal{R}_0(c; g) = Q(c; g)/F(c; g)$, где

$$Q(c; g) = \mathbb{P}(\vartheta > \theta_0, X \leq c) = \int_{\theta_0}^B \Phi\left(\frac{c - \theta}{\sigma}\right) g(\theta) d\theta, \quad (1)$$

$F(c; g)$ – безусловная функция распределения X :

$$F(c; g) = \int_A^B \Phi\left(\frac{c - \theta}{\sigma}\right) g(\theta) d\theta. \quad (2)$$

Как уже отмечалось, оптимальное решение и, в частности, константа $c(g)$ зависят от вида априорной плотности g . Эту плотность можно оценить, если имеется архив предварительных наблюдений. Подобная идея близка по духу к эмпирическому байесовскому подходу, предложенному Г. Роббинсом. Развитие этой идеи применительно к d-апостериорному подходу впервые было осуществлено в работе [3]. Если вид априорной плотности известен с точностью до некоторых параметров, то проблема решается обычным способом. Например, если g также нормальная плотность, то здесь вполне уместны классические оценки параметров модели II дисперсионного анализа со случайными факторами [4].

Задача оценивания плотности g в описанной выше ситуации по архиву наблюдений $\Xi_k = (x_1, \dots, x_k)$ с общей функцией распределения F известна как проблема деконволюции (см. [5–7]). Наша цель состоит в демонстрации возможностей применения методов деконволюции к построению гарантийного решающего правила. В работе [8] нами рассмотрен вопрос построения состоятельного эмпирического аналога для байесовской оценки в рамках предлагаемой вероятностной модели на основе ядерной оценки априорной плотности. Эти аналоги были построены как для случая известной дисперсии наблюдений, так и для случая, когда дисперсию наблюдений также приходилось оценивать по архивным данным (методику такого оценивания можно найти в [9]).

НЕКОТОРЫЕ ВСПОМОГАТЕЛЬНЫЕ ФАКТЫ

1. Свойства функции d-апостериорного риска

В работе [1] было показано, что функция $\mathcal{R}_0(c; g)$ удовлетворяет следующим свойствам.

Лемма 1. При любой плотности g функция $\mathcal{R}_0(c; g), c \in R$,
(a) непрерывна,
(b) возрастает, (c) $\lim_{c \rightarrow -\infty} \mathcal{R}_0(c; g) = 0$, (d) $\lim_{c \rightarrow \infty} \mathcal{R}_0(c; g) = \Pi_1 = P\{\vartheta > \theta_0\}$.

Следовательно, для любого $0 < \beta < \Pi_1$ всегда найдется единственное $c = c(g)$ такое, что $\mathcal{R}_0(c(g); g) = \beta$. В дальнейшем будем предполагать, что для истинного априорного распределения это условие выполнено.

2. Оценка априорной плотности

Подходящая для наших целей ядерная оценка g описана в статье [5]. Из представления (2) видно, что безусловное распределение X есть свертка с плотностью $f = g * h$, где плотность h совпадает с плотностью нормального $\mathcal{N}(0, \sigma^2)$ закона. Оценка g по набору независимых наблюдений Ξ_k , имеющих общую плотность f , строится следующим образом.

Пусть $\psi_g(t)$ – характеристическая функция случайной величины ϑ (преобразование Фурье функции плотности g), $\psi_h(t) = \exp(-t^2/2)$ – характеристическая функция h . Оценка \hat{g}_k для функции плотности g имеет следующий вид:

$$\hat{g}_k(\theta, \Xi_k) = \frac{1}{k\pi} \sum_1^k \int_0^{1/\lambda} \frac{1}{\psi_h(t)} K(t\lambda) \cos\{t(x_j - \theta)\} dt. \quad (3)$$

Заметим, что функция ψ_h удовлетворяет условиям, описанным в [5].

Теорема 1 ([5], Теорема 3.1). *Если $\int_{-\infty}^{\infty} |\psi_g(t)| dt < \infty$, функция $g(\theta)$ непрерывна и ограничена, то при $\lambda_k \rightarrow 0$, $k \rightarrow \infty$, имеет место*

$$\mathbb{E}\left(\sup_{-\infty < \theta < \infty} |\hat{g}_k(\theta, \Xi_k) - g(\theta)|\right) \rightarrow 0.$$

Параметр $\lambda = \lambda_k > 0$ выбирается в зависимости от объема k архива Ξ .

В работе [6] предложено взять $\lambda_k = \sqrt{2/\ln k}$. Ядро K можно выбирать произвольным образом. Как показано в [6], точность аппроксимации зависит от гладкости функции $K(\theta)$ в точке $\theta = 0$.

СОСТОЯЛЬНОСТЬ ЭМПИРИЧЕСКИХ АНАЛОГОВ РИСКА

Используем теперь результат теоремы 1 для построения оценки функции d -риска 1-го рода $\mathcal{R}_0(c; g)$ и изучим ее свойства. Определим константу $c(\hat{g}_k)$ как решение уравнения $\mathcal{R}_0(c; \hat{g}_k) = \beta$, если $\Pi_{1k} = \int_{\theta_0}^B \hat{g}_k(\theta, \Xi_k) d\theta > \beta$, и $c(\hat{g}_k) = B$ в противном случае.

Теорема 2. *Если функция плотности g финитная, то есть непрерывная, принимающая нулевое значение вне некоторого конечного интервала $I = [A, B]$, то при $k \rightarrow \infty$ по вероятности*

i) $\sup_{c \in I} |\mathcal{R}_0(c; \hat{g}_k) - \mathcal{R}_0(c; g)| \xrightarrow{P} 0;$

$$\text{ii)} \quad c(\hat{g}_k) \xrightarrow{P} c(g).$$

Доказательство. В условиях теоремы при вычислениях интегралов, определяющих д-апостериорный риск, область интегрирования можно ограничить интервалом I . В силу (1) для функции $Q(c; g)$ имеем

$$\begin{aligned} \sup_{c \in I} |Q(c; \hat{g}_k) - Q(c; g)| &\leq \sup_{c \in I} \int_{\theta_0}^B \Phi\left(\frac{c - \theta}{\sigma}\right) |\hat{g}_k(\theta, \Xi_k) - g(\theta)| d\theta \\ &\leq (B - \theta_0) \sup_{\theta \in [\theta_0, B]} |\hat{g}_k(\theta, \Xi_k) - g(\theta)| \xrightarrow{P} 0, \end{aligned}$$

так как сходимость в среднем влечет сходимость по вероятности.

Аналогично для функции распределения (2):

$$\sup_{c \in I} |F(c; \hat{g}_k) - F(c; g)| \leq (B - A) \sup_{\theta \in [A, B]} |\hat{g}_k(\theta, \Xi_k) - g(\theta)|.$$

Рассмотрим константу $l = \inf_{c \in I} F(c)$. Из формулы (2) следует, что $l > 0$.

Для каждого $\varepsilon > 0$ справедливо следующее неравенство:

$$\begin{aligned} P \left\{ \sup_{c \in I} \left| \frac{\hat{Q}(c)}{\hat{F}(c)} - \frac{Q(c)}{F(c)} \right| > \varepsilon \right\} &\leq P \left\{ \sup_{c \in I} |\hat{F}(c) - F(c)| > \varepsilon \right\} \\ &+ P \left\{ \sup_{c \in I} \left(\left| \frac{\hat{Q}(c)}{\hat{F}(c)} \hat{F}(c) - \frac{Q(c)}{F(c)} \hat{F}(c) \right| \right) > \varepsilon(l - \varepsilon) \right\} \end{aligned}$$

Действительно, если мы определим $\Delta(c) = \left| \frac{\hat{Q}(c)}{\hat{F}(c)} - \frac{Q(c)}{F(c)} \right|$, то из неравенства

$\sup_{c \in I} \Delta(c) > \varepsilon$ следует, что

$$\begin{aligned} \sup_{c \in I} (\Delta(c) \hat{F}(c)) &> \varepsilon \inf_{c \in I} \hat{F}(c) = \varepsilon \inf_{c \in I} (\hat{F}(c) - F(c) + F(c)) \\ &\geq \varepsilon \inf_{c \in I} (F(c)) - \varepsilon \sup_{c \in I} |\hat{F}(c) - F(c)| \geq \varepsilon l - \varepsilon^2, \end{aligned}$$

где последнее неравенство выполняется, если $\sup_{c \in I} |\hat{F}(c) - F(c)| \leq \varepsilon$.

Заметим, что

$$\begin{aligned} \sup_{c \in I} (\Delta(c) \hat{F}(c)) &\leq \sup_{c \in I} |\hat{Q}(c) - Q(c)| + \sup_{c \in I} \left(\left| \frac{Q(c)}{F(c)} \hat{F}(c) - Q(c) \right| \right) \\ &\leq \sup_{c \in I} |\hat{Q}(c) - Q(c)| + \sup_{c \in I} \left| \frac{Q(c)}{F(c)} \right| \sup_{c \in I} |\hat{F}(c) - F(c)|, \end{aligned}$$

Следовательно, неравенство $\sup_{c \in I} (\Delta(c) \hat{F}(c)) > \varepsilon(l - \varepsilon)$ влечет

$$\sup_{c \in I} |\hat{F}(c) - F(c)| > \frac{\varepsilon(l - \varepsilon)}{2} \text{ или } \sup_{c \in I} |\hat{Q}(c) - Q(c)| > \frac{\varepsilon(l - \varepsilon)}{2}.$$

Если зафиксировать $0 < \varepsilon < l$, то при $\gamma = \min(\varepsilon, \frac{\varepsilon(l - \varepsilon)}{2})$ получим

$$\begin{aligned} P \left\{ \sup_{c \in I} \left| \frac{\hat{Q}(c)}{\hat{F}(c)} - \frac{Q(c)}{F(c)} \right| > \varepsilon \right\} &\leq 2P \left\{ \sup_{c \in I} |\hat{F}(c) - F(c)| > \gamma \right\} \\ &+ P \left\{ \sup_{c \in I} |\hat{Q}(c) - Q(c)| > \gamma \right\}. \end{aligned}$$

Как было доказано ранее,

$$P \left\{ \sup_{c \in I} |\hat{Q}(c) - Q(c)| > \gamma \right\} \rightarrow 0 \text{ и } P \left\{ \sup_{c \in I} |\hat{F}(c) - F(c)| > \gamma \right\} \rightarrow 0,$$

следовательно, $P \left\{ \sup_{c \in I} \left| \frac{\hat{Q}(c)}{\hat{F}(c)} - \frac{Q(c)}{F(c)} \right| > \varepsilon \right\} \rightarrow 0$. Утверждение i) доказано.

Перейдем к доказательству ii). Выберем $\varepsilon > 0$. Так как функция риска $\mathcal{R}_0(c; g)$ непрерывно возрастает по c и $\mathcal{R}_0(c(g); g) = \beta$, найдется $\gamma > 0$ такое, что $\mathcal{R}_0(c; g) > \beta + \gamma$ для всех $c > c(g) + \varepsilon$. По свойствам функции \mathcal{R}_0 в случае $\Pi_{1k} < \beta$ выполняется условие $\mathcal{R}_0(B; \hat{g}_k) < \beta$, то есть всегда $\mathcal{R}_0(c(\hat{g}_k); \hat{g}_k) \leq \beta$. Следовательно, событие $c(\hat{g}_k) > c(g) + \varepsilon$ влечет

$$\mathcal{R}_0(c(\hat{g}_k); g) > \beta + \gamma \geq \mathcal{R}_0(c(\hat{g}_k); \hat{g}_k) + \gamma.$$

Поэтому

$$\begin{aligned} P\{c(\hat{g}_k) > c(g) + \varepsilon\} &\leq P\{\mathcal{R}_0(c(\hat{g}_k); g) - \mathcal{R}_0(c(\hat{g}_k); \hat{g}_k) > \gamma\} \\ &\leq P\left\{\sup_{c \in I} |\mathcal{R}_0(c; g) - \mathcal{R}_0(c; \hat{g}_k)| > \gamma\right\} \rightarrow 0 \end{aligned}$$

в силу утверждения i). Докажем теперь, что $P\{c(\hat{g}_k) < c(g) - \varepsilon\} \rightarrow 0$.

Из теоремы 1 следует также, что

$$\Pi_{1k} = \int_{\theta_0}^B \hat{g}_k(\theta, \Xi_k) d\theta \xrightarrow{P} \Pi_1$$

Так как $\beta - \Pi_1 < 0$, то,

$$P\{\Pi_{1k} < \beta\} \leq P\{\Pi_{1k} - \Pi_1 < \beta - \Pi_1\} \rightarrow 0.$$

При этом, если $\Pi_{1k} < \beta$, то по построению $\mathcal{R}_0(c(\hat{g}_k); g) = \beta$.

Выберем $\gamma > 0$ так, чтобы $\mathcal{R}_0(c; g) < \beta - \gamma$ для всех $c < c(g) - \varepsilon$.

Если $\Pi_{1k} < \beta$, то событие $c(\hat{g}_k) < c(g) + \varepsilon$ влечет

$$\mathcal{R}_0(c(\hat{g}_k); g) < \beta - \gamma = \mathcal{R}_0(c(\hat{g}_k); \hat{g}_k) - \gamma.$$

Поэтому

$$\begin{aligned} P\{c(\hat{g}_k) < c(g) - \varepsilon\} &\leq P\{\mathcal{R}_0(c(\hat{g}_k); g) - \mathcal{R}_0(c(\hat{g}_k); \hat{g}_k) < -\gamma\} \\ &+ P\{\Pi_{1k} < \beta\} \rightarrow 0. \end{aligned}$$

Теорема доказана.

ЧИСЛЕННЫЕ ИЛЛЮСТРАЦИИ

Рассмотрим несколько примеров, в которых реализованы описанные выше методы. В части этих примеров априорная плотность не будет удовлетворять условию на носитель. Такие плотности выбраны специально, чтобы проанализировать возможность будущего обобщения теоремы 2. В этих примерах каждая вероятностная модель априорного распределения, в принципе, также имеет некоторый набор гиперпараметров. Модельные испытания были осуществлены для конкретных значений этих гиперпараметров

Пример 1. Пусть параметр θ имеет априорное экспоненциальное распределение с плотностью $g(\theta) = e^{-\theta}, \theta > 0$. Найдем аналитическое выражение для функции d-апостериорного риска $\mathcal{R}_0(c; g)$. Вид числителя $\mathcal{R}_0(c; g)$ после интегрирования по частям и простых алгебраических преобразований приобретает следующий вид:

$$\begin{aligned} Q(c, g) &= \int_{\theta_0}^{\infty} \Phi\left(\frac{c - \theta}{\sigma}\right) e^{-\theta} d\theta = \Phi\left(\frac{c - \theta_0}{\sigma}\right) e^{-\theta_0} + \int_{\theta_0}^{\infty} \varphi\left(\frac{c - \theta}{\sigma}\right) e^{-\theta} d\theta \\ &= \Phi\left(\frac{c - \theta_0}{\sigma}\right) e^{-\theta_0} - e^{-c + \frac{\sigma^2}{2}} \Phi\left(\frac{c - \theta_0}{\sigma} - \sigma\right). \end{aligned}$$

Подставив сюда $\theta_0 = 0$, найдем функцию распределения $F(c, g)$

$$F(c, g) = \int_0^{\infty} \Phi\left(\frac{c - \theta}{\sigma}\right) g(\theta) d\theta = \Phi\left(\frac{c}{\sigma}\right) - e^{-c + \frac{\sigma^2}{2}} \Phi\left(\frac{c}{\sigma} - \sigma\right).$$

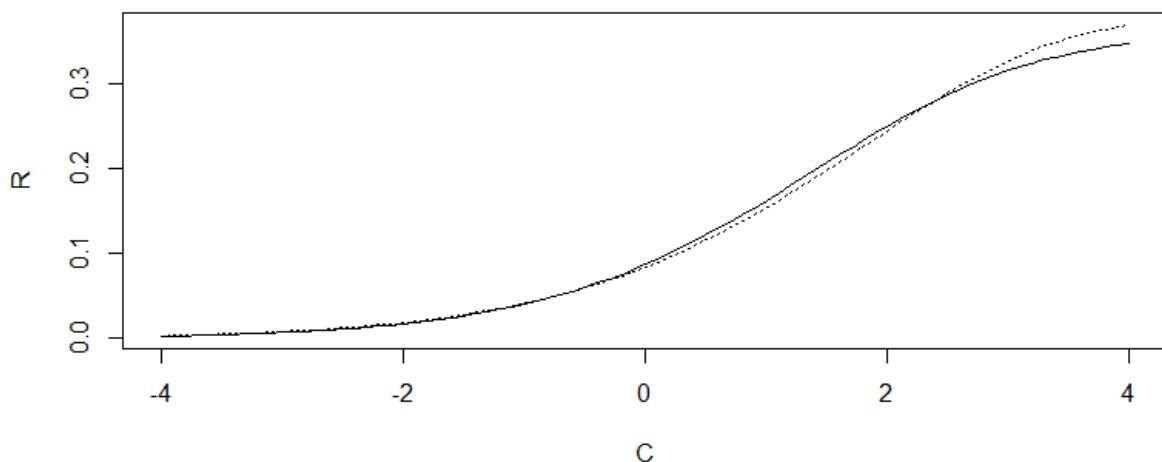


Рис. 1. Оценки функции риска при экспоненциальном априори.

Итак, функция d-апостериорного риска 1-го рода имеет следующее представление:

$$\mathcal{R}_0(c; g) = \frac{\exp(-\theta_0)\Phi((c - \theta_0)/\sigma) - \exp(-c + \frac{\sigma^2}{2})\Phi((c - \theta_0)/\sigma - \sigma)}{\Phi(c/\sigma) - \exp(-c + \frac{\sigma^2}{2})\Phi(c/\sigma - \sigma)}.$$

Модельные испытания проводились для $\sigma = 1$ с границей $\theta_0 = 1$.

На рис. 1 (и во всех последующих для других примеров) приведены графики истинной функции риска (сплошная линия) и функции риска с оценкой априорной плотности, построенной по $k = 700$ архивным данным (пунктирная линия). В этом примере использовано равномерное ядро $K(t) = 1$. Область изменения θ при вычислениях интегралов была выбрана из условия наилучшего приближения ядерной оценки функции распределения к эмпирической функции распределения; в данном случае была взята область $\theta \in [0,5]$.

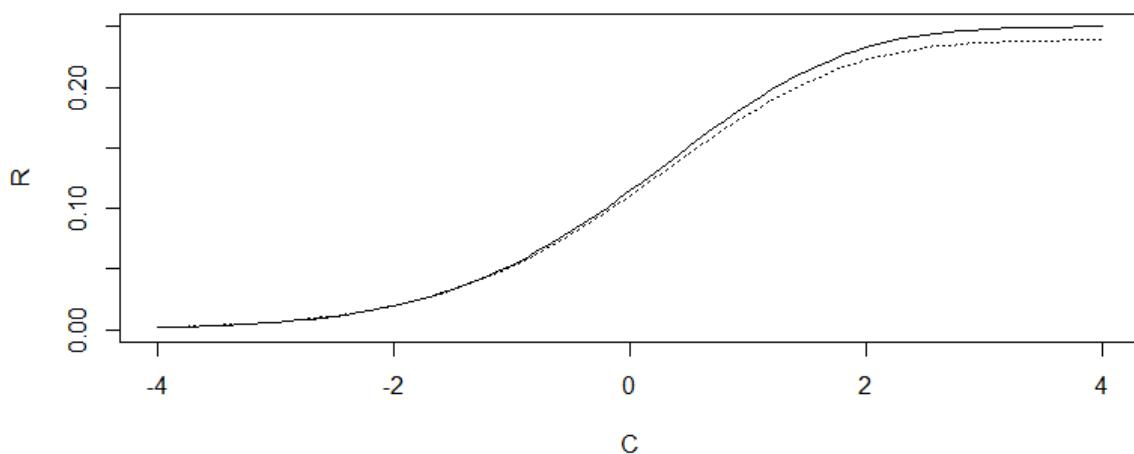


Рис. 2. Оценка функции риска для равномерного априори.

Пример 2. Рассмотрим теперь априорное распределение с равномерной плотностью $g(\theta) = \frac{1}{2}$, $\theta \in [-1,1]$. Функция риска имеет следующий вид:

$$\mathcal{R}_0(c; g) = \int_{\theta_0}^1 \Phi\left(\frac{c-\theta}{\sigma}\right) d\theta / \int_{-1}^1 \Phi\left(\frac{c-\theta}{\sigma}\right) d\theta.$$

Заметим, что производная нормальной плотности $\varphi'(x) = -x\varphi(x)$. Следовательно, $\int \Phi(x) dx = x\Phi(x) + \varphi(x) + \text{Const}$. Введем обозначение $\Phi i(x) = x\Phi(x) + \varphi(x)$. Функция риска равна

$$\mathcal{R}_0(c; g) = \frac{\Phi((c - \theta_0)/\sigma) - \Phi\left(\frac{(c - 1)}{\sigma}\right)}{\Phi((c + 1)/\sigma) - \Phi((c - 1)/\sigma)}.$$

Как и в предыдущем примере, модельные испытания проводились для $\sigma = 1$ с границей $\theta_0 = 0.5$.

Пример 3. В заключительном примере рассмотрена априорная плотность, состоящая из смеси двух нормальных плотностей с общей единичной дисперсией и разными математическими ожиданиями λ_1, λ_2 соответственно:

$$g(\theta) = p\varphi(\theta - \lambda_1) + (1 - p)\varphi(\theta - \lambda_2),$$

где φ – плотность стандартного нормального (0,1) распределения, $0 < p < 1$.

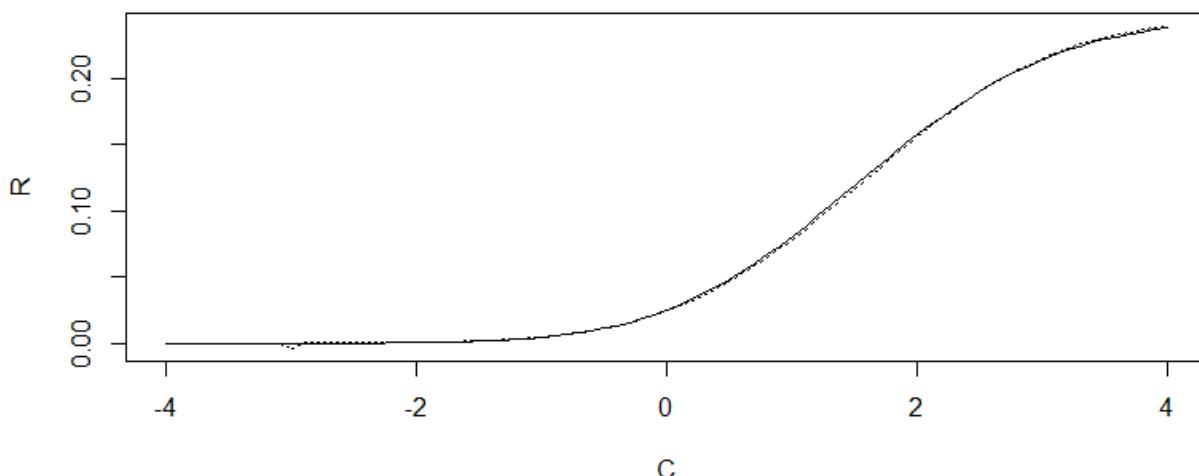


Рис. 3. Оценка функции риска с априори в виде смеси нормальных.

Выберем границу между гипотезами $\theta_0 = 1$. Найдем функцию распределения X , то есть знаменатель \mathcal{R}_0 . Выражение $\int_{-\infty}^{\infty} \Phi\left(\frac{c-\theta}{\sigma}\right) \varphi(\theta - \lambda)d\theta$ есть функция распределения свертки двух нормальных распределений, поэтому

$$\int_{-\infty}^{\infty} \Phi\left(\frac{c-\theta}{\sigma}\right) g(\theta)d\theta = p\Phi\left(\frac{c - \lambda_1}{\sqrt{1 + \sigma^2}}\right) + (1 - p)\Phi\left(\frac{c - \lambda_2}{\sqrt{1 + \sigma^2}}\right).$$

Числитель $Q(c; g)$ функции риска $\mathcal{R}_0(c; g)$ в данном случае не имеет простого представления. При модельных испытаниях этот числитель находился приближенными методами. Итак,

$$\mathcal{R}_0(c; g) = \frac{\int_1^\infty \Phi\left(\frac{c-\theta}{\sigma}\right) (p\varphi(\theta - \lambda_1) + (1-p)\varphi(\theta - \lambda_2)) d\theta}{p\Phi\left(\frac{c-\lambda_1}{\sqrt{1+\sigma^2}}\right) + (1-p)\Phi\left(\frac{c-\lambda_2}{\sqrt{1+\sigma^2}}\right)}.$$

В этом примере значения параметров при модельных испытаниях $\sigma^2 = 1$, $p = 0.4$, $\lambda_1 = 1.2$, $\lambda_2 = -1$. Область интегрирования $\theta \in [-4, 4]$. Графики функций риска приведены на рис. 3

ЗАКЛЮЧЕНИЕ

Эмпирические байесовские методы, использующие ядерную оценку априорной плотности, вполне уместны при решении задач построения гарантийных процедур в d-апостериорном подходе. Проведенные модельные испытания при разных априорных предположениях показали, что, в принципе, требуется достаточно большой объем архивных данных k . В статье не описаны осуществленные попытки построения оценок при малых $k \leq 100$, которые показали не очень хорошие результаты. Кроме того, эти испытания убеждают, что утверждения о состоятельности оценок риска и критической константы, скорее всего, справедливы для более широкого класса априорных плотностей. В частности, сравнение эмпирических аналогов трех моделей показало, что наилучший результат достигается для смеси нормальных распределений.

Благодарности

Работа выполнена за счет средств Программы стратегического академи-ческого лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»).

СПИСОК ЛИТЕРАТУРЫ

1. Simushkin D.S., Simushkin S.V., Volodin I.N. On the d-posterior approach to the multiple testing problem // Journal of Statistical Computation and Simulation. 2021. Vol. 91, No. 4. P. 651–666.

2. Симушкин С.В. Оптимальные d-гарантийные процедуры различия двух гипотез// Деп. ВИНИТИ АН ССР. 1981, № 5547-81. 47 с.
 3. Симушкин С.В. Эмпирический d-апостериорный подход к проблеме гарантированности статистического вывода// Известия ВУЗов. Математика. 1983, № 11. С. 42–58.
 4. Шеффе Г. Дисперсионный анализ. М.: «Физматлит», 1980. 512 с.
 5. Liu M.C., Taylor R.L. A consistent nonparametric density estimator for the deconvolution problem // Canadian Journal of Statistics. 1989. Vol. 17, No. 4. P. 427–438.
 6. Carroll R.J., Hall P. Optimal rates of convergence for deconvolving a density // J. Am. Stat. Assoc. 1988. Vol. 83, No. 404. P. 1184–1186.
 7. Stefanski L.A., Carroll R.J. Deconvolving kernel density estimators // Statistics, 1990. Vol. 21, No. 2. P. 169–184.
 8. Zaarour E., Simushkin S.V. Consistency of the Empirical Bayesian Analogue of the Regression Estimation // Lobachevskii Journal of Mathematics. 2024. Vol. 45, No. 1. P. 551–554.
 9. Meister A. Density estimation with normal measurement error with unknown variance // Statistica Sinica. 2006, Vol. 16. P. 195–211.
-

EMPIRICAL ANALOGUES OF STATISTICAL TESTS WITH GUARANTEED CONCLUSION

E. A. Zaarour¹ [0009-0008-8585-4174], S. V. Simushkin² [0000-0003-2544-0267]

^{1, 2}Kazan Federal University, Kazan, Russia

¹zrwrz05@gmail.com, ²smshkn@gmail.com

Abstract

Methods of kernel estimation of a priori density in the deconvolution problem are used to construct guaranteed procedures for distinguishing between two one-sided hypotheses. The situation is considered when the observed random variable is the sum of an unknown parameter and a centered normal error with a known variance. Consistent empirical estimates are constructed for the d-posterior risk function. The convergence of the corresponding critical constant to the optimal value is established.

The accuracy of the procedures is illustrated numerically on three variants of the prior distribution.

Keywords: *Empirical Bayesian approach, deconvolution problem, guaranteed statistical inference, d-posterior approach.*

REFERENCES

1. *Simushkin D.S., Simushkin S.V., Volodin I.N.* On the d-posterior approach to the multiple testing problem // *Journal of Statistical Computation and Simulation*. 2021. Vol. 91, No. 4. P. 651-666.
 2. *Simushkin D.S.* Optimal d-guarantee procedures for distinguishing two hypotheses // Dep. VINITI AN USSR. 1981, № 5547-81. 47 p.
 3. *Simushkin D.S.* Empirical d- posterior approach to the problem of guarantee of statistical inference // *Izvestiya VUZov. Mathematics*. 1983, № 11. P. 42–58.
 4. *Scheffe G.* The analysis of variance. N-Y.: J. Wiley & Sons, 1980. 512 p.
 5. *Liu M.C., Taylor R.L.* A consistent nonparametric density estimator for the deconvolution problem // *Canadian Journal of Statistics*. 1989. Vol. 17, No. 4. P. 427–438.
 6. *Carroll R.J., Hall P.* Optimal rates of convergence for deconvolving a density // *J. Am. Stat. Assoc.* 1988. Vol. 83, No. 404. P. 1184–1186.
 7. *Stefanski L.A., Carroll R.J.* Deconvolving kernel density estimators // *Statistics*, 1990. Vol. 21, No. 2. P. 169–184.
 8. *Zaarour E., Simushkin S.V.* Consistency of the Empirical Bayesian Analogue of the Regression Estimation // *Lobachevskii Journal of Mathematics*. 2024. Vol. 45, No. 1. P. 551–554.
 9. *Meister A.* Density estimation with normal measurement error with unknown variance // *Statistica Sinica*. 2006, Vol. 16. P. 195–211.
-

СВЕДЕНИЯ ОБ АВТОРАХ



СИМУШКИН Сергей Владимирович – 1956 года рождения, учился в Казанском государственном университете. В настоящее время работает доцентом на кафедре математической статистики К(П)ФУ.

Область научных интересов: гарантийные процедуры статистического вывода-проверка гипотез, точечное и доверительное оценивание, процедуры множественного тестирования в байесовской парадигме.

Sergey Vladimirovich SIMUSHKIN – born in 1956, studied at Kazan State University. Currently, he works as an associate professor at the Department of Mathematical Statistics of Kazan Federal University.

Research interests: optimal guarantee procedures for ensuring statistical inference – hypothesis testing, point and confidence estimation, multiple testing procedures in the Bayesian paradigm.

email: smshkn@gmail.com

ORCID: 0000-0003-2544-0267



ЗААРУР Эзеддин Абдулмуин – родился в Сирии в 1995 году, аспирант четвертого курса по кафедре математической статистики Казанского федерального университета.

Область научных интересов: регрессионный анализ, байесовская регрессия, оптимальные статистические процедуры, программирование.

Ezeddeen Abdulmuneen ZAAROUR – born in Syria in 1995, fourth-year postgraduate student at Kazan Federal University, Department of Mathematical Statistics, specialty 1.1.4 Probability Theory and Mathematical Statistics.

Research interests: regression analysis, Bayesian regression, optimal statistical procedures – hypothesis testing, programming.

email: zrwrz05@gmail.com

ORCID: 0009-0008-8585-4174

Материал поступил в редакцию 25 мая 2025 года

ЦИФРОВОЙ ДВОЙНИК ПАРКОВОЧНОГО ПРОСТРАНСТВА

Р. Н. Минниханов¹ [0000-0001-9166-2955], Т. Р. Баторшин² [0009-0001-9995-8854],

Р. М. Габбазов³ [0009-0009-9187-9121], Р. И. Фахразиев⁴ [0009-0008-3463-7663],

А. С. Катасёв⁵ [0000-0002-9446-0491], М. В. Дагаева⁶ [0000-0002-5444-9669],

И. Р. Бадрутдинов⁷ [0009-0001-6343-4709]

¹Академия наук Республики Татарстан, г. Казань, 420111, Россия

^{1, 6}ГБУ «Безопасность дорожного движения», г. Казань, 420059, Россия

²Казанский государственный энергетический университет, г. Казань, 420066, Россия

^{3-5, 7}Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, г. Казань, 420111, Россия

¹Priemnaya.Minnihanov@tatar.ru, ²Timur.Batorshin.02@gmail.com,

³GabbazovRM@stud.kai.ru, ⁴FakhrazievRI@stud.kai.ru, ⁵ASKatasev@kai.ru,

⁶ITS.Center.kzn@gmail.com, ⁷BadrutdinovIR@stud.kai.ru

Аннотация

Рост уровня урбанизации и автомобилизации приводит к дефициту парковочных мест, что вызывает заторы, повышение выбросов и снижение качества жизни. Традиционные методы организации парковочного пространства не обеспечивают эффективного решения этой проблемы, что требует применения инструментов анализа данных и прогнозирования.

В работе рассмотрено использование цифрового двойника парковочной системы города Казани. Проведены фильтрация и интеграция данных, выполнены кластеризация точек интереса и корреляционный анализ факторов, влияющих на заполняемость парковок. Для прогнозирования уровня загруженности обучены и сравнены модели линейной регрессии, дерева решений, случайного леса, XGBoost, MLP и LSTM. Наилучшие результаты продемонстрировала модель случайного леса. Разработанный прототип цифрового двойника обеспечивает мониторинг и сценарное моделирование, что делает его эффективным инструментом для оптимизации парковочного пространства и принятия управленческих решений.

Ключевые слова: цифровой двойник, парковки, заполняемость парковочных мест, машинное обучение, точки интереса, оптимизация.

ВВЕДЕНИЕ

Быстрый рост уровня урбанизации и автомобилизации в современных городах привел к серьезным проблемам транспортной инфраструктуры, которые проявляются в виде повышенной нагрузки на парковочное пространство. Дефицит парковочных мест способствует обострению заторов на дорогах, росту экологического загрязнения и снижению мобильности населения, что отрицательно сказывается на качестве жизни и вызывает недовольство горожан.

Классические подходы к организации парковочного пространства, такие как формирование платных зон, точечное строительство новых парковок и корректировка тарифной политики, лишь частично смягчают проблему и не гарантируют оптимального решения. Увеличение числа парковочных мест может привести к дополнительным расходам, а необоснованное повышение тарифов – к общественному недовольству. Таким образом, важны не только количественная корректировка, но и эффективное использование существующего парковочного ресурса на основе анализа данных о заполняемости парковочных мест и прогнозирования загруженности парковок.

Для решения этих задач необходим инструмент, позволяющий моделировать и оптимизировать поведение парковочной системы в различных условиях. Таким инструментом может служить цифровой двойник парковочного пространства, позволяющий исследовать эту проблему без использования реальных ресурсов и обосновать управленческие решения на основе аналитических прогнозов. Таким образом, требуется разработка цифрового двойника для анализа и оценки использования парковочного пространства. Также необходимо исследовать факторы, влияющие на его эффективность: пространственную доступность, близость к точкам интереса, взаимосвязь с другими парковками и уровень их загруженности.

1. ОБЗОР ЛИТЕРАТУРЫ

В последние годы концепция цифровых двойников находит все более широкое применение в сфере управления транспортной инфраструктурой, включая

организацию парковочного пространства. Цифровой двойник – это виртуальная модель объекта или системы, которая позволяет в реальном времени собирать и обрабатывать данные, а также проводить сценарное моделирование. Тем не менее, большинство существующих исследований сосредоточено на задачах мониторинга и накопления статистики, тогда как направления, связанные с оптимизацией процессов и прогнозированием, остаются проработанными в меньшей степени.

Так, в ряде исследований акцент сделан на применение цифровых двойников для мониторинга занятости парковочного пространства и сбора статистики. В статье [1] авторы использовали IoT-технологии с ультразвуковыми датчиками и облачными сервисами, добившись высокой точности распознавания занятости мест. В работе [2] были применены алгоритмы компьютерного зрения YOLOv7 для детекции автомобилей в связке с BIM-моделями, что позволило детектировать автомобили и визуализировать систему в 3D-среде. Несмотря на различие технологий, все эти решения направлены прежде всего на контроль текущей ситуации без моделирования сценариев или оптимизации работы парковочной системы.

В ряде исследований предприняты попытки перейти от простого мониторинга к прогнозированию. Авторы статьи [3] разработали цифровой двойник для старых жилых районов с выраженным дефицитом парковочных мест. Их система использовала комбинацию данных с сенсоров, CAD и BIM-моделей, а также методы машинного обучения. В частности, они применили LSTM-модель для прогнозирования спроса на парковку, это позволило оценивать альтернативные сценарии, такие как изменение схемы организации мест или перераспределение потоков. Тем не менее данное решение было ограничено анализом отдельных случаев и не ориентировалось на поиск оптимального распределения ресурсов.

Наиболее близкой к задачам оптимизации и прогнозирования является работа [4], в которой авторы использовали цифровой двойник в сочетании с методами многокритериальной оптимизации (Pareto-оптимизация, марковские процессы принятия решений, алгоритм роя частиц). Полученные результаты показали сокращение времени поиска парковки, энергопотребления, а также заторов. Тем не менее результаты остались в пределах имитационной среды и не были подтверждены на реальных данных, что ограничивает их применимость.

Таким образом, существующие исследования показывают, что цифровые двойники парковок в первую очередь направлены на мониторинг и сбор статистики, а отдельные примеры прогнозирования и оптимизации носят экспериментальный характер. Поэтому разработка цифрового двойника, способного моделировать динамику занятости, прогнозировать востребованность новых парковок и анализировать влияние различных управлеченческих решений на систему в целом, является актуальной задачей.

2. ПОЛУЧЕНИЕ И ПОДГОТОВКА ДАННЫХ ДЛЯ АНАЛИЗА

Для построения цифрового двойника парковочного пространства города Казани использовались разнородные данные, отражающие характеристики парковок и факторы внешней среды. Учитывались сведения о расположении парковочных зон и их заполняемости, а также пространственные признаки, полученные на основе анализа ближайших объектов городской инфраструктуры. Такой подход позволил сформировать основу для выявления закономерности спроса и разработки прогнозных моделей.

Для получения данных об объектах городской среды различных категорий от торговых и сервисных точек до объектов здравоохранения, туризма и образования было принято решение обратиться к проекту OpenStreetMap [5], из которого с помощью Overpass API были выгружены точки интереса (POI). Их пространственное распределение дало возможность оценить, какие зоны города потенциально притягивают автомобилистов и формируют дополнительную нагрузку на парковки (см. рис. 1).

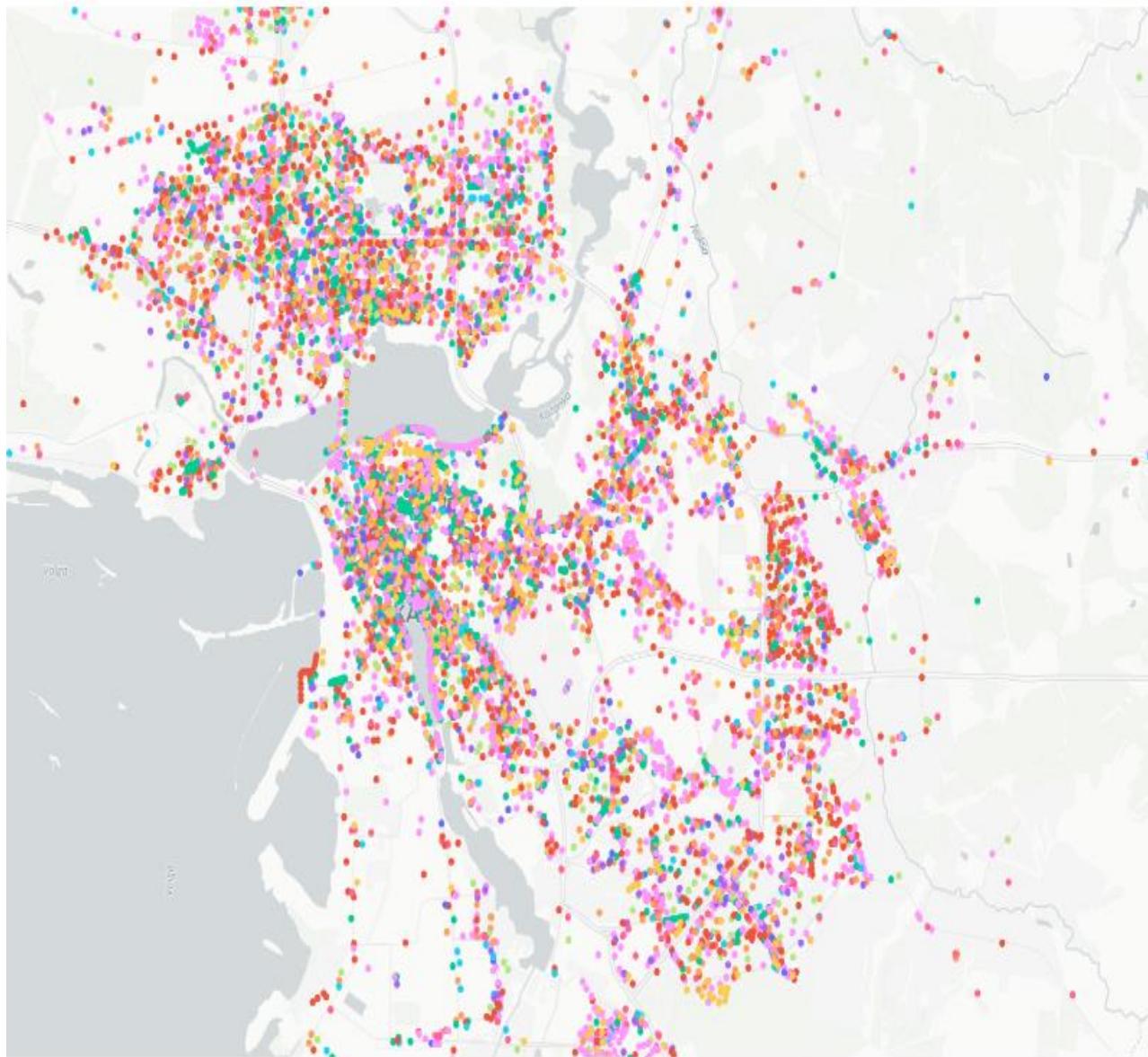


Рис. 1. Карта расположения POI в городе Казани

Данные о парковках города Казани и их заполняемости были получены с помощью сервиса parkingkzn [6]. Сбор данных проводился в течение месяца, что позволило накопить сведения о загруженности парковок в разное время суток и дни недели. Рассмотрим особенности подготовки данных для анализа.

Оба набора данных прошли многоэтапную предобработку. Для POI были выполнены фильтрация по релевантным категориям, устранение дубликатов и исправление ошибок геопозиционирования. Для данных о парковках – синхронизация временных рядов, удаление пропусков и сглаживание шумов, возникающих при кратковременных сбоях фиксации.

Дополнительно была проведена интеграция пространственных и временных признаков: для каждой парковки рассчитывалось количество POI определенных типов в радиусе 200 метров, а также формировались новые атрибуты: время суток, день недели, ценовая категория и доля свободных мест.

Для выявления скрытых закономерностей в пространстве распределения объектов был применен метод кластеризации k-means [7]. На рисунке 2 представлена визуализация результатов кластеризации.

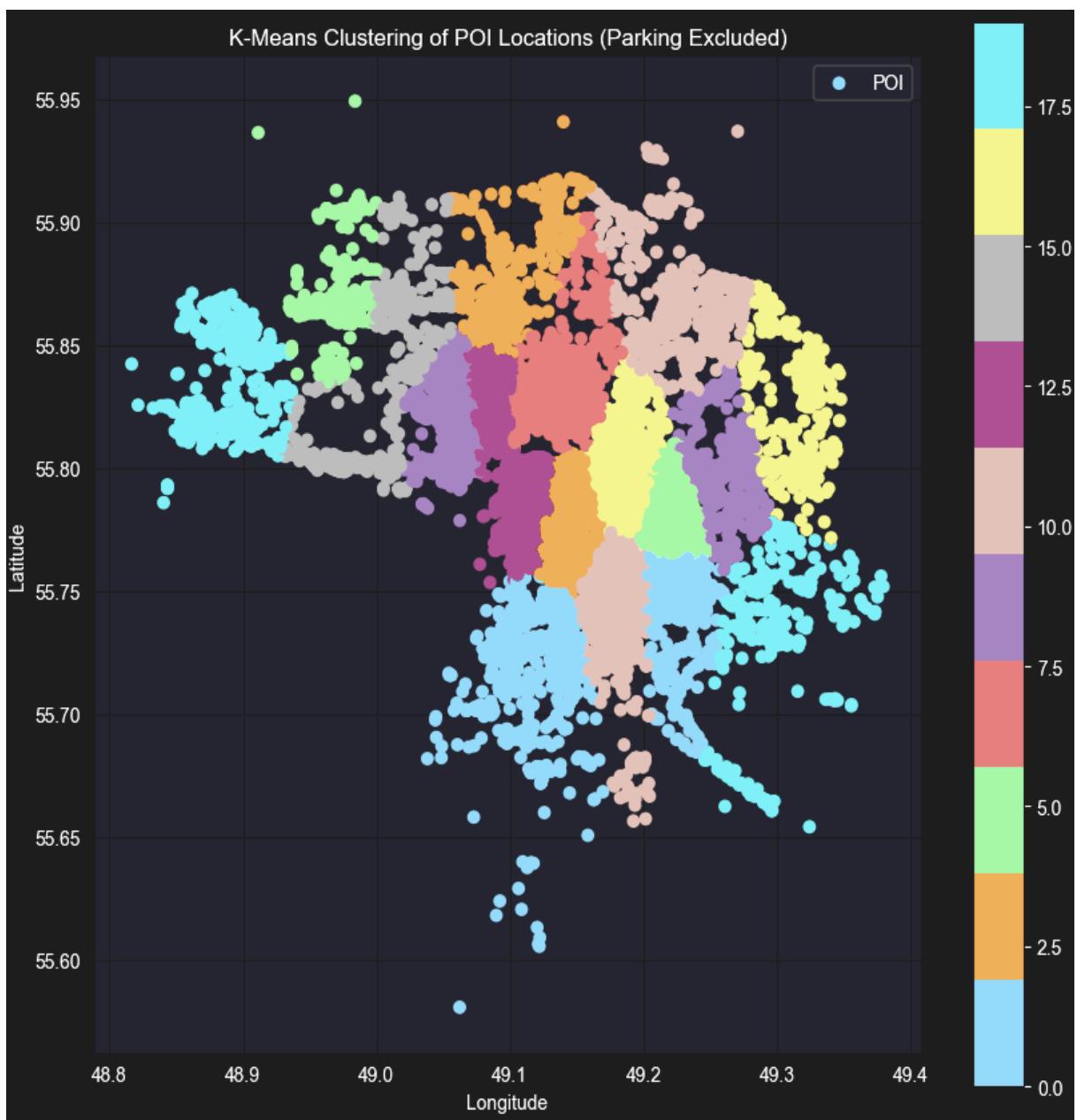


Рис. 2. Визуализация разделения POI на кластеры алгоритмом k-means

В результате удалось сгруппировать точки интереса в 20-ти кластерах, каждый из которых отражал специфический «портрет» городской среды, от зон деловой активности до жилых кварталов. Эта информация была использована в качестве дополнительного фактора при прогнозировании заполняемости парковок, что позволило учесть не только абсолютное число объектов, но и их типологическую структуру.

Следующим этапом стал анализ взаимосвязей между признаками. Для этого была построена матрица корреляций, которая позволила выявить параметры, наиболее сильно влияющие на заполняемость парковок (см. рис. 3).

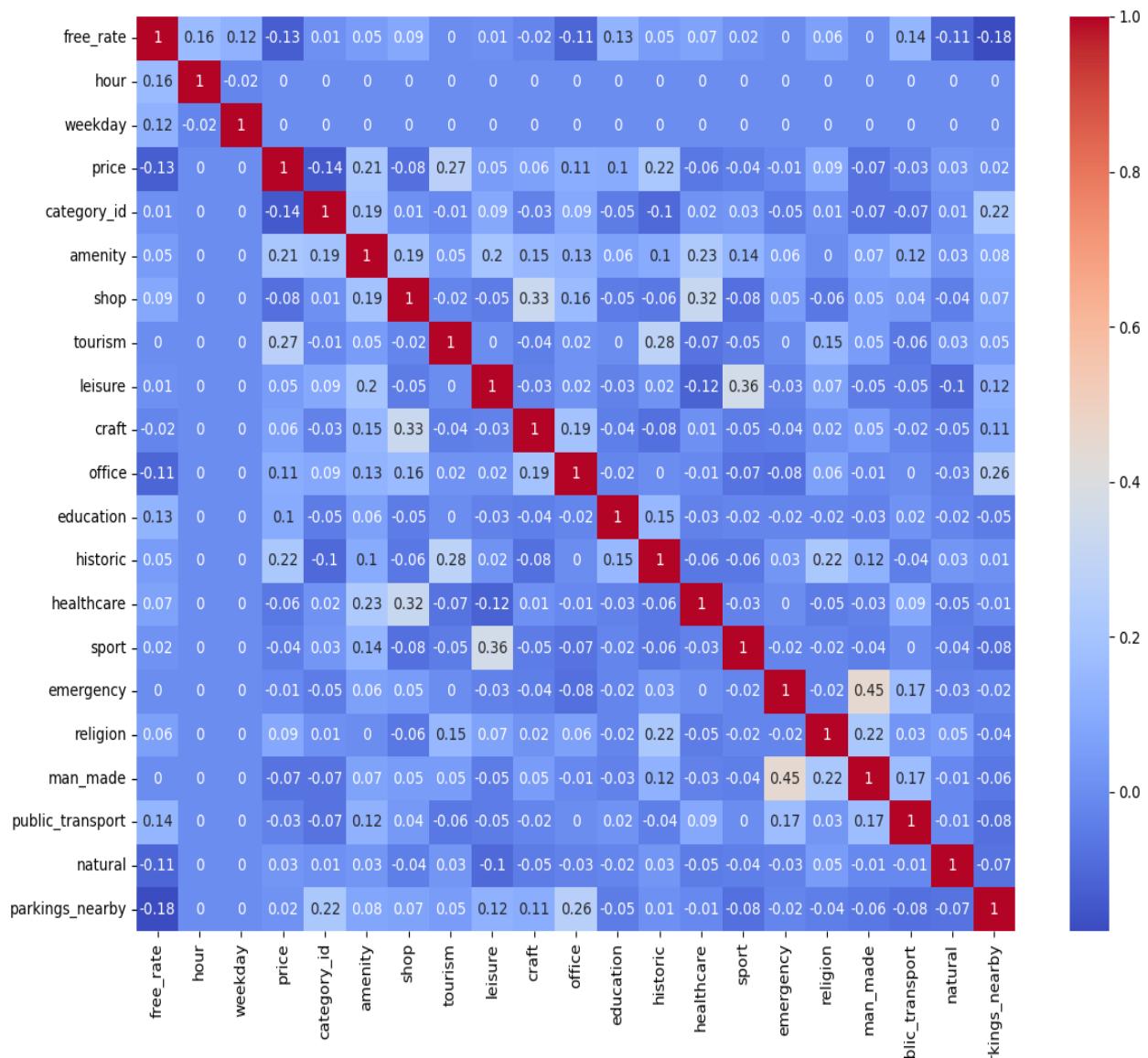


Рис. 3. Матрица корреляции

Как видно из рисунка, высокая корреляция наблюдалась между количеством торговых объектов, расположенных вблизи парковки, и ее фактической загруженностью, а также между ценовой категорией и долей свободных мест. Совмещение результатов корреляционного анализа с кластеризацией ROI дало возможность выделить подмножество признаков, наиболее информативных для прогнозной модели.

Таким образом, подготовка данных включала не только очистку и фильтрацию, но и формирование новых признаков, проведение кластеризации и корреляционного анализа. Такой комплексный подход обеспечил построение репрезентативной выборки, отражающей как внутреннюю динамику парковочных зон, так и влияние городской инфраструктуры. Это создало прочную основу для последующего моделирования сценариев и прогнозирования востребованности парковочных мест.

3. ПОСТРОЕНИЕ И ОЦЕНКА МОДЕЛИ ПРОГНОЗИРОВАНИЯ

После подготовки и анализа данных следующим этапом стало построение моделей, способных прогнозировать уровень заполняемости парковочных зон. Основная задача состояла в том, чтобы сравнить эффективность различных методов машинного обучения и определить, какие из них лучше подходят для практического применения. Для сравнительного анализа были выбраны следующие модели: линейная регрессия [8], дерево решений [9], случайный лес [10], XGBoost [11], многослойный персепtron (MLP) [12, 13] и рекуррентная нейронная сеть LSTM [14, 15]. Такой выбор позволил охватить как классические методы, так и более сложные модели, учитывающие нелинейные зависимости.

Все модели обучались на наборе данных, включавшем временные признаки (день недели, время суток), пространственные характеристики (количество точек интереса различных типов, принадлежность к кластерам), а также информацию о ценовой политике и категории парковки.

На рисунке 4 представлены результаты моделирования.

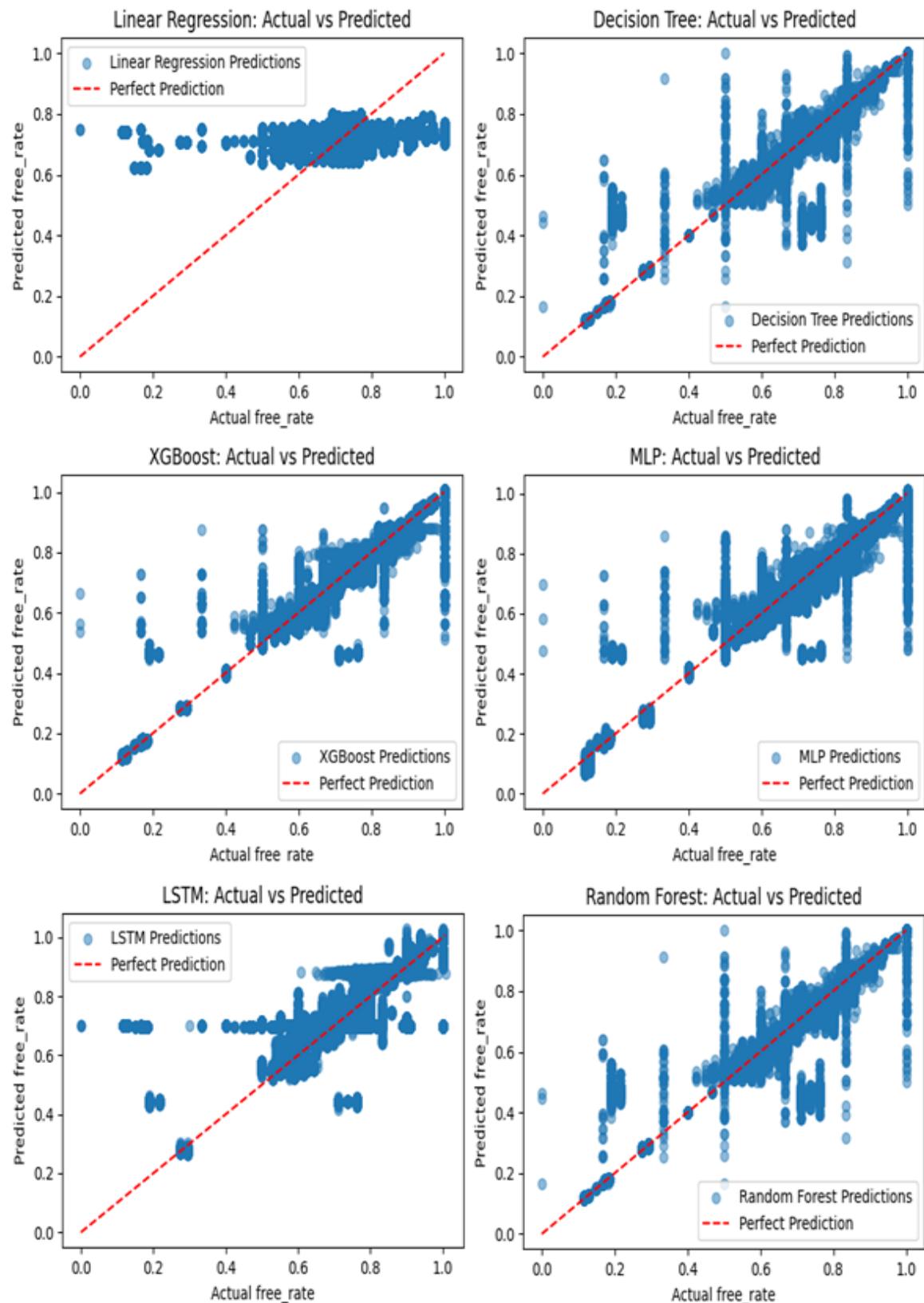


Рис. 4. Графики сравнения реальных данных по заполняемости парковочных мест и данных, полученных моделями

Для оценки качества прогнозирования применялись метрики MSE, RMSE, R2 и MAPE, что обеспечило комплексное сравнение моделей с разных сторон: как по точности приближения фактических значений, так и по способности улавливать закономерности в данных.

На рисунке 5 представлены результаты сравнения построенных моделей по указанным метрикам.

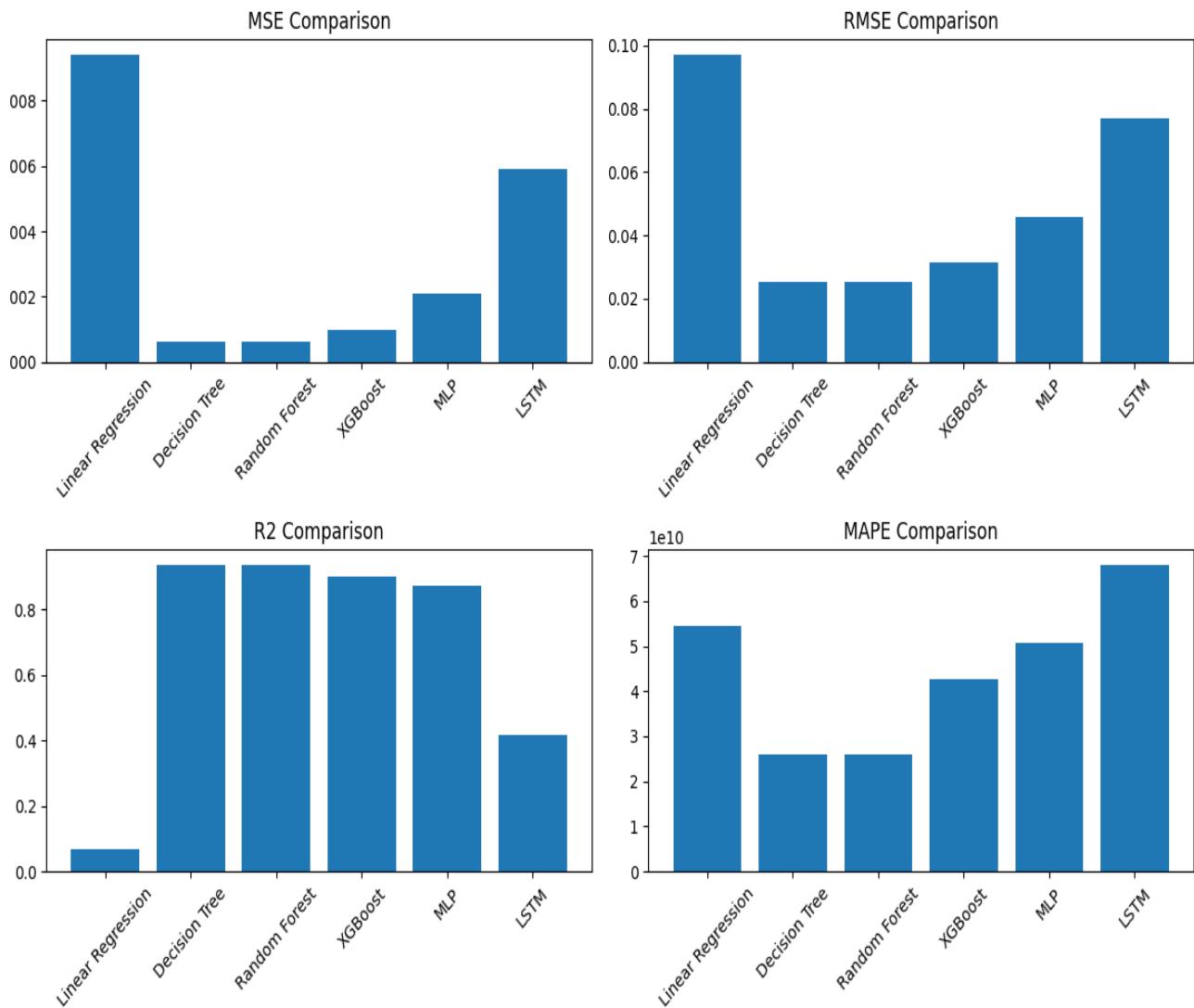


Рис. 5. Графики сравнения метрик точности построенных моделей

Результаты экспериментов показали, что наиболее высокую точность продемонстрировала модель случайного леса, которая оказалась наиболее устойчивой к переобучению. Более низкие значения точности были у XGBoost и дерева решений. Нейросетевые модели оказались менее результативными. LSTM лучше

справлялась с описанием временной динамики по сравнению с MLP, однако их прогностическая способность уступала ансамблевым методам. Наименее удовлетворительные результаты продемонстрировала модель линейной регрессии, что обусловлено ее ограниченной возможностью учитывать нелинейные взаимосвязи между признаками.

Сравнительный анализ построенных моделей позволил сделать вывод о том, что случайный лес является наиболее подходящим методом для решения задачи прогнозирования заполняемости парковочных пространств.

4. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ЦИФРОВОГО ДВОЙНИКА

Для реализации цифрового двойника была разработана программа, объединяющая данные о парковках и окружающей инфраструктуре. На рисунке 6 представлена карта, формируемая программой, где каждая парковка визуализируется в виде отдельного объекта с возможностью просмотра ее характеристик и текущего уровня загруженности.

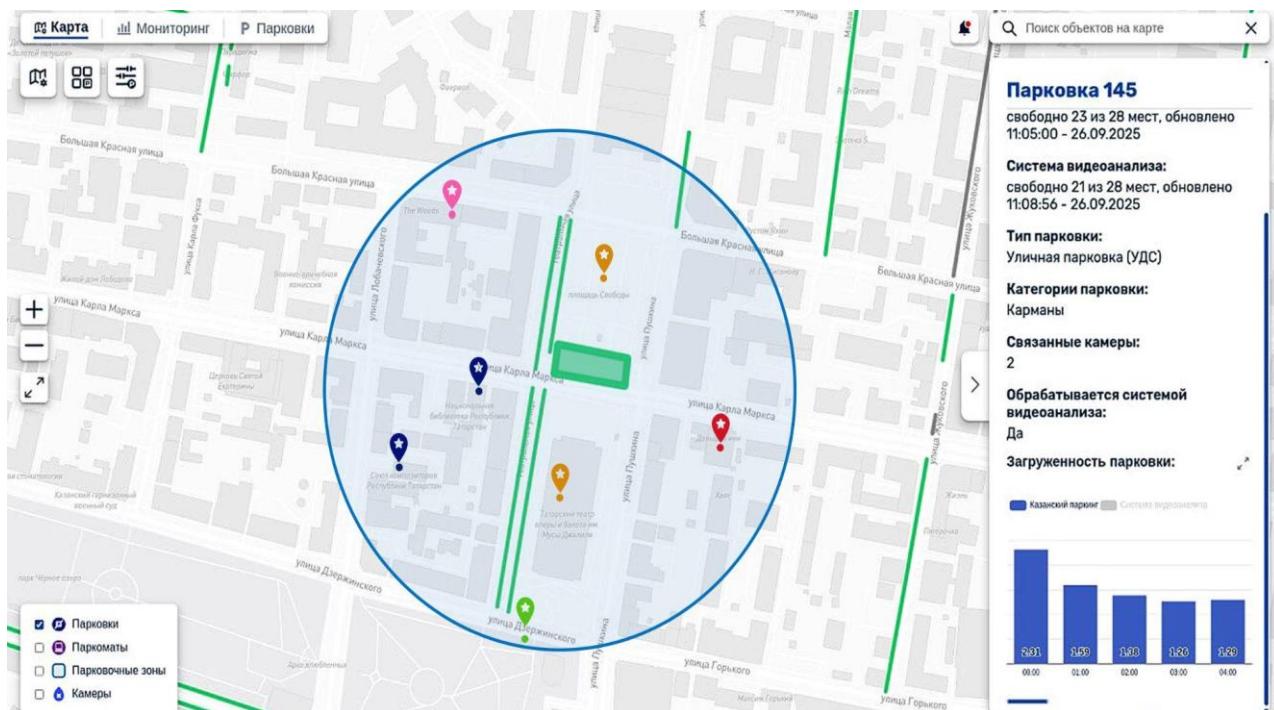


Рис. 6. Интерфейс работы системы

Система интегрирует данные о количестве свободных мест, категории парковки, ее типе и технических параметрах. Пользователям доступна информация о точках интереса (POI), расположенных вблизи каждой парковочной зоны. Для

каждой парковки рассчитывается радиус окружения, в пределах которого анализируется количество магазинов, учреждений, объектов культуры, здравоохранения и других объектов.

Система также предоставляет инструменты для взаимодействия пользователя с цифровым двойником, предусмотрена возможность добавления новых объектов (парковок или точек интереса), а также изменения стоимости парковок и их вместимости. Это позволяет проводить сценарное моделирование, оценивая, как те или иные изменения повлияют на загрузку парковочной зоны.

Дополнительно реализована визуализация динамики загруженности в разные периоды времени. Пользователь может просматривать графики изменения уровня занятости парковки в течение суток и недели. Такая функциональность делает систему не только инструментом мониторинга, но и полноценной аналитической платформой для поддержки управленческих решений. На рисунке 7 представлен пример такого графика.



Рис. 7. График зависимости количества занятых парковочных мест от времени

Разработанный прототип цифрового двойника демонстрирует, каким образом можно объединять данные из различных источников, представлять их в удобной визуальной форме и использовать для аналитики. Система позволяет не

только отслеживать текущую ситуацию, но и моделировать изменения за счет изменения параметров парковок и окружения. Это превращает ее из инструмента мониторинга в платформу для поддержки управленческих решений.

ЗАКЛЮЧЕНИЕ

Рассмотрена проблема дефицита парковочных мест и показано, что традиционные методы ее решения не эффективны. Проведенный обзор литературы подтвердил, что существующие цифровые двойники парковок в основном ориентированы на мониторинг и сбор статистики, в то время как прогнозирование и оптимизация остаются недостаточно проработанными. На примере города Казани разработан прототип цифрового двойника, основанный на данных OpenStreetMap и сервиса «Казанский паркинг». После очистки и анализа данных были построены модели прогнозирования заполняемости парковок, среди которых наилучшие результаты продемонстрировал алгоритм случайного леса. Разработанная система объединяет мониторинг и аналитику с возможностью сценарного моделирования, что превращает ее в инструмент поддержки управленческих решений. Полученные результаты подтверждают перспективность применения цифровых двойников для повышения эффективности управления городской парковочной инфраструктурой.

СПИСОК ЛИТЕРАТУРЫ

1. Amusan A.A., Ogunleye G.A. A Digital Twin-Enabled Smart Car Park Management System: Architecture and Impact on Emission Reduction // FUOYE Journal of Engineering and Technology. 2024. Vol. 9. No. 4. P. 629–635.
2. Coching J.K. et al. Digital Twinning Mechanism and Building Information Modeling for a Smart Parking Management System // Smart Cities. 2025. Vol. 8. No. 5. P. 146.
3. Zou Y. et al. A Digital Twin prototype for smart parking management // ECPPM 2022-eWork and eBusiness in Architecture, Engineering and Construction 2022. CRC Press, 2023. P. 250–257.
4. Chen W., Wang X., Wu M. Intelligent Parking Service System Design Based on Digital Twin for Old Residential Areas // Electronics. 2024. Vol. 13. No. 23. P. 4597.

5. Open Street Map: сайт [Электронный ресурс]. Режим доступа: <https://www.openstreetmap.org/> (дата обращения: 25.08.2025).
6. Горпарковки Казань. Официальный сайт сервиса парковок [Электронный ресурс]. Режим доступа: <https://parkingkzn.ru/ru/> (дата обращения: 25.08.2025).
7. *Aliguliyev R., Tahirzada S.F.* Performance comparison of k-means, parallel k-means and k-means++ // Reliability: Theory & Applications. 2025. Vol. 20. No. S7 (83). P. 169–176.
8. *Базилевский М.П.* Решение оптимизационной задачи оценивания моделей полносвязной линейной регрессии // Моделирование и анализ данных. 2024. Т. 14. № 1. С. 121–134.
9. *Пырнова О.А., Катасёв А.С.* Построение и оценка эффективности модели дерева решений для прогнозирования успеваемости обучающихся // Инженерный вестник Дона. 2024. № 4 (112). С. 578–584.
10. *Utkin L., Konstantinov A.* Random survival forests incorporated by the Nadaraya-Watson regression // Informatics and Automation. 2022. Vol. 21. No. 5. P. 851–880.
11. *Chen T., Guestrin C.* XGBoost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016. P. 785–794.
12. *Матвеева К.А., Минниханов Р.Н., Катасёв А.С.* Сверточная нейросетевая модель акустического обнаружения аварийно-спасательных машин // Вестник Технологического университета. 2024. Т. 27. № 1. С. 76–80.
13. *Комарцова Л.Г., Кадников Д.С.* Исследование генетических алгоритмов для обучения многослойного персептрона // Нейрокомпьютеры: разработка, применение. 2010. № 12. С. 12–19.
14. *Hochreiter S., Schmidhuber J.* Long short-term memory // Neural computation. 1997. Vol. 9. No. 8. P. 1735–1780.
15. *Емалетдинова Л.Ю., Вильданов Н.Р., Катасёв А.С.* Использование нейросетевой модели TCN-LSTM для прогнозирования значений временного ряда // Научно-технический вестник Поволжья. 2023. № 6. С. 62–64.

DIGITAL TWIN OF PARKING SPACE

R. N. Minnikhanov¹ [0000-0001-9166-2955], T. R. Batorshin² [0009-0001-9995-8854],

R. M. Gabbazov³ [0009-0009-9187-9121], R. I. Fakhraziev⁴ [0009-0008-3463-7663],

A. S. Katasev⁵ [0000-0002-9446-0491], M. V. Dagaeva⁶ [0000-0002-5444-9669],

I. R. Badruttinov⁷ [0009-0001-6343-4709]

¹Tatarstan Academy of Sciences, Kazan, 420111, Russia

^{1, 6}SBI "Road Safety", Kazan, 420059, Russia

²Kazan State Power Engineering University, Kazan, 420066, Russia

^{3-5, 7}Kazan National Research Technical University named after A.N. Tupolev-KAI, Kazan, 420111, Russia

¹Priemnaya.Minnihanov@tatar.ru, ²Timur.Batorshin.02@gmail.com,

³GabbazovRM@stud.kai.ru, ⁴FakhrazievRI@stud.kai.ru, ⁵ASKatasev@kai.ru,

⁶ITS.Center.kzn@gmail.com, ⁷BadruttinovIR@stud.kai.ru

Abstract

Increasing urbanization and motorization lead to a shortage of parking spaces, resulting in congestion, increased emissions, and a declining quality of life. Traditional parking management methods are ineffective in addressing this issue, necessitating the use of data analysis and forecasting tools. This paper examines the use of a digital twin of the Kazan parking system. Data was filtered and integrated, points of interest were clustered, and a correlation analysis of factors influencing parking occupancy was performed. Linear regression, decision tree, random forest, XGBoost, MLP, and LSTM models were trained and compared to predict occupancy levels. The random forest model demonstrated the best results. The developed digital twin prototype enables monitoring and scenario modeling, making it an effective tool for parking space optimization and management decision-making.

Keywords: *digital twin, parking, parking space occupancy, machine learning, points of interest, optimization.*

REFERENCES

1. Amusan A.A., Ogunleye G.A. A Digital Twin-Enabled Smart Car Park Management System: Architecture and Impact on Emission Reduction // FUOYE Journal of Engineering and Technology. 2024. Vol. 9. No. 4. P. 629–635.

2. *Coching J.K. et al.* Digital Twinning Mechanism and Building Information Modeling for a Smart Parking Management System // Smart Cities. 2025. Vol. 8. No. 5. P. 146.
3. *Zou Y. et al.* A Digital Twin prototype for smart parking management // ECPPM 2022-eWork and eBusiness in Architecture, Engineering and Construction 2022. CRC Press, 2023. P. 250–257.
4. *Chen W., Wang X., Wu M.* Intelligent Parking Service System Design Based on Digital Twin for Old Residential Areas // Electronics. 2024. Vol. 13. No. 23. P. 4597.
5. Open Street Map: sajt [Elektronnyj resurs].
URL: <https://www.openstreetmap.org/> (last access: 25.08.2025).
6. Gorparkovki Kazan'. Oficial'nyj sajt servisa parkovok [Elektronnyj resurs].
URL: <https://parkingkzn.ru/ru/> (last access: 25.08.2025).
7. *Aliguliyev R., Tahirzada S.F.* Performance comparison of k-means, parallel k-means and k-means++ // Reliability: Theory & Applications. 2025. Vol. 20. No. S7 (83). P. 169–176.
8. *Bazilevsky M.P.* Reshenie optimizacionnoj zadachi ocenivaniya modelej polnosvyaznoj linejnoj regressii // Modeling and data analysis. 2024. Vol. 14. No. 1. P. 121–134.
9. *Pyrnova O.A., Katasev A.S.* Postroenie i ocenka effektivnosti modeli dereva reshenij dlya prognozirovaniya uspevaemosti obuchayushchihся // Engineering Bulletin of the Don. 2024. No. 4 (112). P. 578–584.
10. *Utkin L., Konstantinov A.* Random survival forests incorporated by the Nadaraya-Watson regression // Informatics and Automation. 2022. Vol. 21. No. 5. P. 851–880.
11. *Chen T., Guestrin C.* XGBoost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016. P. 785–794.
12. *Matveeva K.A., Minnikhanov R.N., Katasev A.S.* Svertochnaya nejrosetevaya model' akusticheskogo obnaruzheniya avarijno-spasatel'nyh mashin // Bulletin of the Technological University. 2024. Vol. 27. No. 1. P. 76–80.

13. Komartsova L.G., Kadnikov D.S. Issledovanie geneticheskikh algoritmov dlya obucheniya mnogoslojnogo perseptrona // Neurocomputers: development, application. 2010. No. 12. P. 12–19.
 14. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. Vol. 9. No. 8. P. 1735–1780.
 15. Emaletdinova L.Yu., Vildanov N.R., Katasev A.S. Ispol'zovanie nejrosetevoj modeli TCN-LSTM dlya prognozirovaniya znachenij vremennogo ryada // Scientific and Technical Bulletin of the Volga Region. 2023. No. 6. P. 62–64.
-

СВЕДЕНИЯ ОБ АВТОРАХ



МИННИХАНОВ Рифкат Нургалиевич – д. т. н., проф., Президент Академии наук Республики Татарстан, директор ГБУ «Безопасность дорожного движения».

Rifkat Nurgaliyevich MINNIKHANOV – Doctor of Technical Sciences, Professor, President of the Tatarstan Academy of Sciences, Director of the SBI "Road Safety".

email: Priemnaya.Minnihanov@tatar.ru

ORCID: 0000-0001-9166-2955



БАТОРШИН Тимур Русланович – магистрант кафедры автоматизации технологических процессов и производств Казанского государственного энергетического университета.

Timur Ruslanovich BATORSHIN – Master's student of the Department of Automation of Technological Processes and Production of the Kazan State Power Engineering University.

email: Timur.Batorshin.02@gmail.com

ORCID: 0009-0001-9995-8854



ГАББАЗОВ Руслан Марселеевич – аспирант кафедры автоматизированных систем обработки информации и управления Казанского национального исследовательского технического университета им. А.Н. Туполева-КАИ (КНИТУ-КАИ).

Ruslan Marselevich GABBASOV – postgraduate student of the Department of Automated Systems for Information Processing and Control of the Kazan National Research Technical University named after A.N. Tupolev-KAI (KNRTU-KAI).

email: GabbazovRM@stud.kai.ru

ORCID: 0009-0009-9187-9121



ФАХРАЗИЕВ Рузель Ильдарович – магистрант кафедры систем информационной безопасности КНИТУ-КАИ.

Ruzel Ildarovich FAKHRAZIEV – Master's student of the Department of Information Security Systems of the KNRTU-KAI.

email: FakhrazievRI@stud.kai.ru

ORCID: 0009-0008-3463-7663



КАТАСЁВ Алексей Сергеевич – д. т. н., проф., профессор кафедры систем информационной безопасности КНИТУ-КАИ.

Alexey Sergeevich KATASEV – Doctor of Technical Sciences, Professor, Professor of the Department of Information Security Systems of the KNRTU-KAI.

email: ASKatasev@kai.ru

ORCID: 0000-0002-9446-0491



ДАГАЕВА Мария Витальевна – к. т. н., начальник Центра разработки и сопровождения информационных систем ГБУ «Безопасность дорожного движения».

Maria Vitalievna DAGAEVA – Candidate of Technical Sciences, Head of the Center for Development and Maintenance of Information Systems of the SBI "Road Safety".

email: ITS.Center.kzn@gmail.com

ORCID: 0000-0002-5444-9669



БАДРУТДИНОВ Инзиль Ринатович – магистрант кафедры прикладной математики и информатики КНИТУ-КАИ.

Inzil Rinatovich BADRUTDINOV – Master's student at the Department of Applied Mathematics and Informatics of the KNRTU-KAI.

email: BadruttinovIR@stud.kai.ru

ORCID: 0009-0001-6343-4709

Материал поступил в редакцию 28 августа 2025 года

УДК 004.94:794.02

ОНЛАЙН-ИНСТРУМЕНТ TULA ДЛЯ БАЛАНСИРОВКИ ВИДЕОИГР

В. Р. Раҳманкулова¹ [0009-0006-8425-616X], В. В. Кугуракова² [0000-0002-1552-4910]

^{1, 2}Казанский федеральный университет, г. Казань, Россия

¹raxmankulova.v@mail.ru, ²vлада.kugurakova@gmail.com

Аннотация

Разработан инструмент Tula для балансировки видеоигр. Его необходимость обоснована растущими требованиями к качеству и экономической эффективности в индустрии видеоигр, особенно в аспектах управления внутриигровой экономикой и логикой игрового мира. Проанализированы существующие инструменты и подходы к балансировке игр, выявлены их ограничения, на основе которых построен функционал нового инструмента, интегрирующего функции современных решений и предоставляющего расширенные возможности для анализа и тестирования игровых параметров, включая генерацию прототипов через описание классов и симуляцию в реальном времени. Описаны технологическая база и архитектура инструмента. Рассмотрены ключевые аспекты реализации: отзывчивость интерфейса, непрерывное обновление данных и безопасность. Проведенный сравнительный анализ с известным инструментом Machinations показал преимущества в корректности обработки данных, удобстве интерфейса и гибкости модификации прототипов.

Ключевые слова: видеоигры, игровой процесс, игровые механики, игровой баланс, игровой дизайн, *Machinations*.

ВВЕДЕНИЕ

С каждым днем требования к качеству и экономической эффективности создания видеоигр возрастают. Важной составляющей успеха в индустрии является как разработка общей логики игрового мира и игровых механик, влияющих на удержание внимания игроков, так и калибровка внутриигровой экономики и значений параметров игровых проектов, определяющие распределение ресурсов в игровом мире [1]. Поэтому особое значение приобретает балансировка числовых параметров всех игровых элементов. Цель представляемого проекта – упростить создание сбалансированных прототипов игровых

механик, а в рамках исследований по формализации игровых механик [2] – упростить создание сбалансированных прототипов видеоигр.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ БАЛАНСА В ВИДЕОИГРАХ

Определение и анализ понятия

Понятие «игровой баланс» традиционно ассоциируется с равным распределением игровых ресурсов, обеспечивающим честную и мотивирующую игру. Существуют различные подходы к определению игрового баланса, поэтому имеет смысл привести некоторые из них, опираясь на последние исследования и практики в области разработки видеоигр.

В [3] на основе интервью с игровыми дизайнерами проанализирована концепция игрового баланса. Авторы подчеркивают важность достижения гармонии между интерактивностью, сюжетом и игровым прогрессом.

В [4] предложено выделение количественных параметров для оценки игр, значение которых зависит от контекста.

В [1] дан детальный анализ понятия игрового баланса с разграничением узкого математического понимания и более широкого дизайнераского представления.

В [5] представлены различные подходы практиков к балансировке, такие как акцентирование на принципе справедливости в многопользовательских играх, применение подхода к балансировке многопользовательских игр вокруг игрока, поиск и сохранение равновесия сил в игре.

Существует несколько широко известных научных работ, посвященных балансировке игр. Авторы этих работ используют или расширяют этот термин, часто принимая определения практиков [6], и оценивают динамические подходы к балансировке игры, утверждая, что они должны удовлетворять как минимум трем основным требованиям: адаптироваться к начальному уровню игрока, отслеживать эволюцию и регрессию его производительности, а также поддерживать игру «правдоподобной». Такая оценка показывает, что адаптивные подходы могут быть более эффективными и приводить к большей удовлетворенности пользователей, чем традиционная, неадаптивная, заранее заданная балансировка игры с использованием статических уровней сложности.

В многопользовательских играх неравенство в навыках может привести к разочарованию и нежелательным впечатлениям. Существуют подходы к балансировке для выравнивания игрового поля (например, предоставление игрокам с низкими показателями помощи в прицеливании), но неясно, как различные варианты дизайна влияют на индивидуальное восприятие игроков. В [7] предложено использовать пространство дизайна механик балансировки, а как один из итогов определена балансировка как неотъемлемая часть игры, не нарушающая основной геймплей.

По нашему мнению, не отрицая предлагаемые подходы теоретиков и практиков геймдизайна, необходимо подчеркнуть, что определение игрового баланса непосредственно тесно связано с процессом проектирования механик игры. И все же в узком смысле понимания балансировки игрового баланса можно рассматривать как методику настройки и анализа числовых параметров элементов прототипа игровой механики. Основная его цель – гарантировать, что идеи геймдизайна будут не только реализованы, но и функционально достижимы.

Сравнительный анализ существующих инструментов

Анализ часто используемых инструментов для разработки видеоигр позволяет выделить как их преимущества, так и ограничения.

Бесплатные инструменты для обработки данных в виде таблиц (например, популярный инструмент Google Sheets¹) предоставляют базовые возможности, включая разнообразную обработку данных. Однако, несмотря на широкие аналитические возможности, эти инструменты обладают ограниченным функционалом для разработки прототипов игровых механик из-за их шаблонной и строго математической природы. С развитием и усложнением игровых прототипов становится очевидной потребность в более интуитивно понятных и визуально ориентированных инструментах.

¹ Google Sheets – spreadsheet application included as part of the free, web-based Google Docs Editors suite offered by Google. <https://www.google.com/sheets/>

В отличие от табличных приложений, завоевавшая бесспорную популярность в своем сегменте бесконечная интерактивная онлайн-доска Miro² имеет ограничения в анализе данных и разработке сложных моделей игровых механик. Хотя инструмент Miro очень эффективен для визуального прототипирования в режиме совместной работы, он менее подходит для обработки больших схем или выполнения сложных расчетов.

Браузерный инструмент Machinations³, разработанный специально для проектирования и балансировки игровых прототипов, позволяет моделировать игровые процессы, проводить эксперименты и собирать количественные данные из симуляций игровых механик и систем, отображая результаты в виде диаграмм. Сравнение методов документирования требований к поведению NPC (Non-Player Character — неигровой персонаж в видеоиграх) показало, что использование динамических диаграмм в Machinations значительно ускоряет процесс работы над проектом, уменьшая необходимость в заполнении объемных текстовых документов [8].

Однако, несмотря на неоспоримые преимущества, Machinations имеет ряд ограничений, таких как:

- проблемы с производительностью, визуализацией и масштабированием сложных прототипов;
- сложность в понимании взаимосвязей между сущностями из-за их комплексной структуры;
- вероятность ошибок при запуске симуляций из-за отсутствия проверки валидности числовых значений в полях;
- невозможность внесения изменений во время симуляции;
- ограниченное количество шагов, доступных для проведения симуляции, сужающее функционал;
- сложный и перегруженный интерфейс.

² Miro – visual workspace where teams manage projects, design products. <https://miro.com/>

³ Machinations – game design tool for balancing. <https://machinations.io/>

КОНЦЕПЦИЯ И АРХИТЕКТУРА НОВОГО ИНСТРУМЕНТА

Принципы проектирования

Из анализа существующих инструментов стало очевидным, что для разработки будущего инструмента критично важно обеспечить простую и понятную визуализацию данных для прототипирования, а также интегрировать функционал для анализа и сравнения метрик. Предлагаемый нами инструмент объединяет лучшие функции современных решений, предоставляя интуитивно понятный интерфейс и расширенные возможности для тестирования и анализа игровых механик. Дополнительные возможности итогового приложения включают:

- гибкая настройка весов, позволяющая регулировать различные игровые параметры, такие как скорость, сложность и уровень ресурсов, улучшая тем самым гибкость проектирования;
- визуализация игровых элементов в виде узлов и связей между ними, что облегчает восприятие и анализ структуры игры;
- тестирование в реальном времени – функция, позволяющая наблюдать за изменениями в динамике игры напрямую во время тестирования прототипа;
- аналитические инструменты, которые позволяют изучать эффективность игровых стратегий и оптимизировать распределение ресурсов для улучшения баланса и интерактивности игры;
- функциональность для генерации прототипов видеоигр на основе описания классов игровых механик в онлайн-редакторе кода;
- поддержка экспорта данных, например, в формат JSON.

Предлагаемый инструмент опирается на формальный подход к пространственно-временному моделированию игровых систем [2], что обеспечивает математически обоснованную балансировку параметров.

Сценарии работы

На основе спроектированного концепта инструмента был выстроен следующий сценарий работы пользователя.

- **Авторизация:** Пользователь вводит данные для аутентификации через электронную почту или аккаунт на GitHub⁴.
- **Выбор доски:** После входа в систему пользователь переходит на личную страницу, где выбирает доску для начала работы.
- **Выбор режима:** Пользователь выбирает режим построения прототипа – ручной или генерация путем описания классов в онлайн редакторе.
- **Построение прототипа:** Пользователь создает прототип механики игры или редактирует сгенерированный прототип, перетаскивая узлы и устанавливая связи между ними на интерактивной доске, при этом вводя числовые характеристики для симуляции.
- **Настройка параметров аналитики:** Пользователь выбирает количество итераций, время в секундах и количество игровых сессий для запуска симуляций путем заполнения соответствующих полей на интерактивной панели.
- **Выполнение симуляции:** Пользователь запускает симуляцию прототипа и выбирает нужный узел, «щелкая» по нему для анализа значений узла в результате пройденной игровой сессии.
- **Анализ результатов:** Пользователь анализирует полученные значения и подбирает необходимую числовую характеристику для дальнейшей работы.
- **Выгрузка данных:** По завершении анализа пользователь может выгрузить код в формате JSON или схему в расширении PNG для дальнейшей разработки игры.

В это время инструмент работает по следующему сценарию.

1. **Аутентификация:** Инструмент проверяет предоставленные учетные данные, отправляя запрос в систему Clerk⁵, и в случае успешной аутентификации предоставляет доступ к функционалу инструмента.

⁴ GitHub – облачная платформа для хостинга IT-проектов и совместной разработки.

⁵ Clerk – authentication and user management system. <https://clerk.com>

2. **Личные доски:** При входе инструмент предлагает выбрать доску из существующих или же создать новую, а также добавить доску в избранное. Система отправляет запрос в систему Clerk, сопоставляя токен пользователя с доступными ему досками.
3. **Коллаборация:** При выборе доски инструмент предлагает возможность пригласить других участников для совместной работы над выбранной доской путем интеграции с библиотекой Liveblocks⁶ и системой Clerk для авторизованных аккаунтов.
4. **Выбор режима:** Инструмент реагирует на запрос пользователя о выборе режима построения прототипа, предоставляя возможность выбора между ручным режимом и генерацией на основе описания классов в онлайн редакторе.
5. **Создание прототипа:** Инструмент предоставляет интерфейсную панель на интерактивной доске для создания узлов.
6. **Настройка параметров аналитики:** Инструмент сохраняет введенные значения в менеджере состояний⁷ и автоматически запускает симуляцию прототипа с заданными параметрами через установленный интервал.
7. **Выполнение симуляции:** Система временно сохраняет данные после каждой симуляции игрового прототипа в менеджер состояний. По завершении каждой игровой сессии она отправляет POST-запрос на сервер, предоставляя информацию о значениях узлов на каждом шаге итерации для последующего анализа, вся информация сохраняется в базе данных.
8. **Анализ значений:** После выбора узла путем клика по нему инструмент отправляет GET-запрос на сервер с идентификатором узла, запрашивая

⁶ Liveblocks – real-time collaboration infrastructure for building performant collaborative experiences. <https://liveblocks.io/>

⁷ Менеджер состояний (statemanager) – это инструмент, используемый в программировании для управления состоянием приложения. В контексте веб-разработки и разработки приложений состояние относится к любым данным или информации, которые могут изменяться во время работы приложения, например это данные пользователя, настройки интерфейса, информация о текущей странице и т.д.

информацию. Система сопоставляет значения запроса с информацией в базе данных и затем отправляет данные клиенту, который, в свою очередь, отображает полученные данные в виде линейной диаграммы.

9. **Предоставление данных:** Инструмент предоставляет пользователю возможность анализировать полученные результаты и выгружать для дальнейшей работы код в JSON или схему в PNG. Выбор формата JSON обусловлен несколькими причинами, в частности тем, что он является широко используемым и понятным форматом данных, а его использование в качестве формата экспорта механики игры позволяет значительно сократить время, затрачиваемое на документирование игровой системы.

Описанный сценарий работы созданного инструмента представляет собой четко структурированный процесс, который обеспечивает удобство и гибкость при создании и анализе прототипов.

Архитектурные решения

Переходя к описанию технологической базы инструмента, остановимся на выборе программных платформ и методов обработки данных. В процессе проектирования архитектуры интерактивного приложения возник ряд типичных проблем, таких как:

- отзывчивость интерфейса и быстрая загрузка данных – важно, чтобы переходы между экранами и загрузка данных происходили быстро, чтобы обеспечить плавное взаимодействие пользователя с приложением и избежать задержек;
- непрерывное обновление данных – данные в приложении должны обновляться быстро и незаметно для пользователя, чтобы не нарушать его работу с приложением.

Для бесшовного решения этих проблем предложена следующая архитектура (рис. 1).

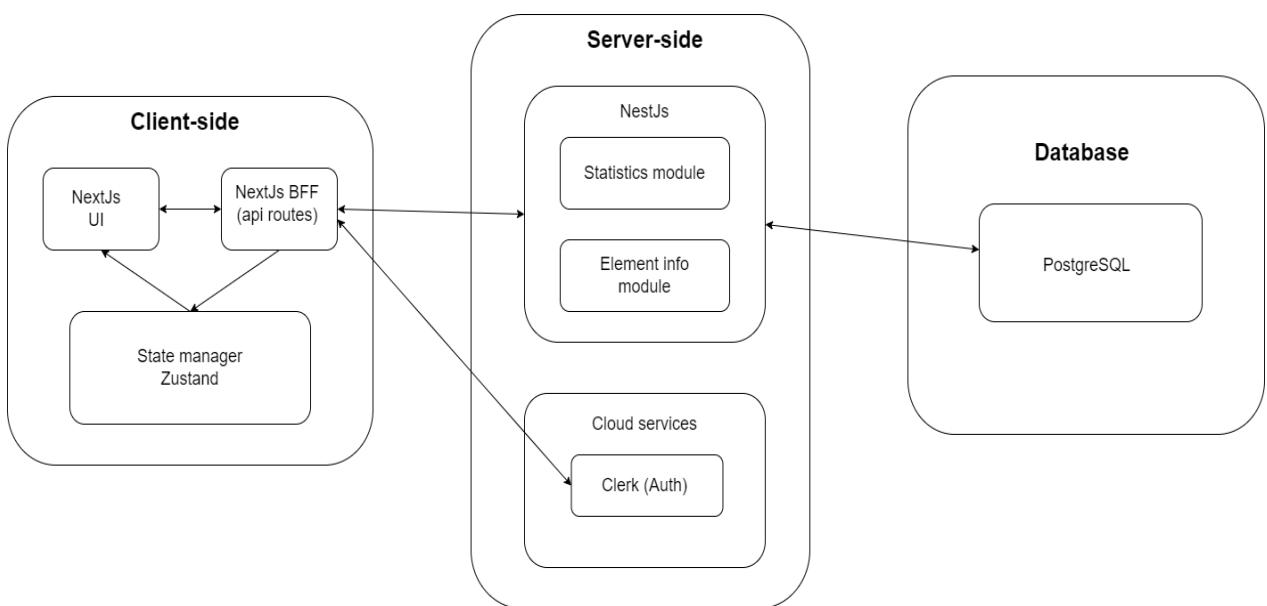


Рис. 1. Архитектура онлайн-инструмента

Клиентская часть инструмента разработана с использованием фреймворка Next.js⁸, который обеспечивает удобную интеграцию с разнообразными библиотеками для создания пользовательских интерфейсов и применяет подход BFF⁹, предоставляя каждому интерфейсу возможность взаимодействовать с серверным компонентом через API для обеспечения необходимой функциональности. Next.js позволяет реализовать предварительную загрузку страниц и асинхронную загрузку данных, что способствует быстрым переходам между экранами и минимизации времени ожидания.

Для глобального управления состоянием использована библиотека Zustand¹⁰, предназначенная для хранения и обновления временных данных в процессе работы инструмента. Благодаря минималистичному синтаксису Zustand создание модульных хранилищ данных для различных компонентов происходит без создания громоздкой архитектуры, позволяя компонентам обновляться автоматически без явного вызова методов перерисовки.

⁸ Next.js – React Framework for the web applications documentation. <https://nextjs.org/>

⁹ BFF (сокр. от англ. Backend For Frontend) – архитектурный подход, позволяющий разделить фронтенд и бэкенд.

¹⁰ Zustand – state manager for the web applications documentation.
<https://docs.pmnd.rs/zustand/>

На серверной стороне для повышения безопасности внедрен инструмент Clerk, механизм которого позволяет выполнять аутентификацию и проверять доступ пользователя к определенным страницам до полной загрузки контента.

В качестве фреймворка для выполнения математических вычислений, расчета статистики и обработки информации об игровых элементах выбран Nest.js.

Для обработки и хранения данных задействована база данных PostgreSQL¹¹, к которой происходит обращение через Nest.js с использованием ORM Prisma¹². PostgreSQL обеспечивает надежное хранение данных о состояниях игровых элементов, получаемых в ходе симуляций, что позволяет эффективно анализировать и использовать эту информацию для статистических расчетов.

РЕАЛИЗАЦИЯ И ФУНКЦИОНАЛЬНОСТЬ ИНСТРУМЕНТА

В качестве имени разработанного инструмента выбрано рабочее название, отражающее главную цель, – балансирование игровых параметров. На санскрите *тула* ((तुला¹³, Tula) означает «весы» или «баланс», и это слово можно транслитерировать на латиницу как Tula.

Особенности реализации

После входа в систему пользователю представляется главная страница инструмента (рис. 2).

¹¹ PostgreSQL – open source database documentation. <https://www.postgresql.org/>

¹² ORM (сокр. от англ. Object-Relational Mapping) Prisma – объектно-реляционное отображение, позволяющее работать как с реляционными (PostgreSQL, MySQL, SQL Server, SQLite), так и нереляционной (MongoDB) базами данных с помощью JavaScript или TypeScript без использования SQL.

¹³ Деванагари — слоговое письмо, используемое для записи санскрита, хинди и других языков Индии.

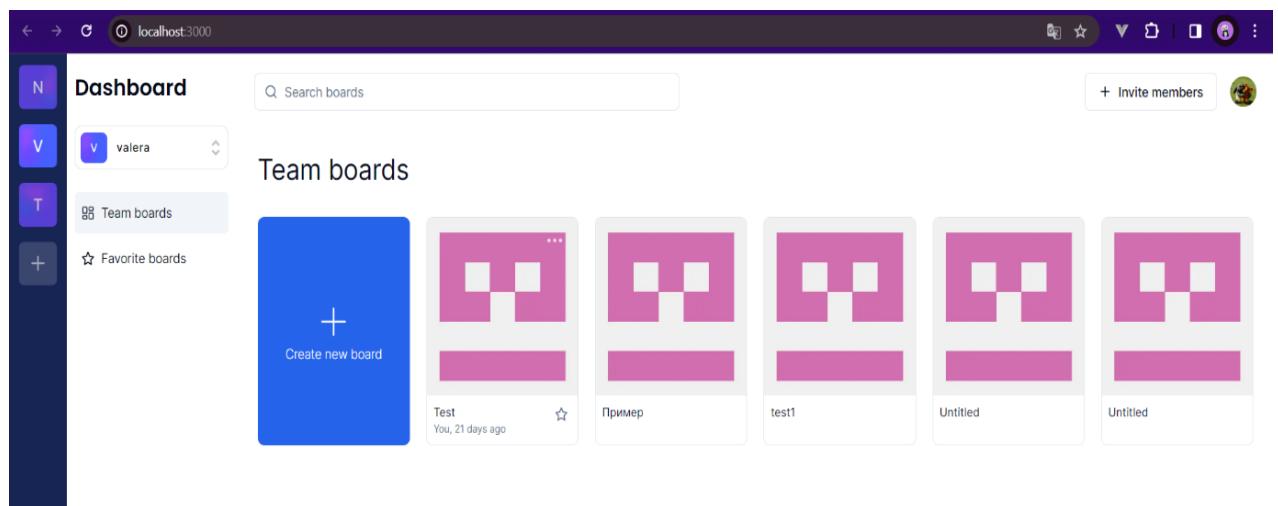


Рис. 2. Главная страница инструмента Tula

Боковая панель отображает список команд, что облегчает разделение задач между разными проектами. На этой панели можно создавать отдельные доски для каждой команды, приглашать участников для совместной работы и эффективно управлять коллaborацией. Доски предназначены для последующей работы и управления проектами.

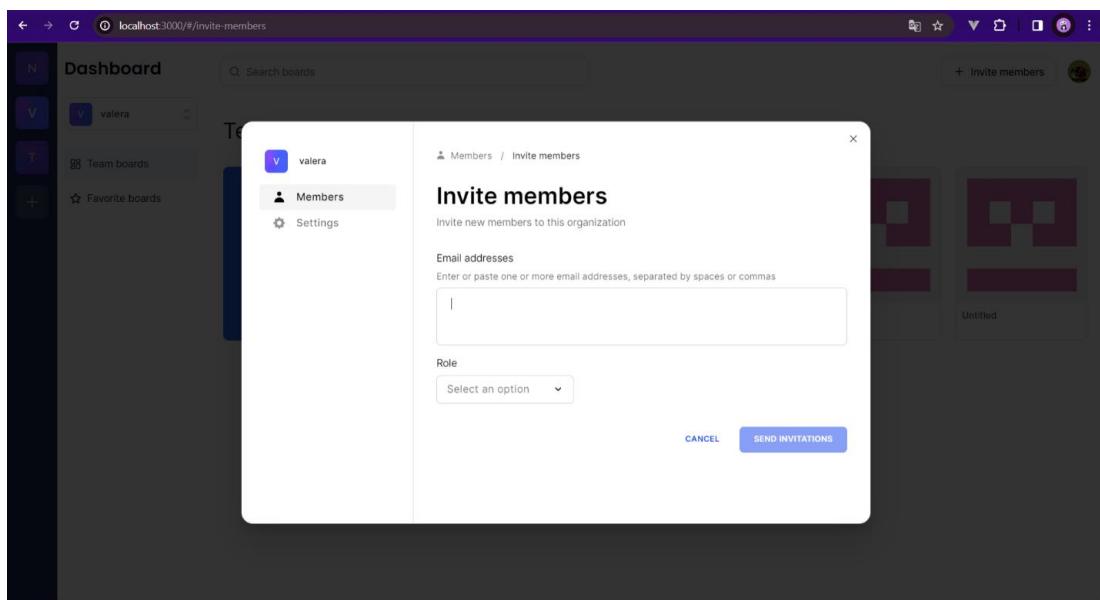


Рис. 3. Окно для отправки приглашений

Создав доску, пользователь имеет возможность отправить приглашения другим пользователям для совместной работы (рис. 3). Инструмент поддерживает реализацию совместной работы в режиме реального времени через

WebSocket¹⁴. Эта функция синхронизирована с Liveblocks для гарантии стабильности.

Архитектурно система автоматически управляет переходом в оффлайн-режим, восстановлением связи после ее потери и другими подобными сценариями, обеспечивая непрерывное функционирование.

Графовая система узлов

Игровая механика и симуляция изображаются с помощью диаграмм. Инструмент предоставляет пользователю визуальный язык для описания взаимодействия игровых элементов. Этот язык, базирующийся на наборе узлов, был разработан с учетом опыта использования схем в инструменте Machinations. Пользовательские отзывы (например, в [9]) подчеркивали сложность восприятия схем в Machinations, что стало толчком для внесения улучшений и будет описано ниже при проведении сравнения.

На левой панели инструмента (рис. 4) представлены иконки и описания компонентов, используемых в игровых схемах.

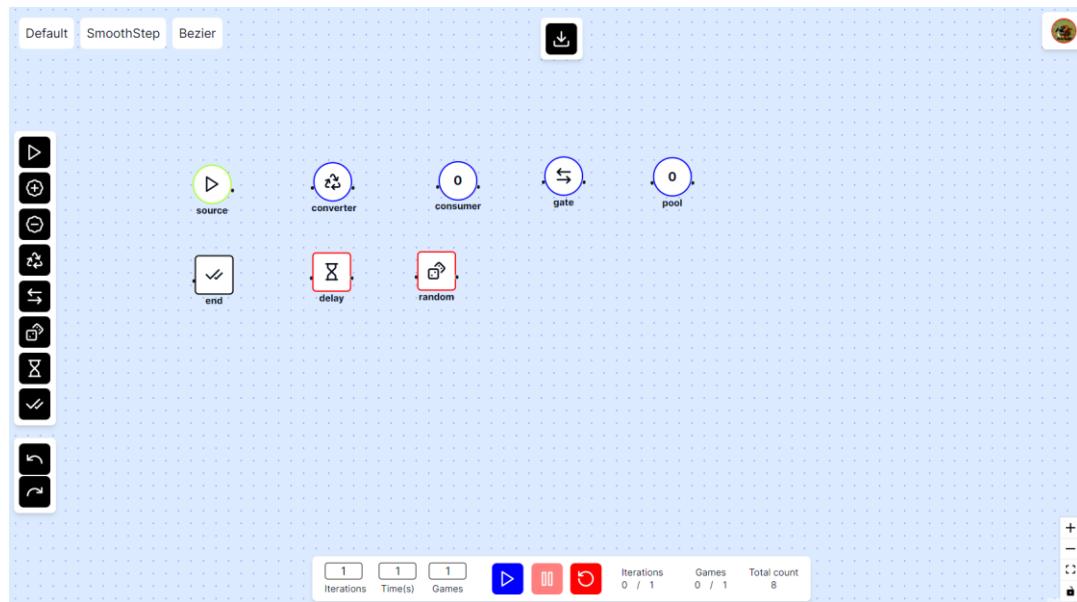


Рис. 4. Интерфейс доски

¹⁴ WebSocket – технология, которая позволяет клиенту установить двухстороннюю («дуплексную») связь с сервером.

Каждый узел выполняет определенную функцию в механике игры:

- Source – генерирует ресурсы;
- Pool – служит для хранения ресурсов;
- Consumer – потребляет ресурсы;
- Converter – преобразует один вид ресурса в другой;
- Gate – активирует процессы при выполнении заданных условий;
- Random – инициирует действия на основе случайных событий;
- Delay – обеспечивает времененную задержку в процессах;
- End – обозначает конечную точку игры.

Представленный язык спецификации обладает интуитивной простотой, что облегчает его освоение будущими пользователями. Это делает его доступным для широкого круга участников разработки разной направленности.

Для улучшения визуализации механики игры предоставлены три вида соединений, позволяющие связывать компоненты путем перетаскивания ручек от одного элемента к другому (рис. 5–7).

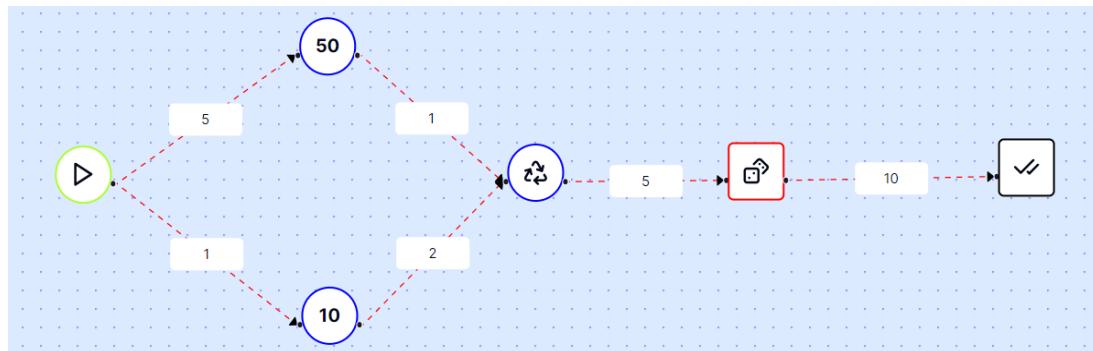


Рис. 5. Соединение по кратчайшему пути

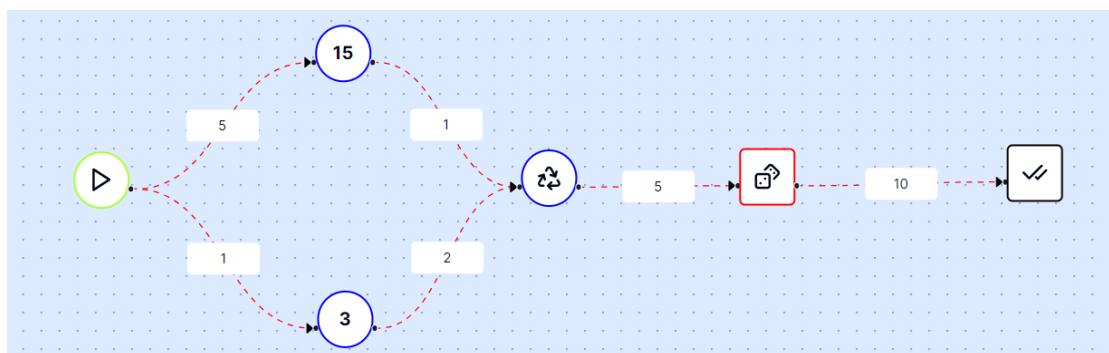


Рис. 6. Соединение кривой Безье

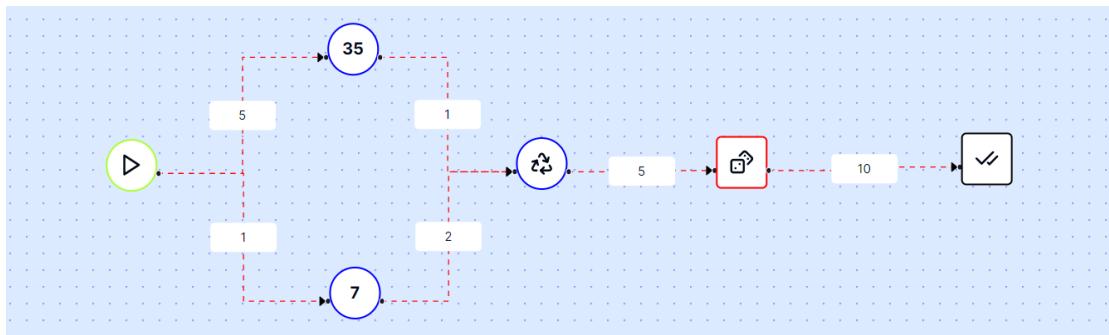


Рис. 7. Соединение прямыми линиями

Разработанный функционал позволяет гибко настраивать тестовые условия, включая количество шагов и продолжительность выполнения симуляции (рис. 8).



Рис. 8. Панель для запуска симуляций

Далее опишем панель подробнее.

- Первое поле служит для установки количества итераций, которые должна выполнить система в течение одной игровой сессии.
- Второе поле задает продолжительность одной итерации в секундах.
- Третье поле определяет количество игровых сессий, которые предстоит симулировать. Каждая сессия включает в себя ранее заданное количество итераций, что позволяет анализировать разнообразные распределения ресурсов в ходе игры.
- Кнопки управления на панели позволяют начать визуализацию симуляции, приостановить и возобновить процесс на определенном этапе. Последняя кнопка предназначена для сброса всех данных и начала симуляции с первой итерации первой сессии.
- Боковая часть панели отображает счетчик выполненных итераций и игровых сессий, а также показывает общее количество элементов, используемых в схеме, включая соединения и компоненты.

Алгоритм случайного распределения ресурсов

В разработке игровых систем ключевую роль играет учет случайных событий, которые вносят элемент непредсказуемости в игровой процесс.

Для более точного моделирования реальных условий игры был разработан компонент, который позволяет случайно распределять ресурсы (рис. 9). Этот компонент дает возможность задавать количество ресурсных единиц для распределения на входе. На выходе он связывается с другими элементами, которые принимают ресурсы с разной вероятностью. Соединения, выходящие из этого компонента, показывают пропорции распределения ресурсов. Например, при распределении 10 единиц ресурса 1/10 единицы перенаправляется по первому соединению, 2/10 – по второму и третьему, а 5/10 – по четвертому.

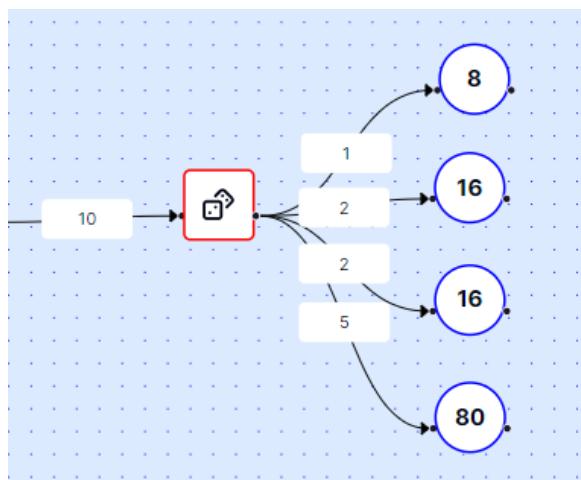


Рис. 9. Компонент для случайного распределения

Алгоритм случайного распределения ресурсов состоит из следующих этапов:

- 1) вычисление вероятностей;
- 2) распределение значений по вероятности;
- 3) распределение оставшихся ресурсов;
- 4) фиксация результата распределения.

На этом этапе вычисляется вероятность отправки ресурса по каждой из ветвей. Для этого сначала суммируются значения всех ребер, чтобы получить общее значение, которое доступно для распределения. Затем каждое значение ветви делится на общее значение, чтобы определить вероятность для каждой ветви

$$p_i = e_i / \sum_{i=1}^n e_i,$$

где i – натуральное число, p_i – вероятность отправки ресурса по i -й ветви, а e_i – значение i -го ребра.

Далее происходит распределение ресурсов с учетом вычисленных вероятностей:

- 1) генерируется случайное число от 0 до оставшегося количества ресурсов;
- 2) полученное число умножается на вероятность определения количества ресурсов, которые будут отправлены по этой ветви; результат округляется в меньшую сторону с помощью функции Math.floor;
- 3) определенное количество ресурсов вычитается из общего доступного количества и отправляется по текущей ветви;
- 4) процесс повторяется для каждой ветви.

Если после основного распределения остались нераспределенные ресурсы, они равномерно распределяются между всеми ветвями. Дополнительно используются случайные числа для обеспечения разнообразного и случайного распределения оставшихся ресурсов.

После завершения всех этапов распределения возвращается массив, где каждый элемент представляет количество ресурсов, отправленных по каждой ветви.

Аналитические возможности

Для визуализации потенциальных исходов игровой сессии использована линейная диаграмма. Инструмент автоматически рассчитывает среднее, медианное, минимальное и максимальное значения для каждого узла на каждом этапе итерации, что позволяет проводить всесторонний анализ баланса:

- среднее значение: позволяет оценить общую тенденцию изменений во время симуляции, отражая основные тенденции данных;
- медианное значение: помогает выявить аномалии и сравнить их со средним значением, что способствует обнаружению выбросов;
- минимальное и максимальное значения: использованы для определения диапазона возможных изменений значений, что критически важно для понимания крайних границ изменений параметров.

Выполнение итераций и симуляций дает возможность вносить изменения в числовые характеристики системы, проводить эксперименты и детально анализировать получаемые результаты, что обеспечивает основу для информированного принятия решений по настройке баланса игровых параметров.

СРАВНЕНИЕ С MACHINATIONS

Практическое применение

Демонстрацию работы инструмента на примере создания прототипа игровой механики, включая настройку параметров и анализ результатов, можно сопроводить сравнением функциональности с аналогичными решениями в Machinations.

В качестве примера рассмотрим механику производства, в которой игрок получает деревья и камень для строительства домов.

- 1) Должен быть создан источник генерации ресурсов, который описывает добычу материалов игроком.
- 2) Должны быть добавлены в систему два узла для хранения ресурсов (Камни и Дерево).
- 3) Конвертер «Построить Дом» будет преобразовывать ресурс «Дерево» в «Дома».

- 4) Нужно установить связи между узлами и указать количество единиц ресурса, участвующих в игровой механике.

Дадим общее описание механики: игрок за одну игровую итерацию добывает 1 единицу Wood (Дерево) и 2 единицы Stone (Камень). Когда количество Дерева достигает 5 единиц, функция Build House (Построить Дом) извлекает 5 ресурсов из пула Дерево и преобразует их в один ресурс Дома. Этот новый ресурс хранится в пуле построенных Домов.

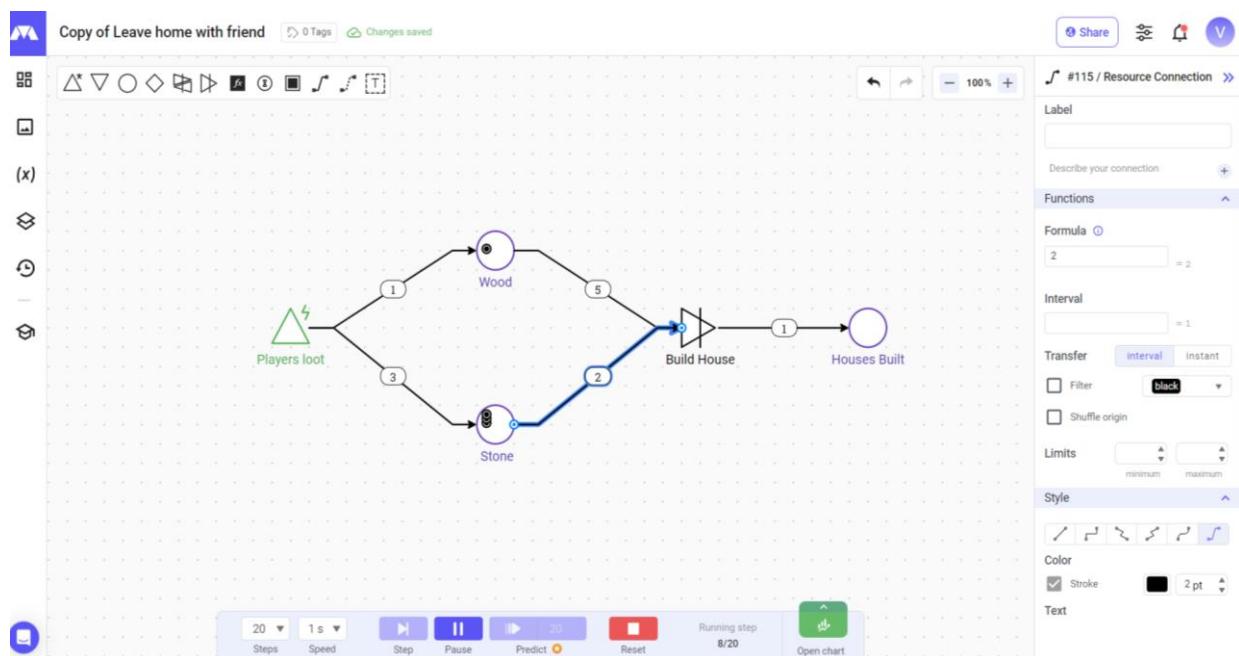


Рис. 10. Прототип игровой механики в Machinations

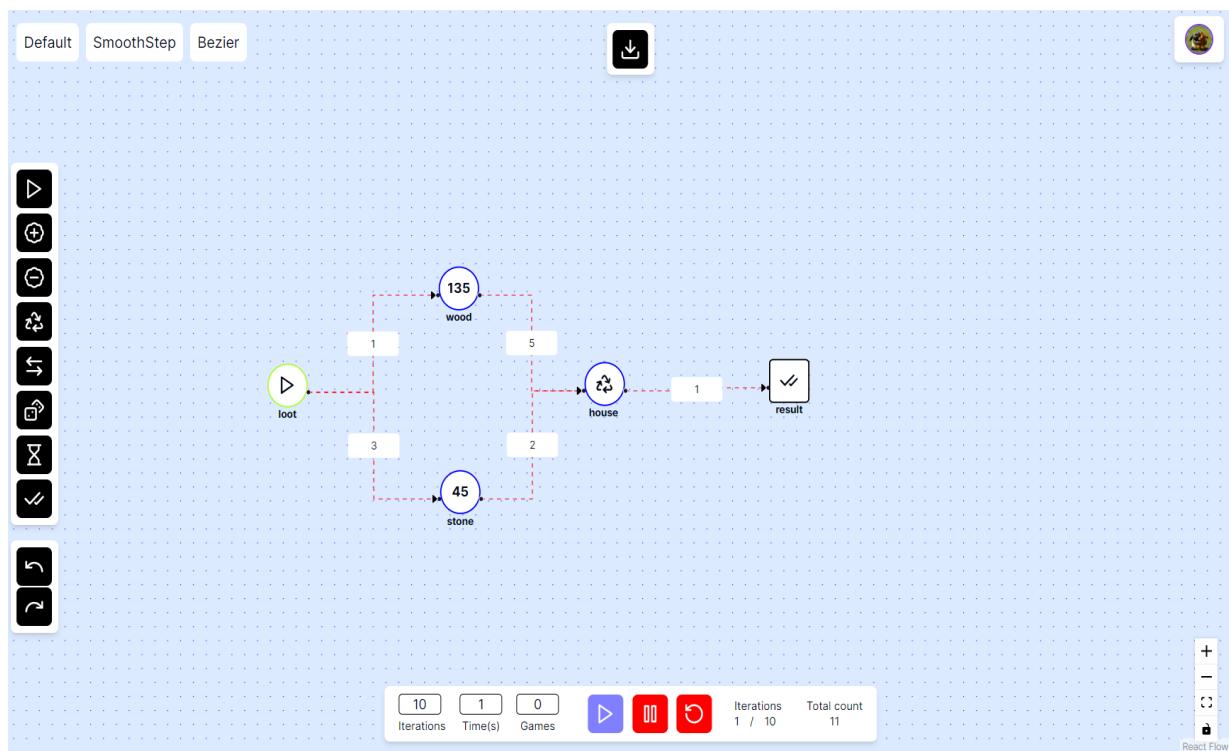


Рис. 11. Прототип игровой механики в Tula

Было принято решение убрать описательную панель, которая находилась справа от интерфейса (см. рис. 10 и 11), после того как были обнаружены проблемы с ее громоздкостью и неэффективным использованием.

В ходе улучшений системы внедрены дополнительные функциональные возможности, в том числе отмена последнего действия и возврат к предыдущему состоянию. Эти возможности осуществляются благодаря механизму хранения истории операций, реализованному с помощью стейтменеджера, что обеспечивает необходимые инструменты для создания прототипов игровых механик.

Валидация полей

Отсутствие валидации полей для ввода значений является существенным недостатком инструмента Machinations, обнаруженным в ходе проектирования. Система позволяет вводить текстовые значения (рис. 12) без какой-либо проверки или уведомления об ошибке, что приводит к запуску игровой симуляции с неверными данными и к некорректному поведению системы.

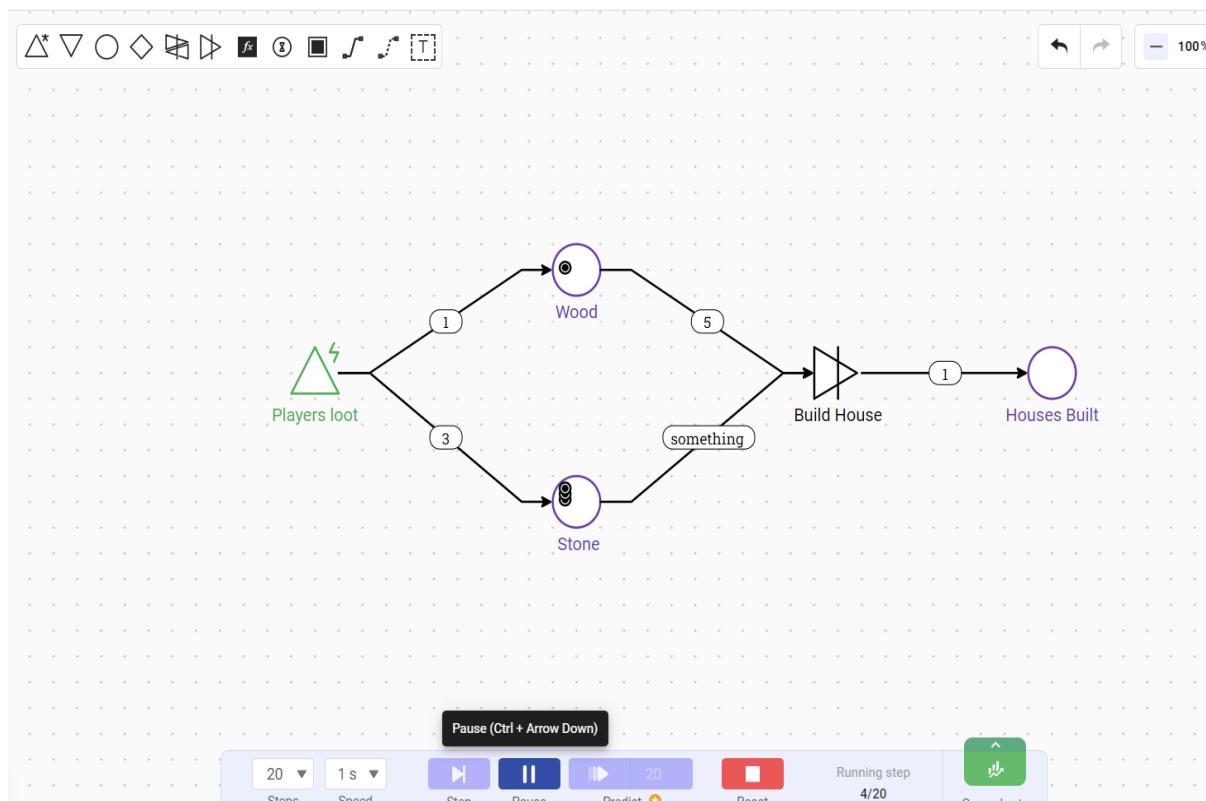


Рис. 12. Невалидные значения (в Machinations)

В инструмент Tula был внедрен механизм проверки валидации строковых значений. Если в поле ребра указано нечисловое значение, система отображает ошибку “Must be a numeric!” и блокирует возможность запуска симуляции (рис. 13). Кнопка запуска остается неактивной до тех пор, пока некорректное значение не будет исправлено.

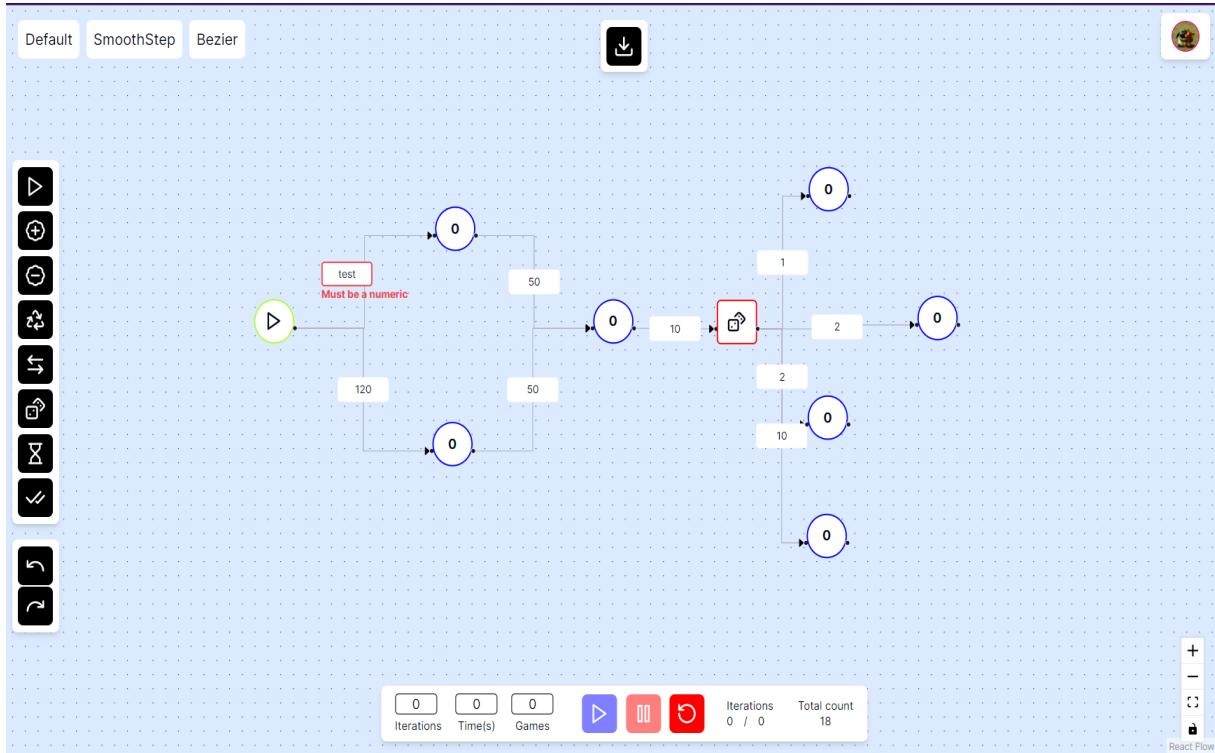


Рис. 13. Невалидные значения (в Tula)

Удаление узлов

Частые изменения в прототипах – это нормальное явление. Поэтому важно иметь возможность быстро удалять узлы и связи между ними.

В инструменте Machinations при удалении узлов (см. рис. 14) связи остаются на месте.

Если узел удаляется из середины цепочки, связь между смежными узлами не обновляется автоматически, что требует ручного удаления и создания новых связей, что отнимает много времени.

В разработанном инструменте Tula эта проблема была решена: после удаления узла связи автоматически пересчитываются и обновляются (рис. 15).

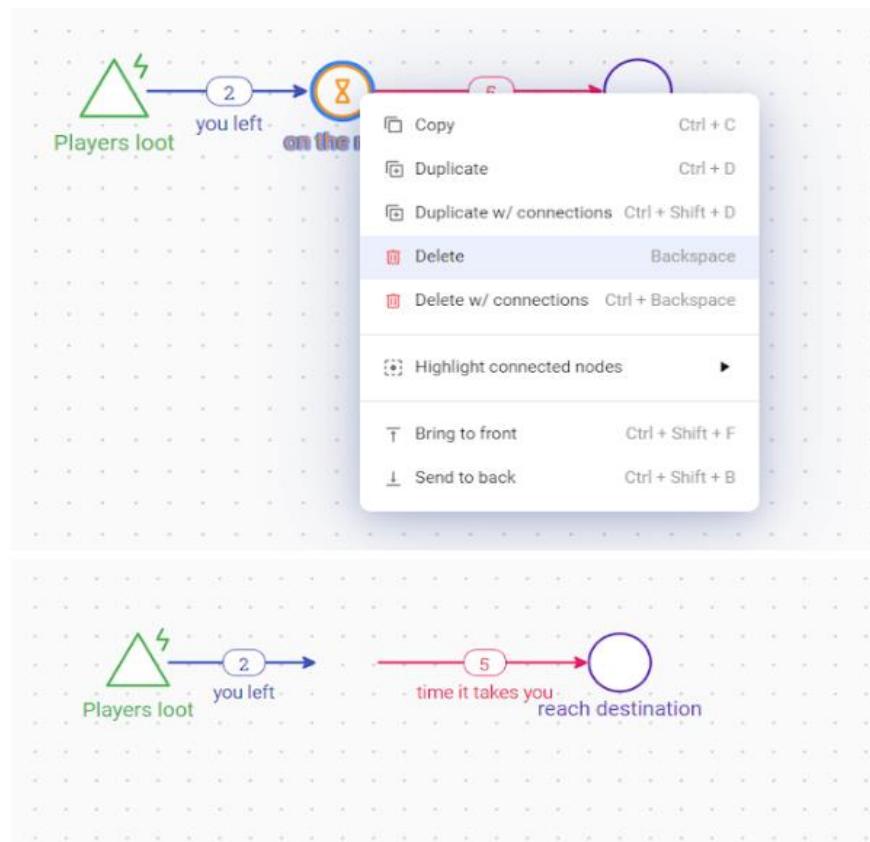


Рис. 14. Удаление узлов в Machinations

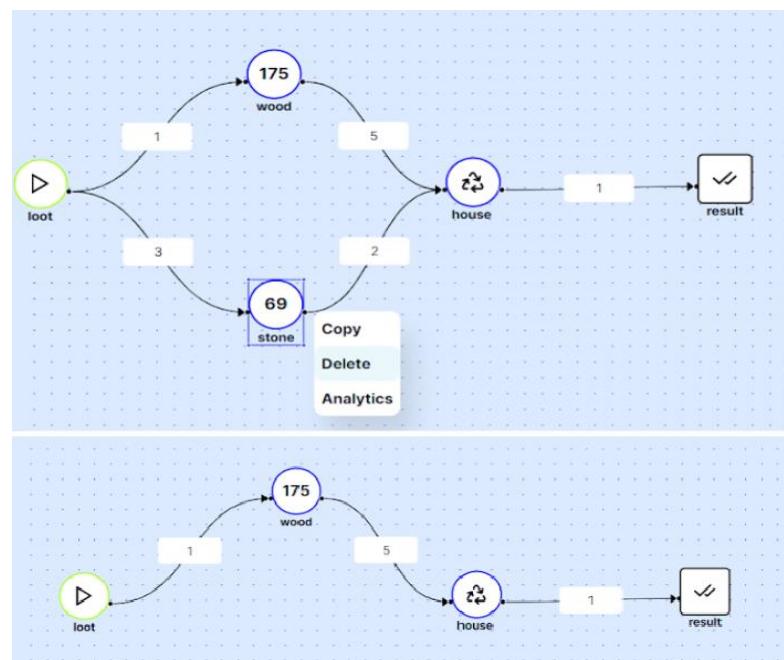


Рис. 15. Удаление узлов в Tula

Анализ значений

При анализе функциональности инструмента Machinations выяснилось, что пользователи часто сталкиваются со сложностью и избыточным объемом информации на аналитических диаграммах (рис. 16), что усложняет их понимание. С учетом отмеченной проблемы процесс вывода результатов аналитических диаграмм был упрощен (рис. 17), что делает их более доступными и понятными для пользователей.



Рис. 16. Результаты анализа значений игрового узла в Machinations

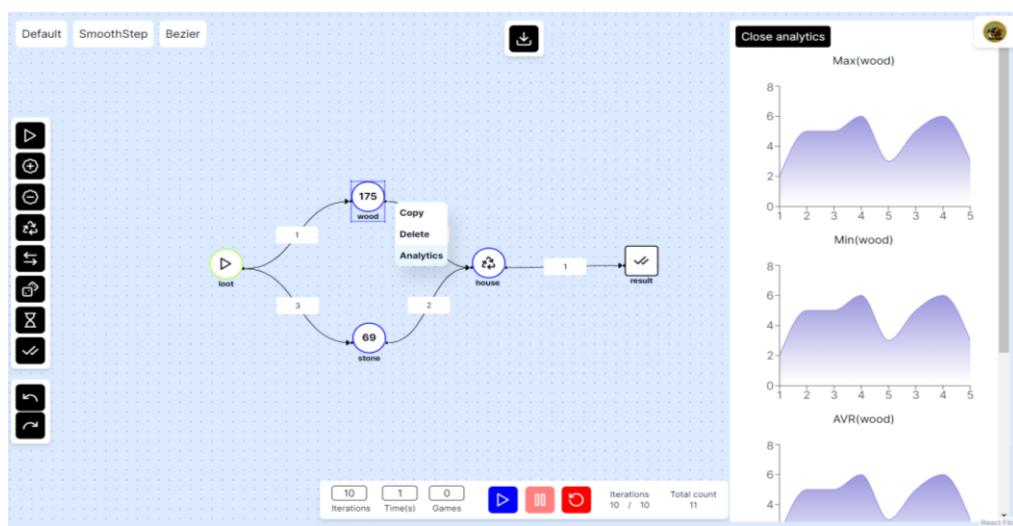


Рис. 17. Результаты анализа значений игрового узла в Tula

Технология также была улучшена путем добавления возможности отображения базовой диаграммы одним кликом на соответствующий узел после этапа игровой сессии для проведения анализа. Это позволяет пользователям мгновенно просматривать диаграмму, которая ясно указывает анализируемый узел и отображает выходные значения в форме понятных диаграмм.

Дополнительный функционал

В рамках созданной системы планируется разработка функционала, который облегчит генерацию первичной структуры игровой механики, на основе описания классов как на популярных языках программирования (например, JavaScript), так и при помощи формального описания (обычным текстом или, например, в JSON). Благодаря такому механизму будет предоставлен удобный кодовый редактор для создания исходной схемы игровой механики с использованием уже существующих описаний элементов.

Эта инициатива находит подтверждение в [3], где подробно рассмотрен потенциал прототипирования игр с применением семантического анализа и машинного обучения на базе текстовой документации. Информация, содержащаяся в документации, может быть эффективно использована для автоматической генерации первичного прототипа.

ЗАКЛЮЧЕНИЕ

Представлен подробный анализ и разработан инструмент для балансировки видеоигр. Продемонстрированы теоретические основы, практическое применение и сравнительный анализ, подчеркивающий значимость инструмента для индустрии видеоигр. Учитывая представленные перспективы развития реализованного инструмента TULA, мы видим его потенциал для дальнейшего улучшения и расширения функциональности. Развитие инструмента сыграет важную роль в индустрии разработки видеоигр, что обеспечивает разработчиков более эффективными средствами для создания сбалансированных видеоигр. Разработанный инструмент TULA является частью более широкого исследования по формализации игровых механик и их верификации. В

перспективе планируется его интеграция с методами автоматической генерации прототипов на основе формальных спецификаций [2], что позволит создать комплексную среду разработки и балансировки игр.

Благодарности

Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»).

СПИСОК ЛИТЕРАТУРЫ

1. Сахибгареева Г.Ф., Кугуракова В.В., Большаков Э.С. Инструменты балансирования игр // Электронные библиотеки. 2023. Т. 26, № 2. С. 225–251.
2. Кугуракова В.В. Формальный подход к пространственно-временному моделированию игровых систем // Ученые записки Казанского университета. Серия Физико-математические науки. 2024. Т. 166. № 4. С. 532–554.
3. Rouse R. Game Design: Theory and Practice: Theory and Practice. Jones & Bartlett Learning, 2004.
4. Game Balance Concepts [Электронный ресурс].
URL: <https://gamebalanceconcepts.wordpress.com> (17.09.25)
5. Becker A., Görlich D. What is game balancing? An examination of concepts // ParadigmPlus. 2020. Vol. 1, No. 1. P. 22–41.
6. Andrade G. et al. Dynamic game balancing: An evaluation of user satisfaction // Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. 2006. Vol. 2, No. 1. P. C. 3–8.
7. Gonçalves D. et al. The Trick is to Stay Behind?: Defining and Exploring the Design Space of Player Balancing Mechanics // Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024. P. 1–16.
8. Сахибгареева Г.Ф., Кугуракова В.В., Большаков Э.С. Генерация и балансирование игровых механик видеоигр // Научный сервис в сети Интернет: труды XXIV Всероссийской научной конференции. 2022. С. 455–485.
9. Галимзянов Г.Р. Разработка инструмента автоматической корректировки внутриигровых параметров: выпускная квалификационная работа 09.03.04. Казань, 2021. 35 с.

TULA ONLINE TOOL FOR BALANCING VIDEO GAMES

V. R. Rakhmankulova¹ [0009-0006-8425-616X], V. V. Kugurakova² [0000-0002-1552-4910]

^{1, 2}Institute of Information Technologies and Intelligent Systems, Kazan Federal University, 35 Kremlyovskaya st., Kazan, 420008

¹raxmankulova.v@mail.ru, ²vlada.kugurakova@gmail.com

Abstract

This paper presents the development of Tula, a tool for video game balancing. The necessity for such a tool is substantiated by the growing requirements for quality and cost-effectiveness in the video game industry, particularly in managing in-game economy and game world logic. The study analyzes existing tools and approaches to game balancing, identifying their limitations, which informed the design of the new tool's functionality. The presented tool integrates features of contemporary solutions while providing enhanced capabilities for game parameter analysis and testing, including prototype generation via class descriptions and real-time simulation. The technological foundation and architecture of the tool are described in detail. Key implementation aspects are discussed: interface responsiveness, continuous data synchronization, and security. Comparative analysis with Machinations revealed advantages in data processing correctness, interface convenience, and prototype modification flexibility.

Keywords: *video games, gameplay, game mechanics, game balance, game design, Machinations.*

Acknowledgment: This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program ("PRIORITY-2030").

REFERENCES

1. Sahibgareeva G.F., Kugurakova V.V., Bolshakov E.S. Video Game's Mechanics Generation and Balancing // In CEUR Workshop Proceeding, 2022. P. 455–485.
 2. Kugurakova V.V. A Formal Approach to Spatio-Temporal Modeling of Game Systems // Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki. 2024. Vol. 166, Is. 4. P. 532–554 (In Russ.)
-

3. *Rouse R.* Game Design: Theory and Practice: Theory and Practice. Jones & Bartlett Learning, 2004.
4. Game Balance Concepts.
URL: <https://gamebalanceconcepts.wordpress.com> (17.09.25)
5. *Becker A., Görlich D.* What is game balancing? An examination of concepts // ParadigmPlus. 2020. Vol. 1, No. 1. P. 22–41.
6. *Andrade G. et al.* Dynamic game balancing: An evaluation of user satisfaction // In Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. 2006. Vol. 2, No. 1. P. 3–8.
7. *Gonçalves D. et al.* The Trick is to Stay Behind?: Defining and Exploring the Design Space of Player Balancing Mechanics // In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024. P. 1–16.
8. *Sahibgareeva G.F., Kugurakova V.V., Bolshakov E.S.* Game Balance Tools // Russian Digital Libraries Journal. 2023. Vol. 26, No. 2. P. 225–251.
9. *Galimzyanov G.R.* Development of a tool for automatic adjustment of in-game parameters [Razrabotka instrumenta avtomaticheskoi korrektirovki vnutriigrovyh parametrov]: graduation. qualification. work on Bachelor, specialty Software Engineering. 2021. Kazan Federal University. 35 p. (In Russ.)

СВЕДЕНИЯ ОБ АВТОРАХ



РАХМАНКУЛОВА Валерия Рашидовна – выпускник бакалавриата Института ИТИС КФУ.

Valeria Rashidovna RAKHMANKULOVA – graduate at the Department of Software Engineering of the Institute of ITIS KFU.

email: raxmankulova.v@mail.ru

ORCID: 0009-0006-8425-616X



КУГУРАКОВА Влада Владимировна – доцент, кандидат технических наук, зав. кафедрой индустрии разработки видеоигр Института ИТИС КФУ.

Vlada Vladimirovna KUGURAKOVA, Ph.D. of Engineering Sciences, Head of the Video Game Development Industry Department of ITIS KFU.

email: vlada.kugurakova@gmail.com

ORCID: 0000-0002-1552-4910

Материал поступил в редакцию 1 сентября 2025 года

УДК 004.81

АЛГОРИТМИЧЕСКИЙ ФРЕЙМВОРК ДЛЯ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИОННОГО ЯДРА ВЕБ-СТРАНИЦЫ

Х. Салем¹ [0000-0002-9143-5231], А. С. Тощев² [0000-0003-4424-6822]

¹Университет Иннополис, г. Иннополис, Россия

²Казанский федеральный университет, г. Казань, Россия

¹h.salem@innopolis.ru, ²atoschev@kpfu.ru

Аннотация

Представлен новый точный алгоритм МСЕ извлечения основного содержимого с новостных веб-сайтов. Предложенный алгоритм использует анализ структуры объектной модели документа (DOM) и метрики плотности контента для идентификации и извлечения информационного ядра веб-страницы. Реализованный подход объединяет три ключевые особенности: максимальное количество прямых дочерних элементов с текстом, максимальное текстовое содержимое без дочерних элементов, содержащих текст, и ближайшее расположение к средней глубине узла. Алгоритм продемонстрировал лучшую производительность по сравнению с существующими решениями, такими как Boilerpipe и Readability, достигая 99.96% точности, 99.69% полноты и 99.80% F1-меры на использованном комплексном наборе данных из 500 разнообразных веб-страниц. Языково-независимый дизайн делает алгоритм особенно эффективным для извлечения мультиязычного контента, включая языки со сложной структурой, такие, например, как арабский.

Ключевые слова: NLP, извлечение данных, языково-независимый алгоритм, RAG (Retrieval-Augmented Generation).

Введение

Современные веб-страницы содержат обилие разнообразной информации [1], где основной текст окружен шумовыми элементами — меню, рекламой, шапками, подвалами [2]. Задача выделения информационного ядра сводится к отделению тела статьи от незначимых компонентов [3], как показано на рис. 1. Отделение информационного ядра от шаблонных элементов представляет серьезную

проблему [4], особенно критичную для функции browsing (поиск в вебе) в приложениях с использованием какой-либо большой языковой модели LLM (ChatGPT, Perplexity, DeepSeek) [5, 6].

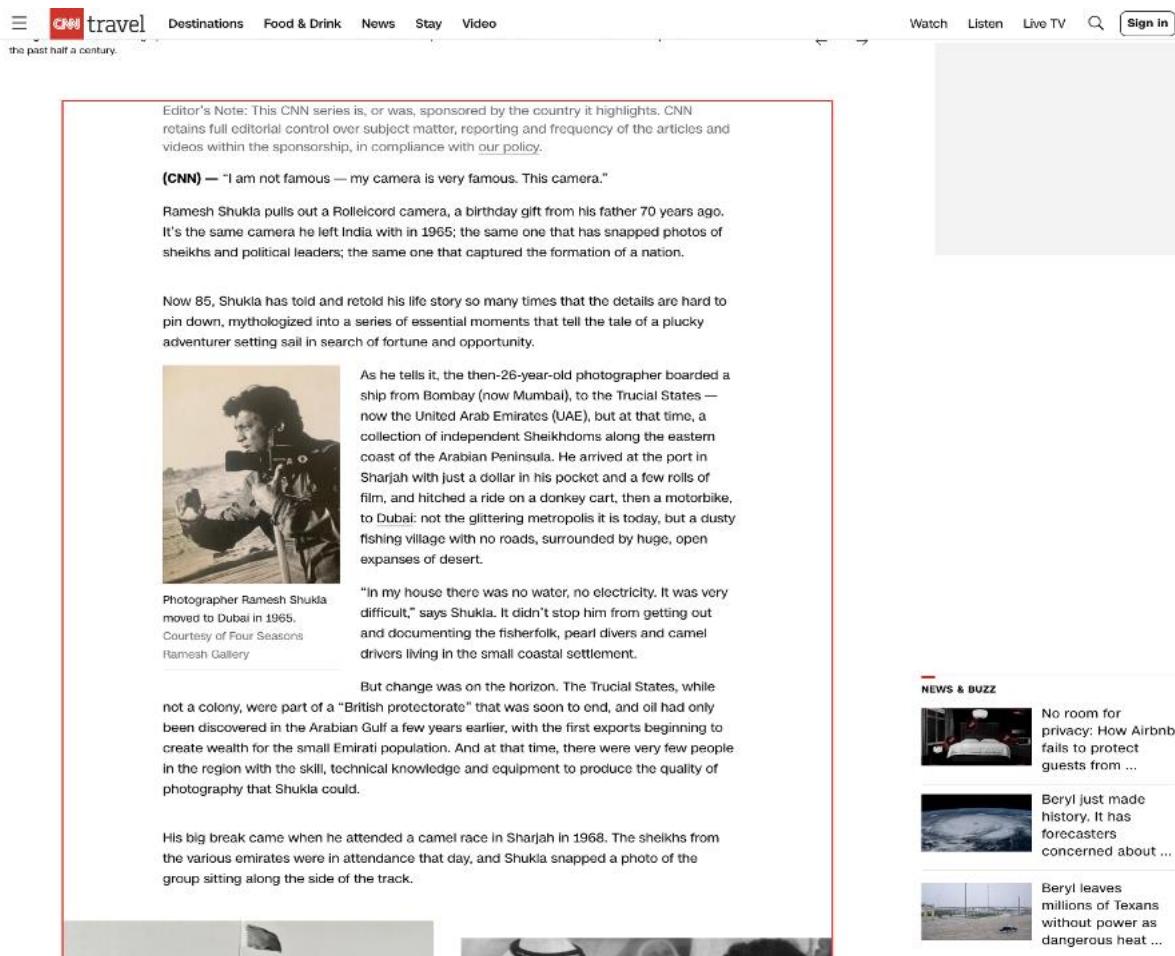


Рис. 1. Пример обнаружения информационного ядра в новостной статье (обведено красным)

Системы RAG (генерации с подключением поиска/извлечения) сначала собирают текст со страниц, а затем передают его языковой модели, чтобы она давала более точные ответы. Однако LLM-сервисы не являются дешевыми: их бизнес-модель основана на оплате токенов, а точность извлечения влияет на стоимость. Популярные библиотеки для извлечения текста с веб-страниц (Boilerpipe, Readability, Trafilatura) нередко захватывают посторонние элементы, что приводит к росту объема данных и увеличению затрат.

Настоящая работа предлагает алгоритм, определяющий информационное ядро веб-страницы и использующий связи в DOM, эвристические методы и анализ плотности текста с фильтрацией ненужных элементов.

ОПИСАНИЕ АЛГОРИТМА

Алгоритм решает задачу идентификации элемента в веб-документе с наибольшим количеством дочерних элементов и слов. Эта идея основана на гипотезе, что именно такие элементы служат основными контейнерами значимого текста для анализа, резюмирования или улучшения контента. Для проверки и практической реализации гипотезы была использована типичная структура сайтов статей и блогов, где основной блок обычно включает несколько дочерних элементов “`<div>`” или “`<p>`” с текстом. Например, современные CMS с визуальными редакторами генерируют HTML, следующий данной структуре (рис. 2).

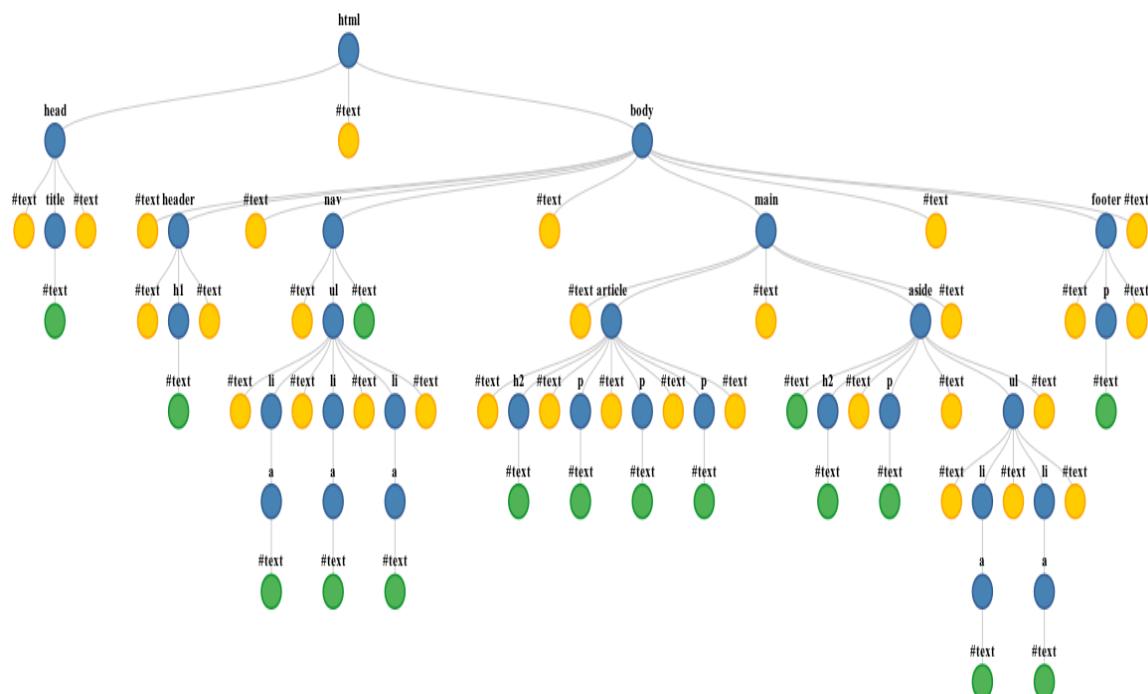


Рис. 2. Максимальное количество прямых дочерних элементов с текстом

Пусть N – множество всех узлов в дереве DOM. Для каждого узла $n \in N$ определим следующие величины (1-3):

$$C(n) = \text{количество прямых дочерних элементов узла } n, \quad (1)$$

$$T(n) = \text{общая длина текста прямых дочерних элементов узла } n. \quad (2)$$

Мы стремимся найти узел n_{\max} , который максимизирует комбинацию $C(n)$ и $T(n)$:

$$n_{\max} = \arg \max_{n \in N} (C(n), T(n)). \quad (3)$$

Еще одной характерной особенностью информационного ядра является наличие значительного объема текста в узле без текстовых дочерних элементов. Другими словами, элемент html “div” с крупным текстовым блоком рассматривается как более вероятный кандидат на основной контент. Формально для узла $n \in N$ обозначим через $L(n)$ длину текста узла. Пусть $H(n)$ — функция, принимающая значение 1, если у узла есть дочерние элементы с текстом, и 0 — в противном случае. Мы ищем узел n_{\max} , максимизирующий $L(n)$ при условии отсутствия дочерних текстовых элементов (4).

$$n_{\max} = \arg \max_{n \in N} (L(n) (1-H(n))). \quad (4)$$

Последняя выявленная нами особенность связана с расположением узла на глубине, близкой к средней по дереву DOM. Она основана на наблюдении, что глубина HTML-структурь отражает положение элемента на странице: слишком поверхностные узлы обычно содержат навигацию или служебные блоки, а чрезмерно глубокие — второстепенные элементы. Пусть $D(n)$ обозначает глубину узла n в DOM (корневой узел имеет глубину 0). Среднюю глубину M определим как (5).

$$M = |N|/2, \quad (5)$$

где $|N|$ — общее количество узлов. Определим узел n_{closest} с глубиной, ближайшей к M :

$$n_{\text{closest}} = \operatorname{argmin}_{n \in N} |D(n) - M|. \quad (6)$$

Итоговый набор узлов основного содержимого S (7) получается объединением узлов, идентифицированных тремя особенностями:

$$S = \{n_{\max-\text{children}}, n_{\max-\text{text}}, n_{\text{closest}-\text{middle}}\}. \quad (7)$$

Для определения первичного узла основного содержимого используем геометрический подход, основанный на векторах признаков. Каждый узел представим как 3D-вектор [количество дочерних элементов, количество слов, глубина]. Центроид с множества m 3-мерных векторов $x_i = [x_{i1}, x_{i2}, x_{i3}]$ зададим так (8).

$$c_j = \left(\frac{1}{m}\right) \sum_{i=1}^m x_i \text{ для } j = 1, 2, 3. \quad (8)$$

Для каждой компоненты c_j центроида имеем (9)

$$c_j = \left(\frac{1}{m}\right) \sum_{i=1}^m x_{ij}, \text{ для } j = 1, 2, 3, \quad (9)$$

где $x_i = [\text{Количество дочерних элементов с текстом } i, \text{ Количество слов } i, \text{ Глубина } i]$. Для измерения близости используем манхэттенское расстояние (10)

$$d(a, b) = |a_1 - b_1| + |a_2 - b_2| + |a_3 - b_3|. \quad (10)$$

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА

Алгоритм МСЕ был валидирован на репрезентативной выборке веб-документов [7, 13] размером 500 документов (уровень доверия 95%, предел ошибки 5%). Была также использована стратифицированная случайная техника выборки по языку, типу содержимого и сложности макета. Оценка проводилась с использованием стандартных метрик: точность, полнота, F1-мера и accuracy. Сравнительный анализ с алгоритмами Boilerpipe [8] и Readability [9] показал превосходящую производительность МСЕ по всем метрикам (табл. 1). Результаты сравнения обобщены и представлены в табл. 1.

Табл. 1. Метрики классификации (основной датасет)

Words Count Text	Words Count MCE	Words Count Readability	Words Count Boilerpipe	TP-C	FP-C	FN-C	TP-R	FN-R
232	232	247	131	232	0	0	232	0
220	220	239	257	220	0	0	220	0
208	208	187	199	208	0	0	208	21
432	434	404	437	434	2	0	434	30
425	425	442	341	425	0	0	425	0
220	220	152	172	220	0	0	220	68
<hr/>								

В табл. 2-4 приведены полученные значения метрик Precision, Recall, F1-score и Accuracy соответственно для методов MCE, Readability, Boilerpipe.

Табл. 2. Метрики классификации для метода MCE

Original Text	Precision	Recall	F1-score	Accuracy
232	1	1	1	1
220	1	1	1	1
208	1	1	1	1
432	0.995	1	0.998	0.995
425	0.962	1	0.981	0.962
220	0.764	0.866	0.764	0.764

Табл. 3. Метрики классификации для метода Readability

Original Text	Precision	Recall	F1-score	Accuracy
232	0.939	1	0.969	0.939
220	0.920	1	0.959	0.920
208	0.908	1	0.951	0.908
432	0.935	1	0.967	0.935
425	0.962	1	0.980	0.962
220	0.782	0.878	0.782	0.782

Табл. 4. Метрики классификации для метода Boilerpipe

Original Text	Precision	Recall	F1-score	Accuracy
232	0.565	0.722	0.565	0.565
220	0.874	0.933	0.874	0.874
208	1	0.957	0.978	1
432	0.988	1	0.994	0.988
425	1	0.802	0.890	1
220	1	0.782	0.878	1

Табл. 5 обобщает усредненные по множеству запусков метрики производительности (Precision, Recall, F1-Score, Accuracy) трех алгоритмов классификации: MCE, Readability и Boilerpipe и дает оценку эффективности каждого алгоритма.

Табл. 5. Метрики производительности классификации для датасета

Metric	MCE	Readability	Boilerpipe
Precision	0.9996%	0.9236%	0.9425%
Recall	0.9969%	0.9419%	0.8919%
F1-Score	0.9980%	0.9183%	0.8997%
Accuracy	0.9965%	0.8654%	0.8343%

Дополнительная оценка на наборе данных Webz.io [11], который представляет собой репозиторий публичных новостных статей с еженедельными выпусками тематических датасетов, содержащих около 1000 новостных статей с богатой метаинформацией, подтвердила эффективность всех алгоритмов, перечисленных выше, на различных типах контента (см. табл. 6).

Табл. 6. Метрики производительности классификации для датасета Webz.io

Metric	MCE	Readability	Boilerpipe
Precision	0.9615%	0.9136%	0.9225%
Recall	0.9597%	0.9219%	0.8819%
F1-Score	0.9468%	0.9133%	0.8897%
Accuracy	0.9212%	0.8554%	0.8143%

Как видим, алгоритм MCE сохраняет лидерство с Precision (96.15%), Recall (95.97%) и F1-Score (94.68%), хотя эти значения примерно на 3–4 процентных пункта ниже, чем его производительность на основном датасете. Алгоритмы Readability и Boilerpipe вновь демонстрируют сопоставимые, но более слабые результаты. Примечательно, что все три алгоритма демонстрируют снижение Accuracy на 2–5% на этом датасете, что позволяет предположить наличие более сложных случаев классификации в контенте Webz.io.

Анализ ошибок выявил ложные положительные результаты (неправильная классификация рекламных блоков, навигационных меню, комментариев) и ложные отрицательные результаты (пропуск контента из-за нестандартной структуры HTML, динамического контента, сложных макетов). Языковой анализ показал высокую производительность на латинских языках (> 95%), сопоставимую на кириллических языках, и особую эффективность на арабском языке (94.2%). Алгоритм реализован в расширении Chrome Elkateb[12] и развернут в производстве на <https://gptafser.com>, где сгенерировал JSON-LD для 90000 страниц за 12 мин, улучшив индексацию и увеличив трафик.

ЗАКЛЮЧЕНИЕ

Предложенный алгоритм МСЕ продемонстрировал превосходство над существующими решениями с показателями точности 96.15% и полноты 95.97% при организации поиска. Ключевыми преимуществами алгоритма МСЕ являются языковая независимость, высокая точность, вычислительная эффективность и открытый исходный код. МСЕ превосходит Readability на 4.79% и Boilerpipe на 3.90% по точности, соответственно на 3.78% и 7.78% по полноте. Ограничения включают затруднения с динамическим контентом JavaScript, снижение точности на нестандартных макетах и оптимизацию только для текстового контента.

МСЕ находит применение в системах RAG, агрегации новостного контента и мобильной оптимизации. Техническая реализация обеспечивает пропускную способность до 1000 страниц в минуту и использование памяти 50–100 МБ для типичной веб-страницы.

СПИСОК ЛИТЕРАТУРЫ

1. Jach T., Kaczmarek M., Kaczmarek T. Web content extraction: A survey of techniques and applications // Information Sciences. 2021. Vol. 570. P. 378–400.
<https://doi.org/10.1016/j.ins.2021.04.014>
2. Brown K., Davis L. Content density metrics for web page analysis // Information Retrieval Journal. 2020. Vol. 23, No. 4. P. 512–530.
<https://doi.org/10.1007/s10791-020-09380-4>
3. Gottron T. Content extraction from web pages // Proceedings of the 2008 ACM Symposium on Applied Computing. 2008. P. 1160–1164.
<https://doi.org/10.1145/1363686.1363939>

4. Insa D., Silva J., Tomás C. Using content extraction for web page classification // Information Processing & Management. 2013. Vol. 49, No. 1. P. 235–250.
<https://doi.org/10.1016/j.ipm.2012.05.005>
5. Qi X., Zhang Y., Wang L. Investigating the impact of content extraction on sentiment analysis // Information Processing & Management. 2024. Vol. 61, No. 1. 103245. <https://doi.org/10.1016/j.ipm.2023.103245>
6. Zhang W., Liu X. Machine learning approaches to content extraction // Pattern Recognition. 2022. Vol. 125. 108456.
<https://doi.org/10.1016/j.patcog.2022.108456>
7. Seitl F., Kovarik T., Mirshahi S. et al. Assessing the quality of information extraction // arXiv preprint. 2024. arXiv:2404.04068.
<https://arxiv.org/abs/2404.04068>.
8. Kohlschütter C. Boilerpipe: A Python library for extracting text from HTML // GitHub Repository. 2010. <https://github.com/misja/python-boilerpipe>
9. Mozilla Foundation. Readability: A Python library for extracting article content from HTML // GitHub Repository. 2020.
<https://github.com.mozilla/readability>
10. Bevendorff J., Gupta S., Kiesel J., Stein B. An empirical comparison of web content extraction algorithms // Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023. P. 2594–2603. <https://doi.org/10.1145/3539618.3591920>
11. Webz.io. Webz.io Free News Datasets // Webz.io. 2023.
<https://webz.io/free-news-datasets>
12. Research Team. Elkateb: Browser Extension for Content Extraction // Browser Extension. 2024. <https://github.com/elkateb/extension>
13. Бобырь М.В., Милостная Н.А., Булатников В.А. The fuzzy filter based on the method of areas' ratio // Applied Soft Computing. 2022. Vol. 117. 108449.
<https://doi.org/10.1016/j.asoc.2022.108449>

AN ALGORITHMIC FRAMEWORK FOR ACCURATELY EXTRACTING MAIN CONTENT FROM NEWS WEBSITES

H. Salem¹ [0000-0002-9143-5231], A. S. Toschev² [0000-0003-4424-6822]

¹Innopolis University, Innopolis, Russia

²Kazan Federal University, Kazan, Russia

¹h.salem@innopolis.ru, ²atoschev@kpfu.ru

Abstract

A new precise MCE algorithm for extracting the main content from news websites is presented. The proposed algorithm uses analysis of the Document Object Model (DOM) structure and content density metrics to identify and extract the informational core of a web page. The implemented approach combines three key features: the maximum number of direct child elements containing text, the maximum textual content without child elements containing text, and the closest position to the average node depth. The algorithm demonstrated superior performance compared to existing solutions such as Boilerpipe and Readability, achieving 99.96% precision, 99.69% recall, and 99.80% F1-score on a comprehensive dataset of 500 diverse web pages. Its language-independent design makes the algorithm particularly effective for extracting multilingual content, including languages with complex structures such as Arabic.

Keywords: *NLP, Data Extraction, Language-Independent Algorithm, RAG (Retrieval-Augmented Generation).*

REFERENCES

1. Jach T., Kaczmarek M., Kaczmarek T. Web content extraction: A survey of techniques and applications // Information Sciences. 2021. Vol. 570. P. 378–400.
<https://doi.org/10.1016/j.ins.2021.04.014>
2. Brown K., Davis L. Content density metrics for web page analysis // Information Retrieval Journal. 2020. Vol. 23, No. 4. P. 512–530.
<https://doi.org/10.1007/s10791-020-09380-4>
3. Gottron T. Content extraction from web pages // Proceedings of the 2008 ACM Symposium on Applied Computing. 2008. P. 1160–1164.
<https://doi.org/10.1145/1363686.1363939>

4. Insa D., Silva J., Tomás C. Using content extraction for web page classification // Information Processing & Management. 2013. Vol. 49, No. 1. P. 235–250.
<https://doi.org/10.1016/j.ipm.2012.05.005>
5. Qi X., Zhang Y., Wang L. Investigating the impact of content extraction on sentiment analysis // Information Processing & Management. 2024. Vol. 61, No. 1. 103245.
<https://doi.org/10.1016/j.ipm.2023.103245>
6. Zhang W., Liu X. Machine learning approaches to content extraction // Pattern Recognition. 2022. Vol. 125. 108456.
<https://doi.org/10.1016/j.patcog.2022.108456>
7. Seitl F., Kovarik T., Mirshahi S. et al. Assessing the quality of information extraction // arXiv preprint. 2024. arXiv:2404.04068.
<https://arxiv.org/abs/2404.04068>
8. Kohlschütter C. Boilerpipe: A Python library for extracting text from HTML // GitHub Repository. 2010.
<https://github.com/misja/python-boilerpipe>
9. Mozilla Foundation. Readability: A Python library for extracting article content from HTML // GitHub Repository. 2020.
<https://github.com.mozilla/readability>
10. Bevendorff J., Gupta S., Kiesel J., Stein B. An empirical comparison of web content extraction algorithms // Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023. P. 2594–2603. <https://doi.org/10.1145/3539618.3591920>
11. Webz.io. Webz.io Free News Datasets // Webz.io. 2023.
<https://webz.io/free-news-datasets>
12. Research Team. Elkateb: Browser Extension for Content Extraction // Browser Extension. 2024.
<https://github.com/elkateb/extension>
13. Bobyr M.V., Milostnaya N.A., Bulatnikov V.A. The fuzzy filter based on the method of areas' ratio // Applied Soft Computing. 2022. Vol. 117. 108449.
<https://doi.org/10.1016/j.asoc.2022.108449>

СВЕДЕНИЯ ОБ АВТОРАХ



Хамза САЛЕМ – аспирант, Университет Иннополис / Лаборатория программной инженерии, г. Иннополис.

Hamza SALEM – PhD student, Innopolis University / Software Engineering Lab, Innopolis.

email: h.salem@innopolis.ru

ORCID: 0000-0002-9143-5231



ТОЩЕВ Александр Сергеевич – доцент, к. н., КФУ / Институт информационных технологий и интеллектуальных систем / Кафедра программной инженерии, г. Казань.

Alexander Sergeevich TOSCHEV – Associate Professor, Ph.D., KFU / Institute of Information Technologies and Intelligent Systems / Department of Software Engineering, Kazan.

email:atoschev@kpfu.ru

ORCID: 0000-0003-4424-6822

УДК 621.396.969+519.6:621.396.969

МОДЕЛИРОВАНИЕ СЦЕНАРИЕВ РАБОТЫ РАДИОЛОКАЦИОННОЙ СИСТЕМЫ ДЛЯ КЛАССИФИКАЦИИ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ И ПТИЦ НА ОСНОВЕ МИКРОДОПЛЕРОВСКИХ СИГНАТУР В СРЕДЕ ENGEE

А. Е. Семичастнов¹ [0009-0003-2301-6267], Д. А. Балакин² [0000-0003-0497-7153]

^{1, 2}ООО ЦИТМ ЭКСПОНЕНТА, г. Москва, 117418, Россия

^{1, 2}Московский энергетический институт, г. Москва, 111250, Россия

¹semichastnov777@yandex.ru, ²dabalakin@yandex.ru;

Аннотация

Рассмотрен метод классификации беспилотных летательных аппаратов (БПЛА) и птиц по радарным измерениям при различных сценариях их движения. Актуальность решения этой задачи подтверждается сложностью обнаружения БПЛА ввиду небольших их габаритных размеров, высокой маневренности, а также схожей геометрией с птицами. Для решения задачи применен метод классификации, основанный на анализе микродоплеровских сигнатур (МДС), которые отражают динамику движения объектов. Моделирование сценариев выполнено в среде Engee, где разработаны модели радиолокационной системы (РЛС), БПЛА и птиц.

Ключевые слова: беспилотный летательный аппарат, микродоплеровские сигнатуры, радиолокационная система, диаграмма обратного рассеяния, цифровая обработка.

ВВЕДЕНИЕ

Беспилотные летательные аппараты (БПЛА) применяются как в гражданской, так и в военной сферах, что делает актуальным вопрос их надежного обнаружения, в том числе для защиты критически важных объектов [1].

Для обнаружения БПЛА применяются различные методы: радиочастотный анализ, акустические датчики, оптические системы и анализ радиочастотного тракта. Анализ радиочастотного тракта (например, идентификация сигналов

управления) теряет актуальность с развитием автономных БПЛА. Акустические системы чувствительны к шумам окружающей среды, а оптические камеры ограничены дальностью действия и ухудшением работы в плохих погодных условиях. Комплексный подход, объединяющий данные от нескольких источников, позволяет повысить надежность обнаружения за счет взаимной компенсации недостатков отдельных методов. Однако ключевую роль в таких системах играют радиолокационные технологии, обеспечивающие всепогодность и высокую дальность обнаружения [2]. Они позволяют отслеживать объекты на значительных дистанциях и в любых погодных условиях. Однако из-за того, что БПЛА и птицы имеют схожие размеры, часто возникают ложные срабатывания, что создает значительные проблемы. Поэтому требуется создание новых методов, позволяющих повысить эффективность обнаружения и снизить количество ошибок [3].

Радиолокационные методы сталкиваются с проблемой нехватки данных, поскольку сложно получить измерения при различных сценариях поведения объекта. Одним из решений этой проблемы является моделирование. Как правило, такие модели состоят из множества функциональных узлов, каждый из которых представляет собой модель. В связи с этим возникает сложность интеграции разнородных моделей в единую систему. В настоящей статье рассмотрены модель радиолокационной станции, модель формирования диаграммы обратного рассеяния цели и модель цифровой обработки. Представлен процесс интеграции этих моделей в единую систему. В качестве инструмента выступает среда моделирования Engree [4].

МОДЕЛЬ МОДЕЛЬ РАДИОЛОКАЦИОННОЙ СИСТЕМЫ (РЛС)

Структура базовой модели радиолокационной системы (РЛС) подробно рассмотрена в [5]. С учетом поставленной задачи можно выделить следующие основные функциональные узлы (рис. 1): генератор – формирует зондирующий сигнал; передатчик – усиливает и излучает электромагнитную волну в среду распространения; канал распространения – учитывает пространство между РЛС и целью; цель – характеризует отражательную способность исследуемого объекта; приемник – выполняет прием сигнала, отраженного от цели; цифровая обработка – реализует фильтрацию и расчет информативных характеристик (в рассматриваемой работе – микродоплеровские сигнатуры (МДС)).

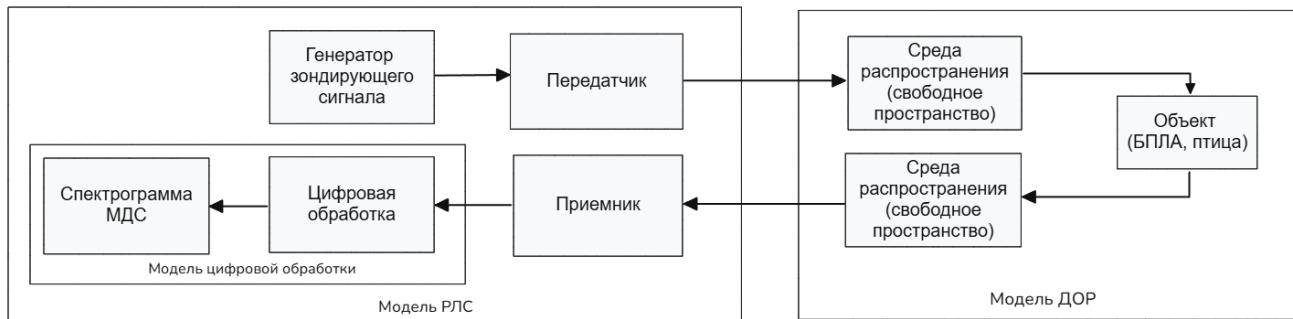


Рис. 1. Структурная схема системной модели для определения МДС

Основные тактико-технические характеристики РЛС (мощность, длительность и период повторения импульса, ширина спектра) задаются исходя из поставленных задач качества разрешения по дальности и скорости движения объектов. В настоящей работе реализована модель РЛС в среде моделирования Engee [4].

МОДЕЛЬ ДИАГРАММЫ ОБРАТНОГО РАССЕЯНИЯ ОБЪЕКТА

Ключевым этапом разработки модели является расчет диаграммы обратного рассеяния (ДОР) детектируемого объекта. Для рассматриваемых объектов (птиц или БПЛА) важно учитывать не только их геометрию, но и динамику движения, так как дополнительные движения (взмахи крыльев или вращение лопастей) обогащают спектrogramму отраженного сигнала периодическими составляющими. Такие компоненты спектра называют микродоплеровскими сигнатурами.

Вычисление ДОР реализуется с помощью подхода, заключающегося в разбиении целевого объекта на простейшие геометрические элементы (примитивы), для которых известно аналитическое выражение. Суммарный сигнал формируется по принципу суперпозиции отраженного сигнала от каждого примитива, рассматриваемого как элементарный излучатель.

Конструкция БПЛА типа «квадрокоптер» позволяет разделить его на следующие компоненты: четыре двухлопастных ротора (заменены на две плоские пластины) и корпус БПЛА, представленный цилиндрическими поверхностями (рис. 2а). ДОР данной модели БПЛА представлена на рис. 2б.

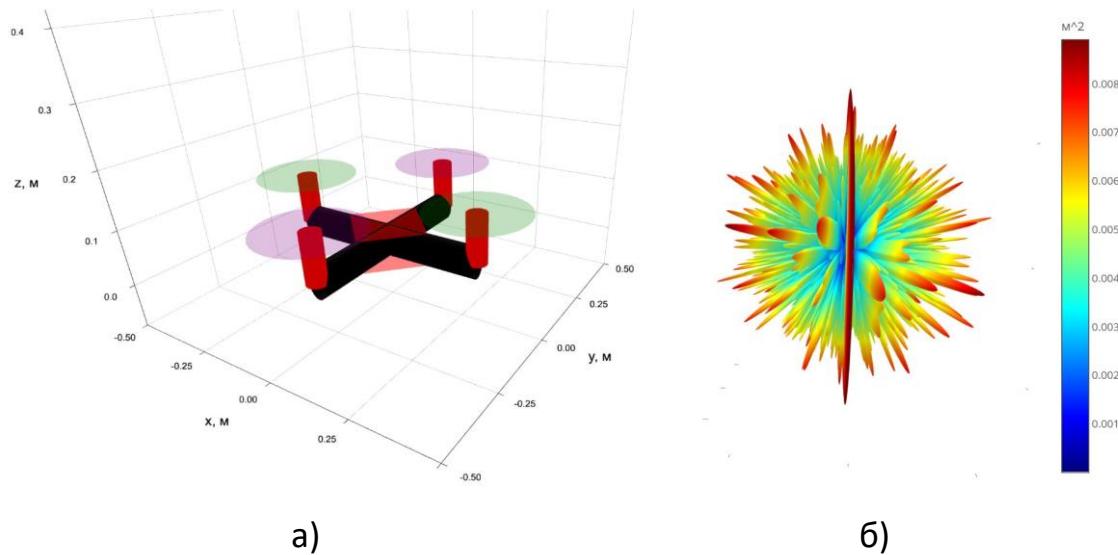


Рис. 2. Цифровая модель (а), диаграмма обратного рассеивания (б) БПЛА типа «квадрокоптер»

При характеристике птицы как объекта обнаружения следует принимать во внимание ее орнитологические характеристики и особенности полета. В строении крыла можно выделить несколько элементов: плечевую кость, которая является верхней частью крыла; предплечье, состоящее из локтевой и лучевой частей, и кисть. Движения крыльев представляют собой подъем и опускание верхней части предплечья относительно суставов. Основными элементами крыла, формирующими микродоплеровские сигнатуры, являются плечевая кость и предплечье. Кисть практически не влияет на спектр отраженного сигнала ввиду небольших размеров, поэтому для простоты кисть и предплечье рассматриваются как единый элемент. Такие элементы с достаточной точностью можно описать с помощью эллиптического тела (эллипсоида), ДОР которого аппроксимируется выражением

$$\sigma_{\text{ellip}}(\theta, \varphi) = \frac{\pi a^2 b^2 c^2}{(a^2 \sin(\theta) \cos^2(\varphi) + b^2 \sin^2(\theta) \sin^2(\varphi) + c^2 \cos^2(\theta))}, \quad (1)$$

где a, b, c – длины полуосей эллипсоида, φ – азимутальный угол, θ – полярный угол. Внешний вид и ДОР птицы представлены на рис. 3.

Выражение (1) позволяет вычислить значения эффективной площади рассеяния элементов птицы для любого направления прихода сигнала по азимутальному и полярному углам без применения ресурсоемких операций, что критически важно при моделировании сценариев движения в режиме реального времени.

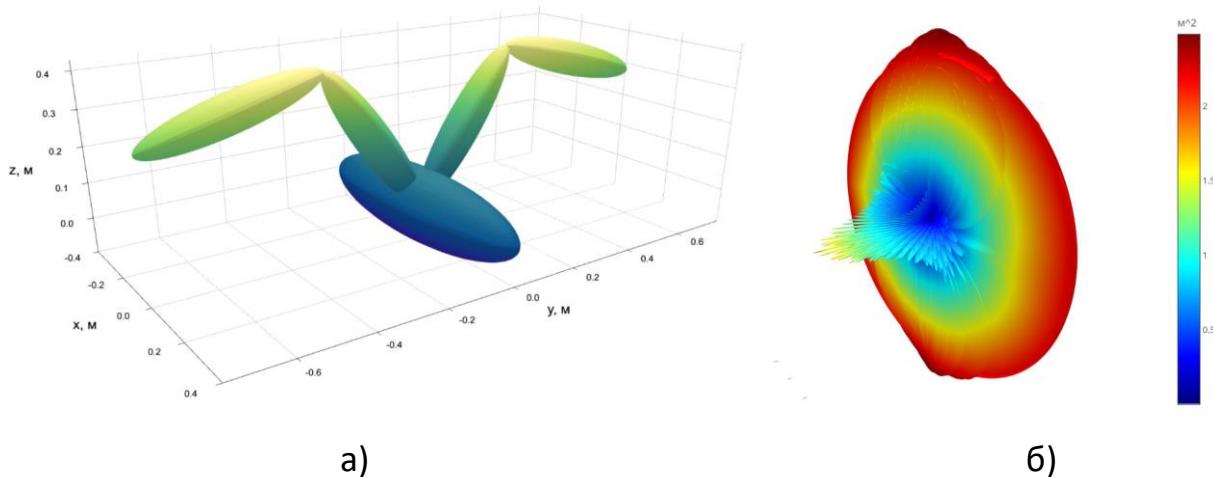


Рис. 3. Цифровая модель (а) и диаграмма обратного рассеивания (б) для птицы

После вычисления ДОР исследуемых объектов применим алгоритмы расчета отраженного сигнала из [6] и выполним запуск сценария, на выходе которого сформируются целевые МДС объектов. Результат моделирования работы РЛС при частоте вращения винта БПЛА, равной 100 Гц, и частоте взмаха крыльев птицы, равной 3 Гц, приведены на рис. 4.

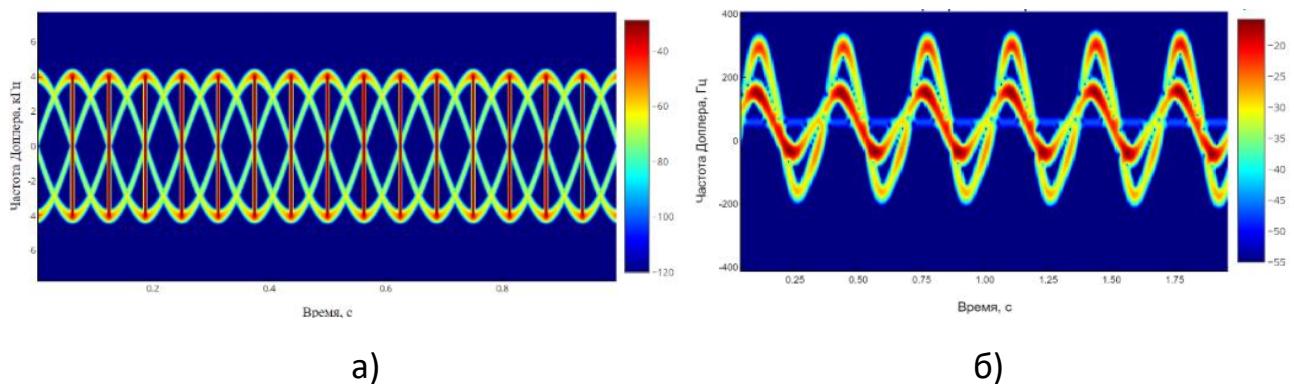


Рис. 4. Спектрограмма отраженного сигнала: а) БПЛА вертолетного типа;
б) птицы

Можно заметить, что характеры МДС для птицы и БПЛА отличаются, что позволяет решить задачу классификации. Благодаря периодической структуре МДС,

выполнив дополнительную обработку, можно извлечь параметры движения объекта: частоту вращения крыльев и скорость вращения винта.

ИССЛЕДОВАНИЕ СЦЕНАРИЕВ РАБОТЫ РЛС

Финальный этап формирования модели состоит в интеграции ранее разработанных структурных частей в общую системную модель. Реализация такой модели в среде моделирования Engee приведена на рис. 5.

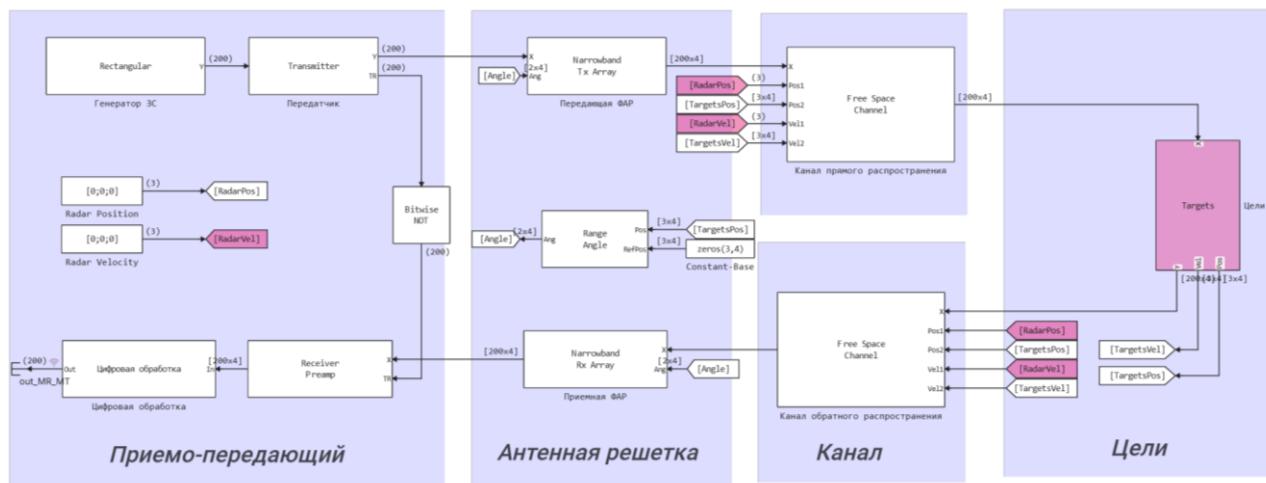
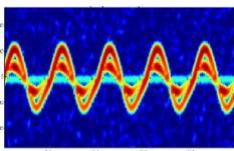
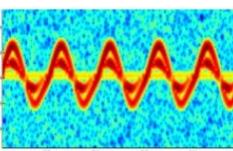
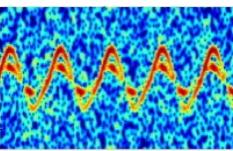
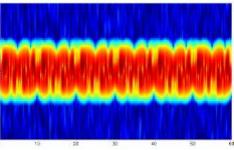
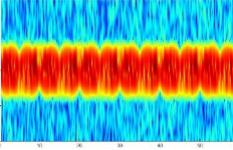
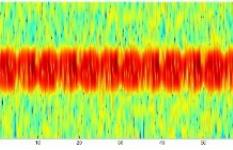


Рис. 5. Системная модель, реализованная в среде моделирования Engee

После реализации всех этапов моделирования требуется провести проверку работоспособности РЛС при решении задач обнаружения БПЛА и птиц. Модель имеет огромное количество параметров, которые можно изменять при расчете сценария. Наиболее важными среди них являются расстояния от РЛС до объекта, траектория движения цели, относительная скорость движения, скорости вращения винтов у БПЛА и взмаха крыльев птиц. В табл. 1 приведены МДС при варьировании дальности до объекта.

Принятие решения о классе объектов по МДС может быть выполнено как оператором, так и реализовано автоматически при помощи встраивания дополнительной обработки на основе сверточных нейронных сетей, способных при достаточной выборке обеспечить высокую точность классификации.

Табл. 1. МДС при варьировании дальности до объекта

Тип объекта	Дальность до объекта		
	200 м	400 м	800 м
Птица			
БПЛА типа квадрокоптер			

ЗАКЛЮЧЕНИЕ

Таким образом, использование метода анализа МДС позволяет классифицировать БПЛА и птиц с применением цифровой модели РЛС и радиолокационных целей. Эта модель имеет большое количество изменяемых параметров системы, что дает возможность исследовать поведение РЛС при различных сценариях движения целей, включая реальные траектории объектов. Такой подход позволяет генерировать синтетические данные отраженного сигнала от различных объектов для обучения искусственных нейронных сетей с целью увеличения точности результата.

СПИСОК ЛИТЕРАТУРЫ

1. Дмитриев В.И., Звонарев В.В., Лисицын Ю.Е. Методика обоснования рациональных способов управления беспилотным летательным аппаратом // Труды МАИ. 2020. № 112. 20 с. <https://doi.org/10.34759/trd-2020-112-16>
2. Карташов В.М. Методы обнаружения-распознавания радиолокационных, акустических, оптических и инфракрасных сигналов беспилотных летательных аппаратов // Радиотехника. 2021. № 2 (205). С. 138–153.
<https://doi.org/10.30837/RT.2021.2.205.15>
3. Park S., Shin S., Kim Y., Matson E.T. Combination of Radar and Audio Sensors for Identification of Rotor-type Unmanned Aerial Vehicles // (UAVs). 2015. Р. 1–4.
<https://doi.org/10.1109/ICSENS.2015.7370533>

4. Российская платформа математических вычислений и динамического моделирования Engee // Engee аккаунт пользователя.

URL: <https://start.engee.com> (дата обращения: 30.05.2025).

5. Richards M.A., Scheer J.A., Holm W.A. Principles of Modern Radar. Vol.1 Basic Principles, 2010. 960 p.

6. Chen V.C. The Micro-Doppler Effect in Radar, Artech House, 2011. 290 p.

SIMULATION OF RADAR OPERATION SCENARIOS FOR CLASSIFYING UNMANNED AERIAL VEHICLES AND BIRDS BASED ON MICRODOPPLER SIGNATURES IN THE ENGEE ENVIRONMENT

A. E. Semichastnov¹ [0009-0003-2301-6267], D. A. Balakin² [0000-0003-0497-7153]

^{1, 2}ETMC EXPONENTA LTD, Moscow, 117418, Russia

^{1, 2}Moscow Power Engineering Institute, Moscow, 111250, Russia

¹Semichastnov777@yandex.ru, ²dabalakin@yandex.ru;

Abstract

The paper considers a method for classifying unmanned aerial vehicles (UAVs) and birds based on radar measurements in various object movement scenarios. The relevance of the problem is explained by the complexity of UAV detection due to its small overall dimensions, high agility, and similar geometry to a bird. To solve the problem, a classification method based on the analysis of micro-Doppler signatures, which reflect the dynamics of object motion, is applied. The scenarios are modelled in the Engee environment, where models of radar system, UAV and birds are developed.

Keywords: *unmanned aerial vehicle, micro-Doppler signatures, radar system, backscatter diagram, digital processing.*

REFERENCES

1. Dmitriev V.I., Zvonarev V.V., Lisitsyn Yu.E. Methodology for substantiating rational methods of unmanned aerial vehicle control // Proceedings of MAI. 2020. No. 112. 20 p. <https://doi.org/10.34759/trd-2020-112-16>
-

2. *Kartashov V.M.* Methods for detecting and recognising radar, acoustic, optical and infrared signals from unmanned aerial vehicles // Radio Engineering. 2021. No. 2 (205). P. 138–153. <https://doi.org/10.30837/RT.2021.2.205.15>
 3. *Park S., Shin S., Kim Y., Matson E.T.* Combination of Radar and Audio Sensors for Identification of Rotor-type Unmanned Aerial Vehicles // (UAVs). 2015. P. 1–4. <https://doi.org/10.1109/ICSENS.2015.7370533>
 4. Russian dynamic modeling environment Engee // Engee user account. URL: <https://start.engee.com> (accessed: 30.05.2025).
 5. *Richards M.A., Scheer J.A., Holm W.A.* Principles of Modern Radar Vol. 1. Basic Principles, 2010. 960 p.
 6. *Chen V.C.* The Micro-Doppler Effect in Radar. Artech House, 2011. 290 p.
-

СВЕДЕНИЯ ОБ АВТОРАХ



СЕМИЧАСТНОВ Алексей Евгеньевич – инженер отдела радиотехнических систем ЦИТМ «Экспонента». Сфера научных интересов: цифровая обработка сигнала, моделирование радиолокационных систем.

Alexey Evgenievich SEMICHASTNOV – Engineer, Radio Engineering Systems Department, ETMC EXPONENTA LTD. Research interests: digital signal processing, modelling of radar systems.

email: Semichastnov777@yandex.ru

ORCID: 0009-0003-2301-6267



БАЛАКИН Дмитрий Александрович – к.т.н., руководитель отдела радиотехнических систем ЦИТМ «Экспонента», доцент кафедра ОРТ НИУ МЭИ. Сфера научных интересов: разработка и моделирование радиотехнических систем, исследование алгоритмов обработки сигналов.

Dmitry Alexandrovich BALAKIN – PhD, Head of the Radio Engineering Systems Department at ETMC EXPONENTA LTD, Associate Professor at the Department of Operational Research at the National Research University Moscow Power Engineering Institute. Research interests: development and modelling of radio engineering systems, research into signal processing algorithms.

email: dabalakin@yandex.ru

ORCID: 0000-0003-0497-7153

Материал поступил в редакцию 9 августа 2025 года

УДК 004.89

РАЗРАБОТКА АДАПТИВНОЙ СИСТЕМЫ ГЕНЕРАЦИИ ИГРОВЫХ КВЕСТОВ И ДИАЛОГОВ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

В. Т. Трофимчук¹ [0009-0001-9106-9614], В. В. Кугуракова² [0000-0002-1552-4910]

^{1, 2} Казанский федеральный университет, г. Казань, Россия

¹ vselord.betta@gmail.com, ² vlada.kugurakova@gmail.com

Аннотация

Рассмотрена проблема создания динамических нарративных систем для видеоигр с интерактивностью в реальном времени. Представлены разработка и тестирование компонента интеграции GPT для генерации диалогов, выявившие критическое ограничение облачных решений – задержку в 30 с., неприемлемую для игрового процесса. Предложена гибридная архитектура адаптивной системы, сочетающая LLM с механизмами обучения с подкреплением. Особое внимание уделяется решению проблем консистентности игрового мира и управлению долгосрочным контекстом взаимодействий с NPC через RAG-подход. Обоснован переход к парадигме Edge AI с применением методов квантования для достижения целевой задержки 200–500 мс. Разработаны метрики оценки персонализации и динамической адаптации контента.

Ключевые слова: видеоигры, большие языковые модели, генерация диалогов, диалогогенерация, генерация квестов, квестогенерация, адаптивные квесты, процедурная генерация контента, агентное поведение, игровой искусственный интеллект, машинное обучение в играх.

ВВЕДЕНИЕ

Современная игровая индустрия переживает значительную трансформацию, связанную с интеграцией технологий искусственного интеллекта (ИИ). Согласно данным аналитического ресурса Statista¹, в начале 2025 г. более трети разработчиков игр по всему миру заявили, что их студии уже используют ИИ в различных аспектах разработки. Это свидетельствует о стремительном проникновении нейросетей в геймдев, где они применяются для генерации текстур, написания диалогов, создания анимации и даже замены актеров озвучки.

Согласно обзору [1], в 2023–2025 г. наблюдается экспоненциальный рост исследований по интеграции больших языковых моделей (LLM)² в игровую индустрию, при этом 67% работ посвящены решению проблем адаптивности и персонализации игрового контента (см. табл. 1).

Табл. 1. Статистика применения LLM в играх

Направление применения	Доля исследований, %	Основные решаемые проблемы
Генерация диалогов NPC	42	Консистентность, персонализация
Динамические квесты	28	Баланс сложности, связность сюжета
Поведение NPC	18	Предсказуемость, адаптивность
Процедурный контент	12	Качество, разнообразие

¹ Statista – немецкая компания, специализирующаяся на рыночных и потребительских данных. URL: <https://www.statista.com/>

² LLM (Large Language Model, «большая языковая модель») — тип программы искусственного интеллекта, которая обучена работать с текстом: понимать, генерировать, пересказывать, дополнять, переводить и анализировать его.

Ключевой проблемой современных игр остается создание динамичных и реалистичных игровых миров, где неигровые персонажи (NPC³) демонстрируют адаптивное поведение и способны к осмысленному взаимодействию с игроком. Исследование [2] показало, что 99% геймеров считают, что продвинутый ИИ улучшит геймплей, при этом основные проблемы современных игр включают повторяющиеся диалоги (60%) и неспособность NPC адаптироваться к изменениям в игре.

Внедрение ИИ в игровую индустрию сопровождается рядом этических и юридических вызовов. Согласно отчету GDC⁴ 2024, 84% разработчиков выражают обеспокоенность этикой ИИ, включая копирование стилей без согласия авторов и снижение ценности человеческого труда.

Проблема авторских прав также остается острой. В 2023 г. компания Valve⁵ отклоняла публикацию игр в Steam⁶ из-за опасений относительно юридического статуса художественных элементов, созданных ИИ. В 2024 г. эта компания сформулировала политику относительно таких игр, предусматривающую проверку контента на предмет нарушений авторских прав.

Профессиональные союзы также активно реагируют на распространение ИИ. Американская гильдия артистов телевидения и радио (SAG-AFTRA⁷)

³ Non-Playable Character (NPC) — неигровой персонаж (неиграбельный персонаж) в играх, который не находится под контролем игрока.

⁴ GDC (Game Developers Conference) — ежегодная конференция для профессиональных разработчиков компьютерных игр.

⁵ Valve (Valve Corporation) (также известна как Valve Software) — американская частная компания, занимающаяся разработкой, изданием и цифровой дистрибуцией компьютерных игр.

⁶ Steam — онлайн-сервис цифрового распространения компьютерных игр и программ, разработанный и поддерживаемый компанией Valve.

⁷ SAG-AFTRA (сокр. от Screen Actors Guild-American Federation of Television and Radio Artists) — американский профсоюз, представляющий актеров, журналистов, радиоведущих, записывающих артистов, дублеров, публичных ораторов и других специалистов в области медиа и развлечений.

подала жалобу на создателей Fortnite⁸ за использование ИИ для озвучки персонажа без надлежащего уведомления. Данные примеры демонстрируют необходимость разработки четких правовых рамок для использования ИИ в игровой индустрии.

В настоящей статье представлены результаты разработки компонента интеграции технологии GPT⁹ для генерации диалогов и предложена архитектура расширенной адаптивной системы, сочетающей большие языковые модели с принципами обучения с подкреплением для создания динамичных игровых квестов и диалогов.

ЭВОЛЮЦИЯ ИИ В ИГРАХ: ОТ СКРИПТОВ К НЕЙРОСЕТЯМ

Исторически ИИ в играх ограничивался примитивными алгоритмами поведения «врагов», которые следовали заранее определенным скриптом и паттернам. В 1990-е годы произошел значительный прорыв — игры серии *Baldur's Gate* представили NPC, способных запоминать действия игрока и соответствующим образом реагировать на них в будущем. Это стало важным этапом в развитии неигровых персонажей, которые превратились из статичных элементов декораций в активных участников игрового мира.

Текстовые игры стали важным полигоном для испытания нарративных технологий. Еще в 1970-е годы такие игры, как *Colossal Cave Adventure* и *Zork*, заложили основы интерактивного повествования. Особенностью текстовых форматов была их глубокая нелинейность: проще было реализовать сложную ветвистую структуру с сотнями развлечений и множеством финалов, что до сих пор представляет сложность для современных AAA¹⁰-проектов.

⁸ Все компьютерные игры, упомянутые в настоящем исследовании, описаны в разд. Лудография.

⁹ GPT (Generative Pre-trained Transformer) — семейство больших языковых моделей (LLM), разработанных OpenAI.

¹⁰ AAA (triple-A) — игры высшего уровня производства с большим бюджетом разработки и маркетинга, создаваемые крупными студиями и издательствами (например, *The Last of Us*, *Red Dead Redemption*, *Cyberpunk 2077*).

Современные нейронные сети совершили качественный скачок — они не просто следуют инструкциям, а создают контент: от диалогов и квестов до текстур и озвучки.

Компании активно экспериментируют с различными подходами к интеграции ИИ, например:

- Ubisoft¹¹ использует Ghostwriter¹² для генерации реплик второстепенных персонажей и разрабатывает проект NEO NPC¹³, позволяющий игрокам разговаривать с персонажами голосом;
- Epic Games¹⁴ интегрировала ИИ для озвучки Дарта Вейдера в *Fortnite*, использовав нейросеть, обученную на голосе покойного Джеймса Эрла Джонса с согласия его семьи;
- разрабатываются специализированные инструменты для «оживления» виртуальных персонажей, такие как Inworld¹⁵ и Bitpart¹⁶, которые привлекают ученых и ветеранов игровой индустрии для решения задачи повышения последовательности и контекстной релевантности нейросетей.

Табл. 2 наглядно иллюстрирует трансформацию нарративных и контент-генерационных систем в индустрии, показывая, как ИИ развивался в играх, переходя от примитивных структур к адаптивным современным нейросетям.

¹¹ Ubisoft (ранее Ubi Soft Entertainment SA) — французская компания, специализирующаяся на разработке и издании компьютерных игр.

¹² Ubisoft Ghostwriter — инструмент на базе искусственного интеллекта, разработанный компанией Ubisoft для помощи сценаристам.

¹³ Ubisoft NEO NPC — инструмент для разработчиков, который позволяет создавать более правдоподобное общение с неигровыми персонажами (NPC).

¹⁴ Epic Games (ранее Epic MegaGames и Potomac Computer Systems) — американская компания, занимающаяся разработкой компьютерных игр и программного обеспечения.

¹⁵ Inworld — движок искусственного интеллекта для создания динамичных неигровых персонажей (NPC) и игровых миров.

¹⁶ Bitpart AI — интеллектуальная платформа для автоматизации создания контента. Предоставляет инструменты для легкой генерации текста и мультимедиа.

Табл. 2. Эволюция подходов к созданию игрового контента

Период	Технологии	Характер контента	Примеры
1970–1980-е	Текстовые парсеры	Статические деревья диалогов	<i>Colossal Cave Adventure, Zork</i>
1990–2000-е	Скриптовые системы	Ветвящиеся сценарии с ограниченной вариативностью	<i>Baldur's Gate, Fallout</i>
2010–2020-е	Процедурная генерация (CG)	Динамический контент по шаблонам	<i>Minecraft, The Elder Scrolls V: Skyrim</i>
2020-е и по настоящее время	Большие языковые модели (LLM)	Адаптивный контент, генерируемый в реальном времени	NEO NPC, <i>AI Dungeon</i>

СВЯЗАННЫЕ РАБОТЫ

Бурное развитие LLM в последние годы привело к активным исследованиям в области адаптивной генерации квестов и диалогов для цифровых игр. Актуальные публикации демонстрируют универсальность LLM при создании интерактивных сценариев и коммуникации персонажей, акцентируя внимание как на архитектурных новшествах, так и на оценке пользовательского опыта.

Обзорные статьи (например, [1]) освещают интеграцию LLM в генерацию сюжетов, диалогов NPC и адаптивных квестов, выделяя современные тренды и проблемы, связанные с созданием связного и увлекательного опыта для игрока. В работе [3] подчеркнуто, что новые модели генерируют более естественные и последовательные тексты, однако остаются открытыми вопросы по контролю качества и консистенции данных в коммерческих играх.

Внедрение LLM в игровые движки уже реализовано в ряде прикладных проектов. В статье [4] представлена архитектура генерации квестов и диалогов

на базе GPT-4o¹⁷ и Claude 3.7¹⁸, позволяющая динамично строить ветвящиеся сюжеты и подстраиваться под поведение игрока; также дан сравнительный анализ моделей и пользовательское тестирование. В исследовании [5] описан опыт интеграции LLM (Google Gemini API¹⁹, Sentence-BERT²⁰) для создания органичных диалогов и миссий для NPC с высокой степенью адаптивности.

Персонализированные сценарии успешно реализуются благодаря гибридным подходам. Например, в [6] LLM объединены с графами знаний для процедурной генерации индивидуализированных квестов и диалогов в RPG²¹, что обеспечивает связность нарратива и высокий уровень кастомизации.

В работах отечественных исследователей также уделяется значительное внимание генеративным системам для игр. Например, в [7] представлен подход к генеративной симуляции игрового окружения в реальном времени, что согласуется с целями данного исследования по достижению минимальной задержки. В [8] представлена методология создания корпуса текстов видеоигр на основе универсальной структуры, что может служить основой для обучения и валидации специализированных языковых моделей для игрового контента.

Отметим, что современные работы также уделяют внимание пользовательской оценке новых технологий. В [9] проанализировано качество диалогов, сгенерированных LLM, на примере реальных коммерческих продуктов,

¹⁷ GPT-4o – флагманская мультимодальная модель от OpenAI, выпущенная в мае 2024 г. Она основана на GPT-4, но дополнена поддержкой обработки сразу нескольких видов данных: текста, изображений, аудио и видео.

¹⁸ Claude 3.7 Sonnet — передовая гибридная модель искусственного интеллекта от компании Anthropic, выпущенная в феврале 2025 г.. Она сочетает два режима работы: быстрые, мгновенные ответы для простых запросов и расширенное пошаговое рассуждение для сложных задач, таких как программирование и аналитика.

¹⁹ Google Gemini API — канал взаимодействия с нейросетью Gemini от Google для разработчиков, занимающихся машинным обучением и созданием сервисов на базе больших языковых моделей.

²⁰ Sentence-BERT (SBERT) — модификация архитектуры BERT (Bidirectional Encoder Representations from Transformers), разработанная специально для генерации векторов фиксированной размерности, представляющих смысл целых предложений. Эти вектора (вложения предложений, эмбеддинги) позволяют сравнивать семантическую близость между предложениями.

²¹ RPG (от англ. Role-Playing Game) — ролевая игра, жанр видеоигр, в которых игрок управляет одним или несколькими персонажами.

в частности *Disco Elysium*, с глубоким разбором отзывов игроков и сравнением с ручным написанием. Аналогично в [10] выполнена структурная оценка квестов, созданных LLM, выявлены типичные ошибки моделей и предпочтения аудитории.

В работе [11] предложена архитектура, сочетающая LLM с графиками знаний, что позволяет достичь коэффициента персонализации до 0.87 против 0.62 у чистых LLM-решений [12].

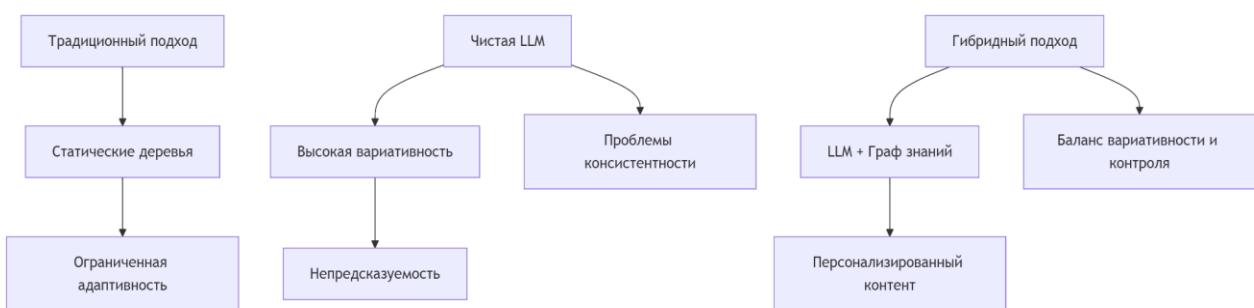


Рис. 1. Сравнение архитектур генерации квестов

На рис. 1 схематично представлено алгоритмическое различие в подходах (традиционный против чистого LLM-решения или гибридного подхода).

Однако необходимо отметить, что гибридный подход требует значительных вычислительных ресурсов для поддержания актуальности графа знаний в реальном времени.

Таким образом, современные исследования характеризуются инновационными архитектурными решениями, междисциплинарными методологиями и акцентом на тщательную практическую оценку LLM-генерации квестов и диалогов как важного инструмента для будущих игровых приложений.

РАЗРАБОТКА КОМПОНЕНТА ИНТЕГРАЦИИ GPT ДЛЯ ГЕНЕРАЦИИ ДИАЛОГОВ

Предлагаемый нами подход отличается от существующих решений тем, что стремится обеспечить реальную интерактивность и автономность в процессе генерации квестов и диалогов. Если большинство современных систем опирается на серверные LLM, требующие постоянного сетевого соединения и централизованной вычислительной инфраструктуры, то предлагаемый метод основан на модели Edge AI²² — исполнении нейросетевых процессов непосредственно на пользовательском устройстве [13].

Концепция нового подхода

Основная цель такого подхода — это минимизация сетевой задержки и достижение отклика в реальном времени, что критически важно для игровых приложений, где скорость реакции напрямую влияет на восприятие интерактивности.

Итак, перечислим целевые метрики:

- снижение задержки до диапазона 200–500 мс по сравнению с ~1800 мс для облачных решений;
- автономность, которая позволяет работать в офлайн-режиме без зависимости от стабильности интернет-соединения;
- защита персональных данных, так как вычисления происходят локально, без передачи пользовательских данных во внешние сервисы [14].

²² Edge AI (Artificial Intelligence at the Edge) — периферийный искусственный интеллект. Это парадигма вычислений, при которой задачи машинного обучения и инференса выполняются непосредственно на периферийных устройствах — смартфонах, камерах, датчиках и другом оборудовании. Вместо того чтобы полагаться на облачные вычисления, Edge AI обрабатывает данные локально, не отправляя их в облако.

Для реализации такой архитектуры требуются агрессивная оптимизация и адаптация LLM под устройства с ограниченными ресурсами. Здесь ключевыми направлениями выступают методы TinyML²³ и компрессии моделей (квантование²⁴, прунинг²⁵, дистилляция²⁶), позволяющие существенно снизить объем вычислений без потери качества текстовой генерации.

Аналогичные идеи использовались в промышленных сценариях для автономных датчиков и систем компьютерного зрения, что подтверждает эффективность принципов Edge AI в условиях ограниченного аппаратного обеспечения [14].

Таким образом, в отличие от предыдущих решений, завязанных на серверные ресурсы и облачные API, предложенная концепция рассматривает генерацию квестов и диалогов как распределенный когнитивный процесс, выполняемый локально, что открывает возможности для взаимодействия с игроками, осуществляемого действительно в реальном времени, и использования новых форм автономного геймплейного поведения NPC.

²³ TinyML — раздел машинного обучения, который позволяет запускать модели искусственного интеллекта на миниатюрных устройствах с низким энергопотреблением, таких как микроконтроллеры.

²⁴ Квантование модели — мощный метод оптимизации модели, который уменьшает объем памяти и вычислительные затраты нейронной сети (NN), преобразуя ее веса и активации из чисел с плавающей запятой высокой точности (например, 32-битное число с плавающей запятой или FP32) в типы данных с более низкой точностью, такие как 8-битные целые числа (INT8).

²⁵ Прунинг нейронных сетей — метод сжатия (уменьшения расхода памяти и вычислительной сложности) сети за счет устранения части параметров в предобученной модели.

²⁶ Дистилляция знаний (Knowledge Distillation) — метод оптимизации моделей и сжатия в машинном обучении (ML), при котором компактная «студенческая» модель обучается воспроизводить производительность более крупной и сложной «учительской» модели.

Архитектура решения

В рамках предварительной работы [12] был разработан компонент Integrator для интеграции технологии GPT в видеоигры, представляющий собой библиотеку на C#, упакованную в NuGet-пакет²⁷.

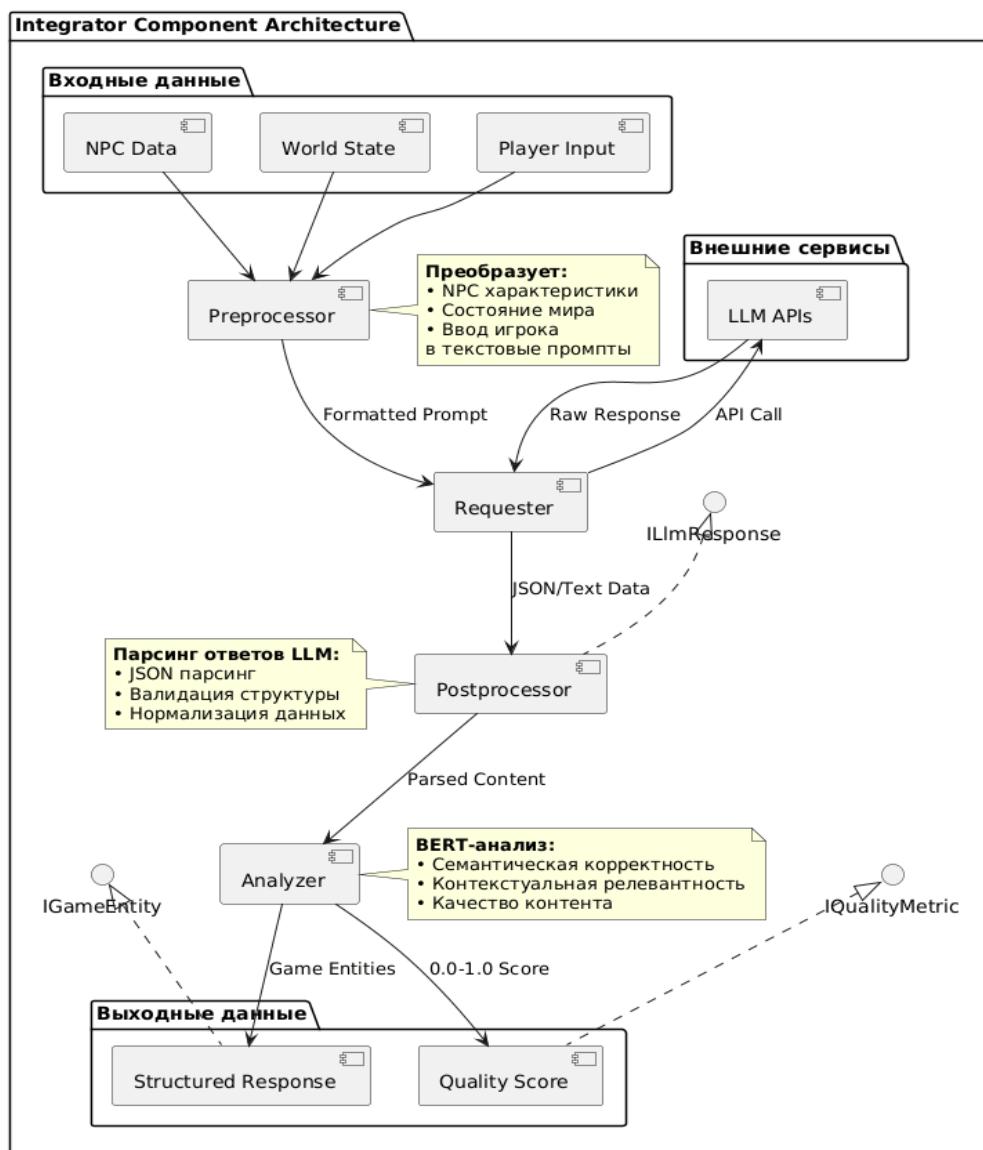


Рис. 2. Архитектура библиотеки

²⁷ NuGet-пакет – ZIP-файл с расширением .nupkg, содержащий скомпилированный код (DLL), другие связанные файлы и метаданные о пакете, такие как версия и описание. Он служит для управления зависимостями и повторного использования кода в .NET-проектах, позволяя разработчикам обмениваться библиотеками и инструментами через систему NuGet.

Архитектура компонента (см. рис. 2) включает четыре ключевых модуля:

- Preprocessor — преобразует сущности игрового мира (NPC, окружение) в текстовые промпты для LLM;
- Requester — осуществляет взаимодействие с API языковых моделей через RESTful-интерфейс;
- Postprocessor — выполняет парсинг ответов LLM и преобразует их в структурированные игровые сущности;
- Analyzer — нейросетевая модель на основе BERT, оценивающая корректность сгенерированного контента.

Необходимые интерфейсы на рис. 2 помечены как i<class>.

Входные данные NPC Data, World State, Player Input составляют Narrative Environment — параметры окружающей среды, влияющие на формирование запроса, и будут детально описаны ниже.

Выходные данные состоят из:

- Structure Response — структурированный ответ, например ветвящийся диалог, полученный в результате диалогогенерации²⁸ (рис. 3).
- Quality Score — оценка синтаксической корректности и семантической согласованности.

²⁸ Диалогогенерация — это автоматическое создание диалогов и разговоров между персонажами с помощью технологий искусственного интеллекта, нейронных сетей и обработки естественного языка.

```
Entry text: Приветствую, странник! Я - сэр Альфред, рыцарь этих земель. Чем могу быть вам полезен?  
Entry child count: 2  
  
Dialogue varient: 1  
Player name: Игрок  
Player text: Здравствуйте, сэр Альфред. Я путешественник, и мне не помешает помочь опытного воина.  
NPC name: Альфред  
NPC text: Рад, что вы обратились ко мне. Как я могу помочь вам на вашем пути?  
Dialogue node childs count: 2  
  
Dialogue varient: 1.1  
Player name: Игрок  
Player text: Я бы хотел присоединиться к вашему отряду и помочь в борьбе с этими угрозами.  
NPC name: Альфред  
NPC text: Это благородное предложение, молодой человек. Я буду рад принять вас в наши ряды. Ваша помощь нам очень пригодится.  
Dialogue node childs count: 0  
  
Dialogue varient: 1.2  
Player name: Игрок  
Player text: Понимаю. Я постараюсь держаться подальше от этих мест, пока ситуация не станет более безопасной.  
NPC name: Альфред  
NPC text: Я понимаю ваши опасения. Однако, если вы решите остаться, я гарантирую вашу безопасность под моим присмотром.  
Dialogue node childs count: 0
```

Рис. 3. Пример сгенерированного ветвистого диалога, выведенного в лог

Представленная архитектура предполагает минимальную задержку (кроме задержки самого облачного GPT API) и полную автономность за счет внедрения библиотеки прямо в код игрового движка.

Реализация и тестирование

Компонент был успешно интегрирован в демонстрационный проект на Unity (см. рис. 4), где была проиллюстрирована возможность генерации ветвящихся диалогов с учетом характеристик NPC (профессия, личностные черты, социальные связи).



Рис. 4. Демонстрационный проект использования библиотеки

Табл. 3. Синтаксическая корректность (грамматика)

Сценарий	Сгенерирован- ный текст	Оценка (вероятность)	Комментарий
Корректно	<i>Я видел его вчера, когда он шел по мосту.</i>	$P(\text{text}) \approx 0.9$ (высокая)	Грамматически правильное предложение
Некор- ректно	<i>Я видел его вчера он мосту шел по.</i>	$P(\text{text}) \approx 0.07$ (низкая)	Нарушение порядка слов и пунктуации. LLM присваивает низкую вероятность такой последовательности токенов, что позволяет от- фильтровать этот результат
Корректно	<i>Ты возьмешь этот меч?</i>	$P(\text{text}) \approx 0.88$ (высокая)	Корректная форма глагола и вопросительное предло- жение
Некор- ректно	<i>Ты этот меч брать?</i>	$P(\text{text}) \approx 0.15$ (низкая)	Ошибка согласования/ спряжения глагола

Для обеспечения корректности работы была реализована система валидации (детально описана в [12]), включающая:

- проверку синтаксической корректности P сгенерированного текста (см. табл. 3);
- оценку семантической согласованности с контекстом игры (см. табл. 4);
- контроль соответствия заданным параметрам NPC.

Табл. 4. Семантическая согласованность (Контекст игры)

Сценарий	Сгенерированный текст (ответ Кузнеца)	Оценка (Вероятность)	Комментарий
Корректно (согласованно)	<i>Отлично, это золото – то, что нужно для клинка! Благодарю тебя, эльф.</i>	$P(text)$ (золото) ≈ 0.92 (высокая)	Золото то, что нужно кузнецу
Некорректно (несогласованно)	<i>Ты принес мне ржавые сапоги? Это не поможет мне выковать серебряный клинок.</i>	$P(text)$ (ржавые сапоги) ≈ 0.13 (низкая)	Ржавые сапоги не то, что нужно кузнецу

Результаты этой валидации и являются Quality Score, определяющие качество сгенерированного контента.

Под семантической согласованностью в дальнейшем будем понимать согласованность сгенерированного контента с контекстом игры (например, для NPC прописана история, что он является обычным деревенским кузнецом, а в контексте истории мира говорится, что мир находится в эпохе раннего средневековья, но не указано, что мир является магическим фэнтези, следовательно, магии, которая может имитировать или обходить современные технологические процессы, нет. Если такому кузнецу принести ржавые предметы и попросить переплавить, то он откажется, так как не сможет обработать данные предметы).

Для оценки семантической согласованности часто используют внешние механизмы валидации (например, *Knowledge Graph*²⁹ или *Правила игры*³⁰) в дополнение к внутренней вероятностной оценке LLM. Это гарантирует, что даже синтаксически корректный, но контекстно-неправильный ответ (как в примере с «ржавыми сапогами») будет отброшен.

Проблемы интеграции

При практической интеграции возникли сложности с управлением зависимостями NuGet-пакетов внутри Unity, особенно для библиотек с нативными компонентами (например, ONNX Runtime³¹ для оценочной нейросети BERT). Это создало значительный риск нестабильности и усложнило процесс сборки.

Анализ производительности

Тестирование с использованием облачного API GPT-3.5-Turbo³² выявило критическую проблему с временем отклика (*latency*).

Для сценария генерации одного полного диалогового хода (фраза NPC и три варианта ответа для Игрока) были получены данные (табл. 5), позволившие выявить зависимости времени генерации.

Общее время задержки T_{total} , исходя из полученных данных, может быть выражено как

$$T_{\text{total}} = T_{\text{network}} + T_{\text{pre}} + T_{\text{API}} + T_{\text{post}} + T,$$

где T_{network} — время выполнения отправки запроса нейросети, T_{pre} — время составления промта, T_{API} — время выполнения запроса нейросетью,

²⁹ Knowledge Graph (граф знаний) — семантическая сеть, в которой хранится информация о разных сущностях и взаимосвязях между ними.

³⁰ «Правила игры» (в контексте семантического анализа) — правила и соглашения, которые определяют, какие высказывания считаются правильными или адекватными в контексте. Это понятие связано с концепцией «языковых игр» Людвига Витгенштейна, которая объясняет как язык функционирует в рамках социальных контекстов, которые рассматриваются как игры.

³¹ ONNX (Open Neural Network Exchange) — открытая библиотека программного обеспечения для построения нейронных сетей глубокого обучения.

³² GPT-3.5-Turbo — это улучшенная версия языковой модели GPT-3.5, выпущенная 1 марта 2023 г. Модель оптимизирована для разговорного чата и может имитировать человеческие ответы.

T_{post} — время парсинга ответа, пришедшего от нейросети, T — время выполнения синтаксического и семантического анализа.

Табл. 5. Параметры генерации

Параметр генерации	Значение	Описание параметра
Количество генерируемых вариантов (N_{variants})	3	Варианты ответа NPC
Длина генерируемого варианта (L_{variants})	3	Длина ответа NPC (в предложениях)
Среднее время генерации	30 с	Время от отправки запроса до получения итогового ответа

Конфигурация устройства, на котором проводились вычисления:

- CPU Intel Core i5 10300H;
- ОЗУ 32 ГБ памяти;
- GPU GeForce GTX 1650 от Nvidia.

Было установлено, что время выполнения запроса нейросетью T_{API} является доминирующим фактором и прямо пропорционально количеству вариантов и их длине:

$$T_{\text{API}} \sim N_{\text{variants}} \cdot L_{\text{variants}},$$

где N_{variants} — количество генерируемых вариантов, а L_{variants} — длина генерируемого варианта.

Задержка в 30 с является неприемлемой для интерактивного игрового процесса в реальном времени, для которого целевое время отклика должно составлять менее 1 с [15]. Последнее значение выбрано как следствие следующих критериев.

1. Разрушение иммерсивности и «магии» игры. Игры, особенно с глубоким повествованием, стремятся погрузить игрока в свой мир.

- При диалогогенерации, если Игрок задает вопрос NPC, но вместо мгновенного, живого ответа получает 30 с молчания, то пропадает эффект иммерсивности – впрочем, возможны решения, позволяющие скрыть эту паузу, например анимация idle³³-действий, например, почесывание затылка или болтание руками, как будто NPC раздумывает, но эти «заплатки» будут явно вычисляться опытным Игроком.
 - При квестогенерации³⁴ действие Игрока должно запустить новую цепочку событий, и снова ожидание в 30 с для выбора задания полностью разрушает *нarrативный импульс*³⁵ и *драматический момент*³⁶. Напряжение рассеивается, а интерес сменяется раздражением.
2. Нарушение игрового потока (Flow State)³⁷. Игровой процесс строится на игровом цикле «действие → реакция → новое действие». Мозг игрока находится в *состоянии потока*, где он быстро реагирует на изменения в мире. Задержка в 30 с — это обрыв данного цикла. Игрок совершает действие, а обратная связь отсутствует. За это время Игрок потеряет фо-

³³ Idle (от англ. «бездействие», «простой») — состояние персонажа в игре, когда игрок не предпринимает никаких действий; idle-анимации — циклические движения персонажа в этом состоянии, придающие ему живость и естественность.

³⁴ Квестогенерация — процедурная генерация квестов, заданий и миссий в компьютерных играх с помощью заранее прописанных алгоритмов.

³⁵ Нарративный импульс — сочетание истории игры (сюжета, персонажей, мира) и игровых механик, которые работают вместе для создания глубокого погружения и эмоционального вовлечения игрока.

³⁶ Драматический момент — ключевая, напряженная сцена внутри самой игры, которая обладает сильной эмоциональной нагрузкой и зачастую является переломным пунктом в сюжете или игровом процессе.

³⁷ Игровой поток (Flow State) — это состояние полного погружения и сосредоточенности на игре, когда игрок теряет ощущение времени и чувствует глубокое удовлетворение от процесса.

кус и стратегический настрой. Вернуться в состояние потока после такого разрыва очень сложно. Игрок чувствует себя не активным участником, а пассивным наблюдателем, ожидающим загрузки.

3. Тактическое и стратегическое ожидание. В реальном времени каждая секунда на счету.

- В диалоге: Игрок может ожидать ответа, чтобы принять решение: атаковать, торговаться, использовать убеждение. 30 с — это время, которое делает любое тактическое планирование бесмысленным.
- В квестах: если квест генерируется на основе действий игрока (например, «враги захватили деревню, потому что вы ушли»), задержка в полминуты означает, что игровая ситуация может кардинально измениться еще до того, как квест будет получен. Это создает противоречивую и нелогичную ситуацию.

4. Психологический порог терпения и восприятие. Исследования UX-партнеров в IT-продуктах [15] выявили универсальные принципы, актуальные и для игровой индустрии:

- 0.1 с: ощущается как мгновенная реакция;
- 1.0 с: естественная, непрерывающая задержка; поток мысли пользователя не прерывается;
- 10 с: максимальный предел для удержания внимания на задаче; пользователь начинает терять концентрацию;
- 30 с и более: это воспринимается не как задержка, а как простой системы или сбой. Игрок с большой вероятностью решит, что игра «зависла», и попытается перезапустить ее или вовсе ее закроет.

Почему целевой показатель — менее 1 с? Именно показатель (от 100 мс до 1 с) соответствует требованиям для поддержания интерактивности в реальном времени, таким как:

- *непосредственность* — реакция системы происходит почти одновременно с действием игрока;
 - *сохранение потока* — мозг не успевает переключиться на другую задачу;
-

- *естественность* – такая задержка сопоставима с реакцией живого собеседника или откликом интерфейса на нажатие кнопки.

Технические и практические следствия для разработки

Понимание временного ограничения (*100 мс на генерацию шага*) требует не прямолинейного, а гибридного и оптимизированного применения LLM, таких как

- кэширование и предварительная генерация – заранее (*но все же во время игрового процесса*) генерируются возможные реплики и варианты квестов, которые быстро подставляются в нужный момент;
- локальные и оптимизированные модели – использование меньших, более быстрых моделей, которые работают локально (или ближе к игровому серверу), чтобы минимизировать сетевые задержки;
- асинхронная генерация – генерация контента в фоне, пока игрок занят другими делами, а не в критический момент диалога;
- строгие ограничения по длине (*tokens*) – генерация очень коротких ответов, которые выполняются за доли секунды;
- offline-генерации (вместо online-генерации) – когда LLM используется для предварительного создания контента (*offline*), который затем используется игровым движком в режиме реального времени вместо синхронной генерации (*online*) во время игровой сессии.

Перечисленные рассуждения послужили основным аргументом для пересмотра архитектуры.

Переход к удаленному сервису (Server-Side AI)

Для обхода проблем с интеграцией зависимостей и изоляции игрового клиента от ресурсоемких процессов система была переведена в формат удаленного сервиса (Microservice Architecture).

Отметим преимущества серверной архитектуры.

- *Изоляция зависимостей*

Игровой клиент (Unity) взаимодействует только по HTTP/HTTPS, полностью изолируясь от сложных C# библиотек и нативных зависимостей.

- *Централизация вычислений*

Сервер, как единая точка обработки, выполняет все ресурсоемкие этапы (Preprocessor, Requester, Postprocessor, Analyzer).

- *Кросс-платформенность*

Любой игровой клиент, поддерживающий HTTP-запросы, может использовать сервис.

Архитектура удаленного сервиса

На UML-диаграмме развертывания (рис. 6) представлена детальная архитектура удаленного сервера, взаимодействующего с микросервисами.

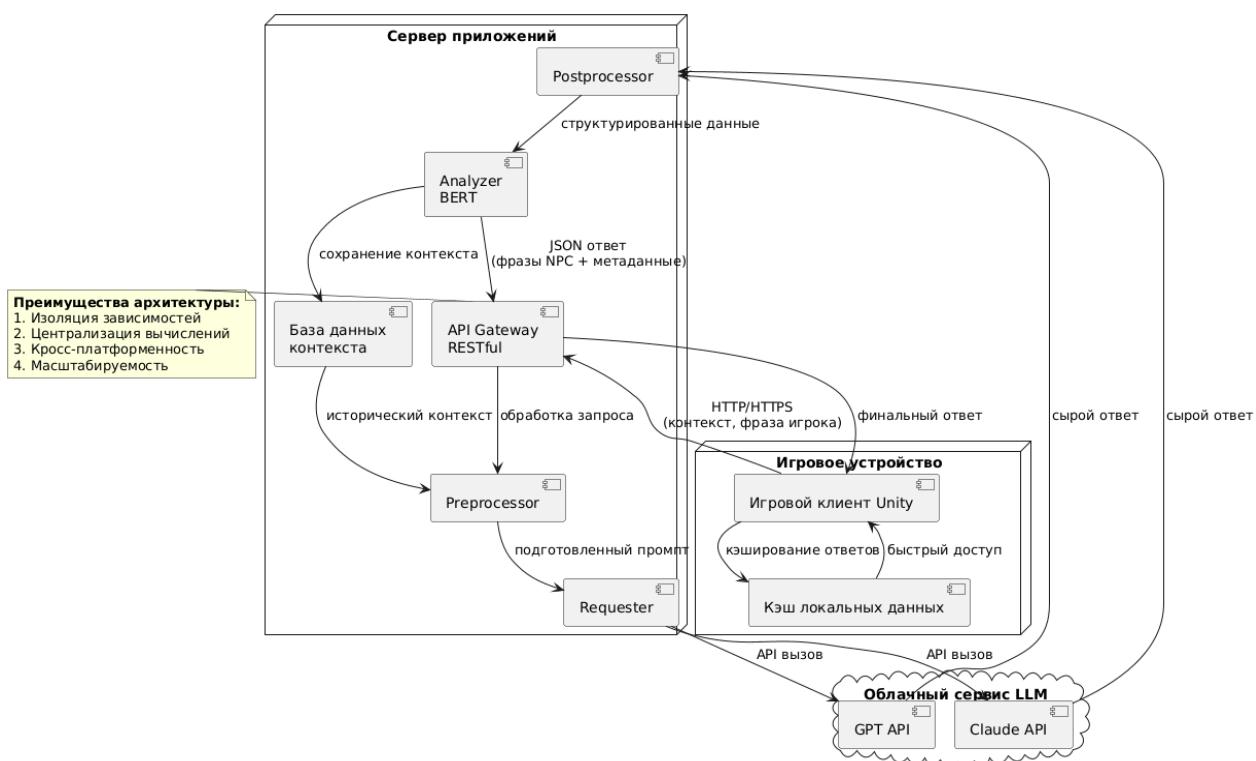


Рис. 6. Развёртывание удаленного сервера с микросервисами

Игровой клиент формирует минимальный запрос, включающий контекстные данные (состояние мира, ID квеста, фраза игрока). Сервер принимает, обрабатывает запрос, вызывает облачный API LLM, анализирует ответ и возвращает клиенту финальный структурированный объект (например, JSON с готовыми фразами NPC и метаданными).

Несмотря на решение проблем интеграции, серверная архитектура по-прежнему страдает от высокой задержки, вызванной облачным API LLM и необходимостью сетевого обмена с удаленным сервером. По приблизительным оценкам, даже с оптимизированным сервером задержка остается на прежнем уровне, что говорит о том, что нужно изменять подход к генерации данных, и требует дальнейших шагов для достижения целей в реальном времени.

ЛОКАЛЬНАЯ МОДЕЛЬ (EDGE AI)

В текущей стадии проекта использование облачного API GPT-3.5-Turbo для генерации диалогового хода дало среднюю задержку в 30 с. Это критическое препятствие: для поддержания погружения (иммерсивности) и сохранения состояния потока (Flow State) игрока время отклика системы на интерактивные действия должно быть менее одной с. Задержка, превышающая 1–2 с, психологически воспринимается пользователем не как ожидание ответа, а как системный сбой, что полностью разрушает интерактивный опыт, особенно в диалоговых системах.

Именно поэтому центральной задачей дальнейшего развития является достижение интерактивности в реальном времени. Это достижимо исключительно через радикальное устранение сетевой задержки и переход к парадигме Edge AI — исполнению LLM непосредственно на вычислительном устройстве пользователя.

Обоснование необходимости Edge AI и целевые показатели

Переход на локальную архитектуру Edge AI является не просто оптимизацией, а сменой фундаментального подхода, позволяющей решить три критические проблемы облачных решений:

1. Проблема задержки

Облачное решение всегда подвержено вариативности сетевого трафика, времени обработки на сервере и скорости токенизации. Локальное исполнение исключает эти переменные. Наша цель — 200–500 мс на генерацию, что вплотную приближается к порогу человеческого восприятия «мгновенного» ответа. Достижение этого показателя требует не только быстрой модели, но и тесной интеграции с аппаратными ускорителями (GPU/NPU).

2. Проблема автономности и стабильности

Локальная модель гарантирует работу в офлайн-режиме, что критически важно для портативных устройств или пользователей с нестабильным интернет-соединением. Система становится независимой от внешних сервисов, их тарифов и потенциальных сбоев.

3. Проблема приватности данных

Вычисления, происходящие на устройстве пользователя, полностью исключают передачу чувствительных данных (контекст игры, история диалогов) на внешние серверы, обеспечивая высокий уровень конфиденциальности.

Стратегия оптимизации на основе методологии TinyML

Для внедрения LLM в игровое окружение с ограниченными ресурсами (VRAM среднего игрового ПК ≤ 8 GB) необходима агрессивная стратегия микроптимизации, вдохновленная принципами TinyML.

1. Выбор и адаптация базовой модели

Выбор начнется с анализа Open-Source LLM (например, Llama 3 8B³⁸, Mistral 7B³⁹), которые показали высокую производительность при относительно малом размере.

- Критерий качества — модель должна обладать достаточной «глубиной

³⁸ Meta-Llama-3-8B — языковая модель с 8 млрд параметров, разработанная компанией Meta в рамках семейства моделей Meta Llama 3.

³⁹ Mistral 7B — LLM с 7.3 млрд параметров от французского стартапа Mistral AI.

знаний» для поддержания согласованного ролевого поведения NPC и генерации логичных квестов.

- *Критерий ресурсов* – модель должна быть способна работать в режиме Inference⁴⁰ на пользовательском оборудовании (например, на GPU с объемом VRAM 8 ГБ или даже на мощном CPU в режиме *CPU Offload*⁴¹).

2. Технологии компрессии

Ключевым шагом было применение квантования. Модели с миллиардами параметров в формате FP32 занимают десятки гигабайт.

- *4-битное квантование (INT4)*

Это наиболее радикальный и необходимый метод. Он позволяет сократить требования к памяти на 75%. При этом может быть проведен тщательный анализ потерь качества (NLG Evaluation⁴²) для гарантирования сохранения «голоса» NPC и нарративной логики после компрессии. Мы использовали техники GPTQ⁴³ или QLoRA⁴⁴ для пост-тренировочного квантования (PTQ).

⁴⁰ Inference — рабочий режим эксплуатации обученной нейросети, при котором она выполняет свою основную функцию (генерацию диалогов и квестов) на основе входных данных в реальном времени.

⁴¹ CPU Offload — стратегия работы с большими нейросетевыми моделями, которая позволяет запускать их на GPU с ограниченной памятью за счет хранения части данных в оперативной памяти CPU, что спасает от ошибок нехватки памяти, но значительно снижает скорость выполнения.

⁴² NLG Evaluation — совокупность методов (как автоматических, так и основанных на оценке человеком) для измерения качества текста, созданного нейросетями, что особенно важно для таких задач, как генерация игровых диалогов и квестов, где принципиальны связность, увлекательность и соответствие игровому миру.

⁴³ GPTQ — метод послеобученного квантования, который позволяет сжать большие языковые модели до 3–4 бит на параметр с минимальной потерей качества.

⁴⁴ QLoRA — метод эффективной тонкой настройки, который позволяет адаптировать большие модели на одном GPU с ограниченной памятью.

- *Прунинг и дистилляция*

В качестве дополнительного шага были рассмотрены методы структурного прунинга и дистилляции для создания специализированной, легковесной модели, настроенной исключительно на задачи генерации диалогов и квестов.

3. Технологический стек для игровой интеграции

Для достижения максимальной скорости Inference и кросс-платформенной совместимости необходимо использовать специализированные движки вывода нейросетей (*Inference Engines*) и низкоуровневые API:

- *Inference Engine*

Модель была конвертирована в формат, оптимизированный для Inference (например, GGUF⁴⁵ или ONNX). Использование библиотек-оберток (например, C#-обертки над llama.cpp) позволило нам получить доступ к высокопроизводительному коду на C++, который использует SIMD-инструкции⁴⁶ и CPU Offload для эффективной работы даже при отсутствии мощного GPU.

- *Аппаратное ускорение (Hardware Acceleration)*

Внедрение фреймворков (таких как DirectML⁴⁷ на Windows или Metal Performance Shaders⁴⁸ на macOS/iOS) позволило задейство-

⁴⁵ GGUF (GPT-Generated Unified Format) — формат файла для хранения моделей машинного обучения, специально разработанный для быстрой загрузки и выполнения на центральных процессорах (CPU).

⁴⁶ SIMD-инструкции — «одна инструкция, множество данных» (single instruction, multiple data), что означает возможность выполнять одну и ту же операцию параллельно над несколькими элементами данных с помощью одной команды.

⁴⁷ DirectML — это низкоуровневый API для машинного обучения (МО) от Microsoft, который использует аппаратное ускорение на графических процессорах (GPU) для выполнения задач МО.

⁴⁸ Metal Performance Shaders (MPS) — набор высокооптимизированных шейдеров, предоставляемых Apple в виде фреймворка для приложений, созданных с использованием Metal API.

вать весь потенциал GPU для параллельных вычислений, что критически важно для снижения задержки с 500 мс до целевых 200 мс.

- *Асинхронный стриминг*

Для снижения воспринимаемой задержки был реализован механизм Token Streaming⁴⁹, при котором текст начинает отображаться на экране сразу после генерации первого токена, а не после полного формирования ответа моделью. Это создает иллюзию мгновенного отклика, даже если полная генерация занимает до полус.

АРХИТЕКТУРА АДАПТИВНОЙ СИСТЕМЫ ГЕНЕРАЦИИ КВЕСТОВ И ДИАЛОГОВ

На основе разработанного компонента Edge AI нами предложена гибридная, многоуровневая архитектура (см. рис. 7 и 8) для адаптивной генерации игрового контента.

⁴⁹ Token Streaming (потоковая передача токенов) — технология генерации и отображения текста языковой моделью, при которой токены (минимальные единицы текста) передаются и отображаются последовательно по мере их генерации, а не после завершения формирования полного ответа.

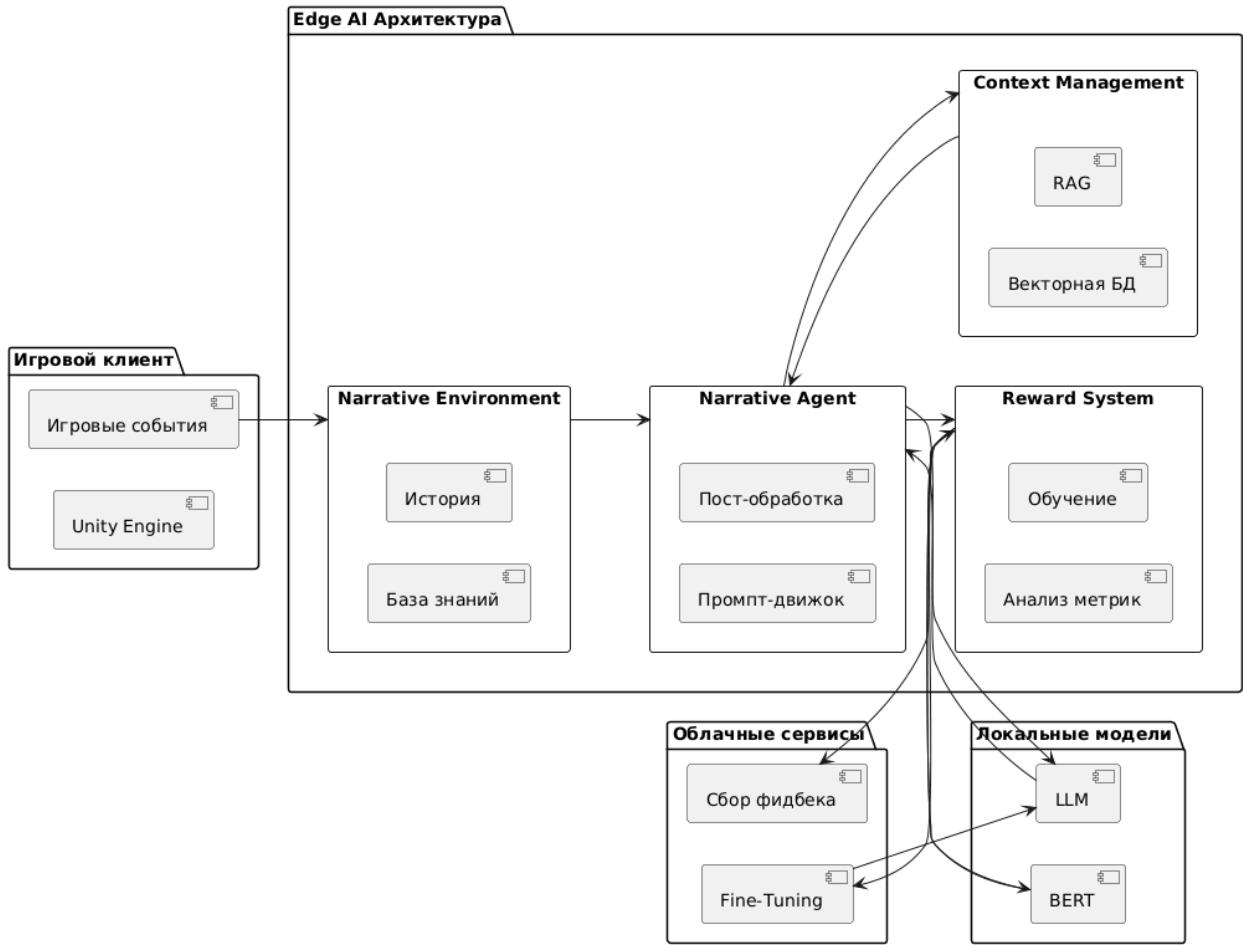


Рис. 7. Упрощенная схема усовершенствованной архитектуры

Система, представленная на рис. 7, уходит от простого промптинга и сочетает мощь LLM с традиционными подходами управления состоянием мира и элементами обучения с подкреплением (Reinforcement Learning, RL).

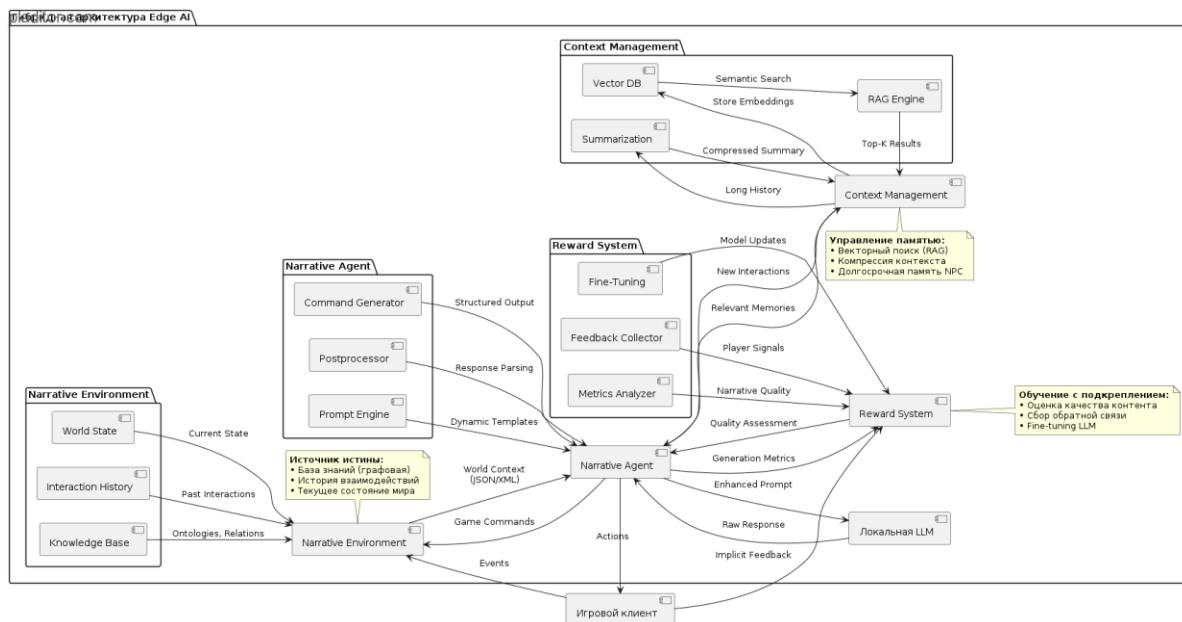


Рис. 8. Подробная схема усовершенствованной архитектуры

Архитектура на рис. 8 состоит из четырех ключевых (Narrative Environment, Narrative Agent, Context Management, Reward System for Narrative), тесно интегрированных модулей, каждый из которых выполняет свою роль в цикле адаптивной генерации.

Далее рассмотрим каждый из четырех модулей.

Модуль управления нарративной средой (Narrative Environment)

Этот модуль является источником истины для всей системы. Его основная задача — хранить, обрабатывать и предоставлять актуальное, консистентное состояние игрового мира.

- **База знаний (Knowledge Base)**

Хранит структурированные данные о мире — онтологии (иерархия существ, предметов), карты отношений (дружба, вражда между NPC и фракциями) и ключевые нарративные точки. Используется графовая база данных (например, Neo4j или реляционная база с графовой структурой) для быстрого извлечения контекста. Для формализации пространственно-временных отношений в игровом мире и структурирования данных о состояниях может быть применен математический аппарат [16], позволяющий создать

строгую формальную модель эволюции игровых систем во времени и пространстве. Это особенно важно для отслеживания положения NPC, изменения состояний локаций и временных зависимостей между игровыми событиями.

- *История взаимодействий* (Interaction History)

Ведет журнал всех действий игрока и сгенерированных диалогов/квестов. Этот журнал критически важен для долгосрочной памяти и предотвращения повторения контента.

- *Формат вывода*

Модуль преобразует сложное состояние мира в структурированный, лаконичный формат (например, JSON или XML) для передачи в Narrative Agent, минимизируя «шум» промпта для LLM.

Нarrативный агент (Narrative Agent)

Это интеллектуальный посредник между состоянием мира и LLM. Его задача — превратить игровые события в эффективные инструкции для генеративной модели и интерпретировать ее «сырой» вывод обратно в игровую логику.

- *Контекстуализация и промптинг.*

Агент принимает данные о состоянии мира от Narrative Environment и, используя динамические шаблоны промптов, формирует запрос к локальной LLM. Этот промпт включает:

- «ролевую» характеристику NPC (характер, цель, знания);
- краткий исторический контекст (из Системы Памяти);
- текущее событие или запрос игрока.

- *Пост-обработка и парсинг.*

После получения ответа от LLM агент использует Postprocessor для парсинга текста и преобразования его в формализованные игровые команды (например, QUEST_START {id: "..."}, DIALOGUE_LINE {text: "..."}, NPC_ACTION {move: "..."}). Это обеспечивает безопасную и предсказуемую интеграцию генеративного контента.

Система управления контекстом и памятью (Context Management)

Этот модуль решает фундаментальную проблему всех LLM — ограниченное окно контекста и потерю долгосрочной памяти.

- *Векторная база данных* (Vector Database)

Используется для хранения всех прошлых диалогов и ключевых событий в виде векторных эмбеддингов.

- *Retrieval-Augmented Generation* (RAG⁵⁰)

При получении нового запроса от игрока система использует RAG-подход: ищет в векторной базе данных наиболее релевантные «воспоминания» (например, 2–3 ключевых диалога) и включает их в промпт для LLM. Это позволяет LLM генерировать ответы, которые согласуются с прошлыми взаимодействиями.

- *Модуль суммаризации*

При переполнении окна контекста или для создания «резюме» долгой истории используется отдельная, более легкая модель для автоматической компрессии старых воспоминаний в краткое, высокоуровневое описание (например, «Игрок выполнил квест по доставке артефакта 5 дней назад»).

Модуль позволяет NPC «помнить» события, произошедшие часы назад.

Система оценки и наград (Reward System for Narrative)

Данный модуль обеспечивает обучение системы и ее адаптацию, используя принципы Reinforcement Learning from Human Feedback (RLHF) или Preference Optimization.

- *Нarrативные метрики*

Включает в себя автоматический Analyzer (возможно, на базе BERT или другой классификационной модели), который оценивает:

⁵⁰ Retrieval-Augmented Generation (RAG) — это архитектурный подход в области искусственного интеллекта, который объединяет большие языковые модели (LLM) с внешней базой знаний для повышения точности и актуальности генерируемых ответов.

- 1) естественность и ролевую консистентность диалога;
 - 2) сбалансированность сгенерированного квеста (сложность, награда);
 - 3) нарративную согласованность (отсутствие противоречий с Narrative Environment).
- *Обратная связь (Feedback Loop)*

Система собирает неявные (время, проведенное в диалоге, выбор действий игрока) и явные (опросы игрока) данные для формирования функции награды (Reward Function). Эта награда используется для тонкой настройки (Fine-Tuning) локальной LLM, улучшая ее способность генерировать контент, который максимально соответствует ожиданиям и стилю игрока, делая систему по-настоящему адаптивной.

МЕТРИКИ ОЦЕНКИ ПЕРСОНАЛИЗИРОВАННОГО КОНТЕНТА

Для оценки качества персонализации генерируемого контента необходим ряд комплексных метрик. В отличие от традиционных метрик, ориентированных исключительно на технические параметры, предлагаемые нами метрики позволяют количественно оценивать степень адаптации контента под индивидуальные особенности игрока.

Метрика персонализации контента для оценки адаптивности игровых систем

Для количественной оценки эффективности персонализации игрового контента в разработанной системе предложена комплексная метрика, построенная по аналогии с подходами, описанными в исследованиях по адаптивным системам и рекомендательным алгоритмам. В работе [6] подчеркнута важность измерения степени соответствия генерируемого контента индивидуальным характеристикам пользователя, однако конкретная математическая формализация не приведена. Восполняет этот пробел предлагаемая метрика персонализации (от 0 до 1)

$$P = \sum_i (R_i \cdot W_i) / N$$

где R_i – релевантность контента i -му параметру пользователя, W_i – вес параметра, N – количество учитываемых параметров.

Интерпретация значений метрики персонализации осуществляется следующим образом:

- $P < 0.3$ – низкий уровень персонализации, контент практически не адаптируется под пользователя;
- $0.3 \leq P < 0.7$ – средний уровень персонализации, учитываются основные предпочтения пользователя;
- $P \geq 0.7$ – высокий уровень персонализации, контент значительно адаптирован под индивидуальные особенности игрока.

Представленная метрика позволяет оценить способность системы адаптировать генерируемые квесты и диалоги под индивидуальные особенности игрока, включая его игровой стиль, принятые ранее решения, предпочтения в прохождении и другие релевантные характеристики.

Метрика динамической адаптации для оценки соответствия контента игровому прогрессу

Для оценки способности системы адаптировать генерируемый контент в соответствии с прогрессом игрока и изменяющимся игровым контекстом предлагаем метрику динамической адаптации

$$D = 1 - \frac{|C_a - C_e|}{C_{\max}},$$

где C_a – фактическая сложность генерируемого контента, C_e – ожидаемая сложность на основе текущего прогресса игрока, C_{\max} – максимально возможная сложность в системе.

Метрика построена на основе анализа подходов, представленных в работе [17], где отмечена важность соответствия сложности контента текущим возможностям игрока.

Интерпретация значений коэффициента динамической адаптации:

- $D < 0.6$ – недостаточная адаптация, дисбаланс сложности;
- $0.6 \leq D < 0.8$ – удовлетворительная адаптация;
- $D \geq 0.8$ – высокая степень адаптации сложности контента.

Дополнительные метрики для комплексной оценки

Можно рассмотреть и другие специализированные метрики, такие как:

- *метрика контекстной непрерывности* – измерение согласованности генерируемого контента с предысторией и текущим состоянием игрового мира;
- *метрика игрового баланса* – оценка сбалансированности наград и сложности в генерируемых квестах;
- *интегральный показатель качества контента* – общий показатель оценки качества, объединяющий уже реализованные и еще не реализованные оценки.

Такая система метрик релевантна гибридной архитектуре, когда появляется возможность тесной интеграции модуля оценки качества с механизмами генерации контента. Это позволит создать замкнутый цикл улучшения качества персонализации на основе объективных количественных данных, что особенно важно для обеспечения долгосрочной вовлеченности и качества игрового опыта.

ЗАКЛЮЧЕНИЕ

Представлено решение критической проблемы современной игровой индустрии — создания динамических и адаптивных нарративных систем, способных генерировать персонализированный контент в реальном времени. Разработанный компонент интеграции GPT продемонстрировал возможность генерации ветвящихся диалогов с учетом характеристик NPC и контекста игрового мира, при этом практическое тестирование выявило критическое ограничение облачных решений — задержку в 30 с, неприемлемую для интерактивных приложений.

Предложенная гибридная архитектура адаптивной системы представляет собой эволюцию от простой интеграции LLM к комплексному решению, сочетающему управление контекстом, векторную память и систему оценки качества. Модульная структура системы обеспечивает технологическую независимость и возможность интеграции различных языковых моделей, а меха-

низм управления контекстом на основе RAG-подхода решает фундаментальную проблему долгосрочной памяти NPC. Переход к парадигме Edge AI с целевой задержкой 200–500 мс открывает путь к достижению интерактивности в реальном времени, что критически важно для сохранения иммерсивности игрового опыта.

Дальнейшее развитие системы связано с несколькими ключевыми направлениями. Разработка и тонкая настройка специализированных языковых моделей для игровых сценариев позволяют повысить качество генерируемого контента при сохранении компактности модели. Интеграция инструментов автоматической балансировки обеспечит оптимальное соответствие сложности контента уровню игрока. Реализация локальных оптимизированных моделей на основе методов квантования и прунинга устранит зависимость от облачных сервисов и связанные с ними задержки. Расширение системы мультимодальными возможностями, включая генерацию голосовых реплик и анимаций, создаст основу для следующего поколения интерактивных игровых персонажей.

Результаты работы демонстрируют практическую осуществимость создания адаптивных игровых миров, где нарративный контент генерируется динамически и персонализируется под каждого игрока, что открывает новые возможности для геймдизайна и повышения вовлеченности аудитории.

Благодарности

Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»).

ЛУДОГРАФИЯ

Bethesda Game Studios (2011). *The Elder Scrolls V: Skyrim* [Action RPG] [Multiplatform], Bethesda Softworks.

BioWare (1998). *Baldur's Gate* [RPG] [Windows, macOS], Interplay Entertainment, Black Isle Studios.

Epic Games (2017). *Fortnite* [Battle Royale, Sandbox] [Multiplatform], Epic Games.

Infocom (1980). Zork [Text adventure] [Apple II, TRS-80, PDP-10], Infocom.

Mojang Studios (2011). Minecraft [Sandbox, survival] [Multiplatform], Mojang Studios, Xbox Game Studios.

Nick Walton (2019). AI Dungeon [Text adventure, AI-generated] [Web, iOS, Android], Latitude.

Will Crowther, Don Woods (1976). Colossal Cave Adventure [Text adventure] [PDP-10].

ZA/UM (2019). Disco Elysium [RPG] [Windows, macOS, PlayStation, Xbox, Nintendo Switch], ZA/UM.

СПИСОК ЛИТЕРАТУРЫ

1. *Gallotta R. et al.* Large language models and games: A survey and roadmap // IEEE Transactions on Games. 2024.
 2. Inworld, Future of NPCs report // inworld [Электронный ресурс] – February 2023. URL: <https://www.inworld.ai/blog/future-of-npcs-report>
 3. *Sweetser P.* Large language models and video games: A preliminary scoping review // Proceedings of the 6th ACM Conference on Conversational User Interfaces. 2024. P. 1–8.
 4. *Wang Q. et al.* GenQuest: An LLM-based Text Adventure Game for Language Learners // arXiv preprint arXiv:2510.04498. 2025.
 5. *Hardiman J.P.W. et al.* AI-powered dialogues and quests generation in role-playing games using Google's Gemini and Sentence BERT framework // Procedia Computer Science. 2024. Vol. 245. P. 1111–1119.
 6. *Ashby T. et al.* Personalized quest and dialogue generation in role-playing games: A knowledge graph-and language model-based approach // Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023. P. 1–20.
 7. *Большаков Э.С., Кугуракова В.В.* Генеративная симуляция игрового окружения в реальном времени // Электронные библиотеки. 2025. Т. 28, № 2. С. 188–212.
 8. *Нурлыгаянов Н.Р., Кугуракова В.В.* Подход к созданию корпуса текстов видеоигр на основе универсальной структуры // Электронные библиотеки. 2024. Т. 27, № 4. С. 578–597.
-

9. *Akoury N., Yang Q., Iyyer M.* A framework for exploring player perceptions of llm-generated dialogue in commercial video games // Findings of the Association for Computational Linguistics: EMNLP 2023. 2023. P. 2295–2311.
 10. *Jin C., Cao P., Zaïane O.* Role-Playing Based on Large Language Models via Style Extraction // International Conference on Neural Information Processing. Singapore: Springer Nature Singapore, 2024. P. 433–447.
 11. *Tseng Y.M. et al.* Two tales of persona in llms: A survey of role-playing and personalization // arXiv preprint arXiv:2406.01171. 2024.
 12. *Трофимчук В.Т.* Разработка компонента для интеграции GPT в видеоигры: выпуск. квалиф. раб. на с.з. Бакалавр, спец. 09.03.04 – Программная инженерия, науч. рук. Хафизов М.Р., Казанский федеральный университет, Институт информационных технологий и интеллектуальных систем, 2024. 54 с.
URL: https://kpfu.ru/student_diplom/10.160.178.20_FLP3APBW54SAM3JYPPT73DBLRFUS75DXQEVT_Z5F6LD7O0KAF7_F_Trofimchuk.pdf
 13. *Abdelrahman E.* Edge AI and Edge Computing: Powering Real-Time Intelligence [Электронный ресурс] // Ultralytics.
URL: <https://www.ultralytics.com/ru/blog/edge-ai-and-edge-computing-powering-real-time-intelligence>
 14. *Кудерин Д.* Edge AI: как работают нейросети на устройствах с ограниченными ресурсами [Электронный ресурс] // TProger.
URL: <https://tproger.ru/articles/edge-ai--kak-rabotayut-nejroseti-na-ustrojstvah-s-ogranichennymi-resursami>
 15. *Martindale J.* Input lag and response time aren't the same. Here's which is more important [Электронный ресурс]. 2024.
URL: <https://www.digitrends.com/computing/input-lag-vs-response-time/>
 16. *Кугуракова В.В.* Формальный подход к пространственно-временному моделированию игровых систем // Ученые записки Казанского университета. Серия Физико-математические науки. 2024. Т. 166, №. 4. С. 532–554.
 17. *Chen B.* Optimization Strategies for Role-Playing Games Based on Large Language Models // Proceedings of the 2nd International Conference on Data Science and Engineering: ICDSE 2025. P. 632–637.
-

18. Сахибгареева Г.Ф., Кугуракова В.В., Большаков Э.С. Инструменты балансирования игр // Электронные библиотеки. 2023. Т. 26, № 2. С. 225–251.
-

DEVELOPMENT OF AN ADAPTIVE SYSTEM FOR GENERATING GAME QUESTS AND DIALOGUES BASED ON LARGE LANGUAGE MODELS

V. T. Trofimchuk¹ [0009-0001-9106-9614], V. V. Kugurakova² [0000-0002-1552-4910]

^{1, 2} Institute of Information Technologies and Intelligent Systems, Kazan Federal University, 35 Kremlyovskaya st., Kazan, 420008

¹ vselord.beta@gmail.com, ² vlada.kugurakova@gmail.com

Abstract

This article addresses the problem of creating dynamic narrative systems for video games with real-time interactivity. It presents the development and testing of a GPT integration component for dialogue generation, which revealed a critical limitation of cloud-based solutions – a 30-second latency unacceptable for gameplay. A hybrid architecture of an adaptive system is proposed, combining LLMs with reinforcement learning mechanisms. Particular attention is given to solving the problems of game world consistency and managing long-term context of NPC interactions through a RAG approach. The transition to the Edge AI paradigm with the application of quantization methods to achieve a target latency of 200–500 ms is substantiated. Metrics for evaluating personalization and dynamic content adaptation have been developed.

Keywords: *video games, large language models, LLM, dialogue generation, quest generation, adaptive quests, procedural content generation, agent behavior, game AI, machine learning in games.*

Acknowledgment: This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program ("PRIORITY–2030").

LUDOGRAPHY

Bethesda Game Studios (2011). The Elder Scrolls V: Skyrim [Action RPG] [Multiplatform], Bethesda Softworks.

BioWare (1998). Baldur's Gate [RPG] [Windows, macOS], Interplay Entertainment, Black Isle Studios.

Epic Games (2017). Fortnite [Battle Royale, Sandbox] [Multiplatform], Epic Games.

Infocom (1980). Zork [Text adventure] [Apple II, TRS-80, PDP-10], Infocom.

Mojang Studios (2011). Minecraft [Sandbox, survival] [Multiplatform], Mojang Studios, Xbox Game Studios.

Nick Walton (2019). AI Dungeon [Text adventure, AI-generated] [Web, iOS, Android], Latitude.

Will Crowther, Don Woods (1976). Colossal Cave Adventure [Text adventure] [PDP-10].

ZA/UM (2019). Disco Elysium [RPG] [Windows, macOS, PlayStation, Xbox, Nintendo Switch], ZA/UM.

REFERENCES

1. *Gallotta R. et al.* Large language models and games: A survey and roadmap // IEEE Transactions on Games. 2024.
2. Inworld. Future of NPCs report // inworld [Electronic resource]. – February 2023. URL: <https://www.inworld.ai/blog/future-of-npcs-report>
3. *Sweetser P.* Large language models and video games: A preliminary scoping review // Proceedings of the 6th ACM Conference on Conversational User Interfaces. 2024. P. 1–8.
4. *Wang Q. et al.* GenQuest: An LLM-based Text Adventure Game for Language Learners // arXiv preprint arXiv:2510.04498. 2025.
5. *Hardiman J.P.W. et al.* AI-powered dialogues and quests generation in role-playing games using Google's Gemini and Sentence BERT framework // Procedia Computer Science. 2024. Vol. 245. P. 1111–1119.

6. *Ashby T. et al.* Personalized quest and dialogue generation in role-playing games: A knowledge graph-and language model-based approach // Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023. P. 1–20.
7. *Bolshakov E.S., Kugurakova V.V.* Generative simulation of a game environment in real time // Russian Digital Libraries Journal. 2025. Vol. 28, No. 2. P. 188–212 (In Russian).
8. *Nurygayanov N.R., Kugurakova V.V.* An approach to creating a corpus of video game texts based on a universal structure // Russian Digital Libraries Journal. 2024. Vol. 27, No. 4. P. 578–597 (In Russian).
9. *Akoury N., Yang Q., Iyyer M.* A framework for exploring player perceptions of llm-generated dialogue in commercial video games // Findings of the Association for Computational Linguistics: EMNLP 2023. P.2295–2311.
10. *Jin C., Cao P., Zaïane O.* Role-Playing Based on Large Language Models via Style Extraction // International Conference on Neural Information Processing. Singapore: Springer Nature Singapore, 2024. P. 433–447.
11. *Tseng Y.M. et al.* Two tales of persona in llms: A survey of role-playing and personalization // arXiv preprint arXiv:2406.01171. 2024.
12. *Trofimchuk V.T.* Development of a component for GPT integration into video games: Bachelor's qualifying work, spec. 09.03.04 – Software Engineering, scientific supervisor Khafizov M.R., Kazan Federal University, Institute of Information Technology and Intelligent Systems, 2024. 54 p.
URL: https://kpfu.ru/student_diplom/10.160.178.20_FLP3APBW54SAM3JYPPT73DBLRFUS75DXQEBT_Z5F6LD700KAF7_F_Trofimchuk.pdf (In Russian)
13. *Abdelrahman E.* Edge AI and Edge Computing: Powering Real-Time Intelligence [Electronic resource] // Ultralytics.
- URL: <https://www.ultralytics.com/ru/blog/edge-ai-and-edge-computing-powering-real-time-intelligence>
14. *Kuderin D.* Edge AI: how neural networks work on devices with limited resources [Electronic resource] // TProger.

URL: <https://tproger.ru/articles/edge-ai--kak-rabotayut-nejroseti-na-ustrojstvah-s-ogranichennymi-resursami> (In Russian).

15. *Martindale J.* Input lag and response time aren't the same. Here's which is more important [Electronic resource]. 2024.

URL: <https://www.digitltrends.com/computing/input-lag-vs-response-time/>

16. *Kugurakova V.V.* A formal approach to spatio-temporal modeling of game systems // Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki. 2024. Vol. 166, No. 4. P. 532–554 (In Russian).

17. *Chen B.* Optimization Strategies for Role-Playing Games Based on Large Language Models // Proceedings of the 2nd International Conference on Data Science and Engineering: ICDSE 2025. P. 632–637.

18. *Sakhibgareeva G.F., Kugurakova V.V., Bolshakov E.S.* Game balancing tools // Russian Digital Libraries Journal. 2023. Vol. 26, No. 2. P. 225–251 (In Russian).

СВЕДЕНИЯ ОБ АВТОРАХ



TРОФИМЧУК Всеволод Тарасович – лаборант-исследователь научно-исследовательской лаборатории Digital Media Lab Института ИТИС КФУ. Область научных интересов – использование LLM для видеоигр в реальном времени.

Vsevolod Tarasovich TROFIMCHUK – laboratory research assistant at Digital Media Lab of the Institute of ITIS KFU. Area of research interests - using LLM for real-time video games.

email: vselord.beta@gmail.com

ORCID: 0009-0001-9106-9614



КУГУРАКОВА Влада Владимировна – кандидат технических наук, и. о. зав. кафедрой индустрии разработки видеоигр Института ИТИС КФУ, руководитель НИЛ Digital Media Lab. Область научных интересов – формальные методы верификации видеоигр.

Vlada Vladimirovna KUGURAKOVA – Ph.D. of Engineering Sciences, Head of the Video Game Development Industry Department of ITIS KFU, Head of Laboratory «Digital Media Lab». The area of scientific interest is formal methods of video game verification.

email: vlada.kugurakova@gmail.com

ORCID: 0000-0002-1552-4910

Материал поступил в редакцию 10 сентября 2025 года