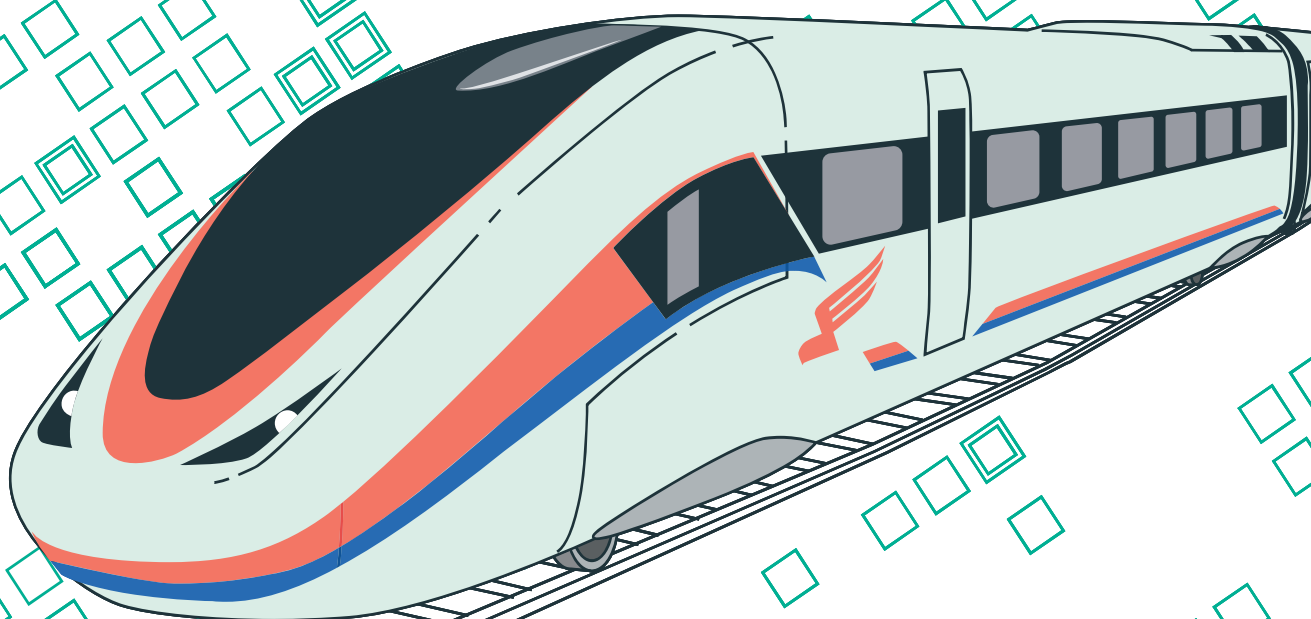


Интеллектуальные технологии на транспорте

Intellectual Technologies
on Transport



Выпуск 2
2026

ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ НА ТРАНСПОРТЕ
(сетевой электронный научный журнал)
Выпуск 2 (46), 2026

ISSN 2413-2527

Сетевой электронный научный журнал, свободно распространяемый через интернет. Публикуются статьи на русском и английском языках с результатами исследований и практических достижений в области интеллектуальных технологий и сопутствующих им научных исследований. Журнал основан в 2015 году.

Учредитель

Федеральное государственное бюджетное образовательное учреждение высшего образования «Петербургский государственный университет путей сообщения Императора Александра I» (ФГБОУ ВО ПГУПС).

Издатель

ООО «Медиа-Сервис» по договору № ЭА87426 от 12.01.2026.

Периодичность выхода — 4 номера в год.

Адрес редакции:

190031, Санкт-Петербург, Московский пр., 9
e-mail: itt-pgups@yandex.ru
Телефон: +7 (812) 457-86-06

Сетевое издание «Интеллектуальные технологии на транспорте (сетевой электронный научный журнал), Intellectual Technologies on Transport» зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций. Свидетельство Эл № ФС77-61707 от 07.05.2015.

Журнал зарегистрирован в Российском индексе научного цитирования (РИНЦ).

Информация предназначена для читателей старше 12 лет.

Выпуски журнала доступны на сайте <http://itt-pgups.ru>

Хомоненко А. Д., д. т. н., проф., С.-Петербург, РФ — главный редактор

Божко Л. М., д. э. н., доц., ПГУПС, С.-Петербург, РФ — заместитель главного редактора

РЕДАКЦИОННЫЙ СОВЕТ

Сопредседатели редакционного совета:

Валинский О. С., к. т. н., ректор ПГУПС, С.-Петербург, РФ
Чаркин Е. И., зам. гендиректора по ИТ ОАО «РЖД», Москва, РФ

Адагуров С. Е., д. т. н., проф., ВНИИЖТ, Москва, РФ
Дудин А. Н., д. ф.-м. н., проф., БГУ, Минск, Беларусь
Корниенко А. А., д. т. н., проф., ПГУПС, С.-Петербург, РФ
Макаренко С. И., д. т. н., доц., ПАО «Интелтех», С.-Петербург, РФ
Меркурьев Ю. А., Dr. Habil., проф., член Латвийской АН, РТУ, Рига, Латвия
Титова Т. С., д. т. н., проф., первый проректор ПГУПС, С.-Петербург, РФ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Александрова Е. Б., д. т. н., проф., ПГУПС, С.-Петербург, РФ
Басыров А. Г., д. т. н., проф., ВКА, С.-Петербург, РФ
Безродный Б. Ф., д. т. н., проф., ООО «1520 Сигнал», Москва, РФ
Благовещенская Е. А., д. ф.-м. н., проф., ПГУПС, С.-Петербург, РФ

Бубнов В. П., д. т. н., проф., С.-Петербург, РФ
Булавский П. Е., д. т. н., доц., ПГУПС, С.-Петербург, РФ
Василенко М. Н., д. т. н., проф., ПГУПС, С.-Петербург, РФ
Глухов А. П., д. т. н., ПГУПС, С.-Петербург, РФ
Гуда А. Н., д. т. н., проф., РГУПС, Ростов-на-Дону, РФ
Ермаков С. Г., д. т. н., проф., ПГУПС, С.-Петербург, РФ
Забаровский В. С., д. т. н., проф., СПбПУ, С.-Петербург, РФ
Канаев А. К., д. т. н., проф., ПГУПС, С.-Петербург, РФ
Котенко А. Г., д. т. н., проф., ВНИИЖТ, Москва, РФ
Куренков П. В., д. э. н., к. т. н., проф., РУТ (МИИТ), Москва, РФ
Лецкий Э. К., д. т. н., проф., РУТ (МИИТ), Москва, РФ
Наседкин О. А., к. т. н., доц., ПГУПС, С.-Петербург, РФ
Никитин А. Б., д. т. н., проф., ПГУПС, С.-Петербург, РФ
Новиков Е. А., д. т. н., доц., ВКА, С.-Петербург, РФ
Охтилев М. Ю., д. т. н., проф., НИО ЦИТ «Петрокомета», С.-Петербург, РФ
Привалов А. А., д. воен. н., проф., Академия войск национальной гвардии, С.-Петербург, РФ
Соколов Б. В., д. т. н., проф., СПб ФИЦ РАН, С.-Петербург, РФ
Таранцев А. А., д. т. н., проф., ИПТ РАН, С.-Петербург, РФ
Утепбергенов И. Т., д. т. н., проф., АУЭС, Алматы, Казахстан
Фазылов Ш. Х., д. т. н., проф., НИИ развития цифровых технологий и ИИ, Ташкент, Узбекистан
Хабаров В. И., д. т. н., проф., СГУПС, Новосибирск, РФ
Ходаковский В. А., д. т. н., проф., ПГУПС, С.-Петербург, РФ
Чехонин К. А., д. ф.-м. н., доц., ХВИЦ ДВО РАН, Хабаровск, РФ

INTELLECTUAL TECHNOLOGIES ON TRANSPORT
(Network electronic scientific journal)

Issue 2 (46), 2026

ISSN 2413-2527

Network electronic scientific journal, open access.
It publishes articles in Russian and English with the
results of research and practical achievements in the field
of intelligent technologies and associated research.
Founded in 2015.

Founder

Federal State Budgetary Educational Institution
of Higher Education “Emperor Alexander I
St. Petersburg State Transport University”

Publisher

Media Service LLC, no. ЭА87426,
12.01.2026.

Frequency of release — 4 issues per year.

Editorial address:

190031, St. Petersburg, Moskovsky ave., 9
e-mail: itt-pgups@yandex.ru
Phone: +7 812 457 86 06

The online journal “Intellectual Technologies
on Transport” is registered by the Federal Service
for Supervision of Communications, Information
Technologies, and Mass Media.
El no. FS77-61707 Testimony from 07.05.2015.

The journal is registered in the Russian Science Citation
Index (RSCI).

The content is for readers over the age of 12.
Issues of the magazine are available at <http://itt-pgups.ru>

Khomonenko A. D., Dr. Sci. in Engineering, Prof.,
St. Petersburg, Russia — Editor-in-Chief

Bozhko L. M., As. Prof., PSTU, St. Petersburg, Russia —
Deputy Editor-in-Chief

EDITORIAL COUNCIL MEMBERS

Co-chairs of the Editorial Council:

Valinsky O. S., PhD in Engineering, rector
of PSTU, St. Petersburg, Russia

Charkin E. I., CIO of JSC “Russian Railways”, Moscow,
Russia

Adadurov S. E., Prof., VNIIZHT, Moscow, Russia

Dudin A. N., Prof., BSU, Minsk, Belarus

Kornienko A. A., Prof., PSTU, St. Petersburg, Russia

Makarenko S. I., As. Prof., Inteltech, St. Petersburg, Russia

Merkuryev Yu. A., Prof., Academician of the Latvian

Academy of Sciences, RTU, Riga, Latvia

Titova T. S., Prof., First Vice-Rector PSTU, St. Petersburg,
Russia

EDITORIAL BOARD MEMBERS

Aleksandrova E. B., Prof., SPbPU, St. Petersburg, Russia

Basyrov A. G., Prof., MSA, St. Petersburg, Russia

Bezrodny B. F., Prof., 1520 Signal LLC, Moscow, Russia

Blagoveshchenskaya E. A., Prof., PSTU, St. Petersburg, Russia

Bubnov V. P., Prof., St. Petersburg, Russia

Bulavsky P. E., As. Prof., PSTU, St. Petersburg, Russia

Vasilenko M. N., Prof., PSTU, St. Petersburg, Russia

Glukhov A. P., PSTU, St. Petersburg, Russia

Guda A. N., Prof., RSTU, Rostov-on-Don, Russia

Ermakov S. G., Prof., PSTU, St. Petersburg, Russia

Zaborovsky V. S., Prof., SPbPU, St. Petersburg, Russia

Kanaev A. K., Prof., PSTU, St. Petersburg, Russia

Kotenko A. G., Prof., VNIIZHT, Moscow, Russia

Kurenkov P. V., Prof., RUT (MIIT), Moscow, Russia

Letsky E. K., Prof., RUT (MIIT), Moscow, Russia

Nasedkin O. A., As. Prof., PSTU, St. Petersburg, Russia

Nikitin A. B., Prof., PSTU, St. Petersburg, Russia

Novikov E. A., As. Prof., MSA, St. Petersburg, Russia

Okhtilev M. Yu., Prof., JSC “Petrokometa”, St. Petersburg,
Russia

Privalov A. A., Prof., Academy of the National Guard Troops,
St. Petersburg, Russia

Sokolov B. V., Prof., SPC RAS, St. Petersburg, Russia

Tarantsev A. A., Prof., IPT RAS, St. Petersburg, Russia

Utepbergenov I. T., Prof., AUPET, Almaty, Kazakhstan

Fazilov Sh. X., Prof., AIRI, Tashkent, Uzbekistan

Khobarov V. I., Prof., STU, Novosibirsk, Russia

Khodakovskiy V. A., Prof., PSTU, St. Petersburg, Russia

Chekhonin K. A., Prof., Khabarovsk FRC RAS,
Khabarovsk, Russia

Содержание

Информационная безопасность и защита данных

Костин М. А., Давыдова Д. М., Петров В. Е.

Реализация RAG-архитектуры для автоматизированной проверки документов
в корпоративных системах хранения с использованием больших языковых моделей 5

Искусственный интеллект и транспортные системы

Зуев Д. В., Бочкарев С. В.

Архитектура цифровых двойников на железнодорожном транспорте 17

Математическое и программное обеспечение вычислительных комплексов и сетей

Забродин А. В., Михеев А. Д., Граур М. А.

Преимущества использования шаблонизаторов Django для бизнеса: баланс между эффективностью
и стоимостью разработки. 36

Стрельников Н. Р., Михайлова С. А., Алимов Н. И.

Инструмент сбора и оценки общественного мнения о транспортной отрасли
из открытых источников. 46

Цуканова Д. А., Забродин А. В.

Реактивное программирование как эволюция архитектурных компромиссов
в управлении изменениями 54

Черкасов С. Д., Хомоненко А. Д.

Исследование влияния фрагментации данных на производительность чтения/записи
и методы ее устранения в MySQL (InnoDB) 64

Математическое моделирование и системный анализ

Бригаднов И. А., Морозов А. В.

Компьютерное моделирование динамических систем на фазовой плоскости 75

Коньков Д. А., Божко Л. М.

Контрастивное самообучение в системах мониторинга транспорта: практический подход
к ранней детекции аномалий 82

Стрелков И. В., Божко Л. М.

Обзор и сравнительный анализ методов моделирования полетных характеристик
беспилотных летательных аппаратов 91

Черняховская Е. С.

Особенности модели выкладки товаров с иерархической категоризацией
на полочном пространстве магазинов и складов 100

Памяти коллеги

Рыжиков Юрий Иванович 109

Contents

Information Security and Data Protection

- Kostin M.A., Davydova D.M., Petrov V.E.**
Implementation of RAG Architecture for Automated Document Verification in Corporate Storage Systems
Using Large Language Models. 5

Artificial Intelligence and Transport Systems

- Zuev D.V., Bochkarev S.V.**
Digital Twin Architecture for Rail Transport 17

Mathematical and Software Support for Computer Complexes and Networks

- Zabrodin A.V., Mikheev A.D., Graur M.A.**
Advantages of Using Django Template Engines for Business: The Balance Between Efficiency
and Cost of Development 36
- Strelnikov N.R., Mikhaylova S.A., Alimov N.I.**
A Tool for Collecting and Assessing Public Opinion on the Transport Industry from Open Sources 46
- Tsukanova D.A., Zabrodin A.V.**
Reactive Programming as an Evolution of Architectural Trade-offs in Managing Changes 54
- Cherkasov S.D., Khomonenko A.D.**
Investigation of Data Fragmentation Impact on Read/Write Performance and Methods
for Its Elimination in MySQL (InnoDB) 64

Mathematical Modelling and System Analysis

- Brigadnov I.A., Morozov A.V.**
Computer Simulation of Dynamic Systems on the Phase Plane. 75
- Konkov D.A., Bozhko L.M.**
Contrastive Self-Supervised Learning in Transportation Monitoring Systems: A Practical Approach
to Early Anomaly Detection 82
- Strelkov I.V., Bozhko L.M.**
Review and Comparative Analysis of Methods for Modeling Flight Characteristics of Unmanned
Aerial Vehicles 91
- Cherniakhovskaia E.S.**
Hierarchical Shelf Space Allocation Model Characteristics in Stores and Warehouses. 100

In memory of colleague

- Ryzhikov Yuri Ivanovich** 109

УДК 004.041+004.056

Реализация RAG-архитектуры для автоматизированной проверки документов в корпоративных системах хранения с использованием больших языковых моделей

- Костин Максим Андреевич** — студент бакалавриата 4-го курса направления 09.03.01 «Информатика и вычислительная техника». Научные интересы: информационные системы, искусственный интеллект, веб-разработка. E-mail: m.kkoston@yandex.ru
- Давыдова Даяна Михайловна** — старший преподаватель кафедры «Информационные и вычислительные системы». Научные интересы: информационные системы, обработка больших данных. E-mail: dayana-0820@bk.ru
- Петров Владимир Евгеньевич** — кандидат воен. наук, доцент кафедры «Информационные и вычислительные системы». Научные интересы: информационные системы, обработка больших данных, моделирование надежности. E-mail: petroffve@mail.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Костин М. А., Давыдова Д. М., Петров В. Е. Реализация RAG-архитектуры для автоматизированной проверки документов в корпоративных системах хранения с использованием больших языковых моделей // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 5–16. DOI: 10.20295/2413-2527-2026-246-5-16

Аннотация. *Предлагается специализированное решение для анализа документов, использующихся в корпоративных системах документооборота. Цель: разработка и практическая реализация RAG-архитектуры для автоматизированной проверки корпоративных документов с использованием локально развернутых больших языковых моделей, обеспечивающей выявление пропущенных обязательных полей, ошибок форматов данных и содержательных противоречий в документах транспортных и логистических систем. Методы: проведены проектирование и программная реализация Java-приложения, интегрирующего модуль извлечения текста из PDF- и DOCX-документов на основе библиотек Apache, векторное хранилище с упрощенными эмбедами на основе частотного анализа слов, алгоритм семантического поиска через вычисление косинусного сходства и клиент взаимодействия с LLM через API сервера Ollama. Результаты: разработанная система продемонстрировала способность к контекстному анализу содержания документов и адаптивность к вариативным форматам представления информации, что позволяет преодолеть ограничения традиционных систем. Экспериментальная проверка выполнена на тестовом наборе корпоративных документов с преднамеренно внесенными ошибками различных типов; оценка эффективности проводилась по метрикам полноты, точности и их среднего, а также по времени отклика системы. Наилучшие результаты показала модель llama3.2:latest при полном исключении передачи конфиденциальных данных за пределы инфраструктуры организации. Практическая значимость: предложенное решение применимо для автоматизации контроля качества документации в корпоративных системах документооборота транспортных предприятий, государственных учреждений и промышленных организаций. Модульная архитектура обеспечивает масштабируемость на другие типы документов и возможность интеграции с существующими информационными системами при минимальных затратах на адаптацию. Использование открытых моделей и локального сервера Ollama снижает зависимость от сторонних облачных сервисов и обеспечивает соответствие требованиям информационной безопасности.*

Ключевые слова: *большие языковые модели, автоматизированная проверка документов, векторный поиск, обработка естественного языка, корпоративные системы, Ollama, извлечение текста, семантический анализ*

2.3.6 — *методы и системы защиты информации, информационная безопасность (технические науки);*

1.2.1 — *искусственный интеллект и машинное обучение (технические науки)*

Введение

Современные корпоративные системы документооборота сталкиваются с беспрецедентным ростом объемов обрабатываемой информации, и традиционные методы проверки документов, основанные на правилах и регулярных выражениях, демонстрируют свою ограниченность при работе с неструктурированными данными и вариативными форматами представления информации. Особенность описанных систем, заключающаяся в неспособности к контекстному анализу содержания, приводит к пропуску содержательных ошибок, некорректному распознаванию сущностей и необходимости постоянной ручной доработки.

Для решения представленных проблем необходимо использовать большие языковые модели (Large Language Models, LLM), которые открыли новые возможности для интеллектуальной обработки документов. Однако не стоит забывать, что прямое применение LLM для анализа корпоративной документации сталкивается с рядом существенных ограничений, в качестве примера которых можно выделить проблемы актуальности информации, тенденцию к «галлюцинациям», отсутствие доступа к специализированным знаниям предметной области и высокие вычислительные затраты. Эти вызовы обуславливают необходимость разработки гибридных подходов, сочетающих достоинства генеративных моделей с точностью традиционных методов информационного поиска.

Для решения подобных проблем необходимо использовать технологию, которая позволит связать механизмы семантического поиска с возможностями генерации текстовых ответов. В качестве такой технологии активно применяются алгоритмы генерации с дополненной выборкой (Retrieval-Augmented Generation, RAG) [1], позволяющие преодолеть ограничения стандартных языковых

моделей. Преимущества RAG заключены в дополнении запроса релевантным контекстом, позволяющим повысить точность обработки корпоративных документов.

Данное исследование посвящено реализации архитектуры приложения на основе RAG-алгоритмов. Таким образом, архитектура системы будет изначально спроектирована для автоматизированной проверки транспортных накладных, заявлений работников и многих других документов, которые ежедневно участвуют в документообороте различных предприятий и организаций. Разработанное решение включает в себя несколько модулей по обработке документов разных форматов (от PDF до DOCX), а также реализации алгоритмов векторного поиска и анализа поступающей в систему информации. Локально спроектированные и связанные модули программы с учетом платформы для использования языковых моделей Ollama позволят обеспечить конфиденциальность информации и минимизируют риски ее утечки во внешние системы.

Целью работы является апробация RAG-архитектуры для автоматизированной проверки документов, обеспечивающей выявление пропущенных значений, идентификацию ошибок формата и содержательных противоречий.

Практическая значимость исследования заключается в создании адаптируемого решения для автоматизации контроля качества корпоративных документов, способного снизить трудозатраты на ручную проверку и минимизировать ошибки в документации.

Проектирование

Для создания системы обработки документов необходимо использовать современные подходы

и методы разработки программного обеспечения. В основе программы, реализующей RAG-архитектуру для работы большой языковой модели, требуется использовать объектно-ориентированный способ организации кода на одном из высокоуровневых языков программирования. Так как в основе большинства приложений банковского сектора, государственных систем документооборота, а также промышленных решений применяется язык программирования Java, то и реализованное приложение было создано на основе его пакетов и системных библиотек. Алгоритм

обработки документов начинается с парсинга текста и формирования фрагментов фиксированного размера — чанков. Для считывания текста из документов необходимо создать утилитный класс Document Processor, в котором будут использоваться библиотеки Apache POI. Последовательное выполнение алгоритма обработки документа через LLM с векторным поиском приведено на рис. 1.

Для функционала работы клиента Ollama и подключения к LLM необходимо реализовать класс, создающий запрос для большой языковой модели

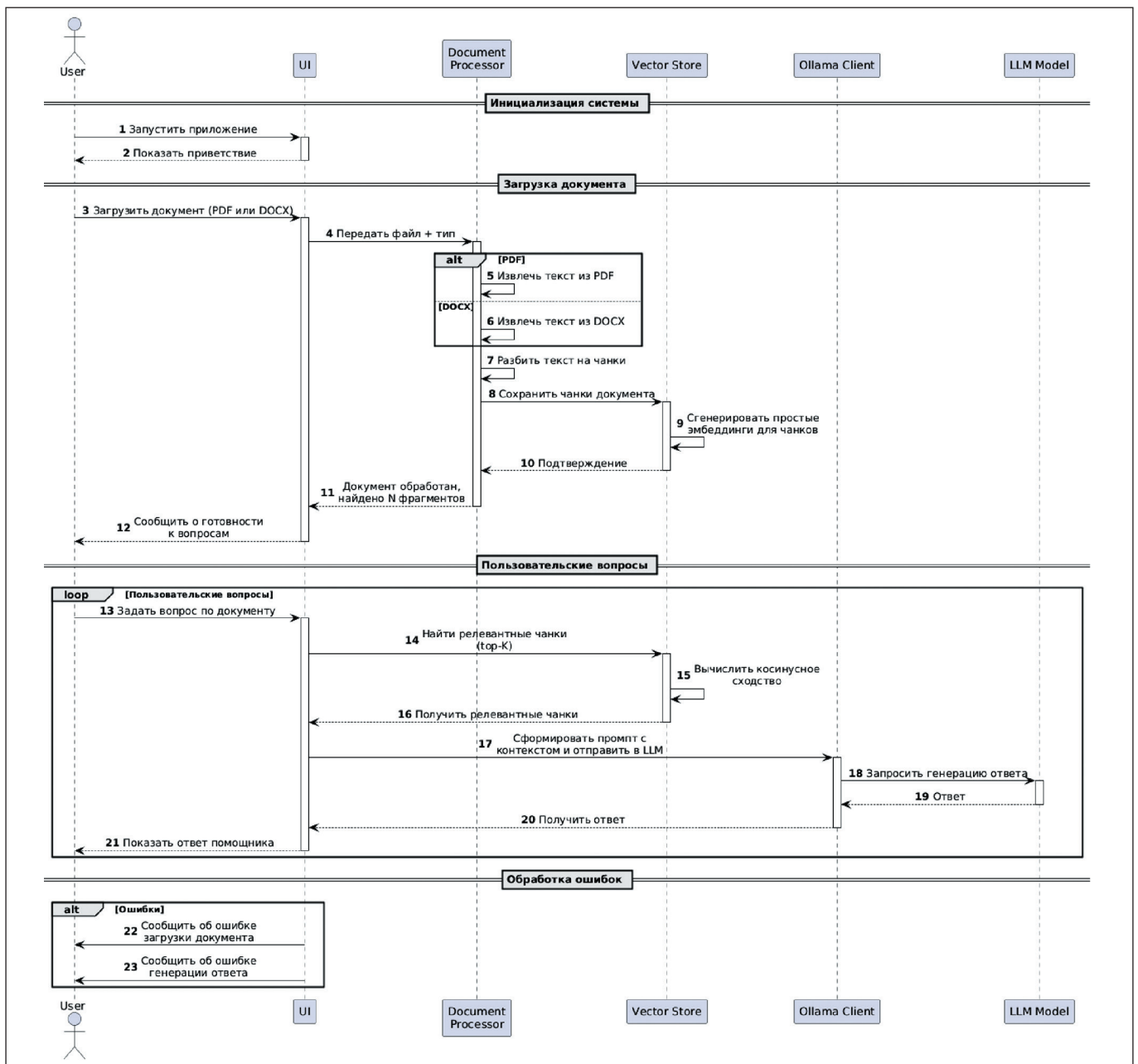


Рис. 1. Диаграмма последовательности системы анализа документации на основе RAG-архитектуры

и десериализующий ее ответ из JSON-формата в строковый тип [2]. В основе векторного хранилища предполагается использовать класс, генерирующий числовое представление фрагментов текста на основе частоты слов, один из методов которого представлен в листинге 1.

Также необходимо отметить и использование метода для вычисления сходства между векторами через измерение угла между двумя ненулевыми векторами, более известного как алгоритм косинусного сходства:

$$\text{sim}(A, B) = \cos(\theta) = \frac{A \times B}{\|A\| \times \|B\|},$$

где $\|A\|$ и $\|B\|$ — нормы векторов;

$A \times B$ — их скалярное произведение.

Реализация метода для нахождения метрики косинусного сходства должна включать вычисление скалярного произведения векторов и их

евклидовых норм с последующим нормированием результата. Значение метрики лежит в диапазоне $[0, 1]$, где 1 соответствует полной идентичности направлений векторов. Подробно метод представлен в листинге 2.

Реализация

В ходе исследования было реализовано приложение на языке программирования Java с использованием сборщика Maven и библиотек Apache для работы с документами различных форматов. Для полноценной работы с сервером Ollama через команду `ollama serve` был запущен сервис [3], к которому созданное Java-приложение обращается через HTTP-запросы [4]. Сервер Ollama работает как фоновая служба и предоставляет API по сетевому адресу, с его помощью вопросы пользователя переадресуются выбранной LLM. На рис. 2 показана диаграмма запуска Ollama-сервера.

```
// Метод для векторного представления текста private List<Double>
generateSimpleEmbedding(String text) {
    // Генерация эмбеддингов
    String[] words = text.toLowerCase().split(«\s+»);
    Map<String, Integer> wordFreq = new HashMap<>();
    // Вычисление частоты слов в строковом массиве
    for (String word : words) {
        wordFreq.put(word, wordFreq.getOrDefault(word, 0) + 1);
    }
    List<Double> embedding = new ArrayList<>();
    // Фиксированный размер эмбеддинга - используется ограничитель (100)
    for (int i = 0; i < 100; i++) {
        embedding.add(0.0);
    }
    int index = 0;
    for (String word : wordFreq.keySet()) {
        if (index >= 100) break;
        embedding.set(index, (double) wordFreq.get(word));
        index++;
    }
    //[2.0, 2.0, 1.0, 1.0, 0.0, 0.0, 0.0, ..., 0.0]
    return embedding;
}
```

Листинг 1. Метод создания векторного представления текста

```
private double cosineSimilarity(List<Double> vec1, List<Double> vec2) {
    double dotProduct = 0.0;
    double norm1 = 0.0;
    double norm2 = 0.0;
    for (int i = 0; i < vec1.size(); i++) {
        dotProduct += vec1.get(i) * vec2.get(i);
        norm1 += Math.pow(vec1.get(i), 2); //норма вектора 1 и 2
        norm2 += Math.pow(vec2.get(i), 2);
    }
    return dotProduct / (Math.sqrt(norm1) * Math.sqrt(norm2));
}
```

Листинг 2. Реализация алгоритма косинусного сходства двух векторов

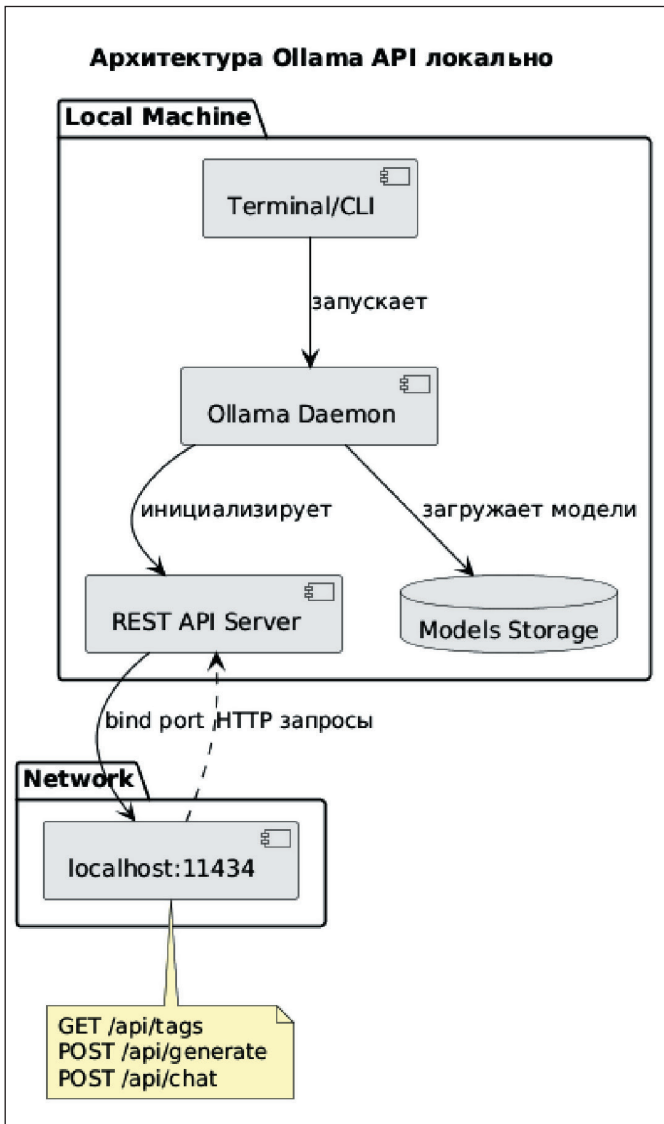


Рис. 2. Запуск сервиса Оллама для инициализации фоновой службы

В разработанном приложении предоставляется выбор LLM по желанию пользователя. Для демонстрации RAG-алгоритма и работы векторного поиска использовались несколько моделей: phi3:mini, qwen3, llama3.2. К преимуществам моделей, развернутых локально, можно отнести повышенную безопасность и контроль за передаваемой информацией. Так как доступ к модели и данным, которые она обрабатывает, есть только у владельцев инфраструктуры и ПО, использующих ее, утечка данных минимизируется [3]. Однако в ходе исследования был проведен анализ выбранных моделей, описанный в следующем разделе. Клиентская часть системы взаимодействует с сервером Ollama через HTTP/HTTPS-протокол, используя Java HttpClient [5], введенный в Java 11. Для каждого запроса формируется JSON-объект с параметром temperature, установленным на уровне 0,1, что обеспечивает детерминированность ответов модели. Это критически важно для задач проверки документов, где требуются высокая точность и воспроизводимость результатов. Параметр top_p = 0,9 в JSON ограничивает выбор токенов наиболее вероятными кандидатами, снижая вероятность генерации нерелевантного контента. Полная реализация метода для обращения к LLM отражена в листинге 3.

В ходе реализации был создан и GUI пользователя с возможностью отправления в систему нескольких видов документов [6]. Производительность обработки текста и скорость ответа зависят

```

public String generateResponse(String prompt) throws Exception {
    String jsonPayload = String.format(“”””
        {
            “model”: “%s”,
            “prompt”: “%s”,
            “stream”: false,
            “options”: {
                “temperature”: 0.1,
                “top_p”: 0.9
            }
        }
        “”””, modelName, escapeJson(prompt));
    HttpRequest request = HttpRequest.newBuilder()
        .uri(URI.create(GENERATE_URL))
        .header(“Content-Type”, “application/json”)
        .POST(HttpRequest.BodyPublishers.ofString(jsonPayload))
        .build();
    HttpResponse<String> response = client.send(request, HttpResponse.BodyHandlers.ofString());
    if (response.statusCode() != 200) {
        throw new RuntimeException(“HTTP Error: “ + response.statusCode() + “ - “ + response.body());
    }
    JsonObject jsonResponse = JsonParser.parseString(response.body()).getAsJsonObject();
    return jsonResponse.get(“response”).getAsString();
}

```

Листинг 3. Клиентский метод для обращения к языковой модели и возвращения ее ответа

от типа используемой модели: чем больше языковая модель, тем больший объем знаний и алгоритмов в ней заложен, что существенно влияет на результат. В разработанном решении использовались небольшие языковые модели, при всех своих недостатках гарантирующие разбор отправленного на анализ текста. В качестве примера при тестировании системы использовался классический документ любой организации — заявление о предоставлении отпуска, представленный на рис. 3, в котором заранее была предусмотрена ошибка.

Языковая модель успешно справилась с ошибкой и вернула системе результат анализа (рис. 4), в котором подробно описываются недочеты документа. Помимо единичного ответа, пользователю предоставляется и возможность продолжить беседу с LLM, так как в системе предусмотрено хране-

ние истории текущего чата, основанное на компонентах библиотеки Java Swing [6].

Программная реализация векторного представления текста, основанная на разбиении на слова, создании словаря и формировании вектора фиксированной длины, показана в алгоритмической схеме на рис. 5.

Алгоритм векторизации текста включает четыре основных этапа:

1. Токенизация текста. На данной стадии алгоритма исходная текстовая строка приводится к нижнему регистру для обеспечения регистронезависимости, после чего разбивается на отдельные слова по пробельным символам. Результатом этапа является упорядоченный массив токенов (слов).

2. Построение частотного словаря. Следующим этапом осуществляется последовательный

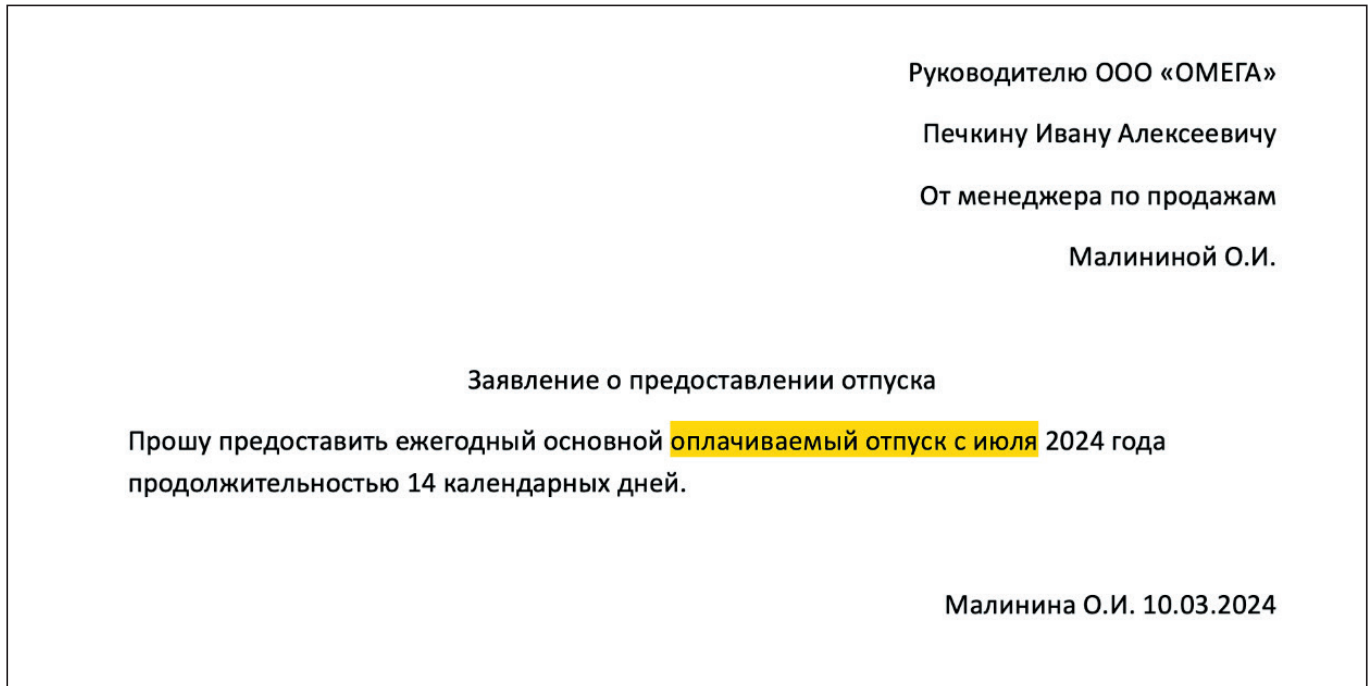


Рис. 3. Шаблон документа с ошибкой, который был отправлен в систему

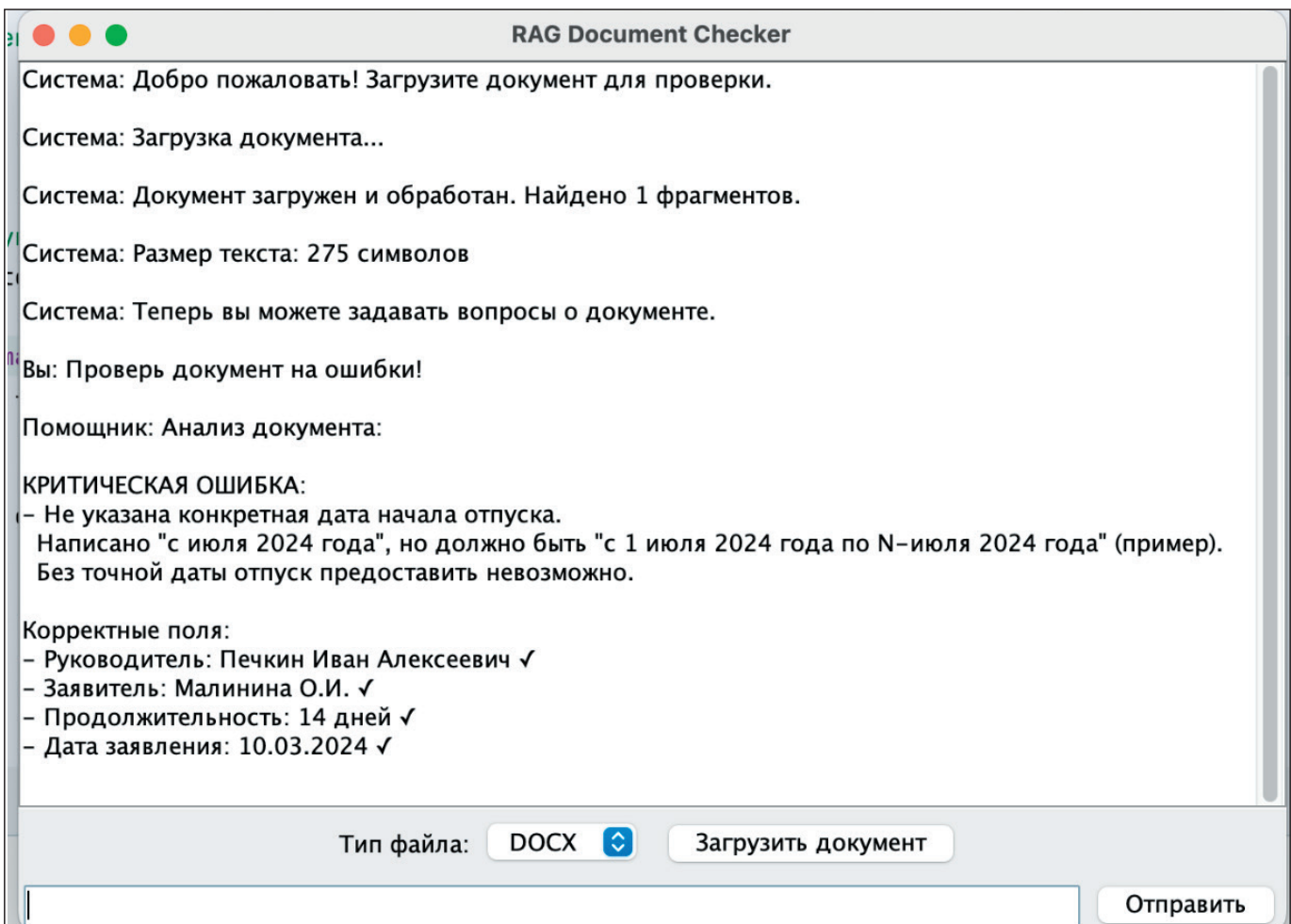


Рис. 4. Результат работы системы на LLM llama:3.2:latest после обработки документа с ошибкой

Алгоритм работы VectorStore

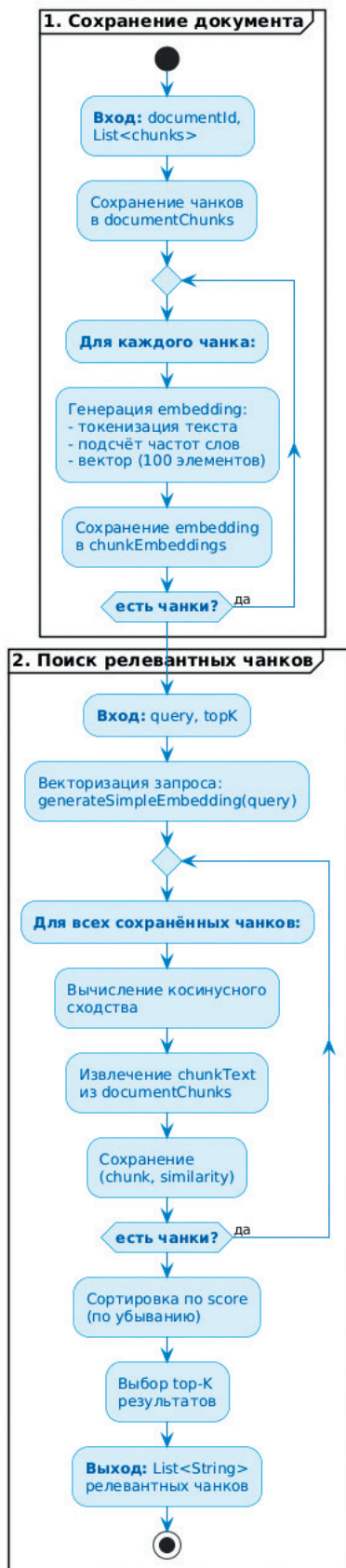


Рис. 5. Реализация метода векторного представления текста с ограничением длины вектора 100 символов

обход массива слов с подсчетом частоты вхождения каждого уникального токена [7]. Для хранения частотного распределения используется ассоциативный массив (словарь), где ключом выступает слово, а значением — количество его вхождений в текст.

3. Инициализация вектора фиксированной размерности. На этой стадии алгоритма создается числовой вектор заранее определенной длины (100 измерений), все элементы которого инициализируются нулевыми значениями. Длина вектора фиксируется, чтобы обеспечить совместимость с векторной базой данных.

4. Проставление значений вектора. Заключительный этап алгоритма включает заполнение исходного массива специальными значениями, которые показывают частоту встречаемых слов. Данный процесс происходит вследствие размещения строковых переменных в векторе последовательно, начиная с нулевой позиции. Так происходит до тех пор, пока либо не закончатся все уникальные слова из общего перечня, либо не будет достигнут допустимый размер вектора.

Результат алгоритма (он же описанный вектор) представляет собой числовую характеристику выбранного фрагмента текста. В таком массиве первые позиции содержат частотные характеристики, а оставшиеся хранят нулевые значения.

Так как обработка данных, которые будут заполнять вектор, не может быть безграничной, важной частью программной системы является эффективное управление памятью. Для минимизации потребления оперативной памяти, которая активно задействуется для локальных языковых моделей, при работе с крупными файлами реализован механизм потоковой обработки данных. Суть механизма состоит в разделении поступающей информации по блокам и их последующей обработке. Предлагаемый подход позволит обрабатывать крупные массивы входной информации, так как с ним программное решение не будет загружать сразу весь массив в память сервера или персонального устройства: вместо этого данные будут обрабатываться последовательно по блокам [8].

Дополнительным решением, принятым при разработке системы для минимизации потребляемых ресурсов, на которое необходимо обратить внимание, стала возможность управления количеством чанков для одного обрабатываемого фрагмента текста. Описанное количество регулируется через параметр, который может редактироваться при запуске ПО.

Отдельно стоит упомянуть встроенный в Java механизм своевременного освобождения ресурсов, который активно задействуется во всех классах программы. Этот механизм реализован с использованием конструкции `try-with-resources` [9], позволяющей автоматически закрывать сетевые соединения и потоки ввода-вывода при возникновении нестандартных ситуаций.

Можно заключить, что архитектура разработанного решения спроектирована с учетом принципов модульности и открытости/закрытости, что обеспечивает возможность расширения функционала без модификации существующих стабильных компонентов. Добавление новых типов проверок документов реализуется посредством модификации системного промпта, направляемого большой языковой модели. Это позволяет гибко адаптировать логику анализа под специфические требования предметной области без изменения программного кода приложения [10].

Интеграция дополнительных источников знаний осуществляется через расширение интерфейса векторного хранилища, что открывает возможность подключения внешних баз знаний, онтологий или специализированных словарей для повышения точности семантического поиска.

Поддержка мультимодальных документов, содержащих изображения, таблицы и структурированные данные, предусмотрена на уровне модульной архитектуры процессора документов: добавление нового парсера для конкретного формата требует реализации соответствующего обработчика, совместимого с общим интерфейсом извлечения текста, при этом остальные компоненты системы, включая векторное хранилище и клиент взаимодействия с языковой моделью, остаются неизменными.

Описанный подход обеспечивает масштабируемость решения и возможность его адаптации к разнообразным сценариям использования в корпоративных системах документооборота различных отраслей.

Экспериментальная проверка

Экспериментальная проверка разработанной системы проводилась на наборе тестовых документов, включающем заявления на отпуск, служебные записки и транспортные накладные. В документы преднамеренно вносились различные типы ошибок: пропущенные обязательные поля (дата, подпись, Ф.И.О.), орфографические и грамматические ошибки, несоответствия форматов дат и числовых значений, логические противоречия в содержании.

В ходе тестирования были оценены базовые языковые модели, доступные Ollama и не требующие высоких вычислительных ресурсов. К ним относятся Phi-3 Mini, Qwen3b и Llama 3.2. Тесты показали, что Phi3 обладает высокой скоростью обработки, но имеет ограничения в выявлении сложных смысловых противоречий по сравнению с более сбалансированными моделями Qwen3 и Llama 3.2. Наилучшей моделью, обеспечивающей оптимальную производительность при низком потреблении оперативной памяти, стала Llama 3.2 с 3 млрд параметров.

Для оценки эффективности системы использовались метрики полноты ответов, для выявления ошибок от общего числа внесенных ошибок, и точности, характеризующей долю корректно выявленных ошибок от общего числа обнаруженных недочетов. Результаты показали, что среднее значение характеристик точности и полноты находится на уровне 0,85 при выявлении пропущенных фрагментов текста и около 0,78 при обнаружении противоречий. Данные характеристики неотъемлемы от времени отклика системы, которое вычислялось при тестировании текстового документа в 2000 символов. Время отклика системы с языковой моделью Llama 3.2 при проверке различного функционала представлено на рис. 6.

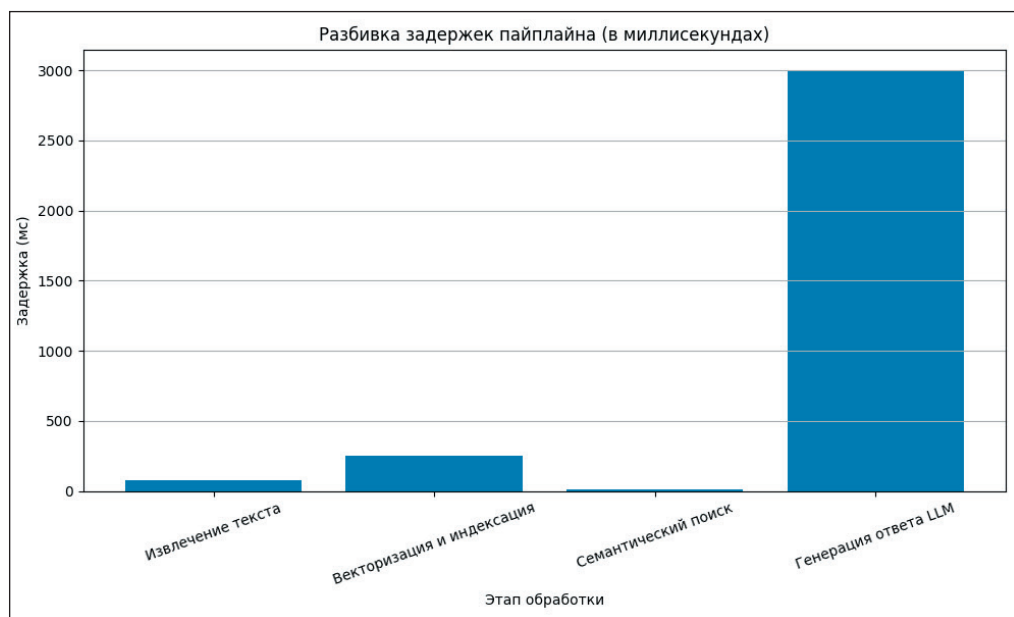


Рис. 6. Результаты тестирования отклика программной системы при использовании модели Llama 3.2

В результате проведенного анализа можно заметить, что общее время работы созданного программного решения не выше 5 сек. при любом режиме работы (от извлечения текста до генерации ответа), следовательно, нормы интерактивного режима работы успешно выполняются.

Заключение

В ходе настоящего исследования было реализовано программное многомодульное решение на языке программирования Java. Основа архитектуры этого приложения включает в себя RAG-алгоритмы, методы для взаимодействия с локальными языковыми моделями, а также специализированные функции для парсинга текста из документов и их последующей обработки. Главное достижение заключается в реализации локального контекстного анализа текстовой информации, поступающей на вход в систему. Разработанное приложение может не только искать ошибки, но и разбирать недочеты в содержании поступающей информации. Эта характеристика позволяет как преодолеть ограничения стандартных языковых моделей, так и сформировать предпосылки для использования данной разработки в уже существующих системах документооборо-

та с повышенными требованиями к конфиденциальности данных.

Система отвечает в среднем за 5 сек., что позволяет использовать ее в интерактивном режиме. При этом не приходится переписывать ядро системы, так как модульная архитектура упрощает добавление новых типов документов. Вместе с тем текущая реализация имеет ряд ограничений: упрощенные эмбединги на основе частотного анализа уступают современным трансформерным моделям в точности семантического поиска, а поддержка многоязычных текстов и документов со сложной версткой требует дополнительной доработки модуля извлечения данных.

Дальнейшее повышение точности выявления смысловых противоречий целесообразно связать с дообучением модели на специализированном корпусе документов транспортной отрасли.

Научная и практическая ценность работы заключается в доказательстве технической реализуемости эффективных систем верификации на базе открытых технологий, что полностью соответствует современным тенденциям суверенизации ИТ-инфраструктуры и снижает зависимость корпоративного сектора от коммерческих облачных API.

СПИСОК ИСТОЧНИКОВ

1. Ротман Д. RAG и генеративный ИИ. Создаем собственные RAG-пайплайны с помощью LlamaIndex, Deep Lake и Pinecon = RAG-Driven Generative AI: Build custom retrieval augmented generation pipelines with LlamaIndex, Deep Lake, and Pinecone. СПб.: Питер, 2025. 320 с.
2. Java SE/JDK Version 26 API Specification: Module java.base. URL: <http://docs.oracle.com/en/java/javase/26/docs/api/java.base/module-summary.html> (дата обращения: 30.01.2026).
3. Ollama's Documentation. URL: <http://docs.ollama.com> (дата обращения: 02.02.2026).
4. Шилдт Г. Java. Полное руководство. 12-е изд. = Java: The Complete Reference. Twelfth Edition / пер. с англ. и ред. Ю. Н. Артеменко. СПб.: Диалектика, 2023. 1344 с.
5. Хорстманн К. С. Java. Библиотека профессионала. Т. 1. Основы. 10-е изд. = Core Java. Volume I — Fundamentals. Tenth Edition / пер. с англ. и ред. И. В. Берштейна. М.: Вильямс, 2016. 864 с.
6. Портянкин И. А. Swing. Эффективные пользовательские интерфейсы. 2-е изд. М.: Лори, 2011. 607 с.
7. Huang D., Wang Z. LLMs at the Edge: Performance and Efficiency Evaluation with Ollama on Diverse Hardware // Proceedings of the International Joint Conference on Neural Networks (IJCNN 2025) (Rome, Italy, 30 June — 5 July 2025). Institute of Electrical and Electronics Engineers, 2025. 8 p. DOI: 10.1109/IJCNN64981.2025.11228317
8. Vahaj M., Raza S. M., Nehra V. Retrieval Augmented Generation (RAG) using LLMs // Proceedings of the Annual International Conference on Data Science, Machine Learning and Blockchain Technology (AICDMB 2025) (Mysuru, India, 27–28 June 2025). Institute of Electrical and Electronics Engineers, 2025. 5 p. DOI: 10.1109/AICDMB64359.2025.11277692
9. Блох Д. Java. Эффективное программирование. 3-е изд. = Effective Java. Third Edition / пер. с англ. и ред. И. В. Красикова. СПб.: Диалектика, 2019. 464 с.
10. Claim Verification in the Age of Large Language Models: a Survey / A. Dmonte [et al.] // ArXiv. 2024. Vol. 2408.14317. 9 p. DOI: 10.48550/arXiv.2408.14317

Дата поступления: 26.03.2026

Решение о публикации: 25.05.2026

Implementation of RAG Architecture for Automated Document Verification in Corporate Storage Systems Using Large Language Models

- Maksim A. Kostin** — 4th year Bachelor's Degree Student in 09.03.01 Informatics and Computer Technology. Research interests: information systems, artificial intelligence, web development. E-mail: m.kkoston@yandex.ru
- Dayana M. Davydova** — Senior lecturer of the “Information and Computing Systems” Department. Research interests: information systems, big data processing. E-mail: dayana-0820@bk.ru
- Vladimir E. Petrov** — PhD in Military Sciences, Associate Professor of the “Information and Computing Systems” Department. Research interests: information systems, big data processing, reliability modeling. E-mail: petroffve@mail.ru

Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky ave., Saint Petersburg, 190031, Russia

For citation: Kostin M. A., Davydova D. M., Petrov V. E. Implementation of RAG Architecture for Automated Document Verification in Corporate Storage Systems Using Large Language Models, *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 5–16. DOI: 10.20295/2413-2527-2026-246-5-16 (In Russian)

Abstract. *A specialized solution is offered for analyzing documents used in corporate document management systems. Purpose: to develop and implement a RAG architecture for automated verification of corporate documents*

using locally deployed large language models that identify missing required fields, data format errors, and substantive contradictions in documents of transport and logistics systems. **Methods:** Java application was designed and implemented programmatically, integrating a text extraction module from PDF and DOCX documents based on Apache libraries, a vector storage with simplified embeddings based on frequency analysis of words, a semantic search algorithm through cosine similarity calculation and an LLM interaction client via the Ollama server API. **Results:** the developed system has demonstrated the ability to contextually analyze the content of documents and adaptability to variable information presentation formats, which makes it possible to overcome the limitations of traditional systems. The experimental verification was performed on a test set of corporate documents with intentionally introduced errors of various types; the effectiveness was assessed by the metrics of completeness, accuracy and their average, as well as by system response time. The llama3.2:latest model showed the best results, while completely excluding the transfer of confidential data outside the organization's infrastructure. **Practical significance:** the proposed solution is applicable for automation of documentation quality control in corporate document management systems of transport enterprises, government agencies and industrial organizations. The modular architecture provides scalability to other types of documents and the ability to integrate with existing information systems at minimal cost of adaptation. Using open models and a local Ollama server reduces dependence on third-party cloud services and ensures compliance with information security requirements.

Keywords: large language models, automated document verification, vector search, natural language processing, corporate systems, Ollama, text extraction, semantic analysis

REFERENCES

1. Rothman D. RAG i generativnyy II. Sozdaem sobstvennyye RAG-payplaynyy s pomoshchyu LlamaIndex, Deep Lake i Pinecon [RAG-Driven Generative AI: Build Custom Retrieval Augmented Generation Pipelines with Llamaindex, Deep Lake and Pinecone]. Saint Petersburg, Piter Publishing House, 2025, 320 p. (In Russian)
2. Java SE/JDK Version 26 API Specification: Module java.base. Available at: <http://docs.oracle.com/en/java/javase/26/docs/api/java.base/module-summary.html> (accessed: January 30, 2026).
3. Ollama's Documentation. Available at: <http://docs.ollama.com> (accessed: February 02, 2026).
4. Schildt H. Java. Polnoe rukovodstvo, 12-e izdanie [Java: The Complete Reference. Twelfth Edition]. Saint Petersburg, Dialektika Publishing House, 2023, 1344 p. (In Russian)
5. Horstmann C. S. Java. Biblioteka professionala. Tom 1. Osnovy. Desyatoe izdanie [Core Java. Volume I — Fundamentals. Tenth Edition]. Moscow, Williams Publishing House, 2016, 864 p. (In Russian)
6. Portyankin I. A. Swing. Effektnye polzovatelskie interfeysy [Spectacular User Interfaces]. Moscow, Lori Publishing House, 2011, 607 p. (In Russian)
7. Huang D., Wang Z. LLMs at the Edge: Performance and Efficiency Evaluation with Ollama on Diverse Hardware, *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2025)*, Rome, Italy, June 30 — July 5, 2025. Institute of Electrical and Electronics Engineers, 2025, 8 p. DOI: 10.1109/IJCNN64981.2025.11228317
8. Vahaj M., Raza S. M., Nehra V. Retrieval Augmented Generation (RAG) using LLMs, *Proceedings of the Annual International Conference on Data Science, Machine Learning and Blockchain Technology (AICDMB 2025)*, Mysuru, India, June 27–28, 2025. Institute of Electrical and Electronics Engineers, 2025, 5 p. DOI: 10.1109/AICDMB64359.2025.11277692
9. Bloch J. Java. Effektivnoe programmirovaniye. Tretye izdanie [Effective Java. Third Edition]. Saint Petersburg, Dialektika Publishing House, 2019, 464 p. (In Russian)
10. Dmonte A., et al. Claim Verification in the Age of Large Language Models: A Survey, *ArXiv*, 2024, vol. 2408.14317, 9 p. DOI: 10.48550/arXiv.2408.14317

Received: March 26, 2026

Accepted: May 25, 2026

УДК 004.94:656.2

Архитектура цифровых двойников на железнодорожном транспорте

Зуев Денис Владимирович — кандидат техн. наук, генеральный директор. Научные интересы: нейронные сети, цифровые двойники. E-mail: zuevdv@gmail.com

Бочкарев Сергей Владимирович — кандидат техн. наук, технический директор. Научные интересы: нейронные сети, цифровые двойники. E-mail: bochkareffsv@yandex.ru

ООО «Синтез АТ», Россия, 199004, Санкт-Петербург, В. О., Средний пр., д. 28/2

Для цитирования: Зуев Д. В., Бочкарев С. В. Архитектура цифровых двойников на железнодорожном транспорте // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 17–35. DOI: 10.20295/2413-2527-2026-246-17-35

Аннотация. Представлен углубленный системный анализ архитектурных подходов к построению и внедрению цифровых двойников на железнодорожном транспорте, что является ключевым этапом цифровой трансформации отрасли. **Цель:** систематизация архитектурных подходов к созданию цифровых двойников, анализ ключевых технологических решений и выработка рекомендаций по их интеграции в единую цифровую экосистему управления инфраструктурой и подвижным составом. **Методы:** проведено сравнительное исследование ведущих мировых платформ (Siemens Railigent X, Bentley iTwin, Dassault 3DEXPERIENCE) и отечественных разработок, в ходе которого выявлены ограничения существующих решений в части адаптации к специфике железнодорожной автоматики и телемеханики. **Результаты:** в отличие от традиционных фрагментарных информационных систем предложена концепция «единого источника правды», базирующаяся на бесшовной интеграции технологий BIM, PLM и SCADA, что позволяет обеспечить сквозное управление жизненным циклом транспортных активов. Научная новизна исследования заключается в разработке и детализации пятиуровневой иерархической модели цифрового двойника, которая объединяет уровень физических объектов и датчиков, каналы передачи данных, платформу хранения и цифровых моделей, аналитический слой и интерфейсы визуализации. **Практическая значимость:** обусловлена описанием механизмов перехода от планово-предупредительного ремонта к предиктивному и риск-ориентированному управлению. Показано, что внедрение предложенной архитектуры позволяет прогнозировать критические неисправности (например, дефекты буксовых узлов) за несколько месяцев до их возникновения и проводить имитационное моделирование различных эксплуатационных сценариев. Результаты исследования могут быть использованы при проектировании интеллектуальных систем управления инфраструктурой и подвижным составом для повышения безопасности и экономической эффективности перевозочного процесса.

Ключевые слова: цифровой двойник, железнодорожный транспорт, многоуровневая архитектура, предиктивная аналитика, промышленный интернет вещей, BIM-технологии, управление жизненным циклом, мониторинг технического состояния, интеллектуальные транспортные системы, риск-ориентированный подход

2.9.8 — интеллектуальные транспортные системы (технические науки); **1.2.2** — математическое моделирование, численные методы и комплексы программ (технические науки)

Введение

Цифровая трансформация железнодорожного транспорта характеризуется переходом от фрагментарных информационных систем к интегрированным интеллектуальным платформам, обеспечивающим сквозное управление жизненным циклом технических систем. Одним из ключевых направлений данной трансформации является внедрение технологии цифровых двойников, позволяющей объединить физические объекты инфраструктуры и подвижного состава с их виртуальными моделями и аналитическими сервисами в едином информационном пространстве.

Актуальность исследования обусловлена ростом требований к надежности, безопасности и эффективности функционирования железнодорожной сети, а также необходимостью перехода к риск-ориентированному и предиктивному управлению. Традиционные подходы, основанные на регламентном обслуживании и разрозненных системах мониторинга, не обеспечивают своевременного выявления деградационных процессов и не позволяют учитывать комплексное влияние факторов эксплуатации.

В этих условиях цифровой двойник рассматривается как киберфизическая система, обеспечивающая непрерывную связь между реальным объектом, потоками диагностических данных и математическими моделями его поведения. Архитектурно такие системы представляют собой многоуровневые решения, включающие уровень датчиков и телеметрии, уровень передачи данных, платформу хранения и обработки, аналитические модули и пользовательские интерфейсы управления.

Несмотря на активное развитие зарубежных платформ (Siemens Railigent X, Bentley iTwin, Dassault 3DEXPERIENCE), а также появление отечественных решений, остается ряд нерешенных задач: отсутствие унифицированной архитектуры, сложность интеграции с BIM/PLM-системами, высокая зависимость от качества данных и ограниченная адаптация к специфике железнодорожной автоматики и телемеханики.

Целью статьи является систематизация архитектурных подходов к построению цифровых двойни-

ков на железнодорожном транспорте, анализ ключевых технологических решений и определение направлений их интеграции в единую цифровую экосистему управления инфраструктурой и подвижным составом.

Типовая многослойная архитектура цифрового двойника

Цифровой двойник железной дороги — это многокомпонентная система, объединяющая данные от физических активов железнодорожной инфраструктуры с их виртуальной моделью и аналитикой в реальном времени [1–5]. Реализация цифрового двойника железной дороги разделяется на уровни, каждый из которых отвечает за свою функцию обработки данных [4, 5]. Примерная схема такой архитектуры приведена на рисунке.

1. Уровень физических объектов и датчиков.

Включает сами железнодорожные активы (напольно-технологическое оборудование, пути, инженерные сооружения) и установленные на них специализированные датчики. Датчики в зависимости от назначения измеряют вибрацию, токи, напряжения, положения механизмов и др. Например, на современном локомотиве могут быть акселерометры, датчики температуры букс, токоприемников и тормозных систем [6]. На железнодорожной инфраструктуре применяются стационарные и мобильные системы диагностики, датчики которых позволяют измерять диагностические параметры активов и накапливать данные для мониторинга активов [7]. Этот уровень обеспечивает непрерывный мониторинг состояния объектов в полевых условиях.

2. Уровень связи и передачи данных.

Данные с датчиков передаются через встроенные контроллеры и шлюзы в коммуникационную сеть. Используются бортовые системы сбора телеметрии (например, «телеметрические модули» локомотивов) и промышленные контроллеры (ПЛК) на объектах инфраструктуры [8]. Связь осуществляется по каналам сотовой связи (GSM-R, 4G/5G), радиоканалам, спутниковым каналам или через проводные сети. Например, в грузовых локомотивах DB Cargo



Рисунок. Архитектура цифрового двойника

данные ~7000 параметров собираются системой TechLok и передаются почти в реальном времени по GSM на платформу аналитики [9]. Каналы связи зачастую дублируются для надежности. За рубежом появляется тренд использования сетей 5G/FRMCS для высокоскоростной передачи больших объемов данных с поездов [2]. Важную роль играют полевые шлюзы IoT — устройства, агрегирующие потоки сенсорных данных и передающие их на верхние уровни с первичной фильтрацией.

3. Уровень платформы данных и цифровой модели. Полученные сырые данные поступают в централизованную платформу, где выполняются их обработка и хранение [2, 5, 10]. Здесь располагается сама цифровая модель — виртуальное представление объектов. Платформа включает базы данных (например, исторический архив, облачное хранилище больших данных) и модели объектов: BIM-модели инфраструктуры, CAD-модели подвижного состава, схемы систем, а также физические модели для симуляции. Сырые данные очищаются, нормализуются и сохраняются в структурированном виде (таблицы баз данных,

файлы) [10]. Платформа обеспечивает интерфейсы для доступа к данным верхним приложениям. В архитектуре академических исследований этот слой называют сервисом данных [11]. Он служит «единым источником правды» о состоянии системы: интегрирует разнородные данные (проектные данные BIM/GIS, телеметрию сенсоров, оперативную информацию). Например, на проекте цифрового двойника железной дороги Transpennine (Великобритания) платформа Bentley iTwin объединила данные из 60 разных систем — от BIM-проектов до GIS-слоев, предоставив 1300 пользователям единый доступ к актуальной информации [11]. В российских проектах также создаются интеграционные платформы. Так, ОАО «РЖД» разработало «Доверенную среду локомотивного комплекса», объединяющую данные о тяговом составе в единую систему [6].

4. Аналитический уровень и приложения. На основе собранных данных работают сервисы аналитики, моделирования и управления [10]. Сюда относятся алгоритмы предиктивной аналитики (прогнозирование отказов, остаточного

ресурса), модули оптимизации (например, расчет оптимального расписания поездов, их маршрутов), симуляционные модели (имитация работы станции или участка железнодорожной линии), а также прикладные приложения для конечных пользователей. Часто именно здесь реализуется «ядро» цифрового двойника — вычислительные модели, принимающие данные реального времени и выдающие рекомендации. Например, приложение Siemens Railigent X Health States анализирует в облаке данные о состоянии букс и ходовых частей поездов с помощью AI-моделей и выдает оценку состояния узлов (в виде светофора: зеленый/желтый/красный) для поддержки решений о ремонте [12]. Другое приложение, Remaining Useful Life, рассчитывает остаточный ресурс компонентов. В аналитическом уровне могут использоваться нейронные сети, статистические модели, алгоритмы распознавания образов (например, для анализа видеоданных с камер инспекции путей). Прогнозное обслуживание — одна из ключевых задач: по данным о вибрациях и температурах алгоритмы способны заранее выявить зарождающуюся неисправность (например, дефект буксы за 6–7 месяцев до отказа) [9]. Также в этом слое реализуются симуляции «что, если», например моделирование последствий введения нового расписания или строительства новой станции. Платформы высшего уровня (например, Dassault 3DEXPERIENCE) позволяют на цифровом двойнике прогонять различные сценарии для поддержки решений [13].

5. Уровень визуализации и управления. Верхний уровень — это интерфейс для пользователей и интеграция с системами управления [10]. Он предоставляет дашборды, 3D-визуализацию, отчеты и команды. Современные цифровые двойники часто имеют 3D-представление, например визуализацию железнодорожной сети на карте с наложением данных о движении поездов и инцидентах в реальном времени [10]. Используются технологии AR/VR для иммерсивного отображения, вплоть до виртуальных диспетчерских залов. На этом уровне осуществляется и обратная связь: операторы могут через интерфейс задавать команды системе — от настройки порогов срабаты-

вания до отправки управляющих воздействий на объекты. Например, Railigent позволяет при выявлении проблемы автоматически формировать наряд на ремонт в системе управления обслуживанием (CMMS) через открытый API [12]. В перспективе некоторые решения могут закрывать полный контур управления: цифровой двойник напрямую задействован в автоматизированном диспетчерском управлении или автоведении поездов, хотя пока чаще он используется как система поддержки решений.

Таким образом, архитектура цифрового двойника объединяет уровень IoT (Интернет вещей), платформу данных и уровень интеллектуальных сервисов с интерфейсом для человека. В исследованиях предлагаются сходные схемы — от четырехслойной архитектуры (подключение данных → сервисы данных → аналитика → визуализация) [11] до более сложных, включающих слои интеграции с внешними системами и управления жизненным циклом. Ключевая особенность заключается в наличии сквозной связи от физического сенсора до цифровой модели и обратно.

Основные аппаратно-программные компоненты

Развертывание цифрового двойника опирается на широкий спектр аппаратных и программных средств, обеспечивающих сбор и обработку данных.

1. Датчики и системы мониторинга. Как отмечалось, на подвижном составе и инфраструктуре устанавливаются разнообразные датчики, например, акселерометры и инклинометры для отслеживания вибраций и крена [14], термодатчики на буксах и двигателях, датчики износа тормозных колодок, датчики давления в тормозной магистрали, датчики тока тяговых двигателей, GPS/ГЛОНАСС для точного позиционирования. В пути широко применяются стационарные диагностические системы (их можно считать «датчиками» инфраструктуры), например, автоматический контроль нагрева букс, система обнаружения дефектов колес, датчики схода колесных пар, метеостанции около пути (фиксируют температуру рельса,

ветер, осадки), а также системы технической диагностики и мониторинга устройств железнодорожной автоматики. Кроме того, существуют встроенные системы диагностики самих поездов — бортовые ОТК. Все эти источники информации формируют массив данных для цифрового двойника объекта контроля. Пример масштаба: цифровой двойник нового поезда CRRC получает данные более чем с 2000 датчиков, установленных на поезде. Аналогично, инфраструктурный двойник моста или тоннеля может получать потоки от сотен датчиков (деформации, нагрузки, сенсоров вибрации).

2. Промышленные контроллеры и шлюзы.

Для сбора сигналов с датчиков и передачи их дальше используются контроллеры и IoT-шлюзы. Например, в депо при создании отечественного комплекса диагностики ОКТА разработчики сначала пробовали стандартные ПЛК Siemens SIMATIC и модули ICP DAS [8]. Эти устройства обычно хороши для локального мониторинга, но оказались ограничены по гибкости интеграции с облаком (нет поддержки современных веб-протоколов, сложность расширения) [5]. Поэтому был сделан акцент на собственные контроллеры с возможностью двусторонней связи через интернет. IoT-шлюз — это мини-компьютер или специализированный модем, который агрегирует данные от множества датчиков по локальным интерфейсам (например, Modbus, CAN, RS-485) и отправляет их пакетом через интернет-протоколы (HTTP(S), MQTT и другие) на сервер. В случае локомотива таким шлюзом может быть бортовой компьютер телеметрии с SIM-картой, а на инфраструктурном объекте — промышленный маршрутизатор с подключением к оптоволоконной сети. Шлюзы также могут выполнять предварительную обработку (фильтрацию, сглаживание данных, локальный анализ на простые аномалии), чтобы снизить нагрузку на канал связи [9]. Современные тренды включают перенос AI в контроллер для мгновенной реакции: например, датчики на пути могут локально определять дефект и сразу сообщать о нем, не передавая гигабайты исходных сырых данных [9].

3. Системы SCADA и телеметрия. Традиционные железнодорожные SCADA-системы (диспет-

черские системы контроля и сбора данных) продолжают играть свою роль. Например, энергетическая SCADA отслеживает состояние контактной сети, тяговых подстанций и может передавать эти данные в цифровой двойник инфраструктуры. Также к источникам данных можно отнести системы ЖАТ (автоматики и телемеханики) — данные релейных шкафов, состояние светофоров, стрелок, показания рельсовых цепей. Цифровой двойник интегрируется с системами источников данных через адаптеры (например, подключение к системам и устройствам ЖАТ, к системам верхнего уровня диспетчеризации). Таким образом, цифровой двойник не заменяет SCADA, а дополняет ее, предоставляя более высокий уровень анализа. В отрасли прослеживается переход от разрозненных SCADA к объединенным ситуационным центрам на базе цифровых двойников, где информация из разных подсистем собирается на общей модели для комплексного отображения и принятия решений.

4. Облачные платформы и хранилища данных. Большинство современных цифровых двойников реализованы с использованием облачных технологий. Это позволяет хранить и обрабатывать огромные массивы данных, использовать масштабируемые вычисления для AI/ML. Примером может служить платформа Siemens MindSphere, выступающая как облако IoT: она собирает данные от подключенных устройств (поездов, датчиков инфраструктуры) и предоставляет сервисы для их обработки [15]. Специализированное решение Siemens Railigent X фактически работает как надстройка над MindSphere, ориентированная на железнодорожные данные [15]. Облако обеспечивает и интеграцию сторонних сервисов: Railigent открыт для подключения приложений от других поставщиков. Так, компания Voith встроила в Railigent модуль мониторинга автосцепок Scharfenberg [9], а компания SKF подключила систему контроля букс Insight Rail [9]. Это достигается за счет стандартных API и микросервисной архитектуры облачной платформы. В России по понятным причинам ставка делается на отечественные облачные решения (например, платформы на базе российских ЦОД). Архитектура платформы

ОКТА, упомянутой выше, построена по принципу: все модули диагностических стендов депо управляются напрямую из облака в режиме онлайн [8]. То есть конфигурации, обновления алгоритмов, сбор данных — все централизовано, что решает проблемы территориальной рассредоточенности и ускоряет тиражирование изменений.

5. Аналитическое ПО и модели. Программные компоненты, выполняющие анализ, — это ядро цифрового двойника. Здесь используются библиотеки машинного обучения (например, TensorFlow, PyTorch для построения нейросетей), инструменты Big Data (Hadoop, Spark) для параллельной обработки потоков, специализированные алгоритмы обработки сигналов (цифровая фильтрация временных рядов, спектральный анализ для детекции дефектов). В области железных дорог есть и готовые аналитические решения: например, Alstom HealthHub — платформа от Alstom, аналог Railigent, «заточенная» под сбор данных с поездов и инфраструктуры для диагностики состояния [8]. Также применяются цифровые модели для имитации — от микроуровня (модель динамики экипажа для прогнозирования износа колес и рельсов) до макроуровня (имитация графика движения поездов). Например, платформа имитационного моделирования AnyLogic использована HCLTech для построения «когнитивного» цифрового двойника сортировочной станции: все данные IoT, АСУ, геоплатформ собираются и в режиме реального времени питают имитационную модель AnyLogic, которая отображает текущее состояние и прогнозирует развитие ситуации [16]. Благодаря таким моделям операторы могут прогонять различные сценарии (например, «что, если этот путь закроется?») и получать прогноз последствий.

6. Средства визуализации и взаимодействия. Пользователь взаимодействует с цифровым двойником через специализированные интерфейсы. Это могут быть веб-порталы с дашбордами, графиками, картами либо настольные приложения (например, 3D-Viewer для инфраструктурных моделей). Тенденция — отображение информации на геопространственной основе (карте или 3D-модели местности). Например, цифровой двойник пассажирского

движения может визуализировать на карте сеть, показывать движущиеся поезда и раскрашивать станции по уровню пассажиропотока в реальном времени [10]. Другой пример — 3D-модель тоннеля или моста, на которой цветом отображается степень износа конструкций по данным датчиков. Для персонала могут создаваться AR-приложения: взглянув через планшет на локомотив, специалист видит «наложение» — визуализацию внутренних параметров и подсказки по ремонту. На уровне управления предприятием используются центры управления: видеостены, совмещающие схему дороги, видео с камер, графики состояния техники. Цифровые двойники все чаще интегрируются с диспетчерскими системами. Например, программный комплекс для управления перевозками получает данные прогноза цифрового двойника и может корректировать график движения для предотвращения сбоев. Компания Dassault Systèmes отмечает, что их платформа объединяет планирование перевозок, управление инфраструктурой, ремонтами и даже расчет работы персонала в едином цифровом пространстве [17], то есть диспетчер, инженер по инфраструктуре и начальник депо работают с общим виртуальным «макетом» железной дороги, видя актуальные данные и прогнозы.

Сравнение ключевых платформ и решений

В настоящее время несколько крупных промышленных платформ лидируют в сфере цифровых двойников железных дорог. Рассмотрим их особенности и архитектурные подходы.

Siemens Railigent X

Siemens Railigent X — пакет приложений для обслуживания железнодорожных активов, работающий поверх IoT-платформы MindSphere [15]. Сервис решает задачи предиктивного обслуживания и мониторинга подвижного состава и смежной инфраструктуры. Архитектура сервиса выглядит следующим образом: Railigent собирает данные с оборудования поездов через бортовые системы диагностики и внешние датчики (например, стационарные пункты контроля колес) и передает данные в облако MindSphere, где работают

аналитические сервисы. Отличительная особенность сервиса — богатая библиотека приложений от Siemens для конкретных задач: состояние буксовых узлов, прогноз остаточного ресурса, мониторинг дверей, климатических систем и т. д. В архитектуру заложена открытость API: Railigent интегрируется с системами клиентов (например, с корпоративными CMMS для автоматического оформления заявок на ремонт [12]). Также Siemens допускает подключение сторонних модулей: есть примеры, когда к Railigent подключены решения по мониторингу автосцепок и система контроля вибрации букс для расширения функций [9]. Аппаратная часть включает в себя сенсоры (например, вибродатчики) и коммуникационные устройства (промышленные маршрутизаторы для подвижного состава). Railigent используется на высокоскоростных поездах Velaro и городских поездах Thameslink [9]. Немецкая DB Cargo применяет Railigent совместно с платформой GE для анализа данных со своих локомотивов [9]. В новых депо Siemens внедряет концепцию «цифрового депо» — все данные о прибывающем поезде (неисправности, износ) автоматически собираются в Railigent и прямо распределяются в задания ремонтным бригадам на планшеты [18].

Bentley OpenRail (платформа iTwin)

Bentley OpenRail и смежные решения (Bentley AssetWise, платформа iTwin) — это семейство продуктов, ориентированных на инфраструктурные цифровые двойники. Решение сфокусировано на жизненном цикле железнодорожной инфраструктуры — от проектирования (BIM-модели путей, мостов, тоннелей, станций) до эксплуатации и технического обслуживания объектов. Bentley продвигает концепцию «федеративного» цифрового двойника, где платформа iTwin объединяет различные данные: BIM-проекты, геодезические данные, GIS, таблицы и датчики. При этом данные не сводятся в одну монолитную модель, а хранятся как взаимосвязанные ресурсы с привязкой к пространству и времени. Например, в проекте модернизации Transpennine (Network Rail, Великобритания) на базе iTwin удалось свести данные более чем

из 60 разрозненных систем в единую цифровую среду, дополнив их 600 слоями GIS-данных [11]. Это дало всем участникам (более 1300 специалистов) доступ к актуальной информации в реальном времени, ускорив принятие решений и сэкономив ~20 тысяч человеко-часов за первые шесть месяцев [11].

К компонентам Bentley OpenRail относятся: OpenRail Designer — ПО для 3D-проектирования железных дорог (позволяет создать подробную модель пути, контактной сети и других элементов железнодорожной инфраструктуры), ContextCapture — инструмент для создания реалистичных 3D-моделей местности (из лазерного сканирования, съемки дроном), ProjectWise — система общих данных (CDE) для управления документацией и моделями. Над ними работает облачная служба Bentley iTwin, которая поддерживает версиюность и обновление цифрового двойника по мере изменений в реальном мире.

Особенности решения Bentley OpenRail — акцент на стандарты OpenBIM (IFC). Bentley поддерживает формат IFC Rail, обеспечивая интероперабельность данных инфраструктуры [19]. Кроме вышеупомянутого Transpennine, решение Bentley OpenRail применяется в проекте обновления сигнализации на вокзале Бристоль Темпл Мидс (Великобритания). С помощью OpenRail была создана 3D-модель путевого развития и сигналов, проведены виртуальные «проезды» машиниста для проверки видимости сигналов. Это позволило выявить и исправить ошибки размещения сигналов до строительства, сократить выезды на место и, по мнению Network Rail, свести месяцы работы к одному совещанию благодаря наглядному согласованию в цифровой модели [11]. Экономический эффект Bentley OpenRail — устранение переделок, ускорение ввода в эксплуатацию новых систем. Bentley также предлагает AssetWise Digital Twin для эксплуатации — это уровень интеграции с датчиками, например подключение беспроводных датчиков вибрации на мостах к модели AssetWise для предиктивного мониторинга состояния конструкций (решения такого типа

эксплуатируются на железных дорогах Италии и других стран ЕС).

Dassault Systèmes 3DEXPERIENCE

Dassault Systèmes 3DEXPERIENCE — платформа, изначально созданная для интеграции системы управления жизненным циклом продукта и 3D-моделирования, которая в настоящее время адаптирована под транспортную отрасль как основа «виртуальных двойников». Dassault продвигает термин Virtual Twin, подразумевая расширение концепции цифрового двойника до моделирования целой системы систем с учетом окружающей среды [13].

Основной фокус платформы делается на сквозном управлении всеми аспектами железнодорожной системы — от проектирования и производства подвижного состава и инфраструктуры до планирования перевозок и технического обслуживания. 3DEXPERIENCE представляет собой единое цифровое пространство, объединяющее всех участников и данные [13].

В основе платформы лежит интеграция продуктов Dassault: CATIA (3D-проектирование), DELMIA (моделирование производственных и операционных процессов), SIMULIA (инженерный анализ и симуляция), ENOVIA (управление данными и процессами). На железной дороге это выливается в то, что все службы — производственные, инфраструктурные, операционные — работают на общей модели. Например, Alstom использует 3DEXPERIENCE для управления конструкторскими данными и технологическими процессами при производстве вагонов [13]. А компания KiwiRail (Новая Зеландия) внедрила платформу для оптимизации грузовых перевозок, планирования парка и маршрутов [13].

Dassault делает упор на системную модель — виртуальный двойник учитывает не только отдельный объект, но и всю связанную систему и даже внешние факторы. Например, при выборе места для новой станции виртуальный двойник города позволяет проиграть множество сценариев, учитывая пассажиропотоки, подвоз пассажиров, влияние на окружающую инфраструктуру, и найти оптимальное решение [13]. Платформа

поддерживает 360-градусную визуализацию операций для операторов и совместную работу: все данные, идеи, ограничения находятся в одном пространстве, доступном безопасно с любого устройства [13]. Интеграция данных: виртуальный двойник 3DEXPERIENCE поглощает данные предиктивного обслуживания, состояния движения, планов путей, чтобы оптимизировать расписание и операции по перевозочному процессу [17]. Реализована связь с реальными датчиками: поток данных в реальном времени позволяет поддерживать «постоянный цикл улучшений», когда данные эксплуатации возвращаются в модель для анализа и корректировки проекта [13].

Европейские операторы перевозок запустили пилотные проекты по использованию виртуального двойника в управлении инфраструктурными мегапроектами. Например, проект Rail Baltica рассматривает использование 3DEXPERIENCE для унификации стандартов и совместного ведения огромного объема данных разных подрядчиков [17].

В целом Dassault предлагает унифицировать BIM, GIS, модели эксплуатации и обслуживания на одной платформе, разрывая узкие специализированные «силосы» данных между организациями [17]. Такой подход особенно актуален при создании транснациональных коридоров и повышении интероперабельности.

Отечественные решения

После 2022 года доступ российских железнодорожных компаний к западным платформам затруднен, поэтому акцент сместился на развитие отечественных программно-аппаратных комплексов. В ОАО «РЖД» цифровые двойники выделены как ключевое направление цифровой трансформации [20]. Основным направлением выбрана предиктивная диагностика локомотивов и инфраструктуры, а также создание цифровых моделей сети для планирования.

1. «Умный локомотив». Комплекс систем от РЖД, в котором применяются технологии IoT, нейронных сетей и цифровых двойников для мониторинга состояния узлов локомотива [1]. Данные с бортовых систем локомотивов передаются

в единую платформу, где AI-модели анализируют отклонения параметров, выявляют аномалии и формируют уведомления о состоянии оборудования [1]. Такая система уже позволила перейти от реактивного ремонта «по событию» к обслуживанию по состоянию. По оценкам, использование цифрового двойника локомотива может снизить затраты на его обслуживание до 20–30 % за счет своевременного планирования ремонтов и предотвращения серьезных отказов [1].

2. «Доверенная среда локомотивного комплекса». Это цифровая платформа РЖД для хранения и обмена информацией между всеми участниками, обеспечивающими работу локомотивов [1]. Она служит единой базой технических данных локомотивного парка. Такая платформа является фундаментом для построения цифровых двойников техники и связана с системой «Умный локомотив», обеспечивая интеграцию данных от локомотиворемонтных заводов, депо, эксплуатационных подразделений.

3. Цифровой двойник инфраструктуры. РЖД декларировали планы создать цифровую модель всей инфраструктуры сети. В рамках этих планов было реализовано несколько проектов по формированию цифрового двойника железнодорожной инфраструктуры на основе технической документации из единой корпоративной платформы ведения технической документации (ЕКП ТД) и данных из открытых картографических источников [21–23]. Полученные модели позволяют оперативно актуализировать информацию об объектах и использовать ее для оптимизации работ по техническому обслуживанию. Пока полная модель не создана, но локальные фрагменты внедряются. Так, в рамках развития ЕКП ТД в части технологии цифрового двойника реализованы следующие средства автоматизации: управление жизненным циклом аппаратуры и напольно-технологического оборудования железнодорожной автоматики [24, 25] и построение исполненной кабельной сети [26].

4. Аппаратные разработки. Компания «ТВЕМА» осваивает производство датчиков мониторинга пути. Создана система ОКТА для депо,

включающая облачную платформу и набор диагностических модулей (стационарные и мобильные стенды) [8]. В системе каждый стенд оснащен своими сенсорами и контроллерами, но они удаленно управляются из облака, загрузка новых алгоритмов, изменение настроек — все делается через интернет. Это решает проблему территориальной рассредоточенности депо: инженеру не требуется ехать за тысячу километров, чтобы обновить программное обеспечение на стенде, все происходит централизованно. Благодаря ОКТА удалось сократить простои в реальных депо, обеспечить полную прослеживаемость ремонта [8]. До 2022 года часть аппаратных разработок пытались строить на импортных решениях от Siemens, но столкнулись с их закрытостью и слабой интеграцией с облаком, поэтому был сделан выбор в пользу кастомных решений [8]. Российские разработчики стремятся создавать более открытые, программно-ориентированные системы, чтобы не столкнуться с ограничениями проприетарных западных стандартов.

В целом российские системы пока развиваются по нескольким направлениям: локальные (для депо, для конкретных узлов) и глобальные (масштаб компании ОАО «РЖД»). Российским компаниям приходится наверстывать то, что у Siemens, Alstom нарабатывалось годами. Плюсом является возможность учесть мировой опыт: новые системы сразу проектируются с учетом современных требований гибкости (поддержка REST API, веб-технологий, модульность) [8].

Китайские инициативы в части цифровых двойников

Китай активно внедряет цифровые двойники как в подвижном составе, так и в инфраструктуре, опираясь на собственную электронную базу и технологии больших данных. Основные направления — интеллектуальные поезда, высокоскоростные магистрали, диспетчеризация на основе AI.

1. Цифровой двойник поезда CINOVA2.0 (CRRC). В 2023 году корпорация CRRC представила первый в Китае умный межгородской электропоезд с технологиями ADAS (Advanced Driver Assistance) и Digital Twin [27]. Этот поезд

снабжен множеством сенсоров и бортовых систем, интегрированных в единую интеллект-систему. Поезд оснащен более чем 2000 датчиками, а для анализа данных применяются примерно 130 алгоритмов машинного обучения. Цифровой двойник поезда работает параллельно с реальным поездом: в реальном времени получает телеметрию и просчитывает состояние каждого узла, прогнозируя вероятность отказов. Например, на основе вибрации и температуры подшипников он предупреждает о приближающемся износе, на основе данных о профиле пути и нагрузке — оптимизирует управление тягой для экономии энергии. Таким образом, CRRC стремится к полной цифровой копии поезда, которую можно использовать и для автономного управления (ADAS предоставляет машинисту рекомендации, а в перспективе — автопилот). Также цифровой двойник повышает безопасность, например моделирует сценарии экстренного торможения на разном состоянии рельсов, чтобы оптимально настроить антиблокировочные системы.

2. Высокоскоростные магистрали. В Китае реализовано несколько пилотных проектов цифровых двойников на линиях ВСМ. Например, на новом высокоскоростном вокзале Гуанчжоу Байюнь была разработана концепция цифрового двойника станции с интеграцией BIM-модели, системы управления пассажиропотоками, IoT-датчиков оборудования станции и алгоритмов оптимизации расписания прибытия поездов. Цель — «интеллектуальный вокзал», где цифровой двойник в режиме реального времени отслеживает все системы и поток пассажиров и помогает операторам принимать решения. Также China State Railway Group (оператор сети) объявляла, что ведет создание Единой цифровой платформы диспетчеризации: собираются данные от всех поездов, инфраструктуры и внешней среды (погода, состояние энергосети) и на основе AI рассчитываются оптимальные сценарии управления движением. Уже сейчас на некоторых линиях в диспетчерских применяется Decision Support System с элементами цифрового двойника, который прогнозирует конфликтные ситуации в графике и предлагает диспетчеру решения.

3. Аппаратная часть. Китайские компании (например, Huawei, CASCO) разработали собственные IoT-контроллеры для железных дорог, промышленные 5G-модемы для подвижного состава и пр. Повсеместно на железной дороге внедряется связь 5G. Например, в новом метро Гуанчжоу запущена система, где поезд передает большой поток данных (видео с камер, данные сенсоров) в центр по сети 5G, и на основе этого цифровой двойник поезда и тоннеля обеспечивает автоматическое вождение (практически уровень АТО GoA4). Благодаря местным технологиям искусственного интеллекта китайские железные дороги активно внедряют системы мониторинга, основанные на компьютерном зрении: камеры на путях и поездах, объединенные в цифровой двойник, автоматически распознают препятствия, повреждения оборудования и др. Это повышает безопасность и уменьшает зависимость от человеческого фактора.

Отличительная черта китайского подхода — масштаб и централизованность. Реализуется концепция «Цифровая железная дорога», в которой вся национальная сеть рассматривается как единая киберфизическая система. Огромные инвестиции направлены на стандартизацию данных и интеграцию: например, ведется разработка национального стандарта данных для цифровых двойников железной дороги (аналог европейских RailML, IFC Rail). Китайские проекты часто достигают впечатляющих цифр: так, на одном из полигонов отмечена генерация 1 Тб данных с 40 км путей в час при плотном оснащении датчиками [9]. Для того чтобы справиться с такими потоками, уделяется внимание технологиям фильтрации и периферийным вычислениям.

Интеграция цифрового двойника с BIM- и PLM-системами и системами мониторинга и аналитики

Цифровой двойник приносит максимальную пользу, когда он не изолирован, а связан со всеми информационными контурами железной дороги — от систем проектирования (BIM/PLM) до оперативной диспетчеризации и управления жизненным циклом активов (EAM/RCM). Современные платформы это учитывают.

1. BIM-системы и ГИС. На этапе проектирования инфраструктуры создаются подробные информационные модели объектов — пути, станции, мосты, тоннели и прочие активы железнодорожной инфраструктуры. Цифровой двойник может строиться на основе этих BIM-моделей, актуализируя их по данным эксплуатации. Например, Национальная компания железных дорог Люксембурга (CFL) все проекты ведет в BIM и ставит цель к 2035 году оцифровать всю существующую инфраструктуру в виде BIM-моделей, связав их с операционными данными [28]. Это позволит вносить изменения (например, результаты лазерного сканирования состояния туннеля) напрямую в цифровую модель, поддерживая ее синхронность с реальностью. Интеграция с ГИС дает геоконтекст: например, цифровой двойник может соотносить данные о поездах с информацией о рельефе, климатических зонах и т. д. Bentley и другие поддерживают слияние BIM-систем и ГИС: к цифровому двойнику были подключены сотни слоев ГИС (топография, инженерные сети, экологические ограничения) для улучшения решений [28]. При эксплуатации связывать BIM с реальными данными позволяют технологии реальности: регулярное сканирование обновляет 3D-модель, отражая износ, ремонты. Таким образом, BIM-модель эволюционирует в цифровой двойник, проживая вместе с объектом весь жизненный цикл.

2. PLM-системы управления активами. Для подвижного состава и оборудования важно связывать цифровой двойник с данными PLM/PDM, то есть конструкторской и ремонтной документацией. Dassault идет именно по этому пути: виртуальный двойник подключен к полной цифровой модели изделия. Это значит, что любая деталь локомотива в двойнике «знает» свой серийный номер, чертеж, материалы, историю изготовителя [13]. Когда двойник выявляет проблему, система сразу может подтянуть из PLM-системы аналоги случаев, чертежи для ремонта и даже внести предложение конструкторам по улучшению [13]. Системы управления активами (EAM/MCM), например Maximo и SAP PM, также интегрируются: цифровой двойник автоматически генерирует в них заявки, обновля-

ет статус работ. В Railigent, как уже упоминалось, есть возможность через API формировать рабочие задания в CMMS при срабатывании определенных условий [12]. То есть связка «цифровой двойник — EAM» реализует концепцию предиктивного обслуживания на практике — от датчика до конкретной ремонтной операции.

3. Мониторинг и диспетчеризация. Оперативное управление движением поездов и мониторинг состояния инфраструктуры — традиционные процессы, которые выигрывают от использования цифровых двойников. Например, виртуальный двойник инфраструктуры может эмулировать работу всей сети в ускоренном режиме, прогнозируя конфликты в расписании. Dassault сообщает, что их решения позволяют учитывать данные трафика, состояние инфраструктуры и подвижного состава вместе, оптимизируя расписание движения и распределение ресурсов [17]. Виртуальный двойник способен оценивать, к примеру, влияние закрытия перегона на ремонт: он смоделирует перераспределение поездопотоков, найдет узкие места и предложит оптимальный план. Также в диспетчерских системах появляются решающие автоматы на базе AI, использующие данные двойника. Например, в Китае на магистрали Пекин — Шанхай тестировалась система, которая в режиме реального времени перестраивает график при отклонениях, поддерживая максимальную пропускную способность. Цифровой двойник тут служит «песочницей» для AI — он постоянно прогоняет возможные варианты действий и выбирает лучший без риска для реальных поездов.

4. Предиктивная аналитика и Big Data. Многие железнодорожные компании накапливают огромные массивы исторических данных о поломках, задержках, погодных условиях, которые не всегда эффективно используются. Цифровой двойник с интегрированной системой анализа больших данных позволяет выявлять скрытые зависимости. Например, нидерландская ProRail создала DataLab, где на данных цифрового двойника сети обучаются модели для прогнозирования отказов стрелок, сбоях сигнализации, даже появления посторонних на путях [9]. Модель проникновения

на пути, учитывающая окружение, погоду, праздники, смогла предсказать места, где вероятны нарушения, и полиция, принимая меры превентивно, снизила инциденты на проблемных участках на 50–100% [9]. Такой междисциплинарный анализ стал возможен, когда данные разрозненных систем (охрана, движение поездов, погода) объединены в рамках единого цифрового представления.

5. Стандартизация данных. Интеграция множества систем требует общих стандартов. В BIM это IFC Rail (международный формат для инфраструктуры), в ж/д данных — формат RailML для

обмена данными расписаний, топологии, сигнализации. Современные платформы поддерживают эти стандарты, что облегчает совместную работу. Dassault, например, продвигает идею, что их платформа поможет унифицировать стандарты взаимодействия глобально [17]. В Европе инициативы Shift2Rail и Europe's Rail ставят задачу создать референсную архитектуру цифрового двойника железной дороги, чтобы любой компонент (модель вагона, модель инфраструктуры, алгоритм прогноза) мог быть подключен по единым интерфейсам.

Таблица

Сравнение ключевых платформ цифровых двойников

Платформа/решение	Основное направление применения	Архитектурные особенности и компоненты	Примеры внедрения/использования
Siemens Railigent X	Операции обслуживания подвижного состава и мониторинг инфраструктуры	<ul style="list-style-type: none"> – облачная IoT-платформа для сбора телеметрии поездов и путевых датчиков; – встроенные AI-приложения для предиктивной аналитики; – открытые API для интеграции с CMMS/ERP; – поддержка внешних модулей (например, Voith, SKF) через партнерскую экосистему 	<ul style="list-style-type: none"> – 5000 дверей поездов Thameslink на обслуживании через Railigent (автоматический мониторинг с ML); – высокоскоростные поезда Velaro (контроль редукторов, букс, движков); – снижение отказов благодаря прогнозам; – DB Cargo: оснащение 2000 локомотивов телеметрией TechLok с передачей ~7000 сигналов, аналитика через GE RailConnect и Railigent
Bentley OpenRail & iTwin	Цифровые двойники инфраструктуры и управление жизненным циклом железнодорожных объектов	<ul style="list-style-type: none"> – BIM-система и ГИС-платформа: объединение инженерных 3D-/4D-моделей (проектов) с данными датчиков и эксплуатационными сведениями; – Federated data — связь более чем с 10 источниками вместо единой БД: ProjectWise CDE, ContextCapture для 3D-реальности, iTwin для облачной консолидации; – поддержка открытых стандартов (IFC Rail) для совместимости; – API для интеграции с корпоративными системами 	<ul style="list-style-type: none"> – Network Rail: проект Transpennine — консолидация данных 60 систем в единый цифровой двойник (100 км линии, 600+ GIS-слоев), что сэкономило ~1 млн фунтов за счет ускорения доступа к данным; – Bristol Signal Renewal: виртуальная прокатка сценариев размещения сигналов на 3D-модели пути; – обнаружение ошибок до стройки; – железные дороги Италии: интеграция Bentley AssetWise с беспроводными датчиками на мостах — предиктивный SHM вместо плановых осмотров
Dassault 3DEXPERIENCE (Virtual Twin)	«Система систем»: сквозное планирование, моделирование и эксплуатация всей железнодорожной инфраструктуры	<ul style="list-style-type: none"> – единая платформа: объединение людей, данных, процессов в одном пространстве (3D-модели (CATIA), симуляция процессов (DELMIA), анализ (SIMULIA)); – концепция Virtual Twin: модель учитывает окружение и всю инфраструктуру, позволяет имитировать любое изменение и видеть его влияние на инфраструктуру 	<ul style="list-style-type: none"> – Alstom: применение 3DEXPERIENCE для виртуального двойника производства поездов — оптимизация технологий, снижение брака, синхронизация поставок в реальном времени; – KiwiRail: цифровая платформа для планирования грузовых перевозок — улучшение планирования флота и расписаний на основе единой модели; – пилот пассажирской железной дороги: моделирование сценариев размещения новой станции в виртуальном городе — отработка десятков вариантов для информированного решения

Платформа/решение	Основное направление применения	Архитектурные особенности и компоненты	Примеры внедрения/использования
Wabtec/GE RailConnect 360	Asset Performance Management для локомотивов и грузовых вагонов	<ul style="list-style-type: none"> – потомок платформы GE Predix на железной дороге: сбор данных о состоянии локомотивов в облако, аналитика по отказам; – интеграция с оперативными данными (рейсы, профили путей) для оптимизации эксплуатации, например рекомендации по экономичному ведению (Trip Optimizer); – модуль Fleet: панель состояния парка, приоритизация ремонтов по состоянию 	<ul style="list-style-type: none"> – применяется крупными грузовыми операторами в США (Union Pacific, BNSF и др.) для перехода к техобслуживанию по состоянию; – DB Cargo: тестирование RailConnect 360 параллельно с Railigent, получая преимущества от обеих систем (Railigent глубже по компонентам, GE — по топливной эффективности и надежности локомотивов); – Kazakhstan Temir Zholy: проект с Wabtec в рамках контракта 2023 года по оснащению локомотивов и депо системой Predix APM для снижения простоев
Отечественные решения	<ul style="list-style-type: none"> – предиктивная диагностика локомотивов; – цифровые модели инфраструктуры 	<ul style="list-style-type: none"> – собственные платформы: «Умный локомотив» — IoT + AI для оценки состояния узлов, аномалий, прогноз отказов локомотивов; – облачные решения: платформа ОКТА — централизованное управление диагностическими устройствами депо, дистанционная настройка и сбор данных; – BIM-/ГИС-инициативы: масштабное сканирование инфраструктуры, создание 3D-моделей, связь с системами Главгосэкспертизы и Ростройконтроля 	<ul style="list-style-type: none"> – РЖД: пилот по цифровому двойнику Восточного полигона — интеграция данных о пропускной способности, состояниях инфраструктуры для оптимизации графика; – локомотивное депо: 10 депо подключены к облаку — каждое изменение методики ремонта мгновенно транслируется на все, что дает экономию ~5 млн руб. на командировках специалистов при масштабировании на 100 депо; – мосты и тоннели: внедрение датчиков и моделей на мостовые сооружения (ожидается продление ресурса инженерных сооружений на 15–20 % при том же уровне безопасности за счет раннего обнаружения дефектов)
Китай (CRRC, CASCO)	<ul style="list-style-type: none"> – интеллектуальные поезда; – высокоскоростная сеть; – автоматизация управления движением 	<ul style="list-style-type: none"> – IoT-насыщенность: в новых поездах тысячи датчиков, 5G-модули — поток данных обрабатывается бортовым Edge и передается в облако. Встроенные в борт алгоритмы AI для безопасности (ADAS); – платформы Big Data: национальная сеть центров данных железных дорог, анализ терабайтов данных в реальном времени. Использование технологий Baidu/Huawei для ML и Computer Vision на видео с камер инфраструктуры; – интеграция с диспетчеризацией: единые централизованные системы, моделирующие движение поездов по всей стране с учетом состояния каждого элемента инфраструктуры; – оптимизация энергоэффективности и пропускной способности посредством цифрового двойника всей сети 	<ul style="list-style-type: none"> – CRRC CINOVA2.0: первый intercity-поезд с цифровым двойником — более 2000 датчиков, 130 алгоритмов — полный контроль над состоянием в реальном времени, повышение надежности и возможность автономного вождения; – ВСМ Пекин — Чжанцзякоу: пилот 5G + DT — вся линия оцифрована, цифровой двойник обеспечивает автопилотирование поездов на участках (GoA3) и дистанционный контроль тоннелей (с 3D-моделями); – метро Шанхая: система управления движением с AI-диспетчером, использующим цифровой двойник линий — увеличение интервалов движения на 10 % без новых путей, снижение сбоев на 15 % (данные CASCO, 2021)

Сравнение ключевых решений по цифровым двойникам

Выше приведена сводная таблица, сравнивающая ключевые платформы цифровых двойников и их архитектурные особенности.

Приведенные в таблице примеры — лишь небольшая часть реализованных проектов. Рассмотренные проекты показывают потенциал технологий. В действительности платформы часто пересекаются: например, инфраструктурный

цифровой двойник может потребовать интеграции с данными подвижного состава для комплексного анализа. Поэтому границы между решениями условны — идет конвергенция к интегрированным экосистемам. Как отмечают эксперты, цифровой двойник железной дороги — это не отдельное ПО, а совокупность взаимодействующих систем, объединенных архитектурно для достижения общей цели: повышения эффективности, безопасности и экономичности железнодорожных перевозок в реальном времени.

Заключение

В статье проведен комплексный анализ архитектуры цифровых двойников железнодорожного транспорта и рассмотрены ключевые аппаратно-программные компоненты, обеспечивающие их функционирование. Показано, что эффективная реализация цифрового двойника возможна только при наличии сквозной интеграции всех уровней — от датчиков и телеметрии до аналитических моделей и систем управления.

Основной вывод заключается в том, что цифровой двойник представляет собой не отдельную информационную систему, а интеграционную платформу, объединяющую разнородные источники данных, математические модели и прикладные сервисы в рамках единой архитектуры. Такой подход обеспечивает переход от реактивного к предиктивному управлению и позволяет реализовать риск-ориентированные стратегии [29] обслуживания технических систем.

Сравнительный анализ зарубежных и отечественных решений показал, что мировые плат-

формы ориентированы на экосистемный подход и глубокую интеграцию с BIM, PLM и IoT, тогда как российские разработки находятся на этапе активного формирования, но обладают потенциалом за счет гибкости, открытости архитектуры и адаптации к отраслевой специфике.

Отдельно установлено, что ключевыми факторами эффективности цифровых двойников являются:

- качество и полнота исходных данных;
- уровень интеграции с существующими системами;
- применение методов предиктивной аналитики и искусственного интеллекта;
- масштабируемость и модульность архитектуры.

К ограничениям текущего этапа развития следует отнести высокую сложность внедрения, необходимость стандартизации данных и значительные требования к вычислительной инфраструктуре.

Перспективные направления развития включают:

- интеграцию цифровых двойников с методами интеллектуального управления (в том числе на основе POMDP [29]);
- развитие edge-аналитики и распределенных вычислений;
- формирование отраслевых стандартов данных и интерфейсов.

Таким образом, цифровые двойники являются базовой технологией формирования интеллектуальной железнодорожной инфраструктуры и создают предпосылки для повышения надежности, безопасности и экономической эффективности перевозочного процесса в условиях цифровой трансформации отрасли.

СПИСОК ИСТОЧНИКОВ

1. Цифровые двойники // РЖД Цифровой. 03.12.2022. URL: <http://rzdigital.ru/technology/tsifrovye-dvoyniki/> (дата обращения: 12.02.2026).
2. Прохоров А. Н., Лысачев М. Н. Цифровой двойник: анализ, тренды, мировой опыт: корпоративное издание / науч. ред. А. И. Боровков. 1-е изд., испр. и доп. М.: АльянсПринт, 2020. 400 с.
3. Зуев Д. В., Бочкарев С. В. Цифровой двойник инфраструктуры железнодорожной автоматике и телемеханики ОАО «РЖД» // Транспорт Российской Федерации. 2022. № 3 (100). С. 19–22.
4. Преимущества технологии цифрового двойника инфраструктуры / С. С. Юханов, Д. В. Зуев, С. В. Бочкарев, А. А. Федоров // Автоматика, связь, информатика. 2021. № 4. С. 25–27. DOI: 10.34649/AT.2021.4.4.002

5. Об утверждении временной концепции технологии цифрового двойника инфраструктурного комплекса ОАО «РЖД»: распоряжение ОАО «РЖД» от 09.06.2021 № 1281/р.
6. Колесников К. Цифровые двойники и их применение в железнодорожной отрасли // *vc.ru*. 24.08.2020. URL: <http://vc.ru/transport/152429-cifrovye-dvoyniki-i-ih-primenenie-v-zheleznodorozhnoy-otrasli> (дата обращения: 12.02.2026).
7. Ефанов Д. В., Лыков А. А. Основы построения и принципы функционирования систем технического диагностирования и мониторинга устройств железнодорожной автоматики и телемеханики: учебное пособие. СПб.: ПГУПС, 2012. 59 с.
8. От «недиагностируемого утюга» к цифровому двойнику // Хабр. 23.09.2025. URL: <http://habr.com/ru/articles/949184> (дата обращения: 12.02.2026).
9. Barrow K. Bridging the Big Data Gap // *Railway Age*. 11.07.2018. URL: <http://www.railwayage.com/news/bridging-the-big-data-gap/> (дата обращения: 11.01.2026).
10. Rail Digital Twin and Deep Learning for Passenger Flow Prediction Using Mobile Data / Y. Ou [et al.] // *Electronics*. 2025. Vol. 14, iss. 12. Art. 2359. 16 p. DOI: 10.3390/electronics14122359
11. Bentley Systems. Digital Twins Drive Better Outcomes Across Rail Lifecycle. 2022. 14 p. URL: <http://www.bentley.com/wp-content/uploads/ebook-digital-twins-rail-asset-lifecycle-en.pdf> (дата обращения: 12.02.2026).
12. Siemens Railigent X Health States: Optimized Maintenance Decisions Through AI // *Railway USA*. 23.05.2024. URL: <http://railway-usa.com/news/81317-railigent-x-health-states-optimized-maintenance-decisions-through-ai> (дата обращения: 12.02.2026).
13. Luczak M. Dassault Systèmes Spotlights ‘Virtual Twin’ Tech at Railway Interchange 2025 // *Railway Age*. 21.05.2025. URL: <http://www.railwayage.com/news/dassault-systemes-spotlights-virtual-twin-tech-at-railway-interchange-2025/> (дата обращения: 12.02.2026).
14. Zhang Z.-Y., Shang D., Su. S. Digital Twin in Railway Industry: A Bibliometric Analysis and Systematic Review // *Digital Twin*. 2026. Vol. 3, iss. 1. Art. no. 2533858. 25 p. DOI: 10.1080/27525783.2025.2533858
15. Mobility Software Suite X // Siemens Mobility. URL: <http://www.mobility.siemens.com/us/en/portfolio/digital-solutions-software/digital-services/railigent-x.html> (дата обращения: 12.02.2026).
16. How Cognitive Digital Twins Enable Predictive Rail Management // The AnyLogic Company’s Blog. 01.07.2025. URL: <http://medium.com/@anylogic/how-cognitive-digital-twins-enable-predictive-rail-management-7c90cc963067> (дата обращения: 14.11.2025).
17. Chen M. Unleashing the Benefits of Virtual Twins in Rail // Dassault Systèmes Blog. 03.07.2023. URL: <http://blog.3ds.com/industries/business-services/unleashing-the-benefits-of-virtual-twins-in-rail> (дата обращения: 12.02.2026).
18. Siemens расширяет цифровое депо для региональных поездов в Дортмунде // *Железные дороги мира*. 08.05.2024. URL: <http://zdmira.com/news/siemens-rasshiryet-tsifrovoe-depo-dlya-regionalnykh-poezdov-v-dortmunde> (дата обращения: 12.02.2026).
19. Digital Twin Railway: The Future of Railway Management // BibLus — News and insights on AEC industry and BIM. 06.12.2024. URL: <http://biblus.accasoftware.com/en/digital-twin-railway-the-future-of-railway-management> (дата обращения: 12.02.2026).
20. Зуев Д. В., Бочкарев С. В. Цифровой двойник инфраструктуры ОАО «РЖД» // *Автоматика, связь, информатика*. 2020. № 11. С. 11–14. DOI: 10.34649/AT.2020.11.11.001
21. Ядро ЕКП ТД. Цифровой двойник ОАО «РЖД». Цифровое моделирование инфраструктуры Трансэнерго (Сервис ЦД-ТЭ) (ЕКП ТД. Сервис ЦД-ТЭ): свидетельство о государственной регистрации программы для ЭВМ № 2023686974 Российская Федерация. № 2023686296 / Д. В. Зуев [и др.]; заявл. 30.11.2023; опубл. 11.12.2023; заявитель ОАО «Российские железные дороги». 1 с.
22. Ядро ЕКП ТД. Цифровой двойник ОАО «РЖД». Цифровое моделирование активов Центральной дирекции инфраструктуры (ЦДИ) (Сервис ЦД-ЦДИ) (ЕКП ТД. Сервис ЦД-ЦДИ): свидетельство о государственной

регистрации программы для ЭВМ № 2023660018 Российская Федерация. № 2023619409 / Д. В. Зуев [и др.]; заявл. 12.05.2023; опубл. 17.05.2023; заявитель ОАО «Российские железные дороги». 1 с.

23. Ядро ЕКП ТД. Цифровой двойник инфраструктуры хозяйства автоматики и телемеханики. Сервис работы с BIM моделью в цифровом двойнике (Сервис ЦДШ-BIM) (ЕКП ТД Сервис ЦДШ-BIM): свидетельство о государственной регистрации программы для ЭВМ № 2022667187 Российская Федерация. № 2022666465 / Д. В. Зуев [и др.]; заявл. 07.09.2022; опубл. 15.09.2022; заявитель ОАО «Российские железные дороги». 1 с.

24. Бочкарев С. В. Методика оценки и анализа числа срабатываний релейной аппаратуры железнодорожной автоматики и телемеханики // Информационные технологии. 2023. Т. 29, № 1. С. 51–56. DOI: 10.17587/it.29.51-56

25. Бочкарев С. В. Расчет сроков замены релейной аппаратуры железнодорожной автоматики и телемеханики // Информационные технологии. 2023. Т. 29, № 8. С. 437–444. DOI: 10.17587/it.29.437-444

26. Зуев Д. В., Бочкарев С. В., Федоров А. А. Методология формирования исполненной кабельной трассы на двухниточном плане станции // Автоматика, связь, информатика. 2022. № 5. С. 21–24. DOI: 10.34649/AT.2022.5.5.001

27. CRRC's New Smart Intercity Train CINOVA2.0 Was Officially Launched // CRRC Group Official Website. 19.05.2023. URL: http://www.crrecg.cc/en/2023-05/30/article_29BA86E691944B61A73EB514217EE0D9.html (дата обращения: 12.02.2026).

28. К 2035 году будет создан цифровой двойник железнодорожной инфраструктуры Люксембурга // Цифровое Строительство. 27.12.2024. URL: <http://digital-build.ru/news/k-2035-godu-budet-sozdan-czifrovoj-dvojn timer-zheleznodorozhnoj-infrastruktury-lyuksemburga> (дата обращения: 12.02.2026).

29. Бочкарев С. В., Бубнов В. П., Зуев Д. В. Применение метода риск-ориентированной диагностики для технических средств ЖАТ // Автоматика, связь, информатика. 2026. № 3. С. 23–26. DOI: 10.62994/AT.2026.3.3.004

Дата поступления: 28.04.2026

Решение о публикации: 25.05.2026

Digital Twin Architecture for Rail Transport

Denis V. Zuev

— PhD in Engineering, General Director. Research interests: neural networks, digital twins.
E-mail: zuevdv@gmail.com

Sergey V. Bochkarev

— PhD in Engineering, Technical Director. Research interests: neural networks, digital twins.
E-mail: bochkareffsv@yandex.ru

Sintez AT LLC, 28/29 Sredny ave., Vasilievsky Island, Saint Petersburg, 199004, Russia

For citation: Zuev D. V., Bochkarev S. V. Digital Twin Architecture for Rail Transport, *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 17–35. DOI: 10.20295/2413-2527-2026-246-17-35 (In Russian)

Abstract. *An in-depth system analysis of architectural approaches to the construction and implementation of digital twins in railway transport is presented, which is a key stage in the digital transformation of the industry. **Purpose:** to systematize architectural approaches to the creation of digital twins, analyze key technological solutions and develop recommendations for their integration into a single digital ecosystem of infrastructure and rolling stock management. **Methods:** a comparative study of the world's leading platforms (Siemens Railigent X, Bentley iTwin, Dassault 3DEXPERIENCE) and domestic developments was conducted,*

which revealed the limitations of existing solutions in terms of adaptation to the specifics of railway automation and telemechanics. **Results:** unlike traditional fragmented information systems, the concept of a “single source of truth” is proposed, based on the seamless integration of BIM, PLM and SCADA technologies, which allows for end-to-end lifecycle management of transport assets. The scientific novelty of the research lies in the development and detailing of a five-level hierarchical model of the digital twin, which combines the level of physical objects and sensors, data transmission channels, a storage platform and digital models, an analytical layer and visualization interfaces. **Practical significance:** due to the description of the mechanisms of transition from planned preventive maintenance to predictive and risk-oriented management. It is shown that the implementation of the proposed architecture makes it possible to predict critical malfunctions (for example, defects in axle boxes) several months before their occurrence and to conduct simulation of various operational scenarios. The research results can be used in the design of intelligent infrastructure and rolling stock management systems to improve the safety and economic efficiency of the transportation process.

Keywords: digital twin, railway transport, multilayer architecture, predictive analytics, industrial internet of things, BIM technologies, product lifecycle management, condition monitoring, intelligent transport systems, risk-based approach

REFERENCES

1. Tsifrovye dvoyniki [Digital Twins], *RZhD Tsifrovoy [RZD.Digital]*. Published online at December 03, 2022. Available at: <http://rzddigital.ru/technology/tsifrovye-dvoyniki/> (accessed: February 12, 2026). (In Russian)
2. Prokhorov A. N., Lysachev M. N. Tsifrovoy dvoynik: analiz, trendy, mirovoy opyt: korporativnoe izdanie [Digital Twin: Analysis, Trends, and Global Experience: Corporate Edition]. Moscow, AlyansPrint Publishing House, 2020, 400 p. (In Russian)
3. Zuev D. V., Bochkarev S. V. Tsifrovoy dvoynik infrastruktury zheleznodorozhnoy avtomatiki i telemekhaniki OAO “RZhD” [Digital Twin of Railway Automation and Telemechanics Infrastructure of JSC RZD], *Transport Rossiyskoy Federatsii [Transport of the Russian Federation]*, 2022, no. 3 (100), pp. 19–22. (In Russian)
4. Yukhanov S. S., et al. Preimushchestva tekhnologii tsifrovogo dvoynika infrastruktury [Benefits of Infrastructure Digital Twin Technology], *Avtomatika, svyaz, informatika [Automation, Communications, Informatics]*, 2021, no. 4, pp. 25–27. DOI: 10.34649/AT.2021.4.4.002 (In Russian)
5. Ob utverzhdenii vremennoy kontseptsii tekhnologii tsifrovogo dvoynika infrastruktornogo kompleksa OAO “RZhD” [On the Approval of the Temporary Concept of the Digital Twin Technology of the Infrastructure Complex of JSC “Russian Railways”]: Order of JSC “Russian Railways” dated June 09, 2021 no. 1281/r. (In Russian)
6. Kolesnikov K. Tsifrovye dvoyniki i ikh primenenie v zheleznodorozhnoy otrasli [Digital Twins and Their Application in the Railway Industry], *vc.ru*. Published online at August 24, 2020. Available at: <http://vc.ru/transport/152429-cifrovye-dvoyniki-i-ih-primenenie-v-zheleznodorozhnoy-otrasli> (accessed: February 12, 2026). (In Russian)
7. Efanov D. V., Lykov A. A. Osnovy postroeniya i printsipy funktsionirovaniya sistem tekhnicheskogo diagnostirovaniya i monitoringa ustroystv zheleznodorozhnoy avtomatiki i telemekhaniki: uchebnoe posobie [Fundamentals of Construction and Principles of Functioning of Technical Diagnostics and Monitoring Systems for Railway Automation and Telemechanics Devices: A Textbook]. Saint Petersburg, PSTU, 2012, 59 p. (In Russian)
8. Ot “nediagnostiruemogo utyuga” k tsifrovomu dvoyniku [From an “Undiagnosed Iron” to a Digital Double], *Khabr [Habr]*. Published online at September 23, 2025. Available at: <http://habr.com/ru/articles/949184> (accessed: February 12, 2026). (In Russian)
9. Barrow K. Bridging the Big Data Gap, *Railway Age*. Published online at July 11, 2018. Available at: <http://www.railwayage.com/news/bridging-the-big-data-gap/> (accessed: January 11, 2026).
10. Ou Y., et al. Rail Digital Twin and Deep Learning for Passenger Flow Prediction Using Mobile Data, *Electronics*, 2025, vol. 14, iss. 12, art. no. 2359, 16 p. DOI: 10.3390/electronics14122359

11. Bentley Systems. Digital Twins Drive Better Outcomes Across Rail Lifecycle. 2022. 14 p. Available at: <http://www.bentley.com/wp-content/uploads/ebook-digital-twins-rail-asset-lifecycle-en.pdf> (accessed: February 12, 2026).
12. Siemens Railigent X Health States: Optimized Maintenance Decisions Through AI, *Railway USA*. Published online at May 23, 2024. Available at: <http://railway-usa.com/news/81317-railigent-x-health-states-optimized-maintenance-decisions-through-ai> (accessed: February 12, 2026).
13. Luczak M. Dassault Systèmes Spotlights ‘Virtual Twin’ Tech at Railway Interchange 2025, *Railway Age*. Published online at May 21, 2025. Available at: <http://www.railwayage.com/news/dassault-systemes-spotlights-virtual-twin-tech-at-railway-interchange-2025/> (accessed: February 12, 2026).
14. Zhang Z.-Y., Shang D., Su. S. Digital Twin in Railway Industry: A Bibliometric Analysis and Systematic Review, *Digital Twin*, 2026, vol. 3, iss. 1, art. no. 2533858, 25 p. DOI: 10.1080/27525783.2025.2533858
15. Mobility Software Suite X, *Siemens Mobility*. Available at: <http://www.mobility.siemens.com/us/en/portfolio/digital-solutions-software/digital-services/railigent-x.html> (accessed: February 12, 2026).
16. How Cognitive Digital Twins Enable Predictive Rail Management, *The AnyLogic Company’s Blog*. Published online at July 01, 2025. Available at: <http://medium.com/@anylogic/how-cognitive-digital-twins-enable-predictive-rail-management-7c90cc963067> (accessed: November 14, 2025).
17. Chen M. Unleashing the Benefits of Virtual Twins in Rail, *Dassault Systèmes Blog*. Published online at July 03, 2023. Available at: <http://blog.3ds.com/industries/business-services/unleashing-the-benefits-of-virtual-twins-in-rail> (accessed: February 12, 2026).
18. Siemens rasshiryayet tsifrovoe depo dlya regionalnykh poezdov v Dortmunde [Siemens Expands Digital Depot for Regional Trains in Dortmund], *Zheleznye dorogi mira*. Published online at May 08, 2024. Available at: <http://zdmira.com/news/siemens-rasshiryayet-tsifrovoe-depo-dlya-regionalnykh-poezdov-v-dortmunde> (accessed: February 12, 2026). (In Russian)
19. Digital Twin Railway: The Future of Railway Management, *BibLus — News and insights on AEC industry and BIM*. Published online at December 06, 2024. Available at: <http://biblus.accasoftware.com/en/digital-twin-railway-the-future-of-railway-management> (accessed: February 12, 2026).
20. Zuev D. V., Bochkarev S. V. Tsifrovoy dvoynik infrastruktury OAO “RZhD” [The Digital Twin of the Russian Railways Infrastructure], *Avtomatika, svyaz, informatika [Automation, Communications, Informatics]*, 2020, no. 11, pp. 11–14. DOI: 10.34649/AT.2020.11.11.001 (In Russian)
21. Zuev D. V., et al. Yadro EKP TD. Tsifrovoy dvoynik OAO “RZhD”. Tsifrovoe modelirovanie infrastruktury Transenergo (Servis TsD-TE) (EKP TD. Servis TsD-TE) [EKP TD Core. Digital Twin of JSC “Russian Railways”. Digital Modeling of TransEnergo Infrastructure (CD-TE Service)]. Certificate of State registration of a computer program RU no. 2023686974. Published at December 11, 2023. 1 p. (In Russian)
22. Zuev D. V., et al. Yadro EKP TD. Tsifrovoy dvoynik OAO “RZhD”. Tsifrovoe modelirovanie aktivov Tsentralnoy direktzii infrastruktury (TsDI) (Servis TsD-TsDI) (EKP TD. Servis TsD-TsDI) [EKP TD Core. Digital Twin of JSC “Russian Railways”. Digital Modeling of Infrastructure Assets (CD-CDI Service)]. Certificate of State registration of a computer program RU no. 2023660018. Published at May 17, 2023. 1 p. (In Russian)
23. Zuev D. V., et al. EKP TD. Tsifrovoy dvoynik infrastruktury khozyaystva avtomatiki i telemekhaniki. Servis raboty s BIM modelu v tsifrovom dvoynike (Servis TsDSh-BIM) (EKP TD Servis TsDSh-BIM) [Digital Twin of Railway Automation Infrastructure. BIM Model Integration Service (CDSh-BIM)]. Certificate of State registration of a computer program RU no. 2022667187. Published at September 15, 2022. 1 p. (In Russian)
24. Bochkarev S. V. Metodika otsenki i analiza chisla srabatyvaniy releynoy apparatury zheleznodorozhnoy avtomatiki i telemekhaniki [Methodology of Estimation and Analysis of the Number of Actuations of Relay Equipment of Railway Automation and Telemechanic], *Informatsionnye tekhnologii [Information Technologies]*, 2023, vol. 29, no. 1, pp. 51–56. DOI: 10.17587/it.29.51-56 (In Russian)

25. Bochkarev S. V. Raschet srokov zameny releynoy apparatury zheleznodorozhnoy avtomatiki i telemekhaniki [Calculation of the Timing of the Replacement of Relay Equipment for Railway Automation and Telemechanic], *Informatsionnye tekhnologii [Information Technologies]*, 2023, vol. 29, no. 8, pp. 437–444. DOI: 10.17587/it.29.437-444 (In Russian)
26. Zuev D.V., Bochkarev S.V., Fedorov A.A. Metodologiya formirovaniya ispolnennoy kabelnoy trassy na dvukhnitochnom plane stantsii [Methodology of Forming a Completed Cable Route on a Two-Line Plan of the Station], *Avtomatika, svyaz, informatika [Automation, Communications, Informatics]*, 2022, no. 5, pp. 21–24. DOI: 10.34649/AT.2022.5.5.001 (In Russian)
27. CRRC’s New Smart Intercity Train CINOVA2.0 Was Officially Launched, *CRRC Group Official Website*. Published online at May 19, 2023. Available at: http://www.crrecg.cc/en/2023-05/30/article_29BA86E691944B61A73EB514217EE0D9.html (accessed: February 12, 2026).
28. K 2035 godu budet sozdan tsifrovoy dvoynik zheleznodorozhnoy infrastruktury Lyuksemburga [By 2035, A Digital Twin of Luxembourg's Railway Infrastructure will be Created], *Tsifrovoe Stroitelstvo [Digital Build]*. Published online at December 27, 2024. Available at: <http://digital-build.ru/news/k-2035-godu-budet-sozdan-czifrovoj-dvoynik-zheleznodorozhnoj-infrastruktury-lyuksemburga> (accessed: February 12, 2026). (In Russian)
29. Bochkarev S. V., Bubnov V. P., Zuev D. V. Primenenie metoda risk-orientirovannoy diagnostiki dlya tekhnicheskikh sredstv ZhAT [Application of the Risk-Oriented Diagnostics Method for Railway Automation Equipment], *Avtomatika, svyaz, informatika [Automation, Communications, Informatics]*, 2026, no. 3, pp. 23–26. DOI: 10.62994/AT.2026.3.3.004 (In Russian)

Received: April 28, 2026

Accepted: May 25, 2026

УДК 004.738.5+004.43

Преимущества использования шаблонизаторов Django для бизнеса: баланс между эффективностью и стоимостью разработки

- Забродин Андрей Владимирович** — кандидат ист. наук, доцент кафедры «Информационные и вычислительные системы». Научные интересы: информационные системы, аналитика данных, парадигмы программирования, облачные технологии. E-mail: zabrodin@pgups.ru
- Михеев Андрей Дмитриевич** — студент 2-го курса бакалавриата направления 09.03.01 «Информатика и вычислительная техника». Научные интересы: автоматизация и клиентская лояльность, веб-разработка. E-mail: miheev.31.andrey@gmail.com
- Граур Максим Александрович** — студент 2-го курса бакалавриата направления 09.03.01 «Информатика и вычислительная техника». Научные интересы: автоматизация и клиентская лояльность, веб-разработка. E-mail: maksimgraur22@mail.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Забродин А. В., Михеев А. Д., Граур М. А. Преимущества использования шаблонизаторов Django для бизнеса: баланс между эффективностью и стоимостью разработки // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 36–45. DOI: 10.20295/2413-2527-2026-246-36-45

Аннотация. Рассматривается проблема выбора технологического стека для веб-проектов малого и среднего бизнеса. **Цель:** на фоне распространения архитектур с разделением фронтенда и бэкенда проанализировать целесообразность использования встроенного шаблонизатора Django как более простого решения для проектов с ограниченным бюджетом, небольшой командой и умеренными требованиями к интерактивности страниц сайта. **Результаты:** показано, что серверный рендеринг в сочетании с запросами одностраничного приложения позволяет сократить архитектурную сложность, снизить затраты на разработку и поддержку приложения, сохранить преимущества SEO и использовать встроенные механизмы безопасности Django. Сделан вывод о том, что для ряда прикладных бизнес-задач Django-шаблоны могут выступать как рациональный архитектурный выбор. **Практическая значимость:** результаты исследования могут быть использованы при проектировании систем массового сбора данных и взаимодействия с ними, построении распределенных поисковых роботов и создании платформ анализа информации на основе многоагентных систем. **Обсуждение:** представленные наблюдения отражают практическую структуру функционирования распределенного поискового робота. В отличие от классических поисковиков предложенный агент-ориентированный подход позволяет перераспределять нагрузку, динамически изменять стратегии обхода при обнаружении блокировок.

Ключевые слова: Django, шаблонизатор, сравнительный анализ архитектур, веб-разработка, наблюдаемость, AJAX, малый бизнес

2.3.5 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки)

Введение

В последние годы в веб-разработке стала стандартом практика разделения клиентской и серверной частей приложения [1]. Фронтенд обычно создают с помощью JavaScript-фреймворков и библиотек React, Vue или Angular. Бэкенд реализуют на отдельных платформах и обеспечивают взаимодействие с клиентом через API. Этот подход оптимален для крупных нагруженных систем, где нужны сложная бизнес-логика и независимое масштабирование компонентов. Однако в проектах малого и среднего бизнеса такую архитектуру веб-приложения используют гораздо чаще, чем требуется [2]. Для типовых задач — корпоративных сайтов, каталогов, личных кабинетов, сервисов записи, небольших интернет-магазинов — важнее другие критерии: скорость разработки, ограниченный бюджет, небольшой размер команды, хорошая поисковая индексация и предсказуемость сопровождения. Разделение фронтенда и бэкенда в таких случаях может провоцировать избыточную сложность: нужно поддерживать две отдельные кодовые базы, требуется проектировать и поддерживать API, необходимо синхронизировать изменения между обеими частями приложения.

Более простой вариант — использовать встроенный шаблонизатор Django [3] для создания веб-приложения с единой серверной логикой. При этом, находясь на одной и той же странице, можно сохранить интерактивность интерфейса при помощи динамического изменения контента посредством AJAX-запросов. Такой подход позволяет балансировать между SPA (Single Page Application — одностраничное веб-приложение) и SSR (Server-Side Rendering — серверный рендеринг, позволяющий командам предоставлять динамический контент), сохраняя простоту и скорость разработки.

Цель статьи — показать, когда использование Django-шаблонов становится обоснованной альтернативой разделенной архитектуре. Основное внимание уделено проектам с ограниченными ресурсами, где критически важны простота реализации, быстрая скорость запуска и удобство поддержки проекта. В рамках исследования проведен сравнительный анализ двух архитектур [4] по ключевым критериям: совокупная стоимость владения (total cost of ownership, TCO), скорость поисковой индексации, уровень встроенной безопасности и затраты на сопровождение.

Новизна исследования заключается в переосмыслении роли встроенного шаблонизатора Django. Вместо того чтобы считать его устаревшим инструментом, он рассматривается как рациональное архитектурное решение для определенного класса бизнес-ориентированных веб-проектов. Ключевое отличие данного подхода заключается в том, что с «технологической современности» фокус смещается на соответствие реальным потребностям бизнеса, включая стоимость сопровождения, требования SEO, уровень безопасности и простоту дальнейшего развития проекта.

Ключевое отличие данного подхода заключается в том, что с «технологической современности» фокус смещается на соответствие реальным потребностям бизнеса, включая стоимость сопровождения, требования SEO, уровень безопасности и простоту дальнейшего развития проекта.

Методология исследования

Django традиционно воспринимается как серверный фреймворк для быстрой разработки [5, 6], а архитектурные исследования акцентируют внимание на масштабируемых системах с API [7]. В проектах с ограниченными ресурсами остаются недостаточно изученными такие аспекты, как совокупная стоимость владения [8], эффективность поисковой индексации [9], организационная сложность сопровождения [10]. Статья восполняет этот пробел, анализируя условия, при которых Django-шаблоны становятся рациональным выбором для создания веб-приложения.

Сравнительный анализ архитектурных подходов

В рамках проведенного сравнительного анализа архитектурных паттернов, результаты которого представлены в табл. 1, отражены ключевые различия между подходами на основе Django-шаблонов и архитектурой с разделением фронтенда и бэкенда на базе React + API. Особое внимание уделено техническим и организационным аспектам реализации.

Критерии сравнения были подобраны на основе анализа профессиональной литературы [5, 9, 10] и практического опыта работы с проектами разного

Таблица 1

Компаративный анализ двух архитектурных паттернов

Параметр сравнения	Архитектура А: Django-шаблоны (SSR)	Архитектура Б: React + API (SPA)
Тип рендеринга	Серверный (SSR)	Клиентский (CSR), возможен гибридный (SSR через Next.js)
Количество кодовых баз	1 (Python)	2 (JavaScript + Python/Java/Go)
Способ передачи данных	Непосредственно в контексте шаблона	Через REST/GraphQL API
Требования к специалистам	1 Middle Python-разработчик	1 Middle React + 1 Middle Backend
Инструментарий отладки	Django Debug Toolbar, логи на сервере	React DevTools + Postman + логи API
SEO из коробки	Да (полный HTML)	Нет (требуется SSR-надстройка)

масштаба. При формировании системы оценки были выделены ключевые факторы: экономическая эффективность (анализ стоимости разработки, внедрения и сопровождения), простота поддержки, поисковая оптимизация, безопасность, скорость вывода на рынок, масштабируемость, кадровые требования.

Гипотеза исследования заключается в том, что для веб-проектов малого и среднего масштаба использование Django-шаблонов может обеспечить более выгодное соотношение цены и качества по сравнению с разделенной архитектурой фронтенда и бэкенда.

Для проверки гипотезы проведем формальное сравнение с использованием метрики TCO за 24 месяца:

$$TCO = DevCost + (TeamSize \times AvgSalary \times 24) + InfraCost + SupportCost,$$

где DevCost — стоимость начальной разработки (человеко-месяцы × ставка);

TeamSize — число разработчиков в команде поддержки;

AvgSalary — средняя зарплата [11];

InfraCost — стоимость хостинга, CDN, мониторинга;

SupportCost — стоимость исправления ошибок и доработок.

Расчет носит ориентировочный характер и используется для сравнительной оценки рассматриваемых подходов разработки, все параметры занесены в табл. 2. Значения ставок и инфраструктурных расходов приведены как усредненные показатели и могут изменяться в зависимости от региона, квалификации специалистов и масштаба проекта.

Проведенное сравнение показывает, что для рассматриваемого класса проектов общая стоимость владения архитектурой React + API может значительно превышать аналогичный показатель для решения на Django-шаблонах. Полученный результат позволяет предварительно подтвердить выдвинутую гипотезу, однако его не следует распространять на иные типы веб-приложений без учета масштаба проекта и состава команды.

Таблица 2

Исходные данные на примере интернет-магазина

Параметр	Django + шаблоны	React + API
Команда разработки, чел.	1 Middle	1 Middle React + 1 Middle Backend
Ставка, тыс. руб./мес.	200	200
Время разработки, мес.	3	5 (из-за интеграции API)
DevCost, тыс. руб.	600	2000
TeamSize, чел.	0,5 (частичная)	1,5 (1 Front + 0,5 Back)
InfraCost, тыс. руб./мес.	10	25 (дополнительные сборки, CDN)
TCO за 24 мес., тыс. руб.	$600 + (0,5 \cdot 200 \cdot 24) + (10 \cdot 24) = 3240$	$2000 + (1,5 \cdot 200 \cdot 24) + (25 \cdot 24) = 9800$

Влияние на SEO: количественная оценка

Проведено экспериментальное сравнение индексации двух идентичных сайтов — тестовых интернет-магазинов на 500 товарных страниц, реализованных на Django-шаблонах и React. Оба сайта имели одинаковую структуру каталога, набор товарных страниц и текстовое наполнение. Различие заключалось в способе формирования HTML-страницы. Сайты были одновременно поданы в Google Search Console и «Яндекс Вебмастер».

Результаты были сформированы через 14 дней после отправки сайтов на индексацию (табл. 3).

Способ рендеринга напрямую влияет на индексацию сайта поисковыми системами. При серверном рендеринге поисковый робот получает готовый HTML-документ, в то время как при клиентском рендеринге значительная часть содержания формируется после выполнения JavaScript. Общая логика различия между серверным и клиентским рендерингом представлена на рис. 1.

Таблица 3

Сравнение индексации сайтов

Показатель	Django (SSR)	React (CSR)
Проиндексировано страниц, количество	487 (97,4 %)	112 (22,4 %)
Среднее время до индексации, дней	1,2	5,8
Позиции в топ-50 по 10 ключевым запросам	7/10	2/10
Обнаружение новых товаров роботом, ч	< 1	> 48



Рис. 1. Схема сравнения серверного рендеринга на Django и клиентского рендеринга на React

Полученные данные позволяют предположить, что сайт на Django-шаблонах может раньше начать получать активный трафик за счет более быстрой индексации страниц.

Безопасность и наблюдаемость архитектурных решений

Для оценки особенностей сопровождения и сложности конфигурации был проведен обзор данных CVE (Common Vulnerabilities and Exposures — список известных уязвимостей и дефектов информационной безопасности) за 2023–2025 годы для Django, а также типичных компонентов разделенной SPA-архитектуры: React, React DOM и серверных API-фреймворков. Ана-

лиз носит сравнительно-иллюстративный характер (табл. 4).

В разделенной архитектуре проекта уязвимости могут возникать на разных уровнях системы: в настройках CORS, механизмах авторизации API, обработке CSRF-защиты, зависимостях клиентской части или серверного фреймворка. В Django часть стандартных механизмов защиты берется «из коробки», что снижает число разделенных точек контроля для небольшой команды разработки.

Различия в наблюдаемости были также проверены на примере преднамеренно внесенной ошибки, связанной с замедленным запросом к базе данных (рис. 2).

Таблица 4

Обзор и сравнение данных CVE

Компонент	Количество CVE (критических и высоких)	Исправлено за < 7 дней, %	Требует обновления двух частей
Django	12	100	Нет
React + React DOM	8	75	Да
Express / Spring	27	60	Да
Связка React + Backend	35	0	Да

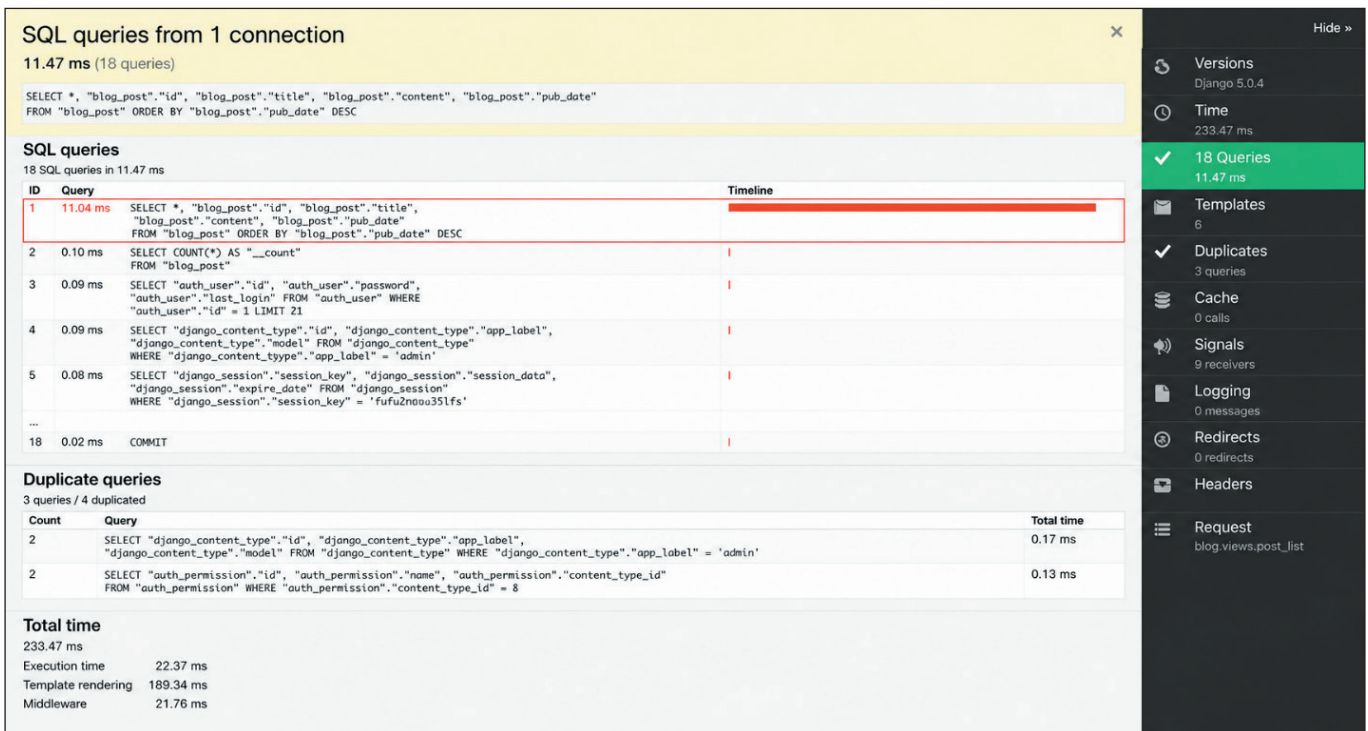


Рис. 2. Скриншот работы Django Debug Toolbar

В варианте Django + Debug Toolbar разработчик получает возможность непосредственно увидеть дублирующиеся SQL-запросы. В варианте React + API наблюдаемость аналогичной проблемы требует сопоставления данных клиентской части, сетевых запросов и серверных логов.

Реализация SPA-подобного взаимодействия средствами Django-шаблонов

Проведена экспериментальная реализация интернет-магазина с использованием Django 5.0 и встроенных шаблонов [12], асинхронных запросов, а также эндпоинтов (endpoint), возвращающих страницу с передачей только блока контента.

Результаты нагрузочного тестирования показаны в табл. 5 и на рис. 3–5. Общие рекомендации по оптимизации производительности шаблонов и снижению нагрузки на сервер подробно разбираются в [5].

Ключевые результаты эксперимента позволяют сделать следующие выводы: плавность переходов между страницами в представленном примере оказалась сопоставимой с вариантом на React, при этом SEO-показатели сохраняются на высоком уровне благодаря выполнению загрузки страницы средствами серверного рендеринга. Кроме того, единая кодовая база снижает риск дублирования маршрутизации,

Таблица 5

Результаты нагрузочного тестирования

Метрика	Django (полная перезагрузка)	Django (AJAX/SPA-подход)	React (CSR, без Next.js)
First Contentful Paint (FCP), с	0,90	0,90	1,80 (ожидание JS)
Time to Interactive (TTI), с	1,20	1,10	2,40
Скорость перехода между страницами, с	0,80 (перезагрузка)	0,12 (подмена блока)	0,10
SEO-готовность (Lighthouse SEO)	100	100	78

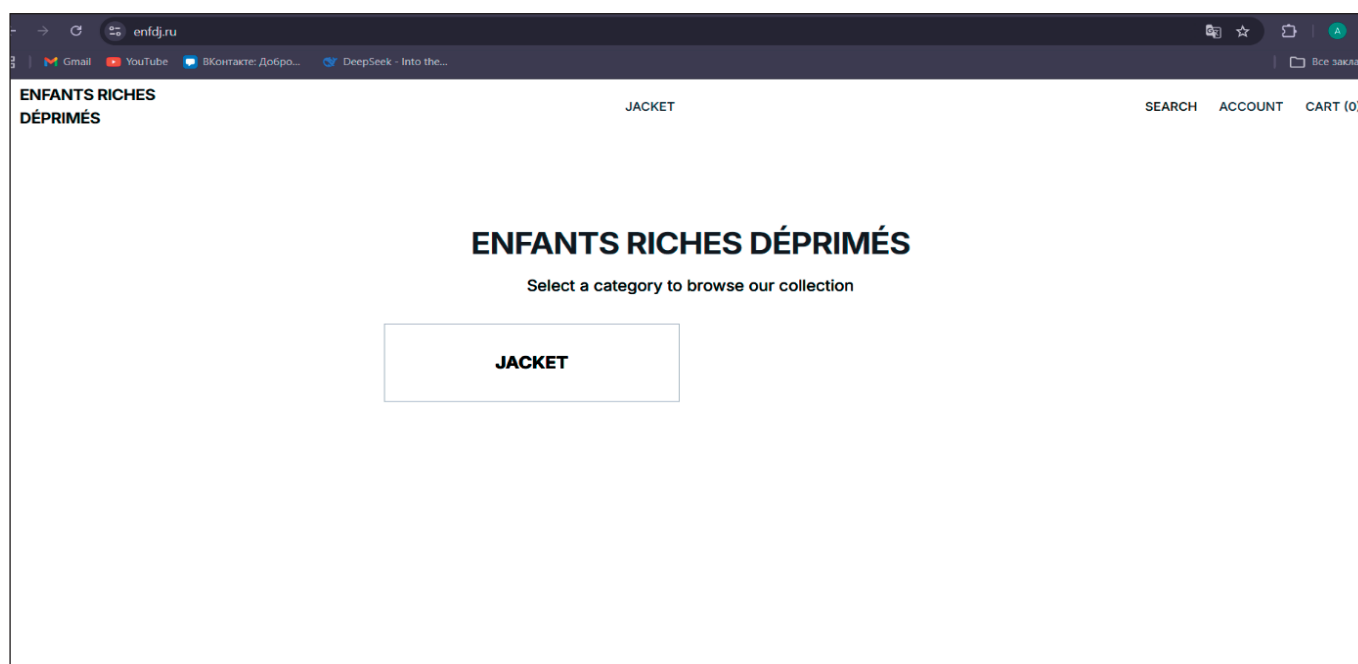


Рис. 3. Главная страница сайта — серверный рендеринг

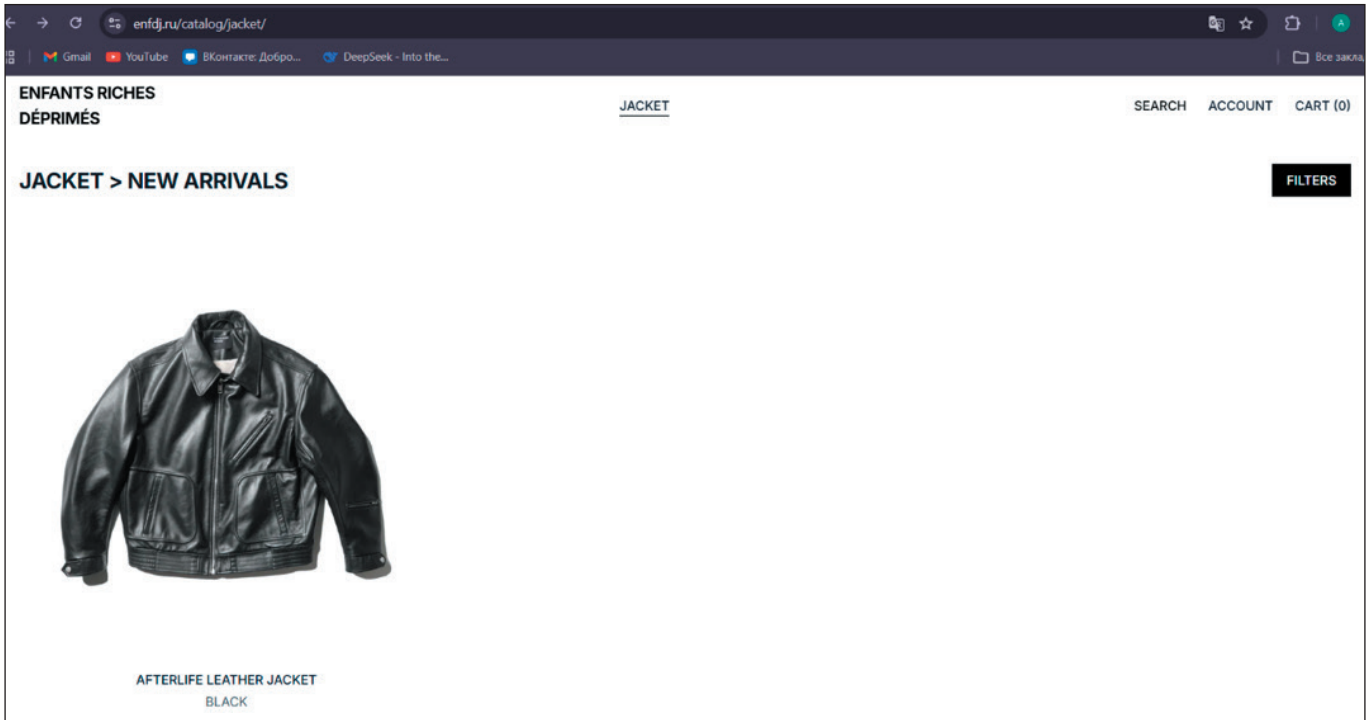


Рис. 4. Каталог — подгрузка через AJAX

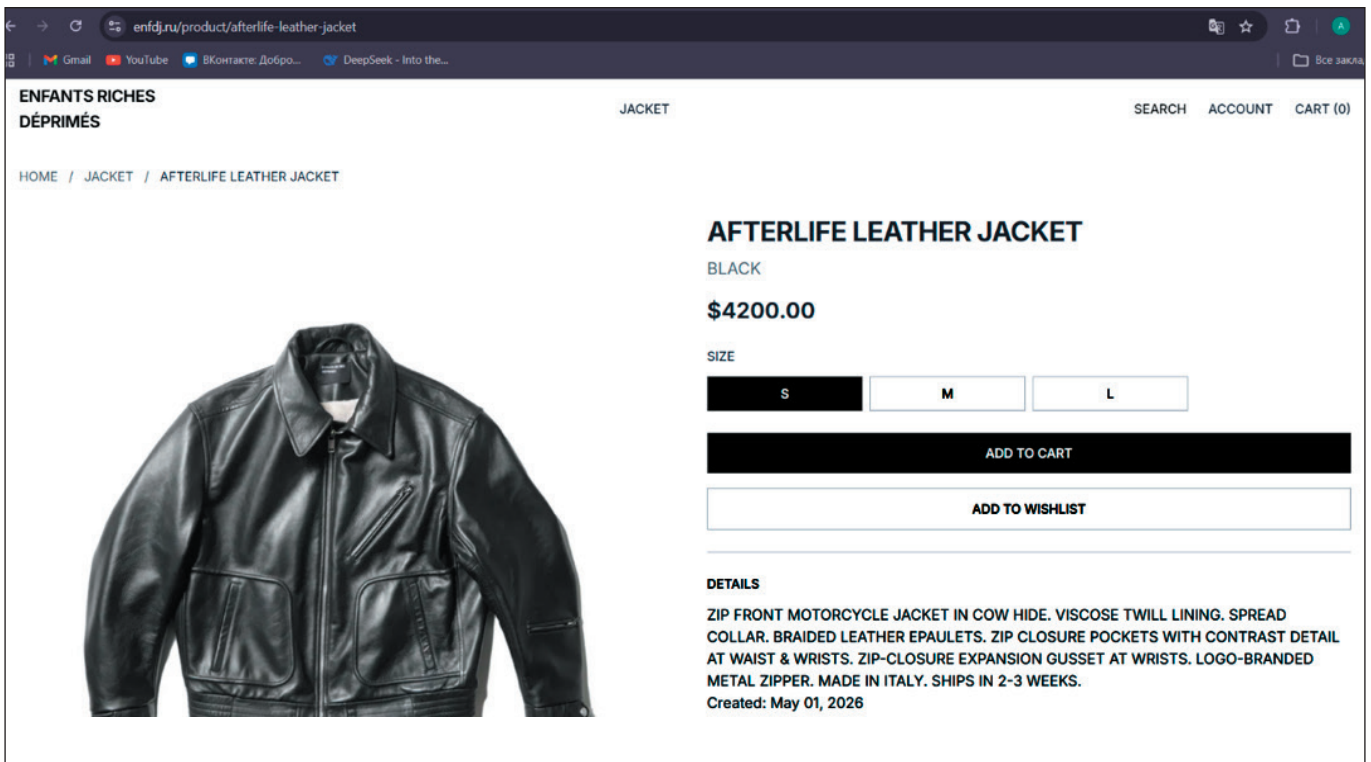


Рис. 5. Карточка товара — динамический контент с изменением URL

валидации и логики приложения, а реализация SPA-подобной навигации не потребовала выделения отдельного фронтенд-разработчика. На

основе проведенного сравнительного анализа сформулированы рекомендации по выбору архитектуры (табл. 6).

Рекомендации по выбору архитектуры зависимости от требований проекта

Условия проекта	Предпочтительный подход
Небольшая команда разработки и ограниченный бюджет	Django-шаблоны
Корпоративный сайт, каталог, сервис записи, MVP	Django-шаблоны
Высокие требования к SEO и скорости запуска	Django-шаблоны или гибридный SSR-подход
Сложная клиентская логика и большое число интерактивных состояний	React/Vue/Angular + API
Независимое масштабирование фронтенда и бэкенда	Разделенная архитектура

Анализ количественных и качественных характеристик — от совокупной стоимости владения и скорости индексации до безопасности и наблюдаемости — показывает, что выбор архитектуры не сводится к противопоставлению «современный vs устаревший».

Представленные в табл. 6 сценарии наглядно демонстрируют: для проектов с ограниченными ресурсами, типовыми бизнес-задачами и умеренными требованиями к интерактивности подход на основе Django-шаблонов обеспечивает лучший баланс между эффективностью разработки и долгосрочными затратами, тогда как разделенная архитектура оправдана лишь при наличии четких требований к сложной клиентской логике или необходимости независимого масштабирования компонентов.

Заключение

Проведенное исследование позволяет рассматривать встроенный шаблонизатор Django как практически значимый инструмент для веб-проектов малого и среднего бизнеса. При ограниченном бюджете,

небольшой команде разработки и умеренных требованиях к интерактивности серверный рендеринг снижает архитектурную сложность, уменьшает объем работ и упрощает последующую поддержку приложения. Сопоставление с разделенной архитектурой React + API выявляет преимущества Django-шаблонов в тех случаях, когда проект не требует сложной клиентской логики и независимого масштабирования фронтенда. Наиболее существенными оказываются единая кодовая база, предсказуемая поисковая индексация, меньшее число точек взаимодействия уровней приложения и более простая экранизация ошибок. Вместе с тем Django-шаблоны имеют собственные ограничения. Они менее удобны при разработке интерфейсов с высокой степенью интерактивности, офлайн-режимом работы и большим количеством независимых frontend-компонентов. Следовательно, выбор между Django-шаблонами и SPA-архитектурой должен определяться не технологической новизной решения, а задачами проекта [13], ресурсами команды и требованиями к дальнейшему сопровождению проекта.

СПИСОК ИСТОЧНИКОВ

1. Фаулер М. Шаблоны корпоративных приложений = Patterns of Enterprise Application Architecture / пер. с англ. Испр. изд. СПб.: Диалектика, 2020. 544 с.
2. Neshet G. SPA Are Dead, Long Live SPA // Semaphore Blog. 23.08.2023. URL: <http://semaphore.io/blog/single-page-applications> (дата обращения: 14.05.2026).
3. Дронов В. А. Django 4. Практика создания веб-сайтов на Python. СПб.: БХВ-Петербург, 2023. 800 с.
4. Наумов Р. В. Программирование Питон. Выбор веб-фреймворка // Достижения науки и образования. 2016. № 12 (13). С. 25–26.
5. Django Documentation. Version 5.0. URL: <http://docs.djangoproject.com/en/5.0> (дата обращения: 16.05.2026).
6. Лутц М. Изучаем Python: авторитетный курс объектно-ориентированного программирования. Пятое издание = Learning Python. Fifth Edition: в 2 т. / пер. с англ. и ред. Ю. Н. Артеменко. СПб.: Диалектика, 2020.

7. Jadhav S. B., Gholve S. S. Django vs. FastAPI: A Comparative Study for High-Performance Web Applications // International Journal of Advance and Applied Research. 2025. Vol. 6, no. 23. Pp. 47–51. DOI: 10.5281/zenodo.15119179
8. Software Engineering Cost Estimation using COCOMO II Model / К. Htay [et al.] // International Journal of Trend in Scientific Research and Development. 2019. Vol. 3, iss. 5. Pp. 2326–2329. DOI: 10.31142/ijtsrd28019
9. Основы поисковой оптимизации сайтов, использующих JavaScript // Google for Developers. URL: <http://developers.google.com/search/docs/crawling-indexing/javascript/javascript-seo-basics> (дата обращения: 17.05.2026).
10. Achille A. Server-Side Rendering vs. Client-Side Rendering: A Guide for Web Development // BairesDev Blog. 13.05.2026. URL: <http://www.bairesdev.com/blog/server-side-client-rendering-web-development> (дата обращения: 15.05.2026).
11. Зарплаты IT-специалистов в первой половине 2025: +2 %, рост замедлился // Хабр. 14.08.2025. URL: <http://habr.com/ru/specials/936618> (дата обращения: 14.05.2026).
12. The Django Template Language: Django Documentation. URL: <http://docs.djangoproject.com/en/5.1/ref/templates/language> (дата обращения: 17.05.2026).
13. Verbina E. Which is the Best Python Web Framework: Django, Flask, or FastAPI? // PyCharm Blog. 18.02.2026. URL: <http://blog.jetbrains.com/pycharm/2025/02/django-flask-fastapi> (дата обращения: 13.05.2026).

Дата поступления: 19.05.2026

Решение о публикации: 22.05.2026

Advantages of Using Django Template Engines for Business: The Balance Between Efficiency and Cost of Development

- Andrey V. Zabrodin** — PhD in History, Associate Professor of the “Information and Computing Systems” Department. Research interests: information systems, data analytics, programming paradigms, cloud technologies. E-mail: zabrodin@pgups.ru
- Andrey D. Mikheev** — 2nd year Bachelor’s Degree Student in 09.03.01 Informatics and Computer Technology. Research interests: automation and customer loyalty, web development. E-mail: miheev.31.andrey@gmail.com
- Maksim A. Graur** — 2nd year Bachelor’s Degree Student in 09.03.01 Informatics and Computer Technology. Research interests: automation and customer loyalty, web development. E-mail: maksimgraur22@mail.ru

Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky ave., Saint Petersburg, 190031, Russia

For citation: Zabrodin A. V., Mikheev A. D. Graur M. A. Advantages of Using Django Template Engines for Business: The Balance Between Efficiency and Cost of Development, *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 36–45. DOI: 10.20295/2413-2527-2026-246-36-45 (In Russian).

Abstract. *The problem of choosing a technology stack for small and medium-sized business web projects is considered. **Purpose:** against the background of the proliferation of architectures with separation of frontend and backend, to analyze the feasibility of using the built-in Django template engine as a simpler solution for projects with a limited budget, a small team and moderate requirements for the interactivity of site pages. **Results:** it is shown that server-side rendering combined with single-page application queries can reduce architectural complexity, reduce application development and support costs, preserve SEO benefits, and use Django’s built-in security mechanisms. It is concluded that Django templates can act as a rational architectural choice for a*

number of applied business tasks. **Practical significance:** the research results can be used in the design of mass data collection systems and interaction with them, the construction of distributed search robots and the creation of information analysis platforms based on multi-agent systems. **Discussion:** the presented observations reflect the practical structure of the functioning of a distributed search robot. Unlike classical search engines, the proposed agent-based approach allows you to redistribute the load and dynamically change the bypass strategies when blocking is detected.

Keywords: Django, template engine, comparative architecture analysis, web development, observability, AJAX, small business

REFERENCES

1. Fowler M. Shablony korporativnykh prilozheniy [Patterns of Enterprise Application Architecture]. Saint Petersburg, Dialektika Publishing House, 2020, 544 p. (In Russian)
2. Neshor G. SPA Are Dead, Long Live SPA, *Semaphore Blog*. Published online at August 23, 2023. Available at: <http://semaphore.io/blog/single-page-applications> (accessed: May 14, 2026).
3. Dronov V. A. Django 4. Praktika sozdaniya veb-saytov na Python [Creating Websites with Python]. Saint Petersburg, BHV-Peterburg Publishing House, 2023, 800 p. (In Russian)
4. Naumov R. V. Programmirovaniye Piton. Vybor veb-freymvorka [Python Programming. Choosing a Web Framework], *Dostizheniya nauki i obrazovaniya*, 2016, no. 12 (13), pp. 25–26. (In Russian)
5. Django Documentation. Version 5.0. Available at: <http://docs.djangoproject.com/en/5.0> (accessed: May 16, 2026).
6. Lutz M. Izuchaem Python: avtoritetnyy kurs obektno-orientirovannogo programmirovaniya. Pyatoe izdanie [Learning Python. Fifth Edition]. In 2 volumes. Saint Petersburg, Dialektika Publishing House, 2020. (In Russian)
7. Jadhav S. B., Gholve S. S. Django vs. FastAPI: A Comparative Study for High-Performance Web Applications, *International Journal of Advance and Applied Research*, 2025, vol. 6, no. 23, pp. 47–51. DOI: 10.5281/zenodo.15119179
8. Htay K., et al. Software Engineering Cost Estimation using COCOMO II Model, *International Journal of Trend in Scientific Research and Development*, 2019, vol. 3, iss. 5, pp. 2326–2329. DOI: 10.31142/ijtsrd28019
9. Understand the JavaScript SEO Basics, *Google for Developers*. Available at: <http://developers.google.com/search/docs/crawling-indexing/javascript/javascript-seo-basics> (accessed: May 17, 2026).
10. Achille A. Server-Side Rendering vs. Client-Side Rendering: A Guide for Web Development, *BairesDev Blog*. Published online at May 13, 2026. Available at: <http://www.bairesdev.com/blog/server-side-client-rendering-web-development> (accessed: May 15, 2026).
11. Zarplaty IT-spetsialistov v pervoy polovine 2025: +2%, rost zamedlilsya [Salaries of IT Specialists in the First Half of 2025: +2%, Growth Slowed Down], *Khabr [Habr]*. Published online at August 14, 2025. Available at: <http://habr.com/ru/specials/936618> (accessed: May 14, 2026). (In Russian)
12. The Django Template Language: Django Documentation. Available at: <http://docs.djangoproject.com/en/5.1/ref/templates/language> (accessed: May 17, 2026).
13. Verbina E. Which is the Best Python Web Framework: Django, Flask, or FastAPI? *PyCharm Blog*. Published online at February 18, 2026. Available at: <http://blog.jetbrains.com/pycharm/2025/02/django-flask-fastapi> (accessed: May 13, 2026).

Received: May 19, 2026

Accepted: May 22, 2026

УДК 004.8

Инструмент сбора и оценки общественного мнения о транспортной отрасли из открытых источников

- Стрельников Никита Рафаилевич** — курсант 4-го курса направления 09.05.01 «Применение и эксплуатация автоматизированных систем специального назначения». Научные интересы: обработка естественного языка, методы машинного обучения, анализ данных. E-mail: vka@mil.ru
- Михайлова Светлана Александровна** — преподаватель кафедры математического и программного обеспечения специальных космических комплексов. Научные интересы: распределенные системы, архитектура приложений, методы машинного обучения, анализ данных. E-mail: vka@mil.ru
- Алимов Наиль Ильгизович** — кандидат техн. наук, старший преподаватель кафедры математического и программного обеспечения специальных космических комплексов. Научные интересы: методы и алгоритмы обработки данных, параллельные и распределенные вычисления. E-mail: vka@mil.ru

Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Стрельников Н. Р., Михайлова С. А., Алимов Н. И. Инструмент сбора и оценки общественного мнения о транспортной отрасли из открытых источников // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 46–53. DOI: 10.20295/2413-2527-2026-246-46-53

Аннотация. *Представлено исследование о применении методов интеллектуального анализа текстовой информации из открытых источников для формирования управленческих решений на примере анализа отзывов населения о транспортной отрасли. Цель: обоснование подхода к интеллектуальному анализу текстовых отзывов пассажиров о транспортной отрасли, направленного на трансформацию неструктурированных пользовательских высказываний в структурированные данные, пригодные для формирования управленческих решений. Методы: использованы методы обработки естественного языка, машинное обучение и технологии интеграции данных из разнородных источников. Результаты: подчеркивается эффективность предложенного конвейера обработки текста, включающего классификацию по виду транспорта и тематике, анализ тональности и извлечение именованных сущностей для выявления ключевых проблем и трендов общественного мнения. Практическая значимость: повышение оперативности реагирования транспортных регуляторов на проблемы пассажиров, оптимизация сервиса и переход к управлению на основе данных. Исследование имеет важное значение для развития цифровых технологий в транспортной отрасли и повышения эффективности управления пассажирскими перевозками в условиях цифровой трансформации.*

Ключевые слова: интеллектуальный анализ данных, обработка естественного языка, Python, транспорт, пассажиры, тональность, машинное обучение, медианпространство

2.3.5 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки); 1.2.1 — искусственный интеллект и машинное обучение (технические науки)

Введение

Цифровое развитие глобально меняет взаимодействие между акторами сферы услуг [1]. Транспортная отрасль не является исключени-

ем и внедряет цифровые технологии во все сферы деятельности [2]. Пользователи транспортной отрасли, в свою очередь, активно реагируют на

нововведения и буквально в режиме реального времени оценивают качество оказываемых им услуг [3]. Наиболее полными и объективными в данной ситуации являются отзывы пользователей, публикуемые в открытых источниках: в мессенджерах, на форумах, в тематических рубриках и чатах [4]. Именно в этих источниках выявляются главные системные проблемы: недовольство логикой распределения пассажиропотока, некорректная логистика транспорта и многие другие проблемы, которые активно обсуждаются пользователями между собой и являются ценным источником сведений.

Однако объем публикуемой информации становится настолько велик, что ее ручная обработка далее невозможна и необходимо применение автоматизированных систем обработки и анализа информации. Современные технологии обработки естественного языка [5] и машинного обучения [6] предлагают инструменты для трансформации неструктурированного текстового потока в структурированное знание, пригодное для аналитики и принятия решений.

Целью статьи является разработка и обоснование подхода к интеллектуальному анализу текстовых отзывов пассажиров о транспортной отрасли. В задачи исследования входит систематизация типов источников данных и методов их сбора, проектирование конвейера обработки текста, включающего этапы очистки, классификации, анализ тональности и извлечение информации, демонстрация работоспособности подхода на практических примерах, формулирование практической значимости для транспортной отрасли.

Источники данных и методы их сбора

Источники данных, содержащие информацию о транспортной ситуации и мнениях пассажиров, можно разделить на два основных типа [7]:

1. **Официальные (структурированные):** нормативные документы, формирующиеся в рамках транспортного предприятия, данные системы отслеживания движения транспорта на основе ГЛОНАСС или GPS, официальные статистические данные.

2. **Неофициальные (слабоструктурированные):** средства массовой информации (СМИ), результа-

ты опросов, социальные сети, мессенджеры, открытые форумы и порталы для обсуждения качества предоставляемых услуг.

Извлечение информации из подобных источников может осуществляться различными методами. В отличие от структурированных ресурсов, предоставляющих уже классифицированные данные по определенным категориям, слабоструктурированные источники представляют собой более сложную задачу для обработки.

Для организации сбора информации из слабоструктурированных источников применяются методы интеллектуальной обработки текстов и анализа неструктурированных данных [8]. Основным инструментом в этом случае выступает скрапинг — технология автоматизированного сбора данных из источников, размещенных в открытом доступе в сети Интернет. После скачивания текста необходимо выполнить его семантический анализ с применением методов обработки естественного языка, что позволяет не только извлечь из массива сообщений конкретные сущности, такие как маршруты, временные промежутки и типы транспорта, но и провести классификацию тональности высказываний (сентимент-анализ) для определения общего настроения пассажиров.

Telegram, как наиболее известный и активно используемый мессенджер, является одним из самых популярных способов обсуждения транспорта в сети Интернет. В нем существуют как официальные каналы перевозчиков, так и неофициальные сообщества, созданные пользователями. Эти каналы содержат прямые текстовые посты, комментарии, медиафайлы с подписями. Для сбора данных из Telegram используется асинхронная библиотека Telethon, предоставляющая доступ к интерфейсу прикладного программирования мессенджера.

Первоначальный отбор источников производился на основе принципа верификации [9]. Для отсеивания откровенно маргинальных или анонимных каналов с низкой степенью достоверности в выборку включались только те ресурсы, которые были официально зарегистрированы в качестве СМИ в реестре Роскомнадзора. Это позволило обеспечить минимальный порог ответственности авторов

за публикуемые сведения и гарантировать определенный уровень регулярности выхода материалов.

Для формирования первичного списка каналов использовались открытые веб-каталоги и агрегаторы Telegram, в частности tgstat.ru. С его помощью был собран датасет, включающий более 4000 открытых телеграм-каналов. В выборку включены каналы 43 различных тематических категорий, что позволяет отслеживать, как транспортные проблемы обсуждаются в контексте городской жизни, экономики, происшествий и других сфер.

Информация собиралась за период с 1 января 2022 года по 17 ноября 2025 года. Такой длительный временной срез позволяет не только оценить текущее состояние настроений пассажиров, но и проанализировать динамику их изменений в зависимости от сезона, экономической ситуации или резонансных событий.

В результате проведенных работ по сбору и первичной агрегации данных была сформирована база, в которой на текущий момент хранится 1 026 163 поста. Этот массив неструктурированных текстовых данных является основой для последующего семантического анализа и выявления закономерностей, связанных с удовлетворенностью пассажиров общественным транспортом.

Конвейер обработки и анализа текста

Обработка и анализ текста выполняются за семь этапов (рис. 1).

Предобработка и очистка данных выполняется в несколько этапов. Текст приводится к единому нижнему регистру, удаляются символы пунктуации и ссылки. Затем текст токенизируется и приводится к морфологической норме с помощью лемматизации для сохранения смысловой нагрузки и группирования различных форм слова (например, слова «опоздали» и «опаздывает» приводятся к общей лемме «опаздывать»), что позволяет выполнять более точный семантический и частотный анализ. В заключение удаляются стоп-слова на базе расширенного списка, адаптированного для транспортной отрасли.

Классификация по виду транспорта определяет, к какой сфере относится сообщение (обществен-



Рис. 1. Схема работы конвейера обработки и анализа текста

ный, железнодорожный или воздушный транспорт), а классификация по тематике/проблеме соотносит текст с категорией проблемы (расписание, комфорт, безопасность, цена, работа персонала и т. д.). Это позволяет автоматически тегировать обращения и агрегировать статистику по типам проблем. Данная задача эффективно решается с помощью предварительно обученной модели XLM-RoBERTa. Особенность этой модели в том, что она в первую очередь предназначена для задач, в которых для принятия решений используется все предложение, что в поставленной задаче дает максимальное преимущество.

Анализ тональности определяет эмоциональную окраску высказываний: позитивная, нейтральная или негативная. Для русского языка используется предобученная модель RuBERT-tiny2-russian-sentiment, дополнительно обученная на датасете, размеченном экспертами в транспортной области с использованием фреймворка Transformers.

Извлечение именованных сущностей выделяет в тексте значимые объекты, такие как названия перевозчиков и подрядчиков, станции метро, вокзалы, аэропорты, улицы, города, номера поездов, номера маршрутов автобусов, авиарейсы и многие другие. Для данной задачи на русском языке используется модель RuBERT-base-cased.

Заключительным этапом конвейера сбора и анализа текстов является векторизация датасета с целью дальнейшего выполнения задач анализа и принятия решения. Библиотека Sentence-Transformers преобразует текст в эмбединги. Эмбединги вместе с исходным текстом, метаданными (источник, время) и результатами классификации сохраняются в базе данных. Для хранения используется система управления базами данных (СУБД) PostgreSQL с расширением pgvector. Это позволяет в рамках одной СУБД выполнять как

обычные запросы по метаданным, так и семантический поиск по векторной близости.

Программная архитектура и реализация

Система реализуется в виде набора микросервисов, управляемых Docker Compose, что обеспечивает модульность и масштабируемость (рис. 2).

Сервис сбора данных — набор скриптов на Python, использующих Telethon для парсинга каналов и размещения в очередь для обработки.

Сервис предобработки принимает сырой текст, выполняет очистку и лемматизацию.

Сервис машинного обучения — это отдельные сервисы, обернутые в FastAPI, который предоставляет REST API для классификации, анализа тональности и извлечения именованных сущностей. Внутри используют загруженные модели из Hugging Face: для бинарной классификации —

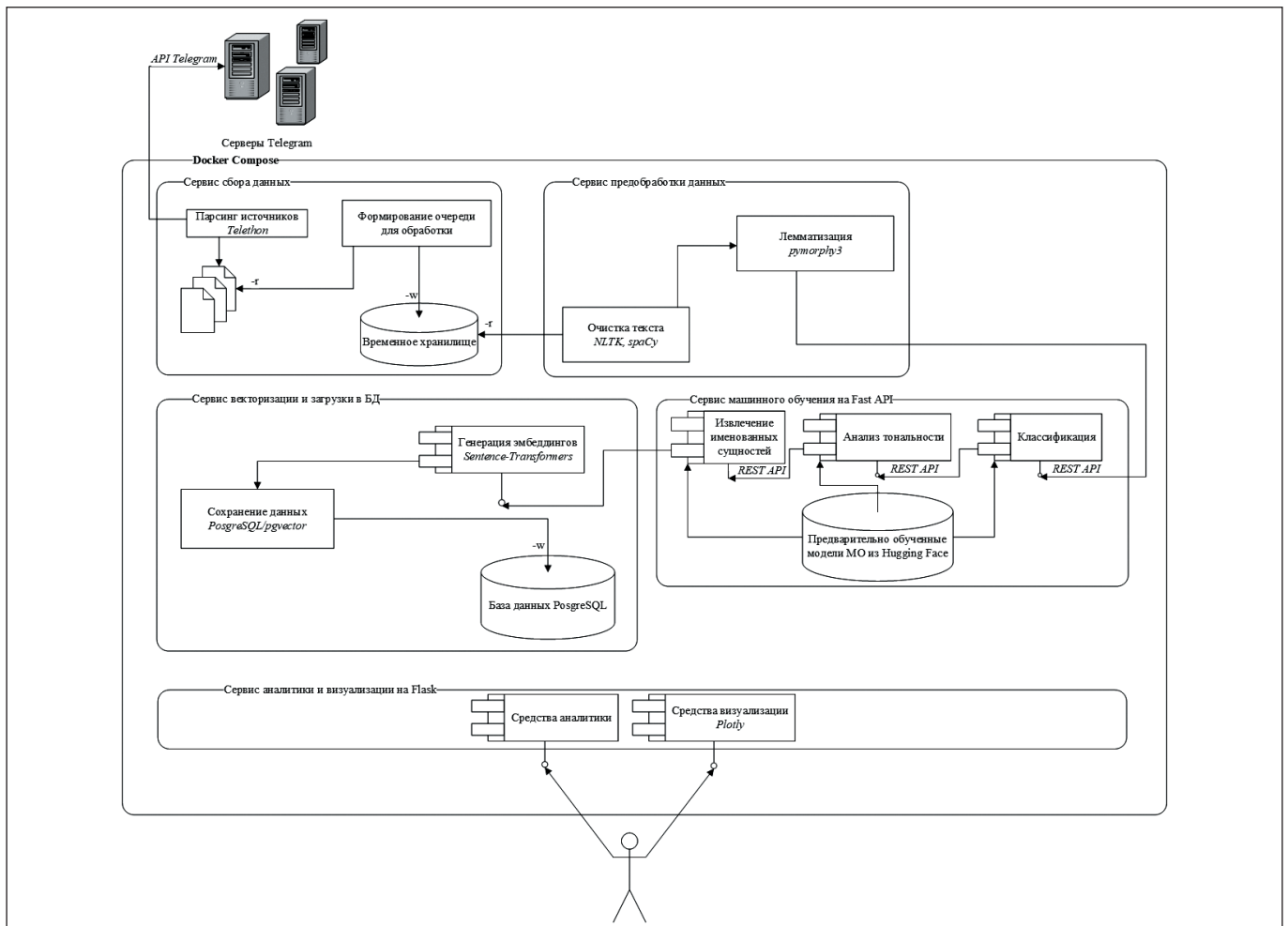


Рис. 2. Схема работы системы обработки данных и взаимодействия с пользователями

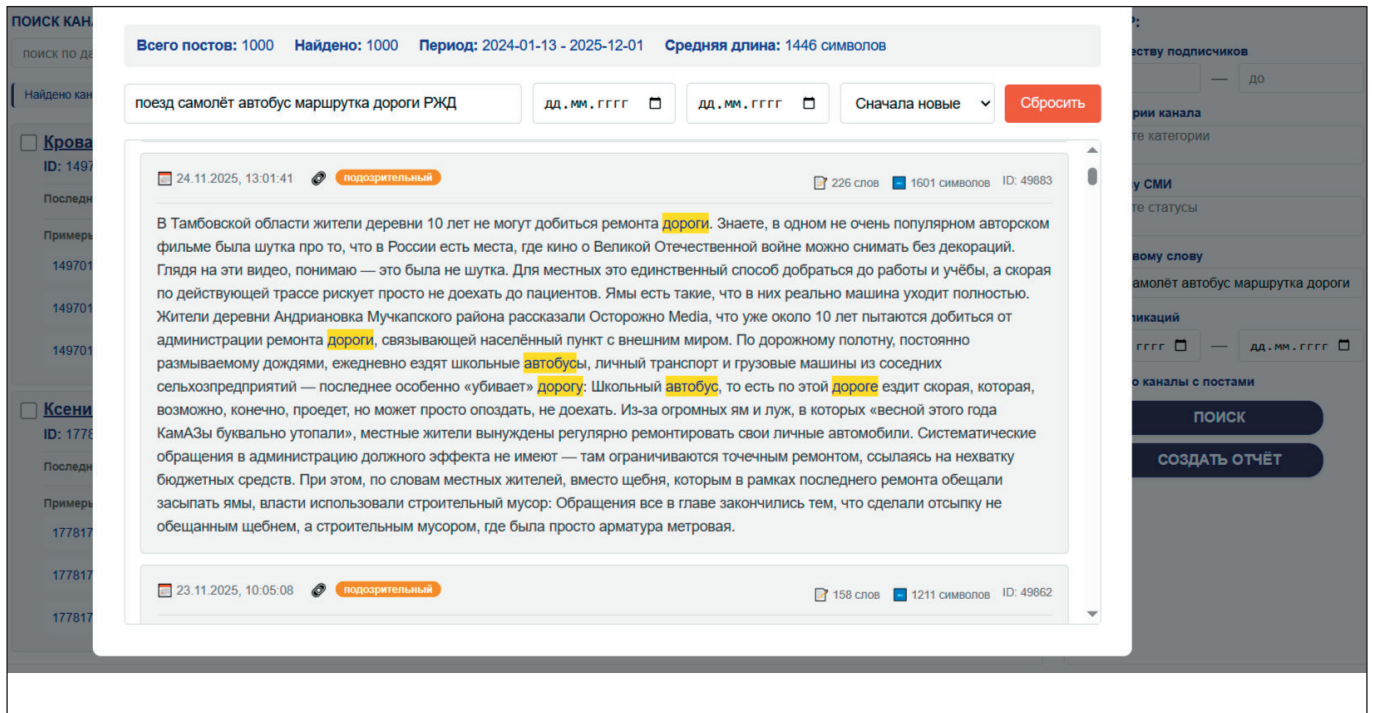


Рис. 4. Страница веб-приложения, предназначенная для отображения результатов поиска по ключевым словам

скрипты в сервис сбора данных. Таким образом, возможно выполнение анализа на большей выборке данных, что способствует формированию статистически корректной картины.

Заключение

Представленный инструмент, построенный на основе конвейера обработки текста, включающего классификацию по виду транспорта и тематике, анализ тональности и извлечение именованных сущностей, позволяет продемонстрировать перспективы применения анализа информации из открытых источников с целью оценки общественного мнения о транспортной отрасли.

СПИСОК ИСТОЧНИКОВ

1. Цифровые пассажирские сервисы: будущее транспорта обсудили на ЦИПРе // Ассоциация «Цифровой транспорт и логистика». URL: <http://www.dtla.ru/news/tsifrovye-passazhirskie-servisy-budushchee-transporta-obsudili-na-tsipre/> (дата обращения: 15.01.2026).
2. 80 % компаний транспортной отрасли используют цифровые технологии, но потенциал роста есть // Цифровая индустрия промышленной России. URL: <http://cipr.ru/news/80-kompanij-transportnoj-otrasli-ispolzuyut-cifrovye-tehnologii-no-potencial-rosta-est> (дата обращения: 15.01.2026).
3. Анализ комментариев в социальных сетях и мессенджерах как метод оценки социальной результативности цифровых городских сервисов / О. Г. Филатова [и др.] // International Journal of Open Information Technologies. 2024. Т. 12, № 11. С. 103–110.

4. What do Riders Say and Where? The Detection and Analysis of Eyewitness Transit Tweets / O. Kabbani [et al.] // Journal of Intelligent Transportation Systems. 2023. Vol. 27, iss. 3. Pp. 347–363. DOI: 10.1080/15472450.2022.2026773
5. Lui Y., Li Y., Li W. A Natural Language Processing Approach for Appraisal of Passenger Satisfaction and Service Quality of Public Transportation // IET Intelligent Transport Systems. 2019. Vol. 13, iss. 11. Pp. 1701–1707. DOI: 10.1049/iet-its.2019.0054
6. Максютин П. А., Шульженко С. Н. Обзор методов классификации текстов с помощью машинного обучения // Инженерный вестник Дона. 2022. № 12.
7. Zannat K. E., Choudhury C. F. Emerging Big Data Sources for Public Transport Planning: A Systematic Review on Current State of Art and Future Research Directions // Journal of the Indian Institute of Science. 2019. Vol. 99, iss. 4. Pp. 601–619. DOI: 10.1007/s41745-019-00125-9
8. Chowdhury S., Alzarrad A. Applications of Text Mining in the Transportation Infrastructure Sector: A Review // Information. 2023. Vol. 14, iss. 4. Art. 201. DOI: 10.3390/info14040201
9. Коновалова М. В. Когнитивные аспекты верификации в интернет-медиадискурсе // Лингвокультурология. 2019. № 13. С. 125–131.

Дата поступления: 06.04.2026

Решение о публикации: 22.05.2026

A Tool for Collecting and Assessing Public Opinion on the Transport Industry from Open Sources

Nikita R. Strelnikov

— 4th year Cadet in 09.05.01 Application and Operation of Automated Systems for Special Purposes. Research interests: natural language processing, machine learning methods, data analysis. E-mail: vka@mil.ru

Svetlana A. Mikhaylova

— Lecturer at the Department of mathematical and software support for special space systems. Research interests: natural language processing, machine learning methods, data analysis. E-mail: vka@mil.ru

Nail I. Alimov

— PhD in Engineering, Senior Lecturer at the Department of mathematical and software support for special space systems. Research interests: natural language processing, machine learning methods, data analysis. E-mail: vka@mil.ru

Mozhaisky Military Aerospace Academy, 13, Zhdanovskaya str., Saint Petersburg, 197198, Russia

For citation: Strelnikov N. R., Mikhaylova S. A., Alimov N. I. A Tool for Collecting and Assessing Public Opinion on the Transport Industry from Open Sources, *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 46–53. DOI: 10.20295/2413-2527-2026-246-46-53 (In Russian)

Abstract. *The study presents the development of a tool for intelligent analysis of passenger feedback on transport. **Purpose:** is to create an automated monitoring and semantic analysis system aimed at transforming unstructured user statements into structured data for managerial decision-making. **Methods:** modern information technologies were used, including natural language processing methods, machine learning, and integration of data from heterogeneous sources. **Results:** emphasize the effectiveness of the proposed text processing pipeline, which includes classification by transport mode and topic, sentiment analysis, and named entity recognition, for identifying key issues and public opinion trends. **Practical significance:** lies in increasing the responsiveness of transport regulators to passenger problems, optimizing service, and transitioning to data-driven management.*

The research is important for the development of digital technologies in the transport industry and improving the efficiency of passenger transportation management in the context of digital transformation.

Keywords: *data mining, natural language processing, Python, transport, passengers, sentiment, machine learning, media space*

REFERENCES

1. Tsifrovye passazhirskie servisy: budushchee transporta obsudili na TsIPRe [Digital Passenger Services: the Future of Transport Discussed at CIPR], *Sayt Assotsiatsii "Tsifrovoy transport i logistika" [Website of the Digital Transport and Logistics Association]*. Published online on June 02, 2025. Available at: <http://www.dtla.ru/news/tsifrovye-passazhirskie-servisy-budushchee-transporta-obsudili-na-tsipre/> (accessed: January 15, 2026). (In Russian)
2. 80 % kompaniy transportnoy otrasli ispolzuyut tsifrovye tekhnologii, no potentsial rosta est [80% of Companies in the Transportation Industry Use Digital Technologies, but There is Growth Potential], *Tsifrovaya industriya promyshlennoy Rossii [Digitalization of Industrial Russia]*. Published online on June 30, 2025. Available at: <http://cipr.ru/news/80-kompanij-transportnoj-otrasli-ispolzuyut-czifrovye-tehnologii-no-potenczial-rosta-est/> (accessed: January 15, 2026). (In Russian)
3. Filatova O. G., et al. Analiz kommentariyev v sotsialnykh setyakh i messendzherakh kak metod otsenki sotsialnoy rezultativnosti tsifrovyykh gorodskikh servisov [Analysis of Comments on Social Networks and Messengers as a Method of Evaluating the Social Effectiveness of Digital Urban Services], *International Journal of Open Information Technologies*, 2024, vol. 12, no. 11, pp. 103–110. (In Russian)
4. Kabbani O., et al. What do Riders Say and Where? The Detection and Analysis of Eyewitness Transit Tweets, *Journal of Intelligent Transportation Systems*, 2023, vol. 27, iss. 3, pp. 347–363. DOI: 10.1080/15472450.2022.2026773
5. Liu Y., Li Y., Li W. A Natural Language Processing Approach for Appraisal of Passenger Satisfaction and Service Quality of Public Transportation, *IET Intelligent Transport Systems*, 2019, vol. 13, iss. 11, pp. 1701–1707. DOI: 10.1049/iet-its.2019.0054
6. Maksyutin P. A., Shuljenko S. N. Obzor metodov klassifikatsii tekstov s pomoshchyu mashinnogo obucheniya [An Overview of Text Classification Methods Using Machine Learning], *Inzhenernyy vestnik Dona [Engineering Journal of Don]*, 2022, no. 12. (In Russian)
7. Zannat K. E., Choudhury C. F. Emerging Big Data Sources for Public Transport Planning: A Systematic Review on Current State of Art and Future Research Directions, *Journal of the Indian Institute of Science*, 2019, vol. 99, iss. 4, pp. 601–619. DOI: 10.1007/s41745-019-00125-9
8. Chowdhury S., Alzarrad A. Applications of Text Mining in the Transportation Infrastructure Sector: A Review, *Information*, 2023, vol. 14, iss. 4, art. 201, 24 p. DOI: 10.3390/info14040201
9. Konovalova M. V. Kognitivnye aspekty verifikatsii v Internet-mediadiskurse [Cognitive Aspects of Verification on the Internet Media Discourse], *Lingvokulturologiya*, 2019, no. 13, pp. 125–131. (In Russian)

Received: April 06, 2026

Accepted: May 22, 2026

УДК 004.415.2

Реактивное программирование как эволюция архитектурных компромиссов в управлении изменениями

Цуканова Дарья Александровна

— студент 4-го курса направления 09.03.01 «Информатика и вычислительная техника». Научные интересы: реактивное программирование, разработка высоконагруженных информационных систем, парадигмы программирования. E-mail: GinD.Nova@yandex.ru

Забродин Андрей Владимирович

— кандидат ист. наук, доцент кафедры «Информационные и вычислительные системы». Научные интересы: информационные системы, аналитика данных, парадигмы программирования, облачные технологии. E-mail: zabrodin@pgups.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Цуканова Д. А., Забродин А. В. Реактивное программирование как эволюция архитектурных компромиссов в управлении изменениями // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 54–63. DOI: 10.20295/2413-2527-2026-246-54-63

Аннотация. Представлено исследование эволюции моделей реактивного программирования от теоретических основ функционального реактивного программирования до современных практических реализаций в библиотеках реактивных потоков и UI-фреймворках. **Цель:** систематизировать подходы к управлению зависимостями и распространению изменений в программных системах. **Методы:** сравнительный анализ архитектурных решений в FRP, библиотеках реактивных потоков и UI-фреймворках. **Результаты:** показано, что реактивные модели не устраняют сложность управления изменениями во времени, а перераспределяют ее между кодом разработчика, механизмами выполнения и инструментами разработки. Выявлены ключевые различия между push-моделью, основанной на явном распространении событий и управлении подписками, и pull-моделью, использующей автоматическое отслеживание зависимостей. **Практическая значимость:** заключается в уточнении архитектурных компромиссов различных моделей реактивности при разработке пользовательских интерфейсов и серверных систем обработки данных.

Ключевые слова: реактивное программирование, функциональное реактивное программирование, push-модель, pull-модель, управление зависимостями, архитектурные паттерны

2.3.5 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки)

Введение

Реактивность в программировании — это способ автоматически обновлять состояние системы в ответ на изменение входных данных. Существует множество подходов к реализации этого принципа, каждый из которых накладывает свои архитектурные ограничения и определяет синтаксис кода. Однако работы по сравнению реактивных подхо-

дов остаются фрагментарными, сосредоточенными на конкретных инструментах. Они не раскрывают общих принципов. Кроме того, отсутствует единая классификация, охватывающая реализации от теории до практики.

Цель статьи — систематизация моделей реактивности через анализ того, как они распределяют

сложность управления изменениями между кодом разработчика, механизмом выполнения и инструментами.

Основная гипотеза: реактивные модели не устраняют сложность управления изменениями, а перемещают ее с одного уровня на другой. Выбор модели определяется архитектурными особенностями системы и квалификацией разработчиков, поскольку каждый подход требует компенсировать сложность в разных местах.

Задачи:

- проследить эволюцию реактивных моделей от теоретических основ до современных реализаций;
- выявить фундаментальные компромиссы между различными моделями реактивности;
- проанализировать, как эти компромиссы проявляются в архитектуре конкретных инструментов и фреймворков.

Научная новизна исследования заключается в систематизации моделей реактивности с точ-

ки зрения распределения сложности между кодом разработчика, механизмами выполнения и инструментами разработки. Показано, что push-, pull- и гибридные модели различаются не только способом распространения изменений, но и тем, на каком уровне локализуются архитектурные риски, а также кто несет ответственность за согласованность данных.

Императивное управление состоянием: ограничения парадигмы

До реактивного программирования доминировал императивный подход. Разработчик явно описывал последовательность действий при каждом событии: получить событие, обновить состояние, найти зависимые части интерфейса, обновить их вручную. Это работало для простых программ, но с ростом масштаба возникали проблемы [1].

Основная сложность — ручное управление графом зависимостей. Как показано на рисунке, при изменении переменной *A* разработчик обязан найти

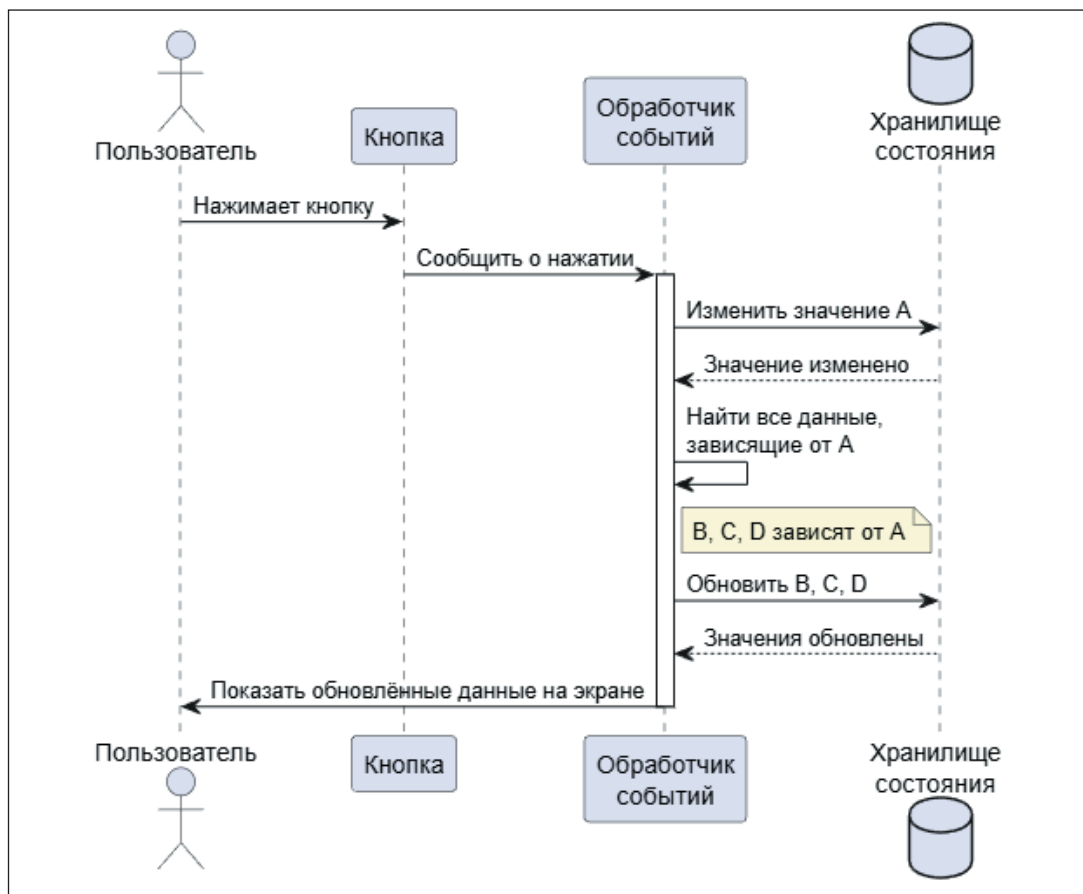


Рисунок. Императивное обновление состояния: ручное управление зависимостями

все зависящие от нее элементы (B , C , D), обновить их в правильном порядке и лишь затем перерисовать экран. При десятках и сотнях взаимозависимых переменных легко пропустить элемент или нарушить порядок обновлений, что приводит к рассогласованности данных и трудноуловимым ошибкам [2, 3].

Классический пример императивного подхода — паттерн Observer: наблюдатель подписывается на изменения субъекта. Это требует явного управления подписками (риск утечек памяти), не гарантирует порядка обновлений при сложных зависимостях и перекладывает ответственность за согласованность данных на разработчика. С ростом графа зависимостей задача становится неуправляемой.

Эти ограничения императивного управления состоянием задали вектор поиска альтернативных моделей: идеи dataflow и функционального программирования, оформившиеся к 1960–1970-м годам, впоследствии легли в основу функционального реактивного программирования (Functional reactive programming, FRP).

Functional reactive programming: формальная модель времени

Функциональное реактивное программирование было формализовано Коном Эллиоттом и Полом Худаком в [4]. Вместо последовательности действий при событиях они предложили моделировать интерактивные системы как математические функции над временем.

Центральная идея: изменяющиеся величины представляются не переменными, которые обновляются, а непрерывными функциями времени с четкой денотационной семантикой (математическим значением, независимым от реализации).

Эллиотт ввел два базовых понятия, на которых строится вся система:

- events (события) — дискретные значения, возникающие в отдельные моменты времени (например, клики мышью или нажатия клавиш);
- behaviors (поведения) — значения, заданные для каждого момента времени и меняющиеся непрерывно (например, текущие координаты курсора или текущее время).

Различие между behavior и event принципиально [4]. Behavior отвечает на вопрос «каково значение сейчас?», event представляет «что произошло?». Первое непрерывно, второе дискретно.

Ключевое преимущество FRP — композируемость, обеспечиваемая функциональными абстракциями (map, аппликативные функторы, функции высшего порядка). Это позволяет строить сложные поведения из простых компонентов с корректностью, проверяемой статической типизацией. Денотационная семантика дает возможность математически доказывать свойства программ в отличие от операционной семантики, описывающей лишь механизм выполнения.

На практике идеи FRP столкнулись с тремя проблемами.

Первая — расхождение модели с реальностью: непрерывная функция времени против дискретных событий, пакетной обработки и ограниченных ресурсов [5]. Реализации вынуждены вводить эвристики (батчинг, дискретизацию, кэширование) [6], что отдаляет систему от идеальной семантики.

Вторая проблема заключается в сложности взаимодействия с внешним миром (сеть, файловая система, асинхронность) [7], которая требует дополнительных механизмов (подписки, очистка ресурсов, буферизация), иначе возникают утечки памяти и состояния гонки.

Третья проблема — трудности отладки: автоматическое распространение изменений делает порядок вычислений неявным; разработчик не может проследить последовательность обновлений, что требует специальных инструментов профилирования.

Прагматичные адаптации: от математической строгости к практической применимости

Внедрение классического FRP в реальные системы выявило существенные препятствия, которые побудили разработчиков искать более прагматичные решения. Главное отступление от первоначальной концепции коснулось временной модели: вместо непрерывного времени индустрия перешла на обработку последовательности

отдельных событий. Reactive Extensions (Rx), разработанные компанией Microsoft в начале 2000-х годов, переформатировали FRP под новую парадигму — работу с дискретными событиями [8–10]. В этой модели observable функционирует как последовательность значений, снабженная набором функциональных трансформаций (select, where, combine) [8], где поставщик данных уведомляет заинтересованные компоненты о происходящих изменениях. Главным строительным блоком выступает event; система теперь отвечает на вопрос «какие события зафиксированы?» вместо «какое значение существует в момент времени t ?», и исходное понятие непрерывного поведения (behavior) в практическом применении Rx почти полностью вытеснено событийной моделью [5].

Reactive Extensions развивают концепцию наблюдателя, адаптируя ее для функциональных инструментов. Это позволяет выражать обработку потоков данных в декларативной форме и автоматизировать управление подписками. Такой подход лучше соответствует характеру современных вычислений и упрощает интеграцию асинхронных операций. Однако это требует переосмысления исходной семантики и увеличивает разрыв между практической реализацией и теоретическим идеалом FRP.

Классический Rx столкнулся с проблемой обратной реакции (backpressure) при работе с высокоскоростными потоками данных. Это привело к разработке более совершенных реализаций, таких как Project Reactor [11], которые интегрируют управление обратной реакцией в свою архитектуру.

Современные реактивные системы разработали набор проверенных шаблонов для управления сложностью асинхронных потоков данных [12]. Это сделало event-driven-архитектуру более предсказуемой для разработчиков.

Модели распространения изменений и их различия

После появления Reactive Extensions реактивные системы разделились на два фундаментально разных подхода к управлению зависимостями и распространению изменений: push-модель и pull-модель [13]. Сравнение моделей приведено в таблице.

В push-модели источник данных активно управляет процессом распространения изменений. Он знает о своих подписчиках (наблюдателях) и при каждом изменении самостоятельно инициирует цепочку уведомлений. Данные «проталкиваются» от источника к потребителям.

Таблица

Сравнение push- и pull-моделей

Аспект	Push-модель	Pull-модель
Инициатор изменений	Источник данных активно уведомляет наблюдателей	Реактивное вычисление неявно «вытягивает» значение при выполнении
Регистрация зависимостей	Явная подписка: observable.subscribe()	Неявная: возникает как побочный эффект чтения значения
Граф зависимостей	Статический, определяется явно разработчиком	Динамический, строится во время выполнения
Управление наблюдателями	Источник хранит список наблюдателей и уведомляет каждого	Источник заранее не знает о потребителях
Поток управления	От источника к наблюдателям	От наблюдателей к источникам
Момент уведомления	Определяет источник (уведомление отправляется немедленно)	Определяет наблюдатель (запрос данных по мере необходимости)
Примеры реализации	Reactive Extensions (RxJS, RxJava, Rx.NET), Project Reactor, event bus-системы и pub/sub-паттерны	SolidJS, MobX, Knockout, Svelte, Vue 3 (реактивные ref/effect), Angular Signals

В pull-модели зависимости фиксируются автоматически во время вычисления. Разработчик не подписывается вручную, а просто использует реактивное значение. Система, в свою очередь, самостоятельно отслеживает зависимости и повторно запускает вычисления при изменении источника.

Компромиссы между моделями реактивного программирования

Push-модель естественна для асинхронных сценариев, например непредсказуемых событий, которые могут идти от сети или пользователей. Она обеспечивает явный поток данных: видно, откуда и куда они идут.

Обратная сторона — нужно строго следить за подписками: каждая должна вовремя завершаться, иначе появляются утечки памяти. При множественных одновременных источниках возможны каскадные уведомления и лишние вычисления. Поэтому современные библиотеки предлагают батчинг и координацию, но использовать их — работа разработчика.

Pull-модель решает проблему подписок более радикально: зависимости возникают при чтении, обновляются при изменении источников и автоматически удаляются. Система сама оптимизирует порядок обновлений через граф зависимостей, гарантируя единственное вычисление за цикл с согласованными данными. Однако эта автоматизация работает только синхронно. При асинхронности (`await`, `callback`) контекст отслеживания теряется, становится невозможно восстановить связь между чтением и источником. Требуются специальные обертки или явное управление эффектами, что возвращает часть сложности. Поэтому pull доминирует в синхронных сценариях: вычисления состояний UI и производных значений, локальная мемоизация.

Гибридные системы, поддерживающие обе модели, сталкиваются с новыми вопросами: как `observable` интегрируется с сигналами, как гарантировать согласованность при смешивании синхронных и асинхронных обновлений. Каждый ответ добавляет концептуальную нагрузку.

Это не недостаток конкретных реализаций — это фундаментальные компромиссы, вытекающие из самой природы моделей. Сложность управления изменениями во времени не исчезает, а перераспределяется. Push перекладывает ее на разработчика, pull — на реализацию системы, а гибридные подходы требуют понимания взаимодействия обеих моделей одновременно.

Современные реализации: от теории к практике

Эволюция моделей реактивности от теоретических основ FRP привела к появлению разнообразных практических реализаций в различных областях программирования. Наиболее заметное развитие произошло в двух направлениях: UI-фреймворки, где реактивность решает задачу автоматического отражения изменений данных в интерфейсе, и серверные системы, где реактивность помогает управлять асинхронными потоками данных.

Реактивность в UI-фреймворках

В контексте пользовательских интерфейсов реактивность демонстрирует несколько принципиально разных стратегий локализации сложности.

1. Минимальная реактивность с явным управлением.

Один из подходов — отказаться от автоматического отслеживания зависимостей на уровне данных, локализовав реактивность на уровне более крупных единиц, таких как компоненты. Данный подход используется в React. Его механизмы — Virtual DOM и reconciliation — предполагают, что компонент описывается как функция входных параметров и локального состояния, при изменении которого React перестраивает виртуальное дерево и сравнивает его с предыдущим, находя минимальный набор изменений для применения к DOM-браузера [14, 15].

Преимущество подхода — в простоте ментальной модели и отсутствии неявной магии. Недостатком является необходимость ручной оптимизации, то есть разработчик должен явно мемоизировать вычисления и компоненты,

перечислять зависимости в массивах хуков, из-за чего сложность переносится в код разработчика.

2. Автоматическая реактивность через глобальное отслеживание.

Противоположным подходом можно считать автоматический перехват всех потенциальных источников изменений. Наиболее известным примером реализации данной стратегии можно считать Angular (до версии 17), который использовал Zone.js — библиотеку, перехватывающую все асинхронные операции и автоматически иницилирующую процесс проверки изменений по их завершении [16].

Такой механизм помогает избавить разработчика от явного управления состояниями, но при этом проверяется все дерево компонентов на каждое изменение. Это приводит к автоматизму без явных подписок, но также может вызвать избыточную работу и непредсказуемость момента обновления. Данный подход переносит сложность в саму систему, что в некоторых ситуациях влияет на производительность.

3. Pull-модель с автоматическими зависимостями.

Третья стратегия — автоматическое отслеживание зависимостей на уровне отдельных данных, а не компонентов. SolidJS и Angular Signals (с версии 17) реализуют эту модель: компоненты инициализируются один раз, после чего обновляются только конкретные DOM-узлы, зависящие от изменившихся сигналов. Зависимости фиксируются автоматически при чтении [17–19].

Преимущество стратегии — минимальный объем обновлений без ручной оптимизации. Недостаток — необходимость понимать модель выполнения (setup и render) и ограничения при работе с асинхронностью. Сложность локализована в понимании неявной модели.

4. Компиляция реактивности.

Четвертая стратегия — анализ зависимостей статически на этапе сборки. Svelte анализирует код, выявляет реактивные зависимости и генерирует оптимальный императивный код обновлений без дополнительных библиотек во время выполнения. Разработчик пишет декларативно, компи-

лятор превращает это в эффективные обновления DOM-браузера [20].

Преимущество — отсутствие накладных расходов при выполнении, оптимальный код. Недостаток — ограниченность возможностями статического анализа, меньшая динамичность. Сложность локализована в компиляторе и инструментах сборки.

Реактивность в серверных системах и потоках данных

За пределами UI реактивное программирование нашло применение в обработке асинхронных потоков данных, где доминирует push-модель.

1. Reactive Streams и backpressure.

Спецификация Reactive Streams (реализованная в Project Reactor, RxJava, Akka Streams) формализует асинхронную обработку потоков с явным управлением обратным давлением [8, 11]. Когда производитель генерирует данные быстрее, чем потребитель их обрабатывает, система должна координировать скорость через механизм запросов (request/demand). Это push-модель с элементами pull: потребитель явно запрашивает определенное количество элементов, которые источник затем «проталкивает».

Сложность локализована в протоколе взаимодействия между производителем и потребителем. Преимущество — контроль над потреблением ресурсов в распределенных системах. Недостаток — необходимость явно проектировать стратегии backpressure.

2. Реактивные базы данных.

Firebase, RethinkDB и подобные системы предоставляют реактивные запросы — подписки на изменения данных в реальном времени [21, 22]. Клиент подписывается на запрос один раз и автоматически получает обновления при изменении результата на сервере. Это чистая push-модель, где база данных активно уведомляет подписчиков.

Сложность перенесена на сервер: необходимость отслеживать активные подписки; эффективно вычислять, какие подписки затронуты изменением; управлять сетевыми соединениями. Преимущество заключается в простоте

клиентского кода. Недостаток — в масштабируемости сервера при большом количестве подписок.

3. Потокковая обработка данных.

Apache Kafka Streams, Apache Flink представляют модель непрерывной обработки потоков событий. Данные моделируются как бесконечные последовательности, к которым применяются трансформации (map, filter, aggregate). Это push-модель на уровне распределенной системы [23, 24].

Сложность локализована в координации распределенного состояния: партиционирование данных, гарантии доставки, управление временными окнами для агрегаций. Реактивность здесь сталкивается с теми же фундаментальными вопросами, что и в UI-фреймворках: как распространять изменения, как гарантировать согласованность, как управлять ресурсами.

Общая закономерность

Независимо от области применения (UI, серверные системы, базы данных) реактивные системы сталкиваются с одними и теми же фундаментальными компромиссами. Push-модель наиболее уместна там, где важна асинхронность и непредсказуемость событий, pull-модель — для детализации обновлений и синхронности вычислений. Гибридные подходы пытаются совместить преимущества ценой дополнительной концептуальной нагрузки.

СПИСОК ИСТОЧНИКОВ

1. Moseley B., Marks P. Out of the Tar Pit // Proceedings of the BCS Software Practice Advancement Conference (SPA '2006) (Bedfordshire, England, 26–29 March 2006). 66 p. URL: <http://curtclifton.net/papers/MoseleyMarks06a.pdf> (дата обращения: 26.04.2026).
2. Java Concurrency на практике = Java Concurrency in Practice / Б. Гетц [и др.]; пер. с англ. А. Логунова. СПб.: Питер, 2026. 464 с.
3. Meijer E. The Curse of the Excluded Middle // Communications of the ACM. 2014. Vol. 57, iss. 6. Pp. 50–55. DOI: 10.1145/2605176
4. Elliott C., Hudak P. Functional Reactive Animation // ACM SIGPLAN Notices. 1997. Vol. 32, no. 8. Pp. 263–273. DOI: 10.1145/258948.258973
5. Salvaneschi G., Mezini M. Reactive Behavior in Object-Oriented Applications: An Analysis and a Research Roadmap // Proceedings of the 12th Annual International Conference on Aspect-Oriented Software Development (AOSD '13) (Fukuoka, Japan, 24–29 March 2013) / H. Masuhara [et al.] (eds). New York: Association for Computing Machinery, 2013. Pp. 37–48. DOI: 10.1145/2451436.2451442

Заключение

Таким образом, реактивное программирование эволюционировало от формальной математической модели Эллиотта до множества практических реализаций, каждая из которых решает одну и ту же задачу — автоматическое распространение изменений, но делает разные компромиссы.

Push-модель (Reactive Extensions, Reactive Streams) дает контроль и асинхронность ценой явного управления подписками и координации обновлений. Pull-модель (Signals) дает автоматизм и эффективность ценой ограничений при работе с асинхронностью. Компиляция (Svelte) переносит сложность в инструменты сборки ценой ограниченной статического анализа.

Фундаментальная сложность управления изменениями во времени не устраняется — она перераспределяется между кодом разработчика, механизмами выполнения системы и инструментами разработки. Понимание этого распределения важно для практического применения реактивных моделей: разработчик должен учитывать не только возможности выбранного инструмента, но и связанную с ним зону ответственности, а также возможные классы ошибок.

Выбор модели реактивности определяет архитектурные ограничения системы и способ мышления разработчика. Тезис, сформулированный во введении, получает конкретное подтверждение в анализе современных реализаций.

6. Wan Z., Taha W., Hudak P. Real-Time FRP // ACM SIGPLAN Notices. 2001. Vol. 36, no. 10. Pp. 146–156. DOI: 10.1145/507635.507654
7. Czaplicki E., Chong S. Asynchronous Functional Reactive Programming for GUIs // ACM SIGPLAN Notices. 2013. Vol. 48, no. 9. Pp. 411–422. DOI: 10.1145/2462156.2462161
8. Нуркевич Т., Кристенсен Б. Реактивное программирование с применением RxJava. Разработка асинхронных событийно-ориентированных приложений = Reactive Programming with RxJava: Creating Asynchronous, Event-based Application / пер. с англ. А. А. Слинкина. М.: ДМК Пресс, 2017. 358 с.
9. ReactiveX Observable Documentation. URL: <http://reactivex.io/documentation/observable.html> (дата обращения: 25.04.2026).
10. RxJS API Docs. URL: <http://rxjs.dev/guide/overview> (дата обращения: 25.04.2026).
11. Project Reactor Documentation. URL: <http://projectreactor.io/docs> (дата обращения: 25.04.2026).
12. Кун Р., Ханафи Б., Аллен Дж. Реактивные шаблоны проектирования = Reactive Design Patterns / пер. с англ. С. Черникова. СПб.: Питер, 2018. 416 с.
13. Elliott C. M. Push-Pull Functional Reactive Programming // Haskell 2009: Proceedings of the Second ACM SIGPLAN Symposium on Haskell (Edinburgh, Scotland, 03 September 2009) / S. Weirich (ed.). New York: Association for Computing Machinery, 2009. Pp. 25–36. DOI: 10.1145/1596638.1596643
14. Кумар Т. React. К вершинам мастерства: создание быстрых, производительных и интуитивно понятных веб-приложений = Fluent React: Build Fast, Performant, and Intuitive Web Applications / пер. с англ. Астана: АЛИСТ, 2025. 368 с.
15. React DOM: React Reference Overview. URL: <http://react.dev/reference/react> (дата обращения: 26.04.2026).
16. What is Angular? URL: <http://angular.dev/overview> (дата обращения: 26.04.2026).
17. Angular Signals: In-depth Guides. URL: <http://angular.dev/guide/signals> (дата обращения: 26.04.2026).
18. Solid Overview. URL: <http://docs.solidjs.com> (дата обращения: 26.04.2026).
19. Signals: Preact Guide. URL: <http://preactjs.com/guide/v10/signals> (дата обращения: 26.04.2026).
20. Svelte Overview: Svelte Docs. URL: <http://svelte.dev/docs/svelte/overview> (дата обращения: 26.04.2026).
21. Wingerath W., Ritter N., Gessert F. Real-Time and Stream Data Management: Push-Based Data in Research and Practice. Cham: Springer, 2019. 86 p. DOI: 10.1007/978-3-030-10555-6
22. Changefeeds in RethinkDB. URL: <http://rethinkdb.com/docs/changefeeds/python> (дата обращения: 26.04.2026).
23. Kafka Streams: Apache Kafka Docs. URL: <http://kafka.apache.org/43/streams> (дата обращения: 26.04.2026).
24. Apache Flink Documentation. URL: <http://nightlies.apache.org/flink/flink-docs-stable> (дата обращения: 26.04.2026).

Дата поступления: 18.05.2026

Решение о публикации: 22.05.2026

Reactive Programming as an Evolution of Architectural Trade-offs in Managing Changes

- Darya A. Tsukanova** — 4th year Bachelor’s Degree Student in 09.03.01 Informatics and Computer Technology. Research interests: reactive programming, development of high-load information systems, programming paradigms. E-mail: GinD.Nova@yandex.ru
- Andrey V. Zabrodin** — PhD in History, Associate Professor of the “Information and Computing Systems” Department. Research interests: information systems, data analytics, programming paradigms, cloud technologies. E-mail: zabrodin@pgups.ru

Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky ave., Saint Petersburg, 190031, Russia

For citation: Tsukanova D.A., Zabrodin A.V. Reactive Programming as an Evolution of Architectural Trade-offs in Managing Changes, *Intellectual Technologies on Transport*, 2026, No. 2 (46), Pp. 54–63. DOI: 10.20295/2413-2527-2026-246-54-63 (In Russian)

Abstract. *The paper presents a study of the evolution of reactive programming models from the theoretical foundations of functional reactive programming to modern practical implementations in reactive thread libraries and UI frameworks. **Purpose:** to systematize approaches to dependency management and dissemination of changes in software systems. **Methods:** comparative analysis of architectural solutions in FRP, reactive flow libraries and UI frameworks. **Results:** it is shown that reactive models do not eliminate the complexity of managing changes over time, but redistribute it between the developer's code, execution mechanisms, and development tools. The key differences between the push model based on explicit event propagation and subscription management and the pull model using automatic dependency tracking have been identified. **Practical significance:** it consists in clarifying the architectural trade-offs of various reactivity models in the development of user interfaces and server-side data processing systems.*

Keywords: *reactive programming, functional reactive programming, push model, pull model, dependency management, architectural patterns*

REFERENCES

1. Moseley B., Marks P. Out of the Tar Pit, *Proceedings of the BCS Software Practice Advancement Conference (SPA '2006)*, Bedfordshire, England, March 26–29, 2006. 66 p. Available at: <http://curtclifton.net/papers/MoseleyMarks06a.pdf> (accessed: April 26, 2026).
2. Goetz B., et al. Java Concurrency na praktike [Java Concurrency in Practice]. Saint Petersburg, Piter Publishing House, 2026, 464 p. (In Russian)
3. Meijer E. The Curse of the Excluded Middle, *Communications of the ACM*, 2014, vol. 57, iss. 6, pp. 50–55. DOI: 10.1145/2605176
4. Elliott C., Hudak P. Functional Reactive Animation, *ACM SIGPLAN Notices*, 1997, vol. 32, no. 8, pp. 263–273. DOI: 10.1145/258948.258973
5. Salvaneschi G., Mezini M. Reactive Behavior in Object-Oriented Applications: An Analysis and a Research Roadmap. In: Masuhara H., et al. (eds) *Proceedings of the 12th Annual International Conference on Aspect-Oriented Software Development (AOSD '13)*, Fukuoka, Japan, March 24–29, 2013. New York, Association for Computing Machinery, 2013, pp. 37–48. DOI: 10.1145/2451436.2451442
6. Wan Z., Taha W., Hudak P. Real-Time FRP, *ACM SIGPLAN Notices*, 2001, vol. 36, no. 10, pp. 146–156. DOI: 10.1145/507635.507654
7. Czaplicki E., Chong S. Asynchronous Functional Reactive Programming for GUIs, *ACM SIGPLAN Notices*, 2013, vol. 48, no. 9, pp. 411–422. DOI: 10.1145/2462156.2462161
8. Nurkiewicz T., Christensen B. Реактивное программирование с применением RxJava. Разработка асинхронных событийно-ориентированных приложений [Reactive Programming with RxJava: Creating Asynchronous, Event-based Application]. Moscow, DMK Press Publishing House, 2017, 358 p. (In Russian)
9. ReactiveX Observable Documentation. Available at: <http://reactivex.io/documentation/observable.html> (accessed: April 25, 2026).
10. RxJS API Docs. Available at: <http://rxjs.dev/guide/overview> (accessed: April 25, 2026).
11. Project Reactor Documentation. Available at: <http://projectreactor.io/docs> (accessed: April 25, 2026).
12. Kuhn R., Hanafee B., Allen J. Реактивные шаблоны проектирования [Reactive Design Patterns]. Saint Petersburg, Piter Publishing House, 2018, 416 p. (In Russian)

13. Elliott C. M. Push-Pull Functional Reactive Programming. In: *Weirich S. (ed.) Haskell 2009: Proceedings of the Second ACM SIGPLAN Symposium on Haskell*, Edinburgh, Scotland, September 03, 2009. New York, Association for Computing Machinery, 2009, pp. 25–36. DOI: 10.1145/1596638.1596643
14. Kumar T. React. K vershinam masterstva: sozдание bystrykh, proizvoditelnykh i intuitivno ponyatnykh veb-prilozheniy [Fluent React: Build Fast, Performant, and Intuitive Web Applications]. Astana, ALIST Publishing House, 2025, 368 p. (In Russian)
15. React DOM: React Reference Overview. Available at: <http://react.dev/reference/react> (accessed: April 26, 2026).
16. What is Angular? Available at: <http://angular.dev/overview> (accessed: April 26, 2026).
17. Angular Signals: In-depth Guides. Available at: <http://angular.dev/guide/signals> (accessed: April 26, 2026).
18. Solid Overview. Available at: <http://docs.solidjs.com> (accessed: April 26, 2026).
19. Signals: Preact Guide. Available at: <http://preactjs.com/guide/v10/signals> (accessed: April 26, 2026).
20. Svelte Overview: Svelte Docs. Available at: <http://svelte.dev/docs/svelte/overview> (accessed: April 26, 2026).
21. Wingerath W., Ritter N., Gessert F. Real-Time and Stream Data Management: Push-Based Data in Research and Practice. Cham, Springer, 2019, 86 p. DOI: 10.1007/978-3-030-10555-6
22. Changefeeds in RethinkDB. Available at: <http://rethinkdb.com/docs/changefeeds/python> (accessed: April 26, 2026).
23. Kafka Streams: Apache Kafka Docs. Available at: <http://kafka.apache.org/43/streams> (accessed: April 26, 2026).
24. Apache Flink Documentation. Available at: <http://nightlies.apache.org/flink/flink-docs-stable> (accessed: April 26, 2026).

Received: May 18, 2026

Accepted: May 22, 2026

УДК 004.65

Исследование влияния фрагментации данных на производительность чтения/записи и методы ее устранения в MySQL (InnoDB)

Черкасов Сергей Дмитриевич¹ — студент бакалавриата 2-го курса направления 38.03.05 «Бизнес-информатика». Научные интересы: информационные системы, разработка и исследование характеристик приложений для работы с базами данных. E-mail: cherkasovsergeidx@mail.ru

Хомоненко Анатолий Дмитриевич^{1,2} — доктор техн. наук, профессор, профессор кафедры «Информационные и вычислительные системы»; профессор кафедры математического и программного обеспечения. Научные интересы: информационные системы, обработка больших данных, вероятностное моделирование геоинформационных систем, генетические алгоритмы, информационная безопасность. E-mail: khomon@mail.ru

¹Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

²Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Черкасов С. Д., Хомоненко А. Д. Исследование влияния фрагментации данных на производительность чтения/записи и методы ее устранения в MySQL (InnoDB) // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 64–74. DOI: 10.20295/2413-2527-2026-246-64-74

Аннотация. *Предлагается экспериментально обоснованный подход к оценке влияния фрагментации данных на производительность СУБД MySQL (InnoDB) и выбору метода ее устранения. Цель: исследование зависимости времени выполнения операций чтения и записи от степени фрагментации данных и определение наиболее эффективного способа дефрагментации. Методы: анализ архитектуры InnoDB (B+-деревья, page split); создание экспериментального стенда на базе MySQL 8.0 в Docker; генерация тестовой таблицы объемом 500 тыс. записей; моделирование фрагментации через массовые DELETE, UPDATE и хаотичные INSERT; нагрузочное тестирование sysbench; измерение времени запросов, латентности, IOPS и data_free. Результаты: фрагментация замедляет SELECT в 2,5–3,5 раза, INSERT и UPDATE — в 2–3 раза. Рост data_free с 2 до 38 Мб коррелирует с падением TPS на 58%. OPTIMIZE TABLE восстанавливает производительность до 95–98% от исходного уровня. Практическая значимость: методики мониторинга и плановой дефрагментации применимы для транспортных информационных систем с высокой интенсивностью обновлений данных (трекинг, логистика).*

Ключевые слова: *фрагментация данных, MySQL, InnoDB, производительность, page split, optimize table, B+-дерево, data_free, интеллектуальные транспортные системы*

2.3.5 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки); **2.9.8** — интеллектуальные транспортные системы (технические науки)

Введение

Современные информационные системы, особенно в сфере транспорта и логистики, обрабатывают огромные объемы данных в режиме реального времени. Системы мониторинга подвижного состава, трекинга грузов, управления перевозками генерируют терабайты записей ежедневно.

При этом производительность системы управления базами данных (СУБД) становится критическим фактором, определяющим качество работы всего приложения.

Одной из скрытых, но существенных причин деградации производительности является *фрагментация данных* [1]. В отличие от явных ошибок (неоптимальные запросы, отсутствие индексов) фрагментация накапливается незаметно в процессе эксплуатации. Массовые удаления (например, очистка устаревших трекинговых данных), частые обновления, хаотичные вставки записей приводят к тому, что физическая структура данных на диске перестает соответствовать логической.

Движок InnoDB — наиболее распространенный в MySQL (по умолчанию с версии 5.5) [2]. Он используется в высоконагруженных проектах, включая онлайн-торговлю, банковские системы, а также в транспортных информационных системах. Однако, несмотря на его популярность, влияние фрагментации на реальную эксплуатационную производительность изучено недостаточно, а методы борьбы с ней часто применяются хаотично или игнорируются.

Особенно остро проблема фрагментации стоит в OLTP-системах (online transaction processing), где важна высокая скорость выполнения операций в режиме реального времени. В статье исследу-

ется влияние фрагментации данных на производительность операций чтения и записи в MySQL (InnoDB) для определения эффективных способов ее устранения, в том числе применительно к системам обработки данных на транспорте. Для этого решались следующие задачи:

- рассмотрение архитектуры движка InnoDB, механизмов хранения и индексации [1, 3];
- анализ видов и причины возникновения фрагментации данных и индексов [2, 4, 5];
- разработка экспериментального стенда и тестовой базы данных;
- проведение контролируемого эксперимента с фиксацией ключевых метрик производительности в трех состояниях (без фрагментации, с фрагментацией, после оптимизации);
- сравнение полученных метрик и интерпретация результатов [6, 7];
- формулирование практических рекомендаций для эксплуатации высоконагруженных систем, включая транспортные [8–10].

Теоретическая основа: архитектура InnoDB и механизмы фрагментации

Для понимания природы фрагментации приведем кратко основные архитектурные решения InnoDB [1, 3]. Схема архитектуры хранения данных InnoDB показана на рис. 1.

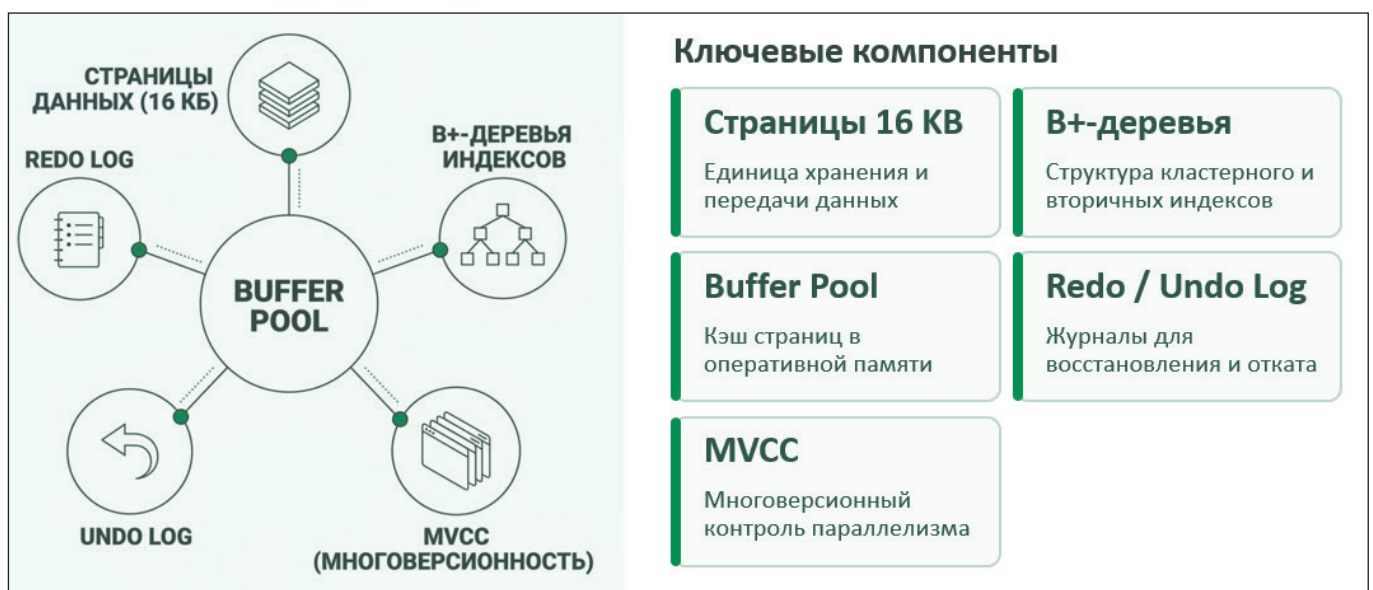


Рис. 1. Схема архитектуры хранения данных InnoDB

Полная схема архитектуры включает табличное пространство, сегменты, экстенды, страницы, буферный пул и журналы Redo/Undo [1].

Страничная организация данных

InnoDB хранит данные и индексы на диске блоками фиксированного размера — страницами (pages), по умолчанию 16 Кб [1, 2]. Каждая страница содержит записи таблицы или узлы индексов. При выполнении запросов страницы подгружаются в буферный пул (Buffer Pool) в оперативной памяти [3]. Эффективность работы системы напрямую зависит от того, насколько плотно заполнены страницы и насколько последовательно они расположены на диске [5].

B+-деревья

И первичный ключ (кластерный индекс), и вторичные индексы в InnoDB реализованы в виде B+-деревьев [1, 2]. Листовые страницы кластерного индекса содержат полные строки данных. Листовые страницы вторичных индексов содержат значения индексированных полей и значения первичного ключа [11]. B+-деревья обеспечивают логарифмическую сложность поиска, но чувствительны к степени заполнения страниц [3, 5].

Кластерный индекс и последовательная вставка

Если первичный ключ генерируется как AUTO_INCREMENT, новые записи вставляются в конец B+-дерева — в правую крайнюю страницу [2, 4]. Это минимизирует разбиения страниц. Если же первичный ключ не является последовательным (например, UUID, случайные числа, хеши), новые записи могут попадать в середину дерева, вызывая частые разбиения [5]. Для транспортных систем, использующих распределенные идентификаторы (GUID) для синхронизации между филиалами, это особенно актуально [8, 10].

Page split (разбиение страницы) — ключевой механизм фрагментации

Когда вставка новой записи происходит в уже заполненную страницу, InnoDB выполняет разби-

ение страницы: примерно половина записей перемещается на новую страницу [1, 4]. В результате:

- образуется пустое пространство (фрагментация внутри страниц) — страница становится заполненной менее чем на 50–60 % [2];
- нарушается физическая смежность логически последовательных записей [5];
- увеличивается глубина B+-дерева и количество операций ввода-вывода [6, 11];
- снижается эффективность буферного пула, так как в кэш попадает больше пустых страниц [3].

Именно page split является ключевым механизмом, порождающим фрагментацию индексов и данных [4, 5]. Особенно интенсивно page split происходят:

- при массовых INSERT с непоследовательными ключами [2];
- при UPDATE, увеличивающих размер строки (например, заполнение поля VARCHAR) [11];
- при DELETE, создающих дыры, которые затем хаотично заполняются [4].

Виды и причины фрагментации в InnoDB

В контексте InnoDB выделяют два основных типа фрагментации: данных и индексов [2, 4].

Фрагментация данных возникает в кластерном индексе из-за частых операций DELETE и UPDATE, изменяющих размер строк [1, 5]. Удаленные строки помечаются как мертвые (внутренний флаг DELETE_MARK), но занимаемое ими физическое пространство не возвращается операционной системе немедленно [2]. При последующих вставках InnoDB может переиспользовать эти дыры, однако это приводит к несвязному расположению данных — строки, логически принадлежащие одному диапазону, физически разбросаны по разным страницам [4, 6].

Фрагментация индексов характерна для вторичных индексов и также возникает при page split, обновлениях ключей и удалениях [1, 11]. Фрагментированный индекс имеет много частично заполненных страниц (коэффициент заполнения может падать до 30–40 %), что увеличивает количество страниц, которые нужно прочитать при сканировании индекса [4, 7]. Для запросов, использую-

щих вторичный индекс с последующим доступом к кластерному (так называемый обратный поиск), фрагментация особенно критична [2, 6].

На рис. 2 для наглядности приведены три варианта фрагментации: внутренняя, внешняя и индексная.

Внутренняя фрагментация возникает внутри выделенного блока памяти, когда выделенное пространство используется не полностью, оставляя пустоты (гэпы), что снижает плотность хранения. Внешняя фрагментация проявляется на уровне расположения данных на диске, когда строки таблицы или блоки данных разбросаны по несмежным страницам, что приводит к избыточным перемещениям головок диска. Индексная фрагментация характерна для структур B-tree: при вставке записей не по порядку.

Основные причины фрагментации в реальных системах

1. Массовые DELETE (особенно 30% и более записей), например, периодическая очистка устаревших трекингových данных или удаление завершенных заказов [8, 9].
2. UPDATE индексируемых полей фактически выполняется как DELETE + INSERT, что порождает фрагментацию сразу в двух индексах [2, 5].
3. Хаотичные INSERT — вставка записей с неупорядоченными значениями первичного ключа (характерно для распределенных транспортных систем, где ID генерируются на разных узлах) [8, 10].
4. Отсутствие плановой дефрагментации в течение длительной эксплуатации (месяцы и годы) [4, 7].

Экспериментальная методика

Для получения объективных результатов разработан контролируемый эксперимент. Эксперимент проводился в трех состояниях: без фрагментации (эталон), с искусственно созданной фрагментацией и после выполнения оптимизации [6].

Окружение

Контролируемый эксперимент проводился с использованием в окружении следующих составляющих и характеристик:

- СУБД: MySQL 8.0.33 (движок InnoDB) [1];
- развертывание: Docker (изолированная среда, контейнер с 4 vCPU и 8 Гб RAM);
- тестовая база данных: fragmentation_test;
- основная таблица: orders (моделирует таблицу заказов или записей трекинга) [8];
- объем данных: более 500 тыс. записей.

Структура таблицы:

```

sql
CREATE TABLE orders (
  id BIGINT PRIMARY KEY AUTO_INCREMENT,
  user_id INT NOT NULL,
  product_id INT NOT NULL,
  amount DECIMAL(10,2),
  status VARCHAR(20),
  created_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
  INDEX idx_user(user_id),
  INDEX idx_product(product_id)
) ENGINE=InnoDB;
    
```

Выбор такой структуры обусловлен тем, что она типична для OLTP-систем [2, 3]: есть первичный



Рис. 2. Варианты фрагментации

ключ-автоинкремент, два внешних логических поля для фильтрации, числовой атрибут (amount) и статус. Вторичные индексы позволяют оценить влияние фрагментации на индексные сканирования [6, 11].

Инструменты нагрузочного тестирования и анализа:

- sysbench — для генерации смешанной OLTP-нагрузки [6] (подготовка, запуск, очистка);
- INFORMATION_SCHEMA — для мониторинга data_length, index_length, data_free [1];
- EXPLAIN ANALYZE — для анализа реального плана выполнения запросов и оценки стоимости [2];
- системные утилиты: iostat, vmstat — для оценки I/O [5];
- MySQL Workbench — для выполнения SQL-скриптов и визуализации.

Этапы эксперимента

Этап 1. Создание эталонного состояния (без фрагментации):

- создание таблицы и индексов [1];
- заполнение таблицы (500 тыс. строк) с помощью хранимой процедуры generate_orders();
- выполнение OPTIMIZE TABLE для гарантии отсутствия фрагментации [4];
- замер метрик производительности [6].

Этап 2. Моделирование фрагментации:

- последовательное выполнение операций:

- удаление 33% записей (DELETE FROM orders WHERE id % 3 = 0);
- обновление 20% записей (UPDATE orders SET status = 'archived' WHERE id % 5 = 0);
- вставка 100 тыс. новых записей (частично заполняют образовавшиеся дыры, создавая хаотичную структуру) [4, 5];
- замер метрик производительности [6].

Этап 3. Оптимизация (устранение фрагментации) [4, 7]:

- выполнение OPTIMIZE TABLE orders;
- повторный замер метрик [6].

Измеряемые метрики:

- время выполнения SELECT (точечные запросы по индексу user_id и диапазонные по id) [2, 6];
- скорость INSERT (операций в секунду) [5];
- скорость UPDATE неиндексируемого поля (amount) и индексируемого поля (status) [11];
- латентность (p95, p99 по данным sysbench) [6];
- количество операций чтения / записи на диск (IOPS) [5];
- размер таблицы и индексов на диске (через INFORMATION_SCHEMA) [1];
- показатель data_free (незанятое пространство внутри таблицы) [4].

Результаты эксперимента

Состояние таблицы по данным INFORMATION_SCHEMA

На рис. 3 приведен результат запроса к INFORMATION_SCHEMA.TABLES.

The screenshot shows a MySQL Workbench window with a query editor and a result grid. The query is:

```

1 • SELECT
2     table_name,
3     data_length,
4     index_length,
5     data_free
6 • FROM information_schema.tables
7 • WHERE table_name = 'orders';
    
```

The result grid displays the following data:

table_name	data_length	index_length	data_free
orders	107479040	31457280	125829120
NULL	NULL	NULL	NULL

Рис. 3. Проверка состояния таблицы с помощью запроса

Ключевые показатели данных и индексов [1, 4] приведены в табл. 1.

Рост data_free с 2 до 38 Мб свидетельствует о появлении значительного количества пустого пространства внутри страниц — это прямой индикатор фрагментации [2, 4].

Влияние на время выполнения запросов

Измерения проводились для трех типовых запросов:

- запрос 1 (SELECT по вторичному индексу):

```
sql
SELECT * FROM orders WHERE user_id = 100;
```

- запрос 2 (INSERT новой записи):

```
sql
INSERT INTO orders (user_id, product_id, amount, status)
VALUES (1001, 501, 250.50, 'new');
```

- запрос 3 (UPDATE статуса):

```
sql
UPDATE orders SET status = 'processed' WHERE
id = 1000;
```

Результаты экспериментального оценивания среднего времени выполнения запросов приведены в табл. 2.

Нагрузочное тестирование (sysbench)

С помощью sysbench проводилось смешанное тестирование oltp_read_write с 8 потоками в течение 60 с. [6]. Результаты приведены в табл. 3.

Интерпретация:

- TPS упал в 2,37 раза при фрагментации [5, 6];
- латенция P95 выросла в 2,4 раза [6];
- IOPS выросли в 2,6 раза — система вынуждена читать больше разрозненных страниц с диска, несмотря на кэширование [2, 5].

Сравнительный анализ и интерпретация

Основные метрики, используемые для сравнительного анализа и интерпретации результатов, приведены в табл. 4.

Ключевые выводы из измерений [4–6]:

1. SELECT страдает сильнее всего — замедление в 3,2 раза. Причина: сканирование фрагментированного индекса idx_user требует чтения большего количества страниц, многие из которых заполнены менее чем на 50% [2, 4].
2. INSERT и UPDATE замедляются в 2–3 раза. Основной вклад вносит механизм page split: при вставке новых записей в неплотные страницы

Таблица 1

Ключевые показатели данных и индексов

Состояние	data_length, Мб	index_length, Мб	data_free, Мб
Без фрагментации	~180	~45	~2
С фрагментацией	~195	~52	~38
После OPTIMIZE	~178	~44	~1

Таблица 2

Среднее время выполнения запросов

Состояние БД	SELECT (user_id), мс	INSERT, мс	UPDATE (status), мс
Без фрагментации	12	7	10
С фрагментацией	38	21	28
После OPTIMIZE	14	8	11

Таблица 3

Результаты нагрузочного тестирования

Состояние БД	TPS, транзакций/с	P95 latency, мс	IOPS (чтение)
Без фрагментации	185	28	340
С фрагментацией	78	67	890
После OPTIMIZE	179	30	365

Таблица 4

Основные метрики для сравнительного анализа и интерпретации результатов

Метрика	Без фрагментации	С фрагментацией	После OPTIMIZE
SELECT (по индексу), мс	12	38 (↑ 217%)	14
INSERT, мс	7	21 (↑ 200%)	8
UPDATE (индекс), мс	10	28 (↑ 180%)	11
TPS (sysbench)	185	78 (↓ 58%)	179
P95 latency, мс	28	67 (↑ 139%)	30
Размер данных + индексов, Мб	225	247 (↑ 10%)	222
data_free, Мб	2	38 (↑ 1800%)	1

InnoDB вынуждена выполнять дополнительные операции перераспределения [1, 5].

3. OPTIMIZE TABLE восстанавливает производительность до 95–98% от исходного уровня [4, 7]. Небольшое остаточное отставание связано с тем, что после оптимизации структура данных становится даже более плотной, чем в исходной (за счет устранения начальных дыр) [2].

4. data_free — наиболее наглядный индикатор фрагментации [4]. Его рост с 2 до 38 Мб коррелирует с падением TPS на 58% [6]. Рекомендуется проводить дефрагментацию при data_free > 20% от data_length.

На рис. 4 приведена диаграмма, визуализирующая замедление всех типов операций на фрагментированной таблице и полное восстановление после OPTIMIZE TABLE.

Наиболее критичное замедление наблюдается для операций чтения, что объясняется сканированием фрагментированных страниц вторичных индексов и ростом числа случайных операций ввода-вывода [1, 11]. В то же время выполнение команды OPTIMIZE TABLE позволяет практически полностью восстановить исходную производительность.

Методы устранения фрагментации
OPTIMIZE TABLE

Команда OPTIMIZE TABLE orders перестраивает таблицу и все ее индексы, освобождая незанятое пространство и уплотняя данные [1, 4]. Для InnoDB эта команда:

- создает новую пустую таблицу;
- копирует в нее данные из исходной в порядке первичного ключа;

- перестраивает все вторичные индексы;
- удаляет старую таблицу и переименовывает новую [2, 7].

В данном эксперименте OPTIMIZE TABLE сократил время выполнения SELECT в 2,7 раза по сравнению с фрагментированным состоянием, а размер таблицы уменьшил на 10% (с 247 до 222 Мб).

Недостаток метода: требует блокировки таблицы на время выполнения (для InnoDB — не полной блокировки, но значительной нагрузки на I/O) [4, 6]. Рекомендуется проводить в периоды низкой активности системы (например, ночью) [8, 10].

ALTER TABLE ... ENGINE=InnoDB

Альтернативный способ ALTER TABLE orders ENGINE=InnoDB [1] заставляет InnoDB пересоздать таблицу заново аналогично OPTIMIZE

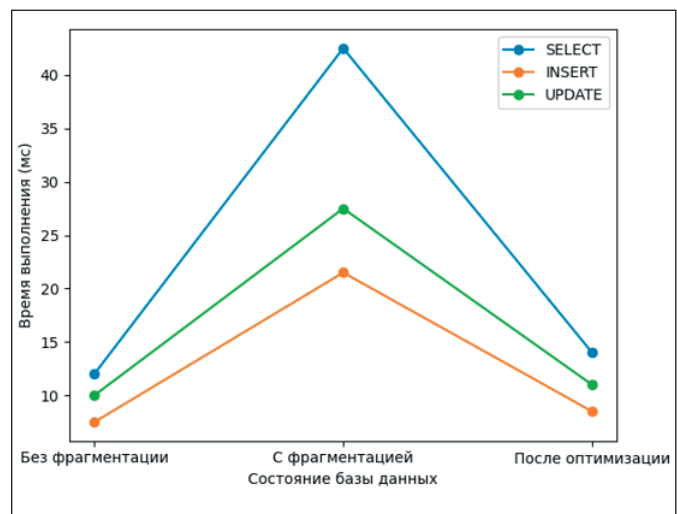


Рис. 4. Влияние фрагментации на производительность MySQL InnoDB

TABLE [4, 7]. По эффективности сопоставим. Может использоваться, если OPTIMIZE по каким-то причинам недоступен (например, в некоторых репликационных конфигурациях) [2].

Профилактические меры (архитектурные решения)

1. Использование последовательных первичных ключей (AUTO_INCREMENT). Избегать случайных значений (UUID, хеши) в качестве PK [2, 5].

2. Настройка innodb_fill_factor (доступен в MariaDB и некоторых сборках MySQL) — резервирование места (например, 90%) на страницах для будущих обновлений, что снижает частоту page split [4].

3. Регулярный мониторинг через INFORMATION_SCHEMA.TABLES — обращать внимание на поле data_free [1, 6].

4. Партиционирование больших таблиц — разделение данных на независимые физические сегменты [3, 10].

Мониторинг фрагментации в production

Простейший SQL-запрос для мониторинга (использовался в эксперименте):

```
sql
SELECT
  table_name,
  ROUND(data_length / 1024 / 1024, 2) AS data_mb,
  ROUND(index_length / 1024 / 1024, 2) AS index_mb,
  ROUND(data_free / 1024 / 1024, 2) AS free_mb,
  ROUND((data_free / data_length) * 100, 2) AS free_pct
FROM information_schema.tables
WHERE table_schema = 'fragmentation_test'
AND table_name = 'orders';
```

Эмпирическое правило [4]: если free_pct > 20% — рекомендуется дефрагментация. При free_pct > 40% дефрагментация обязательна, так как производительность уже существенно деградировала [6, 7].

Выводы

В результате проведенного экспериментального исследования установлено:

1. Фрагментация данных в InnoDB приводит к снижению производительности в 2–3,5 раза в за-

висимости от типа операции [4, 5]. Наиболее сильно страдают запросы чтения, использующие вторичные индексы (замедление до 3,2 раза) [2, 6].

2. Операции записи (INSERT, UPDATE индексируемых полей) замедляются в 2–3 раза из-за увеличения числа page split и большего размера индексов [1, 11].

3. Метод OPTIMIZE TABLE полностью восстанавливает производительность до 95–98% от исходного уровня, а размер таблицы уменьшает на 10–15% за счет уплотнения страниц [4, 7].

4. Показатель data_free из INFORMATION_SCHEMA.TABLES является простым и надежным индикатором фрагментации [1, 4]. Его рост с 2 до 38 Мб в эксперименте коррелирует с падением TPS на 58% [6].

5. Фрагментация накапливается нелинейно — первые массовые DELETE и UPDATE создают дыры, а последующие хаотичные вставки закрепляют фрагментацию, делая структуру максимально неоптимальной [2, 5].

Для транспортных информационных систем, где интенсивность обновлений (статусы заказов, координаты трекинга) и удалений (очистка истории) высока, проблема фрагментации стоит особенно остро [8–10]. Игнорирование плановой дефрагментации приводит к постоянному росту времени отклика и, как следствие, нарушению SLA [6, 10].

Практические рекомендации для транспортных систем

На основе проведенного исследования сформулированы следующие практические рекомендации для администраторов баз данных и разработчиков:

Регламентные меры:

1. Настроить автоматический сбор метрик data_free для всех таблиц с интенсивными изменениями (таблицы заказов, трекинга, статусов) [4, 8].

2. Выполнять дефрагментацию (OPTIMIZE TABLE) в периоды минимальной нагрузки, например с 2:00 до 4:00, с частотой 1 раз в 1–2 недели для высоконагруженных таблиц [6, 10].

3. Использовать партиционирование для больших таблиц (более 10 млн записей), чтобы

дефрагментировать только изменяемые партиции, а не всю таблицу [3, 9].

Архитектурные решения:

1. Применять последовательные первичные ключи (AUTO_INCREMENT) [2, 5]. Если требуется глобальный уникальный идентификатор (например, для синхронизации между филиалами), добавить отдельное поле uuid с индексом, но PK оставить AUTO_INCREMENT [1, 4].

2. Минимизировать обновления индексируемых полей. Если поле часто меняется, рассмотреть возможность выноса его в отдельную таблицу или замены статусной модели на журналирование событий (журнал изменений) [7, 11].

3. Разделять горячие и холодные данные — старые записи перемещать в таблицы-архивы, что снижает фрагментацию в оперативной таблице [8, 10].

Операционный мониторинг:

1. Контролировать размер буферного пула (innodb_buffer_pool_size). При фрагментации эффективность кэша падает, поэтому увеличение буферного пула может временно компенсировать падение производительности, но не устраняет причину [3, 6].

2. Использовать EXPLAIN ANALYZE для запросов, которые стали выполняться дольше обычно. Часто проблема кроется во фрагментированном индексе, а не в самом запросе [2, 5].

СПИСОК ИСТОЧНИКОВ

1. The InnoDB Storage Engine Oracle // MySQL 8.0 Reference Manual. URL: <http://dev.mysql.com/doc/refman/8.0/en/innodb-storage-engine.html> (дата обращения: 13.05.2026).
2. Schwartz B., Zaitsev P., Tkachenko V. High Performance MySQL: Optimization, Backups, and Replication. Third Edition. Sebastopol (CA): O'Reilly Media, 2012. 826 p.
3. Хомоненко А. Д., Цыганков В. М., Мальцев М. Г. Базы данных: учебник для вузов / под ред. А. Д. Хомоненко. 6-е изд., доп. СПб.: Корона-Век, 2009. 736 с.
4. Zaitsev P. MySQL File System Fragmentation Benchmarks // Percona Blog. 22.03.2008. URL: <http://www.percona.com/blog/mysql-file-system-fragmentation-benchmarks> (дата обращения: 13.05.2026).
5. Huang S.-M., Chang L.-P. Exploiting Page Correlations for Write Buffering in Page-Mapping Multichannel SSDs // ACM Transactions on Embedded Computing Systems. 2016. Vol. 15, iss. 1. Art. no. 12. 25 p. DOI: 10.1145/2815622
6. Малютин А. Г., Лаврухин А. А., Окишев А. С. Архитектурные аспекты реализации корпоративной информационной системы мониторинга и учета ресурсов // Известия Транссиба. 2017. № 4 (32). С. 130–141.
7. Павлов Д. В. Реляционная распределенная система управления базами данных с автоматической масштабируемостью // Вестник Уфимского государственного авиационного технического университета. 2012. Т. 16, № 3 (48). С. 143–152.
8. Малыгин Д. С. Проблемы производительности реляционных баз данных в распределенных архитектурах и стратегии их решения // Современные наукоемкие технологии. 2024. № 10. С. 61–71. DOI: 10.17513/snt.40173

Заключение

Фрагментация данных в MySQL (InnoDB) является скрытым, но управляемым фактором деградации производительности [4, 5]. Как показывает эксперимент, ее негативное влияние на чтение и запись может быть весьма существенным: замедление в 2–3,5 раза. Однако современные средства (OPTIMIZE TABLE, ALTER TABLE ... ENGINE=InnoDB) позволяют эффективно восстанавливать производительность [4, 7], а превентивные архитектурные решения (последовательный первичный ключ, мониторинг data_free) — долгое время поддерживать базу данных в оптимальном состоянии [2, 6].

Особую значимость исследование имеет для интеллектуальных транспортных систем, где высока интенсивность обновлений (GPS-треки, статусы заказов, логистические события) и удалений (очистка истории) [8–10]. Внедрение регулярной дефрагментации должно стать обязательной частью регламента эксплуатации таких систем [6].

Результаты работы могут быть полезны администраторам баз данных, разработчикам и техническим архитекторам, применяющим MySQL в производственных средах с высокой интенсивностью изменений данных [3, 4, 7].

9. Харченко П. А. Совершенствование системы и технологии эксплуатации подвижного состава на основе анализа информации средств регистрации параметров движения, диагностики и мониторинга технического состояния локомотива // Вестник Ростовского государственного университета путей сообщения. 2025. № 1 (97). С. 201–210. DOI: 10.46973/0201-727X_2025_1_201

10. Безфамильный Д. Г. Концептуальные подходы к оптимизации жизненного цикла разработки программного обеспечения // Современные инновации, системы и технологии. 2026. Т. 6, № 2. С. 1027–1034. DOI: 10.47813/2782-2818-2026-6-2-1027-1034

11. Нихтер Д. Настройка производительности MySQL. Секреты и приемы / пер. с англ. СПб.: БХВ-Петербург, 2023. 340 с.

Дата поступления: 21.05.2026

Решение о публикации: 21.05.2026

Investigation of Data Fragmentation Impact on Read/Write Performance and Methods for Its Elimination in MySQL (InnoDB)

Sergey D. Cherkasov¹

— 2nd year Bachelor’s Degree Student in 38.03.05 Business Informatics. Research interests: information systems, development and research of characteristics of applications for working with databases. E-mail: cherkasovsergeidx@mail.ru

Anatoly D. Khomonenko^{1,2}

— Dr. Sci. in Engineering, Full Professor, Professor of the “Information and Computing Systems” Department; Professor of the Department of Mathematical and Software Engineering. Research interests: information systems, big data processing, probabilistic modeling of geographic information systems, genetic algorithms, information security. E-mail: khomon@mail.ru

¹Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky ave., Saint Petersburg, 190031, Russia

²Mozhaisky Military Aerospace Academy, 13 Zhdanovskaya str., Saint Petersburg, 197198, Russia

For citation: Cherkasov S. D., Khomonenko A. D. Investigation of Data Fragmentation Impact on Read/Write Performance and Methods for Its Elimination in MySQL (InnoDB), *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 64–74. DOI: 10.20295/2413-2527-2026-246-64-74 (In Russian)

Abstract. *An experimentally substantiated approach to assessing the impact of data fragmentation on MySQL (InnoDB) performance and selecting a method for its elimination is proposed. **Purpose:** to study the dependence of read and write operation execution time on the degree of data fragmentation and to determine the most effective defragmentation method. **Methods:** analysis of InnoDB architecture (B+ trees, page split); creation of an experimental testbed based on MySQL 8.0 in Docker; generation of a test table with 500 000 records; simulation of fragmentation through massive DELETE, UPDATE, and chaotic INSERT operations; load testing using sysbench; measurement of query time, latency, IOPS, and data_free. **Results:** fragmentation slows down SELECT by 2,5–3,5 times, INSERT and UPDATE by 2–3 times. An increase in data_free from 2 to 38 MB correlates with a 58% drop in TPS. OPTIMIZE TABLE restores performance to 95–98% of the original level. **Practical significance:** monitoring and scheduled defragmentation techniques are applicable to transport information systems with high data update intensity (tracking, logistics).*

Keywords: *data fragmentation, MySQL, InnoDB, performance, page split, optimize table, B+ tree, data_free, intelligent transport systems*

REFERENCES

1. The InnoDB Storage Engine Oracle, *MySQL 8.0. Reference Manual*. Available at: <http://dev.mysql.com/doc/refman/8.0/en/innodb-storage-engine.html> (accessed: May 13, 2026).
2. Schwartz B., Zaitsev P., Tkachenko V. *High Performance MySQL: Optimization, Backups, and Replication*. Third Edition. Sebastopol (CA), O'Reilly Media, 2012, 826 p.
3. Khomonenko A. D., Tsygankov V. M., Maltsev M. G. *Bazy dannykh: uchebnik dlya vuzov [Databases: A Textbook for Universities]*. Saint Petersburg, Korona-Vek Publishing House, 2009, 736 p. (In Russian)
4. Zaitsev P. *MySQL File System Fragmentation Benchmarks, Percona Blog*. Published online at March 22, 2008. Available at: <http://www.percona.com/blog/mysql-file-system-fragmentation-benchmarks> (accessed: May 13, 2026).
5. Huang S.-M., Chang L.-P. Exploiting Page Correlations for Write Buffering in Page-Mapping Multichannel SSDs, *ACM Transactions on Embedded Computing Systems*, 2016, vol. 15, iss. 1, art. 12, 25 p. DOI: 10.1145/2815622
6. Malyutin A. G., Lavrukhin A. A., Okishev A. S. *Arkhitekturnye aspekty realizatsii korporativnoy informatsionnoy sistemy monitoringa i ucheta resursov [Architectural Aspects of the Implementation of the Corporate Information System for Monitoring and Resources Accounting]*, *Izvestiya Transsiba [Journal of Transsib Railway Studies]*, 2017, no. 4 (32), pp. 130–141. (In Russian)
7. Pavlov D. V. *Relyatsionnaya raspredelennaya sistema upravleniya bazami dannykh s avtomaticheskoy masshtabiruemostyu [Distributed Relational Database Management System with Automatic Scalability]*, *Vestnik Ufimskogo gosudarstvennogo aviatsionnogo tekhnicheskogo Universiteta [Vestnik of Ufa State Aviation Technical University]*, 2012, Vol. 16, no. 3 (48), pp. 143–152. (In Russian)
8. Malygin D. S. *Problemy proizvoditelnosti relyatsionnykh baz dannykh v raspredelennykh arkhitekturakh i strategii ikh resheniya [Performance Issues of Relational Databases in Distributed Architectures and Strategies for Resolution]*, *Sovremennye naukoemkie tekhnologii [Modern High Technologies]*, 2024, no. 10, pp. 61–71. DOI: 10.17513/snt.40173 (In Russian)
9. Kharchenko P. A. *Sovershenstvovanie sistemy i tekhnologii ekspluatatsii podvizhnogo sostava na osnove analiza informatsii sredstv registratsii parametrov dvizheniya, diagnostiki i monitoringa tekhnicheskogo sostoyaniya lokomotiva [Improvement of the System and Technology of Rolling Stock Operation Based on the Analysis of Information from Motion Parameter Registration, Diagnostics and Monitoring of the Technical Condition of the Locomotive]*, *Vestnik Rostovskogo Gosudarstvennogo Universiteta Putej Soobshcheniya*, 2025, no. 1 (97), pp. 201–210. DOI: 10.46973/0201–727X_2025_1_201 (In Russian)
10. Bezfamilnyi D. G. *Kontseptualnye podkhody k optimizatsii zhiznennogo tsikla razrabotki programmno obespечeniya [Conceptual Approaches to Optimizing the Software Development Lifecycle]*, *Sovremennye innovatsii, sistemy i tekhnologii [Modern Innovations, Systems and Technologies]*, 2026, vol. 6, no. 2, pp. 1027–1034. DOI: 10.47813/2782-2818-2026-6-2-1027-1034 (In Russian)
11. Nichter D. *Nastroyka proizvoditelnosti MySQL. Sekrety i priemy [Efficient MySQL Performance: Best Practices and Techniques]*. Saint Petersburg, BHV-Peterburg Publishing House, 2023, 340 p. (In Russian)

Received: May 21, 2026

Accepted: May 21, 2026

УДК 004.942

Компьютерное моделирование динамических систем на фазовой плоскости

Бригаднов Игорь Альбертович¹

— доктор физ.-мат. наук, профессор. Профессор кафедры «Информационные и вычислительные системы». Научные интересы: математическое и компьютерное моделирование, теория принятия решений, численный анализ. E-mail: brigadnov@mail.ru

Морозов Алексей Валентинович²

— кандидат физ.-мат. наук, профессор. Профессор кафедры высшей математики. Научные интересы: математическое и компьютерное моделирование, численный анализ. E-mail: vka@mil.ru

¹Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

²Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Бригаднов И. А., Морозов А. В. Компьютерное моделирование динамических систем на фазовой плоскости // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 75–81. DOI: 10.20295/2413-2527-2026-246-75-81

Аннотация. Исследование динамических систем на фазовой плоскости является одним из распространенных методов теории колебаний и широко используется в инженерной и научной практике. **Цель:** совместить элементы аналитического исследования с компьютерным моделированием. **Результаты:** предлагается и обсуждается расчетно-графическая работа по дисциплине «Моделирование систем», которая читается в ряде вузов страны. **Практическая значимость:** изучение динамических систем способствует развитию интеллекта, креативности и формированию профессиональных компетенций, что повышает мотивацию к научно-исследовательской деятельности активной части студентов.

Ключевые слова: динамическая система на плоскости, фазовые портреты, аналитически-численное исследование

2.3.1 — системный анализ, управление и обработка информации, статистика (технические науки);

1.2.2 — математическое моделирование, численные методы и комплексы программ (технические науки)

Введение

На кафедре «Информационные и вычислительные системы» Петербургского государственного университета путей сообщения Императора Александра I традиционно читается курс «Моделирование систем», видное место в котором занимает анализ динамических систем (ДС) как аналитическими, так и численными методами. Студенты знакомятся с различными способами описания ДС, их классификацией, с элементами качественной (геометриче-

ской) теории дифференциальных уравнений на плоскости, основными понятиями теории устойчивости, дискретными отображениями и их различными техническими приложениями. Важное место занимает изучение возможных структур фазовых портретов ДС, а также их зависимость от параметров. Данные вопросы составляют содержание теории бифуркаций, отражающей закон перехода количества в качество [1]. Курс предусматривает выполнение

студентами расчетно-графической работы (РГР), состоящей из пяти заданий, целью которой является усвоение основных положений теории через анализ геометрических образов. Каждое задание содержит аналитическую часть, основанную на лекциях, и численную, выполненную при помощи открытого отечественного программного продукта WinSet [2]. Результатом РГР является научное исследование, имеющее творческую направленность, поскольку все задания строго индивидуальны.

Линейные динамические системы

В первом задании РГР исследуется автономная линейная ДС с постоянными коэффициентами вида

$$\begin{cases} \dot{x} = ax + by \\ \dot{y} = cx + dy. \end{cases} \quad (1)$$

Анализируя корни характеристического уравнения

$$\lambda^2 - I_1\lambda + I_2 = 0,$$

где $I_1 = a + d$ и $I_2 = ad - bc$ — первый и второй инварианты матрицы ДС, определяется тип единственной стационарной точки, или положения равновесия $(0,0)$, в которой $\dot{x} = \dot{y} = 0$. А именно вычисляются дискриминант $D = I_1^2 - 4I_2$ и корни $\lambda_{1,2} = \frac{1}{2}(I_1 \mp \sqrt{D})$ характеристического уравнения.

Далее в соответствии с классификацией Пуанкаре [1] устанавливается, что:

1. При $D \geq 0$ (пара вещественных корней):
 - если $\lambda_1 < 0$ и $\lambda_2 < 0$, тогда точка $(0,0)$ — устойчивый узел;
 - если $\lambda_1 > 0$ и $\lambda_2 > 0$, тогда точка $(0,0)$ — неустойчивый узел;
 - если $\lambda_1\lambda_2 < 0$, тогда точка $(0,0)$ — седло.
2. При $D < 0$ (пара комплексно-сопряженных корней):
 - если $I_1 < 0$, тогда точка $(0,0)$ — устойчивый фокус;
 - если $I_1 > 0$, тогда точка $(0,0)$ — неустойчивый фокус;
 - если $I_1 = 0$ (чисто мнимые корни), тогда точка $(0,0)$ — центр.

Далее находятся собственные векторы $Z^{(1)}$ и $Z^{(2)}$ для собственных чисел λ_1 и λ_2 соответственно и записывается аналитическое решение в векторном виде:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = C_1 \begin{pmatrix} Z_1^{(1)} \\ Z_2^{(1)} \end{pmatrix} e^{\lambda_1 t} + C_2 \begin{pmatrix} Z_1^{(2)} \\ Z_2^{(2)} \end{pmatrix} e^{\lambda_2 t},$$

где C_1, C_2 — произвольные константы.

После этого находится уравнение фазовых траекторий в неявном виде $F(x, y) = 0$ путем аналитического решения дифференциального уравнения, порожденного ДС (1), при помощи стандартной подстановки $y = xz(x)$:

$$y' = \frac{cx + dy}{ax + by}, \quad xz' + z = \frac{c + dz}{a + bz}.$$

Исследование завершается построением фазового портрета ДС (1) (рис. 1).

Нелинейные динамические системы

Во втором задании РГР исследуется автономная нелинейная ДС вида

$$\begin{cases} \dot{x} = P(x, y) \\ \dot{y} = Q(x, y). \end{cases}$$

После нахождения стационарных точек или положений равновесия как корней системы уравнений $\{P(x, y) = 0, Q(x, y) = 0\}$ производится линеаризация ДС в окрестности каждой стационарной точки и определяется ее тип по аналогии с первым заданием. Линеаризация ДС в окрестности какой-либо стационарной точки (x_*, y_*) имеет стандартный вид:

$$\begin{cases} \dot{x} = \frac{\partial P(x_*, y_*)}{\partial x} x + \frac{\partial P(x_*, y_*)}{\partial y} y = ax + by \\ \dot{y} = \frac{\partial Q(x_*, y_*)}{\partial x} x + \frac{\partial Q(x_*, y_*)}{\partial y} y = cx + dy. \end{cases}$$

Исследование завершается построением фазового портрета (рис. 2).

Например, классическое дифференциальное уравнение Дуффинга [3]

$$\ddot{x} + p_1\dot{x} + p_2x + p_3x^2 + p_4x^3 = 0,$$

играющее в качественной теории обыкновенных дифференциальных уравнений (ОДУ) важную феноменологическую роль, после стандартной замены переменной $y = \dot{x}$ сводится к нелинейной ДС вида

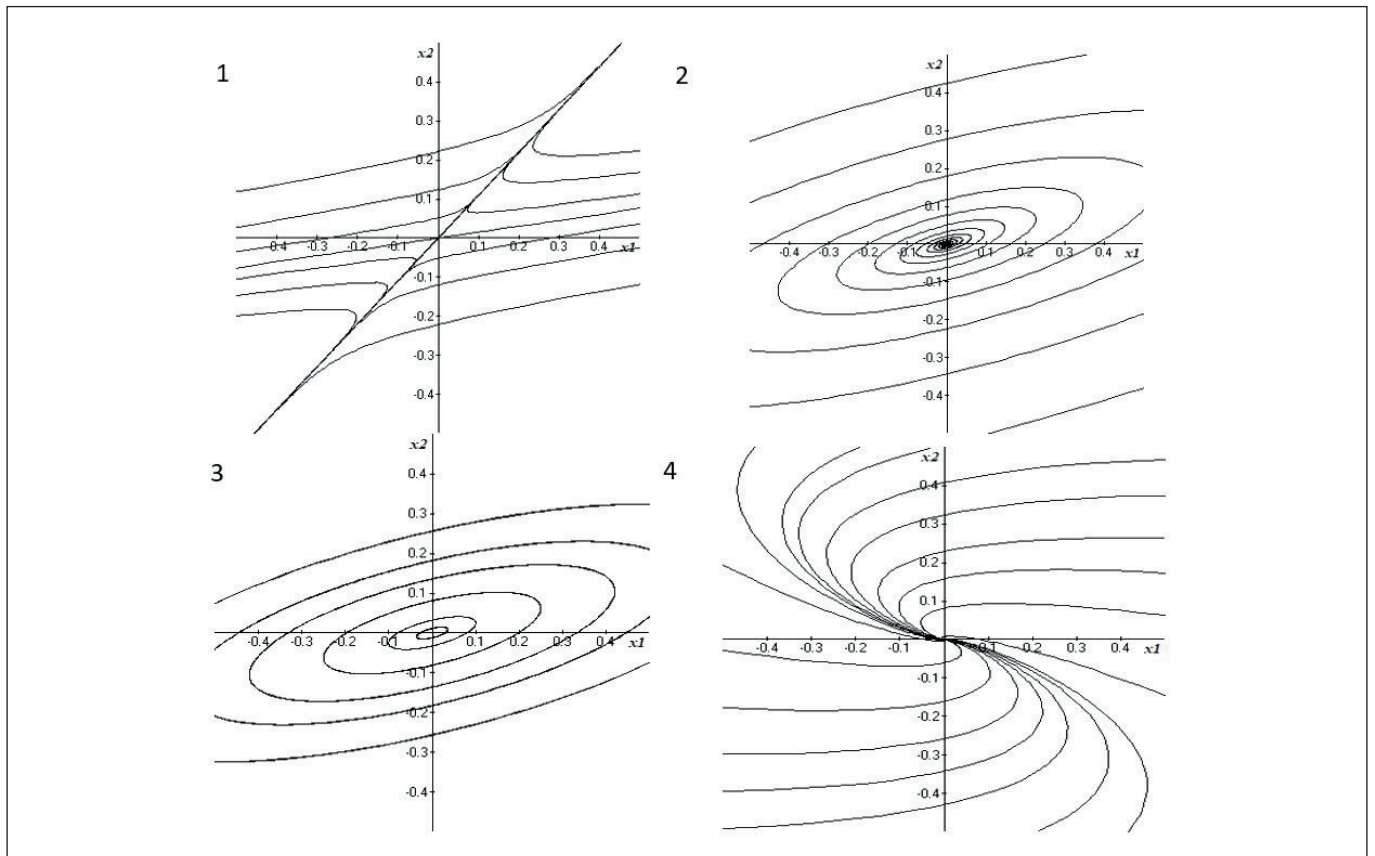


Рис. 1. Некоторые фазовые портреты линейных ДС:

1 — седло; 2 — устойчивый фокус; 3 — центр; 4 — неустойчивый узел

$$\begin{cases} \dot{x} = P(x, y) = y \\ \dot{y} = Q(x, y) = -p_2x - p_3x^2 - p_4x^3 - p_1y. \end{cases}$$

Студенты аналитически находят стационарные точки этой ДС, проводят в них линеаризацию и устанавливают типы положений равновесия. Далее в среде WInSet они строят фазовые портреты консервативных при $p_1 = 0$ и диссипативных при $p_1 > 0$ динамических систем (рис. 2).

Бифуркационный анализ динамических систем

В третьем задании РГР студенты исследуют поведение автономной нелинейной ДС с параметром p . Рассмотрим конкретные примеры такого задания.

Пример 1. Аналитически-численное исследование ДС

$$\begin{cases} \dot{x} = -y - x^2 \\ \dot{y} = px - x^3y \end{cases} \quad (2)$$

показывает, что при $p < 0$ в системе имеется три положения равновесия: седло в начале координат

и два фокуса (один устойчивый, другой неустойчивый). При $p > 0$ существует одно положение равновесия типа «центр» в начале координат, при этом векторное поле обладает симметрией относительно оси ординат (рис. 3).

Таким образом, при переходе слева направо по оси параметра p через критическое значение $p = 0$ качественно меняется поведение ДС, то есть происходит бифуркация слияния трех стационарных точек в одно положение равновесия типа «центр» в начале координат.

Пример 2. Аналитически-численное исследование более сложной ДС

$$\begin{cases} \dot{x} = y - x(px^2 + y^2 - 1) \\ \dot{y} = -px - y(px^2 + y^2 - 1) \end{cases} \quad (3)$$

показывает, что единственная стационарная точка $(0,0)$ при $p < -1$ является седлом, при $-1 < p \leq 0$ — неустойчивым узлом. При этом фазовые траектории асимптотически ложатся слева

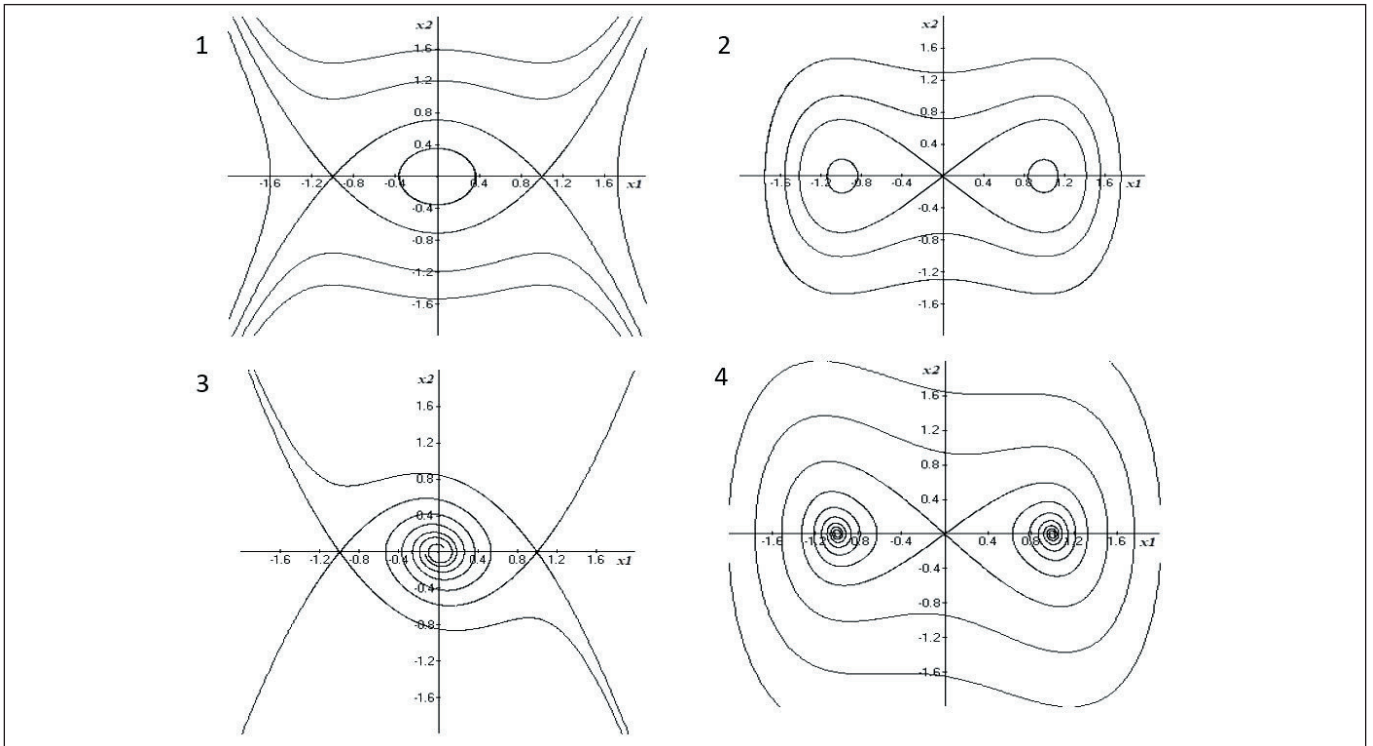


Рис. 2. Фазовые портреты: 1, 2 — консервативные ДС; 3, 4 — диссипативные ДС

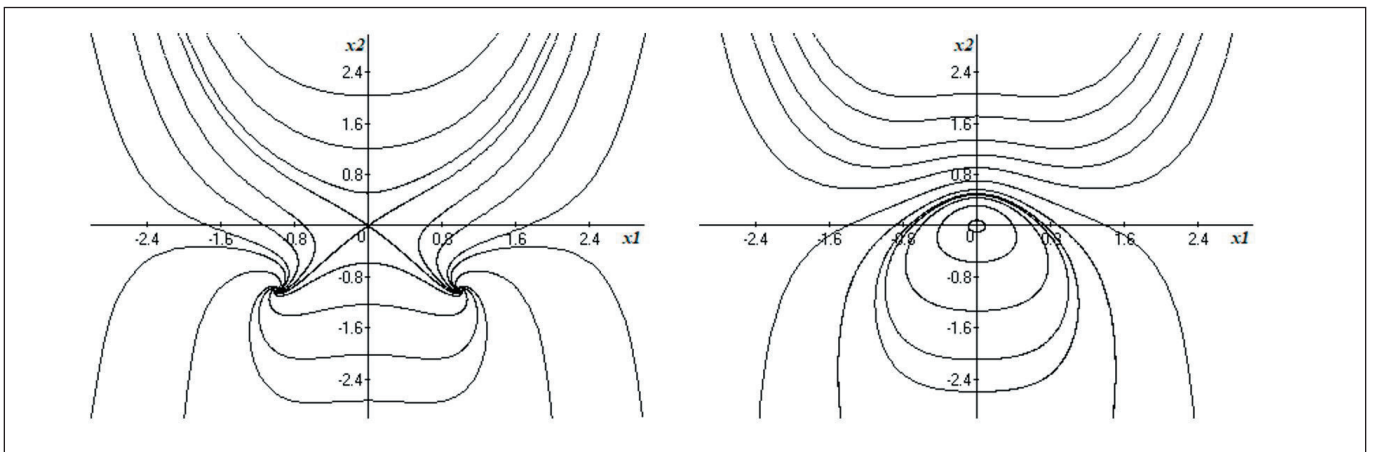


Рис. 3. Фазовые портреты ДС (2) при $p < 0$ (слева) и при $p > 0$ (справа)

на прямую $y = -1$ и справа на прямую $y = 1$. При переходе слева направо по оси параметра p через критическое значение $p = 0$ из пары прямых $y = \pm 1$ рождается предельный цикл — эллипс (рис. 4), что отвечает бифуркации из бесконечности.

Выполнение третьего задания требует изучения дополнительных вопросов, связанных с теорией бифуркаций [4, 5], и здесь основную помощь должен оказать преподаватель.

Аттракторы динамических систем

Четвертое задание РГР является по-настоящему исследовательским и весьма трудоемким, поскольку требует длительной и кропотливой внеаудиторной работы за компьютером. Результатом исследования является карта динамических режимов заданной нелинейной ДС с внешним воздействием [6].

Для примера рассмотрим неавтономный осциллятор Дуффинга с гармоническим воздействием [7]:

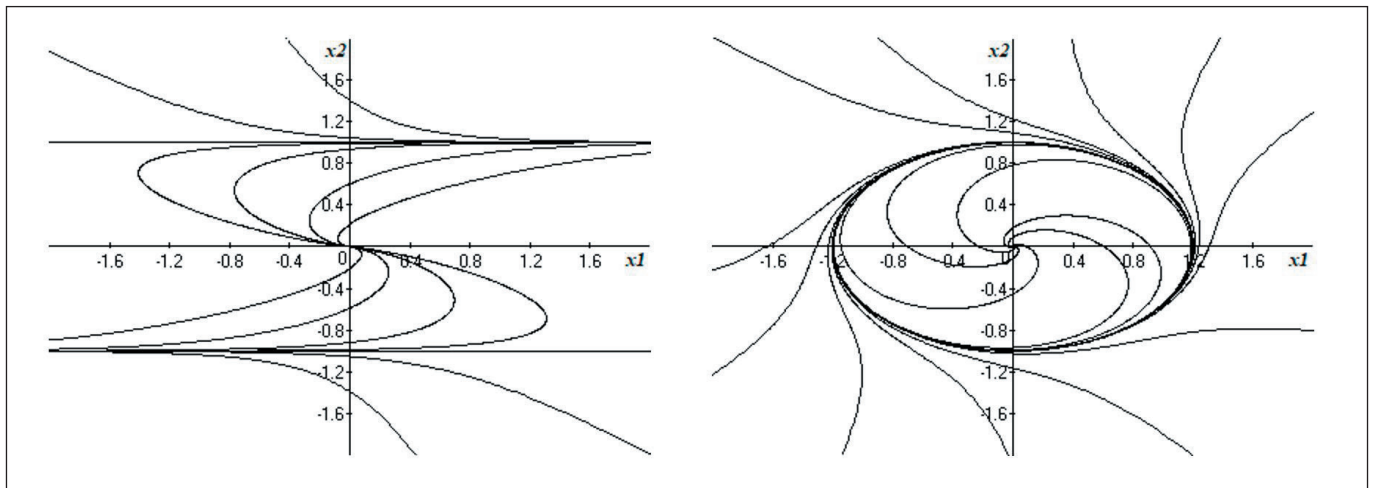


Рис. 4. Фазовые портреты ДС (3) при $p < 0$ (слева) и при $p > 0$ (справа)

$$\begin{cases} \dot{x} = y \\ \dot{y} = -\lambda y - x^3 + A \sin(t). \end{cases}$$

В этой ДС существуют как устойчивые колебания различных периодов и амплитуд, так и хаотический режим при возрастании амплитуды гармонического воздействия, что отчетливо наблюдается на фазовой плоскости, а именно: для параметра $\lambda = 0,35$ с ростом амплитуды A последовательно происходят бифуркации удвоения периодов [7], переходящие при $A = 6,7$ в хаотический режим, или детерминированный хаос [8, 9]. На оси значений параметра A имеются аттракторы — интервалы устойчивого колебания, которые иногда могут перекрываться, что указывает на возможность генерации внешним гармоническим воздействием в этих областях как регулярных колебаний, так и хаотических режимов.

Таким образом, выполнение четвертого задания знакомит студента с современными научно-исследовательскими проблемами, связанными с теорией странных аттракторов и фрактальных структур.

Линейные динамические системы с T -периодическими коэффициентами

В пятом задании исследуется неавтономная линейная ДС (1) с T -периодическими коэффициентами — функциями времени, для которой начало координат является единственной стационарной точкой. Опираясь на теорию Флоке [10], имеющую

большое прикладное значение [11, 12], численно находится матрица монодромии

$$M = \begin{pmatrix} x_1(T) & x_2(T) \\ y_1(T) & y_2(T) \end{pmatrix}$$

в результате интегрирования ДС в среде WinSet на одном общем периоде T для двух начальных условий $\{x_1(0) = 1, y_1(0) = 0\}$ и $\{x_2(0) = 0, y_2(0) = 1\}$. После этого находятся собственные числа матрицы монодромии и по величине их модуля определяется устойчивость или неустойчивость стационарной точки $(0,0)$, а именно: если $|\lambda_1| < 1$ и $|\lambda_2| < 1$, тогда начало координат является устойчивым положением равновесия, иначе — нет. Работа завершается построением фазового портрета ДС для 15–20 периодов, экспериментально подтверждая аналитический результат по анализу динамики неавтономной системы.

Заключение

Материал статьи носит научно-методический характер и может использоваться не только в дисциплине «Моделирование систем». Задания являются комплексными и включают элементы качественной теории ОДУ и теории бифуркаций, слабо отраженные в современных учебных программах. Статья преследует цель побудить студентов к получению фундаментальных знаний через компьютерное моделирование, поскольку именно наглядные образы часто являются стимулом к познанию нового. Сегодня необходимо учить не рецептам решения задач, а их исследованию.

СПИСОК ИСТОЧНИКОВ

1. Морозов А. В., Бригаднов И. А. Математические основы теории систем: динамические системы. СПб.: Северо-Западный государственный заочный технический университет, 2006. 214 с.
2. Морозов А. Д., Драгунов Т. Н. Визуализация и анализ инвариантных множеств динамических систем. М.; Ижевск: Институт компьютерных исследований, 2003. 304 с.
3. Булекбаев Д. А., Морозов А. В. Численное исследование аттракторов в нестационарном уравнении Дуффинга // Перспективы науки. 2024. № 8 (179). С. 32–38.
4. Треногин В. А. Обыкновенные дифференциальные уравнения. М.: Физматлит, 2009. 312 с.
5. Арнольд В. И. Геометрические методы в теории обыкновенных дифференциальных уравнений. 4-е изд. М.: МЦНМО, 2012. 384 с.
6. Аносов Д. В. Дифференциальные уравнения: то решаем, то рисуем. 3-е изд., стер. М.: МЦНМО, 2016. 200 с.
7. Морозов А. В. Качественная теория дифференциальных уравнений — основная составляющая теории динамических систем // Труды Военно-космической академии имени А. Ф. Можайского. 2014. Вып. 642. С. 177–184.
8. Хакен Г. Синергетика: иерархии неустойчивостей в самоорганизующихся системах и устройствах / пер. с англ. Ю. А. Данилова; под ред. Ю. Л. Климонтовича. М.: Мир, 1985. 419 с.
9. Ueda Y. Steady Motions Exhibited by Duffing's Equation: A Picture Book of Regular and Chaotic Motions // New Approaches to Nonlinear Problems in Dynamics / P. J. Holmes (ed.). Philadelphia (PA): Society for Industrial and Applied Mathematics, 1980. Pp. 311–322.
10. Филиппов А. Ф. Введение в теорию дифференциальных уравнений: учебник. 2-е изд., испр. М.: КомКнига, 2007. 240 с.
11. Montagnier P., Paige C. C., Spiteri R. J. Real Floquet Factors of Linear Time-Periodic Systems // Systems and Control Letters. 2003. Vol. 50, iss. 4. Pp. 251–262. DOI: 10.1016/S0167-6911(03)00158-0
12. Oka T., Kitamura S. Floquet Engineering of Quantum Materials // Annual Review of Condensed Matter Physics. 2019. Vol. 10. Pp. 387–408. DOI: 10.1146/annurev-conmatphys-031218-013423

Дата поступления: 14.04.2026

Решение о публикации: 20.05.2026

Computer Simulation of Dynamic Systems on the Phase Plane

- Igor' A. Brigadnov**¹ — Dr. Sci. in Physics and Mathematics, Professor, Full Professor of the “Information and Computing Systems” Department. Research interests: mathematical and computer modeling, decision theory, numerical analysis. E-mail: brigadnov@mail.ru
- Alexey V. Morozov**² — PhD in Physics and Mathematics, Professor, Full Professor of the Department of Higher Mathematics. Research interests: mathematical and computer modeling, numerical analysis. E-mail: vka@mil.ru

¹ Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky ave., Saint Petersburg, 190031, Russia

² Mozhaisky Military Aerospace Academy, 13 Zhdanovskaya str., Saint Petersburg, 197198, Russia

For citation: Brigadnov I.A., Morozov A.V. Computer Simulation of Dynamic Systems on the Phase Plane, *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 75–81. DOI: 10.20295/2413-2527-2026-246-75-81 (In Russian)

Abstract. *The study of dynamical systems on the phase plane is one of the common methods of oscillation theory and is widely used in engineering and scientific practice. Purpose: to combine the elements of analytical research with computer modeling. Results: computational and graphical work on the discipline “Systems Modeling” is proposed and discussed, which is taught in a number of universities in the country. Practical significance: the study of dynamic systems contributes to the development of intelligence, creativity and the formation of professional competencies, which increases the motivation for research activities of the active part of students.*

Keywords: *dynamic system on a plane, phase portraits, analytically-numerical study*

REFERENCES

1. Morozov A. V., Brigadnov I. A. *Matematicheskie osnovy teorii sistem: dinamicheskie sistemy* [Mathematical Foundations of Systems Theory: Dynamical Systems]. Saint Petersburg, North-Western State Correspondence Technical University, 2006, 214 p. (In Russian)
2. Morozov A. D., Dragunov T. N. *Vizualizatsiya i analiz invariantnykh mnozhestv dinamicheskikh sistem* [Visualization and Analysis of Invariant Sets of Dynamical Systems]. Moscow, Izhevsk, Institute of Computer Research, 2003, 304 p. (In Russian)
3. Bulekbaev D. A., Morozov A. V. *Chislennoe issledovanie attraktorov v nestatsionarnom uravnenii Duffinga* [Numerical Study of Attractors in the Non-Stationary Duffing Equation], *Perspektivy nauki [Science Prospects]*, 2024, no. 8 (179), pp. 32–38. (In Russian)
4. Trenogin V. A. *Obyknovennye differentsialnye uravneniya* [Ordinary Differential Equations], Moscow, Fizmatlit Publishing House, 2009, 312 p. (In Russian)
5. Arnold V. I. *Geometricheskie metody v teorii obyknovennykh differentsialnykh uravneniy* [Geometric Methods in the Theory of Ordinary Differential Equations]. Moscow, Moscow Center for Continuous Mathematical Education, 2012, 384 p. (In Russian)
6. Anosov D. V. *Differentsialnye uravneniya: to reshaem, to risuem* [Differential Equations: Solving and Drawing]. Moscow, Moscow Center for Continuous Mathematical Education, 2016, 200 p. (In Russian)
7. Morozov A. V. *Kachestvennaya teoriya differentsialnykh uravneniy — osnovnaya sostavlyayushchaya teorii dinamicheskikh sistem* [Qualitative Theory of Differential Equations as a Basic Component of Dynamic Systems Theory], *Trudy VoЕННО-kosmicheskoy akademii imeni A. F. Mozhayskogo [Proceedings of the Mozhaisky Military Space Academy]*, 2014, iss. 642, pp. 177–184. (In Russian)
8. Haken H. *Sinergetika: ierarkhii neustoychivostey v samoorganizuyushchikhsya sistemakh i ustroystvakh* [Advanced Synergetics]. Moscow, Mir Publishers, 1985, 419 p. (In Russian)
9. Ueda Y. *Steady Motions Exhibited by Duffing’s Equation: A Picture Book of Regular and Chaotic Motions*. In: *Holmes P. J. (ed.). New Approaches to Nonlinear Problems in Dynamics*. Philadelphia (PA), Society for Industrial and Applied Mathematics, 1980, pp. 311–322.
10. Filippov A. F. *Vvedenie v teoriyu differentsialnykh uravneniy: uchebnik* [Introduction to the Theory of Differential Equations: Textbook]. KomKniga Publishing House, 2007, 240 p. (In Russian)
11. Montagnier P., Paige C. C., Spiteri R. J. *Real Floquet Factors of Linear Time-Periodic Systems*, *Systems and Control Letters*, 2003, vol. 50, iss. 4, pp. 251–262. DOI: 10.1016/S0167-6911(03)00158-0
12. Oka T., Kitamura S. *Floquet Engineering of Quantum Materials*, *Annual Review of Condensed Matter Physics*, 2019, vol. 10, pp. 387–408. DOI: 10.1146/annurev-conmatphys-031218-013423

Received: April 14, 2026

Accepted: May 20, 2026

УДК 004.85

Контрастивное самообучение в системах мониторинга транспорта: практический подход к ранней детекции аномалий

- Коньков Дмитрий Александрович** — аспирант кафедры «Информационные и вычислительные системы». Научные интересы: контрастивное обучение, самообучающиеся системы, обеспечение безопасности информационных систем. E-mail: dmitry.konkov@letka.space
- Божко Леся Михайловна** — доктор экон. наук, профессор кафедры «Информационные и вычислительные системы». Научные интересы: математическое моделирование, численные методы, комплексы программ, управление в организационных системах. E-mail: lemib@rambler.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Коньков Д. А., Божко Л. М. Контрастивное самообучение в системах мониторинга транспорта: практический подход к ранней детекции аномалий // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 82–90. DOI: 10.20295/2413-2527-2026-246-82-90

Аннотация. Для выявления дефектов и предупреждения аварий в транспортных системах используются датчики, применяющие модели на основе классических конструкций или пороговых правил, что чревато ложными тревогами или запаздыванием. Модели контрастивного самообучения позволяют снять указанные недостатки, но требуют доработки под прикладные задачи. **Цель исследования:** предложить и описать практический подход к использованию методов контрастивного самообучения для задач раннего обнаружения аномалий в системах мониторинга транспорта. **Методы:** применены методы контрастивного обучения с функцией потерь *InfoNCE*, легковесные сверточные энкодеры (*ID-CNN/ID-ResNet*), вариационные автокодировщики (*VAE*) для реконструкционного контроля, механизмы внимания для оценки вкладов сенсоров, адаптивная калибровка порогов на основе экспоненциального скользящего среднего (*EMA*), кластеризация представлений для множественных нормальных режимов и сценарии тестирования на полуреальных данных. **Результаты:** выявлены особенности мультисенсорных временных рядов, описаны ограничения полевых устройств и определены пути уменьшения ложных срабатываний при ограниченной разметке. Предложены алгоритмические блоки для практического внедрения: аугментации временных рядов, адаптивная калибровка порогов, *attention*-механизмы для объяснимости и валидация на полуреальных данных. Разработан гибридный критерий аномальности контрастивной и реконструкционной оценке аномальности. **Практическая значимость:** внедрение контрастивного блока с адаптивной калибровкой обеспечивает снижение времени обнаружения аномалий и сокращение частоты ложных срабатываний по сравнению с базовыми *VAE* и пороговыми системами. **Обсуждение:** рекомендуется интеграция контрастивных модулей и адаптивной калибровки в существующие системы мониторинга транспорта с использованием трехуровневой системы тревог и объяснений по каналам.

Ключевые слова: контрастивное самообучение, ранняя детекция, мультисенсорный мониторинг, мониторинг транспорта, адаптивная калибровка

1.2.2 — математическое моделирование, численные методы и комплексы программ (технические науки)

Введение

Современные транспортные системы все активнее используют датчики для оперативного выявления дефектов и предупреждения аварий: акселерометры и микрофоны на рельсах, датчики вибрации и температуры на вагонах, датчики тока и GPS [1]. При этом на практике возникают проблемы нехватки размеченных данных, смены условий эксплуатации (погодные нагрузки, тип поезда, скорость) и ограниченных вычислительных ресурсов на краю сети [2]. Классические модели на основе реконструкции (AE, VAE [3]) или пороговых правил часто либо дают ложные тревоги, либо запаздывают [4]. Контрастивное самообучение [5], напротив, позволяет формировать более различимые признаки при минимальной разметке, но оно требует доработки под прикладные ограничения. В статье рассматривается один из возможных вариантов взаимодействия инструментов контрастивного самообучения для практических задач обнаружения аномалий на ранних стадиях развития в системах мониторинга на транспорте.

Основная часть

Математическая формализация задачи обнаружения аномалий в мультисенсорных временных рядах для контрастивного самообучения нейронной сети [5, 6] будет выглядеть следующим образом.

Пусть $X = \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^{d \times T}$, где d — число сенсорных каналов, T — длина окна. Энкодер E переводит сигнал в представление $z = E(x) \in \mathbb{R}^m$. Функция аномальности $A(x)$ сравнивается с порогом τ : $A(x) \geq \tau \Rightarrow$ аномалия, $A(x) < \tau \Rightarrow$ норма.

Архитектура для практической задачи внедрения модели контрастивного самообучения представлена следующим образом.

1. Сбор и предобработка.

Скользящие окна длиной 1–10 с. Набор аугментаций [5], для каждой $\forall t \in T$: $\tilde{x} = x(t)$. Примеры аугментаций: $\tilde{x} = x + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$; временной сдвиг: $\tilde{x}[t] = x[t + \delta]$, $\delta \in [-\Delta, \Delta]$; маскирование каналов [7]: $\tilde{x}_i = m_i x_i$.

2. Компактный энкодер + контекст (на краю).

Энкодер E (1D-CNN/1D-ResNet) дает $z = E(x) \in \mathbb{R}^m$.

Практический паттерн: отдельные CNN для каналов $z_i = \text{CNN}_i(x_i)$, $z_{\text{concat}} = [z_1, \dots, z_d]$.

Здесь x_i — сигнал одного сенсорного канала (например, датчика температуры номер i); CNN_i — отдельная сверточная сеть для этого канала; z_i — компактное представление (эмбединг) этого канала; z_1, \dots, z_d — представления всех d каналов; $[\cdot]$ — операция конкатенации (склеивания векторов); z_{concat} — итоговый объединенный вектор всех каналов.

Контекстный модуль (LSTM/Transformer light) дает конечное z .

3. Контрастивный модуль для представлений.

Косинусное сходство [5]:

$$\text{sim}(z_i, z_j) = \frac{z_i^\top z_j}{\|z_i\| \|z_j\|}.$$

InfoNCE-потеря [5] для пары (позитив z^+ , негативы $\{i = 1\}^{\wedge K}$):

$$\mathcal{L}_{\text{InfoNCE}} = -\log \frac{\exp(\text{sim}(z, z^+) / \kappa)}{\exp(\text{sim}(z, z^+) / \kappa) + \sum_{i=1}^K \exp(\text{sim}(z, z_i^-) / \kappa)},$$

где κ — температурный параметр. Негативы хранятся в буфере B размера K (FIFO).

4. Простая генеративная проверка (VAE) и комбинированный критерий VAE [3] реконструирует x^\wedge .

MSE-реконструкция:

$$\text{MSE}(x, \hat{x}) = \frac{1}{dT} \sum_{i=1}^d \sum_{t=1}^T (x_{i,t} - \hat{x}_{i,t}).$$

KL-дивергенция для нормального апостериорного $\mathcal{N}(\mu, \text{diag}(\sigma^2))$:

$$\mathcal{L}_{\text{KL}} = \frac{1}{2} \sum_{j=1}^m (\mu_j^2 + \sigma_j^2 - \log \sigma_j^2 - 1).$$

Полная VAE-потеря [3, 7]:

$$\mathcal{L}_{\text{VAE}} = \text{MSE}(x, \hat{x}) + \beta \mathcal{L}_{\text{KL}},$$

где β — коэффициент регуляризации. Формулировка с весом β при KL-члене соответствует β -VAE; при $\beta = 1$ она совпадает с классическим VAE из [2].

Контрастивная оценка аномальности и реконструкционная оценка аномальности определяются как

$$S_{\text{contr}}(x) = 1 - \max_i \text{sim}(E(x), c_i),$$

$$S_{\text{recon}}(x) = \text{MSE}(x, \hat{x}).$$

Здесь контрастная оценка аномальности должна расти при удалении представления от нормального поведения. Поэтому в качестве опорного множества берутся прототипы нормы (центроиды нормальных кластеров $\{c_i\}$, см. п. 6), а не буфер негативов: $S_{\text{contr}}(x) = 1 - \max_i \text{sim}(E(x), c_i)$. Для нормального объекта $\text{sim} \rightarrow 1$ и $S_{\text{contr}} \rightarrow 0$, для аномального — $S_{\text{contr}} \rightarrow 1$, что согласуется с InfoNCE и решающим правилом $A(x) \geq \tau$. Перед взвешиванием оценки аномальности S_{contr} и S_{recon} нормируются (например, робастной стандартизацией по нормальной выборке), чтобы коэффициент α имел смысл сопоставимого веса; значение α подбирается по валидации и проверяется ablation-сравнением (проверка вклада компонентов путем их поочередного отключения) с $\alpha = 0$ (только VAE) и $\alpha = 1$ (только контраст).

Комбинированный критерий [3, 5]:

$$A(x) = \alpha S_{\text{contr}}(x) + (1 - \alpha) S_{\text{recon}}(x), \alpha \in [0, 1].$$

Решение о тревоге: $A(x) \geq \tau$.

5. Attention для объяснимости.

Для каналов $x = [x_1, \dots, x_d]$ вычисляем скрытые представления h_i и attention-веса

$$a_i = \frac{\exp(Wh_i + b)}{\sum_{j=1}^d \exp(Wh_j + b)}.$$

Вклад i -го канала в общую аномальность: $C_i = a_i \times A(x)$.

Веса внимания интерпретируются как индикатор относительной значимости канала, а не как строгая каузальная мера вклада; для эксплуатации они служат подсказкой оператору о наиболее «подозрительном» канале.

6. Адаптация и калибровка на месте.

Онлайн-оценка статистик с ЕМА [10, 11]:

$$\mu_t = \gamma \mu_{t-1} + (1 - \gamma) A(x_t),$$

$$v_t = \gamma v_{t-1} + (1 - \gamma) (A(x_t) - \mu_t)^2,$$

$$\tau_t = \mu_t + \lambda \sqrt{v_t},$$

где $\gamma \in (0, 1)$ (например, 0,9–0,99);

λ — множитель для контроля FPR (обычно 2–3).

При множественности нормальных режимов используется кластеризация представлений [7, 8] с центроидами c_i и $d(E(x_t), c_i)$ и статистиками (μ_i, v_i) , где v_i — ЕМА-дисперсия оценки аномальности в i -м режиме. Для нового окна выбирается ближайший центроид $j = \text{argmin}_i d(E(x_t), c_i)$, и порог берется по его статистикам: $\tau_t = \mu_j + \lambda \sqrt{v_j}$. Тем самым обозначения согласованы с обычным режимом — всюду используется дисперсия v и \sqrt{v} .

7. Тестирование и валидация в полевых условиях.

Рекомендуемые сценарии: синтетические инъекции дефектов, полуреальные тесты (пилоты на 1–2 состава/участка).

8. Метрики.

Стандартные метрики [9] оценивают качество классификации:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Precision (точность) показывает, какая доля от всех тревог модели реально являлась аномалией (насколько можно доверять срабатыванию), **Recall** (полнота) — какую долю реальных аномалий модель обнаружила (ничего не пропустила), а **F1** объединяет их в одно число через гармоническое среднее, штрафую за сильный перекося в любую сторону.

TTD (Time to Detect) измеряет задержку между реальным началом аномалии и моментом ее обнаружения моделью: чем меньше, тем быстрее можно отреагировать.

FPR (False Positive Rate) нормирует количество ложных срабатываний на длину маршрута, позволяя корректно сравнивать системы на маршрутах разной протяженности: чем ниже значение, тем надежнее система в эксплуатации.

9. Примеры ожидаемой эффективности и пример расчета экономии.

На референсных сценариях внедрение контрастного блока с адаптивной калибровкой обычно дает снижение TTD на 30–60% и снижение FPR на 25–50% по сравнению с простыми VAE/

пороговыми системами (при корректной калибровке). Приводимые значения являются прогнозными оценками на основе референсных диапазонов и подлежат подтверждению в полевых испытаниях. Экономическая оценка перехода от базового метода base к новому:

$$E = C_{FP} \times (FPR_{base} - FPR_{new}) \times D + C_{delay} \times (TTD_{base} - TTD_{new}) \times N, \tag{1}$$

где C_{FP} — стоимость одной ложной проверки; C_{delay} — стоимость задержки на единицу времени;

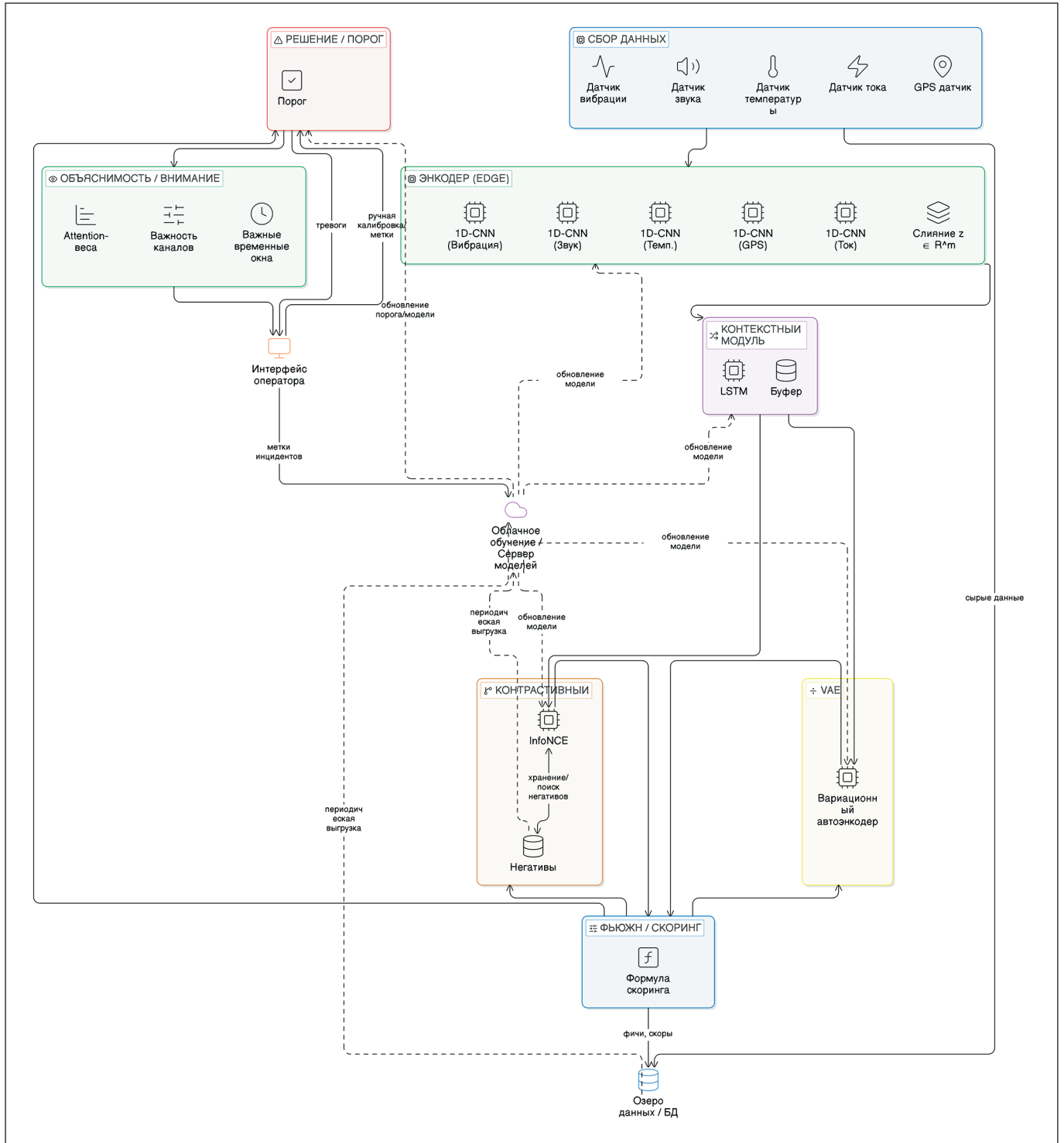


Рисунок. Блок-схема практической системы мониторинга (сбор → энкодер → контекст → contrastive + VAE → пороговая логика → объяснение)

D — длина участка, км;

N — число дефектов за период.

Пример подстановки: при $FPR_{base} = 0,18/км$, $FPR_{new} = 0,09/км$, $D = 100$ км, $C_{FP} = 20\,000$ руб. получаем компоненту экономии по FP: $E_{FP} = 20\,000 \times (0,18 - 0,09) \times 100 = 180\,000$ руб.

Для интеграции в эксплуатацию рекомендует-ся [10, 11]:

- собирать при выездах технических бригад в течение нескольких часов (1–2) данные в штатных режимах работы;
- логировать attention-объяснения в отчетах;
- централизованно провести предобучение и локальную тонкую настройку на узлах;
- построить интерфейс с тремя уровнями тревоги и объяснением по каналам.

Практическая реализация системы мониторинга требует интеграции нескольких компонентов в единый конвейер обработки данных. Архитектура должна обеспечивать как качество детекции, так и возможность развертывания на устройствах с ограниченными ресурсами. Предлагаемая архитектура может быть представлена в виде последовательности функциональных блоков, как показано на рисунке.

На рисунке представлена последовательность обработки данных в предлагаемой системе. Мультисенсорные данные (вибрация, температура, ток, GPS) [1] поступают с транспортного средства и сегментируются скользящими окнами (блок «Сбор и предобработка»). Каждое окно подвергается аугментации (шум, сдвиг, маскирование каналов) для формирования позитивных пар. Легковесный энкодер (1D-CNN/ResNet) извлекает локальные признаки по каналам, которые объединяются и передаются в контекстный модуль (LSTM/Transformer с пониженной сложностью) [12, 13], формирующий финальное представление z .

Далее z направляется в два параллельных блока: контрастный модуль (InfoNCE-потеря [5] с буфером негативных примеров) для обучения различимых представлений и VAE [3] для реконструкционной проверки. Контрастная оценка аномальности S_{contr} и реконструкционная оценка аномальности S_{recon} взвешенно комбинируются

в финальный критерий аномальности $A(x)$. Пороговая логика (с адаптивной калибровкой через ЕМА) принимает решение о тревоге. Attention-механизм вычисляет вклады каждого сенсора в общую аномальность и формирует объяснения для оператора (какой канал вызвал тревогу и насколько). Вся схема рассчитана на развертывание на краевых устройствах с ограниченными ресурсами (отсюда легковесные архитектуры и онлайн-калибровка).

Анализ методов обнаружения аномалий с учетом специфики транспортной сферы, на основании логических и математических правил постановки задачи, а также выработанных выше метрик оценки позволил свести результаты такого анализа в единую таблицу.

Из таблицы можно сформулировать ряд ключевых выводов:

1. Пороговые правила простые, хорошо объяснимые, но требуют много ручной настройки и дают много ложных срабатываний.
2. VAE и LSTM Autoencoder улучшают показатели, но страдают от проблем с объяснимостью (особенно LSTM).
3. Изолированный лес не требует меток, но имеет посредственные показатели по TTD и FPR.
4. Контрастное обучение значительно улучшает TTD и FPR с хорошей объяснимостью.
5. Гибридный подход (предлагаемый авторами) сочетает преимущества контрастного обучения и VAE, достигая наилучших показателей по TTD и FPR при высокой объяснимости.

Таблица ниже наглядно демонстрирует превосходство предлагаемого гибридного подхода над существующими методами по ключевым метрикам производительности. Приведенные в таблице значения являются ориентировочными (референсными) диапазонами и требуют подтверждения в полевых испытаниях; в частности, преимущество гибрида над отдельными компонентами (только VAE, только контраст) должно быть подтверждено ablation-экспериментом.

Примеры расчета ожидаемой эффективности

Экономическая оценка внедрения проводилась на основе формулы (1).

Сравнение методов обнаружения аномалий в транспортном мониторинге

Метод	Основа	Требуемые метки, %	TTD, с	FPR, 1/км	Объяснимость
Пороговые правила	Ручные эксперты	> 50	180–300	0,15–0,25	Отлично
Классический VAE	Реконструкция	< 5	120–240	0,12–0,18	Средняя
LSTM Autoencoder	Временной контекст	< 5	90–180	0,10–0,15	Низкая
USAD	Adversarial + реконструкция	0	90–180	0,10–0,14	Низкая
Изолированный лес	Аномалия в пространстве	0	120–180	0,14–0,20	Средняя
Контрастивное обучение	Различимые представления	1–5	60–120	0,08–0,10	Высокая (attention)
Гибридный подход	Контрастив + VAE	1–5	48–120	0,08–0,09	Отлично

Пояснения к таблице:

- метод — названия различных подходов к обнаружению аномалий;
- основа — краткое описание принципа работы метода;
- требуемые метки — количество размеченных данных, необходимых для обучения;
- TTD — время до обнаружения аномалии в секундах (меньше = лучше);
- FPR — частота ложных срабатываний на километр пути (меньше = лучше);
- объяснимость — насколько понятны и интерпретируемы результаты.

Пример 1 (компонента ложных срабатываний).

При $FPR_{base} = 0,18/км$, $FPR_{new} = 0,09/км$, $D = 100 км$, $C_{FP} = 20\,000 руб.$ экономия по компоненте ложных проверок составляет $E_{FP} = 20\,000 \times (0,18 - 0,09) \times 100 = 180\,000 (руб.)$.

При $FPR_{base} = 0,18/км$, $FPR_{new} = 0,09/км$, $D = 100 км$, $C_{FP} = 20\,000 руб.$ экономия по компоненте ложных проверок составляет $E_{FP} = 20\,000 \times (0,18 - 0,09) \times 100 = 180\,000 (руб.)$.

Пример 2 (компонента времени обнаружения).

При $TTD_{base} = 180 с$, $TTD_{new} = 90 с$, $N = 50$ дефектов за период, $C_{delay} \approx 83,33 руб./с$ экономия по компоненте сокращения времени составляет $E_{TTD} \approx 83,33 \times (180 - 90) \times 50 \approx 375\,000 (руб.)$.

Суммарная ожидаемая экономия для данного сценария: $E_{total} = 180\,000 + 375\,000 = 555\,000 (руб.)$ за период наблюдения. В качестве base здесь взяты значения порогового/VAE-метода по верхней границе таблицы; при выборе другого базового метода значения изменятся. Сдвиг $FPR 0,18 \rightarrow 0,09$ соответствует диапазону снижения, указанному в п. 9.

Заключение

В результате предложенный гибридный подход, основанный на комбинации контрастивного самообучения и вспомогательного модуля реконструкции (VAE), а также дополненный attention-механизмом и адаптивной калибровкой, нацелен на снижение числа ложных срабатываний и задержек при детекции. Согласно сравнительному анализу (см. таблицу) и расчетам на референсных сценариях, ожидается снижение времени до обнаружения аномалии (TTD) и доли ложных тревог (FPR) относительно базовых методов, что создает основу для прямого экономического эффекта; количественное подтверждение требует полевых испытаний.

Полученные данные позволяют сделать вывод, что для современных систем мониторинга транспорта [1, 10] на первый план выходит не сложность самой модели, а задачи ее грамотной интеграции, полевой валидации и поддержки полного жизненного цикла. Перспективными направлениями для дальнейшей работы

становятся поиск путей оптимизации архитектуры для устройств с критически малым объемом ресурсов, а также разработка методов адаптации модели к новым типам подвижного состава в режиме онлайн, без необходимости полного переобучения.

СПИСОК ИСТОЧНИКОВ

1. Vehicle-as-a-Sensor Approach for Urban Track Anomaly Detection / V. Sruc [et al.] // *Sensors*. 2025. Vol. 25, iss. 21. Art. 6679. 24 p. DOI: 10.3390/s25216679
2. Thill M., Konen W., Bäck T. Online Anomaly Detection on the Webscope S5 Dataset: A Comparative Study // *Proceedings of the 2017 IEEE Workshop on Evolving and Adaptive Intelligent Systems (EAIS 2017)* (Ljubljana, Slovenia, 31 May — 2 June 2017). Institute of Electrical and Electronics Engineers, 2017. 8 p. DOI: 10.1109/EAIS.2017.7954844
3. Kingma D. P., Welling M. Auto-Encoding Variational Bayes // *Proceedings of the Second International Conference on Learning Representations (ICLR 2014)* (Banff, Canada, 14–16 April 2014). 14 p. DOI: 10.48550/arXiv.1312.6114
4. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding / K. Hundman [et al.] // *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '18)* (London, United Kingdom, 19–23 August 2018). New York: Association for Computing Machinery, 2018. Pp. 387–395. DOI: 10.1145/3219819.3219845
5. A Simple Framework for Contrastive Learning of Visual Representations / T. Chen [et al.] // *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)* (Virtual Event, 13–18 July 2020). *Proceedings of Machine Learning Research*. 2020. Vol. 119. Pp. 1597–1607.
6. USAD: Unsupervised Anomaly Detection on Multivariate Time Series / J. Audibert [et al.] // *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'20)*, (Virtual Event, 6–10 July 2020). New York: Association for Computing Machinery, 2020. Pp. 3395–3404. DOI: 10.1145/3394486.3403392
7. Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection / B. Zong [et al.] // *Proceedings of the Sixth International Conference on Learning Representations (ICLR 2018)* (Vancouver, Canada, 30 April — 3 May 2018). 19 p. URL: <http://openreview.net/forum?id=BJLHbb0-> (дата обращения: 09.05.2026).
8. Deng A., Hooi B. Graph Neural Network-Based Anomaly Detection in Multivariate Time Series // *Proceedings of the 35th Conference on Artificial Intelligence (AAAI 2021)* (Virtual Event, 2–9 February 2021). *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. Vol. 35, no. 5. Pp. 4027–4035. DOI: 10.1609/aaai.v35i5.16523
9. On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study / G.O. Campos [et al.] // *Data Mining and Knowledge Discovery*. 2016. Vol. 30, iss. 4. Pp. 891–927. DOI: 10.1007/s10618-015-0444-8
10. Federated Learning: Challenges, Methods, and Future Directions / T. Li [et al.] // *IEEE Signal Processing Magazine*. 2020. Vol. 37, no. 3. Pp. 50–60. DOI: 10.1109/MSP.2020.2975749
11. Schneider J. Wenig P., Papenbrock T. Distributed Detection of Sequential Anomalies in Univariate Time Series // *The International Journal on Very Large Data Bases*. 2021. Vol. 30, iss. 4. Pp. 579–602. DOI: 10.1007/s00778-021-00657-6
12. Long Short Term Memory Networks for Anomaly Detection in Time Series / P. Malhotra [et al.] // *Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015)* (Bruges, Belgium, 22–24 April 2015). i6doc.com Publishing, 2015. Pp. 89–94.
13. Robust Anomaly Detection for Multivariate Time Series Through Stochastic Recurrent Neural Network / Y. Su [et al.] // *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19)* (Anchorage, AK, USA, 4–8 August 2019). New York: Association for Computing Machinery, 2019. Pp. 2828–2837. DOI: 10.1145/3292500.3330672

Дата поступления: 24.05.2026

Решение о публикации: 26.05.2026

Contrastive Self-Supervised Learning in Transportation Monitoring Systems: A Practical Approach to Early Anomaly Detection

Dmitry A. Konkov

— Postgraduate Student of the “Information and Computing Systems” Department. Research interests: contrastive learning, self-supervised systems, information-security assurance for information systems. E-mail: dmitry.konkov@letka.space

Lesya M. Bozhko

— Dr. Sci. in Economics, Professor of the “Information and Computing Systems” Department. Research interests: mathematical modeling, numerical methods, program complexes, management in organizational systems. E-mail: lemib@rambler.ru

Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky ave., Saint Petersburg, 190031, Russia

For citation: Konkov D. A., Bozhko L. M. Contrastive Self-Supervised Learning in Transportation Monitoring Systems: A Practical Approach to Early Anomaly Detection, *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 82–90. DOI: 10.20295/2413-2527-2026-246-82-90 (In Russian)

Abstract. *To detect defects and prevent accidents in transport systems, sensors are used that use models based on classical designs or threshold rules, which is fraught with false alarms or delays. **Purpose:** to propose and describe a practical approach to the use of contrastive self-supervised learning methods for early detection of anomalies in transport monitoring systems. **Methods:** contrastive learning methods with the InfoNCE loss function, lightweight convolutional encoders (1D-CNN/1D-ResNet), variational autoencoders (VAE) for reconstruction control, attention mechanisms for assessing sensor contributions, adaptive calibration of thresholds based on the exponential moving average (EMA), representation clustering for multiple normal modes, and testing scenarios on semi-real data are applied. **Results:** features of multisensory time series have been identified, limitations of field devices have been described, and ways to reduce false positives with limited markup have been identified. Algorithmic blocks for practical implementation are proposed: time series augmentation, adaptive threshold calibration, attention mechanisms for explainability and validation on semi-real data. A hybrid anomaly criterion based on contrastive and reconstruction scores has been developed. **Practical significance:** the introduction of a contrast unit with adaptive calibration provides a decrease in the time for detecting anomalies and a decrease in the frequency of false positives compared to basic VAE and threshold systems. **Discussion:** integration of contrast modules and adaptive calibration into existing transport monitoring systems using a three-level alarm system with per-channel explanation.*

Keywords: *contrastive self-supervised learning, early detection, multi-sensor monitoring, transport monitoring, adaptive calibration*

REFERENCES

1. Sruk V., et al. Vehicle-as-a-Sensor Approach for Urban Track Anomaly Detection, *Sensors*, 2025, vol. 25, iss. 21, art. no. 6679, 24 p. DOI: 10.3390/s25216679
2. Thill M., Konen W., Bäck T. Online Anomaly Detection on the Webscope S5 Dataset: A Comparative Study, *Proceedings of the 2017 IEEE Workshop on Evolving and Adaptive Intelligent Systems (EAIS 2017)*, Ljubljana, Slovenia, May 31 — June 02, 2017. Institute of Electrical and Electronics Engineers, 2017, 8 p. DOI: 10.1109/EAIS.2017.7954844
3. Kingma D. P., Welling M. Auto-Encoding Variational Bayes, *Proceedings of the Second International Conference on Learning Representations (ICLR 2014)*, Banff, Canada, April 14–16, 2014, 14 p. DOI: 10.48550/arXiv.1312.6114

4. Hundman K., et al. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '18)*, London, United Kingdom, August 19–23, 2018. New York, Association for Computing Machinery, 2018, pp. 387–395. DOI: 10.1145/3219819.3219845
5. Chen T., et al. A Simple Framework for Contrastive Learning of Visual Representations, *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, Virtual Event, July 13–18, 2020. *Proceedings of Machine Learning Research*, 2020, vol. 119, pp. 1597–1607.
6. Audibert J., et al. USAD: Unsupervised Anomaly Detection on Multivariate Time Series, *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '20)*, Virtual Event, July 06–10, 2020. New York, Association for Computing Machinery, 2020, pp. 3395–3404. DOI: 10.1145/3394486.3403392
7. Zong B., et al. Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection, *Proceedings of the Sixth International Conference on Learning Representations (ICLR 2018)*, Vancouver, Canada, April 30 – May 03, 2018, 19 p. Available at: <http://openreview.net/forum?id=BJLHbb0-> (accessed: May 09, 2026).
8. Deng A., Hooi B. Graph Neural Network-Based Anomaly Detection in Multivariate Time Series, *Proceedings of the 35th Conference on Artificial Intelligence (AAAI 2021)*, Virtual Event, February 02–09, 2021. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, no. 5, pp. 4027–4035. DOI: 10.1609/aaai.v35i5.16523
9. Campos G. O., et al. On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study, *Data Mining and Knowledge Discovery*, 2016, vol. 30, iss. 4, pp. 891–927. DOI: 10.1007/s10618-015-0444-8
10. Li T., et al. Federated Learning: Challenges, Methods, and Future Directions, *IEEE Signal Processing Magazine*, 2020, vol. 37, no. 3, pp. 50–60. DOI: 10.1109/MSP.2020.2975749
11. Schneider J. Wenig P., Papenbrock T. Distributed Detection of Sequential Anomalies in Univariate Time Series, *The International Journal on Very Large Data Bases*, 2021, vol. 30, iss. 4, pp. 579–602. DOI: 10.1007/s00778-021-00657-6
12. Malhotra P., et al. Long Short Term Memory Networks for Anomaly Detection in Time Series, *Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015)*, Bruges, Belgium, April 22–24, 2015. i6doc.com Publishing, 2015, pp. 89–94.
13. Su Y., et al. Robust Anomaly Detection for Multivariate Time Series Through Stochastic Recurrent Neural Network, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19)*, Anchorage, AK, USA, August 04–08, 2019. New York, Association for Computing Machinery, 2019, pp. 2828–2837. DOI: 10.1145/3292500.3330672

Received: May 24, 2026

Accepted: May 26, 2026

УДК 51-74

Обзор и сравнительный анализ методов моделирования полетных характеристик беспилотных летательных аппаратов

Стрелков Иван Владимирович

— аспирант кафедры «Информационные и вычислительные системы». Научные интересы: математическое моделирование, численные методы, беспилотные летательные аппараты. E-mail: ivan_strelkov99@mail.ru

Божко Леся Михайловна

— доктор экон. наук, профессор кафедры «Информационные и вычислительные системы». Научные интересы: математическое моделирование, численные методы, комплексы программ, управление в организационных системах. E-mail: lemib@rambler.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Стрелков И. В., Божко Л. М. Обзор и сравнительный анализ методов моделирования полетных характеристик беспилотных летательных аппаратов // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 91–99. DOI: 10.20295/2413-2527-2026-246-91-99

Аннотация. С расширением сфер применения беспилотных летательных аппаратов достоверное прогнозирование их поведения на этапе проектирования становится все более актуальной задачей. Решить ее возможно с использованием соответствующих методов моделирования. **Цель:** выявить и сравнить методы, применяемые в моделировании полетных характеристик беспилотных летательных аппаратов. **Результаты:** рассмотрена модель пространственного движения с шестью степенями свободы (6DoF, Six Degrees of Freedom), проведено сравнение основных методов, используемых для моделирования полетных характеристик (метода Эйлера, классического метода Рунге — Кутты четвертого порядка и адаптивной схемы Рунге — Кутты — Фельберга), проанализированы алгоритмы планирования маршрута (A^* , RRT*, генетический алгоритм и метод роя частиц). Установлено, что для большинства исследовательских задач метод Рунге — Кутты четвертого порядка оказывается разумным компромиссом по точности и вычислительной стоимости, тогда как на режимах с разномасштабной динамикой предпочтительны адаптивные схемы. Сформулированы рекомендации по подбору метода в зависимости от класса задачи и допустимой погрешности. **Практическая значимость:** результаты работы применимы при создании программных средств моделирования полета и выборе вычислительного ядра для конкретного класса беспилотных летательных аппаратов.

Ключевые слова: математическое моделирование, численные методы, беспилотный летательный аппарат, БПЛА, моделирование полета, метод Рунге — Кутты, оптимизация траектории, алгоритм A^* , метод роя частиц

1.2.2 — математическое моделирование, численные методы и комплексы программ (технические науки)

Введение

За последние десять лет беспилотные летательные аппараты (БПЛА) прошли путь из преимущественно военной техники в массовый гражданский инструмент: аэрофотосъемка полей, осмотр линий

электропередачи и трубопроводов, доставка небольших грузов в труднодоступные районы, мониторинг состояния инфраструктуры. На стадии проектирования нового аппарата инженер должен заранее

оценить его аэродинамические свойства, устойчивость и реакцию на управляющие воздействия: натурные летные испытания стоят дорого, занимают время и связаны с риском потерять прототип [1, 2].

Цифровая модель дает возможность воспроизвести полет на настольном компьютере, меняя массу и геометрию аппарата, атмосферные условия, настройки регулятора, без того, чтобы физически строить каждую итерацию прототипа. Задача при этом разбивается на несколько взаимосвязанных блоков: уравнения движения, схема их численного интегрирования, синтез контура управления и планирование маршрута [3]. Для каждого из них существует свое семейство алгоритмов, и подбор конкретных в значительной мере определяется требованиями по точности и скорости конкретной задачи.

Спектр доступных решений охватывает как простые линейные аналитические приближения, описывающие малые отклонения от установившегося режима, так и полные нелинейные схемы, воспроизводящие штопор, сваливание и турбулентные порывы [4, 5]. Имеющиеся обзоры [2, 6] зафиксировали состояние области на момент их выхода, но стремительное развитие вычислительной базы и появление новых адаптивных алгоритмов делают повторное сравнительное рассмотрение оправданным.

Цель статьи — выявить и сравнить методы, применяемые в моделировании полетных характеристик беспилотных летательных аппаратов, что позволило бы определить ключевые подходы к воспроизведению летных характеристик БПЛА. В исследовании, помимо формулировки уравнений динамики, приведены их численное решение и алгоритмы построения оптимального маршрута.

Математические модели динамики полета БПЛА

Пространственное движение БПЛА описывается в рамках классической механики твердого тела двумя векторными уравнениями для поступательного и вращательного движения [1]. При формализации вводят две системы координат: земную инерциальную NED (North, East, Down, «север, восток, вниз») и связанную систему, жестко прикрепленную к корпусу аппарата [3].

Взаимная ориентация связанной и земной систем задается тремя углами Эйлера: угол крена φ отвечает за вращение вокруг продольной оси, угол тангажа θ — за отклонение носа выше или ниже горизонта, угол рыскания ψ — за поворот в горизонтальной плоскости. Перевод вектора между системами делается умножением на матрицу направляющих косинусов 3×3 , элементы которой — тригонометрические функции этих углов [3, 6]. При $\theta = \pm 90^\circ$ матрица вырождается (так называемый карданный замок); обойти это вырождение позволяет кватернионное описание ориентации.

Поступательное движение центра масс БПЛА описывается в проекциях на оси связанной системы координат следующим векторным уравнением [1, 5]:

$$m \left(\frac{dV}{dt} + \omega \times V \right) = F, \quad (1)$$

где m — масса аппарата;

V — вектор линейной скорости в связанной системе координат;

ω — вектор угловой скорости;

F — суммарный вектор действующих сил (аэродинамических, гравитационных и тяговых).

Вращение аппарата описывается уравнением моментов:

$$I \left(\frac{d\omega}{dt} \right) + \omega \times (I\omega) = M, \quad (2)$$

где I — тензор инерции аппарата (симметричная матрица 3×3);

M — результирующий момент относительно центра масс, складывающийся из аэродинамического момента, момента тяги двигателей и моментов от отклонения рулевых поверхностей [1, 5].

Аэродинамические нагрузки вычисляют через безразмерные коэффициенты, которые зависят от угла атаки α , угла скольжения β , чисел Маха и Рейнольдса, а также от текущих отклонений элеронов, руля высоты и руля направления. Подъемная сила — произведение скоростного напора, площади крыла и коэффициента подъемной силы C_L , который при малых углах атаки линеен по

а. Лобовое сопротивление складывается из профильной составляющей C_{D0} и индуктивной добавки, пропорциональной квадрату C_L [3, 7]. Сами зависимости коэффициентов от параметров режима получают экспериментально в продувках в аэродинамической трубе либо численно, средствами вычислительной гидродинамики (computational fluid dynamics, CFD); в расчетные модели их подставляют как таблицы или как полиномиальные аппроксимации [7].

Шесть уравнений поступательного и вращательного движения вместе с кинематикой углов Эйлера складываются в модель с шестью степенями свободы (6DoF, Six Degrees of Freedom): три линейные скорости (u, v, w), три угловые скорости (p, q, r), координаты центра масс (x, y, z) и углы (φ, θ, ψ). Итого 12 обыкновенных дифференциальных уравнений (ОДУ) первого порядка [1, 5]. Линеаризуя систему в окрестности прямолинейного горизонтального полета, ее разложение по собственным модам дает быструю короткопериодическую моду и медленную фугоиду в продольном канале плюс голландский шаг и спиральную моду — в боковом [3, 8].

Численные методы интегрирования уравнений движения

Замкнутого аналитического решения система (1)–(2) не имеет: аэродинамические коэффициенты входят в правые части как нелинейные функции углов α, β и числа Маха, а управляющие отклонения меняются во времени по закону, задаваемому регулятором. Траектория получается численным интегрированием с фиксированным или переменным шагом h . Критерий выбора схемы двоякий: шаг должен быть достаточно мелким, чтобы погрешность за время моделирования не накопилась до величины, сравнимой с характерными амплитудами переходного процесса, и одновременно достаточно крупным, чтобы общее число обращений к правой части не выходило за вычислительный бюджет бортового или настольного контроллера [4].

Явная схема Эйлера — простейшая из одношаговых методов: значение вектора состояния на

шаге $n + 1$ получается одной подстановкой правой части в текущей точке [4, 9]:

$$y_{n+1} = y_n + h \times f(t_n, y_n),$$

где h — шаг интегрирования.

Разложение точного решения в ряд Тейлора показывает, что локальная ошибка усечения эйлеровой схемы имеет порядок $O(h^2)$, глобальная — $O(h)$. Для динамики БПЛА это означает, что на прямолинейном крейсерском участке схема еще приемлема: отклонение от эталона после минуты полета остается на уровне единиц метров, но на участке с активным маневром, где правая часть системы быстро меняется, тот же шаг приводит к накоплению отклонения до десятков метров за несколько десятков секунд, и для содержательной симуляции схема уже непригодна.

Существенно больший запас точности при умеренных затратах дает классическая схема Рунге — Кутты четвертого порядка (Runge — Kutta 4th order, RK4). На каждом шаге правая часть вычисляется четырежды: в начале отрезка, дважды в его середине с разными оценками промежуточного состояния и в конце [9]:

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ k_2 &= f\left(t_n + \frac{h}{2}, y_n + k_1 \frac{h}{2}\right), \\ k_3 &= f\left(t_n + \frac{h}{2}, y_n + k_2 \frac{h}{2}\right), \\ k_4 &= f(t_n + h, y_n + k_3 h). \end{aligned}$$

Итоговое приращение составляется как взвешенная сумма четырех оценок с весами $1 : 2 : 2 : 1$, нормированными на шесть:

$$y_{n+1} = y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4).$$

Локальная ошибка усечения RK4 имеет порядок $O(h^5)$, глобальная — $O(h^4)$, то есть при одинаковом шаге h погрешность оказывается на четыре порядка меньше эйлеровой. При этом вычислительная стоимость одного шага возрастает лишь вчетверо, что делает RK4 выгодной схемой во всем диапазоне крейсерских режимов [9, 10]. Слабое место

фиксированного шага проявляется на двух крайних режимах: на установившемся горизонтальном полете заданный заранее h оказывается избыточно мелким и ведет к неоправданному числу обращений к правой части, а на резких маневрах, напротив, недостаточен: интегратор не успевает отследить быстро меняющиеся переменные без роста погрешности.

Когда фиксированная сетка становится ограничением, применяют вложенные пары Рунге — Кутты. В работе используется схема Фельберга RK45 (Runge — Kutta — Fehlberg): на каждом шаге она одновременно дает две оценки приращения (четвертого и пятого порядков), а их разность трактуется как апостериорная оценка локальной ошибки [9]. Следующий шаг пересчитывается по правилу $h_{\text{new}} = h \times s \times (\text{tol}/\text{err})^{(1/5)}$ с эмпирическим коэффициентом запаса $s = 0,9$; рост и уменьшение шага ограничены интервалом $[h/10; 5h]$, что предохраняет от выхода за пределы устойчивости при резких возмущениях. В автономном сравнительном эксперименте на типовых задачах БПЛА самолетного типа такая схема при $\text{tol} = 10^{-6}$ сохраняет порядок погрешности, сопоставимый с классическим RK4 при шаге 0,01 с., но достигает его за существенно меньшее число обращений к правой части, укрупняя шаг там, где динамика позволяет.

Сопоставление трех схем сведено в таблице. В практике моделирования БПЛА самолетного типа RK4 служит рабочей лошадкой: его детерминированная вычислительная нагрузка удобна для бортовых систем реального времени, а при шаге 0,01 с. обеспечиваемая точность достаточна для задач проектирования контуров стабилизации и расчета траекторий [10]. Адаптивные схемы типа RK45 берут на себя режимы, в которых характерное время процесса меняется на порядок

и более: штопор, восстановление после срыва, жесткая посадка.

Алгоритмы оптимизации траектории полета

Планирование маршрута формализуется как задача нахождения последовательности пространственных точек $\{x_k\}$, минимизирующей функционал потерь $J(x_1, \dots, x_N)$ при ограничениях типа неравенств (зоны запрета полета, максимальная перегрузка, допустимый угол крена) и конечных условиях в стартовой и финишной точках. В качестве функционала обычно выступают длина пути, время прохождения или расход энергии на борту [11].

Для дискретизированного пространства эффективен поиск A^* : он обходит узлы графа в порядке, определяемом оценочной функцией [12]:

$$f(n) = g(n) + h(n),$$

где $g(n)$ — фактическая стоимость перехода из стартового узла к узлу n ;

$h(n)$ — эвристическая нижняя оценка оставшегося пути от n до цели.

При выполнении условия допустимости ($h(n)$ не превосходит истинного расстояния от n до цели) поиск завершается на оптимальном решении [12]. Ограничение на практике связано с плотностью сетки: при кубической дискретизации рабочей зоны число узлов растет как $O(L^3/\Delta^3)$, где Δ — шаг сетки, и для трехмерных задач планирования над пересеченной местностью это становится лимитирующим фактором.

Если пространство поиска непрерывно или имеет высокую размерность, перспективнее выборочные методы семейства RRT (Rapidly-exploring Random Tree). Идея такова: из стартовой точки выращивается дерево, каждая новая вершина выбирается случайно из рабочей области и затем

Таблица

Сравнительная характеристика численных методов интегрирования

Метод	Порядок точности	Вычислений f на шаг	Область применения
Метод Эйлера	1	1	Предварительная оценка, отладка моделей
Метод RK4	4	4	Основное моделирование, системы реального времени
Метод RK45 (адаптивный)	4–5	6	Маневры, переменная динамика полета

подтягивается к ближайшей вершине дерева на расстояние Δ . Базовый RRT быстро покрывает пространство, но найденный путь обычно неоптимален. Модификация RRT* после каждого добавления вершины перепроверяет, не дает ли новый узел более короткого пути для уже присоединенных соседей, и при необходимости переподключает ребра (rewiring). В пределе при $n \rightarrow \infty$ такой путь асимптотически сходится к оптимальному [13, 14].

Метаэвристика генетического алгоритма (ГА) хорошо приспособлена к негладким и многоэкстремальным целевым функциям, характерным для задач облета с учетом ветровых полей. Траектория кодируется в виде хромосомы (вектора параметров промежуточных точек): популяция из N особей эволюционирует в течение заданного числа поколений с операторами селекции (турнирной или пропорциональной), скрещивания (одноточечного, двухточечного или арифметического blend с коэффициентом α) и мутации с вероятностью p_m . В разработанной реализации по умолчанию используются популяция из 100 особей, турнирная селекция с размером турнира 3, blend-скрещивание и вероятность мутации 0,1; сходимость отслеживается по истории лучшей приспособленности, элитизм сохраняет двух лучших особей между поколениями [15].

Альтернатива биологического происхождения — метод роя частиц (particle swarm optimization, PSO). Каждая частица хранит собственную позицию в пространстве параметров, скорость и персональный рекорд p_i . Рой в целом запоминает глобальный рекорд g . На каждой итерации скорость частицы пересчитывается с учетом инерции предыдущей скорости, направленного движения к p_i и направленного движения к g [16]:

$$v_i = wv_i + c_1 r_1 (p_i - x_i) + c_2 r_2 (g - x_i),$$

где w — коэффициент инерции;

c_1 и c_2 — когнитивный и социальный коэффициенты, задающие баланс между эксплуатацией личного опыта и коллективного;

r_1 и r_2 — равномерно распределенные случайные числа на $[0, 1]$, делающие поиск стохастическим.

В программной реализации применена линейно убывающая инерция ($0,9 \rightarrow 0,4$ по итерациям), фиксированные $c_1 = c_2 = 1,5$ и ограничение скорости на уровне половины диапазона переменных, что препятствует выходу частиц за пределы рабочей области. Для траекторий умеренной размерности (10–20 параметров) сходимость достигается за 50–100 итераций при рое из 50 частиц [16].

Устойчивую работу показывает двухэтапная стратегия. Сначала планировщик (A^* на регулярной сетке или RRT* в непрерывном пространстве) строит геометрически допустимый маршрут, обходящий зоны запрета и статические препятствия. Полученная последовательность точек служит начальным приближением для метаэвристики (ГА или PSO), которая уже с учетом динамических ограничений БПЛА сглаживает траекторию, оптимизирует ее по выбранному критерию (расход энергии, время полета) и при необходимости добавляет горизонтальные обходы ветровых зон. Такое разделение сводит вычислительно затратную глобальную оптимизацию к локальной доработке уже осмысленного маршрута [11, 14].

Программные средства моделирования полета БПЛА

Инструментом, на протяжении многих лет удерживающим доминирующее положение, остается связка MATLAB/Simulink: пакеты Aerospace Blockset и Control System Toolbox предоставляют готовые модели атмосферы, датчиков и типовые звенья регуляторов, а графический редактор упрощает отладку многоконтурных схем [5, 8]. Недостатки тоже известны: закрытый формат и стоимость коммерческой лицензии, не всегда подъемная для малых коллективов и научных целей. Открытая альтернатива — Python: NumPy закрывает матричные операции, SciPy содержит готовые интеграторы и оптимизаторы, Matplotlib покрывает двумерную и трехмерную визуализацию. В одном из примеров разработанного программного комплекса [17] перечисленные схемы интегрирования (Euler, Improved Euler, RK2, RK4, RKF45) и метаэвристики (ГА, PSO) реализованы как отдельные классы с общим

интерфейсом, что дает возможность сопоставлять методы на одной и той же задаче при единых входных данных.

На стыке моделирования и реальной эксплуатации работают открытые автопилотные стеки ArduPilot и PX4. Их режим SITL (Software-in-the-Loop) запускает тот же исполняемый код, что и на борту, но с подменой аппаратных интерфейсов на сокет, по которым данные состояния и управления передаются во внешний симулятор (Gazebo, jMAVSim, FlightGear). Наземные станции Mission Planner и QGroundControl работают одинаково с виртуальным и реальным бортом по протоколу MAVLink, что позволяет переносить отлаженные в симуляции миссии без модификаций. Для исследовательских задач это дает возможность автоматизированного сравнения телеметрии SITL с результатами работы собственной модели на одних и тех же входных сигналах.

Анализ методов моделирования

Изучение особенностей рассмотренных методов не ставит целью выбор «лучшего» подхода: его обнаружение зависит от конкретной задачи и поставленных критериев. Для бортовой реализации, где вычислительный цикл должен укладываться в строго детерминированное время, разумным остается фиксированный шаг RK4, поскольку его стоимость заранее известна и не зависит от текущего режима. В кабинетных условиях, где требование реального времени отсутствует, удобнее адаптивный RK45: он сам подстраивает шаг под текущую динамику, экономя вычисления на спокойных участках и укрупняя сетку там, где можно.

Конкретизируем выявленные закономерности. Отдача от повышения порядка интегратора быстро убывает: переход от Эйлера к RK2 сокращает ошибку на два-три порядка, переход от RK2 к RK4 — еще на три-четыре порядка, дальнейшее усложнение (формулы седьмого-восьмого порядков типа Дорманда — Принса) при сопоставимом шаге дает уже не пропорциональный прирост точности, а кратно меньший, причем стоимость шага растет пропорционально числу обращений

к правой части. Адаптивные вложенные схемы выигрывают у схем с фиксированным шагом на траекториях с разнородной динамикой, и этот выигрыш измеряется десятками процентов по числу обращений к правой части. Плата за экономию — непредсказуемость длительности отдельного цикла, ограничивающая применение адаптации в системах реального времени. Среди планировщиков маршрута A^* и стохастические метаэвристики (ГА, PSO) не заменяют друг друга: графовый поиск гарантирует оптимальность на дискретной сетке при допустимой эвристике, а метаэвристики работают там, где явная дискретизация невозможна или ведет к «проклятию размерности» (платой за гибкость является отсутствие математической гарантии на конечном числе итераций).

Заключение

Сравнение методов, применяемых в моделировании полетных характеристик беспилотных летательных аппаратов, по вычислительной трудоемкости, точности и применимости к различным режимам полета позволяет систематизировать инструменты численного воспроизведения динамики БПЛА самолетного типа. Модель с шестью степенями свободы (12 ОДУ для линейных скоростей, угловых скоростей, координат и углов Эйлера) остается стандартом «по умолчанию» для задач проектирования контура стабилизации и анализа устойчивости. Среди классических схем интегрирования метод RK4 при шаге порядка 10^{-2} с. обеспечивает компромисс точности и предсказуемой вычислительной нагрузки, приемлемый для большинства задач. Адаптивная схема Фельберга RK45 оправдана там, где диапазон характерных времен процесса превышает один порядок, прежде всего на маневрах с резким изменением угла атаки или тяги.

В части оптимизации маршрута обоснована двухуровневая архитектура: графовый планировщик (A^* или RRT*) формирует геометрически допустимое начальное приближение, метаэвристика (ГА или PSO) выполняет его энергетическую и кинематическую доводку. Такое разделение снижает вычислительную сложность без потери качества

решения. Дальнейшие исследования могут быть направлены на разработку автоматического выбора пары «интегратор — планировщик» в зависимости от типа БПЛА и параметров полетного зада-

ния, а также интеграцию программного комплекса с открытыми автопилотными стеками через SITL для инструментальной проверки получаемых траекторий на пилотажных моделях.

СПИСОК ИСТОЧНИКОВ

1. Лебедев А. А., Чернобровкин Л. С. Динамика полета беспилотных летательных аппаратов: учебное пособие для вузов. 2-е изд., перераб. и доп. М.: Машиностроение, 1973. 616 с.
2. Управление и наведение беспилотных маневренных летательных аппаратов на основе современных информационных технологий / под ред. М. Н. Красильщикова и Г. Г. Себрякова. М.: Физматлит, 2003. 280 с.
3. Бюшгенс Г. С., Студнев Р. В. Аэродинамика самолета. Динамика продольного и бокового движения. М.: Машиностроение, 1979. 352 с.
4. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы: учебник. 12-е изд. М.: Лаборатория знаний, 2024. 639 с.
5. Stevens B. L., Lewis F. L., Johnson E. N. Aircraft Control and Simulation: Dynamics, Controls Design, and Autonomous Systems. Third Edition. Hoboken (NJ): Wiley-Blackwell, 2015. 768 p.
6. Etkin B., Reid L. D. Dynamics of Flight: Stability and Control. Third Edition. Hoboken (NJ): Wiley, 1996. 400 p.
7. Мхитарян А. М. Аэродинамика: учебник для вузов. 2-е изд., перераб. и доп. М.: КНОРУС, 2012. 448 с.
8. Ким Д. П. Теория автоматического управления: учебник для вузов. Т. 1. Линейные системы. 2-е изд., испр. и доп. М.: Физматлит, 2016. 312 с.
9. Хайпер Э., Нёрсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи / пер. с англ. под ред. С. С. Филиппова. М.: Мир, 1990. 512 с.
10. Боднер В. А. Системы управления летательными аппаратами: учебник для вузов. М.: Машиностроение, 1973. 504 с.
11. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учебное пособие. 2-е изд. М.: Изд-во МГТУ им. Н. Э. Баумана, 2017. 446 с.
12. Алгоритмы: построение и анализ. Третье издание = Introduction to Algorithms. Third Edition / Кормен Т. Х. [и др.]; пер. с англ. и ред. И. В. Красикова. М.: Вильямс, 2013. 1328 с.
13. Karaman S., Frazzoli E. Sampling-Based Algorithms for Optimal Motion Planning // International Journal of Robotics Research. 2011. Vol. 30, no. 7. Pp. 846–894. DOI: 10.1177/0278364911406761
14. LaValle S. M. Planning Algorithms. Cambridge: Cambridge University Press, 2006. 844 p.
15. Гладков Л. А., Курейчик В. В., Курейчик В. М. Генетические алгоритмы: учебник. 2-е изд., испр. и доп. М.: Физматлит, 2010. 368 с.
16. Kennedy J., Eberhart R. Particle Swarm Optimization // Proceedings of ICNN '95 — International Conference on Neural Networks (Perth, Australia, 27 November — 1 December 1995). Vol. 4. Institute of Electrical and Electronics Engineers, 1995. Pp. 1942–1948. DOI: 10.1109/ICNN.1995.488968
17. Beard R. W., McLain T. W. Small Unmanned Aircraft: Theory and Practice. Second Edition. Princeton (NJ): Princeton University Press, 2012. 320 p.

Дата поступления: 20.05.2026

Решение о публикации: 26.05.2026

Review and Comparative Analysis of Methods for Modeling Flight Characteristics of Unmanned Aerial Vehicles

Ivan V. Strelkov

— Postgraduate Student of the “Information and Computing Systems” Department. Research interests: mathematical modeling, numerical methods, unmanned aerial vehicles. E-mail: ivan_strelkov99@mail.ru

Lesya M. Bozhko

— Dr. Sci. in Economics, Professor of the “Information and Computing Systems” Department. Research interests: mathematical modeling, numerical methods, program complexes, management in organizational systems. E-mail: lemib@rambler.ru

Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky ave., Saint Petersburg, 190031, Russia

For citation: Strelkov I. V., Bozhko L. M. Review and Comparative Analysis of Methods for Modeling Flight Characteristics of Unmanned Aerial Vehicles, *Intellectual Technologies on Transport*, 2026, no. 2 (46), pp. 91–99. DOI: 10.20295/2413-2527-2026-246-91-99 (In Russian)

Abstract. *With the expansion of the scope of unmanned aerial vehicles, reliable forecasting of their behavior at the design stage is becoming an increasingly urgent task. This problem can be solved using appropriate modeling methods. Purpose: to identify and compare the methods used in simulating the flight characteristics of unmanned aerial vehicles. Results: a model of spatial motion with six degrees of freedom (6DoF, Six Degrees of Freedom) was considered, a comparison was made of the main methods used to simulate flight characteristics (Euler’s method, the classic Runge — Kutta fourth-order method and the adaptive Runge — Kutta — Felberg scheme), route planning algorithms (A^* , RRT*, genetic particle swarm algorithm and method. It has been established that for most research problems, the Runge — Kutta method of the fourth order turns out to be a reasonable compromise in accuracy and computational cost, while adaptive schemes are preferred in modes with different-scale dynamics. Recommendations are formulated on the selection of the method depending on the class of the problem and the permissible error. Practical significance: the results of the work are applicable to the creation of flight simulation software and the selection of a computing core for a specific class of unmanned aerial vehicles.*

Keywords: *unmanned aerial vehicle, mathematical modeling, flight dynamics, numerical methods, Runge — Kutta method, trajectory optimization, A^* algorithm, particle swarm optimization, flight simulation*

REFERENCES

1. Lebedev A. A., Chernobrovkin L. S. *Dinamika poleta bespilotnykh letatelnykh apparatov: uchebnoe posobie dlya vuzov* [Flight Dynamics of Unmanned Aerial Vehicles: A Textbook for Universities]. Moscow, Mashinostroenie Publishing House, 1973, 616 p. (In Russian)
2. Krasilshchikov M. N., Sebyakov G. G. (eds). *Upravlenie i navedenie bespilotnykh manevrennykh letatelnykh apparatov na osnove sovremennykh informatsionnykh tekhnologiy* [Control and Guidance of Unmanned Maneuverable Aircraft Based on Modern Information Technologies]. Moscow, Fizmatlit Publishing House, 2003, 280 p. (In Russian)
3. Byushgens G. S., Studnev R. V. *Aerodinamika samoleta. Dinamika prodolnogo i bokovogo dvizheniya* [Aerodynamics of an Aircraft. Dynamics of Longitudinal and Lateral Motion]. Moscow, Mashinostroenie Publishing House, 1979, 352 p. (In Russian)
4. Bakhvalov N. S., Zhidkov N. P., Kobelkov G. M. *Chislennyye metody: uchebnik* [Numerical Methods: A Textbook]. Moscow, Laboratoriya Znaniy Publishing House, 2024, 639 p. (In Russian)

5. Stevens B. L., Lewis F. L., Johnson E. N. Aircraft Control and Simulation: Dynamics, Controls Design, and Autonomous Systems. Third Edition. Hoboken (NJ), Wiley-Blackwell, 2015, 768 p.
6. Etkin B., Reid L. D. Dynamics of Flight: Stability and Control. Third Edition. Hoboken (NJ), Wiley, 1996, 400 p.
7. Mkhitaryan A. M. Aerodinamika: uchebnik dlya vuzov [Aerodynamics: A Textbook for Universities]. Moscow, KnoRus Publishing House, 2012, 448 p. (In Russian)
8. Kim D. P. Teoriya avtomaticheskogo upravleniya: uchebnik dlya vuzov. T. 1. Lineynye sistemy [Automatic Control Theory: A Textbook for Universities. Vol. 1. Linear Systems]. Moscow, Fizmatlit Publishing House, 2016, 312 p. (In Russian)
9. Hairer E., Nørsett S., Wanner G. Reshenie obyknovennykh differentsialnykh uravneniy. Nezhestkie zadachi [Solution of ordinary differential equations. Non-rigid problems]. Moscow, Mir Publishers, 1990, 512 p. (In Russian)
10. Bodner V. A. Sistemy upravleniya letatelnyimi apparatami: uchebnik dlya vuzov [Aircraft Control Systems: A Textbook for Universities]. Moscow, Mashinostroenie Publishing House, 1973, 504 p. (In Russian)
11. Karpenko A. P. Sovremennye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennyye prirodoy: uchebnoe posobie [Modern Search Engine Optimization Algorithms. Nature-Inspired Algorithms: A Textbook]. Moscow, Bauman Moscow State Technical University Press, 2017, 446 p. (In Russian)
12. Cormen T. H., et al. Algoritmy: postroenie i analiz. Tret'ye izdanie [Introduction to algorithms. Third Edition]. Moscow, Williams Publishing House, 2013, 1328 p. (In Russian)
13. Karaman S., Frazzoli E. Sampling-Based Algorithms for Optimal Motion Planning, *International Journal of Robotics Research*, 2011, vol. 30, no. 7, pp. 846–894. DOI: 10.1177/0278364911406761
14. LaValle S. M. Planning Algorithms. Cambridge, Cambridge University Press, 2006, 844 p.
15. Gladkov L. A., Kureychik V. V., Kureychik V. M. Geneticheskie algoritmy: uchebnik [Genetic Algorithms: A Textbook]. Moscow, Fizmatlit Publishing House, 2010, 368 p. (In Russian)
16. Kennedy J., Eberhart R. Particle Swarm Optimization, *Proceedings of ICNN '95 — International Conference on Neural Networks*, Perth, Australia, November 27 – December 01, 1995, vol. 4. Institute of Electrical and Electronics Engineers, 1995, pp. 1942–1948. DOI: 10.1109/ICNN.1995.488968
17. Beard R. W., McLain T. W. Small Unmanned Aircraft: Theory and Practice. Second Edition. Princeton (NJ), Princeton University Press, 2012, 320 p.

Received: May 20, 2026

Accepted: May 26, 2026

УДК 656.073

Особенности модели выкладки товаров с иерархической категоризацией на полочном пространстве магазинов и складов

Черняховская
Екатерина
Сергеевна

— кандидат наук, преподаватель. Научные интересы: оптимизация, программирование, распределение полочного пространства.
E-mail: kateryna.czerniachowska@ue.wroc.pl

Вроцлавский университет экономики и бизнеса, Польша, 53-345, Вроцлав, ул. Командорская, 118-120

Для цитирования: Черняховская Е. С. Особенности модели выкладки товаров с иерархической категоризацией на полочном пространстве магазинов и складов // Интеллектуальные технологии на транспорте. 2026. № 2 (46). С. 100–108. DOI: 10.20295/2413-2527-2026-246-100-108

Аннотация. Эффективная складская логистика является основополагающим фактором в поддержании ассортимента товаров на полках магазина и склада. Надежная система управления складом позволяет своевременно пополнять запасы и избежать дефицита популярных товаров. Кроме того, оптимизация складских процессов способствует снижению затрат и повышению общей эффективности розничной торговли. **Цель:** построение и анализ математической модели выкладки товаров с иерархической категоризацией товаров на складском стеллаже и на стеллаже розничного магазина. Максимизация прибыли со стеллажа представлена как целевая функция. **Результаты:** сформулированы следующие категории ограничений: ограничения на полку, ограничения на расположение на нескольких полках, ограничения на товар, ограничения на положение, ограничения на категории и подкатегории. Приводятся особенности, достоинства и недостатки иерархической категоризации товаров. Делается сравнение иерархической категоризации товаров на складах и в обычных магазинах. **Практическая значимость:** задача оптимального размещения товаров на полках имеет важное значение для складской логистики и для розничных продаж, так как помогает улучшить процесс комплектования, доставки заказов, а также видимость товаров на полках. Понимание этих аспектов позволяет складам и дистрибьюторам эффективно управлять запасами, повышать прибыль и снижать затраты, обеспечивая высокую скорость и точность выполнения заказов, увеличивать частоту покупок.

Ключевые слова: математическое моделирование, оптимизация, распределение полочного пространства, складская логистика, категоризация товаров

1.2.2 — математическое моделирование, численные методы и комплексы программ (технические науки)

Введение

От того, как организована складская логистика, напрямую зависит, появятся ли нужные товары в магазинах вовремя. Если запасы пополняются своевременно, магазины не сталкиваются с нехваткой популярных товаров. В свою очередь, это повышает лояльность покупателей [1, 2]. Когда складские процессы автоматизируются и опира-

ются на аналитику, становится проще предсказывать спрос и сокращать расходы. Когда склады и магазины работают слаженно, товары приходят без задержек и быстрее доходят до полок. Следовательно, улучшаются доступность и разнообразие ассортимента. Все это влияет на рост продаж и привлечение клиентов.

Эффективность розничной торговли зависит от производительности цепочки поставок [3]. Внедрение современных технологий автоматизации в распределительных центрах позволяет ускорить сборку и обработку заказов. Сокращение времени доставки повышает уровень удовлетворенности клиентов и укрепляет их лояльность. Компании, активно внедряющие инновации в логистические процессы, становятся более конкурентоспособными на рынке.

Место на полках всегда ограничено, поэтому за ним приходится следить и использовать с максимальной отдачей. Основные цели розничного продавца: увеличение прибыли, снижение расходов, повышение удовлетворенности клиентов.

Различают две ключевые проблемы: выбор ассортимента и распределение полочного пространства. Первая заключается в определении того, какие товары должны быть представлены на ограниченном пространстве, вторая — в организации их фактического размещения на полках и определении оптимального количества единиц каждой позиции [4, 5].

Если грамотно управлять выкладкой, магазин сможет и заработать больше, и снизить свои расходы. Решения о размещении товаров должны опираться на анализ спроса и прибыльности каждой позиции. Хорошо продуманная выкладка делает востребованные товары заметнее и удобнее для покупателя, что улучшает его впечатление от магазина.

Хорошо известно, если соседние товары логично дополняют друг друга, покупатели чаще берут их вместе, что повышает продажи. Так, установлено, что вертикальное расположение товаров приблизительно в два раза эффективнее влияет на производительность магазина, чем горизонтальное [6]. В [7] изучалась связь расстояния между продуктами на полке и их продажами, а также влияние размера полочного пространства на продажи с учетом расположения товаров на полках. Для этой цели была создана методика, основанная на ассоциативных правилах. В соответствии с выводами, сделанными в [8], продажи больше всего зависят от того, на каком уровне на-

ходится товар, а вот размер самой полки влияет слабее — главное, чтобы запас не опускался ниже минимума.

Продажи растут, когда товары расставлены на полках с учетом того, как именно покупатели смотрят и выбирают продукты. Особенно важным является вертикальное размещение на уровне глаз покупателей, поскольку такое положение привлекает внимание к товару и стимулирует импульсные покупки.

Исследования показывают: когда выкладку подстраивают под реальные привычки покупателей, магазины лучше отвечают их ожиданиям. Размещение наиболее востребованных товаров на уровне глаз покупателей способствует более легкому их обнаружению и повышает вероятность покупки.

При этом сильнее всего на продажи влияет именно вертикальное размещение товара, а не то, где он стоит по горизонтали [6, 9–12]. В [9] определены четыре вертикальных уровня полок: уровень шляпы, уровень глаз, уровень руки и уровень колена. Товары на уровне глаз сразу притягивают взгляд покупателя, поэтому магазины продают их лучше [13]. В связи с этим, чтобы получить более высокую выручку, розничные торговцы должны размещать продукты на уровне глаз.

Итог простой: если учитывать, как покупатели осматривают полки, и располагать товары на удобных уровнях, магазин сможет увеличить продажи и улучшить впечатление от покупки. В итоге продуманная выкладка, особенно с учетом высоты полок, помогает увеличить продажи и средний чек.

Постановка задачи

Предлагаемая в статье модель отличается от существующих тем, что вводит иерархический способ распределения товаров на полках, при котором выделены главные категории, а подкатегории товаров можно расположить горизонтально, вертикально, квадратными или прямоугольными блоками на нескольких полках. Так, в [14] приводится одновременная категоризация товаров по вертикали и по горизонтали с присвоением нескольких меток каждому товару. В [15] приводится вложенная

категоризация товаров, в [16, 17] — категоризация товаров по нескольким горизонтальным уровням.

Параметры и индексы, используемые в модели:

- S — количество полок;
- P — количество товаров;
- K — количество категорий;
- G — количество иерархических подкатегорий;
- i, a, b — индекс полки, $i, a, b = 1, \dots, S$;
- j, q — индекс товара, $j, q = 1, \dots, P$;
- k — индекс категории, $k = 1, \dots, K$;
- g — индекс подкатегории, $g = 1, \dots, G$;
- r — индекс положения,

$$r = \left\{ \begin{array}{l} 1, \text{ горизонтально} \\ 2, \text{ вертикально} \\ 3, \text{ лицом вверх или вниз} \end{array} \right\}.$$

Параметры полки i :

- s_i^l — длина;
- s_i^h — высота;
- s_i^d — глубина.

Параметры товара j :

- p_j^w — ширина;
- p_j^h — высота;
- p_j^d — глубина;
- p_j^u — прибыль;
- p_j^k — категория;
- p_j^g — подкатегория;
- p_j^s — лимит поставки;
- f_j^{\min} — минимальное количество единиц товара;

вара;

- f_j^{\max} — максимальное количество единиц товара;

вара;

- s_j^{\min} — минимальное количество полок, на которых можно разместить товар;

которых можно разместить товар;

- s_j^{\max} — максимальное количество полок, на которых можно разместить товар;
- p_{jr}^w — параметр ширины в зависимости от положения,

$$p_{jr}^w = \left\{ \begin{array}{l} p_j^w, \text{ если } r = 1, \text{ ширина при} \\ \text{горизонтальном положении} \\ p_j^d, \text{ если } r = 2, \text{ глубина при вертикальном} \\ \text{положении} \\ p_j^h, \text{ если } r = 3, \text{ высота при положении} \\ \text{лицом вверх или вниз} \end{array} \right\},$$

- p_{jr}^o — бинарный параметр положения,

$$p_{jr}^o = \left\{ \begin{array}{l} 1, \text{ если определенное положение допустимо} \\ 0, \text{ в противном случае} \end{array} \right\}.$$

Параметры категории k и подкатегории g :

- c_k^m — минимальная ширина категории в процентах;
- c_k^t — допуск ширины категории на разных полках в процентах;
- c_g^t — допуск ширины подкатегории на разных полках в процентах.

Переменные задачи:

$$x_{ijr} = \left\{ \begin{array}{l} 1, \text{ если товар } j \text{ положен на полку } i \\ \text{в положении } r \\ 0, \text{ в противном случае} \end{array} \right\} —$$

бинарная переменная расположения товара на полке;

- f_{ijr} — количество единиц товара j на полке i в положении r .

В данном исследовании задача оптимизации полочного пространства на стеллаже с иерархической категоризацией товаров представлена таким образом, что стеллаж визуально разделен на вертикальные категории, что облегчает поиск товаров на стеллаже. В пределах каждой категории стеллаж разделен на вертикальные, горизонтальные или блочные подкатегории. Такое расположение помогает осуществлять поиск товаров в пределах главной категории.

На рисунке представлен пример расположения товаров с иерархической категоризацией на стеллаже с двумя вертикальными категориями. В категории А иерархические подкатегории товаров располагаются вертикально. Товары одной подкатегории располагаются на всех полках, однако граница между подкатегориями заметна. Гибкость границы между подкатегориями контролируется параметром c_g^t , а гибкость границ между категориями, ввиду чего категории также становятся визуально отделяемыми друг от друга на стеллаже, контролируется параметром c_k^t . В категории Б товары могут располагаться горизонтальными и вертикальными рядами, а также прямоугольными блоками на нескольких полках. Товары подкатегории Б1.1 и Б3.2 располагаются квадратными блоками на двух полках. Товары подкатегории Б2.1 поставлены обособленно на полке. Товары подкатегории

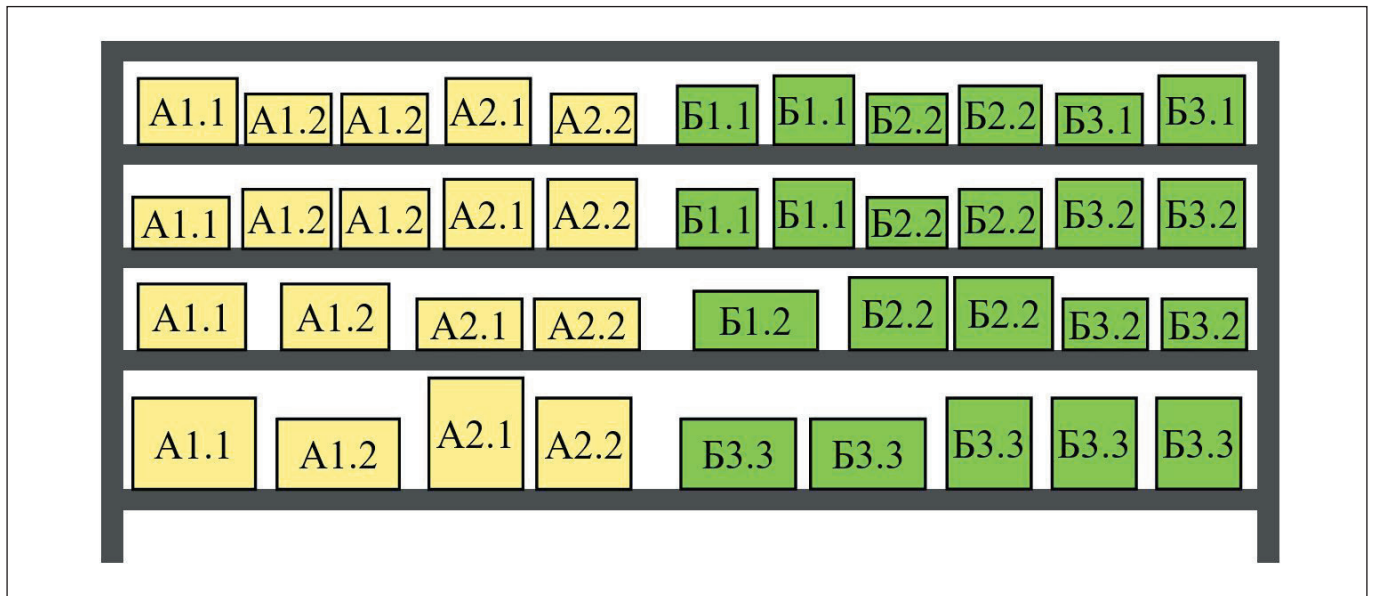


Рисунок. Стеллаж с иерархической категоризацией товаров

Б2.2 располагаются вертикальными блоками на трех полках. Товары подкатегории Б3.1 и Б3.3 располагаются горизонтальными блоками на одной полке. Отсутствует подкатегория Б2.1, зато в категории Б3 имеется целых три уровня, а не два, как в остальных, то есть количество подкатегорий в каждой категории может варьироваться.

Для решения поставленной задачи дистрибьютор должен определить полки для размещения товара, количество каждого товара на полке, а также положение товара на полке, учитывая разного рода ограничения: ограничения на полку, ограничения на расположение на нескольких полках, ограничения на товар, ограничения на положение, ограничения на категории и подкатегории. Целью размещения товаров на стеллаже является максимизация общей прибыли от использования стеллажного пространства.

В данной задаче используются следующие переменные:

- x_{ijr} — товар j размещен на полке i в положении r ;
- f_{ijr} — количество единиц товара j на полке i в положении r .

Проблему можно сформулировать как

$$\max \sum_{j=1}^P \sum_{i=1}^S \sum_{r=1}^R p_j^u f_{ijr}$$

при следующих ограничениях:

1. Ограничения по полке:
 - длина полки — $\forall(i) [\sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr} \leq s_i^w]$;
 - глубина полки в зависимости от положения товара — $\forall(i, j, p_{jr}^d > s_i^d) [f_{ijr} = 0]$;
 - высота полки в зависимости от положения товара — $\forall(i, j, p_{jr}^h > s_i^h) [f_{ijr} = 0]$.
2. Ограничения на расположение товара на нескольких полках:
 - минимальное и максимальное количество полок — $\forall(j) [s_j^{\min} \leq \sum_{i=1}^S \sum_{r=1}^R x_{ijr} \leq s_j^{\max}]$;
 - одинаковое количество единиц одного и того же товара на каждой полке ($\lceil \cdot \rceil$ — округленное значение до ближайшего целого вверх) —

$$\forall(i, j) \left[\sum_{r=1}^R x_{ijr} \leq \left\lceil \frac{\sum_{a=1}^S \sum_{r=1}^R f_{ajr}}{\sum_{a=1}^S \sum_{r=1}^R x_{ajr}} \right\rceil \right];$$

- при расположении подкатегории товара на нескольких полках ее можно поставить только на подряд идущей полке —

$$\forall(g) \forall(j, q : p_j^g = p_q^g, j, q = 1, \dots, P)$$

$$\forall(a, b : |a - b| \neq 1 \wedge a < b, a, b = 1, \dots, S)$$

$$\left[\sum_{r=1}^R x_{ajr} + \sum_{r=1}^R x_{bqr} \leq 1 \right].$$

3. Ограничения на товар:

- минимальное и максимальное количество единиц товара — $\forall(i, j)[x_{ijr} \leq \sum_{r=1}^R f_{ijr} \leq f_j^{\max}]$;
- лимит поставки, если один и тот же товар размещен на нескольких полках, — $\forall(j)[\sum_{i=1}^S \sum_{r=1}^R f_{ijr} \leq p_j^s]$;
- товар поставлен на полку — $\forall(i, j, r)[f_{ijr} \leq x_{ijr} f_j^{\max}]$.

4. Ограничения на положение товара:

- определенное положение товара допустимо — $\forall(i, j, r)[x_{ijr} \leq p_{jr}^o]$;
- допустимо только одно положение для конкретного товара — $\forall(i, j)[\sum_{r=1}^R x_{ijr} \leq 1]$.

5. Ограничения на категории и подкатегории:

- минимальная ширина категории, если категория должна быть на стеллаже, —
- $\forall(i, k)[(\sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr} \geq [s_i \cdot c_k^m]) \vee (\sum_{j=1}^P \sum_{r=1}^R f_{ijr} = 0)]$;
- допуск по ширине категории на разных полках —

$$\forall(k, s)[(\sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr} \geq \frac{\sum_{a=1}^S \sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr}}{\sum_{a=1}^S \min(\sum_{j=1}^P \sum_{r=1}^R x_{ijr}, 1)} \cdot (1 - c_k^t) \wedge \wedge \sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr} \leq \frac{\sum_{a=1}^S \sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr}}{\sum_{a=1}^S \min(\sum_{j=1}^P \sum_{r=1}^R x_{ijr}, 1)} \times \times (1 + c_k^t) \vee (\sum_{j=1}^P \sum_{r=1}^R f_{ijr} = 0)];$$

- допуск по ширине подкатегории на разных полках ($[]$ — округленное значение) —

$$\forall(g, s)[(\sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr} \geq$$

$$\geq \frac{\sum_{a=1}^S \sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr}}{\sum_{a=1}^S \min(\sum_{j=1}^P \sum_{r=1}^R x_{ijr}, 1)} \cdot (1 - c_g^t) \wedge$$

$$\wedge \sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr} \leq \frac{\sum_{a=1}^S \sum_{j=1}^P \sum_{r=1}^R p_{jr}^w f_{ijr}}{\sum_{a=1}^S \min(\sum_{j=1}^P \sum_{r=1}^R x_{ijr}, 1)} \times$$

$$\times (1 + c_g^t) \vee (\sum_{j=1}^P \sum_{r=1}^R f_{ijr} = 0)].$$

6. Ограничение отношений:

- соотношение единиц товара и длины полки — $\forall(i, j, r)[\frac{s_i^l x_{ijr}}{p_{jr}^w} \geq f_{ijr}]$.

7. Переменные решения:

- товар размещен на полке — $\forall(i, j, r)[x_{ijr} \in \{0, 1\}]$;
- количество единиц товара на полке — $\forall(i, j, r)[f_{ijr} = \{f_j^{\min} \dots f_j^{\max}\}]$.

Дискуссия

Практика организации товаров по иерархическим категориям применима как в магазинах, так и на складах. Обычно речь идет о разделении ассортимента на крупные группы, подгруппы и, при необходимости, более мелкие подгруппы. Такой подход облегчает поиск нужных товаров, позволяет лучше контролировать запасы и рационально использовать полочное пространство.

Однако важно понимать: эффективность системы сильно зависит не только от самой структуры, но и от того, насколько сотрудники соблюдают правила ее использования. Иногда даже продуманная схема теряет смысл без регулярного контроля и корректировки под изменения ассортимента.

Основные достоинства:

1. Упрощенная навигация. Для покупателей магазинов наличие категорий облегчает ориен-

тирование, особенно если речь идет о большом ассортименте. В этом случае покупатель быстрее находит нужный товар. На складе такая система помогает сотрудникам быстрее ориентироваться в ячейках и уменьшает количество ошибок при сборке заказов.

2. Рациональное использование пространства. Разделение товаров по группам позволяет компактнее размещать их на полках, избегая пустующих мест или чрезмерного нагромождения. На складах логично распределенные ячейки повышают вместимость стеллажей и упрощают работу с крупными партиями продукции.

3. Контроль запасов. Структурированная выкладка облегчает отслеживание остатков и сроков годности. Особенно это заметно в магазинах, где важно предотвратить дефицит или переполнение полок. На складах такая система снижает количество ошибок при инвентаризации и делает учет более точным.

4. Стимулирование продаж. Размещение взаимосвязанных товаров рядом часто приводит к незапланированным покупкам. Например, покупатели охотнее берут вместе сопутствующие продукты, если они расположены поблизости. А это очень важно для повышения оборота.

5. Снижение времени на сборку заказов. На складе товары, которые часто входят в одни и те же заказы, располагаются рядом. Это заметно ускоряет процесс комплектования и повышает производительность сборщиков.

6. Обучение персонала. Новым сотрудникам проще освоиться в логичной системе, а в периоды пиковой нагрузки можно быстрее привлечь работников из других отделов без снижения эффективности или длительного обучения.

Основные недостатки:

1. Ограниченная гибкость. Любые изменения ассортимента или перераспределение позиций требуют пересмотра всей структуры. Иногда это затрудняет оперативное реагирование на новые потребности или неожиданный спрос.

2. Сложности при расширении ассортимента. С увеличением числа товаров поддержание иерархии становится более трудоемким. Введение

новых позиций может нарушить устоявшуюся логику, что усложняет навигацию.

3. Начальные затраты. Создание и внедрение структуры требует времени и ресурсов: проектирование, обучение персонала, перестройка пространства. Тем не менее при правильной реализации вложения окупаются за счет сокращения ошибок и повышения эффективности.

4. Риск мертвых зон. Некоторые товары могут оказаться в местах с низкой проходимостью, что сказывается на продажах. На складе труднодоступные позиции замедляют сборку и создают дополнительную нагрузку для сотрудников.

5. Сезонные товары и акции. Временное размещение сезонной продукции или акционных товаров может нарушить постоянную структуру. Это требует дополнительного планирования и временного перераспределения зон хранения или торгового пространства.

Методы организации ассортимента

Подходы к классификации зависят от задач, объема ассортимента и особенностей работы персонала и покупателей:

1. Общие методы: деление на крупные группы (например, продукты питания, электроника, косметика), иерархия с подгруппами, сортировка по размеру и весу, размещение по принципу FIFO для скоропортящихся товаров.

2. Складские методы: алфавитная сортировка, сортировка по артикулу, по частоте использования — востребованные товары располагаются ближе к зоне отгрузки.

3. Магазинные методы: бренды, целевая аудитория, ценовые категории, назначение, сезонность, алфавит (например, для книжных магазинов).

Выбор подхода зависит от специфики ассортимента и логистики. Часто приходится искать баланс между удобством для покупателей и эффективностью использования пространства.

Заключение

В настоящем исследовании представлена математическая модель выкладки товаров на стеллаже с применением иерархической категоризации товаров.

Данная модель позволяет организовать товары на полках таким образом, что формируются главные вертикальные категории, а в их пределах подкатегории товаров можно расположить горизонтально, вертикально, квадратными или прямоугольными блоками на нескольких полках. Иерархические подкатегории, как и категории, визуальны заметны на полках стеллажа.

Иерархическая категоризация позволяет формировать вертикальные категории и подкатегории, которые могут располагаться горизонтально, вертикально или блоками. На складах это ускоряет сборку и повышает точность учета, а в магазинах помогает рационально использовать пространство, улучшать визуальное восприятие ассортимента и стимулировать сопутствующие покупки.

При этом система требует регулярного контроля и адаптации под изменения ассортимента. С правильной поддержкой она остается одним из основных инструментов управления запасами

и организации выкладки, повышая удобство для покупателей и эффективность работы персонала.

Описываемая в исследовании модель может применяться как на складских стеллажах, так и на торговом оборудовании в магазинах. В дискуссии делается акцент на достоинствах и недостатках иерархической категоризации товаров, приводится сравнение иерархической категоризации товаров на складах и в обычных магазинах.

Предложенный подход способствует оптимизации управления запасами и организации процесса комплектования заказов на складах и в дистрибутивных центрах. В свою очередь, в магазинах он позволяет эффективно использовать принципы мерчандайзинга при оптимизации полочного пространства. Представленная модель также помогает улучшить видимость и доступность товаров для клиентов, повысить уровень их удовлетворенности при совершении покупок в магазине.

СПИСОК ИСТОЧНИКОВ

1. Yu M., de Koster R. B. M. The Impact of Order Batching and Picking Area Zoning on Order Picking System Performance // *European Journal of Operational Research*. 2009. Vol. 198, iss. 2. Pp. 480–490. DOI: 10.1016/j.ejor.2008.09.011
2. *Facilities Planning*. Fourth Edition / J. A. Tompkins [et al.]. Hoboken (NJ): Wiley, 2010. 864 p.
3. Bhattacharjya J., Ellison A., Tripathi S. An Exploration of Logistics-Related Customer Service Provision on Twitter: The Case of e-Retailers // *International Journal of Physical Distribution and Logistics Management*. 2016. Vol. 46, iss. 6–7. Pp. 659–680. DOI: 10.1108/IJPDLM-01-2015-0007
4. Hansen P., Heinsbroek H. Product Selection and Space Allocation in Supermarkets // *European Journal of Operational Research*. 1979. Vol. 3, iss. 6. Pp. 474–484. DOI: 10.1016/0377-2217(79)90030-4
5. Landa-Silva D., Marikar F., Le K. Heuristic Approach for Automated Shelf Space Allocation // SAC09: Proceedings of the ACM Symposium on Applied Computing (Honolulu, HI, USA, 08–12 March 2009). New York: Association for Computing Machinery, 2009. Pp. 922–928. DOI: 10.1145/1529282.1529482
6. Hansen J. M., Raut S., Swami S. Retail Shelf Allocation: A Comparative Analysis of Heuristic and Meta-Heuristic Approaches // *Journal of Retailing*. 2010. Vol. 86, iss. 1. Pp. 94–105. DOI: 10.1016/j.jretai.2010.01.004
7. Chen Y.-L., Chen J.-M., Tung C.-W. A Data Mining Approach for Retail Knowledge Discovery with Consideration of the Effect of Shelf-Space Adjacency on Sales // *Decision Support Systems*. 2006. Vol. 42, iss. 3. Pp. 1503–1520. DOI: 10.1016/j.dss.2005.12.004
8. Drèze X., Hoch S. J. Purk, M. E. Shelf Management and Space Elasticity // *Journal of Retailing*. 1994. Vol. 70, iss. 4. Pp. 301–326. DOI: 10.1016/0022-4359(94)90002-7
9. Ebster C., Garaus M. *Store Design and Visual Merchandising: Creating Store Space That Encourages Buying*. New York: Business Expert Press, 2011. 150 p.
10. Elbers T. The Effects of In-Store Layout and Shelf Designs on Consumer Behavior. 2016. 22 p. URL: <http://edepot.wur.nl/369091> (дата обращения: 16.03.2025).
11. Valenzuela A., Raghubir P., Mitakakis C. Shelf Space Schemas: Myth or Reality? // *Journal of Business Research*. 2013. Vol. 66, iss. 7. Pp. 881–888. DOI: 10.1016/j.jbusres.2011.12.006

12. Wongkitrungrueng A., Valenzuela A., Sen S. The Cake Looks Yummy on the Shelf up There: The Interactive Effect of Retail Shelf Position and Consumers' Personal Sense of Power on Indulgent Choice // *Journal of Retailing*. 2018. Vol. 94, iss. 3. Pp. 280–295. DOI: 10.1016/j.jretai.2018.07.001
13. Wright B. L. More Sales, Less Effort — Part 1: The Elements of Visual Merchandising // *20/20 Magazine*. 2013. Vol. 40, iss. 13. Pp. 105–108.
14. Czerniachowska K., Lutosławski K., Fojcik M. Heuristics for Shelf Space Allocation Problem with Vertical and Horizontal Product Categorization // *Procedia Computer Science*. 2022. Vol. 207. Pp. 195–204. DOI: 10.1016/j.procs.2022.09.052
15. Czerniachowska K., Michalak K., Hernes M. Heuristics for the Shelf Space Allocation Problem // *OPSEARCH*. 2023. Vol. 60, iss. 2. Pp. 835–869. DOI: 10.1007/s12597-023-00636-1
16. Czerniachowska K., Hernes M. Simulated Annealing Hyper-Heuristic for a Shelf Space Allocation on Symmetrical Planograms Problem // *Symmetry*. 2021. Vol. 13, iss. 7. Art. no. 1182. 25 p. DOI: 10.3390/sym13071182
17. Czerniachowska K. A Genetic Algorithm for the Retail Shelf Space Allocation Problem with Virtual Segments // *OPSEARCH*. 2022. Vol. 59, iss. 1. Pp. 364–412. DOI: 10.1007/s12597-021-00551-3

Дата поступления: 30.03.2026

Решение о публикации: 25.05.2026

Hierarchical Shelf Space Allocation Model Characteristics in Stores and Warehouses

Ekaterina S. Cherniakhovskaia — PhD, lecturer. Research interests: optimization, programming, shelf space allocation. E-mail: kateryna.czerniachowska@ue.wroc.pl

Wroclaw University of Economics and Business, 118-120, Komandorska str., Wroclaw, 53-345, Poland

For citation: Cherniakhovskaia E. S. Hierarchical Shelf Space Allocation Model Characteristics in Stores and Warehouses, *Intellectual Technologies in Transport*, 2026, no. 2 (46), pp. 100–108. DOI: 10.20295/2413-2527-2026-246-100-108. (In Russian)

Abstract. *Efficient warehouse logistics is fundamental to maintaining a well-stocked product range on store shelves and in warehouses. A reliable warehouse management system enables timely replenishment and avoids stockouts of popular products. Furthermore, optimization of warehouse processes helps reduce costs and improve overall retail efficiency. **Purpose:** to present and analyze a mathematical model for product display with hierarchical categorization on warehouse shelves and retail store shelves. Maximizing shelf profit is the objective function. **Results:** the following constraint categories are formulated: shelf constraints, multi-shelf placement constraints, product constraints, position constraints, and category and subcategory constraints. The characteristics, advantages and disadvantages of hierarchical categorization of products are considered. A comparison is made between hierarchical product categorization in warehouses and traditional stores. **Practical significance:** optimal shelf placement is crucial for warehouse logistics and retail sales, as it helps improve order picking, delivery, and shelf visibility. Understanding these aspects allows warehouses and distributors to effectively manage inventory, increase profits, and reduce costs, ensuring high speed and accuracy of order fulfillment and increasing purchase frequency.*

Keywords: *mathematical modelling, optimization, shelf space allocation, warehouse logistics, product categorization*

REFERENCES

1. Yu M., de Koster R. B. M. The Impact of Order Batching and Picking Area Zoning on Order Picking System Performance, *European Journal of Operational Research*, 2009, vol. 198, iss. 2, pp. 480–490. DOI: 10.1016/j.ejor.2008.09.011
2. Tompkins J. A., et al. Facilities Planning. Fourth Edition. Hoboken (NJ), Wiley, 2010, 864 p.
3. Bhattacharjya J., Ellison A., Tripathi S. An Exploration of Logistics-Related Customer Service Provision on Twitter: The Case of e-Retailers, *International Journal of Physical Distribution and Logistics Management*, 2016, vol. 46, iss. 6–7, pp. 659–680. DOI: 10.1108/IJPDLM-01-2015-0007
4. Hansen P., Heinsbroek H. Product Selection and Space Allocation in Supermarkets, *European Journal of Operational Research*, 1979, vol. 3, iss. 6, pp. 474–484. DOI: 10.1016/0377-2217(79)90030-4
5. Landa-Silva D., Marikar F., Le K. Heuristic Approach for Automated Shelf Space Allocation, *SAC09: Proceedings of the ACM Symposium on Applied Computing*, Honolulu, HI, USA, March 08–12, 2009. New York, Association for Computing Machinery, 2009, pp. 922–928. DOI: 10.1145/1529282.1529482
6. Hansen J. M., Raut S., Swami S. Retail Shelf Allocation: A Comparative Analysis of Heuristic and Meta-Heuristic Approaches, *Journal of Retailing*, 2010, vol. 86, iss. 1, pp. 94–105. DOI: 10.1016/j.jretai.2010.01.004
7. Chen Y.-L., Chen J.-M., Tung C.-W. A Data Mining Approach for Retail Knowledge Discovery with Consideration of the Effect of Shelf-Space Adjacency on Sales, *Decision Support Systems*, 2006, vol. 42, iss. 3, pp. 1503–1520. DOI: 10.1016/j.dss.2005.12.004
8. Drèze X., Hoch S. J. Purk, M. E. Shelf Management and Space Elasticity, *Journal of Retailing*, 1994, vol. 70, iss. 4, pp. 301–326. DOI: 10.1016/0022-4359(94)90002-7
9. Ebster C., Garaus M. Store Design and Visual Merchandising: Creating Store Space That Encourages Buying. New York, Business Expert Press, 2011, 150 p.
10. Elbers T. The Effects of In-Store Layout and Shelf Designs on Consumer Behavior. 2016, 22 p. Available at: <http://edepot.wur.nl/369091> (accessed: March 16, 2025).
11. Valenzuela A., Raghubir P., Mitakakis C. Shelf Space Schemas: Myth or Reality?, *Journal of Business Research*, 2013, vol. 66, iss. 7, pp. 881–888. DOI: 10.1016/j.jbusres.2011.12.006
12. Wongkitrungrueng A., Valenzuela A., Sen S. The Cake Looks Yummy on the Shelf up There: The Interactive Effect of Retail Shelf Position and Consumers' Personal Sense of Power on Indulgent Choice, *Journal of Retailing*, 2018, vol. 94, iss. 3, pp. 280–295. DOI: 10.1016/j.jretai.2018.07.001
13. Wright B. L. More Sales, Less Effort — Part 1: The Elements of Visual Merchandising, *20/20 Magazine*, 2013, vol. 40, iss. 13, pp. 105–108.
14. Czerniachowska, K., Lutosławski K., Fojcik M. Heuristics for Shelf Space Allocation Problem with Vertical and Horizontal Product Categorization, *Procedia Computer Science*, 2022, vol. 207, pp. 195–204. DOI: 10.1016/j.procs.2022.09.052
15. Czerniachowska K., Michalak K., Hernes M. Heuristics for the Shelf Space Allocation Problem, *OPSEARCH*, 2023, vol. 60, iss. 2, pp. 835–869. DOI: 10.1007/s12597-023-00636-1
16. Czerniachowska K., Hernes M. Simulated Annealing Hyper-Heuristic for a Shelf Space Allocation on Symmetrical Planograms Problem, *Symmetry*, 2021, vol. 13, iss. 7, art. no. 1182, 25 p. DOI: 10.3390/sym13071182
17. Czerniachowska K. A Genetic Algorithm for the Retail Shelf Space Allocation Problem with Virtual Segments, *OPSEARCH*, 2022, vol. 59, iss. 1, pp. 364–412. DOI: 10.1007/s12597-021-00551-3

Received: March 30, 2026

Accepted: May 25, 2026

Рыжиков Юрий Иванович

10 октября 1933 — 29 апреля 2026

Сотрудники кафедр математического и программного обеспечения специальных космических комплексов, информационно-вычислительных систем и сетей специальных космических комплексов Военно-космической академии имени А.Ф. Можайского и СПб ФИЦ РАН с глубоким прискорбием сообщают, что 29 апреля 2026 года на 93-м году жизни после продолжительной болезни скончался наш коллега, выдающийся советский и российский ученый, заслуженный деятель науки Российской Федерации, почетный профессор академии, доктор технических наук, профессор, полковник в отставке Рыжиков Юрий Иванович.

Ю.И. Рыжиков родился 10 октября 1933 года в городе Ленинграде. В 1952 году по комсомольскому набору поступил в Высшее военно-морское училище инженеров оружия, после четвертого курса в составе факультета был переведен в Черноморское высшее военно-морское училище имени П.С. Нахимова, которое окончил с отличием в 1958 году.

Вся дальнейшая служба Юрия Ивановича проходила в Ленинграде: по ноябрь 1958 года — младший инженер 301-й группы сборочных бригад 6-го управления ВМФ; по сентябрь 1962 года — младший научный сотрудник 16-го института ВМФ (позднее — Морской филиал 12-го ЦНИИ МО); далее — адъюнкт Ленинградской Краснознаменной военно-воздушной инженерной академии имени А.Ф. Можайского.

Пройдя путь от преподавателя до начальника кафедры, профессор Рыжиков внес неоценимый вклад в развитие отечественной вычислительной техники, теории управления запасами и теории массового обслуживания. С 1972 по 1976 год он возглавлял первую в Вооруженных силах СССР кафедру математического обеспечения ЭВМ (ныне — 27-я кафедра), а с 1986 года продолжал трудиться профессором этой кафедры. Основное направление научной деятельно-

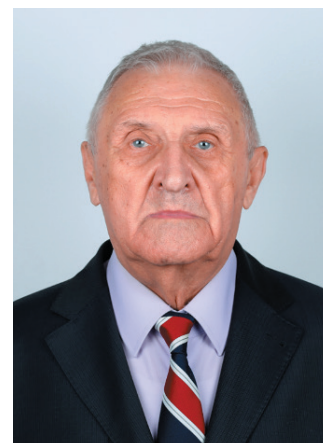
сти Юрия Ивановича было связано с разработкой численных методов теории очередей, управлением запасами и имитационным моделированием сложных систем. Он является

автором первой отечественной монографии по теории управления запасами (издательство «Наука», 1969), переведенной и изданной в Германии, создателем пакета программ «МОСТ» (массовое обслуживание — стационарные задачи), принятого Государственным фондом алгоритмов и программ и используемого более чем в 30 организациях страны. Им получены обобщенная формула Уилсона и алгоритмы для многопродуктовых задач, объединяющие методы управления запасами и теории очередей.

Юрий Иванович — автор более 300 научных и учебно-методических трудов, в том числе около 60 монографий, учебников и учебных пособий. Под его руководством и при его научном консультировании подготовлены два доктора и более 20 кандидатов наук. Многие годы он вел активную научно-педагогическую работу, воспитав научные кадры для оборонной промышленности и космической отрасли.

Заслуги профессора Рыжикова отмечены высокими наградами: 19 медалями и другими ведомственными знаками отличия. Светлая память об Юрии Ивановиче Рыжикове — талантливом ученом, педагоге и офицере — навсегда сохранится в сердцах его коллег, учеников и всех, кто имел честь работать с ним.

*Сотрудники ФГБОУ ВО ПГУПС,
ВКА имени А.Ф. Можайского, СПб ФИЦ РАН,
коллектив редакции журнала
«Интеллектуальные технологии на транспорте»*



RYZHNIKOV Yuri Ivanovich

10.09.1933 — 29.04.2026

The staff of the departments of Mathematical and Software Support, Information and Computing Systems of the Mozhaisky Military Space Academy, together with the staff of the St. Petersburg Federal Research Center of the Russian Academy of Sciences, with deep sorrow announce that on April 29, 2026, at the age of 93, after a prolonged illness, passed away our colleague, an outstanding Soviet and Russian scientist, Honored Scientist of the Russian Federation, Professor Emeritus of the Academy, Doctor of Technical Sciences, Professor, retired Colonel Yuri Ivanovich Ryzhnikov (10.10.1933 — 29.04.2026).

Yu. I. Ryzhnikov was born on October 10, 1933 in Leningrad. In 1952, he entered the Higher Naval School of Weapons Engineers as a Komsomol recruit; after the fourth year, together with his faculty, he was transferred to the Black Sea Higher Naval School named after P. S. Nakhimov, from which he graduated with honors in 1958.

His entire subsequent service took place in Leningrad: junior engineer of the 301st group of assembly brigades of the 6th Directorate of the Navy — until November 1958; junior researcher at the 16th Naval Institute (later the Naval Branch of the Central Research Institute No. 12 of the Ministry of Defense) — until September 1962; and then adjunct at the Leningrad Red Banner Air Force Engineering Academy named after A. F. Mozhaisky.

Progressing from lecturer to department head, Professor Ryzhnikov made an invaluable contribution to the development of national computer technology, inventory control theory, and queueing theory. From 1972 to 1976, he headed the first department of mathematical software for computers in the Armed Forces of the USSR (now Department 27), and from 1986 onwards he continued to work as a professor at this department.

The main direction of Yuri Ivanovich's scientific activity was related to numerical methods of queueing theory, inventory management, and simulation modeling of complex systems. He is the author of the first Russian monograph on inventory control theory (Nauka Publishing House, 1969), which was translated and published in Germany, and the creator of the MOST software package (stationary queueing problems), which was adopted by the State Fund of Algorithms and Programs and operated in more than 30 organizations across the country. He obtained a generalized Wilson formula and algorithms for multi-product problems, combining the methods of inventory management and queueing theory.

Yuri Ivanovich is the author of more than 300 scientific and methodological publications, including about 60 monographs, textbooks, and teaching aids. Under his leadership and with his scientific advice, two doctors and more than 20 candidates of sciences have been trained. For many years, he conducted active scientific and pedagogical work, training scientific personnel for the defense industry and space sector. The merits of Professor Ryzhnikov have been recognized with high awards: 19 medals and other departmental distinctions.

The blessed memory of Yuri Ivanovich Ryzhnikov — a talented scientist, educator and officer — will forever remain in the hearts of his colleagues, students, and all who had the honor of working with him.

Staff of the Emperor Alexander I St. Petersburg State Transport University (PGUPS), the Mozhaisky Military Space Academy, the St. Petersburg Federal Research Center of the Russian Academy of Sciences, and the editorial board of the journal "Intellectual Technologies on Transport"