

## ОГЛАВЛЕНИЕ

Д. И. Багаутдинов, Р. Салман, В. А. Алексеев, Р. М. Усмонов РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ ЭМОЦИОНАЛЬНОГО ТОНА СООБЩЕНИЯ С ИСПОЛЬЗОВАНИЕМ РАЗЛИЧНЫХ АРХИТЕКТУР НЕЙРОННЫХ СЕТЕЙ	396–413
А. И. Гурьянов, А. Ш. Якупов ИНСТРУМЕНТ ПОСЛЕДОВАТЕЛЬНОГО СНЯТИЯ СНИМКОВ АГРЕГИРОВАННЫХ ДАННЫХ ИЗ ПОТОКОВЫХ ДАННЫХ	414–436
А. Р. Давлетгареева, К. А. Едкова НЕЙРОННАЯ СЕТЬ ДЛЯ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ТЕКСТА ПЕСЕН С ПРИМЕНЕНИЕМ МОДЕЛЕЙ OPENAI И CLIP	437–455
И. А. Даминов, А. Ю. Арсенюк, А. С. Тощев КОМБИНИРОВАНИЕ АЛГОРИТМОВ SFM И ORB ПРИ 3D-РЕКОНСТРУКЦИИ	456–465
М. В. Исангулов, А. Р. Хамеджанов, Р. Р. Миннеахметов, Т. Р. Хафизьянов, Э. Ф. Пашаев, Э. Р. Калимуллин СИСТЕМА КОНТРОЛИРУЕМОЙ ГЕНЕРАЦИИ ЛИЦА, ПОСТРОЕННАЯ С ИСПОЛЬЗОВАНИЕМ СЕТИ STYLEGAN2	466–482
Р. А. Климов, А. Ш. Якупов РАЗРАБОТКА СИСТЕМЫ ПОИСКА И ИНДЕКСИРОВАНИЯ КОНТЕНТА АУДИОЗАПИСЕЙ	483–497
Р. В. Мосолов О РАЗРАБОТКЕ NOSQL СУБД GOLDENRACEDB КАК АЛЬТЕРНАТИВЫ GOOGLE FIREBASE	498–517
В. В. Наумова, С. В. Черкасов, В. С. Ерёменко, А. А. Загумёнов ЕСТЕСТВЕННОНАУЧНЫЕ МУЗЕИ В ЦИФРОВОМ ПРОСТРАНСТВЕ ГЕОЛОГИЧЕСКИХ ЗНАНИЙ	518–537

<b>О. А. Невзорова, Е. К. Липачёв, К. С. Николаев</b> <b>СЕМАНТИЧЕСКИЕ СЕРВИСЫ ЦИФРОВОЙ ЭКОСИСТЕМЫ ОНТОМАТН</b> <b>ДЛЯ МАТЕМАТИЧЕСКОГО ОБРАЗОВАНИЯ</b>	<b>538–569</b>
<b>Т. Р. Файзрахманов</b> <b>«ВЫЧИСЛИТЕЛЬНОЕ МЫШЛЕНИЕ» КАК ВТОРАЯ ГРАМОТНОСТЬ В</b> <b>СОВРЕМЕННОМ ЦИФРОВОМ ОБРАЗОВАНИИ И ОБЩЕСТВЕ</b>	<b>570–587</b>

## РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ ЭМОЦИОНАЛЬНОГО ТОНА СООБЩЕНИЯ С ОПРЕДЕЛЕНИЕМ НАИБОЛЕЕ ПОДХОДЯЩЕЙ АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ

Д. И. Багаутдинов<sup>1</sup> [0000-0003-4673-3253], Р. Салман<sup>2</sup> [0000-0002-1552-4910],  
В. А. Алексеев<sup>3</sup> [0000-0002-2208-9515], Р. М. Усмонов<sup>4</sup> [0009-0002-1289-5465]

<sup>1-4</sup> *Институт информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, ул. Кремлевская, 35, г. Казань, 420008*

<sup>1</sup>d4niskabag@yandex.ru, <sup>2</sup>Riham.salman.98@gmail.com, <sup>3</sup>aleksvlad99@gmail.com, <sup>4</sup>rustam190401@mail.ru

### **Аннотация**

Для определения наиболее эффективного подхода к решению задачи классификации эмоционального тона сообщения проведено обучение выбранных моделей нейронной сети на различных наборах обучающих данных. На основе такого показателя, как процентное соотношение правильно данных ответов на тестовом наборе данных, сравнены комбинации наборов обучающих данных и различных моделей, обученных на основе этих данных. Произведено обучение четырех моделей нейронной сети на трех различных наборах обучающих данных. В результате сравнения точности ответов каждой модели, обученной на разных обучающих данных, сделаны выводы о выборе модели нейронной сети, наиболее подходящей для решения поставленной задачи.

**Ключевые слова:** *NLP, sentiment detection, neural networks, comparison of neural network models, LSTM, CNN, BiLSTM.*

### **ВВЕДЕНИЕ**

Распознавание эмоционального тона сообщений [1] является важной задачей в области обработки естественного языка (NLP), поскольку может предоставить ценную информацию о значении и намерении, скрывающимися за тексто-

выми сообщениями. В последние годы использование нейронных сетей для решения этой задачи стало все более популярным благодаря их способности к обучению и обобщению данных [2, 3].

В настоящем исследовании мы сравнили различные модели нейронных сетей и наборы данных с целью их использования для решения задачи распознавания эмоционального тона в сообщениях. Мы исследовали эффективность различных архитектур нейронных сетей, таких как сверточные нейронные сети (CNN), нейронные сети с долгой краткосрочной памятью (LSTM), двунаправленная модель нейросети с долгой краткосрочной памятью (BiLSTM), на разных наборах данных с различными размерами и количеством классов эмоционального тона входных сообщений. В рамках данной работы проведено распознавание пяти базовых классов эмоционального тона сообщений: радость, грусть, нейтральная эмоция, злость, любовь. Для этого мы нашли и выбрали несколько наборов данных сообщений, помеченных соответствующим классом эмоционального тона, от маломасштабных наборов данных с несколькими тысячами примеров до больших наборов данных с 1.6 миллиона примеров. Затем, после обработки обучающих наборов данных, мы обучили несколько моделей нейронных сетей на каждом наборе. Для сравнения качества работы этих моделей мы подготовили набор тестовых сообщений, каждое из которых подавалось на вход обученным нейронным сетям [4, 5]. В конечном итоге комбинация нейронной сети и обучающего набора данных, у которой процентное соотношение верных ответов по тестовому набору сообщений стало наивысшим, признана наиболее эффективной для решения задачи классификации эмоционального тона сообщений.

### **БОТ ДЛЯ АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА**

Итоговой задачей нашего исследования является анализ влияния архитектуры нейронной сети и использованного для ее обучения датасета на примере разработки telegram-бота для предсказания эмоциональной тональности текста. Акцент сделан именно на рассмотрении использованных архитектур нейронной сети.

## Использованный инструментарий

Основным использованным языком программирования являлся Python 3-ей версии, в процессе работы были использованы следующие библиотеки.

**Keras (TensorFlow)** [6] — открытая программная библиотека для машинного обучения, разработанная компанией Google для решения задач построения и тренировки нейронной сети с целью автоматического нахождения и классификации образов с достижением качества человеческого восприятия. Эта библиотека написана на языке Python и обеспечивает взаимодействие с искусственными нейронными сетями и в данный момент является частью TensorFlow [7].

**NumPy** [8] — библиотека для работы с математическими функциями, многомерными массивами и матрицами. В частности, она была использована нами для генерации векторов.

**Pandas** [9] — библиотека для обработки и анализа данных, ее работа с данными строится поверх библиотеки NumPy, являющейся инструментом более низкого уровня.

**Aiogram** [10] — это современный и полностью асинхронный фреймворк для Telegram Bot API, написанный на Python с использованием `asyncio` [11] и `aiohttp`[12].

Также в работе была применена предобученная модель GloVe [13] для получения векторных представлений слов размерности 200. GloVe – это алгоритм обучения без учителя для получения векторных представлений слов. Часть вычислений производилась с использованием платформы **Google Colab** [14] – бесплатном облачном сервисе на основе Jupyter Notebook [15]. Благодаря использованию GPU этого сервиса, время обучения моделей было сокращено в разы [13].

## Данные для обучения

Обучение моделей производилось на трех различных наборах данных. Первый набор состоял только лишь из предложений от первого лица («Я люблю пиццу» и т. п.) [16], что было достаточно серьезным ограничением, и модель плохо распознавала предложения иной структуры, предложения были помечены 6-ю классами. Второй набор данных имел ограничение в количестве классов, которыми он был

размечен, предложения были более разнообразными по своей структуре, но датасет размечен только на два класса – «позитивный» и «негативный» [17]. Последний датасет собран из данных твитов социальной сети Twitter, размеченных 5-ю классами эмоций, поэтому в конечном итоге выбор был сделан в пользу этого обучающего набора данных [18].

### **Обзор архитектур нейронных сетей, использованных для анализа**

В процессе разработки нашего бота, решающего задачу анализа тональности текста, было опробовано несколько архитектур нейронных сетей с целью улучшить результаты анализа тональности текста, а именно, сверточная нейронная сеть (CNN), нейронная сеть с долгой краткосрочной памятью (LSTM), двунаправленная модель нейросети с долгой краткосрочной памятью (BiLSTM).

**CNN.** Изначально была реализована простая архитектура сверточной нейронной сети, или CNN. Состоит данная модель из следующих слоев:

1) Первый слой модели – ‘Embedding’, предназначенный для представления входных данных в виде числового вектора. Этот слой принимает три аргумента: ‘num\_words’, который представляет размер словаря, ‘32’, который представляет размерность числового вектора, и ‘input\_length’, который представляет длину каждой входной последовательности. Этот слой создает обучающую таблицу поиска, которая преобразует каждое слово во входной последовательности в вектор размера 32.

2) Второй слой – это сверточный слой Conv1D, который выполняет одномерную свертку переданной входной последовательности. Этот слой применяет 250 фильтров размером 5 каждый и использует функцию активации ‘relu’. Этот слой помогает идентифицировать значимые признаки во входной последовательности на основе числового вектора, полученного из предыдущего слоя.

3) Третий слой – это слой максимальной подвыборки (GlobalMaxPooling1D Layer), который выполняет глобальную операцию максимальной подвыборки над выходом предыдущего слоя. Этот слой выбирает максимальное значение из каждого набора признаков, созданного слоем Conv1D, и выводит вектор фиксированного размера.

4) Четвертый слой – это полносвязный слой (Dense Layer) с 128 нейронами и функцией активации 'relu'. Этот слой принимает выход предыдущего слоя в качестве входа и применяет к нему линейное преобразование. Выходом этого слоя является вектор размера 128.

5) Последний слой – это еще один полносвязный слой Dense с размерностью выходного пространства 5 и функцией активации 'softmax'. Этот слой принимает выход предыдущего слоя в качестве входа и производит распределение вероятностей по 5 возможным классам эмоционального тона входных сообщений.

Архитектура данной модели представлена на Рис. 1.

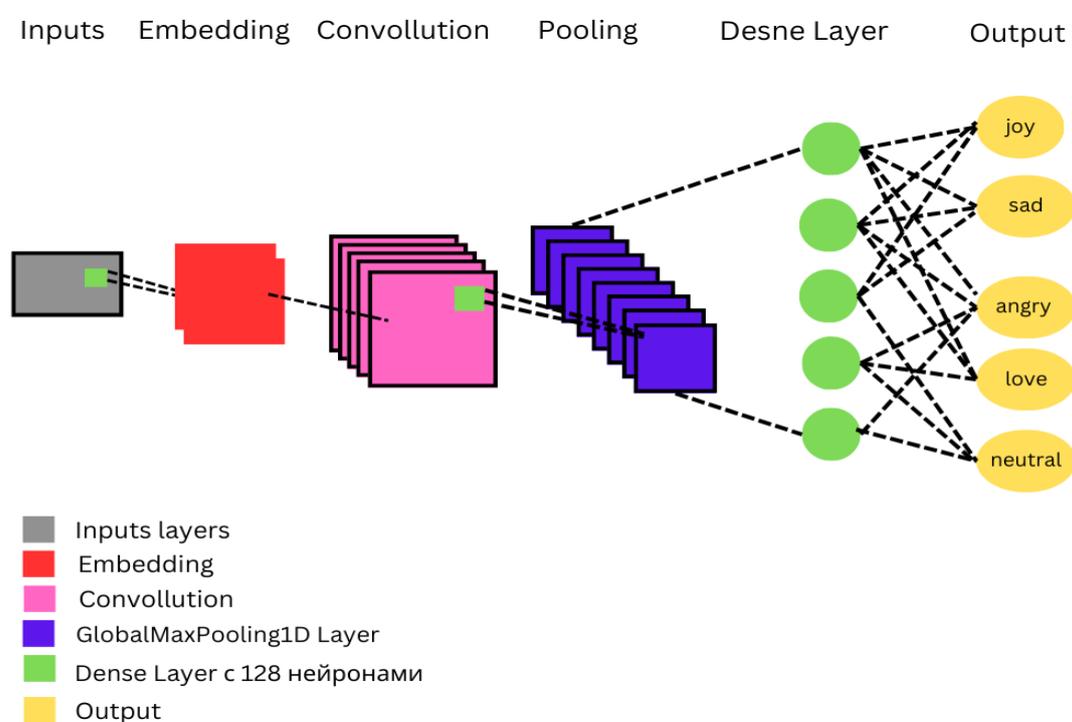


Рис. 1. Архитектура CNN модели

**LSTM.** Следующей архитектурой нейронной сети для реализации поставленной задачи была выбрана архитектура нейронной сети с долгой краткосрочной памятью (LSTM). Состоит данная модель из следующих нейронных слоев (Рис. 2).

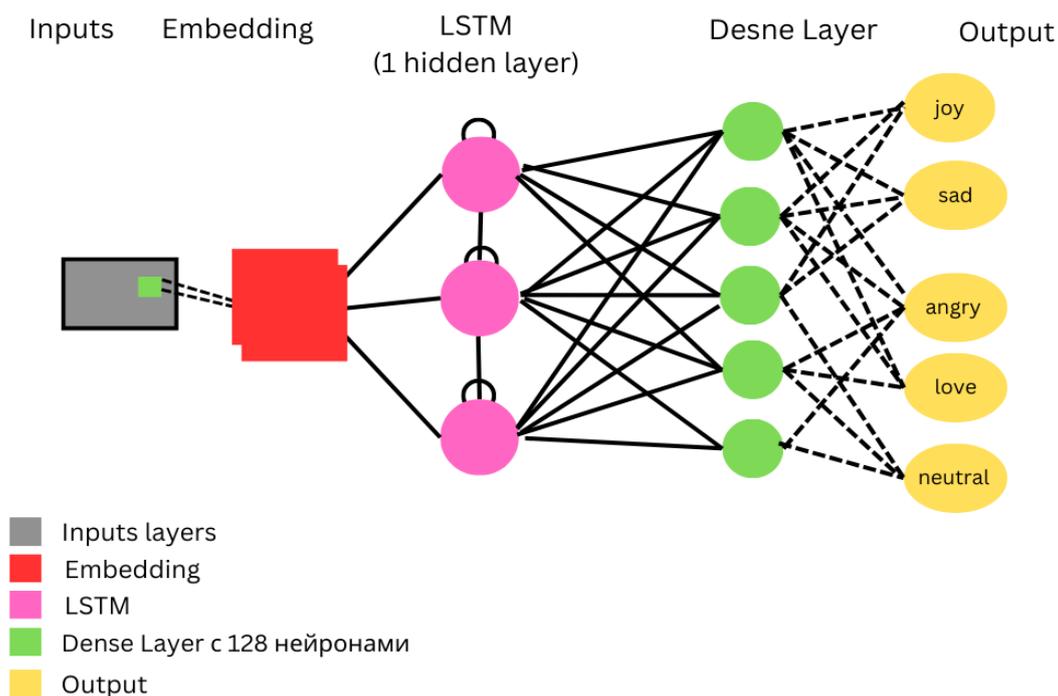


Рис. 2. Архитектура модели LSTM

- 1) Первый слой модели, так же, как у предыдущей, – ‘Embedding’, предназначенный для представления входных данных в виде числового вектора;
- 2) Второй слой – слой LSTM. Это рекуррентный слой нейронной сети, который обрабатывает последовательность числовых векторов, выводимых слоем Embedding. Уровень LSTM имеет 128 нейронов и коэффициент ‘dropout’, равный 0,2 для предотвращения переобучения. Аргумент ‘recurrent\_dropout’ задает частоту выпадения повторяющихся соединений между блоками памяти.
- 3) Третий слой – полносвязный слой (Dense Layer). Этот уровень состоит из 128 нейронов и использует функцию активации ‘relu’. Это полносвязный слой, который принимает выходные данные из слоя LSTM в качестве входных данных и преобразует их в новое представление. Выходом этого слоя является вектор размера 128.
- 4) Последний слой – это еще один полносвязный слой Dense с размерностью выходного пространства 5 и функцией активации ‘softmax’. Этот слой принимает выход предыдущего слоя в качестве входа и производит распределение вероятностей по 5 возможным классам эмоционального тона входных сообще-

ний. В примере, представленном на Рис. 2, указаны именно 5 классов эмоционального тона, так как это значение разнится в зависимости от используемого набора обучающих данных; в данном примере показана модель для обучения на данных с 5-ю классами эмоционального тона сообщений.

**BiLSTM.** В качестве еще одной архитектуры модели нейронной сети была выбрана модель BiLSTM [19] (Двунаправленная LSTM) (см. Рис. 3). Она принимает на вход последовательность данных и последовательно обрабатывает ее слева направо и справа налево, чтобы отслеживать зависимости в обоих направлениях в контексте задачи.

1) Первым слоем модели аналогично предыдущим используемым архитектурам является слой 'Embedding' векторных представлений слов;

2) Далее создаются два LSTM «канала», которые складываются после прохождения через сверточные и рекуррентные слои. Как сказано выше, это рекуррентные слои нейронной сети, которые обрабатывают последовательность числовых векторов, выводимых слоем Embedding (только в данной случае – с двух сторон; имеет 128 нейронов и коэффициент 'dropout', равный 0,2 для предотвращения переобучения). Имеет по 6 нейронов и коэффициент 'dropout', равный 0,3, для предотвращения переобучения;

3) Далее входная последовательность в двух каналах обрабатывается в сверточных слоях с различными размерами ядер (2, 3, 4, 5, 6 и 8), чтобы обнаружить в тексте различные паттерны. Путем прохождения через Conv1D и Dropout (защита от переобучения) слои с различными значениями параметра kernel (как было указано выше, 2, 3, 4, 5, 6 и 8);

4) Далее сверточные слои объединяются в LSTM-слои, и выполняется конкатенация выходов LSTM-слоя и сверточно-рекуррентного слоя, а затем производится понижение размерности через слои MaxPooling1D и Flatten;

5) Наконец, на выходном слое сети используется полносвязный Dense-слой с размерностью выхода 5 (по количеству классифицируемых классов), использующий функцию активации 'softmax', который преобразует входы в вероятности для каждого класса, где модель должна выбрать один из пяти классов предсказания.

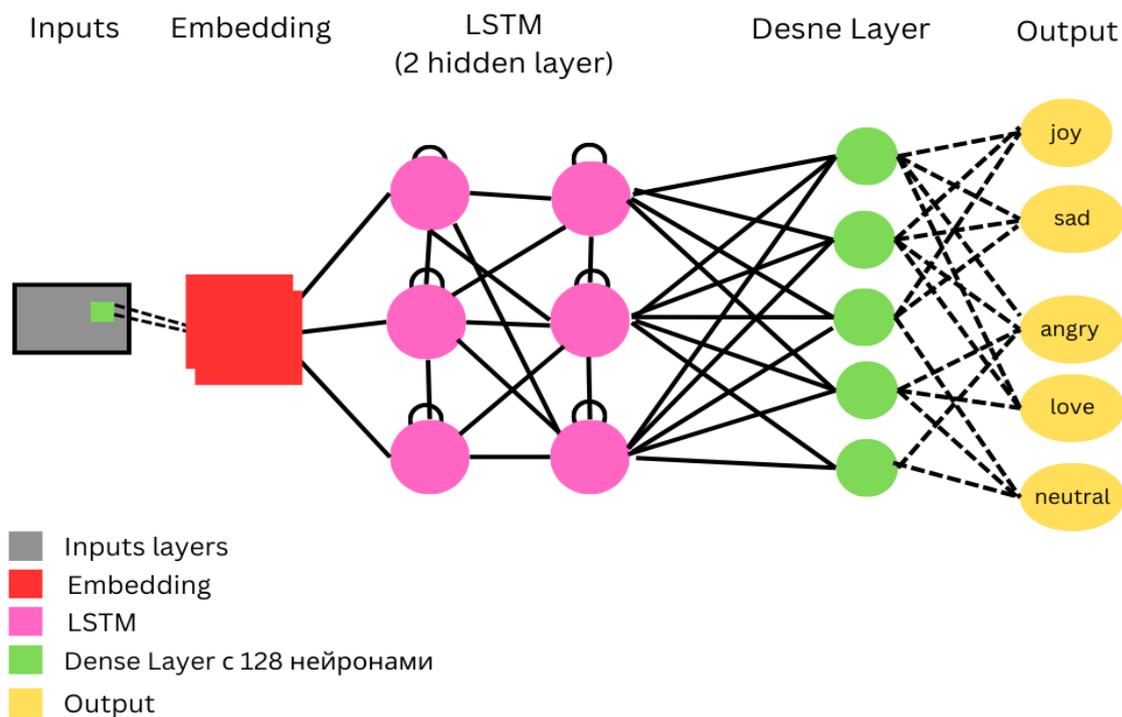


Рис. 3. Архитектура модели BiLSTM

### Результаты, полученные в рамках исследования

Получены следующие результаты, отраженные в Таблице 1. В качестве метрик качества для сравнения сетей выбрана validation accuracy (доля правильных ответов алгоритма). Зеленым цветом в таблице 1 выделено самое большое значение точности, полученное в модели Bi-LSTM.

Таблица 1. Сравнительная таблица accuracy для обученных моделей.

Модель	accuracy	validation accuracy
CNN	0.9907	0.6002
LSTM	0.9109	0.6060
Bi-LSTM	0.6242	0.6149

Далее для сравнения каждой из использованных топологий нейронных сетей на Рисунках 4 и 5 показаны графики обучения для моделей CNN, обученных на 1-ом и 3-ем датасетах, на 5-й и 80-й эпохах соответственно.

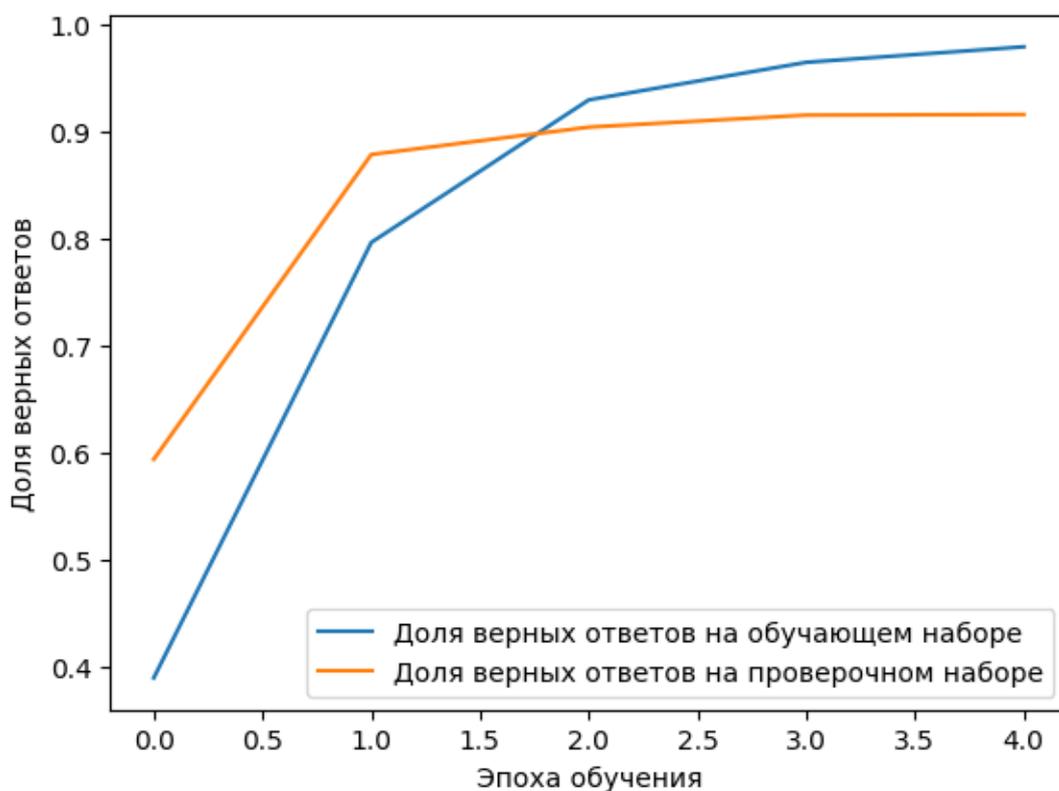


Рис. 4. График обучения модели CNN с 6-ю классами эмоций

На Рис. 5 приведен явный пример «переобучения» модели, так как прогресс в обучении останавливается примерно на 10-ой эпохе, и дальнейшее обучение было бессмысленно. На Рис. 6 показан график обучения на 20-ти эпохах – видим, что ассигасу растет достаточно быстро, но при этом валидационное ассигасу падает, что опять же может говорить о переобучении.

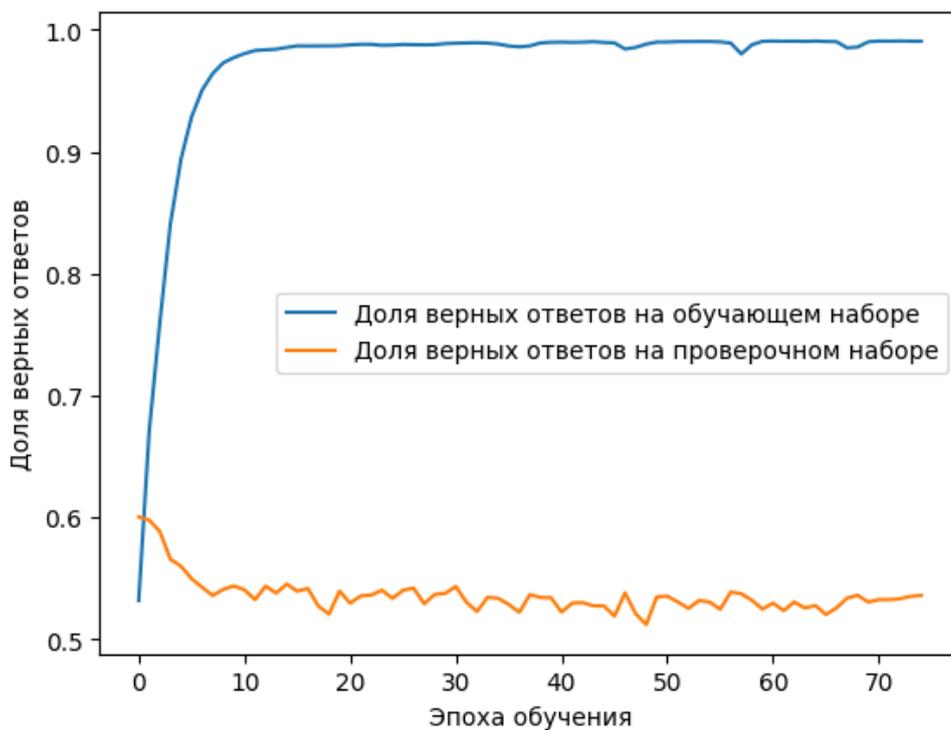


Рис. 5. График обучения модели CNN с-ю 5 классами эмоций

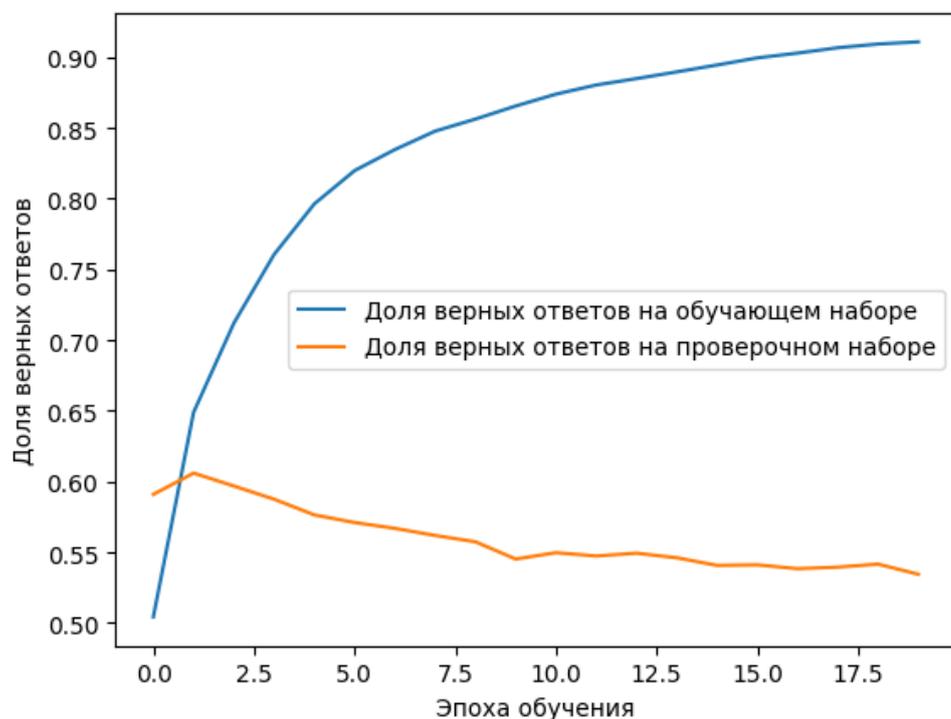


Рис. 6. График обучения модели LSTM с-ю 5 классами эмоций

Самой удачной в динамике обучения оказалась модель Bi-LSTM. На Рис. 7 можно проследить изменение значений метрики точности, которая плавно возрастает и на обучающей, и на валидационной выборках. Как и ожидалось, функция потерь при этом снижается и для валидационной, и для обучающей выборок (Рис. 8).

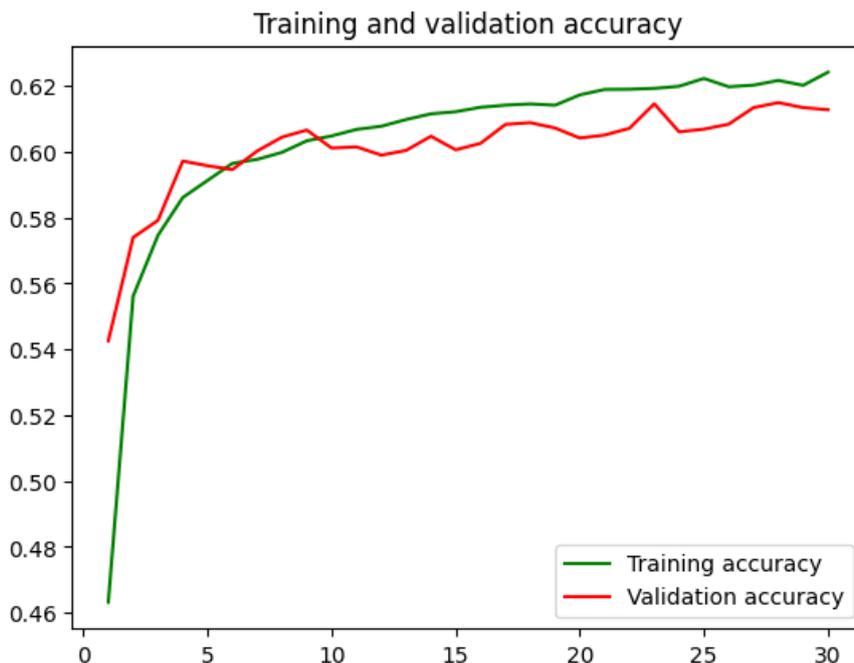


Рис. 7. График обучения модели BiLSTM с 5-ю классами эмоций

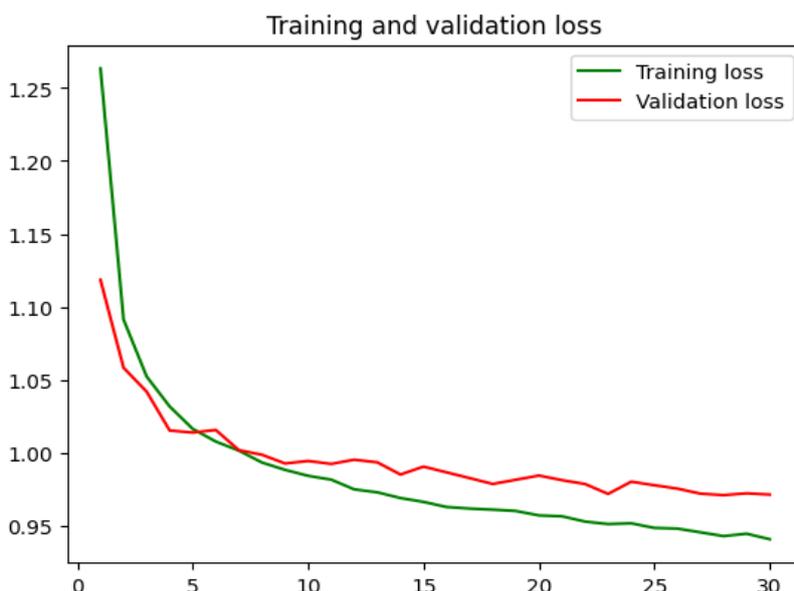


Рис. 8. График функции потерь модели BiLSTM с 5-ю классами эмоций

## Реализация telegram-бота

В результате проведенной работы был создан telegram-бот (Рис. 9), который может распознавать основные эмоции по введенному сообщению. Бот был создан с помощью фреймворка aiogram. Бот работает на английском языке, т. к. модели обучались на датасетах на английском языке.



Рис. 9. а) Пример предсказаний моделью на коротких сообщениях моделью BiLSTM; б) Пример предсказаний моделью на коротких сообщениях моделью с CNN

Эксперименты с ботом проводились с использованием трех моделей нейронных сетей (CNN, LSTM и Vi-LSTM). После этого мы сравнили результаты, которые выдает бот: в следующей таблице показаны различия в результатах для каждой нейросети. В первой модели с CNN бот распознавал 6 эмоций (грусть, сча-

стью, радость, удивление, страх и любовь) (Рис. 9а), а вторая модель с использованием Bi-LSTM распознавала только 5 эмоций (грусть, счастье, радость, нейтральный и любовь) (Рис. 9б), но с большей точностью предсказаний. Также с использованием модели CNN и датасета с 2-мя классами эмоций была получена модель, которая достаточно точно определяла эмоциональный тон сообщения, но использовала лишь 2 класса эмоционального тона сообщений – positive и negative, чего недостаточно для полноценной реализации поставленной задачи

Таблица 2. Пример классификации предложений моделью CNN, LSTM и BiLSTM

	BiLSTM	LSTM	CNN
I feel bored	Sad 😞	Sad	anger
I am sad today	Sad 😞	Sad	sadness
I don't know what I should do	Neutral 😐	Neutral	fear
I have an exam	Neutral 😐	Sad	fear
I got an excellent deal	Happy 😊	Happy	fear
I like summer	Neutral 😐	Neutral	anger
I like food	Sad 😞	Sad	anger
I am happy	Happy 😊	Happy	joy
I like funny videos	Happy 😊	Happy	surprise
I forgot about the exam	Neutral 😐	Neutral	fear
I have a good friend	Neutral 😐	Neutral	joy
This professor has good ideas	Happy 😊	Happy	joy

## ЗАКЛЮЧЕНИЕ

Результатом работы стал Telegram-бот, предсказывающий эмоциональный тон сообщения, введенного пользователем. Так как в результате анализа различных моделей наибольшее валидационное ассурасу было получено при использовании архитектуры Bi-LSTM, к боту в итоге была подключена именно данная модель, обученная на 3-ем наборе данных. Как можно увидеть на Рис. 9, модель с данной архитектурой лучше предсказывает короткие фразы, поданные на вход. Наши результаты показывают, что выбор модели нейронной сети и набора данных может оказать значительное влияние на производительность нейронной сети для решения задачи распознавания эмоционального тона.

## СПИСОК ЛИТЕРАТУРЫ

1. Gosai D. D., Gohil H. J., Jayswal H. S. A Review on a Emotion Detection and Recognition from Text Using Natural Language Processing. URL: [http://www.ripublication.com/ijaer18/ijaerv13n9\\_33.pdf](http://www.ripublication.com/ijaer18/ijaerv13n9_33.pdf)
2. Acheampong F. A., Nunoo-Mensah H., Wenyu Chen. Transformer models for text-based emotion detection: a review of BERT-based approaches. URL: <https://link.springer.com/article/10.1007/s10462-021-09958-2>
3. Naresh S. S., Khethawat S. Emotion Detection from Text URL: <https://arxiv.org/abs/1205.4944>
4. Baccianella S., Esuli A., Sebastiani F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining // Lrec. 2010. P. 2200–2204
5. Akbik A., Blythe D., Vollgraf R. Contextual string embeddings for sequence labeling // Proceedings of the 27th international conference on computational linguistics. 2010. P. 1638–1649.
6. Keras. URL: <https://keras.io/>
7. Tensorflow. URL: <https://www.tensorflow.org/>
8. Numpy. URL: <https://numpy.org/>
9. Pandas. URL: <https://pandas.pydata.org/>
10. Aiogram. URL: <https://aiogram.dev>
11. Asyncio. URL: <https://docs.python.org/3/library/asyncio.html>
12. Aiohttp. URL: <https://docs.aiohttp.org/en/stable/index.html>

13. GloVe: Global Vectors for Word Representation.

URL: <https://nlp.stanford.edu/projects/glove/>

14. Google Collab. URL: <https://colab.research.google.com>

15. Jupyter Notebook.

URL: <https://docs.jupyter.org/en/latest/install/notebook-classic.html>

16. Multi-Class Text Emotion Analysis.

URL: <https://github.com/kanchitank/Text-Emotion-Analysis>

17. Sentiment140 dataset with 1.6 million tweets.

URL: <https://www.kaggle.com/datasets/kazanov/sentiment140>

18. Twitter Emotion Classification.

URL: <https://www.kaggle.com/code/shtrausslearning/twitter-emotion-classification?scriptVersionId=126983369>

19. BiLSTM.

URL: <https://www.kaggle.com/code/mlwhiz/bilstm-pytorch-and-keras>

---

## SOLVING THE PROBLEM OF CLASSIFYING THE EMOTIONAL TONE OF A MESSAGE WITH DETERMINING THE MOST APPROPRIATE NEURAL NETWORK ARCHITECTURE

D. I. Bagautdinov<sup>1</sup> [0000-0003-4673-3253], R. Salman<sup>2</sup> [0000-0002-1552-4910],

V. A. Alekseev<sup>3</sup> [0000-0002-2208-9515], R. M. Usmonov<sup>4</sup> [0009-0002-1289-5465]

<sup>1-4</sup>*Institute of Information Technologies and Intelligent Systems, Kazan (Volga region) Federal University, ul. Kremlyovskaya, 35, Kazan, 420008*

<sup>1</sup>d4niskabag@yandex.ru, <sup>2</sup>Riham.salman.98@gmail.com, <sup>3</sup>aleksvlad99@gmail.com,

<sup>4</sup>rustam190401@mail.ru

### Abstract

To determine the most effective approach for solving the task of classifying the emotional tone of a message, we trained selected neural network models on various sets of training data. Next, based on the performance metric of the percentage of correctly classified responses on a test data set, we compared combinations of training

---

data sets and various models trained on them. During the writing of this article, we trained four neural network models on three different sets of training data. By comparing the accuracy of the responses from each model trained on different training data sets, conclusions were drawn regarding the neural network model best suited for solving the task at hand.

**Keywords:** *NLP, sentiment detection, neural networks, comparison of neural network models, LSTM, CNN, BiLSTM.*

## REFERENCES

1. Gosai D. D., Gohil H. J., Jayswal H. S. A Review on a Emotion Detection and Recognition from Text Using Natural Language Processing. URL: [http://www.ripublication.com/ijaer18/ijaerv13n9\\_33.pdf](http://www.ripublication.com/ijaer18/ijaerv13n9_33.pdf)
2. Acheampong F. A., Nunoo-Mensah H., Wenyu Chen. Transformer models for text-based emotion detection: a review of BERT-based approaches. URL: <https://link.springer.com/article/10.1007/s10462-021-09958-2>
3. Naresh S. S., Khethawat S. *Emotion Detection from Text* URL: <https://arxiv.org/abs/1205.4944>
4. Baccianella S., Esuli A., Sebastiani F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining // *Lrec. 2010. P. 2200–2204*
5. Akbik A., Blythe D., Vollgraf R. Contextual string embeddings for sequence labeling // *Proceedings of the 27th international conference on computational linguistics. 2010. P. 1638–1649.*
6. Keras. URL: <https://keras.io/>
7. Tensorflow. URL: <https://www.tensorflow.org/>
8. Numpy. URL: <https://numpy.org/>
9. Pandas. URL: <https://pandas.pydata.org/>
10. Aiogram. URL: <https://aiogram.dev>
11. Asyncio. URL: <https://docs.python.org/3/library/asyncio.html>
12. Aiohttp. URL: <https://docs.aiohttp.org/en/stable/index.html>
13. GloVe: Global Vectors for Word Representation.  
URL: <https://nlp.stanford.edu/projects/glove/>
14. Google Collab. URL: <https://colab.research.google.com>

15. Jupyter Notebook.

URL: <https://docs.jupyter.org/en/latest/install/notebook-classic.html>

16. Multi-Class Text Emotion Analysis.

URL: <https://github.com/kanchitank/Text-Emotion-Analysis>

17. Sentiment140 dataset with 1.6 million tweets.

URL: <https://www.kaggle.com/datasets/kazanova/sentiment140>

18. Twitter Emotion Classification.

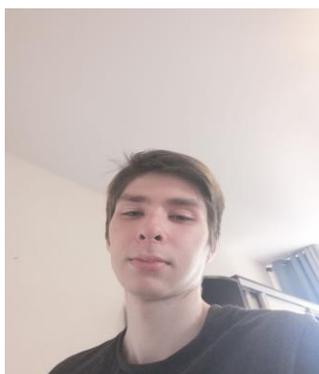
URL: <https://www.kaggle.com/code/shtrausslearning/twitter-emotion-classification?scriptVersionId=126983369>

19. BiLSTM.

URL: <https://www.kaggle.com/code/mlwhiz/bilstm-pytorch-and-keras>

---

## СВЕДЕНИЯ ОБ АВТОРАХ



**БАГАУТДИНОВ Данис Ильмасович** – студент магистратуры Института ИТИС КФУ. Сфера научных интересов – обработка естественных языков, машинное обучение, нейронные сети, алгоритмы и структуры данных.

**Danis Ilmasovich BAGAUTDINOV** – Master's student at the Institute of ITIS KFU. Research interests are natural language processing, machine learning, neural networks, algorithms and data structures.

email: [d4niskabag@yandex.ru](mailto:d4niskabag@yandex.ru)

ORCID 0000-0003-4673-3253



**САЛМАН Рухам** – студентка магистратуры Института ИТИС КФУ. Сфера научных интересов – обработка естественных языков, машинное обучение, нейронные сети.

**Riham SALMAN** – Master's student at the Institute of ITIS KFU. Research interests – natural language processing, machine learning, neural networks.

email: [Riham.salman.98@gmail.com](mailto:Riham.salman.98@gmail.com)

ORCID 0000-0002-1552-4910

---



**АЛЕКСЕЕВ Владислав Алексеевич** – студент магистратуры Института ИТИС КФУ. Сфера научных интересов – обработка естественных языков, машинное обучение, нейронные сети, веб-разработка, разработка эффективных интерфейсов.

**Vladislav Alekseevich ALEKSEEV** – Master's student at the Institute of ITIS KFU. Research interests – natural language processing, machine learning, neural networks, web development, development of effective interfaces.

email: aleksvlad99@gmail.com

ORCID 0000-0002-2208-9515



**УСМОНОВ Рустамджон Муроджонович** – студент магистратуры Института ИТИС КФУ. Сфера научных интересов – обработка естественных языков, машинное обучение, нейронные сети, веб-разработка, разработка эффективных интерфейсов.

**Rustamdzhon Murodzhonovich Usmonov** – Master's student at the Institute of ITIS KFU. Research interests – natural language processing, machine learning, neural networks, web development, development of effective interfaces.

email: rustam190401@mail.ru

ORCID 0009-0002-1289-5465

*Материал поступил в редакцию 23 мая 2023 года*

## ИНСТРУМЕНТ ПОСЛЕДОВАТЕЛЬНОГО СНЯТИЯ СНИМКОВ АГРЕГИРОВАННЫХ ДАННЫХ ИЗ ПОТОКОВЫХ ДАННЫХ

А. И. Гурьянов<sup>1</sup> [0000-0002-9870-7973], А. Ш. Якупов<sup>2</sup> [0000-0002-2333-8819]

<sup>1</sup>Казанский филиал Межведомственного суперкомпьютерного центра Российской академии наук, ул. Лобачевского, 2, г. Казань, 420008;

<sup>1,2</sup>Казанский (Приволжский) федеральный университет, ул. Кремлевская, 35, г. Казань, 420008

<sup>1</sup>armgnv@gmail.com, <sup>2</sup>asyakupov@kpfu.ru

### **Аннотация**

В современном мире потоковые данные получили широкое распространение во многих предметных областях. Высокую актуальность имеет решение задачи обработки потоковых данных в реальном времени, с минимальной задержкой.

При потоковой обработке данных часто применяются различные приближенные алгоритмы, имеющие гораздо более высокую эффективность по времени и памяти, чем точные алгоритмы. Кроме того, часто возникает потребность прогнозирования состояния потока.

Таким образом, в настоящее время существует потребность в инструменте последовательного снятия снимков агрегированных данных из потоковых данных, дающем возможность прогнозирования состояния потока и применения приближенных алгоритмов обработки потоковых данных.

Авторами статьи разработан такой инструмент, рассмотрены архитектура и механизм его функционирования, а также оценены перспективы его дальнейшего развития.

**Ключевые слова:** потоковые данные, потоковая обработка данных, анализ потоковых данных, материализованные представления, потоковые алгоритмы, приближенные алгоритмы, прогнозирование потока.

## **ВВЕДЕНИЕ**

В современном мире информация является важнейшим элементом выживаемости и конкурентоспособности организаций [1]. Данные являются важным стратегическим ресурсом компании. При этом в современном мире в различных предметных областях получили широкое распространение потоковые данные [2]. Поэтому обработка и анализ потоковых данных имеют высокую актуальность [3].

Потоковые данные – данные, непрерывно генерируемые различными источниками [4]. На практике очень часто возникает потребность обработки потоковых данных с минимальной задержкой, в реальном времени [5]. Такая обработка данных называется потоковой.

Потоковая обработка данных широко распространена и имеет высокую актуальность в большом количестве предметных областей [6]. В частности, потоковая обработка данных является необходимым условием существования таких сфер, как интернет вещей и социальные сети [7]. Помимо этого, потоковая обработка данных имеет очень высокую важность в сфере информационной безопасности [8]. В потоковых данных журналов очень важно оперативно находить аномальную активность, которая может являться признаком атаки.

Во многих случаях обработка данных в реальном времени предоставляет значительное конкурентное преимущество [9]. Например, обработка в реальном времени данных о покупках товаров в сети магазинов дает возможность предоставлять пользователям актуальную информацию о том, в каких магазинах товар есть в наличии. Кроме того, применение потоковой обработки данных дает возможность повысить актуальность информации, используемой менеджерами организации для принятия решений [10], что позволяет оперативно реагировать на изменения как внешней, так и внутренней среды организации.

Потоковая обработка данных часто противопоставляется пакетной. При пакетной обработке потоковых данных они сохраняются в некоторое хранилище данных и далее обрабатываются большими пакетами.

Важным фактором востребованности потоковой обработки данных является происходящий в современном мире стремительный рост объема данных [11]. Во многих случаях суммарный объем потоковых данных огромен, и хранение

этих данных для последующего анализа традиционными методами пакетной обработки данных может быть нерентабельным. В то же время, при использовании методов потоковой обработки данных можно хранить лишь малую долю данных потока, на порядки меньшую, чем объем всех данных потока.

Кроме этого, чем больше объем данных, тем выше задержка до получения результата при использовании методов пакетной обработки данных. Эта задержка может достигать многих часов, что во многих случаях неприемлемо, так как полученный с такой задержкой результат в значительной мере теряет актуальность.

В базах данных для хранения результатов запросов применяются материализованные представления (Materialized View). Такие представления дают возможность эффективно использовать результаты сложных запросов без необходимости повторного выполнения запроса при каждом обращении. При этом в большинстве реализаций у материализованных представлений есть значительный недостаток – невозможность их инкрементального обновления на основе изменений в исходных данных. Как правило, единственный способ обновления материализованного представления – полное повторное выполнение запроса, что часто сопряжено со значительными задержками и требует значительных ресурсов, и поэтому несовместимо с потоковой обработкой данных. Такие материализованные представления обновляются или вручную, или в определенное время, или по триггеру, отслеживающему изменения в исходных данных [12].

В то же время, при потоковой обработке данных широко используются приближенные алгоритмы и вероятностные структуры данных. Они дают возможность решать соответствующие задачи со значительно меньшими затратами времени и памяти ценой некоторых потерь в точности. Необходимость их применения при потоковой обработке данных вызвана тем, что для многих широко распространенных задач поиск точного решения требует больших затрат времени и памяти [13]. Из-за этого в ряде случаев поиск точного решения несовместим с потоковой обработкой данных.

Также значительную актуальность имеет задача прогнозирования значений различных метрик потока данных [14]. Результаты прогнозирования, в частности,

могут использоваться организациями для принятия решений. Кроме того, прогнозирование состояния потока может быть применено для раннего обнаружения и предотвращения нештатных ситуаций в таких сферах, как транспорт и интернет вещей [15]. Для решения таких задач применяются экстраполяция, а также различные методы машинного обучения.

Исходя из вышеизложенного, можно заключить, что на рынке существует потребность в инструменте анализа потоковых данных, строящем инкрементально обновляемые материализованные представления этих данных, а также дающем возможность применения приближенных алгоритмов анализа потоковых данных и прогнозирования состояния потока [16, 17].

На основе вышеизложенного нами разработан инструмент, обладающий следующими свойствами:

- дает возможность построения на потоковых данных инкрементально обновляемых материализованных представлений, поддерживающих операции обновления и удаления;
- предоставляет возможность использования приближенных алгоритмов обработки и анализа потоковых данных;
- позволяет прогнозировать состояния потока;
- имеет открытый исходный код;
- является отечественным;
- выпускается под свободной лицензией MIT; возможно бесплатное использование инструмента, в том числе коммерческое, без каких-либо ограничений;
- является расширяемым, с возможностью добавления новых агрегатных функций и коннекторов к источникам данных.

Разработанный инструмент доступен в репозитории GitHub [18].

## СУЩЕСТВУЮЩИЕ ИНСТРУМЕНТЫ ДЛЯ СОЗДАНИЯ ИНКРЕМЕНТАЛЬНО ОБНОВЛЯЕМЫХ МАТЕРИАЛИЗОВАННЫХ ПРЕДСТАВЛЕНИЙ

В настоящее время существует ряд инструментов, частично решающих поставленную выше задачу. Имеются следующие инструменты, дающие возможность создания инкрементально обновляемых материализованных представлений потоковых данных:

- Materialize [19];
- ksqlDB [20];
- Amazon Redshift [21].

Результаты сравнительного анализа данных инструментов приведены в таблице 1.

Таблица 1. Сравнительный анализ инструментов для создания инкрементально обновляемых материализованных представлений

	<b>Materialize</b>	<b>ksqlDB</b>	<b>Amazon Redshift</b>
<b>Лицензия</b>	Business Source License 1.1	Confluent Community License Agreement 1.0	Проприетарная
<b>Открытый исходный код</b>	+	+	–
<b>Бесплатное использование</b>	Со значительными ограничениями	+	–
<b>Обновление и удаление строк</b>	+	–	–
<b>Приближенные алгоритмы</b>	–	–	–
<b>Прогнозирование состояния потока</b>	–	–	–

Источник: составлено авторами по данным [19–23].

Ни один из этих инструментов не поддерживает приближенные алгоритмы обработки потоковых данных и прогнозирование состояния потока.

Единственным инструментом, поддерживающим операции обновления и удаления пришедших ранее записей, является Materialize, позиционируемый как Data Warehouse для потоковых данных. Остальные инструменты рассматривают потоковые данные как append-only, что значительно ограничивает набор возможных сценариев их использования.

Materialize имеет открытый исходный код, однако выпускается под лицензией Business Source License 1.1. Данная лицензия значительно ограничивает бесплатное использование Materialize [23]. В частности, при бесплатном использовании Materialize:

- может работать только в одном процессе;
- не может использовать репликацию.

Эти ограничения делают невозможным бесплатное использование Materialize в большинстве практических сценариев.

Для использования Materialize без ограничений необходимо приобрести коммерческую лицензию. В текущей ситуации приобретение такой лицензии сопряжено со значительными рисками, так как существует ненулевая вероятность недобросовестных, политически мотивированных действий лицензиара. При этом для каждой версии Materialize лицензия через 4 года заменяется на Apache License 2.0, что приводит к снятию ограничений, названных выше. При этом первая версия Materialize была выпущена 14 февраля 2020 года, она перейдет под лицензию Apache License 2.0 только 14 февраля 2024 года.

Использование версии Materialize четырехлетней давности во многих случаях будет являться неприемлемым, во-первых, из-за значительного устаревания и, во-вторых, потенциально из-за наличия известных уязвимостей, которые могут быть использованы злоумышленниками для проведения атаки.

## **АРХИТЕКТУРА РАЗРАБОТАННОГО ИНСТРУМЕНТА**

Диаграмма классов разработанного инструмента представлена в приложении 1. Реализации абстрактных классов Source и Aggregate представлены в приложениях 2 и 3.

Инструмент состоит из трех частей, работающих в отдельных процессах:

1. Главный процесс;
2. Сервер;
3. Процессы источников.

Главный процесс начинает работу в момент запуска инструмента. Его основной задачей является управление жизненным циклом остальных частей инструмента. Главному процессу соответствует класс `MainProcess`.

Сервер запускается главным процессом вскоре после запуска инструмента и отвечает за взаимодействие клиентов с инструментом. Серверу соответствуют классы `ThreadingServer` и `Handler`

Процессы источников содержат основную функциональность инструмента. Они запускаются главным процессом при выполнении запроса на подключение инструмента к источнику данных (`CREATE SOURCE`). Каждому источнику данных соответствует отдельный процесс.

Такая архитектура дает потенциальную возможность реализации в будущем возможности развертывания инструмента на компьютерном кластере, что значительно увеличит масштабируемость инструмента.

На рисунке 1 отражены функционирование и взаимодействие частей инструмента при сценарии использования, включающем в себя последовательное выполнение следующих действий:

1. Запуск инструмента.
2. Создание источника (запрос `CREATE SOURCE`).
3. Создание материализованного представления (`CREATE MATERIALIZED VIEW`).
4. Запрос данных (`SELECT`).

Источник на данном рисунке – это внешний объект, из которого процесс источника запрашивает данные в соответствии с параметрами конфигурации и подключения, переданными при его создании. В роли источника могут выступать, например, база данных или брокер сообщений.

В рамках каждого процесса источника последовательно выполняются следующие действия:

1. Запрос данных из источника.

2. Обновление материализованных представлений на основе изменений в данных.
3. Выполнение запросов клиентов, связанных с данным источником и его материализованными представлениями.

Механизм функционирования процессов источников отражен в приложениях 4 и 5.

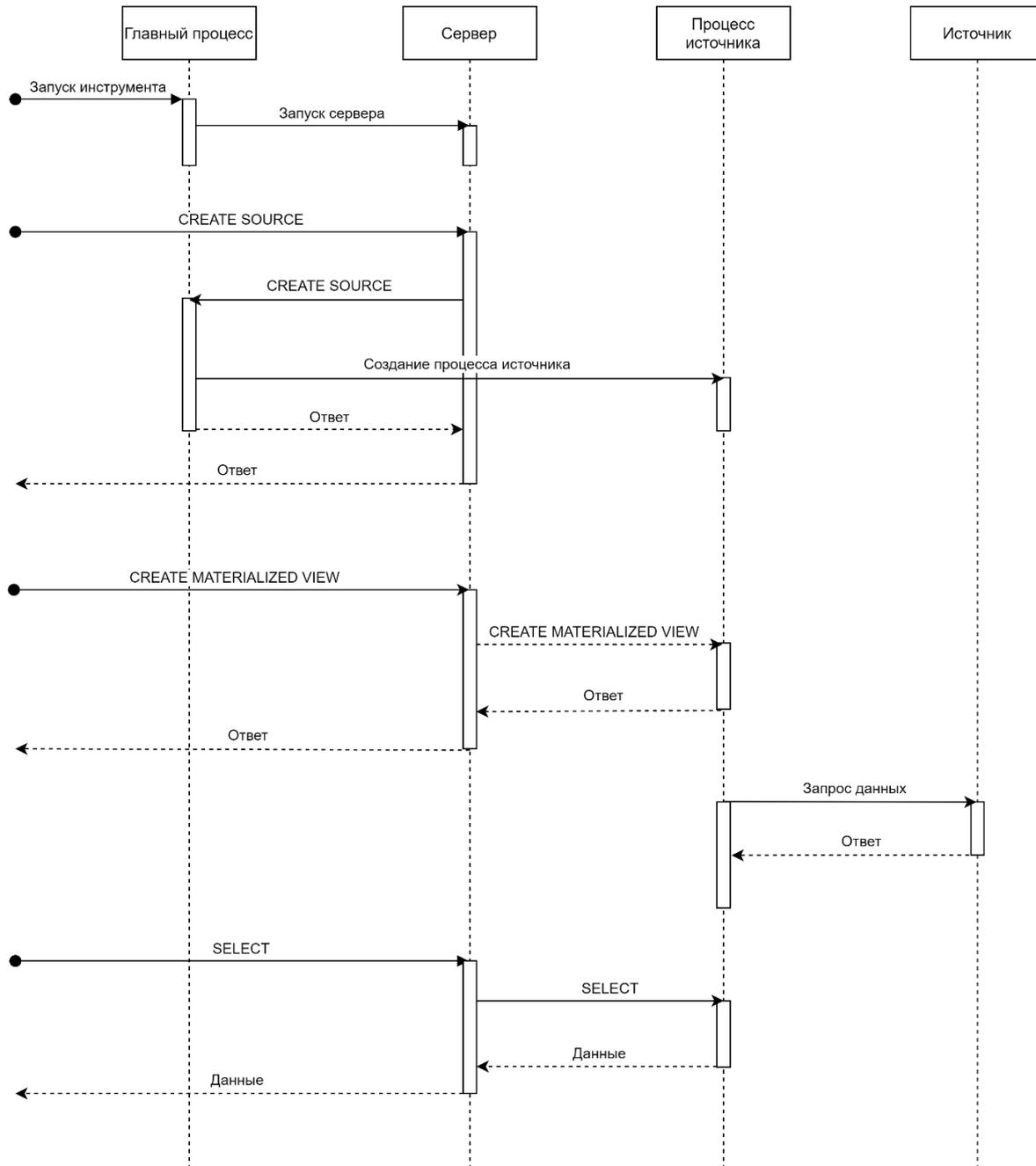


Рис. 1. Механизм функционирования инструмента

В настоящее время в инструменте поддерживаются следующие типы запросов:

- CREATE SOURCE – создание подключения к источнику данных;
- DROP SOURCE – удаление подключения к источнику данных;
- CREATE MATERIALIZED VIEW – создание материализованного представления;
- DROP MATERIALIZED VIEW – удаление материализованного представления;
- SELECT – получение данных;
- SELECT FORECASTED – получение результатов прогнозирования.

### **МАТЕРИАЛИЗОВАННЫЕ ПРЕДСТАВЛЕНИЯ**

В настоящее время в инструменте реализованы два типа материализованных представлений:

1. По умолчанию. Поддерживают операции вставки, изменения и удаления данных.
2. Append-only. Поддерживают только вставку данных, изменение и удаление невозможны. Преимущество этого типа материализованных представлений заключается в том, что в этом случае для ряда агрегатных функций нужно хранить меньше данных, что может привести к значительной экономии памяти.

При создании материализованного представления возможно указать функцию фильтрации. Эта функция аналогична WHERE в SQL, строки, для которых она возвращает false, не попадают в материализованное представление.

### **ПРИБЛИЖЕННЫЕ АЛГОРИТМЫ ОБРАБОТКИ ПОТОКОВЫХ ДАННЫХ**

В рамках инструмента приближенные алгоритмы реализованы в формате агрегатных функций. Последние являются классами, наследующимися от абстрактного класса Aggregate.

В настоящий момент в рамках инструмента реализованы приближенные алгоритмы, решающие ряд задач, распространенных в сфере потоковых данных.

1. Подсчет количества уникальных значений

Эта задача также известна как задача count-distinct. Она имеет высокую актуальность, в том числе, в сфере обработки big data [24].

---

Для решения этой задачи в рамках инструмента реализованы два алгоритма: алгоритм, основанный на фильтре Блума с подсчетом, и HyperLogLog [25, 26]. Эти алгоритмы предназначены для разных сценариев использования: HyperLogLog более эффективен по памяти, чем алгоритм, основанный на фильтре Блума с подсчетом, однако не поддерживает изменение и удаление данных.

## 2. Поиск преобладающего элемента последовательности

Эта задача широко распространена на практике, в частности, в распределенных системах, например, для избрания лидера или координатора системы путем голосования узлов. Инструмент содержит реализацию алгоритма большинства голосов Бойера–Мура, решающего данную задачу [27].

В будущем планируется расширить перечень алгоритмов, реализованных в рамках инструмента.

Количество приближенных алгоритмов и вероятностных структур данных, а также их модификаций, разработанных на данный момент, очень велико, для разных задач наиболее оптимальными являются разные алгоритмы. Поэтому для инструмента анализа потоковых данных, включающего в себя приближенные алгоритмы, важное значение имеет расширяемость.

## **ПРОГНОЗИРОВАНИЕ СОСТОЯНИЯ ПОТОКА**

В разработанном инструменте имеется возможность прогнозирования значений агрегатных функций с помощью различных методов экстраполяции. В дальнейшем планируется реализовать возможность применения для этой задачи методов машинного обучения. Так как существует большое количество различных методов машинного обучения, применимых для этой задачи, и разные методы эффективны в разных ситуациях, планируется сделать прогнозирование значений агрегатных функций расширяемым.

При создании материализованного представления можно включить для него прогнозирование агрегатных функций и задать параметры прогнозирования. По умолчанию прогнозирование отключено, так как для него необходимо кэширование предыдущих состояний материализованного представления, что приводит к дополнительным затратам памяти.

Прогнозирование осуществляется на уровне групп материализованного представления. Прогнозирование группы осуществляется на основе истории ее предыдущих состояний.

При запросе спрогнозированных значений по умолчанию возвращается результат, соответствующий текущему моменту времени. При этом можно получить результат прогнозирования для произвольного момента времени, ограниченного на практике границами применимости используемого алгоритма.

Кроме того, существует возможность обращения к истории состояний для получения снимка состояния материализованного представления в прошлом.

### **ПОДКЛЮЧЕНИЕ К ИСТОЧНИКАМ ДАННЫХ**

Для подключения к различным источникам данных в рамках инструмента используются коннекторы. Коннекторы к источникам данных являются классами, наследуемыми от абстрактного базового класса источника Source.

Существует возможность создания дополнительных коннекторов к источникам данных. Такая возможность имеет очень важное значение, так как разнообразие существующих источников потоковых данных и форматов потоковых данных крайне велико.

При разработке инструмента значительное внимание было уделено реализации поддержки взаимодействия с брокером сообщений Apache Kafka в связи с его высокой востребованностью в сфере потоковых данных [28]. Кроме того, была реализована поддержка инструмента CDC Debezium, что дает возможность обрабатывать поток данных об изменениях в базах данных, поддерживаемых инструментом, в частности, PostgreSQL, Oracle, MySQL.

### **ЗАКЛЮЧЕНИЕ**

Проведенное исследование показало, что в данный момент существует потребность в инструменте последовательного снятия снимков агрегированных данных из потоковых данных, включающем в себя приближенные алгоритмы анализа потоковых данных и прогнозирования состояния потока. При этом данная потребность не удовлетворена в полной мере ни одним из программных продуктов, присутствующих на рынке.

Авторами разработан такой инструмент, рассмотрены архитектура и особенности его функционирования. Также изучены перспективы его дальнейшего развития.

Разработанный инструмент доступен в репозитории GitHub [18].

### СПИСОК ЛИТЕРАТУРЫ

1. Гурьянова Э. А., Гурьянов А. И. Анализ и перспективы рынка SaaS в Российской Федерации // Вестник экономики, права и социологии. 2022. №1. С. 182–185.
2. Kolajo T., Daramola O., Adebisi A. Big data stream analysis: a systematic literature review. // Journal of Big Data. 2019. Vol. 6.  
<https://doi.org/10.1186/s40537-019-0210-7>
3. Маркова В. Д. Влияние цифровой экономики на бизнес // ЭКО. 2018. №12 (534). С. 7–22.
4. Определение потоковой передачи данных // Amazon Web Services (AWS). – URL: <https://aws.amazon.com/ru/streaming-data/> (дата обращения 12.05.2023)
5. Ельченков Р. А., Дунаев М. Е., Зайцев К. С. Прогнозирование временных рядов при обработке потоковых данных в реальном времени // International Journal of Open Information Technologies. 2022. Т. 10, №6. С. 62–69.
6. Алатова Н. В. Управление в экосистеме бизнеса в период цифровой трансформации // Эффективное управление экономикой: проблемы и перспективы. 2022. С. 238–241.
7. Маркова В. Д., Кузнецова С. А. Развитие стратегического менеджмента в цифровой экономике // Вестник Томского государственного университета. Экономика. 2019. №48. С. 217–232. <https://doi.org/10.17223/19988648/48/15>
8. Петренко А. С., Петренко С. А. Технологии больших данных (big data) в области информационной безопасности // The 2018 Symposium on Cybersecurity of the Digital Economy. 2018. С. 248–255.
9. Трофимов В. В., Трофимова Л. А. О концепции управления на основе данных в условиях цифровой трансформации // Петербургский экономический журнал. 2021. №4. С. 149–155. <https://doi.org/10.24412/2307-5368-2021-4-149-155>

10. *Логиновский О. В., Шестаков А. Л., Шинкарев А. А.* Построение современных корпоративных информационных систем // Управление большими системами: сборник трудов. 2019. №81. С. 113–146.

<https://doi.org/10.25728/ubs.2019.81.5>

11. *Alwaisi S. S. A., Abbood M. N., Jalil L. F., Kasim S., Fudzee M. F. M., Hadi R., Ismail M. A. A.* Review on Big Data Stream Processing Applications: Contributions, Benefits, and Limitations // International Journal on Informatics Visualization. 2021. Vol. 5(4). P. 456–460. <https://doi.org/10.30630/joiv.5.4.737>

12. *McSherry F.* View Maintenance: A New Approach to Data Processing // Materialize Blog. 2020. URL: <https://materialize.com/blog/olvm/> (дата обращения 12.05.2023)

13. *Singh A., Garg S., Kaur R., Batra S., Kumar N., Zomaya A. Y.* Probabilistic data structures for big data analytics: A comprehensive review // Knowledge-Based Systems. 2020. Vol. 188. <https://doi.org/10.1016/j.knosys.2019.104987>

14. *Torres J. F., Hadjout D., Sebaa A., Martinez-Alvarez F., Troncoso A.* Deep Learning for Time Series Forecasting: A Survey // Big Data. 2021. Vol 9(1). <https://doi.org/10.1089/big.2020.0159>

15. *Brandt T. L., Grawunder M.* Moving Object Stream Processing with Short-Time Prediction // Proceedings of the 8th ACM SIGSPATIAL Workshop on GeoStreaming. 2017. <https://doi.org/10.1145/3148160.3148168>

16. Incremental Computation in the Database // Materialize. – URL: <https://materialize.com/guides/incremental-computation/> (дата обращения 12.05.2023)

17. Upserts in Differential Dataflow // Materialize Blog. 2020. URL: <https://materialize.com/blog/upserts-in-differential-dataflow/> (дата обращения 12.05.2023)

18. *artemgur/Diplom* // GitHub. URL: <https://github.com/artemgur/diplom> (дата обращения 12.05.2023)

19. Materialize Documentation // Materialize. URL: <https://materialize.com/docs/> (дата обращения 12.05.2023)

20. Data definition // *ksqldb Documentation*. URL: <https://docs.ksqldb.io/en/latest/reference/sql/data-definition/> (дата обращения 12.05.2023)

21. Streaming ingestion // Amazon Redshift.

URL: <https://docs.aws.amazon.com/redshift/latest/dg/materialized-view-streaming-ingestion.html> (дата обращения 12.05.2023)

22. Confluent Community License Agreement // GitHub. 2018.

URL: <https://github.com/confluentinc/ksql/blob/master/LICENSE> (дата обращения 12.05.2023)

23. Materialize Business Source License Agreement // GitHub.

URL: <https://github.com/MaterializeInc/materialize/blob/main/LICENSE> (дата обращения 12.05.2023)

24. *Ting D.* Approximate Distinct Counts for Billions of Datasets // Proceedings of the 2019 International Conference on Management of Data. 2019. P. 69–86.  
<https://doi.org/10.1145/3299869.3319897>

25. *Fan L., Cao P., Almeida J., Broder A.* Summary Cache: A Scalable Wide-Area Web Cache Sharing Protocol // IEEE/ACM Transactions on Networking. 2000. Vol. 8(3). P. 281–293. <https://doi.org/10.1109/90.851975>

26. *Flajolet P., Fusy E., Gandouet O., Meunier F.* HyperLogLog: the analysis of a near-optimal cardinality estimation algorithm // Discrete Mathematics & Theoretical Computer Science. 2007. P. 137–156.

27. *Boyer R.S., Moore J.S.* MJRTY – A Fast Majority Vote Algorithm // Automated Reasoning / ed. Boyer R. S. Dordrecht: Kluwer Academic Publishers, 1991. P. 105–117.  
[https://doi.org/10.1007/978-94-011-3488-0\\_5](https://doi.org/10.1007/978-94-011-3488-0_5)

28. *Singh B., Chaitra B. H.* Comprehensive Review of Stream Processing Tools // International Research Journal of Engineering and Technology. 2020. Vol. 7(5). P. 3537–3540.

## TOOL FOR SEQUENTIAL SNAPSHOTTING OF AGGREGATED DATA FROM STREAMING DATA

A. I. Gurianov<sup>1</sup> [0000-0002-9870-7973], A. S. Yakupov<sup>2</sup> [0000-0002-2333-8819]

<sup>1</sup> Joint Supercomputer Center of the Russian Academy of Sciences, Lobachevskii ul., 2, Kazan, 420008 Russia

<sup>1,2</sup> Kazan (Volga region) Federal University, 35 Kremlevskaya str., Kazan, 420008 Russia

<sup>1</sup>armgnv@gmail.com, <sup>2</sup>asyakupov@kpfu.ru

### **Abstract**

In the modern world, streaming data has become widespread in many subject areas. The task of processing streaming data in real time, with minimal delay, is highly relevant.

In stream processing, data processing, various approximate algorithms are often used, which have much higher time and memory efficiency than exact algorithms. In addition, there is often a need to forecast the state of the stream.

Thus, there is currently a need for a tool for sequential snapshotting of aggregated data from streaming data, enabling flow state prediction and approximate algorithms for stream data processing.

The authors of the article have developed such a tool, reviewed its architecture and mechanism of functioning, and evaluated the prospects for its further development.

**Keywords:** *streaming data, stream processing, stream analysis, materialized views, streaming algorithms, approximate algorithms, stream forecasting.*

### **REFERENCES**

1. Gurianova E. A., Gurianov A. I. Analiz i perspektivy rynka SaaS v Rossijskoj Federacii // Vestnik jekonomiki, prava i sociologii. 2022. №1. S. 182–185.

2. Kolajo T., Daramola O., Adebiji A. Big data stream analysis: a systematic literature review // Journal of Big Data. 2019. Vol. 6.

<https://doi.org/10.1186/s40537-019-0210-7>

3. *Markova V. D.* Vliyanie cifrovoj jekonomiki na biznes // JeKO. 2018. №12 (534). S. 7–22.

4. *Opređenje potokovoj peredachi dannyh* // Amazon Web Services (AWS). URL: <https://aws.amazon.com/ru/streaming-data/>, last accessed 12.05.2023

5. *El'chenkov R. A., Dunaev M. E., Zajcev K. S.* Prognozirovanie vremennyh rjadov pri obrabotke potokovyh dannyh v real'nom vremeni // International Journal of Open Information Technologies. 2022. T. 10, №6. S. 62–69.

6. *Apatova N. V.* Upravlenie v jekosisteme biznesa v period cifrovoj transformacii // *Jeffektivnoe upravlenie jekonomikoj: problemy i perspektivy*. 2022. S. 238–241.

7. *Markova V. D., Kuznecova S. A.* Razvitie strategicheskogo menedzhmenta v cifrovoj jekonomike // Vestnik Tomskogo gosudarstvennogo universiteta. Jekonomika. 2019. №48. S. 217–232. <https://doi.org/10.17223/19988648/48/15>

8. *Petrenko A. S., Petrenko S. A.* Tehnologii bol'shih dannyh (big data) v oblasti informacionnoj bezopasnosti // The 2018 Symposium on Cybersecurity of the Digital Economy. 2018. S. 248–255.

9. *Trofimov V. V., Trofimova L. A.* O koncepcii upravlenija na osnove dannyh v uslovijah cifrovoj transformacii // Peterburgskij jekonomicheskij zhurnal. 2021. №4. S. 149–155. <https://doi.org/10.24412/2307-5368-2021-4-149-155>

10. *Loginovskij O. V., Shestakov A. L., Shinkarev A. A.* Postroenie sovremennyh korporativnyh informacionnyh sistem // *Upravlenie bol'shimi sistemami: sbornik trudov*. 2019. №81. S. 113–146. <https://doi.org/10.25728/ubs.2019.81.5>

11. *Alwaisi S. S. A., Abbood M. N., Jalil L. F., Kasim S., Fudzee M. F. M., Hadi R., Ismail M. A. A.* Review on Big Data Stream Processing Applications: Contributions, Benefits, and Limitations // International Journal on Informatics Visualization. 2021. Vol. 5(4). P. 456–460. <https://doi.org/10.30630/joiv.5.4.737>

12. *McSherry F.* View Maintenance: A New Approach to Data Processing // Materialize Blog. 2020.

URL: <https://materialize.com/blog/olvm/>, last accessed 12.05.2023

13. *Singh A., Garg S., Kaur R., Batra S., Kumar N., Zomaya A. Y.* Probabilistic data structures for big data analytics: A comprehensive review // Knowledge-Based Systems. 2020. Vol. 188. <https://doi.org/10.1016/j.knosys.2019.104987>

14. *Torres J. F., Hadjout D., Sebaa A., Martinez-Alvarez F., Troncoso A.* Deep Learning for Time Series Forecasting: A Survey // *Big Data*. 2021. Vol. 9(1).

<https://doi.org/10.1089/big.2020.0159>

15. *Brandt T. L., Grawunder M.* Moving Object Stream Processing with Short-Time Prediction // *Proceedings of the 8th ACM SIGSPATIAL Workshop on GeoStreaming*. 2017. <https://doi.org/10.1145/3148160.3148168>

16. Incremental Computation in the Database // *Materialize*.

URL: <https://materialize.com/guides/incremental-computation/>, last accessed 12.05.2023

17. Upserts in Differential Dataflow // *Materialize Blog*. 2020.

URL: <https://materialize.com/blog/upserts-in-differential-dataflow/>, last accessed 12.05.2023

18. *artemgur/Diplom* // *GitHub*.

URL: <https://github.com/artemgur/diplom>, last accessed 12.05.2023

19. *Materialize Documentation* // *Materialize*.

URL: <https://materialize.com/docs/>, last accessed 12.05.2023

20. Data definition // *ksqlDB Documentation*.

URL: <https://docs.ksqldb.io/en/latest/reference/sql/data-definition/>, last accessed 12.05.2023

21. Streaming ingestion // *Amazon Redshift*.

URL: <https://docs.aws.amazon.com/redshift/latest/dg/materialized-view-streaming-ingestion.html>, last accessed 12.05.2023

22. *Confluent Community License Agreement* // *GitHub*. 2018. – URL: <https://github.com/confluentinc/ksql/blob/master/LICENSE>, last accessed 12.05.2023

23. *Materialize Business Source License Agreement* // *GitHub*.

URL: <https://github.com/MaterializeInc/materialize/blob/main/LICENSE>, last accessed 12.05.2023

24. *Ting D.* Approximate Distinct Counts for Billions of Datasets // *Proceedings of the 2019 International Conference on Management of Data*. 2019. P. 69–86. <https://doi.org/10.1145/3299869.3319897>

25. *Fan L., Cao P., Almeida J., Broder A.* Summary Cache: A Scalable Wide-Area Web Cache Sharing Protocol // IEEE/ACM Transactions on Networking. 2000. Vol. 8(3). P. 281–293. <https://doi.org/10.1109/90.851975>

26. *Flajolet P., Fusy E., Gandouet O., Meunier F.* HyperLogLog: the analysis of a near-optimal cardinality estimation algorithm // Discrete Mathematics & Theoretical Computer Science. 2007. P. 137–156.

27. *Boyer R. S., Moore J. S.* MJRTY – A Fast Majority Vote Algorithm // Automated Reasoning / ed. Boyer R. S. Dordrecht: Kluwer Academic Publishers, 1991. P. 105–117. [https://doi.org/10.1007/978-94-011-3488-0\\_5](https://doi.org/10.1007/978-94-011-3488-0_5)

28. *Singh B., Chaitra B. H.* Comprehensive Review of Stream Processing Tools // International Research Journal of Engineering and Technology. 2020. Vol. 7(5). P. 3537–3540.

---

#### СВЕДЕНИЯ ОБ АВТОРАХ



**ГУРЬЯНОВ Артем Игоревич** – студент, Казанский (Приволжский) федеральный университет, г. Казань.

**Artem Igorevich GURIANOV** – student, Kazan (Volga region) Federal University, Kazan.

email: [armgnv@gmail.com](mailto:armgnv@gmail.com)

ORCID: 0000-0002-9870-7973



**ЯКУПОВ Азат Шавкатович** – старший преподаватель, Казанский (Приволжский) федеральный университет, г. Казань.

**Azat Shavkatovich YAKUPOV** – Senior Lecturer, Kazan (Volga region) Federal University, Kazan.

email: [asyakupov@kpfu.ru](mailto:asyakupov@kpfu.ru)

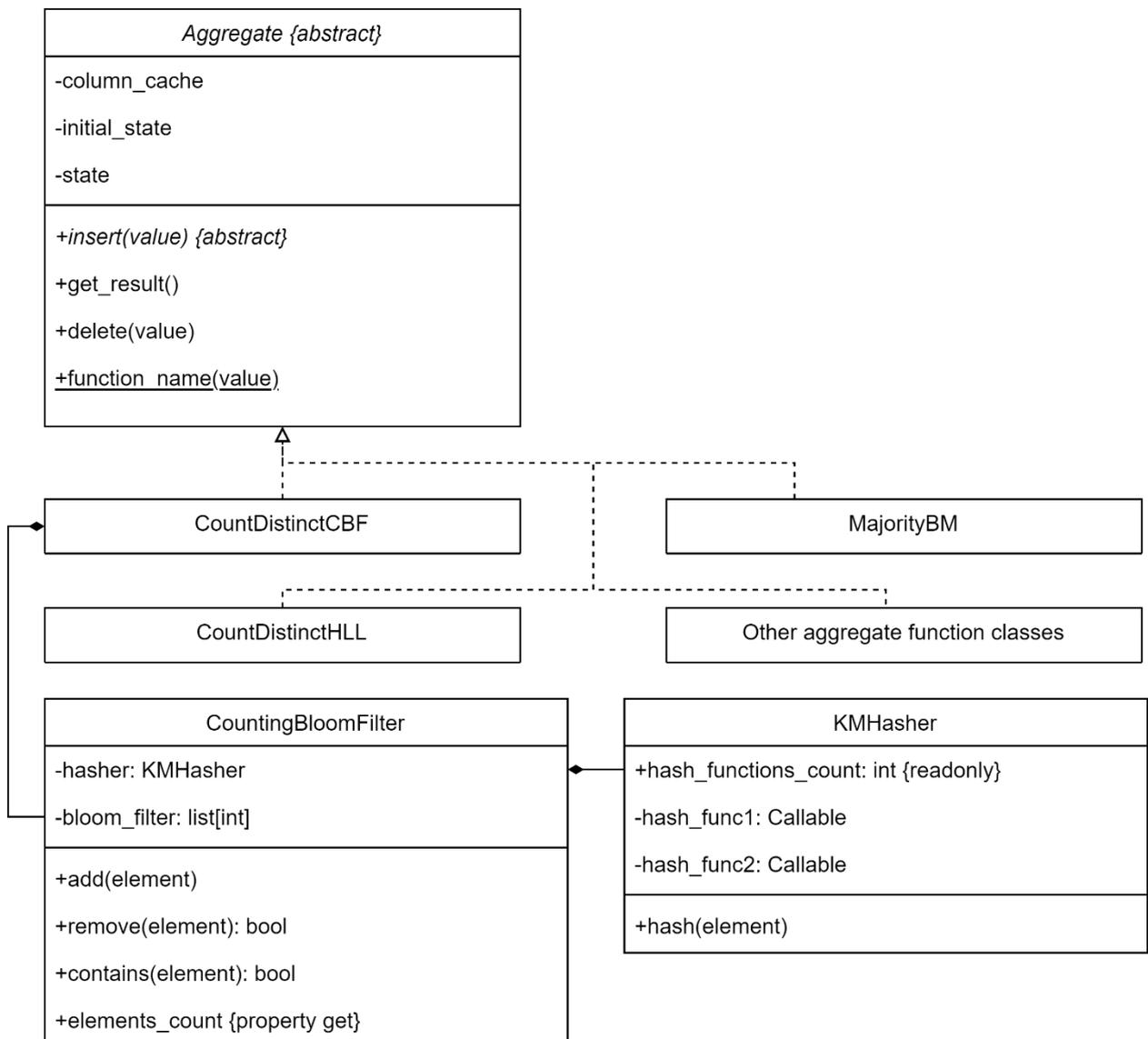
ORCID: 0000-0002-2333-8819

*Материал поступил в редакцию 21 мая 2023 года*

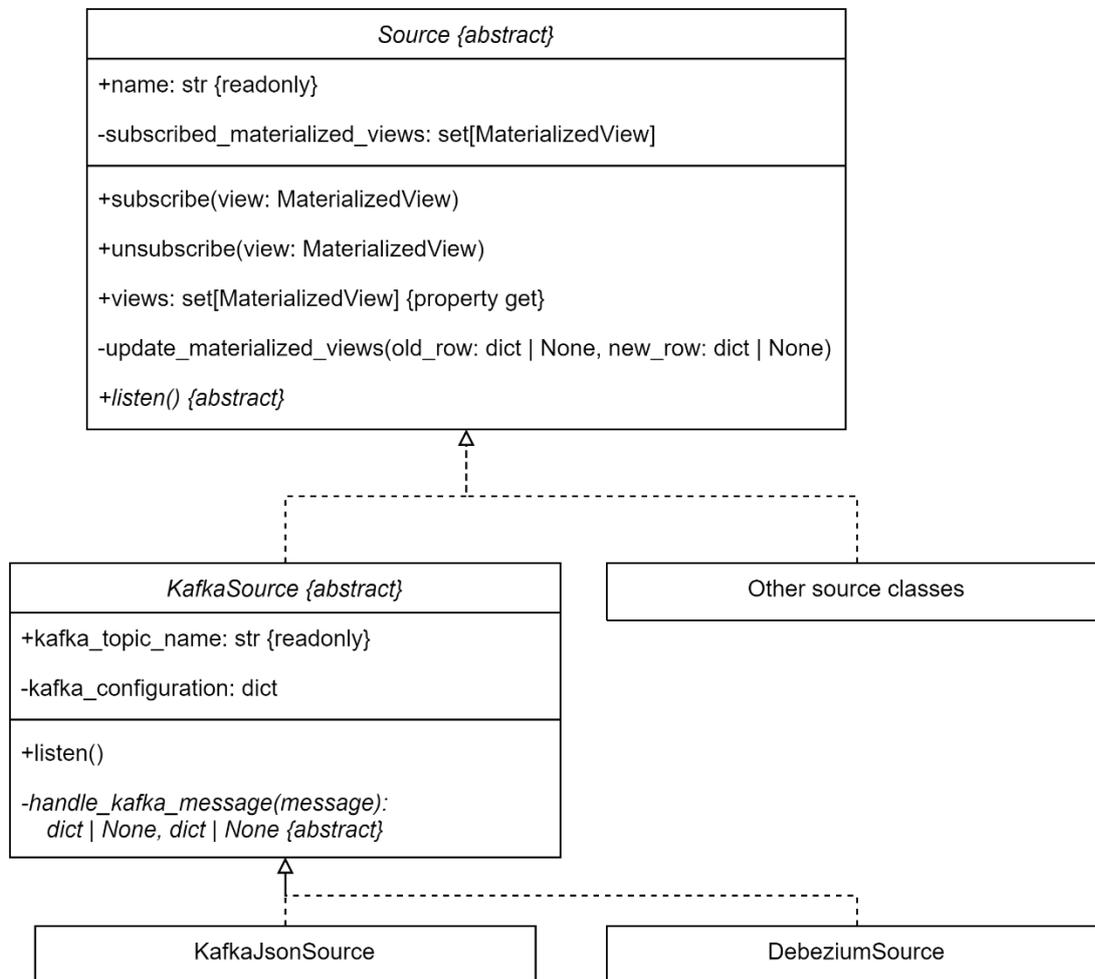
---



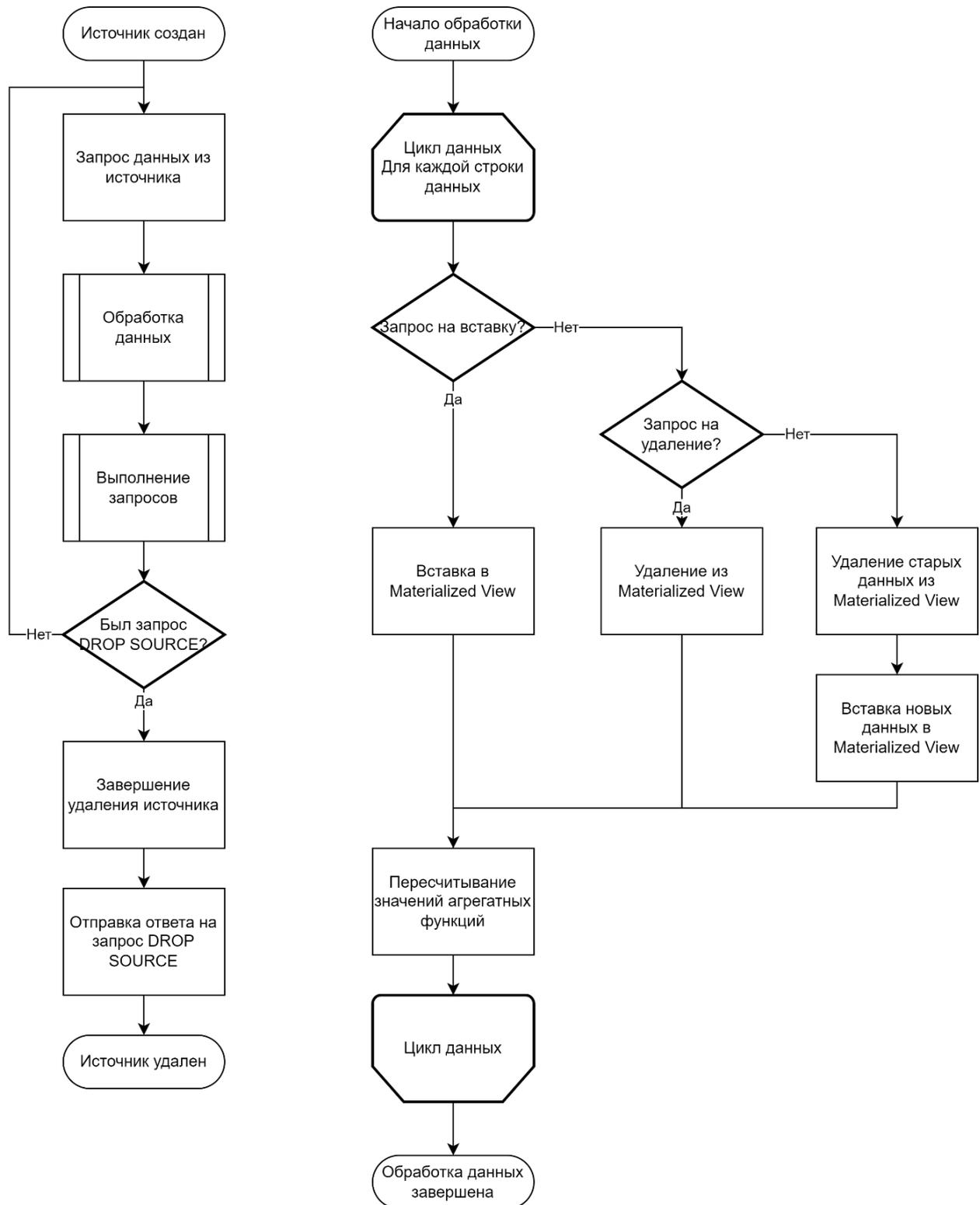
Приближенные алгоритмы обработки потоковых данных



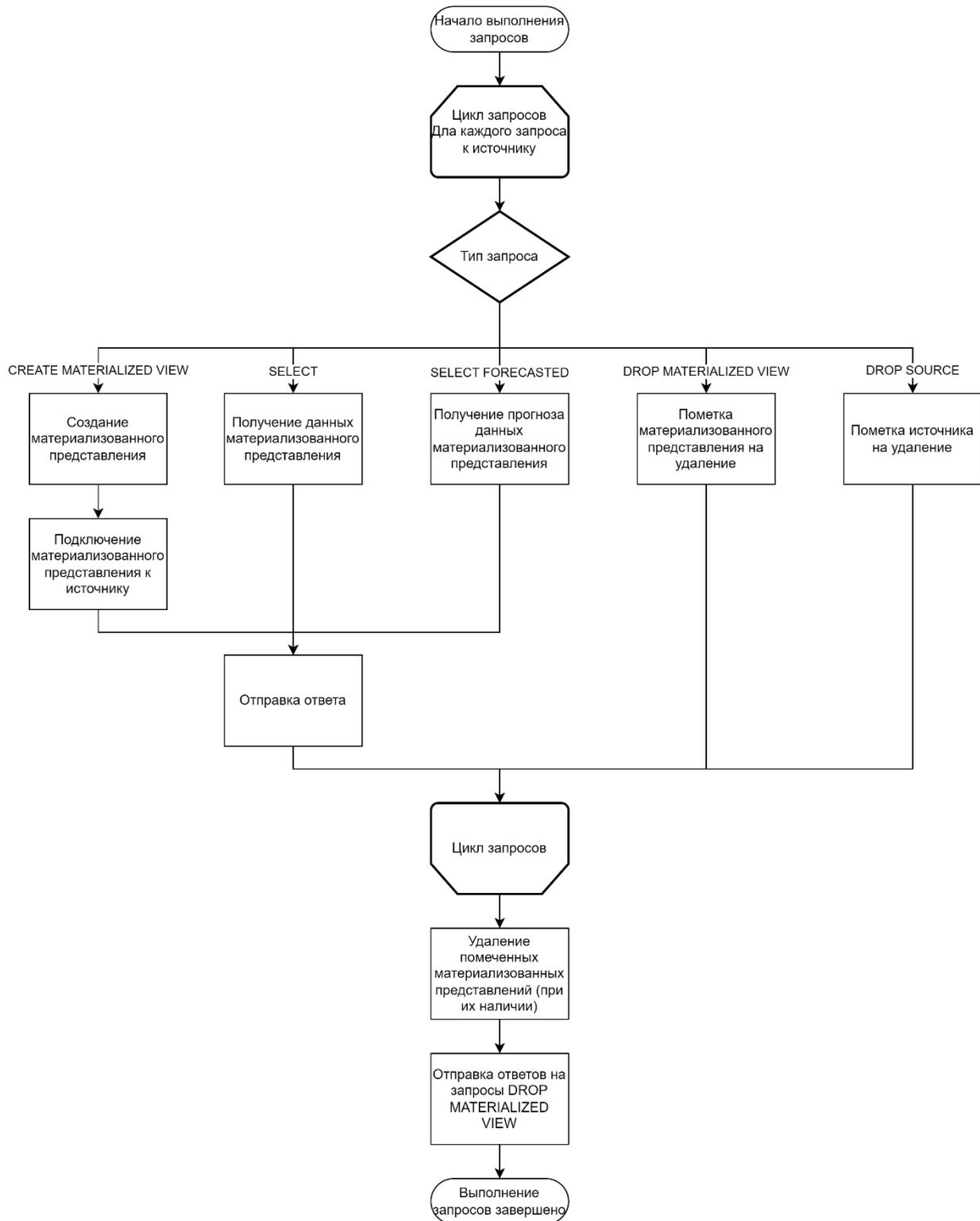
Коннекторы к источникам данных



Механизм функционирования процессов источников



Механизм выполнения запросов в рамках процессов источников



## НЕЙРОННАЯ СЕТЬ ДЛЯ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ТЕКСТА ПЕСЕН С ПРИМЕНЕНИЕМ МОДЕЛЕЙ OPENAI И CLIP

А. Р. Давлетгареева<sup>1</sup> [0009-0008-7258-470X], К. А. Едкова<sup>2</sup> [0009-0005-4706-2254]

<sup>1,2</sup> *Институт информационных технологий и интеллектуальных систем Казанского федерального университета, ул. Кремлевская, 35, г. Казань, 420008*

<sup>1</sup>alsudavletgareeva@gmail.com, <sup>2</sup>ksushka.e21@gmail.com

### **Аннотация**

Исследована эффективность моделей ImageNet diffusion model и CLIP для генерации изображений по текстовому описанию. С использованием различных текстовых вводов на разных параметрах проведены два эксперимента для определения лучших параметров при генерации изображений на основе текстового описания. Результаты показали, что, хотя ImageNet хорошо справляется с созданием изображений, CLIP лучше обеспечивает соединение текстовых подсказок с релевантными изображениями. Полученные результаты характеризуют высокий потенциал объединения названных моделей для создания высококачественных и контекстно релевантных изображений на основе текстового описания.

**Ключевые слова:** *генерация изображений, ImageNet diffusion model, CLIP, глубокое обучение, нейронные сети, обработка естественного языка.*

### **ВВЕДЕНИЕ**

Искусственный интеллект (ИИ) в области генерации изображений развивался с первых дней возникновения машинного обучения (МО). Первые попытки создания изображений на основе текстовых описаний были предприняты с использованием систем, основанных на правилах, которые были ограничены в своей способности улавливать нюансы и сложности человеческого языка [1]. Однако с появлением методов глубокого обучения, таких как генеративные состязательные сети (GAN) и вариационные автокодеры (VAEs) [2], в этой области наблюдается значительный прогресс.

Генерация изображений на основе текстового описания привлекла значительное внимание в последние годы благодаря своему потенциалу генерировать высококачественные изображения на основе текстовых описаний [2].

Одним из наиболее популярных подходов к созданию изображений из текста является использование сетей GAN [3], которые состоят из двух нейронных сетей: сети-генератора, которая генерирует изображения, и сети-дискриминатора, которая пытается отличить реальные изображения от сгенерированных. Объединяя эти сети в игровой среде, GAN способны создавать высококачественные изображения, которые трудно отличить от реальных.

Еще одним популярным направлением в этой области исследований стала разработка крупномасштабных языковых моделей, таких как GPT-3 [4]. Эти модели способны генерировать текст, который практически неотличим от текста, написанного человеком, и могут использоваться также для генерации текстовых описаний изображений. Это открыло новые возможности для исследований в области генерации изображений, поскольку позволяет генерировать изображения на основе описаний на естественном языке, а не полагаться на заранее определенные правила или шаблоны.

Отметим, что способность генерировать изображения на основе текстовых описаний значительно расширилась за последние годы благодаря достижениям в области методов глубокого обучения и крупномасштабных языковых моделей [2, 5]. Хотя предстоит преодолеть еще много проблем, таких, например, как создание изображений, точно отражающих нюансы человеческого языка, будущее этой области выглядит многообещающим, и можно ожидать много интересных разработок в ближайшие годы [6].

В настоящей статье мы предлагаем подход к созданию изображений на основе текстов песен и используем два алгоритма из репозитория OpenAI [7] для генерации изображений, которые строятся на основе текста. Чтобы достичь желаемого результата, мы представляем текст песни в виде последовательности слов, выделяя ключевые выражения, и кодируем их, используя предварительно обученную языковую модель.

## АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ

Для создания изображения на основе текста песни была предложена архитектура нейронной сети (Рисунок 1), построенная на основе генеративно-сопоставительной нейросети (GAN). GAN – это один из алгоритмов классического МО – обучения без учителя. Его суть заключается в комбинации двух нейросетей: генератора и дискриминатора [8]. Задача генератора (Рисунок 1а) заключается в генерации образов заданной категории, а задача дискриминатора – в распознавании созданных образов.

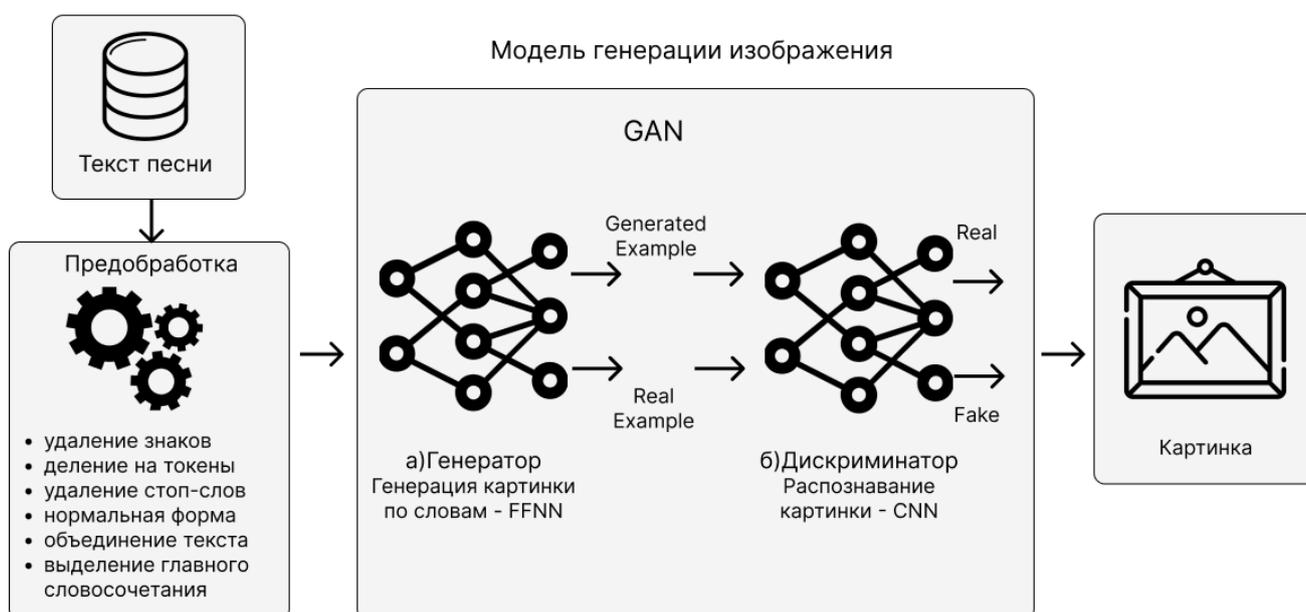


Рис. 1. Архитектура нейронной сети №1

Дискриминатор (Рисунок 1б) использует сверточные нейронные сети (CNN) [9] для распознавания образов на изображении. Чтобы обучить нейронную сеть распознавать образы, необходимо обработать большое количество изображений, на которых присутствуют искомые образы. Генератор начинает формирование изображений с создания произвольного шума, на котором постепенно появляются фрагменты искомого изображения. В качестве генерирующей нейронной сети могут использоваться сети FFNN – нейронные сети прямого распространения [10].

Шаги, которые проходит GAN: генератор получает случайное число и возвращает изображение, а затем передает это сгенерированное изображения дискриминатору вместе с потоком изображений, взятых из фактического набора данных. Дискриминатор принимает набор из реальных и сгенерированных изображений и возвращает вероятности, 0 или 1, где 1 – изображение подлинное, а 0 – фальшивое.

Из-за ограничений нашего вычислительного оборудования мы не смогли использовать предложенную ранее архитектуру нейронной сети для генерации изображений по тексту песни. В связи с этим мы обратили внимание на недавние достижения в области искусственного интеллекта, особенно на решения от компании OpenAI.

Одним из таких решений является модель CLIP (Contrastive Language-Image Pre-Training) [11], которая была разработана для связи естественного языка и изображений. Эта модель использует большой корпус данных текстов и изображений для обучения своей архитектуры, что позволяет ей понимать связь между текстом и изображением. Благодаря этому CLIP может использоваться для различных задач, в том числе для генерации изображений по тексту.

Вторым решением в репозитории компании OpenAI является модель ImageNet diffusion [12], которая позволяет генерировать высококачественные изображения с помощью диффузии. Она обучается на большом наборе изображений (1,3 миллиардах изображений из набора данных ImageNet), что позволяет ей генерировать разнообразные изображения с высокой детализацией и качеством: пиковое отношение сигнал/шум (PSNR) – диффузионная модель ImageNet генерирует изображения с разрешением до 40 дБ, Индекс структурного сходства (SAM) – генерирует изображения с суммой до 0,95, средняя оценка мнений (MOS) – до 5.

Диффузионные модели – это тип генерирующих моделей, которые обучают генерировать изображения путем постепенного добавления шума к изображению, а затем итеративного удаления шума до тех пор, пока изображение не станет неотличимым от реального изображения. Поэтому, в качестве альтернативы модели GAN, мы решили применить интеграцию решений от OpenAI, такие как

модели CLIP (Рисунок 2а) и ImageNet (Рисунок 2б) diffusion для генерации изображений по тексту песни (Рисунок 2). Эти модели показали высокую эффективность в задачах генерации изображений и позволяют создавать качественные визуальные интерпретации текстовых данных [12].

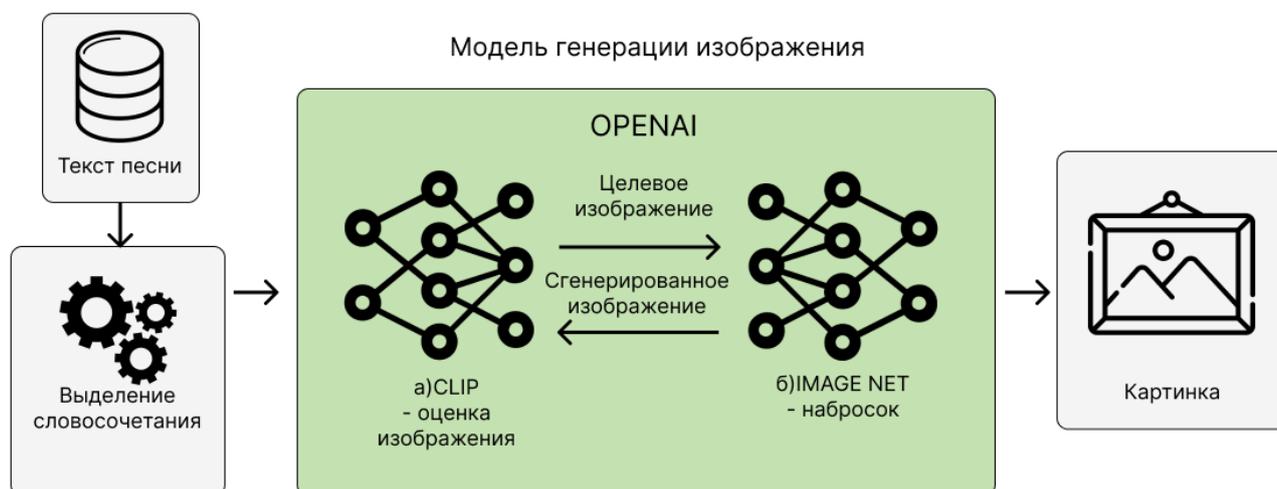


Рис. 2. Архитектура нейронной сети №2

ImageNet – это нейронная сеть, предназначенная для генерации изображений с высоким разрешением. Для генерации изображения использовалась обученная диффузионная модель ImageNet 256 x 256, которая превосходит [12] генеративную модель GAN при генерации изображений.

CLIP [11] – это нейросеть обученная на парах текст-изображение. Таким образом CLIP выстраивает ассоциации на основе текстового описания с соответствующими изображениями.

### ЭКСПЕРИМЕНТЫ И ИХ РЕЗУЛЬТАТЫ

Были проведены два эксперимента на основе различных параметров, для каждого эксперимента использовались 10 текстовых выборок. Модель ImageNet diffusion была обучена на большом наборе данных изображений для генерации качественных изображений из текстовых подсказок. Нейросеть CLIP, с другой стороны, была обучена понимать взаимосвязь между текстом и изображениями, что позволило ей найти наиболее релевантное изображение для данного текстового

описания. Сгенерированные изображения оценивались на основе их качества и соответствия текстовому описанию.

Мы подобрали песни из различных музыкальных жанров на основе личных предпочтений, а при выборе словосочетаний руководствовались их смысловой составляющей, предпочитая использовать конкретные выражения, которые могут вызывать у большинства людей представление ясной и определенной картинки, а не абстрактные конструкции (см. Таблицу 1).

Таблица 1. Песни, выбранные для проведения экспериментов.

№	Исполнитель	Название песни	Выбранный отрывок
1	Smash Mouth	All Star	“shooting stars”
2	Rufus Wainwright	Hallelujah	“hallelujah”
3	The neighborhood	Sweater Weather	“sweater weather”
4	Sia	Snowman	“cry snowman”
5	Sia	Courage To Change	“news on TV”
6	Lady Gaga	Bloody Mary	“bloody mary”
7	Ariana Grande	7 rings	“bottles of bubbles”
8	Artem Kolpakov	The Blue Tractor	“blue tractor on big wheels”
9	Ryan Gosling	City Of Stars	“crowded restaurant”
10	Mark Philippe	Dancer in the Dark	“Dancer in the dark”

Для каждого отрывка было проведено по 2 эксперимента. Было замечено, что с определенными параметрами генерируемые изображения абсолютно не соответствуют предполагаемому результату. В экспериментах рассматривались 2 параметра: clip\_guidance\_scale и range\_scale.

Параметр `clip_guidance_scale` отвечает за масштабирование влияния функции потерь, связанной с моделью CLIP. Модель CLIP используется для оценки сходства между сгенерированными изображениями и их текстовыми описаниями. Чем выше значение `clip_guidance_scale`, тем больший учет будет приниматься вклад функции потерь от модели CLIP при оптимизации генерации изображений. Более высокое значение этого параметра означает, что сгенерированные изображения будут более точно соответствовать заданным текстовым описаниям.

Параметр `range_scale` отвечает за масштабирование влияния функции потерь, связанной с диапазоном значений пикселей в сгенерированных изображениях. Функция потерь, связанная с диапазоном, гарантирует, что значения пикселей в сгенерированных изображениях остаются в допустимом диапазоне, например, от 0 до 1. Чем выше значение `range_scale`, тем больше будет учитываться влияние функции потерь, связанной с диапазоном, при оптимизации генерации изображений, и сгенерированные изображения будут более ограничены в диапазоне значений пикселей.

Для экспериментов значения параметров были выбраны случайным образом, а именно: для первого эксперимента `clip_guidance_scale = 1000`, `range_scale = 100`; для второго эксперимента `clip_guidance_scale = 1500`, `range_scale = 150`.

Рассмотрим пример с генерацией изображения по тексту “shooting stars”. На шагах 4–6 (Рисунок 3) можно было предположить, что алгоритм пытается нарисовать морскую звезду, т. е. он выделяет из словосочетания слово “stars”. Далее с каждым шагом изображение всё меньше похоже на звезду, и результат (Рисунок 4) уже мало похож на морскую звезду, не говоря уже о падающих (“shooting”) звездах.

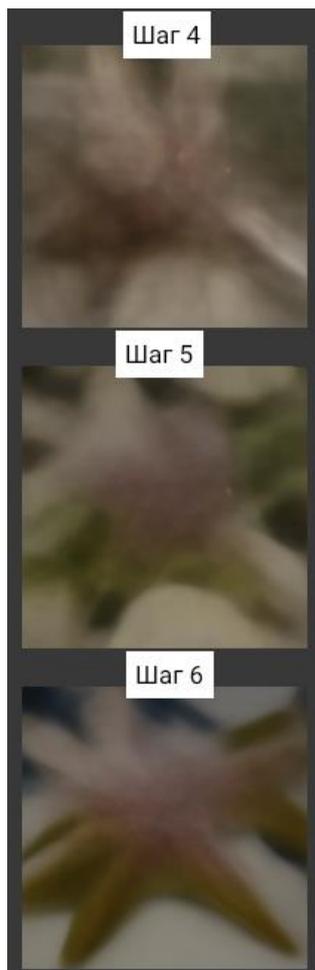


Рис. 3. Результаты генерации изображения для “shooting stars” на шагах 4–6



Рис. 4. Результат эксперимента “shooting stars”

Далее на Рисунке 5 приведены изображения, сгенерированные в результате двух экспериментов для каждой песни. Номер на изображении расшифровывается следующим образом:

<номер песни из таблицы 1>.<номер эксперимента>.

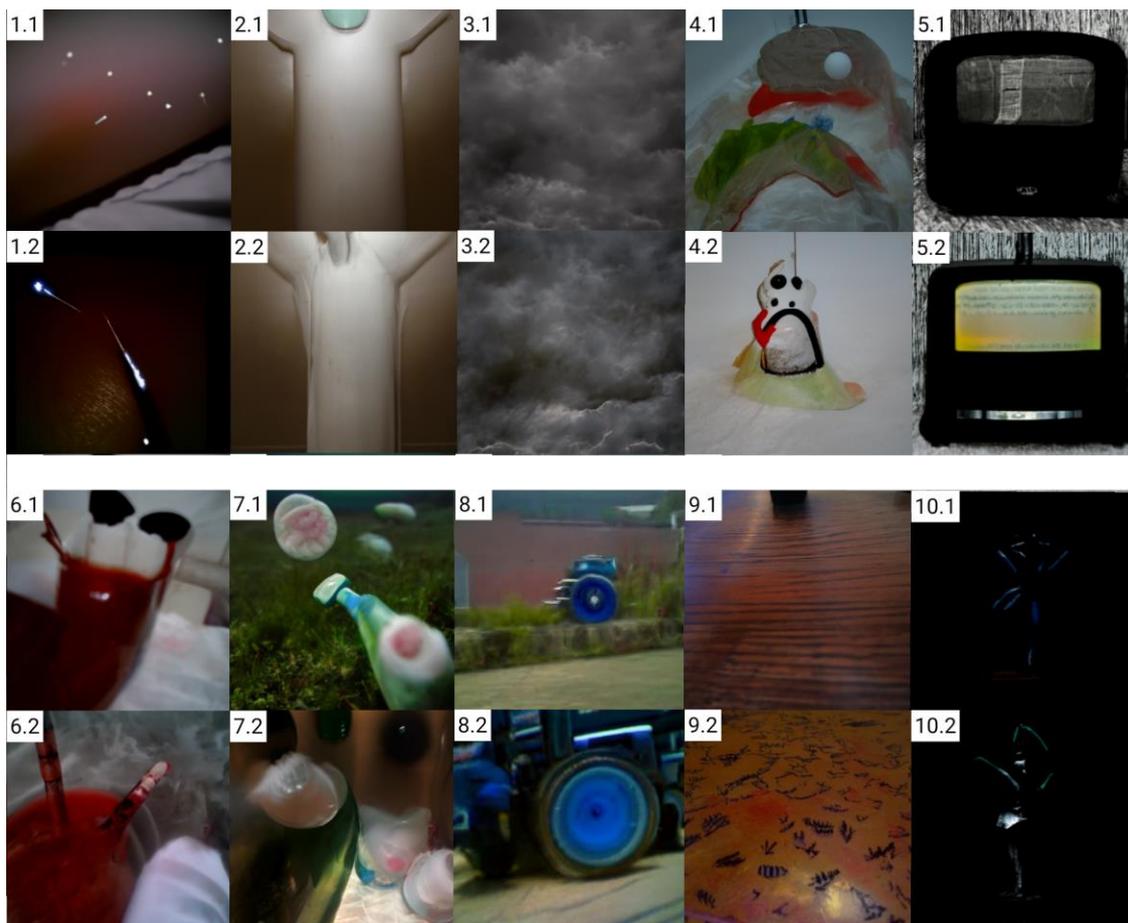


Рис. 5. Результаты работы алгоритма.

Рассмотрим результаты генерации изображений (Рисунок 5):

- 1.1 и 1.2: можно заметить падающие звезды, при этом картинка 1.1 как будто больше похожа на “shooting stars”;
- 2.1 и 2.2: учитывая, что выражение “hallelujah” очень абстрактно, не совсем понятно, что именно пыталась изобразить нейросеть;
- 3.1 и 3.2: “sweater weather” можно перевести как прохладная погода (погода для надевания свитера), алгоритм сгенерировал мрачную погоду, тучи, молнии;
- 4.1 и 4.2: не похожи на снеговика, на первой картинке – пингвин, на втором – что-то отдаленно похожее на снеговика;
- 5.1 и 5.2: можно увидеть телевизор, на втором изображении более четкий текст;
- 6.1 и 6.2: просматривается коктейль «Кровавая Мэри»;

- 7.1 и 7.2: если постараться, то можно разглядеть пузырьки и бутылку, но в целом получилось не очень понятное изображение;
- 8.1 и 8.2: на первой картинке нейронная сеть сгенерировала что-то похожее на трактор, на второй можно увидеть только часть картинки;
- 9.1 и 9.2: на обоих изображениях можно разглядеть какой-то узор, что совершенно не подходит под описание “crowded restaurant”, что в переводе означает «переполненный ресторан»;
- 10.1 и 10.2: на первой картинке можно увидеть контур балерины-танцовщицы в темноте, на второй же более четкий силуэт;

После завершения экспериментов мы приняли решение провести опрос, чтобы оценить соответствие визуального восприятия людей относительно выражений, выделенных в тексте, с изображениями, созданными нейронной сетью. Для оценки сгенерированных изображений использовалась шкала от 0 до 10, где 0 – нет схожести, а 10 – точно совпадает. Опрашиваемая группа: 18 человек в возрасте от 18 до 67. Среди них 2 мужчин и 16 женщин. Результаты опроса представлены в Таблице 2.

Таблица 2. Анализ оценки результатов экспериментов по опросу

	среднее значение	максимум	минимум	среднее отклонение	дисперсия	медиана
1.1 НАЗВАНИЕ - All Star из мультфильма «Шрэк» Smash Mouth ОТРЫВОК - shooting stars Параметры: 1000	7,27	10,00	3,00	1,88	6,02	7,14
1.2 НАЗВАНИЕ - All Star из мультфильма «Шрэк» Smash Mouth ОТРЫВОК - shooting stars Параметры: 1500	6,82	10,00	2,00	2,05	6,56	6,91
1.3 НАЗВАНИЕ - All Star из мультфильма «Шрэк» Smash Mouth ОТРЫВОК - shooting stars Параметры: 500	1,82	8,00	<b>0,00</b>	1,80	6,16	1,81
2.1 НАЗВАНИЕ - Hallelujah из мультфильма «Шрэк» Rufus Wainwright ОТРЫВОК - Hallelujah Параметры: 1000	4,18	9,00	<b>0,00</b>	3,26	12,76	2,63

2.2 НАЗВАНИЕ - Hallelujah из мультфильма «Шрек» Rufus Wainwright ОТРЫВОК - Hallelujah Параметры: 1500	4,36	9,00	<b>0,00</b>	2,69	9,25	2,35
3.1 НАЗВАНИЕ - The neighbourhood Sweater Weather ОТРЫВОК - Sweater weather Параметры:1000	7,36	10,00	4,00	2,33	7,05	4,50
3.2 НАЗВАНИЕ - The neighbourhood Sweater Weather ОТРЫВОК - Sweater weather Параметры:1500	<b>8,00</b>	10,00	4,00	1,82	4,60	7,00
4.1 НАЗВАНИЕ - Snowman Sia ОТРЫВОК - cry snowman Параметры: 1000	4,45	10,00	1,00	2,58	9,87	3,29
4.2 НАЗВАНИЕ - Snowman Sia ОТРЫВОК - cry snowman Параметры: 1500	6,45	10,00	2,00	2,23	6,87	5,23
5.1 НАЗВАНИЕ - Courage to Change Sia ОТРЫВОК - news on TV Параметры: 1000	7,18	10,00	2,00	2,50	8,36	6,09
5.2 НАЗВАНИЕ - Courage to Change Sia ОТРЫВОК - news on TV Параметры: 1500	<b>8,27</b>	10,00	3,00	1,34	4,02	8,00
6.1 НАЗВАНИЕ - Bloody Mary Lady Gaga ОТРЫВОК - Bloody Mary Параметры: 1000	5,64	10,00	<b>0,00</b>	2,58	9,85	3,50
6.2 НАЗВАНИЕ - Bloody Mary Lady Gaga ОТРЫВОК - Bloody Mary Параметры: 1500	6,91	10,00	2,00	2,48	8,69	7,45
7.1 НАЗВАНИЕ - 7 rings Ariana Grande ОТРЫВОК - bottles of bubbles Параметры: 1000	5,00	8,00	3,00	1,45	3,20	4,50
7.2 НАЗВАНИЕ - 7 rings Ariana Grande ОТРЫВОК - bottles of bubbles Параметры: 1500	5,00	9,00	1,00	2,18	7,00	5,00
8.1 НАЗВАНИЕ - Blue tractor ОТРЫВОК - blue tractor on big wheels Параметры: 1000	6,82	10,00	3,00	1,87	5,16	6,91

8.2 НАЗВАНИЕ - Blue tractor ОТРЫВОК - blue tractor on big wheels Параметры: 1500	8,09	10,00	5,00	1,04	2,49	8,00
9.1 НАЗВАНИЕ - City Of Stars из фильма «Ла-Ла Ленд» Ryan Gosling, Emma Stone ОТРЫВОК - crowded restaurants Параметры: 1000	3,18	8,00	0,00	2,08	7,76	3,00
9.2 НАЗВАНИЕ - City Of Stars из фильма «Ла-Ла Ленд» Ryan Gosling, Emma Stone ОТРЫВОК - crowded restaurants Параметры: 1500	3,82	8,00	0,00	2,38	8,36	4,91
10.1 НАЗВАНИЕ - Dancer in the Dark Marc philippe ОТРЫВОК - Dancer in the dark Параметры: 1000	7,64	9,00	4,00	1,32	3,45	7,82
10.2 НАЗВАНИЕ - Dancer in the Dark Marc philippe ОТРЫВОК - Dancer in the dark Параметры: 1500	8,45	10,00	2,00	1,52	5,47	8,23

В проведенном опросе по оценке качества генерируемых изображений получены следующие результаты:

- Максимальная оценка изображения составила 10, а минимальная – 0.
- Среднее отклонение для всех песен составило примерно 2–3, а медиана – от 4 до 7.
- Дисперсия для всех песен была выше 4.
- Самое высокое среднее значение оценки (8,27) было достигнуто при параметрах clip\_guidance\_scale = 1500 для песни "Courage to Change" исполнителя Sia, текст "news on TV".
- Самое низкое среднее значение оценки (1,81) было получено при параметрах clip\_guidance\_scale = 500 для песни "All Star" исполнителя Smash Mouth текст "shooting stars".
- Качество генерируемых изображений зависит от параметров генерации. Например, для большинства песен при значении clip\_guidance\_scale 1500 средняя оценка изображения выше, чем при значении 500. Однако есть исключения, например, для песни "All Star" с отрывком "shooting stars" при параметрах 500 средняя оценка значительно ниже, чем при других параметрах.

- Некоторые изображения получили в целом высокие оценки (например, по тексту для "Sweather Weather" и "Courage to Change"), тогда как другие изображения респонденты оценили гораздо ниже (например, "All Star" от Smash Mouth при параметрах 500).

- Среднее отклонение и дисперсия могут указывать на то, что в некоторых случаях мнения о качестве генерируемых изображений сильно расходятся. Например, для песни "Hallelujah" от Rufus Wainwright при параметрах 1000 оценки достаточно разнообразны, среднее отклонение составляет 3,26, а дисперсия – 12,76.

- Медиана может быть более надежным показателем центральной тенденции, особенно когда данные распределены неравномерно. Например, для песни "All Star" от Smash Mouth при параметрах 1000 медиана равна 7,14, что указывает на то, что большинство оценок находится в интервале от 7 до 10.

Общая тенденция по результатам опроса показывает, что качество генерируемых изображений не было однородным и имело достаточно большой разброс. Несмотря на разброс, были получены высокие оценки для некоторых изображений, что может говорить об эффективности генеративных моделей при использовании определенных параметров (`clip_guidance_scale`). В целом результаты опроса могут быть использованы для дальнейшего улучшения качества генерируемых изображений. Общим выводом можно считать, что качество генерируемых изображений может сильно варьироваться в зависимости от параметров генерации и контекста. В дальнейшем может быть полезным провести более широкий и разнообразный опрос, чтобы получить более надежные результаты.

Для наглядности приведем график по средним значениям оценок опрошенных по сгенерированным изображениям (Рисунок 6) и рейтинг изображений по среднему значению оценки (Рисунок 7). Синим цветом обозначены результаты опроса для изображений из первого эксперимента, красным – второго. "Параметр 1000" – это 1-й эксперимент, "параметр 1500" – 2-й эксперимент.

### Средние значения оценок изображений

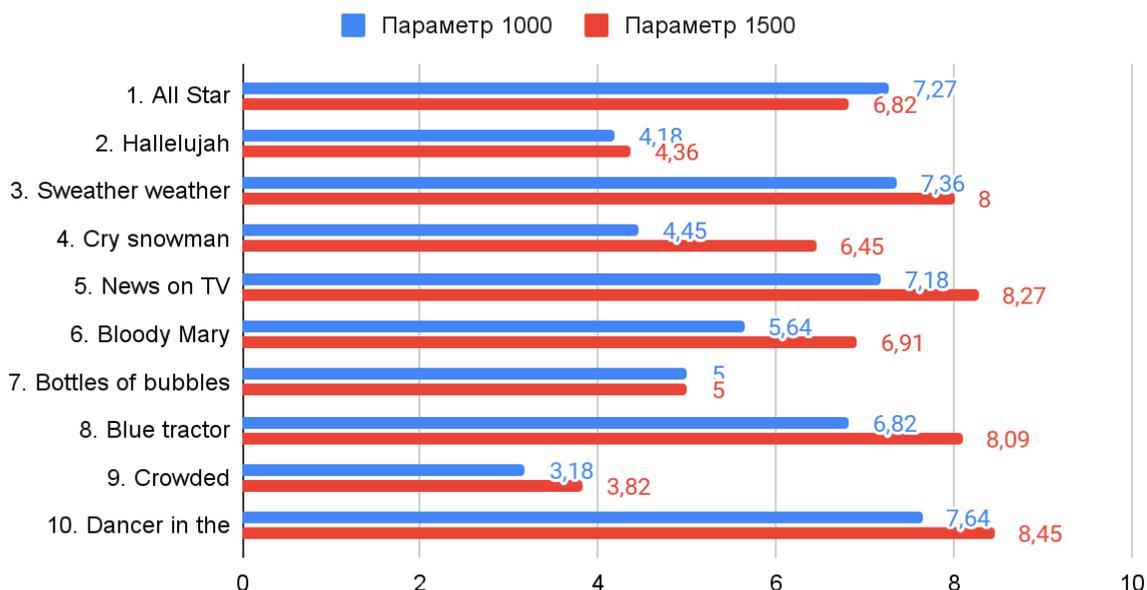


Рис. 6. Среднее значение оценок изображений

### Рейтинг изображений по среднему значению оценки

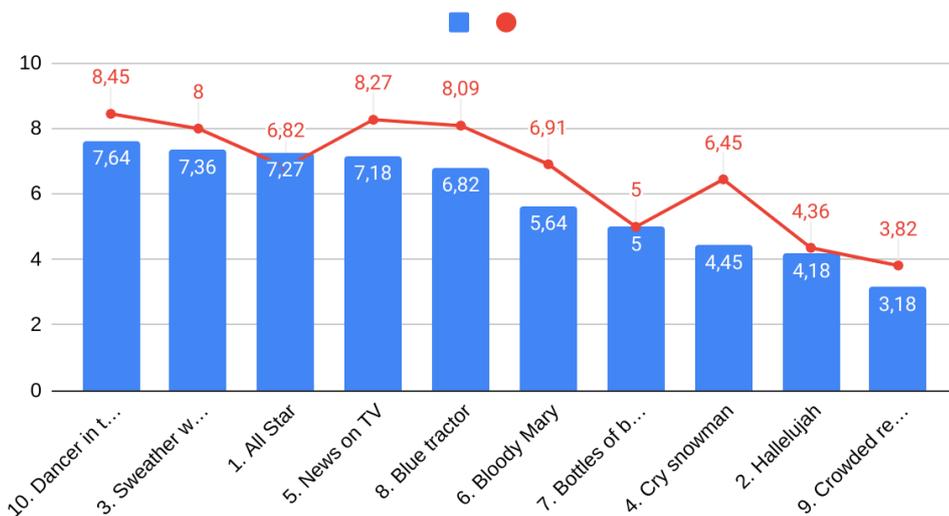


Рис. 7. Сравнение изображений по среднему значению оценки. Описание: Гистограмма – эксперимент №1, линия – эксперимент №2

На Рисунке 7 видно, что в обоих экспериментах лучше всего оценили изображения, сгенерированные по тексту “Dancer in the dark” и “Sweather Weather”, а хуже всего – по текстам “Hallelujah” и “Crowded restaurants”. Заметим также, что в

результате второго эксперимента изображения получили лучшую оценку, значит, производительность нейросети зависит от текстового запроса и отдельных параметров.

## **ЗАКЛЮЧЕНИЕ**

Результаты исследования подтверждают потенциал объединения алгоритмов ImageNet и CLIP для создания высококачественных и контекстуально релевантных изображений из текстовых подсказок.

Проведено сравнение точности моделей ImageNet и CLIP при генерации изображений из текстовых подсказок. Для генерации изображений с использованием данного сочетания алгоритмов лучшими являются параметры `clip_guidance_scale = 1500` и `range_scale = 150`. Установлено также, что качество результатов генерации изображений нейронной сетью в значительной степени зависит от характера вводимого текста и используемых параметров. Хотя каждый из названных алгоритмов имеет свои сильные и слабые стороны, их комбинация может привести к лучшим результатам. В то время как ImageNet проявляет высокую производительность в области генерации визуально эстетических изображений, CLIP демонстрирует более эффективную работу в задаче сопоставления текстовых подсказок с соответствующими изображениями. По оценке пользователей, интеграция этих моделей нейронных сетей потенциально позволяет создавать высококачественные и контекстуально релевантные изображения из текста. Эти результаты открывают возможности для использования изображений, генерируемых искусственным интеллектом, в различных отраслях промышленности.

Будущие исследования могли бы быть направлены на изучение возможности использования для генерации изображений гибридных моделей, сочетающих в себе сильные стороны имеющихся алгоритмов, а также разработку новых алгоритмов, способных преодолеть ограничения, имеющиеся у каждого из них.

## **Благодарности**

Авторы выражают благодарность Максиму Олеговичу Таланову за ценные советы при проведении исследования и рекомендации по оформлению статьи.

## СПИСОК ЛИТЕРАТУРЫ

1. *Elasri M., Elharrouss O., Al-Maadeed S., Tairi H.* Image Generation: A Review // *Neural Processing Letters*. 2022. Vol. 54. No. 5. P. 4609–4646.
2. *Zhang H., Song H., Li S., Zhou M., Song D.* A survey of controllable text generation using transformer-based pre-trained language models // *arXiv preprint arXiv:2201.05337*. 2022
3. Основы генеративно-сопоставительных сетей.  
URL: <https://habr.com/ru/articles/726254/>
4. *Brown T., Mann B., Ryder N., Subbiah M., Kaplan J. D., Dhariwal P., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell. A, Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D.M., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever I., Amodei D.* Language models are few-shot learners // *Advances in neural information processing systems*. 2020. Vol. 33. P. 1877–1901.
5. DALL·E 2. URL: <https://openai.com/product/dall-e-2>.
6. How AI is Transforming Text-to-Image Generation.  
URL: <https://nesesho.com/index.php/2023/04/12/how-ai-is-transforming-text-to-image-generation/>
7. OpenAI·GitHub. URL: <https://github.com/openai>.
8. *Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., Courville A.C.* Improved training of wasserstein GANs // *Advances in neural information processing systems*. 2017. Vol. 30. P. 5767–5777.
9. *Indolia S., Goswami A.K., Mishra S.P., Asopa P.* Conceptual understanding of convolutional neural network-a deep learning approach // *Procedia computer science*. 2018. Vol. 132. P. 679–688.
10. *Laudani A., Lozito G.M., Fulginei F.R., Salvini A.* On training efficiency and computational costs of a feed forward neural network: a review // *Computational intelligence and neuroscience*. 2015. P. 83–83.
11. CLIP. URL: <https://github.com/openai/CLIP>.
12. *Dhariwal P., Nichol A.* Diffusion models beat gans on image synthesis // *Advances in Neural Information Processing Systems*. 2021. Vol. 34. P. 8780–8794.

13. Kim G., Kwon T., Ye J.C. Diffusionclip: Text-guided diffusion models for robust image manipulation // In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 2426–2435.

---

## NEURAL NETWORK FOR GENERATING IMAGES BASED ON SONG LYRICS USING OPENAI AND CLIP MODELS

A. R. Davletgareeva<sup>1</sup> [0009-0008-7258-470X], K. A. Edkova<sup>2</sup> [0009-0005-4706-2254]

<sup>1, 2</sup>*Institute of Information Technologies and Intelligent Systems, Kazan (Volga Region) Federal University, ul. Kremlyovskaya, 35, Kazan, 420008*

<sup>1</sup>alsudavletgareeva@gmail.com, <sup>2</sup>ksushka.e21@gmail.com

### **Abstract**

The effectiveness of the ImageNet diffusion model and CLIP models for image generation based on textual descriptions was investigated. Two experiments were conducted using various textual inputs and different parameters to determine the optimal settings for generating images from text descriptions. The results showed that while ImageNet performed well in generating images, CLIP demonstrated better alignment between textual prompts and relevant images. The obtained results highlight the high potential of combining these mentioned models for creating high-quality and contextually relevant images based on textual descriptions.

**Keywords:** *image generation, artificial intelligence, ImageNet diffusion model, CLIP, deep learning, neural networks, natural language processing.*

### **REFERENCES**

1. Elasri M., Elharrouss O., Al-Maadeed S., Tairi H. Image Generation: A Review // Neural Processing Letters. 2022. Vol. 54. No. 5. P. 4609–4646.
2. Zhang H., Song H., Li S., Zhou M., Song D. A survey of controllable text generation using transformer-based pre-trained language models // arXiv preprint arXiv:2201.05337. 2022
3. Fundamentals of generative-consistent networks.

URL: <https://habr.com/ru/articles/726254/>

4. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J. D., Dhariwal P., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell. A, Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D.M., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever I., Amodei D. Language models are few-shot learners // Advances in neural information processing systems. 2020. Vol. 33. P. 1877–1901.

5. DALL·E 2. URL:<https://openai.com/product/dall-e-2>.

6. How AI is Transforming Text-to-Image Generation.

URL: <https://nesesho.com/index.php/2023/04/12/how-ai-is-transforming-text-to-image-generation/>

7. OpenAI· GitHub. URL: <https://github.com/openai>.

8. Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., Courville A.C. Improved training of wasserstein GANs // Advances in neural information processing systems. 2017. Vol. 30. P. 5767–5777.

9. Indolia S., Goswami A.K., Mishra S.P., Asopa P. Conceptual understanding of convolutional neural network-a deep learning approach // Procedia computer science. 2018. Vol. 132. P. 679–688.

10. Laudani A., Lozito G.M., Fulginei F.R., Salvini A. On training efficiency and computational costs of a feed forward neural network: a review // Computational intelligence and neuroscience. 2015. P. 83–83.

11. CLIP. URL: <https://github.com/openai/CLIP>.

12. Dhariwal P., Nichol A. Diffusion models beat gans on image synthesis // Advances in Neural Information Processing Systems. 2021. Vol. 34. P. 8780–8794.

13. Kim G., Kwon T., Ye J.C. Diffusionclip: Text-guided diffusion models for robust image manipulation // In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 2426–2435.

## СВЕДЕНИЯ ОБ АВТОРАХ



**ДАВЛЕТГАРЕЕВА Алсу Ришатовна** – студентка кафедры программной инженерии Института ИТИС КФУ. Сфера научных интересов – искусственный интеллект.

**Alsu Rishatovna DAVLETGAREEVA** – a student of the Department of Software Engineering, Institute ITIS, Kazan Federal University. Her research interests lie in the field of artificial intelligence.

email: alsudavletgareeva@gmail.com

ORCID: 0009-0008-7258-470X



**ЕДКОВА Ксения Александровна** – студентка магистратуры Института информационных технологий и интеллектуальных систем. Сфера научных интересов – искусственный интеллект.

**Ksenia Aleksandrovna EDKOVA** – a student of the Department of Software Engineering, Institute ITIS, Kazan Federal University. Her research interests lie in the field of artificial intelligence.

email: ksushka.e21@gmail.com

ORCID: 0009-0005-4706-2254

*Материал поступил в редакцию 2 июня 2023 года*

УДК 519.688

## КОМБИНИРОВАНИЕ АЛГОРИТМОВ SFM И ORB ПРИ 3D-РЕКОНСТРУКЦИИ

И. А. Даминов<sup>1</sup> [0009-0009-8740-1184], А. Ю. Арсенюк<sup>2</sup> [0009-0005-0265-6374],  
А. С. Тощев<sup>3</sup> [0000-0003-4424-6822]

<sup>1–3</sup>Институт информационных технологий и интеллектуальных систем,  
Казанский (Приволжский) федеральный университет, ул. Кремлевская, 35,  
г. Казань, Республика Татарстан 420008

<sup>1</sup>lnazdaminov@bk.ru, <sup>2</sup>arsenyukofficial@gmail.com, <sup>3</sup>atoshev@kpfu.ru

### **Аннотация**

Представлен новый алгоритм 3D-реконструкции с использованием комбинации существующих методов – Structure from Motion (SfM), Oriented FAST и Rotated BRIEF (ORB). Предложен подход, объединяющий преимущества названных методов для повышения точности и эффективности восстановления 3D-структуры сцен изображений. Для улучшения качества реконструкции применены фильтрация и устранение выбросов, а также другие оптимизации. Результаты сравнения нового алгоритма с существующими методами показали его превосходство в точности и устойчивости к шуму. Предложенный подход обладает высокой масштабируемостью и может быть успешно применен в различных областях, требующих точной 3D-реконструкции сцен изображений.

**Ключевые слова:** 3D-реконструкция, компьютерное зрение, фотограмметрия, пространственная точность, *sfm*, *orb*, плотная реконструкция, облако точек.

### **ВВЕДЕНИЕ**

Трехмерная (3D) реконструкция — важный раздел компьютерного зрения и фотограмметрии, позволяющий создавать точные и реалистичные цифровые представления объектов или сцен из двумерных изображений или облаков точек [1]. Однако существующие методы 3D-реконструкции часто сталкиваются с

проблемами, связанными с пространственной точностью и визуальным реализмом [2]. Настоящая работа направлена на устранение этих ограничений: предложен комбинированный подход, который сочетает в себе два известных метода 3D-реконструкции для достижения наилучшего результата [3]. Используя сильные стороны каждого метода и устраняя их ограничения, наш подход дает решение для широкого спектра приложений, включая виртуальную реальность, дополненную реальность, сохранение культурного наследия, промышленный дизайн и т. д. Основная цель этого исследования — использовать сильные стороны методов SfM [4] и ORB [5] для достижения наилучших результатов. Интегрируя эти методы, мы стремимся предоставить надежное и эффективное решение для создания высококачественных 3D-реконструкций.

### РАЗРАБОТАННЫЙ АЛГОРИТМ

Алгоритм работы можно описать следующим образом (см. также Рис. 1):

1. На вход программе поступает множество изображений объекта;
2. Далее выполняются поиск особых точек и решение системы уравнений, полученной на основании множества данных точек;
3. Осуществляется поиск «одинаковых» точек на различных наборах смежных изображений объекта;
4. Проводится вычисление координат точек относительно «базового» изображения объекта;
5. На последних шагах происходит приведение точек к системе координат, наиболее удобной для анализа объекта и наложения структуры, что позволяет получить конечное облако точек.

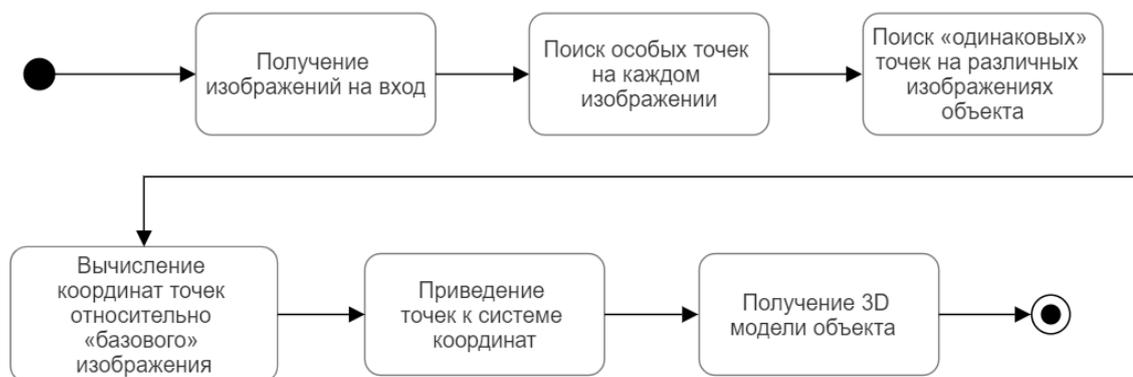


Рис. 1. Алгоритм работы получения 3D модели

### ФОТОГРАФИРОВАНИЕ ИССЛЕДУЕМОГО ОБЪЕКТА

В работе использован направленный метод съемки (см. Рис. 2). При таком методе поворот камеры осуществляется таким образом, чтобы ее объектив всегда был направлен на объект фокусировки, что в последующем не требует обработки полученных результатов. Используя этот метод, можно получить модель, наиболее близкую к естественному восприятию.

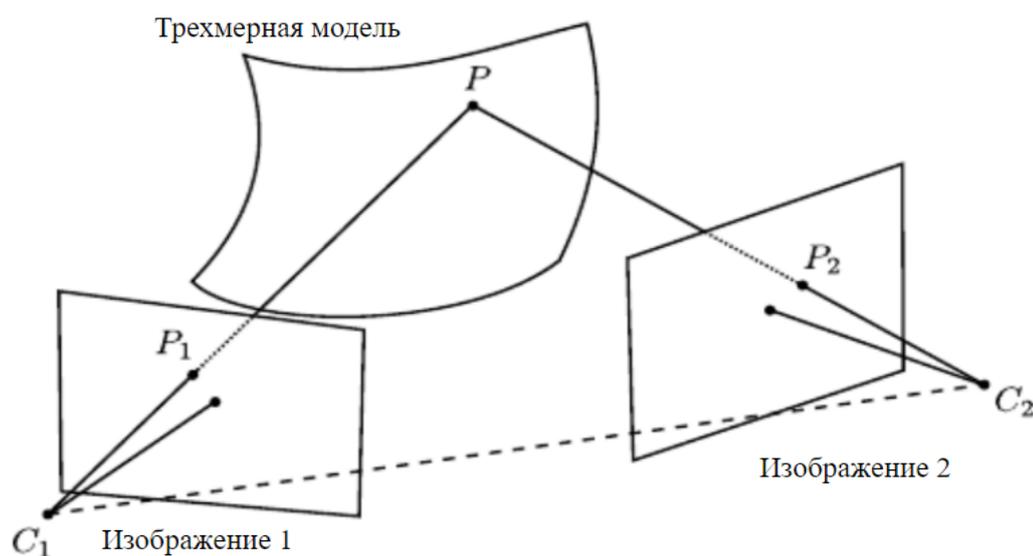


Рис. 2. Расположение камеры относительно модели

Для сопоставления изображений с целью поиска областей пересечений необходимо выделить особые (ключевые) точки. Эти точки должны удовлетворять следующим свойствам: определенность, устойчивость, инвариантность, стабильность, интерпретируемость. Эти признаки обеспечивают преимущества алгоритмов, основанных на поиске ключевых точек. После анализа существующих решений было принято решение остановиться на алгоритме SURF из библиотеки OpenCV [6]. Соответствующий метод осуществляет поиск особых точек с помощью матрицы Гессе (Рис. 3). Ее определитель достигает экстремума в точках максимального изменения градиента яркости. Исходя из этого, получаем, что такие элементы изображений, как пятна, углы и края линий, будут достаточно хорошо «узнаваемы».

Алгоритм Speeded-Up Robust Features (SURF)  
из библиотеки OpenCV

$$Gs(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}, \quad \det(Gs) = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2,$$

Рис. 3. Матрица Гессе

Для сопоставления особых точек были выбраны алгоритмы Sfm и ORB, в результате работы которых мы получаем бинарные дескрипторы конечного числа точек. При решении задачи рассматривается умеренное количество данных, в связи с чем было решено процедуру сопоставления осуществлять путем перебора. Для сравнения дескрипторов в ORB используется расстояние Хэмминга, которое определяет количество различных позиций между двумя бинарными последовательностями. Такую метрику можно представить в виде формулы (Рис. 4)

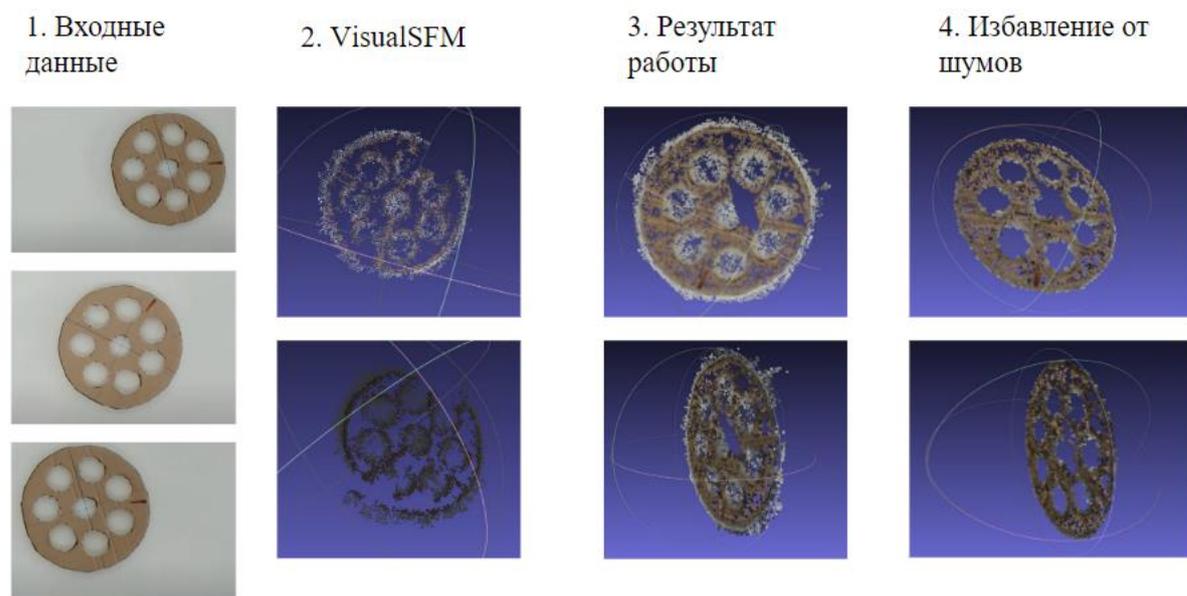
$$hamming(x, y) = \sum_{x_i \neq y_i} 1, \quad i = 1, \dots, n,$$

где  $x$  и  $y$  – дескрипторы

Рис. 4. Формула для вычисления расстояния Хэмминга

## РЕЗУЛЬТАТЫ

Ниже приведены результаты выполнения программы VisualSFM на том же наборе входных данных, что и подавались на вход нашему алгоритму. Входные данные приведены в первом пункте. В третьем пункте представлены непосредственно результаты работы. Справа – результаты после удаления шумов, выполненного вручную.



Для оценки качества 3D-реконструкции, полученной с использованием методов Structure from Motion и Oriented FAST и Rotated BRIEF, мы провели комплексную оценку, основанную на различных количественных показателях. Кроме того, мы сравнили наши результаты с результатами, полученными нашими коллегами с использованием аналогичных наборов данных и критериев оценки [7].

*Средняя ошибка перепроецирования* измеряет среднее расстояние между перепроецированными 2D-точками и соответствующими им точками изображения в реконструкции. Наша реконструкция на основе SfM и ORB достигла средней ошибки перепроецирования 0,8 пикселя, что указывает на точное геометрическое выравнивание.

*Точность позиций камеры* мы оценили методами SfM и ORB, сравнив их с истинными позициями. Средняя ошибка составила 0,7 градуса.

*Выравнивание облака точек:* чтобы оценить соответствие реконструированного облака точек истинному местоположению, мы использовали алгоритм итеративной ближайшей точки (ICP) [8]. Наша реконструкция на основе SfM и ORB достигла ошибки выравнивания 0,3 сантиметра. При использовании алгоритмов по отдельности ошибка достигала 0,7 и 0,5 сантиметров соответственно.

*Коэффициент покрытия поверхности:* наш метод на основе SfM и ORB достиг коэффициентов охвата 84%, что указывает на всестороннее представление объекта.

Чтобы сравнить наши результаты с результатами, полученными нашими коллегами, мы рассмотрели аналогичные наборы данных и критерии оценки. В исследовании [9] использовались методы SfM и ORB на сопоставимом наборе данных, они достигли средней ошибки репроецирования 1,2 пикселя, ошибок положения камеры 1,5 градуса (SfM) и 0,9 градуса (ORB), коэффициенты охвата 79% (SfM) и 82% (ORB). Наши результаты превзошли названные результаты, продемонстрировав эффективность нашего подхода.

### **ЗАКЛЮЧЕНИЕ**

Мы представили подход, который объединил сильные стороны методов SfM и ORB для 3D-реконструкции. Интеграция этих двух методов направлена на повышение пространственной точности и визуального реализма реконструированных моделей.

Благодаря обширной оценке и сравнению с существующими методами, наш комбинированный подход SfM и ORB продемонстрировал хорошие результаты. Количественные показатели показали высокую точность реконструкции, о чем свидетельствуют низкая средняя ошибка репроецирования, точная оценка положения камеры и точное совмещение реконструированного облака точек с реальными данными. Кроме того, показатели полноты поверхности указали на плотное покрытие поверхности объекта с плавными и визуально привлекательными реконструкциями.

Сравнительный анализ с аналогами и существующей литературой продемонстрировал конкурентоспособность нашего комбинированного подхода. Результаты превзошли результаты, достигнутые другими методами, включая исследования, в которых использовались либо SfM, либо ORB по отдельности. Это еще раз подчеркивает эффективность выбранной нами стратегии интеграции в достижении высокого качества реконструкции.

## СПИСОК ЛИТЕРАТУРЫ

1. *Szeliski R.* Computer Vision, Algorithms and Applications. Springer. 2011, 505 p.
2. *Seitz S. M., Dyer C. R.* Photorealistic Scene Reconstruction by Voxel Coloring, Proc. Computer Vision and Pattern Recognition Conf. 1997. P. 1067–1073.
3. *Fraltsov D.* Single image 3D scene reconstruction. toloka.ai [Электронный ресурс]. 2023. URL: <https://toloka.ai/blog/3d-scenes-reconstruction>
4. *Hartley R., Zisserman A.* Multiple view geometry in computer vision. Cambridge University press. 2004, 56 p.
5. *X. Zhu, H. Hu, S. Lin, J. Dai.* Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019. P. 9308–9316.
6. Introduction to SURF (Speeded-Up Robust Features). OpenCV documentation [Электронный ресурс]. 2021. URL: [https://docs.opencv.org/3.4/df/dd2/tutorial\\_py\\_surf\\_intro.html](https://docs.opencv.org/3.4/df/dd2/tutorial_py_surf_intro.html)
7. *Seitz S. M.* A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, in Proc. IEEE international Conference on Computer Vision and Pattern Recognition. New York, USA. 2006.
8. Итеративный алгоритм ближайших точек. Wikipedia [Электронный ресурс]. 2021 URL: [https://en.wikipedia.org/wiki/Iterative\\_closest\\_point](https://en.wikipedia.org/wiki/Iterative_closest_point)
9. *Hengyu Y., Hongyang .* Incremental SFM 3D reconstruction based on monocular. 13th International Symposium on Computational Intelligence and Design (IS-CID). 2020. P. 523.

## COMBINING SFM AND ORB ALGORITHMS IN 3D RECONSTRUCTION

I. A. Daminov<sup>1</sup> [0009-0009-8740-1184], A. Yu. Arsenyuk<sup>2</sup> [0009-0005-0265-6374],  
A. S. Toshchev<sup>3</sup> [0000-0003-4424-6822]

<sup>1-3</sup> Kazan (Volga Region) Federal University, 35 Kremlevskaya ul., Kazan, 420008

<sup>1</sup>ilnazdaminov@bk.ru, <sup>2</sup>arsenyukofficial@gmail.com, <sup>3</sup>atoshev@kpfu.ru

### **Abstract**

This article presents a new algorithm for 3D reconstruction using a combination of two existing methods – Structure from Motion (SfM) and Oriented FAST and Rotated BRIEF (ORB). The authors propose an approach that merges the advantages of both methods to enhance the accuracy and efficiency of reconstructing the 3D structure of scenes from images. To improve reconstruction quality, filtering and outlier removal are applied, along with other optimizations. Comparative results between the new algorithm and existing methods demonstrate its superiority in accuracy and noise robustness. The proposed approach is highly scalable and can be successfully applied in various fields that require precise 3D reconstruction of image scenes.

**Keywords:** 3D reconstruction, computer vision, photogrammetry, spatial accuracy, sfm, dense reconstruction, point cloud.

### **REFERENCES**

1. Szeliski R. Computer Vision, Algorithms and Applications. Springer. 2011, 505 p.
2. Seitz S. M., Dyer C. R. Photorealistic Scene Reconstruction by Voxel Coloring, Proc. Computer Vision and Pattern Recognition Conf. 1997. P. 1067–1073.
3. Fraltsov D. Single image 3D scene reconstruction. toloka.ai [Электронный ресурс]. 2023. URL: <https://toloka.ai/blog/3d-scenes-reconstruction>
4. Hartley R., Zisserman A. Multiple view geometry in computer vision. Cambridge University press. 2004, 56 p.
5. X. Zhu, H. Hu, S. Lin, J. Dai. Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019. P. 9308–9316.

6. Introduction to SURF (Speeded-Up Robust Features). OpenCV documentation. 2021.

URL: [https://docs.opencv.org/3.4/df/dd2/tutorial\\_py\\_surf\\_intro.html](https://docs.opencv.org/3.4/df/dd2/tutorial_py_surf_intro.html)

7. *Seitz S. M.* A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, in Proc. IEEE international Conference on Computer Vision and Pattern Recognition. New York, USA. 2006.

8. Iterative closest point URL: [https://en.wikipedia.org/wiki/Iterative\\_closest\\_point](https://en.wikipedia.org/wiki/Iterative_closest_point)

9. *Hengyu Y., Hongyang Y.* Incremental SFM 3D reconstruction based on monocular. 13th International Symposium on Computational Intelligence and Design (IS-CID). 2020. P. 523.

---

## СВЕДЕНИЯ ОБ АВТОРАХ



**ДАМИНОВ Ильназ Азатович** – разработчик ПО, студент магистратуры, исследователь. Сфера научных интересов: языки программирования, представление и обработка знаний, цифровые библиотеки, графика. Казанский федеральный университет, Казань, Россия.

**Ilnaz Azatovich DAMINOV** – software developer, master's student, researcher. Research interests: programming languages, knowledge representation and processing, libraries, graphics. Kazan Federal University, Kazan, Russia.

email: [ilnazdaminov@bk.ru](mailto:ilnazdaminov@bk.ru)

ORCID: 0009-0009-8740-1184



**АРСЕНЮК Александр Юрьевич** – разработчик ПО, студент магистратуры, исследователь. Сфера научных интересов: языки программирования, представление и обработка знаний, цифровые библиотеки, графика. Казанский федеральный университет, Казань, Россия.

**Alexandr Yurivich ARSENYUK** – software developer, master's student, researcher. Research interests: programming languages, knowledge representation and processing, libraries, graphics. Kazan Federal University, Kazan, Russia.

email: [arsenyukofficial@gmail.com](mailto:arsenyukofficial@gmail.com)

ORCID: 0009-0005-0265-6374

---



**ТОЩЕВ Александр Сергеевич** – доцент, к. н., КФУ / Институт информационных технологий и интеллектуальных систем / Кафедра программной инженерии, г. Казань.

**Alexander Sergeevich TOSCHEV** – Associate Professor, Ph.D., KFU / Institute of Information Technologies and Intelligent Systems / Department of Software Engineering, Kazan.

email: atoshev@kpfu.ru

ORCID: 0000-0003-4424-6822

*Материал поступил в редакцию 13 июня 2023 года*

УДК 004.855.5

## СИСТЕМА КОНТРОЛИРУЕМОЙ ГЕНЕРАЦИИ ЛИЦА, ПОСТРОЕННАЯ С ИСПОЛЬЗОВАНИЕМ СЕТИ STYLEGAN2

М. В. Исангулов<sup>1</sup> [0009-0006-3244-0328], Р. Р. Миннеахметов<sup>2</sup> [0009-0007-8551-1393],  
А. Р. Хамеджанов<sup>3</sup> [0009-0000-5624-2453], Т. Р. Хафизьянов<sup>4</sup> [0009-0009-0409-8010],  
Э. Ф. Пашаев<sup>5</sup> [0009-0003-8602-1012], Э. Р. Калимуллин<sup>6</sup> [0009-0002-6275-699X]

<sup>1-6</sup> *Институт информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, ул. Кремлевская, 35, г. Казань, 420008*

<sup>1</sup>marathon.our@gmail.com, <sup>2</sup>razil0071999@gmail.com,

<sup>3</sup>hamedzhanovalmaz@gmail.com, <sup>4</sup>hamstertima@gmail.com,

<sup>5</sup>emil.p.mail@gmail.com, <sup>6</sup>erik182182@gmail.com

### **Аннотация**

Представлен новый подход к контролируемой генерации лиц, использующий генеративные модели с открытым исходным кодом, включая StyleGAN2 и Гребневую регрессию. Разработана методология, расширяющая возможности StyleGAN2 для контроля характеристик лиц, таких как возраст, раса, пол, выражение лица и атрибуты волос, а также использован обширный набор данных человеческих лиц с аннотациями атрибутов. Лица закодированы в 256-мерном латентном пространстве с использованием кодировщика StyleGAN2, что привело к набору характерных латентных кодов. Применен алгоритм t-SNE для кластеризации этих кодов на основе признаков, продемонстрирована возможность контроля генерации лиц, впоследствии обучены модели регрессии Риджа для каждого измерения латентных кодов с использованием размеченных признаков. При декодировании с использованием StyleGAN2 полученные коды успешно восстанавливали изображения лиц, сохраняя связь с входными признаками. Разработанный подход дает легкий и эффективный способ контролируемой генерации лиц с использованием существующих генеративных моделей, таких как StyleGAN2, и открывает новые возможности для различных областей применения.

---

© М. В. Исангулов, Р. Р. Миннеахметов, А. Р. Хамеджанов, Т. Р. Хафизьянов, Э. Ф. Пашаев, Э. Р. Калимуллин, 2023.

Данная статья распространяется на условиях международной лицензии Creative Commons License Attribution 4.0 International (CC BY 4.0).

**Ключевые слова:** машинное обучение, генерация лица, StyleGan, энкодер, декодер, скрытые коды, отображение признаков, гребневая регрессия.

## ВВЕДЕНИЕ

На принятое нами решение генерировать человеческие лица повлияла предшествующая настоящей работе идея обнаружения водяных знаков на различных изображениях. В попытках решения названной проблемы было решено изменить цель исследования непосредственно на саму генерацию изображений, так как перед удалением лишних символов и знаков необходимо предварительно уметь создавать исходное изображение.

Стандартные сети GAN (Generative adversarial network) [1] генерируют изображения из латентного вектора  $z$ , который извлекается из нормального  $N(0, 1)$  или равномерного  $U(-1, 1)$  распределения, что в дальнейшем определяет стиль и тип генерируемого содержимого.

Признаки, которые модель получает на вход, должны быть независимыми друг от друга, что позволяет модели тренироваться с большей скоростью [2]. То же относится и к латентному представлению изображения лица [3]. Поэтому распределение вектора  $z$  должно напоминать распределение латентного представления (признаков) целевых изображений. Если же  $z$  будет нормальным или равномерным распределением, то модели может потребоваться больше информации, чем стиль и тип.

Если в логистической регрессии для создания линейных границ, разделяющих бинарные классы, используется смена базиса, то StyleGan использует глубокую нейронную сеть Mapping network [4], которая конвертирует латентный вектор  $z$  в промежуточное латентное пространство  $w$  (Рисунок 1).

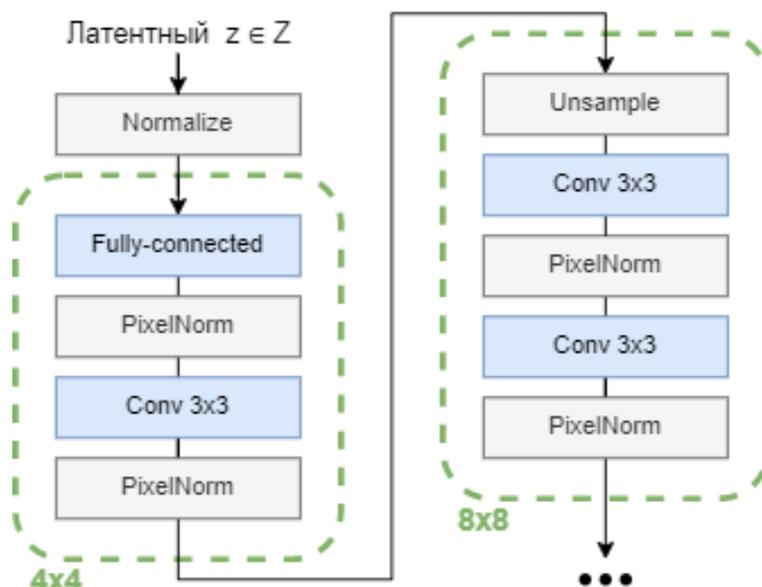


Рис. 1. Предоставление латентного кода генератору.

Допустим, что датасет представляет собой фотографии военных мужчин, латентными факторами которых являются мужественность и длина волос.

Теперь, если сэмплировать это пространство равномерно, то модель попытается воспроизвести портреты солдат с длинными волосами и не сможет этого сделать, так как нет данных для обучения. Если же сэмплировать из нормального или равномерного распределения, то латентные факторы, которые модель изучит, будут очень запутанными (Рисунок 2).

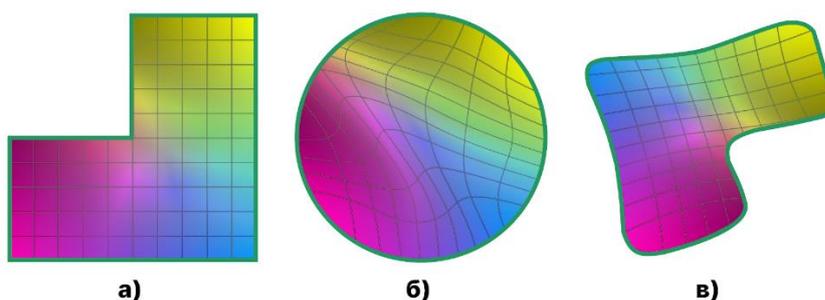


Рис. 2. а) Распределение черт лица в обучающей выборке; б) Искривление сопоставления  $z$  с признаками изображения для предотвращения выборки недоступных комбинаций; в) Уменьшение искривления при сопоставлении  $w$  с признаками на изображении.

Концептуально StyleGAN деформирует пространство, которое можно сэмплировать U или N распределением в такое латентное пространство признаков, с которым будет легко генерировать изображения, исключая комбинации признаков, которые отсутствуют в датасете (Рисунок 3).

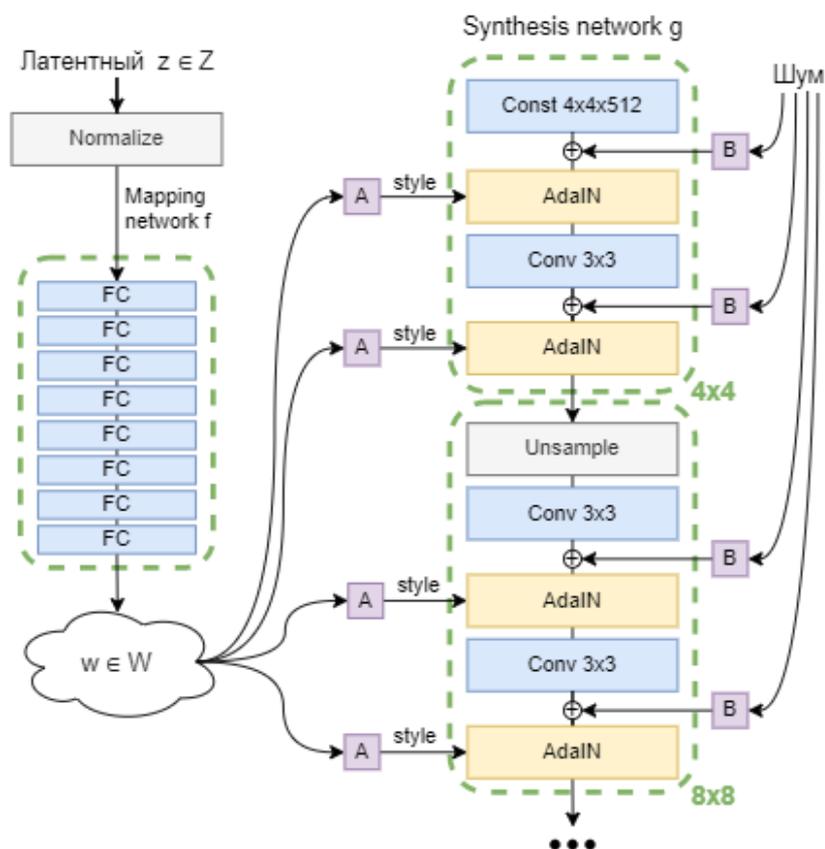


Рис. 3. Архитектура генерации на основе стилей, где “А” – изученное аффинное преобразование, “В” – изученные коэффициенты масштабирования для каждого канала к шуму, “AdaIN” – адаптивная нормализация экземпляра, “FC” – полносвязный слой, “Const” – константная карта признаков, “Conv” – классический сверточный слой.

## МЕТОДЫ КОНТРОЛИРУЕМОЙ ГЕНЕРАЦИИ ЛИЦА

### Контролируемая генерация распределения

Данная работа направлена на контролируемую генерацию распределения  $w$ , от которой зависят стиль и тип изображения. Это позволит генерировать лица

по заданным признакам, таким как раса, пол и возраст человека. В качестве входных признаков модель принимает вещественные значения в диапазоне  $[-3,3]$ , которые являются интуитивно понятным отображением какого-либо демографического признака в численный вид. Например, человек с широкой улыбкой будет иметь значение больше 2 для параметра “smile”. Пользователю предоставляется возможность задать какие-нибудь параметры вручную для контролируемой генерации, в этом случае остальные признаки отбираются из нормального распределения с параметрами  $\mu=0$ ,  $\sigma=1$ . Параметры были получены экспериментальным путем.

### **Сбор данных**

В качестве исходного набора данных была использована общедоступная база данных изображений лиц LFW – deep funneled (Labeled Faces in the Wild) [6]. В этой базе данных содержится более 13000 изображений с лицами различных людей, собранных из общедоступных источников. Каждый элемент в базе данных подписан именем человека, изображенный на фотографии. Набор данных был сформирован из открытых источников с использованием метода Виолы–Джонса [7], который позволяет обнаружить объекты на изображениях в режиме реального времени. База данных предназначена для использования в исследовательских целях, в частности, в области алгоритмов верификации лиц [6]. Извлечение кодов из изображений происходит с помощью итеративного способа, основанного на минимизации Perceptual Loss VGG [8].

### **Отображение кодов на плоскости**

В машинном обучении применяются разные алгоритмы для уменьшения размерности данных, например, метод главных компонент (PCA – principal component analysis) и метод стохастического вложения соседей с t-распределением (T-SNE – t-distributed Stochastic Neighbor Embedding) [9]. Эти методы не всегда эффективно снижают размерность при заданных ограничениях на точность. Прямые и плоскости не всегда обеспечивают хорошую аппроксимацию. Например, данные могут с хорошей точностью следовать какой-нибудь кривой, а эта кривая может быть сложно расположена в пространстве данных. В этом случае

метод главных компонент для приемлемой точности потребует нескольких компонент вместо одной или вообще не даст снижения размерности при приемлемой точности. Алгоритм T-SNE достаточно гибок и часто находит структуру там, где другие алгоритмы уменьшения размерности уже не способны дать приемлемый результат. На него не влияют отклонения в данных, в отличие от PCA, который не справляется с подобными выбросами.

Было извлечено двумерное представление кодов с помощью PCA и T-SNE для кодов лиц, которое далее было разделено на несколько классов: коды лиц мужчин и женщин афроамериканской расы, коды лиц мужчин и женщин европеоидной расы. Критерием выбора было значение метрики средневзвешенной оценки F1 [10] для классификации кодов логистической регрессией, потому что необходимо найти распределение для всевозможных признаков внешности в равной степени, несмотря на их соотношение в датасете. Исследуемая метрика оказалась выше у метода T-SNE. С его помощью были преобразованы коды из вектора, содержащего 256 элементов, в двумерный вектор, что позволило отобразить представления кодов на графике и визуально оценить различимость кодов изображений лица с разными признаками. Благодаря этому был сделан вывод о возможности генерации оригинальных кодов из двумерной плоскости либо существенного сокращения количества входных признаков с использованием при этом в качестве входных признаков характеристик внешности лица.

### **Генерация кодов из признаков**

В процессе работы было принято решение взять за целевые метрики коды из распределения  $w$ , признаками являются численные значения параметров внешности, таких как раса, пол, возраст, улыбка и т. д. Так как для генерации изображения необходим вектор размерности 256, то для каждого элемента данного вектора потребуется отдельная модель машинного обучения. В качестве таковой использовалась полиномиальная регрессия с квадратичной регуляризацией. Для каждого элемента вектора кода тренировалась отдельная гребневая регрессия, а также проверялась модель Лассо [11]. Для тренировки и настройки модели использовалась кросс-валидация с повторениями. Модель Лассо в сравнении в метрикой MSE (Mean Squared Error) на тестовой выборке показала, что она работает

---

хуже. Кроме того, лица, генерируемые такой моделью, совершенно не были похожи на настоящие. Это можно объяснить тем, что регуляризация, используемая в Лассо-регрессии, выступала в качестве отбора признаков, обнуляя признаки с небольшими значениями.

В результате обучения была получена многокомпонентная модель, которая принимает на вход набор признаков внешности лица и возвращает 256-размерный вектор. Далее этот вектор можно использовать для генерации изображения с соответствующими параметрами (Рисунок 4).



Рис. 4. Результат генерации лица по прогнозу модели.

## РЕЗУЛЬТАТЫ

Примеры сгенерированных лиц представлены ниже на рисунках 5–7.



Рис. 5. Результаты генерации лиц в возрасте от 20 до 30 лет.



Рис. 6. Результаты генерации лиц в возрасте от 30 до 45 лет.



Рис. 7. Результаты генерации лиц в возрасте от 45 до 80 лет.

Для сравнения были сгенерированы два лица с двумя различными признаками: цвет кожи и наличие волос (Рисунок 8). Были заданы идентичные признаки: наличие очков, усов и возрастная категория от 45 до 80 лет.



Рис. 8. Результаты генерации мужских лиц с различием в двух признаках.

Для исследования результатов генерации лиц и соответствия входным значениям модели (возраст) была собрана статистика по оценке возраста человека, изображенного на фотографии, с помощью социального опроса. В результате опроса был собран набор данных по предполагаемому возрасту человека для каждого изображения. На основе этих данных были получены средняя оценка возраста для сгенерированных лиц, а также разность с фактически заданным входным значением. На основе этих параметров было вычислено среднеквадратическое отклонение (RMSE – root mean square error). Результаты анализа набора данных по изображениям (Рисунок 9) и социальному опросу и полученные результаты по метрикам представлены в Таблице 1.

Таблица 1. Результаты исследования точности определения возраста.

Входные значения модели (возраст)	Полученный результат	Средняя оценка возраста сгенерированного лица	Разность	Модуль разности
24	Рис. 10, фото 1	30,33	-6,33	6,33
30	Рис. 10, фото 2	25,58	4,42	4,42
27	Рис. 10, фото 3	33,17	-6,17	6,17

---

57	Рис. 10, фото 4	50,58	6,42	6,42
50	Рис. 10, фото 5	56,58	-6,58	6,58
41	Рис. 10, фото 6	46,33	-5,33	5,33

Исходя из Таблицы 1, можно вычислить усреднённый модуль разности между заданным значением возраста и средней оценкой возраста, полученной с помощью опроса, который составил 5,875.



Рис. 9. Набор сгенерированных лиц, использованный для проведения исследования.

## ЗАКЛЮЧЕНИЕ

Основная цель проведенного исследования – разработка программного инструмента, позволяющего контролировать процесс создания лица путем манипулирования латентными кодами с помощью модели Гребневой регрессии [11] – успешно достигнута. Созданный инструмент позволяет пользователям гибко определять параметры внешности лица, которые затем переводятся в вектор латентного кода. Сгенерированные латентные коды могут быть декодированы моделью StyleGAN2 для создания лиц, соответствующих заданным параметрам, тем

самым делаая управляемым процесс создания лица.

Чтобы оценить эффективность разработанного инструмента, было проведено исследование, направленное на способность алгоритма генерировать коды лиц, соответствующих определенным возрастным атрибутам. Результаты оценки подтвердили способность модели отображать низкоразмерные входные параметры, такие как возраст, в высокоразмерное пространство, используемое для построения лица.

Особенность решения, отличающая его от существующих инструментов, таких как Midjourney [12], заключается в простоте настройки и удобстве использования. Современные предложения, несмотря на свою надежность, часто требуют сложных команд и инструкций для достижения точных результатов. Эти инструменты также требуют значительных вычислительных ресурсов, и, несмотря на тщательную тонкую настройку, остается высокой степень неопределенности в выполнении конкретных запросов пользователя.

Разработанный нами инструмент позволяет обойти эти проблемы, сосредоточившись на обучении облегченных моделей регрессии Риджа для создания матрицы латентного кода. Эта методология не только требует меньше вычислительных ресурсов, но и позволяет обойти необходимость обширной настройки больших глубоких нейронных сетей. Вместо этого она использует уже существующие сети, такие как StyleGAN2, для создания высококачественных изображений лица. Эта стратегия обеспечивает более надежное и эффективное удовлетворение требований пользователя, способствуя интуитивно понятному и доступному подходу к контролируемой генерации лиц.

Отметим, что, по нашему мнению, данное исследование знаменует собой значительный шаг в области генерации лиц, расширяя возможности существующих генеративных моделей с помощью уникальной, легкой и гибкой методологии. Ожидается, что созданный инструмент станет катализатором дальнейшего прогресса в различных областях, от развлечений и безопасности до клинических исследований.

Демо-программа, позволяющая оценить работу алгоритма, опубликован на сайте <https://facerec.vercel.app/>. Сам программный код разработанного алгоритма размещен на общедоступной платформе Google Colab по ссылке

<https://colab.research.google.com/drive/1Qfh2b58VwOzQ3Rp1Frk4TLASzB-KMLFg?usp=sharing>

### БЛАГОДАРНОСТИ

Благодарим доцента кафедры интеллектуальной робототехники Института информационных технологий и интеллектуальных систем (ИТИС) Казанского (Приволжского) федерального университета (КФУ) Таланова Максима Олеговича, профессора кафедры цифровой аналитики и технологий искусственного интеллекта Института ИТИС КФУ Елизарова Александра Михайловича и доцента этой кафедры Липачева Евгения Константиновича за помощь и курирование данной исследовательской работы.

### СПИСОК ЛИТЕРАТУРЫ

1. Xia W., Zhang Y., Yang Y., Xue J.-H., Zhou B., Yang M.-H. GAN Inversion: A Survey. ArXiv210105278 Cs. 2022. URL: <http://arxiv.org/abs/2101.05278>
2. Bishop C. Pattern Recognition and Machine Learning. Information Science and Statistics. 2006. URL: <https://link.springer.com/book/9780387310732>
3. Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks. ArXiv181204948 Cs Stat. 2019. URL: <http://arxiv.org/abs/1812.04948>
4. Karras T., Hellsten J et al. Analyzing and Improving the Image Quality of StyleGAN. arXiv:1912.04958 Cs. 2019. URL: <https://arxiv.org/pdf/1912.04958.pdf>
5. Kryuchkov M., Khanzhina N., Osmakov I., Ulyanov P. CT images GAN-based augmentation with AdaIN for lung nodules detection // Proceedings of SPIE – The International Society for Optical Engineering: 13, Rome, 02–06 November 2020. Rome, 2020. P. 1160526. <https://doi.org/10.1117/12.2587940>–EDN JYZOEO.
6. Huang G., Ramesh M., Berg T., Learned-Miller E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49. 2018. URL: <http://vis-www.cs.umass.edu/lfw/>
7. Viola P., Jones M. Robust Real-time Object Detection. Second international workshop on statistical and computational theories of vision. 2001. URL: <https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers>

8. *Ledig C., Theis L. et al.* Photo-realistic single image super-resolution using a generative adversarial network. ArXiv:1609.04802v5 Cs. 2016.

URL: <https://arxiv.org/pdf/1609.04802v5.pdf>

9. *Goar V., Kuri M., Kumar R., Senjyu T.* Data Compression and Visualization Using PCA and T-SNE. Advances in Information Communication Technology and Computing. 2019.

URL: [https://www.researchgate.net/publication/344000619\\_Data\\_Compression\\_and\\_Visualization\\_Using\\_PCA\\_and\\_T-SNE](https://www.researchgate.net/publication/344000619_Data_Compression_and_Visualization_Using_PCA_and_T-SNE)

10. *Kolo B.* Binary and Multiclass Classification. Weatherford Press. 2010. URL: <https://www.amazon.com/Binary-Multiclass-Classification-Brian-Kolo/dp/1615800131>

11. *Rawlings J., Pantula S., Dickey D.* Polynomial Regression. Applied Regression Analysis. 1998. URL: [https://link.springer.com/chapter/10.1007/0-387-22753-9\\_8](https://link.springer.com/chapter/10.1007/0-387-22753-9_8)

12. *Козина Н. И., Шиян Н. В., Чалченко М. Р.* Современные достижения в области генерации изображений на примере нейронной сети MIDJOURNEY // Сборник материалов XVI-ой международной очно-заочной научно-практической конференции. М.: Научно-издательский центр «Империя», 2023. С. 121–125.

---

## CONTROLLED FACE GENERATION SYSTEM USING STYLEGAN2 NEURAL NETWORK

**M. Isangulov**<sup>1</sup> [0009-0006-3244-0328], **R. Minneakhmetov**<sup>2</sup> [0009-0007-8551-1393],  
**A. Khamedzhanov**<sup>3</sup> [0009-0000-5624-2453], **T. Khafizyanov**<sup>4</sup> [0009-0009-0409-8010],  
**E. Pashaev**<sup>5</sup> [0009-0003-8602-1012], **E. Kalimullin**<sup>6</sup> [0009-0002-6275-699X]

<sup>1–6</sup> Institute of Information Technologies and Intelligent Systems, Kazan (Volga Region) Federal University

<sup>1</sup>marathon.our@gmail.com, <sup>2</sup>razil0071999@gmail.com,

<sup>3</sup>hamedzhanovalmaz@gmail.com, <sup>4</sup>hamstertima@gmail.com,

<sup>5</sup>emil.p.mail@gmail.com, <sup>6</sup>erik182182@gmail.com

### **Abstract**

A novel approach to supervised face generation using open-source generative models including StyleGAN2 and Ridge Regression is presented. A methodology that extends StyleGAN2 to control facial characteristics such as age, race, gender, facial expression, and hair attributes is developed, and an extensive dataset of human faces with attribute annotations is utilized. The faces were encoded in 256-dimensional latent space using the StyleGAN2 encoder, resulting in a set of characteristic latent codes. We applied the t-SNE algorithm to cluster these feature-based codes, demonstrated the ability to control face generation, and subsequently trained Ridge regression models for each dimension of the latent codes using the labeled features. When decoded using StyleGAN2, the resulting codes successfully reconstructed face images while maintaining the association with the input features. The developed approach provides an easy and efficient way to supervised face generation using existing generative models such as StyleGAN2, and opens up new possibilities for different application areas.

**Keywords:** machine learning, face generation, StyleGan, encoder, decoder, latent codes, feature mapping, ridge regression.

### **REFERENCES**

1. Xia W., Zhang Y., Yang Y., Xue J.-H., Zhou B., Yang M.-H. GAN Inversion: A Survey. ArXiv210105278 Cs. 2022. URL: <http://arxiv.org/abs/2101.05278>
2. Bishop C. Pattern Recognition and Machine Learning. Information Science and Statistics. 2006. URL: <https://link.springer.com/book/9780387310732>
3. Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks. ArXiv181204948 Cs Stat. 2019. URL: <http://arxiv.org/abs/1812.04948>
4. Karras T., Hellsten J et al. Analyzing and Improving the Image Quality of StyleGAN. arXiv:1912.04958 Cs. 2019. URL: <https://arxiv.org/pdf/1912.04958.pdf>
5. Kryuchkov M., Khanzhina N., Osmakov I., Ulyanov P. CT images GAN-based augmentation with AdaIN for lung nodules detection // Proceedings of SPIE – The International Society for Optical Engineering: 13, Rome, 02–06 November 2020. Rome, 2020. P. 1160526. <https://doi.org/10.1117/12.2587940>–EDN JYZOEO.

6. *Huang G., Ramesh M., Berg T., Learned-Miller E.* Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49. 2018.

URL: <http://vis-www.cs.umass.edu/lfw/>

7. *Viola P., Jones M.* Robust Real-time Object Detection. Second international workshop on statistical and computational theories of vision. 2001.

URL: <https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers>

8. *Ledig C., Theis L. et al.* Photo-realistic single image super-resolution using a generative adversarial network. ArXiv:1609.04802v5 Cs. 2016.

URL: <https://arxiv.org/pdf/1609.04802v5.pdf>

9. *Goar V., Kuri M., Kumar R., Senjyu T.* Data Compression and Visualization Using PCA and T-SNE. Advances in Information Communication Technology and Computing. 2019.

URL: [https://www.researchgate.net/publication/344000619\\_Data\\_Compression\\_and\\_Visualization\\_Using\\_PCA\\_and\\_T-SNE](https://www.researchgate.net/publication/344000619_Data_Compression_and_Visualization_Using_PCA_and_T-SNE)

10. *Kolo B.* Binary and Multiclass Classification. Weatherford Press. 2010. URL: <https://www.amazon.com/Binary-Multiclass-Classification-Brian-Kolo/dp/1615800131>

11. *Rawlings J., Pantula S., Dickey D.* Polynomial Regression. Applied Regression Analysis. 1998. URL: [https://link.springer.com/chapter/10.1007/0-387-22753-9\\_8](https://link.springer.com/chapter/10.1007/0-387-22753-9_8)

12. *Kozina N. I., Shiyan N. V., Chalchenko M. R.* Sovremennyye dostizheniya v oblasti generacii izobrazhenij na primere nejronnoj seti MIDJOURNEY // Sbornik materialov XVI-oy mezhdunarodnoj ochno-zaochnoj nauchno-prakticheskoj konferencii. M.: Nauchno-izdatel'skij centr «Imperiya», 2023. S. 121–125.

## СВЕДЕНИЯ ОБ АВТОРАХ



**ИСАНГУЛОВ Марат Вильданович** – студент магистратуры Института информационных технологий и интеллектуальных систем КФУ, г. Казань.

**Marat ISANGULOV** is a master's student of the Institute of Information Technologies and Intelligent Systems of KFU, Kazan.

Email: [marathon.our@gmail.com](mailto:marathon.our@gmail.com)

ORCID: 0009-0006-3244-0328



**МИННЕАХМЕТОВ Разиль Рустемович** – студент магистратуры Института информационных технологий и интеллектуальных систем КФУ, г. Казань.

**Razil MINNEAKHMETOV** is a master's student at the Institute of Information Technologies and Intelligent Systems of KFU, Kazan.

Email: [razil0071999@gmail.com](mailto:razil0071999@gmail.com)

ORCID: 0009-0007-8551-1393



**ХАМЕДЖАНОВ Алмаз Рустамович** – студент магистратуры Института информационных технологий и интеллектуальных систем КФУ, г. Казань.

**Almaz KHAMEDZHANOV** is a master's student of the Institute of Information Technologies and Intelligent Systems of KFU, Kazan.

Email: [hamedzhanovalmaz@gmail.com](mailto:hamedzhanovalmaz@gmail.com)

ORCID: 0009-0000-5624-2453



**ХАФИЗЬЯНОВ Тимур Робертович** – студент магистратуры Института информационных технологий и интеллектуальных систем КФУ, г. Казань.

**Timur KHAfizyanov** is a master's student at the Institute of Information Technologies and Intelligent Systems of KFU, Kazan.

Email: hamstertima@gmail.com

ORCID: 0009-0009-0409-8010



**ПАШАЕВ Эмиль Асифович** – студент магистратуры Института информационных технологий и интеллектуальных систем КФУ, г. Казань.

**Emil PASHAEV** is a master's student at the Institute of Information Technologies and Intelligent Systems of KFU, Kazan.

Email: emil.p.mail@gmail.com

ORCID: 0009-0003-8602-1012



**КАЛИМУЛЛИН Эрнест Ришатович** – студент магистратуры Института информационных технологий и интеллектуальных систем КФУ, г. Казань.

**Ernest KALIMULLIN** is a master's student at the Institute of Information Technologies and Intelligent Systems of KFU, Kazan.

Email: erik182182@gmail.com

ORCID: 0009-0002-6275-699X

*Материал поступил в редакцию 26 июня 2023 года*

## РАЗРАБОТКА СИСТЕМЫ ПОИСКА И ИНДЕКСИРОВАНИЯ КОНТЕНТА АУДИОЗАПИСЕЙ

Р. А. Климов<sup>1</sup> [0009-0003-8566-7946], А. Ш. Якупов<sup>2</sup> [0000-0002-2333-8819]

Казанский (Приволжский) федеральный университет, ул. Кремлевская, 35, г. Казань, 420008

<sup>1</sup>itis.klimov@gmail.com, <sup>2</sup>asyakupov@kpfu.ru

### ***Аннотация***

Статья посвящена разработке системы поиска и индексации аудиофайлов с использованием автоматического распознавания речи (ASR) и Elasticsearch. Проанализированы актуальные системы транскрибирования аудиофайлов на русском языке и выбрана система whisper как лучшая. Создан алгоритм оптимизации скорости транскрибирования с помощью параллелизации процессов обработки файла, продемонстрирована его эффективность. Построена система на микросервисной архитектуре, способная индексировать контент аудиофайлов и их метаданные для поиска. Результаты исследования показали, что предложенный подход может быть применен для создания эффективных и гибких систем поиска и аналитики аудиоинформации.

***Ключевые слова:*** транскрибирование, индексирование, параллелизация, микросервисы, масштабируемость.

### **ВВЕДЕНИЕ**

В современном мире, который характеризуется стремительным ростом информационных технологий и объемов данных, особое внимание уделяется аудиоинформации. В условиях постоянного увеличения количества аудиозаписей в образовании, медицине, музыкальной индустрии, журналистике и других областях актуальность и необходимость эффективного поиска по контенту аудиозаписей становятся все более очевидными. Это утверждение подтверждается различными исследованиями.

В конце 2021 года компания Amazon — один из крупнейших провайдеров хранилищ медиаконтента — создала и опубликовала в открытом доступе (open-

source) свою систему поиска по медиаконтенту mediaSearch [1]. Однако эта система специализируется на сервисах Amazon, которые в настоящее время невозможно использовать в России из-за санкций. В то же время, адаптация системы под open-source и такие аналоги, как openShift, является непростой задачей, так как архитектура системы не является модульной – все компоненты сильно связаны и не могут быть заменены аналогами без внесения изменений в исходные код и архитектуру.

В исследовании 2023 года [2] учеными из Бангладеша описана проблема отсутствия решений в сфере поиска по медиаконтенту. Они создают систему Sherlock, позволяющую выполнять поиск по содержимому документов и фотографий, индексируя данные одного из крупнейших open-source файлового хранилища — openShift Swift [3].

В исследовании 2022 года, опубликованном на ResearchGate [4], авторы акцентируют внимание на отсутствии практических решений в области поиска по контенту аудиозаписей и подчеркивают значимость дополнительных научных работ в этом направлении, в частности, в контексте библиотек и информационных наук (Library and Information Science, LIS).

Вышесказанное подчеркивает актуальность разработки системы поиска информации по контенту аудиозаписей, которая может заполнить существующий пробел в этой области и удовлетворить растущие потребности пользователей. Исследование и разработка системы, способной осуществлять поиск по содержанию слов в аудиофайлах, позволят пользователям быстро и точно находить нужную информацию, облегчая тем самым работу и улучшая качество принятия решений.

## **ВЫБОР КОМПОНЕНТОВ СИСТЕМЫ**

**Выбор архитектуры.** При выборе архитектуры системы учитывались два ключевых требования, основанных на особенностях данной доменной области:

1) *Возможность гибкой замены компонентов.* Это важно, поскольку технологии транскрибирования аудио сегодня активно развиваются и улучшаются. Поэтому необходима архитектура, которая позволит быстро и без больших усилий

обновлять поисковую систему, заменяя отдельные ее компоненты без внесения изменений в исходный код остальных.

2) *Горизонтальное масштабирование.* Обработка аудио данных может включать работу с большими объемами информации, а транскрибирование требует значительных вычислительных ресурсов. Для обеспечения приемлемого времени обработки больших объемов данных необходимо обеспечить распределение вычислительных задач между различными сервисами или узлами.

Исходя из названных требований, мы выбрали микросервисную архитектуру. Её основные компоненты представлены на рисунке 1.

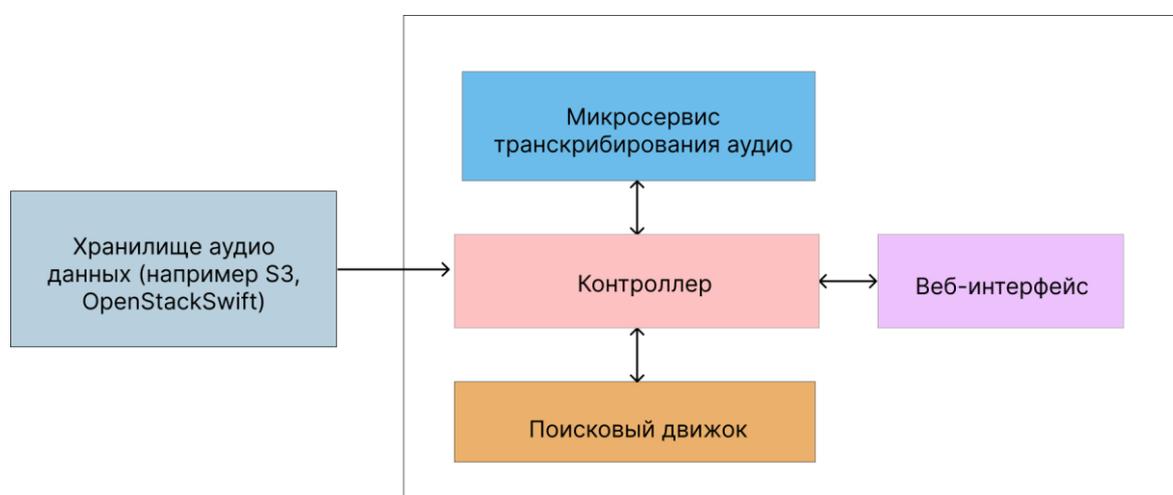


Рис. 1. Высокоуровневая архитектура системы.

**Анализ решений для транскрибирования аудиофайлов.** Для осуществления поиска релевантных файлов нужен инструмент конвертации аудио файлов в текст (транскрибирования). Поэтому следующим шагом исследования стал подбор оптимальной модели ASR (Automatic Speech Recognition). При таком подборе учитывались следующие критерии:

- Высокая точность распознавания – чем точнее перевод из аудио в текст, тем более релевантными будут результаты поиска;
- Поддержка русского языка – итоговый продукт ориентирован на рынок РФ, поэтому нужна поддержка русского языка, который для первой версии системы будет единственным;

- Доступность – наличие возможности использовать модель в РФ;
- Стоимость – у open-source решений будет преимущество перед платными аналогами.

Были рассмотрены 4 решения:

- Google Cloud Speech to Text [5];
- VOSK Speech Recognition Toolkit [6];
- Yandex SpeechKit [7];
- Whisper OpenAI [8].

Google Cloud Speech to Text поддерживает более чем 125 языков мира, в том числе русский, для конвертирования аудио в текст. Доступ осуществляется с использованием API.

Vosk Speech Recognition Toolkit – легковесная open source модель, которая поддерживает более 20 языков и диалектов и автономный режим работы, обеспечивая непрерывную транскрипцию большого словарного запаса.

Yandex SpeechKit – сервис распознавания и синтеза речи платформы Яндекс.Облако. Работает на базе разработанной в Яндекс технологии распознавания речи, поддерживает русский язык.

Whisper OpenAI – система применяется для транскрибирования и перевода, поддерживает около 10 языков, в том числе русский. Обучение модели Whisper проводилось на очень разнообразных данных, что позволяет распознавать не только чистую речь, но и с высокой долей точности понимать уникальные акценты, разговоры при фоновом шуме и технические термины.

В 2022 году опубликована статья [9], где проведено сравнение первых трех решений. Анализ точности транскрипции проведен на основе специализированных метрик WER (Word Error Rate), MER (Match Error Rate), WIL (Word Information Lost) [10]. В качестве экспериментальных данных были использованы две разноплановые аудиозаписи длительностью 20 секунд: запись университетской лекции и телефонный звонок рекламного агента. Для анализа и вычисления был использован язык программирования Python. Из результатов, приведенных в этой статье, видно, что Google Cloud Speech to Text демонстрирует лучшие показатели по всем трем метрикам. Поэтому вторым этапом сравнения был выбор между Google Cloud Speech to Text и Whisper OpenAI, который не был рассмотрен в [11].

---

Сравнение Google Cloud Speech to Text и Whisper OpenAI проводилось нами самостоятельно. За неимением доступа к данным, которые использовались в [9], для второго этапа сравнения были подобраны другие, но максимально схожие с первыми аудио файлы: разноплановые между собой, длительностью также 20 секунд и похожие по тематике. Два аудио файла представляли собой телефонный разговор с сотрудником банка и лекцию по психологии в университете. При подборе внимание уделялось тому, чтобы на записях присутствовали естественные шумы, а дикция говорящих была не идеальной.

Google Cloud Speech to Text позволяет транскрибировать аудио в текст онлайн через платформу Google Cloud. Для анализа Whisper OpenAI на языке программирования Python был написан скрипт, который использует методы системы Whisper для транскрибирования аудио.

Для анализа двух моделей были использованы те же метрики, что и в первом этапе анализа. Для вычисления метрик WER, MER, WIL была использована open-source библиотека JiWER [11]. Также к полученным текстам были применены методы препроцессинга из этой же библиотеки:

- приведение к нижнему регистру;
- удаление лишних пробелов и служебных символов (`\t`, `\n`, `\r` и т.д.);
- приведение текста к определенному виду для вычисления метрик.

В таблицах 1 и 2 приведены результаты вычисления метрик для двух аудиофайлов.

**Таблица 1.** Сравнение метрик для записи лекции по психологии

Модель	WER, %	MER, %	WIL, %
Google Cloud StT	72	72	86
Whisper OpenAI	33	33	51

**Таблица 2.** Сравнение метрик для телефонного разговора с сотрудником банка

Модель	WER, %	MER, %	WIL, %
Google Cloud StT	78	75	88
Whisper OpenAI	8	8	15

Из сравнения значений метрик видно, что Whisper OpenAI существенно более точен по сравнению с Google Cloud Speech to Text, из чего можно сделать предварительный вывод об оптимальности Whisper OpenAI для решения поставленной задачи. В завершение сравним две модели по оставшимся критериям.

**Стоимость.** Если оценивать затраты на использование, то Whisper OpenAI имеет преимущество, так как является открытым решением, в то время, как стоимость использования модели от Google стартует от \$0.024 за минуту аудио файла.

**Доступность.** Даже если выбор будет сделан в пользу Google Cloud Speech to Text, на данный момент оплатить тариф не представляется возможным из-за санкций по отношению к РФ.

**Вывод.** По результатам сравнительного анализа четырех моделей Whisper OpenAI показал наименьшие значения ошибок в транскрипции по всем трем использованным метрикам. Кроме того, эта модель выигрывает и по остальным критериям, по которым проводилось сравнение, и поэтому является оптимальным вариантом ASR модели для использования на русском языке.

**Оптимизация Whisper с использованием параллелизма.** С момента выпуска официальной модели Whisper от OpenAI [8] появились несколько оптимизированных реализаций, включая `whisper.cpp` [12] и `faster-whisper` [13].

На данный момент самой эффективной реализацией является `faster-whisper`, эффективность которой достигается благодаря использованию `CTranslate2` [14] – это библиотека на C++ и Python, предназначенная для эффективной работы с моделями Transformer в процессе вывода. Данный проект включает собственную среду выполнения, применяющую множество методов оптимизации производительности, таких как квантование весов, слияние слоев, переупорядочивание партий и т. д., что позволяет ускорить обработку и снизить потребление памяти моделей Transformer на CPU и GPU. Благодаря этому, `faster-whisper` работает до 4 раз быстрее официальной реализации, используя при этом меньше памяти [13].

Хотя текущая реализация `faster-whisper` обеспечивает возможность параллельного выполнения, она ограничена параллельным транскрибированием разных файлов, в то время как обработка одного файла происходит последовательно. Эта проблема может быть критичной в ряде ситуаций, когда требуется быстро получить транскрипцию одного конкретного файла, например, лекции.

Наше исследование направлено на устранение этого ограничения и повышение производительности транскрибирования путем разделения одного файла на части (чанки) и их параллельной обработки.

Стоит учесть, что в оригинальной модели Whisper используется способ подачи транскрипции предыдущего 30-секундного окна файла в качестве подсказки для следующего окна, чтобы получить непрерывную транскрипцию [15]. В случае разделения аудиофайла на чанки, это свойство не может быть использовано в оптимальной мере, так как контекст между частями теряется. Это может повлиять на качество и непрерывность транскрипции при обработке разделенных аудиофайлов.

Чтобы устранить этот недостаток, можно разделить аудиофайл на чанки, опираясь на результаты алгоритма определения активности голоса, который позволяет найти «независимые» фрагменты речи, разделенные 2–3 секундами тишины. Применение такого подхода позволит сохранить контекст и непрерывность

транскрибации между чанками, так как разделение будет происходить в моменты, когда речи нет, и связь между сегментами минимальна.

**Описание алгоритма разделения файла на чанки:**

1. С помощью python библиотеки 'ffmpeg' [16] к аудиофайлу применяется фильтр 'silencedetect' для определения участков тишины (настройки фильтра: уровень звука – 20 децибел, продолжительность тишины – 2 секунды). На выходе получаем массив, каждое значение в котором является временем начала «тишины» в аудиофайле;

2. Для каждого значения полученного массива вычисляется его «ценность» на основе его расстояния от начала и конца аудиофайла по формуле

$$W = point * (end\_point - point),$$

где *point* – время начала тишины, *end\_point* – время конца аудиофайла.

Ценность точек, наиболее удаленных от начала и конца аудиофайла, обусловлена тем, что такое распределение позволяет создавать чанки примерно одинаковой длины. Это важно, поскольку в параллельной обработке оптимальное использование вычислительных мощностей достигается, когда все параллельные процессы имеют сбалансированную нагрузку. Если чанки имеют разную длину, короткие чанки будут обрабатываться быстрее, и вычислительные мощности, отведенные на их обработку, будут простаивать в ожидании завершения обработки длинных чанков другими процессами;

3. Значения массива сортируются в порядке убывания их ценности;

4. Выбираются  $n-1$  первых значений массива, как «наиболее ценных», где  $n$  – количество доступных процессов для обработки;

5. Полученные значения сортируются в порядке возрастания их временных меток;

6. Аудиофайл разделяется на чанки с использованием выбранных точки тишины в качестве точек разделения.

Общая временная сложность алгоритма  $O(m * \log(m))$ , где  $m$  – количество найденных точек «тишины».

**Результаты тестирования оптимизации.** Тесты выполнялись на MacBook с процессором Apple M1 Pro, 8 ядер, с 16 гигабайтами оперативной памяти с использованием модели "tiny" faster-whisper, версия сборки 0.4.1.

Для транскрибирования была выбрана аудиокнига длительностью 3706 секунд – 1 час, 1 минута и 46 секунд. Результаты представлены в таблице 3.

**Таблица 3.** Результаты тестирования оптимизации транскрибации

Ядра	Время	Скорость
1	134 сек. – 2 мин. 14 сек.	27.66x (без оптимизации)
2	86.56 сек. – 01 мин. 27 сек.	42.81x
3	59.6 сек.	62.18x
4	61 сек.	60.75x
5	59.5 сек.	62.29x
6	58 сек.	63.9x

**Выводы.** При переходе от одного ядра к двум производительность увеличивается значительно, с 27,66x до 42,81x (в 1.54 раза). Однако при дальнейшем увеличении количества ядер прирост производительности становится менее значительным, и в некоторых случаях даже наблюдается небольшое снижение производительности (например, при переходе от 3 к 4 ядрам). Это может быть вызвано неравномерным распределением участков «тишины» аудиофайла, что может приводить к простоям вычислительных ресурсов.

Данный алгоритм можно улучшить, анализируя длительности чанков и подбирая для работы оптимальное количество вычислительных ядер.

**Индексирование контента для поиска.** В качестве поискового движка был выбран Elasticsearch – распределенный, бесплатный и открытый поисковый движок и система аналитики для всех типов данных, включая текстовые, числовые, геопространственные, структурированные и неструктурированные. Высоко адаптивный API запросов Elasticsearch позволяет одновременно использовать фильтрацию, сортировку, постраничный вывод и агрегацию в одном запросе.

Elasticsearch способен легко обрабатывать неструктурированные данные и индексировать файлы JSON. Он автоматически пытается определить сопоставления классов и корректирует новые или удаленные поля. Также он предлагает встроенную функциональность для кластеризации, репликации данных и мгновенного переключения при сбоях, которые прозрачны для пользователя [17].

В роли контроллера потока данных и точкой хода API для хранилищ данных выступает микросервис, написанный на java с использованием фреймворка Spring.

При первой попытке индексирования файла на контроллере происходит первичная настройка elasticsearch. В частности, создается индекс "transcriptions" с настройкой типа данных "completion" для поля текста. Такая конфигурация необходима для поддержки функциональности автодополнения и быстрого предложения результатов поиска на основе символов, введенных пользователем.

Взаимодействие между микросервисами осуществляется с помощью протокола gRPC.

**Описание потока индексирования аудиофайла:**

1. Файл вместе с метаданными передается в систему через эндпоинт /upload по REST или gRPC;
2. Генерируется внутренний идентификатор файла и добавляется в очередь вместе с метаданными файла (например, ссылкой на файл);
3. Файл вместе с идентификатором асинхронно передается на микросервис транскрибирования по gRPC;
4. После получения ответа из микросервиса транскрибирования транскрипция вместе с информацией помещается для индексирования в elasticsearch.

## РЕЗУЛЬТАТЫ

В результате соединения выбранных подходящих компонентов была разработана гибкая и масштабируемая система, способная:

- Эффективно индексировать контент аудиофайлов и их метаданные;
- Выполнять поиск по контенту аудиофайлов;
- Находить для выбранной аудиозаписи наиболее «близкие» по контенту с использованием параметра “more\_like\_this” [18] elasticsearch.

Для демонстрации работы API системы был разработан простой веб-интерфейс, представленный на рис. 2.

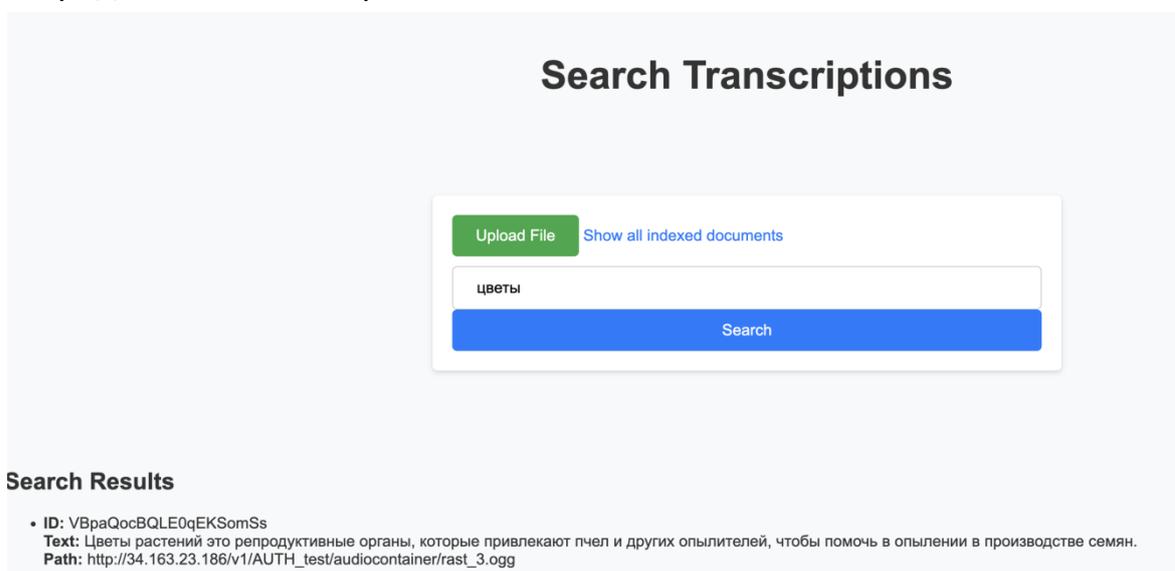


Рис. 2. Веб-интерфейс системы

## ЗАКЛЮЧЕНИЕ

Проведен анализ существующих решений для транскрибации аудио на русском языке, в результате чего было выбрано наиболее эффективное и актуальное решение. Далее был разработан алгоритм для оптимизации скорости транскрибации, основанный на параллелизации процессов, и продемонстрирована его эффективность.

Как результат, создана масштабируемая микросервисная архитектура, позволяющая эффективно индексировать контент аудиозаписей для поиска.

В дальнейшей перспективе планируется продолжить оптимизацию скорости транскрибации, а также сосредоточиться на улучшении пользовательского

опыта за счет добавления инструментов быстрого анализа полученных данных в процессе поиска.

### **СПИСОК ЛИТЕРАТУРЫ**

1. AWS Kendra Transcribe Media Search.  
URL: <https://github.com/aws-samples/aws-kendra-transcribe-media-search>
2. *Noor J., Rownak A., Ratul R., Mondal J.* Sherlock in OSS: A Novel Approach of Content-Based Searching on Object Storage System. 2023.  
URL: <https://arxiv.org/pdf/2303.02105.pdf>.
3. Swift Object Storage.  
URL: <https://www.openstack.org/software/releases/zed/components/swift>
4. *Adrakatti A., Mulia K.R.* Research Challenges of Library and Information Science in retrieving content based Multimedia Information. 2023.  
URL: [https://www.researchgate.net/publication/361107734\\_Research\\_Challenges\\_of\\_Library\\_and\\_Information\\_Science\\_in\\_retrieving\\_content\\_based\\_Multimedia\\_Information](https://www.researchgate.net/publication/361107734_Research_Challenges_of_Library_and_Information_Science_in_retrieving_content_based_Multimedia_Information).
5. Google Speech. URL: <https://console.cloud.google.com/speech/overview>.
6. Vosk. URL: <https://github.com/alphacep/vosk>.
7. Yandex SpeechKit. URL: <https://cloud.yandex.com/en/services/speechkit>.
8. Whisper. URL: <https://github.com/openai/whisper>.
9. *Подопригорова Н. С., Подопригорова С. С., Кан А. Д.* Автоматическое распознавание речи в системе информационного поиска по аудио // Искусственный интеллект в автоматизированных системах управления и обработки данных, Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет). 2022. Т. 2. С. 339–345.
10. *Morris A., Maier V., Green P.* From WER and RIL to MER and WIL. 2004.  
URL: [https://www.isca-speech.org/archive\\_v0/archive\\_papers/inter-speech\\_2004/i04\\_2765.pdf](https://www.isca-speech.org/archive_v0/archive_papers/inter-speech_2004/i04_2765.pdf).
11. JiWER: A Simple and Fast Python Package to Evaluate an Automatic Speech Recognition System. URL: <https://github.com/jitsi/jiwer>
12. Whisper.cpp. URL: <https://github.com/ggerganov/whisper.cpp>
13. Faster-whisper. URL: <https://github.com/guillaumekln/faster-whisper>
14. CTranslate2. URL: <https://github.com/OpenNMT/CTranslate2/>

15. Prompt vs prefix in DecodingOptions.  
URL: <https://github.com/openai/whisper/discussions/117>
  16. FFmpeg. URL: <https://ffmpeg.org/>
  17. Elasticsearch. URL: <https://www.elastic.co/>
  18. Elasticsearch More like this query  
URL: <https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-mlt-query.html>
- 

## DEVELOPMENT OF A SYSTEM FOR SEARCHING AND INDEXING THE CONTENT OF AUDIO RECORDINGS

R. A Klimov<sup>1</sup> [0009-0003-8566-7946], A. Sh. Yakupov<sup>2</sup> [0000-0002-2333-8819]

<sup>1,2</sup> Kazan (Volga Region) Federal University, 35 Kremlevskaya str., Kazan,  
420008

<sup>1</sup>itis.klimov@gmail.com, <sup>2</sup>asyakupov@kpfu.ru

### **Abstract**

The article is devoted to the development of a search and indexing system for audio files using Automatic Speech Recognition (ASR) and Elasticsearch. Current Russian-language audio file transcription systems have been analyzed, and Whisper has been chosen as the best one. An algorithm for optimizing transcription speed using parallelization of file processing processes has been developed, and its effectiveness has been demonstrated. A microservice architecture-based system has been built, capable of indexing audio file content and their metadata for search purposes. The research results show that the proposed approach can be applied to create efficient and flexible systems for searching and analyzing audio information.

**Keywords:** *transcription, indexing, parallelization, microservices, scalability.*

### **REFERENCES**

1. AWS Kendra Transcribe Media Search.  
URL: <https://github.com/aws-samples/aws-kendra-transcribe-media-search>
-

2. Noor J., Rownak A., Ratul R., Mondal J. Sherlock in OSS: A Novel Approach of Content-Based Searching on Object Storage System. 2023.

URL: <https://arxiv.org/pdf/2303.02105.pdf>.

3. Swift Object Storage.

URL: <https://www.openstack.org/software/releases/zed/components/swift>

4. Adrakatti A., Mulia K. R. Research Challenges of Library and Information Science in retrieving content based Multimedia Information.

URL: [https://www.researchgate.net/publication/361107734\\_Research\\_Challenges\\_of\\_Library\\_and\\_Information\\_Science\\_in\\_retrieving\\_content\\_based\\_Multimedia\\_Information](https://www.researchgate.net/publication/361107734_Research_Challenges_of_Library_and_Information_Science_in_retrieving_content_based_Multimedia_Information).

5. Google Speech. URL: <https://console.cloud.google.com/speech/overview>.

6. Vosk. URL: <https://github.com/alphacep/vosk>.

7. Yandex SpeechKit. URL: <https://cloud.yandex.com/en/services/speechkit>.

8. Whisper. URL: <https://github.com/openai/whisper>.

9. Podoprigorova N. S., Podoprigorova S. S., Kan A. D. Automatic Speech Recognition in an Information Retrieval System for Audio // Artificial Intelligence in Automated Control Systems and Data Processing, Bauman Moscow State Technical University (National Research University), 2022. V. 2. P. 339–345.

10. Morris A., Maier V., Green P. From WER and RIL to MER and WIL. 2004. URL: [https://www.isca-speech.org/archive\\_v0/archive\\_papers/inter-speech\\_2004/i04\\_2765.pdf](https://www.isca-speech.org/archive_v0/archive_papers/inter-speech_2004/i04_2765.pdf).

11. JiWER: A Simple and Fast Python Package to Evaluate an Automatic Speech Recognition System. URL: <https://github.com/jitsi/jiwer>

12. Whisper. URL: <https://github.com/openai/whisper>

13. Whisper.cpp. URL: <https://github.com/ggerganov/whisper.cpp>

14. Faster-whisper. URL: <https://github.com/guillaumekln/faster-whisper>

15. CTranslate2. URL: <https://github.com/OpenNMT/CTranslate2/>

16. Prompt vs prefix in DecodingOptions.

URL: <https://github.com/openai/whisper/discussions/117>

17. FFmpeg. URL: <https://ffmpeg.org/>

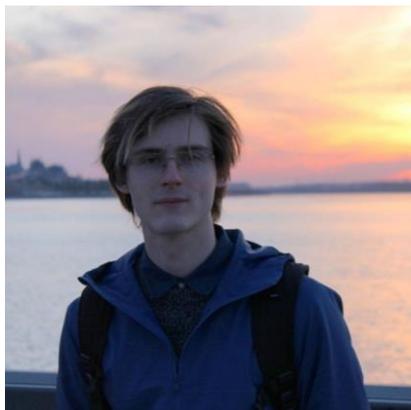
18. Elasticsearch. URL: <https://www.elastic.co/>

19. Elasticsearch More like this query

URL: <https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-mlt-query.html>

---

## СВЕДЕНИЯ ОБ АВТОРАХ



**КЛИМОВ Роман Алексеевич** – магистрант, Казанский (Приволжский) федеральный университет, г. Казань.

**Roman Aleckseevich KLIMOV** is Master's student, Kazan (Volga region) Federal University,

email: [itis.klimov@gmail.com](mailto:itis.klimov@gmail.com);

ORCID: 0009-0003-8566-7946



**ЯКУПОВ Азат Шавкатович** – старший преподаватель, Казанский (Приволжский) федеральный университет, г. Казань.

**Azat Shavkatovich YAKUPOV** – Senior Lecturer, Kazan (Volga region) Federal University, Kazan.

email: [asyakupov@kpfu.ru](mailto:asyakupov@kpfu.ru)

ORCID: 0000-0002-2333-8819

*Материал поступил в редакцию 17 апреля 2023 года*

УДК 004.65

## О РАЗРАБОТКЕ NOSQL СУБД GOLDENRACEDB КАК АЛЬТЕРНАТИВЫ GOOGLE FIREBASE

Р. В. Мосолов<sup>[0000-0002-4399-4397]</sup>

Сеть 70 салонов красоты «GRUSHKA» (Москва)

R.V.Mosolov@ya.ru

### **Аннотация**

Описаны опыт разработки новой нереляционной системы управления базами данных, названной GoldenRaceDB, и предпосылки, располагавшие к ее созданию в контексте проблемы импортозамещения зарубежных технологий. Новая технология реализована на базе серверного окружения Node.js.

Материал статьи предполагает наличие у читателя опыта разработки серверной части системы минимум на одном высокоуровневом языке программирования либо опыта разработки СУБД. Описанная технология не является свободно распространяемой (open source) и была предназначена исключительно для решения задач внутри организации, места ее создания. Однако по прилагаемым в статье листингам можно понять общий вектор создания аналогичной СУБД для разработки собственной упраздненной СУБД на другом высокоуровневом языке программирования.

**Ключевые слова:** система управления базами данных, СУБД, NoSQL СУБД, разработка СУБД, создание СУБД, оператор СУБД, СУБД примеры, запросы в СУБД, средства СУБД, импортозамещение софт, аналоги СУБД.

### **О ПРЕДПОСЫЛКАХ СОЗДАНИЯ**

В рамках курируемого автором в течение последних 2 лет международного хобби-проекта, над которым ведутся работы в качестве разработчика полного цикла (full-stack), встала задача замещения зарубежной технологии Google Firebase, являющейся нереляционной [3] системой управления базами данных (NoSQL СУБД). Таким временным решением стала разработка самописной СУБД, названной GoldenRaceDB в честь организации, в стенах которой она была создана.

Отметим, что разработка собственной СУБД виделась далеко не самым оптимальным решением<sup>1</sup>, и данному решению предшествовал целый ряд причин:

**1. Периодические сбои в работе прежней СУБД.** Как отмечалось ранее, прежней СУБД была Google Firebase. На протяжении двух лет она служила достаточно исправно, однако после запуска спецоперации на Украине, примерно с февраля 2022 г., в ее работе начали наблюдаться довольно частые сбои, вплоть до того, что они начали препятствовать получению доступа к некоторым данным, важным для предметной области.

**2. Нарушение функциональной работы минимум 4 сервисов.** Данный пункт вытекает из п. 1 Списка предпосылок. Было замечено, что СУБД начала оказывать отрицательное влияние в той или иной степени на 4 сервиса (см. рис. 1) трех разрабатываемых веб-приложений организации, где требовалось взаимодействие серверной и клиентской частей.

**3. Нарушение работы 7 из 12 отделов организации хобби-проекта.** Данный пункт вытекает из п. 2 Списка предпосылок. На четырех вышеупомянутых сервисах, как оказалось, за прошедшие два года стала завязана работа свыше половины отделов организации, в рамках которой осуществлялось развитие хобби-проекта. Обнаружение подобного количества взаимосвязей стало точкой принятия решения о срочной *миграции данных*, т. к. доля времени, уделяемая работе с жалобами о неисправности отдельных сервисов, получаемых от пользователей, стала значительно превышать долю времени, уделяемую вопросам поставки новой функциональности для веб-приложений.

**4. Недоступность СУБД MongoDB.** В качестве альтернативы технологии Google Firebase автор рассматривал другую нереляционную СУБД, MongoDB (<https://habr.com/ru/news/t/655291/>), с которой у него имелся некоторый опыт работы ранее. Однако данная СУБД нас тоже «подвела» (см. рис. 2). Данная технология не просто возвращала сбои в работе, она напрочь остановила

---

1 Большинство коммерческих организаций предпочитает использовать промышленные решения вместо самописных.

---

поставки своих услуг в России, предупредив пользователей из России и Беларуси о «немедленном» удалении их данных.

**5. Постоянная статья расходов.** Одной из причин периодических сбоев в работе СУБД Google Firebase являлось то, что хобби-проект начал превышать ранее достаточную для организации квоту ресурсов сервера, где хранились данные организации. По мере роста организации за два года росло и число ее пользователей. Так, например, в первом квартале 2022 г. был снят документальный фильм, размещенный на видеохостинге «Ютуб», собравший за год около 2 млн. просмотров. Выпуск фильма увеличил среднюю посещаемость сайта в 2,5 раза относительно периода, когда аналитика только начала собираться, и была установлена система сбора метрик посещаемости («Яндекс.Метрика»). А на начало 2023 г. статистика посещаемости уже насчитывала 50 тыс. просмотров главной страницы за год при географии пользователей, исчисляемой 83 странами. Поскольку ресурсы VPS<sup>2</sup> еще не были полностью исчерпаны<sup>3</sup>, оплата для использования зарубежной технологии казалось нерациональной с точки зрения экономичности процесса разработки. К тому же, даже несмотря на имеющиеся ресурсы, мы бы не смогли оплатить увеличение квот, поскольку Google Firebase, как и большинство других владельцев зарубежных технологий, стала запрещать оплату услуг с российских банковских карт.

**6. Низкая скорость отклика либо полное его отсутствие на действия пользователей.** Одним из стратегически важных веб-приложений организации являлась разработка Библиотеки 2.0 проекта, который позволит: 1) читать около 400 книг, переведенных в перспективе 5 лет примерно на 20 языков, онлайн; 2) вводить элементы монетизации<sup>4</sup> системы, которая будет осуществляться через продажу абонементов в онлайн-библиотеку. В данном проекте главным компонентом, для использования которого пользователь будет покупать абонемент, является компонент BookContent, отвечающий за генерацию (рендеринг) страниц книг. С системой Google Firebase главная проблема

---

2 Виртуального частного сервера.

3 Оставалось в свободном распоряжении около 59 Гб.

4 Организация, будучи некоммерческой, на март 2023 г. финансируется исключительно на членские взносы и пожертвования.

состояла в том, что она: 1) либо возвращала контент только по прошествии 20 секунд – 2 минут времени; 2) либо не возвращала его совсем. Иначе говоря, в случае публикации данного веб-приложения, мы даже не могли бы гарантировать, что книга, которую решил почитать пользователь, будет загружена. Подобная перспектива была крайне непривлекательной, это было равносильно интернет-магазину, у которого онлайн-продажи являются базисом бизнеса, но при этом технически не работает возможность положить товар в корзину. Механизм, вследствие которого получался такой исход, представлен на рис. 3. Если кратко, данные следствия имели место по той причине, что каждая книга переносится из формата данных DOCX/PDF в JSON не как один модуль, а как совокупность таковых, т. е. каждая книга для удобства хранения делится на главы для последующего упрощения сопровождения. Следовательно, если книга содержит, например, 21 модуль (реальный пример одной из книг), то каждый такой модуль запрашивается у сервера *асинхронно*. Получается, что клиентская часть инициирует 21 запрос на сервер<sup>5</sup>, вероятнее всего, вставая в очередь с другими разработчиками, подключившими данную СУБД к своим проектам. Как следствие ожидание по 21 запросу от сервера вполне может достигать 2 минут, а то и вовсе не завершаться ответом всех модулей книги. В связи с этим мы решили изменить способ получения данных с асинхронного на синхронный и сделать так, чтобы клиент в очереди был всего один<sup>6</sup>. Перевод решения с асинхронного на синхронное и стал решением проблемы ожидания<sup>7</sup>. К данному выводу мы пришли, изучая в [7, с. 160]

---

5 Поиск в API Google Firebase способа, как одним запросом выгрузить сразу все содержимое файлов директории, результата не дал.

6 На самом деле, клиентов как минимум 4, поскольку запросы к серверу используются минимум в 4 сервисах. Добавим к этому, что клиентами могут выступать разные пользователи, что также увеличит объем очереди. Но второй фактор (очереди) пока не столь проблематичным, поскольку посещаемость основного сайта организации обычно не превышает 100 визитов в день, т. е. на данном этапе высокая нагрузка (*англ. high load*) отсутствует.

7 Наблюдательный читатель мог бы заметить, что нами не был описан способ т. н. «ленивой загрузки», при котором главы книг могли бы загружаться при пролистывании экрана монитора вниз. На самом деле, его мы тоже пробовали внедрять, но он оказал несущественное влияние. Если автору не изменяет память, при «ленивой загрузке» или сохра-

один из возможных механизмов распределения трафика. По ряду косвенных признаков было похоже, что Google Firebase, вероятнее всего, придерживается его, а именно, реализует очередь FIFO (*англ.* first in – first out).

**7. Отсутствие в команде экспертизы, связанной с разработкой серверных частей веб-приложения.** В группе разработчиков имелось всего 2 специалиста, оба из которых имеют больше практики с разработкой клиентской части веб-приложений, чем с работой над серверной частью. В связи с этим, делегирование данной задачи было возможным только в случае найма подрядчика. Учитывая некоммерческий статус организации, данного специалиста было бы возможно привлечь лишь на разовой основе<sup>8</sup>, что все равно бы означало впоследствии необходимость освоения концепций программирования серверной части веб-приложений штатными разработчиками.

**8. Отсутствие достаточного объема времени на освоение новой технологии.** Помимо освоения MongoDB как альтернативы Google Firebase, автор также рассматривал вариант подключения PostgreSQL (SQL СУБД). Однако данный вариант, по его оценкам, потребовал бы минимум около 3 месяцев времени на освоение новой технологии, тогда как исправлять проблему с периодическими сбоями в работе 4 сервисов нужно было в срочном порядке. Срок в виде минимум 3 месяцев предполагал:

а. **изучение литературы по основам работы СУБД.** Поскольку автор по специализации является разработчиком клиентской части веб-приложений, что редко требует в профессиональной деятельности взаимодействия с СУБД напрямую (в основном подобное взаимодействие осуществляется через API, предоставляемый разработчиками серверной части), а

---

нялась проблема 20-секундного ожидания, или начинали блокироваться некоторые запросы вследствие квот Google Firebase. Иначе говоря, подобное решение оставалось бы лишь частичным.

8 При средней заработной плате серверного разработчика по Москве на март 2023 г. ежегодных денежных ресурсов, поступивших в бюджет организации с марта 2022 по март 2023, хватило бы лишь на 5 месяцев содержания 1 разработчика серверной части. Учитывая, что разработка ПО не является ее предметной областью (область – издательская деятельность), данную идею было бы крайне сложно инициировать и отстаивать на уровне Совета организации.

также предпочитает начинать освоение любой технологии с теоретической основы, данный шаг для него виделся необходимым. В качестве источника теоретической базы планировалось изучение [3];

б. **затем планировалось освоение дополнительной книги** уже непосредственно по синтаксису новой СУБД, приведенной в [5];

с. **после этого потребовалась бы небольшая практическая работа** по применению данной технологии в целях закрепления теории на практике.

Предлагаемая к рассмотрению новая технология будет базироваться на наборе классических CRUD-операций (*аббр. от англ. Create, Read, Update, Delete*), в целом характерных для СУБД.

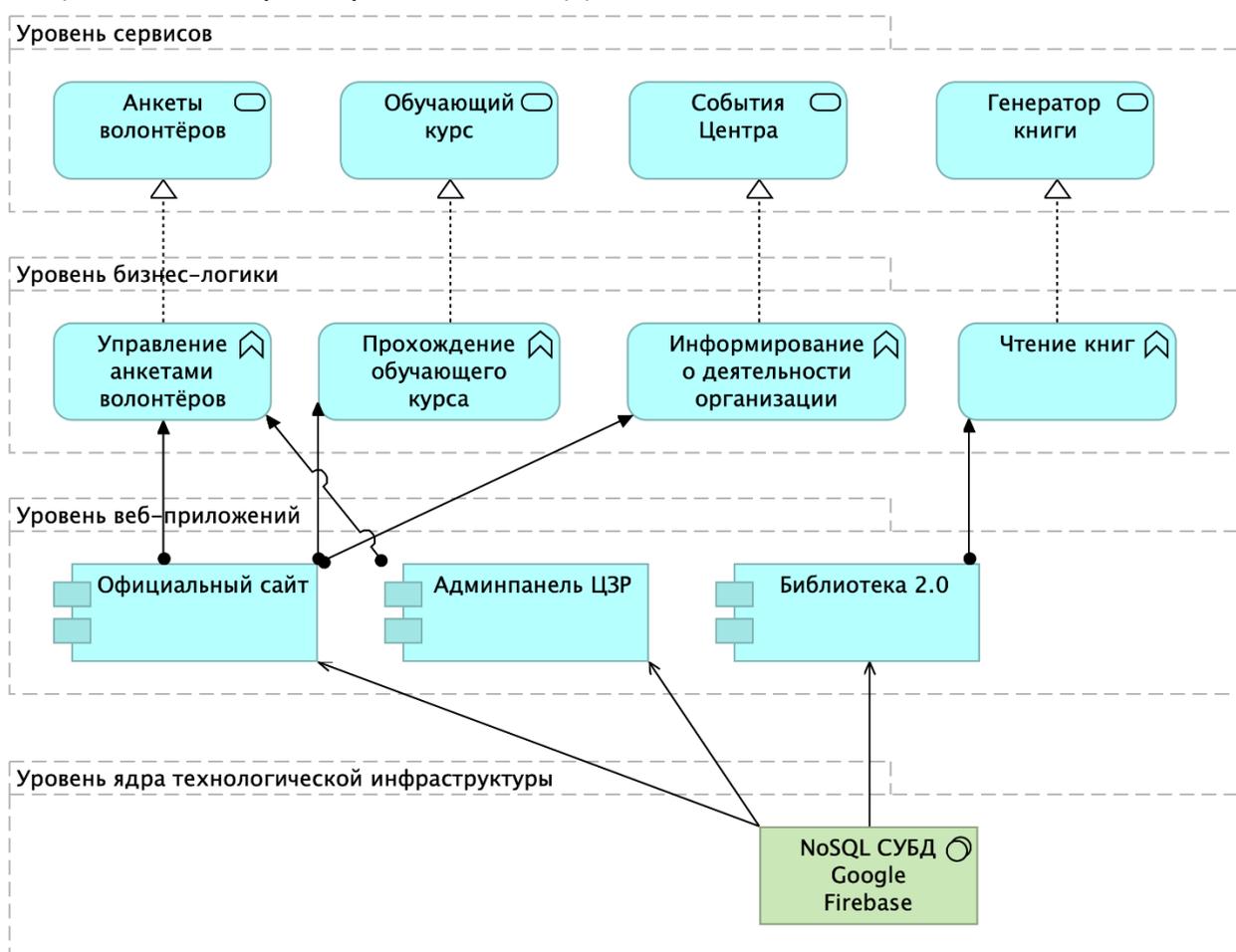


Рис. 1. Высокоуровневая архитектура веб-приложений и сервисов, роль СУБД в ней



Hi [REDACTED]

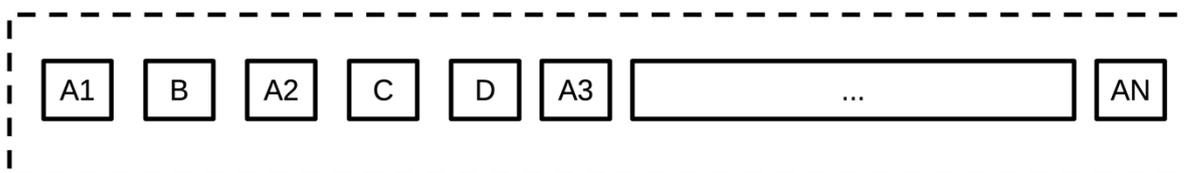
We regret to inform you that due to the new US and international sanctions against Russia and Belarus, your MongoDB Atlas environment, [REDACTED], will be terminated. If you have critical data, we advise you [download backups](#) IMMEDIATELY before they will become permanently unrecoverable.

Regards,

The MongoDB Team

Рис. 2. Уведомление, отправленное администрацией MongoDB разработчикам из России и Беларуси

АСИНХРОННАЯ ОЧЕРЕДЬ (20 СЕК.-2 МИН.)



*Очередь запросов, поступающих в «Google Firebase»*

**Обозначения:** A1, A2, ..., AN – наши запросы; A – совокупность запросов A1, A2, ..., AN; B, C, ..., Z – запросы других разработчиков.

СИНХРОННАЯ ОЧЕРЕДЬ (4-5 СЕК.)



**Комментарий:** в данном случае достаточно всего 1 асинхронного запроса к серверу (A), сервер дальше синхронно объединит все модули в один общий модуль и вернёт его клиенту. Из-за соединения модулей сервер немного потеряет во времени, но итоговый результат всё равно получится минимум в 4-5 раза быстрее, чем при нескольких асинхронных запросах.

Рис. 3. Отличие асинхронной очереди от синхронной

## СТРУКТУРА КОДОВОЙ БАЗЫ

Как можно видеть из структуры репозитория (рис. 4), файловая структура делится на две основных части:

1. **Директорию dbms**, которая содержит бизнес-логику работы СУБД. В модели MVC ее обычно представляют слоем С (*англ.* controller). В ней содержатся методы для манипуляции данными, хранящимися в другой директории.

2. **Директорию db**, хранящую непосредственно базы данных. Именно в данной директории впоследствии отображаются все актуальные данные.

---constants	Константные значения
-----paths.js	Переменные для ускорения доступа к БД
---db	Корневая директория баз данных
-----docs	Аналог Firestore в Google Firebase, базируется на концепции коллекции документов [2]
-----admin	
-----library	
-----readers.json	
-----www	
-----courseSubscribers.json	
-----volunteers.json	
-----files	
-----admin	
-----library	
-----books	
-----creatingMan	
-----creatingMan01.json	
-----creatingMan02.json	
-----...	
-----www	
---dbms	
---node_modules	Содержит основные зависимости проекта
---.gitignore	
---index.js	
---package.json	
---README.md	
---sandbox.js	
---yarn-lock.js	

Рисунок 4. Структура директорий и модулей в репозитории СУБД

Обратите внимание, что после перезаписей внутри баз данных сервер не подтягивает их сразу, поскольку сервер был реализован не на высокоуровневом языке программирования JavaScript (JS), а посредством использования надстройки над JS, т. е. с использованием TypeScript (TS). Последнее означает, что данные могут быть подтянуты только после повторной транспилиации из TS в JS.

Чтобы автоматизировать процесс транспилиации данных, нами используется инструмент pm2 (см. подр. <https://pm2.keymetrics.io/>), позволяющий создавать «демона» (робота), осуществляющего автоматическую процедуру транспилиации раз в час. В принципе, можно запускать его и чаще, например, раз в минуту. Однако следствием слишком частого запуска будут являться более частая недоступность сервера и как следствие работа со сбоями 4 сервисов веб-приложений организации, поскольку каждая такая процедура занимает около 10–15 секунд.

## О МЕТОДАХ КЛАССА, СОДЕРЖАЩЕГО ОСНОВНУЮ ЛОГИКУ

Как упоминалось ранее, логика работы СУБД, отвечающая за манипуляцию данными (см. листинг 1), хранится в директории dbms и представляет собой два класса – Docs и Files – с CRUD-методами. Именуя данные классы, мы старались сделать их привычными для возможных будущих разработчиков, уже имеющих опыт работы с Google Firebase. Тогда как понятие «документ» активно используется в данной СУБД<sup>9</sup>, в его дочернем продукте Firestore, а понятие «файл» – в Storage.

Отличия между двумя данными понятиями, в рамках новой СУБД, достаточно условны, в частности, их сложно привязать к формату данных<sup>10</sup>. Например, данные о пользователях (логин, email и др.) хранятся в формате JSON, а манипулирование ими происходит посредством методов класса Docs.

---

9 Кстати, не только в ней, но и в других нереляционных СУБД тоже, например, в MongoDB. Понятие «документ», в целом, является основополагающим при работе с NoSQL СУБД.

10 Основным форматом обмена данными в GoldenRaceDB является формат JSON, широко распространенный в вебе. Причиной выбора этого формата как основного стало то, что все три системы, разрабатываемые в рамках хобби-проекта, предполагают размещение систем (или как минимум одной из ее версий) в интернет-браузерах.

---

Тогда как главы книг тоже хранятся в JSON, но манипулирование ими происходит уже посредством методов класса Files, поскольку данные файлы предполагаются довольно объемными<sup>11</sup>. Иначе говоря, основными критериями разделения файлов являются следующие два фактора: 1) формат файла; 2) размер файла.

```
const fs = require("fs");
const { __reponame } = require("../constants/paths");

class Docs {
  constructor() {}

  readOne(subdomain, collection, id) {
    return require(`../db/docs/${subdomain}/${collection}`).find(
      (item) => item.id === id
    );
  }

  readAll(subdomain, collection) {
    return require(`../db/docs/${subdomain}/${collection}`);
  }

  create(subdomain, collection, data) {
    fs.writeFileSync(
      __reponame + `/db/docs/${subdomain}/${collection}.json`,
      JSON.stringify([...this.readAll(subdomain, collection), data]),
      "utf8",
      (error, data) => {
        if (error) {
          return error;
        }
        return data;
      }
    );
  }
}
```

```
module.exports = Docs;
```

Листинг 1. Содержимое одного из основных модулей СУБД, отвечающего за манипуляцию данными

## МЕХАНИЗМ РЕЗЕРВНОГО КОПИРОВАНИЯ ДАННЫХ

В процессе разработки новой СУБД мы столкнулись с тем, что данные в процессе ее работы будут храниться на VPS, а не в системах, к которым можно

---

11 Одна книга со всеми главами, объединенными в JSON, в Библиотеке 2.0 «весит» примерно 2–4 Мб, т. е. близка по весу к размеру изображения, снятого на зеркальный фотоаппарат.

---

было бы обеспечить доступ при наличии интернета вроде Git. Последнее создавало риск, что в ряде случаев мы могли просто в одночасье потерять ценные данные, аккумулируемые на протяжении двух лет. В ряде случаев такие данные могли обходиться в весомую сумму денег, например, выступая результатом всевозможных маркетинговых кампаний, на которые выделялся рекламный бюджет.

Развилки, при которых была возможна утеря данных:

1. **Прерывание работы сервера в случае неудачного продления оплаты за содержание VPS.** Автор был наблюдателем прецедента, когда за неуплату VPS все веб-приложения просто выходили из строя (при отсутствии диверсификации веб-приложений по отдельным серверам из-за ограничений в бюджете). При слишком длительном поиске причины их вывода из строя, а такое может случиться, если разработчиков по какой-то причине в команде нет, есть риск, что владельцы VPS просто удалят все наработки организации и продадут очищенный VPS другому заказчику;

2. **Случайный ввод ошибочных команд.** Примером такой команды является `rm -r /home`, введенная сразу после авторизации на VPS, которая начисто удаляет все файлы и директории внутри директории `home`, где на данный момент располагаются репозитории веб-приложений организации. Хотя такой случай маловероятен, для полноты картины он был включен тоже;

3. **Физическая поломка сервера (при отсутствии резервного копирования, осуществляемого провайдером).** В данном случае имеется риск появления форс-мажорных обстоятельств (пожара, наводнения и др.), которые могут оказать физическое отрицательное воздействие на работу сервера и вывести его из строя.

В связи с названными факторами нами было принято решение позаботиться также о резервном копировании данных с новой СУБД. Для этого была выбрана система GitHub, обычно используемая для управления версиями исходного кода, поскольку содержимое директории `db` – набор исходных кодов.

На листинге 2 показана реализация интеграции GoldenRaceDB с GitHub. Главной в нем является строка:

```
exec("git add . && git commit -m \"Create copy of databases\" && git push", ...),
```

она отвечает за ввод в Терминале команды изнутри модуля на Node.js для создания резервной копии баз данных в GitHub.

```
const http = require('node:http');
const exec = require('child_process').exec;
const { v4: uuidv4 } = require("uuid");
const Docs = require("./dbms/Docs");

const hostname = '127.0.0.1';
const port = 3000;

const server = http.createServer((req, res) => {
  res.statusCode = 200;
  res.setHeader('Content-Type', 'text/plain');
  res.end('Hello, World!\n');
});

setInterval(() => {
  const docs = new Docs();
  docs.create("test-db", {
    id: uuidv4(),
    firstName: "Test",
    email: "test@mail.ru",
  });
  exec(
    "git add . && git commit -m \"Create copy of databases\" && git push",
    function(err, stdout, stderr) {
      if (err) {
        console.log(err);
      }
      console.log(stdout);
    }
  );
  return console.log("The setInterval is finished");
}, 5000);

server.listen(port, hostname, () => {
  console.log(`Server running at http://${hostname}:${port}/`);
});
```

Листинг 2. Логика, отвечающая за резервное копирование

## **О РАБОТЕ С БОЛЬШИМИ ФАЙЛАМИ, ГЕНЕРАЦИИ ПУБЛИЧНЫХ URL И РЕГИСТРАЦИИ ПОЛЬЗОВАТЕЛЕЙ**

В процессе разработки кастомной СУБД мы столкнулись с рядом сложностей.

**Первое – работа с большими файлами.** Было обнаружено, что при резервном копировании данных возникают сложности на стадии резервного копирования. Так, например, GitHub устанавливает требование, согласно которому файлы размером свыше 100 Мб нельзя сохранять в репозитории. Иначе говоря, имеется запрет на использование GitHub в качестве хостинга. Данная проблема была решена посредством следующей команды, вводимой в окне терминала:

```
scp -C <PATH_TO_OLD_PLACE> <PATH_TO_NEW_PLACE>,
```

она позволяет скопировать файл с локального компьютера на VPS (для VPS дополнительно необходимо вводить логин и IP-адрес).

**Второе – генерация публичных URL.** Для решения определенных задач нам требовалось формировать публичные ссылки на размещенные в GoldenRaceDB файлы. Решением в данном случае стало использование метода `res.sendFile()` (<https://expressjs.com/ru/api.html#res.sendFile>), предоставляемое серверным фреймворком Express.js [12, с. 117]. Пример реализации логики с использованием данного метода см. в Листинге 3.

```
module.exports.getFilePublicLink = (req, res, next) => {
  try {
    const { path } = req.query;

    if (!req.query.path) {
      res.status(400).json({
        error: 'Please set the path query for the end point',
      });
    }

    const pathChunks = path.split('/');
    const fileFullName = pathChunks[pathChunks.length - 1];
    const pathWithoutFileName = path.split('/') + fileFullName)[0];
    const options = {
      root: './../gr-dbms-videos' + pathWithoutFileName,
      dotfiles: 'deny',
      headers: {
        'x-timestamp': Date.now(),
        'x-sent': true,
      },
    };

    res.sendFile(fileFullName, options, function (error) {
      if (error) {
        next(error);
      } else {
        console.log('Sent:', fileFullName);
      }
    });
  } catch (error) {
    res.status(500).send(error);
  }
};
```

### Листинг 3. Реализация логики для генерации публичных URL с использованием Node.js

**Третье – регистрация пользователей.** В процессе разработки новой СУБД мы столкнулись с тем, что пользователей требовалось каким-то образом регистрировать в системе. Причем речь здесь не просто о добавлении регистрационных данных в БД, а, скорее, о том, как регистрировать email и пароль, гарантируя, что пароль не станет известен кому-либо из тех, у кого есть/будет доступ к СУБД (разработчикам, хакерам [10] и т. д.).

Здесь можно было внедрить реализацию системы шифрования данных, разработанную нами и формализованную в [6, с. 25]. Однако, поскольку время на миграцию данных в новую СУБД было ограничено, а функционал Google Firebase для регистрации новых пользователей ранее уже был апробирован и не имел сбоев в работе, было принято решение пока ограничиться реализацией регистраций учетных записей, имеющейся в Google Firebase.

### ПРЕИМУЩЕСТВА И НЕДОСТАТКИ НОВОЙ СУБД

#### Преимущества:

1. **Простота реализации, возможность самостоятельного внедрения.** Для понимания степени простоты вышеприведенного листинга отметим лишь, что главный модуль технологии Redux [13], широко известной в сообществе разработчиков клиентской части веб-приложений как инструмент для управления состояниями (англ. state), содержал на начальной стадии становления СУБД примерно такой же объем кода.

2. **Были решены основные задачи, стоявшие перед отделом сроком в 1 месяц.** В это время входит не только стадия разработки новой СУБД, но и перевод нуждающихся в ней сервисов на новую СУБД.

3. **Знакомый синтаксис (JS).** Хотя Node.js [1, с. 36] является относительно молодой технологией и сильно уступает ЯП (языку программирования) Java по численности сообщества, тем не менее, синтаксис данной технологии уже был знаком автору, что уменьшало порог входа для ее подключения.

4. **Увеличена скорость отклика основной функции веб-приложения Библиотека 2.0 в 4 раза.**

5. **Небольшой размер кодовой базы.** Если приложения на React.js «славятся» необходимостью переноса за собой директории с обширным списком зависимостей (node\_modules) размером около 250 Мб, что, бесспорно, являлось бы недостатком в средах с высокими требованиями к ограничениям вычислительных ресурсов [8], то Node.js, хотя также требует переноса данной директории, тем не менее, требует для переноса всего около 12 Мб.

6. **Отсутствие необходимости в регулярных расходах за пользование зарубежными технологиями и сервером, что критично для общественных организаций.**

**Недостатки:**

1. **Отсутствие механизма авторизации на уровне СУБД.** Если при использовании других СУБД, например, MongoDB, обычно перед началом использования требуется ввести специальную команду авторизации в кодовой базе, то в предложенном на рассмотрение решении такой авторизации нет, поскольку решение не проектировалось нами как переиспользуемое другими разработчиками ввиду кажущейся простоты его самостоятельной реализации.

2. **Более низкая производительность Node.js как высокоуровневой технологии [11].** Поскольку Node.js изначально строилось как высокоуровневое серверное окружение, то оно в некоторой степени проигрывает СУБД, построенным на низкоуровневых ЯП.

3. **Степень задокументированности.** Технология, будучи новой, пока слабо задокументирована, что может в будущем усложнить сопровождение серверных модулей, завязанных на нее. Частично данную проблему можно решить за счет подключения к проекту технологии TypeScript [9], позволяющей привнести в динамически типизированный JavaScript статическую типизацию. При добавлении данной технологии IDE (см., например, <https://code.visualstudio.com/>) тогда будет подсвечивать API методов.

4. **Отсутствие сообщества разработчиков.** Поскольку технология нова, поиск готовых решений по ней в интернете (например, на портале Stack Overflow) будет затруднительным из-за ограниченного круга работающих с ней разработчиков.

5. **Зависимость от другого иностранного ресурса (GitHub; см. <https://github.com/>).** Как упоминалось выше, механизм резервного копирования данных обеспечивается с завязкой на стороннюю зарубежную систему GitHub, что создает некоторые риски утери резервных копий БД. Однако данное опасение видится избыточным ввиду того, что резервное копирование данных осуществляется в 3 местах: 1) локальная машина автора; 2) VPS; 3) GitHub. Иначе говоря, в случае вывода из строя одного из трех мест хранения данных высоковероятно, что останутся доступными другие два.

6. **Отсутствие дополнительных условий-проверок при манипуляции данными.** Новая СУБД не осуществляет никаких проверок данных, поступающих ей на вход (например, проверку на дублирование данных), что перекладывает написание всех необходимых проверок на разработчика.

7. **Ограниченность небольшим масштабом проекта.** Предложенное решение хорошо подходит для небольших проектов, однако в крупных корпоративных системах, где чаще предпочитают использовать СУБД на базе SQL, оно будет менее предпочтительно.

8. **Задержка при резервном копировании данных может составлять до 24 ч.**

9. **Задержка при обновлении подтягиваемых сервером данных из СУБД может составлять 1 час.**

## **ЗАКЛЮЧЕНИЕ**

Признаем, что описанное выше решение частично наследуется от технологии Google Firebase. Хотя нам неизвестно внутреннее содержимое данной технологии на уровне кодовой базы, полагаем, что рассмотренное решение вполне могло бы стать хорошей альтернативой на российском рынке для небольших организаций. **Главное – оно позволило бы решить две проблемы, стоявшие перед нами: 1) предоставление гарантии получения данных, часть которых является критически важными для организации; 2) а также значительно ускорить работу сервисов (в 1,5–4,5 раза).**

После того, как мы уже внедрили самописную СУБД, нами была обнаружена возможность использования API Яндекс.Диска (см. подробнее

<https://yandex.ru/dev/disk/>), позволяющая решить задачу, как говорят в среде разработчиков, «более элегантно», за счёт хранения некоторых данных на сервере(-ах) Яндекса. Маловероятно, что Яндекс.Диск будет подвержен геополитическому контексту в отношении пользователей из России и Беларуси. Квот на частоту использования API нами также замечено не было (API не апробировалось).

Тем не менее, мы пока решили повременить с данным решением, поскольку есть риск, что впоследствии в одностороннем порядке Яндексом может начать взиматься плата, например, за объём обращений по API, как это уже имело место с Google Firebase. А учитывая некоммерческий статус организации и нестабильность работы волонтеров<sup>12</sup>, **практика работы после февраля 2022 г. показывает, что хотя самописные решения и более сложны на стадии сопровождения, они являются более надёжными и менее зависимыми от действий подрядчиков (создателей др. СУБД и т. п.).**

#### СПИСОК ЛИТЕРАТУРЫ

1. *Аквино К., Ганди Т.* Front-end. Клиентская разработка для профессионалов. Node.js, ES6, REST. СПб.: Питер, 2017. 512 с.: ил. (Серия «Для профессионалов»).
2. *Бэнкер К.* MongoDB в действии. / Пер. с англ. А.А. Слинкина. М.: ДМК Пресс, 2012. 394 с.: ил.
3. *Дейт К. Дж.* Введение в системы баз данных, 8-е издание: Пер. с англ. М.: Издательский дом «Вильямс», 2005. 1328 с.: ил. Парал. тит. англ.
4. *Джуба С., Волков А.* Изучаем PostgreSQL 10 / пер. с англ. А.А. Слинкина. М.: ДМК Пресс, 2019. 400 с.: ил.
5. *Лоре А.* Проектирование веб-API / Пер. с англ. Д.А. Беликова. М.: ДМК Пресс, 2020. 440 с.

---

12 Порядка 40% волонтеров, приходящих в отдел ИТ организации, теряет интерес к работам над проектами в течение 1–4 недель. Полагаем, основной причиной потери мотивации является невозможность использования материальных способов мотивации, оклада и премий, при работе с ними.

---

6. Мосолов Р. В. Программа «История гениального открытия»: автоматизация рутинной деятельности ученых: дис. маг. комп. наук 09.04.04 / Казань, 2021. 47 с.

URL: [https://kpfu.ru/student\\_diplom/10.160.178.20\\_29GG9GW6GT5OBPD9N0DGSIVNS0WANG9\\_PEM-UXEOD82DR2X2Y26\\_F\\_Mosolov.pdf](https://kpfu.ru/student_diplom/10.160.178.20_29GG9GW6GT5OBPD9N0DGSIVNS0WANG9_PEM-UXEOD82DR2X2Y26_F_Mosolov.pdf).

7. Олифер В., Олифер Н. Компьютерные сети. Принципы, технологии, протоколы: Юбилейное издание. СПб.: Питер, 2021. 1008 с.: ил. (Серия «Учебник для вузов»).

8. Портянкин И. Программируем Cloud Native: микросервисы, Docker и Kubernetes. 2022. 177 с.: ил.

9. Розенталс Н. Изучаем TypeScript 3. М.: ДМК Пресс, 2019. 624 с.

10. Скабцов Н. В. Аудит безопасности информационных систем. СПб.: Питер, 2018. 272 с.: ил. (Серия «Библиотека программиста»).

11. Таненбаум Э., Остин Т. Архитектура компьютера. 6-е изд. СПб.: Питер, 2022. 816 с.: ил. (Серия «Классика computer science»).

12. Холмс С. Стек MEAN. Mongo, Express, Angular, Node. СПб.: Питер, 2017. 496 с.: ил. (Серия «Библиотека программиста»).

13. Garreau M., Faurot W. 2018. Redux in Action. Manning Publications. E-book. Accessed: 04 November 2022. Hagos, T. 2021.

---

## ABOUT “GOLDEN RACE DB” DEVELOPMENT (NOSQL DBMS) AS AN ALTERNATIVE OF “GOOGLE FIRABASE”

R. V. Mosolov<sup>[0000-0002-4399-4397]</sup>

*The chain of 70 beauty branches “GRUSHKA” (Moscow)*

R.V.Mosolov@ya.ru

### **Abstract**

In this article, we have described experience of the new NoSQL Database Management System (DBMS) development named as “GoldenRaceDB”. Also, we have described prerequisites that needed to create it in the context of the Russian

software's import problem. The new technology is realised as based on server environment Node.js. To understand this article, a reader should have experience of server-side development by using one of high-level programming language (as minimum) or experience of development custom DBMS. This technical solution is not open source, we have created it to solve our local tasks at our organization exactly, the technology birth place. But a reader can understand a general vector of creating custom resized DBMS by using other high-level programming language.

**Keywords:** *database management system, DBMS, NoSQL DBMS, DBMS development, DBMS creation, DBMS operator, DBMS example, DBMS queries, Russian software import problem, DBMS analogs.*

## REFERENCES

1. Akvino K., Gandi T. Front-end. Klienteskaja razrabotka dlja professionalov. Node.js, ES6, REST. SPb.: Piter, 2017. 512 s.: il. (Serija "Dlja professionalov").
2. Bjenker K. MongoDB v dejstvii / Per. s angl. A.A. Slinkina. M.: DMK Press, 2012. 394 s.: il.
3. Dejt K. Dzh. Vvedenie v sistemy baz dannyh, 8-e izdanie.: Per. s angl. M.: Izdatel'skij dom "Vil'jame", 2005. 1328 s.: il. Paral. tit. angl.
4. Dzhuba S., Volkov A. Izuchaem PostgreSQL 10 / per. s ang. A.A. Slinkina. M.: DMK Press, 2019. 400 s.: il.
5. Lore A. Proektirovanie veb-API / Per. s angl. D.A. Belikova. M.: DMK Press, 2020. 440 s.
6. Mosolov R. V. Programma "Istorija genial'nogo otkrytija": avtomatizacija rutinnoj dejatel'nosti uchenyh : dis. mag. komp. nauk 09.04.04 / Kazan', 2021. 47 s.  
URL: [https://kpfu.ru/student\\_diplom/10.160.178.20\\_29GG9GW6GT5OBPD9N0DGSIVNS0WANG9\\_PEM-UXEOD82DR2X2Y26\\_F\\_Mosolov.pdf](https://kpfu.ru/student_diplom/10.160.178.20_29GG9GW6GT5OBPD9N0DGSIVNS0WANG9_PEM-UXEOD82DR2X2Y26_F_Mosolov.pdf).
7. Olifer V., Olifer N. Komp'juternye seti. Principy, tehnologii, protokoly: Jubilejnoe izdanie. SPb.: Piter, 2021. 1008 s.: il. (Serija "Uchebnik dlja vuzov").
8. Portjankin I. Programmiruem Cloud Native: mikroservisy, Docker i Kubernetes. 2022. 177 s.: il.
9. Rozentals N. Izuchaem TypeScript 3. M.: DMK Press, 2019. 624 s.

10. *Skabcov N.V.* Audit bezopasnosti informacionnyh sistem. SPb.: Piter, 2018. 272 s.: il. (Serija "Biblioteka programmista").
  11. *Tanenbaum Je., Ostin T.* Arhitektura komp'jutera. 6-e izd. SPb.: Piter, 2022. 816 p.: il. (Serija "Klassika computer science").
  12. *Holms S.* Stek MEAN. Mongo, Express, Angular, Node. SPb.: Piter, 2017 496 s.: il. (Serija "Biblioteka programmista").
  13. *Garreau M., Faurot W.* 2018. Redux in Action. Manning Publications. E-book. Accessed: 04 November 2022. Hagos, T. 2021.
- 

### СВЕДЕНИЯ ОБ АВТОРЕ



**МОСОЛОВ Роман Валерьевич** – бакалавр социологии, магистр программной инженерии, разработчик CRM-системы в сети салонов красоты «GRUSHKA» (70 филиалов по России). Альма-матер: Казанский (Приволжский) федеральный университет. Специализация в программировании: разработка клиентской части веб-приложений с использованием технологического стека TS, SASS, React.js, Redux Toolkit, Ant Design, Webpack, Storybook.

**Roman Valerievich MOSOLOV** – Bachelor in Sociology, Master in Computer Science, developer of CRM system at chain of beauty salons "GRUSHKA" (70 branches by Russia). Kazan Federal University is the alma mater. The specialization in software engineering is using the following technological stack: TS, SASS, React.js, Redux Toolkit, Ant Design, Webpack, Storybook.

email: R.V.Mosolov@ya.ru

ORCID: 0000-0002-4399-4397

*Материал поступил в редакцию 8 июля 2023 года*

УДК 004.550

## ЕСТЕСТВЕННОНАУЧНЫЕ МУЗЕИ В ЦИФРОВОМ ПРОСТРАНСТВЕ ГЕОЛОГИЧЕСКИХ ЗНАНИЙ

**В. В. Наумова**<sup>1</sup> [0000-0002-3001-1638], **С. В. Черкасов**<sup>2</sup> [0000-0002-3360-0753],  
**В. С. Ерёменко**<sup>3</sup> [0000-0002-5250-5743], **А. А. Загумённов**<sup>4</sup> [0000-0002-0501-5362]

<sup>1, 3</sup>ФГБУН Государственный геологический музей им. В.И. Вернадского РАН,  
Москва

<sup>4</sup>ФГБУН Институт автоматизи и процессов управления ДВО РАН, Владивосток

<sup>1</sup>naumova\_new@mail.ru, <sup>2</sup>s.cherkasov@sgm.ru, <sup>3</sup>vitaer@gmail.com,

<sup>4</sup>trueepikvic@gmail.com

### **Аннотация**

Охарактеризованы роль музейных данных в научных геологических исследованиях, а также интеграция этих данных в цифровое пространство геологических знаний для более эффективного использования и анализа распределенных геологических и музейных ресурсов и возможности построения цифровых моделей.

**Ключевые слова:** геологические данные, естественнонаучные музеи, открытые данные, интеграция музейных данных, цифровое пространство геологических знаний.

### **ВВЕДЕНИЕ**

Сегодня ИТ-технологии сосредоточены на разработке передовых технологий и рабочих процессов для поддержки успешных открытий всех геологов, геофизиков и геохимиков, чтобы сделать их работу более простой, естественной, мощной. Меньшие затраты времени на обработку данных дают больше свободы в исследованиях и помогают идти в ногу с современными требованиями и возможностями науки. Бурное развитие информационных технологий, в том числе организация хранения больших данных в облаках, их обработки и анализа на территориально распределенных интерактивных вычислительных ресурсах стимулируют также разработку территориально-распределенных тематических информационных систем, в том числе в геологии.

---

© В. В. Наумова, С. В. Черкасов, В. С. Ерёменко, А. А. Загумённов, 2023.

Данная статья распространяется на условиях международной лицензии Creative Commons License Attribution 4.0 International (CC BY 4.0).

Разработка Единого цифрового пространства геологических знаний рассматривается как перспектива развития и интеграции разнотиповых информационных ресурсов и сервисов в современной цифровой среде. Предполагается, что Единое цифровое пространство научных геологических знаний поможет сформировать новые процессы генерации данных и знаний. Ранее нами осуществлялась разработка методических подходов и технологических решений для проектирования такого цифрового пространства, в том числе с использованием методов искусственного интеллекта, на базовой основе созданной ранее в Государственном геологическом музее (ГГМ) им. В.И. Вернадского РАН Информационно-аналитической среды для поддержки и сопровождения научных исследований в геологии, осуществляющей интеграцию территориально распределенной геологической информации с использованием специализированных служб её анализа и обработки.

Целями настоящей работы являются изучение и анализ возможностей информационных ресурсов естественнонаучных музеев в современной цифровой геологической среде.

## **МЕСТО И РОЛЬ ЕСТЕСТВЕННОНАУЧНЫХ МУЗЕЕВ В ЦИФРОВОМ ПРОСТРАНСТВЕ ГЕОЛОГИЧЕСКИХ ЗНАНИЙ**

Естественнонаучные музеи и кернохранилища сохраняют огромный объем каменного материала, что делает их уникальной сокровищницей первичной информации о геологических объектах. Достаточно часто сами геологические объекты, такие, как отработанные месторождения полезных ископаемых, затопленные шахты, ликвидированные скважины, и местонахождения, утраченные в результате природных процессов или деятельности человека, уже не существуют или становятся недоступными. Однако даже для существующих объектов ранее отобранные образцы представляют собой большую ценность, поскольку могут исследоваться современными методами без проведения новых отборов, связанных с дорогостоящими полевыми работами. Именно по указанным причинам научные сотрудники разных организаций обращаются к фондам естественнонаучных музеев. В качестве примера можно привести ряд исследований, проведенных в последние годы с использованием фондов ГГМ им. В.И. Вернадского РАН.

С коллекциями млекопитающих неогена работал научный сотрудник Института проблем экологии и эволюции им. А.Н. Северцова РАН (ИПЭЭ РАН), изучающий экологию ископаемых млекопитающих [1].

Сотрудники Зоологического института РАН изучили фрагмент черепа птерозавра *Istiodactylus latidens* из фондов ГГМ. Экземпляр № ГГМ-1810 представляет собой связанные передние концы рострума и нижнечелюстного сочленения, он был изучен с использованием компьютерной микротомографии. Результаты опубликованы.

С коллекциями пермских насекомых работали палеоэнтомологи Палеонтологического института (ПИН) им. А.А. Борисяк РАН. При переизучении коллекций был установлен новый вид пермского насекомого *Epilestes rasnitsyni* (ГГМ-1844-02/БП-12) [2]).

Коллекцию меловых растений (колл. ГГМ-1835) из Казахстана изучали сотрудники Геологического института РАН [3]. С коллекциями ископаемых млекопитающих (слонов, носорогов, верблюдов) работали сотрудники Южного научного центра РАН. С коллекцией остатков млекопитающих из позднепалеолитической стоянки Сунгирь ознакомились сотрудники кафедры археологии исторического факультета Московского государственного университета им. М.В. Ломоносова. Коллекцию позднепалеозойских растений Якутии изучал сотрудник Института геологии алмаза и благородных металлов Сибирского отделения РАН. С коллекциями ископаемых хищников работали сотрудники ПИН РАН. С коллекциями мезозойских рептилий работал сотрудник ГИН РАН.

С образцами минералов из различных местонахождений России ознакомились и отобрали микропробы для микрозондового анализа сотрудники Минералогического музея им. А.Е. Ферсмана РАН.

Фонды музеев предоставляют коллекционный материал для ознакомления и изучения не только ученым, но и студентам и магистрантам профильных вузов, аспирантам и докторантам. При этом огромное значение приобретают вопросы доступности информации о хранящемся каменном материале.

Естественнонаучные коллекции и содержащаяся в них информация продолжают оцифровываться с ошеломляющей скоростью, также увеличивается и объем данных, которые «рождаются цифровыми». В наших руках появляется все

больше инструментов, помогающих наблюдать, регистрировать и измерять параметры сложных природных систем на основе изучения каменного материала. Растут и возможности распространения информации. Практика обмена богатыми данными, полученными в результате исследований, аналогична той, которая была принята публикующими свои результаты натуралистами столетиями ранее, однако и скорость обмена, и детальность публикуемых данных растут благодаря применению цифровых технологий. Теоретически уже сейчас существуют технологические предпосылки для обмена не обобщенными или интерпретированными, а именно первичными данными во всей их полноте, но информационная инфраструктура для такого обмена пока не реализована.

Ключевые цели работ ГГМ РАН в этом направлении включают:

- создание платформы открытого доступа к данным, связанным с образцами каменного материала;
- предоставление интуитивно понятного интерфейса для обзора и поиска необходимой информации (включая пространственный просмотр), а также загрузки данных;
- предоставление уникальных и постоянных идентификаторов уровня записей и наборов данных для облегчения научного цитирования;
- поддержку пользовательских наборов данных;
- предоставление интерфейса управления прикладными программами (API) для доступа, запроса и загрузки данных в соответствии с признанными стандартами данных;
- возможность основным внешним агрегаторам собирать данные о коллекциях порталов [4].

## ПОРТАЛЫ ОТКРЫТЫХ ДАННЫХ

В течение последнего десятилетия одним из устойчивых трендов развития информатизации музеев стало создание Порталов открытого доступа к данным естественнонаучных музеев. В качестве примеров подобных систем можно привести следующие: Data Portal of Natural History Museum, London [4]; Yale Peabody Museum of Natural History, New Haven (<https://peabody.yale.edu>); Australian Museum, Sydney (<https://australian.museum>) и др.

Лондонский музей естественной истории (NHM, <https://data.nhm.ac.uk>) генерирует и хранит одни из крупнейших глобальных наборов данных, относящихся к биологическому и геологическому разнообразию мира природы.

В качестве программной платформы для реализации Порталов в крупных музеях мира используется платформа CKAN (<https://ckan.org>) – мощная система управления данными с открытым программным кодом, которая делает данные доступными, обеспечивая инструменты для оптимизации данных, их совместного использования, нахождения, представления и хранения.

Платформа CKAN относится к новому типу информационных систем – систем управления данными (DMC), основанных на принципах «открытого доступа» и работах Комитета по данным Международного научного совета CODATA (<https://codata.org>). Объектом хранения такой системы являются наборы данных, т. е. таблицы. Новые данные в систему попадают через интерфейс пользователя или по протоколам обмена метаданными и данными OAI. Для обеспечения уникальности каждого набора данных используется технология присваивания DOI-имен. Таким образом, наборы данных с DOI-именем однозначно идентифицируются и имеют постоянную ссылку для упоминания в интернете и научных публикациях.

Система хранения CKAN включает модуль управления метаданными о хранящихся наборах, которые используются для обмена между DMC-системами. Сами данные хранятся в виде файлов или в специализированном модуле "DataPusher". Система поиска позволяет искать информацию как на уровне наборов данных, так и внутри таблицы через поисковые запросы и фильтры. Для предоставления информации предусмотрено несколько режимов: табличное представление, в виде графиков, на интерактивной карте. CKAN-система, помимо

---

таблиц, способна хранить и предоставлять наборы изображений, веб-страниц, текстовых данных и PDF-файлов.

### **ПОРТАЛ ОТКРЫТЫХ ДАННЫХ ГГМ РАН <http://data.sgm.ru>**

С 2017 года в Государственном геологическом музее им. В.И. Вернадского РАН начаты работы по разработке и реализации Портала открытых данных Музея (Рис. 1) [5]. Основные принципы, которые заложены в Проект разработки этой системы:

1. Открытый доступ к данным музея. Open Access определяется как бесплатный, оперативный, постоянный, полнотекстовый, онлайн-доступ к научным публикациям и данным.

2. Каждая коллекция организована в виде набора данных. Наборы данных – это новое цифровое представление научных данных, которое в нашем случае содержит всю информацию о коллекции.

3. Форма представления коллекции может быть различной: таблицы, ссылки в интернет, фотографии, описания экспонатов, письменные источники к музейным предметам и коллекциям. Особое значение имеет представление монографических коллекций, которые являются основой научных статей и монографий.

4. Представления данных и метаданных согласуются с соответствующими международными стандартами, включая те, которые необходимы для удовлетворения требований организации DataCite (<https://datacite.org>), необходимых для выдачи идентификаторов цифровых объектов DataCite.

5. Стандарты – одна из ключевых составляющих инфраструктуры музейных данных. Введение стандартов обеспечивает совместимость на уровне данных и программных средств, позволяет избежать потерь информации и открывает новые возможности по интеграции данных и их совместной обработке [6]. В решаемой задаче используется формат на базе Dublin Core (<https://www.dublincore.org>).

6. Общедоступный API должен использоваться для создания всего функционала портала и интеграции с другими ресурсами.

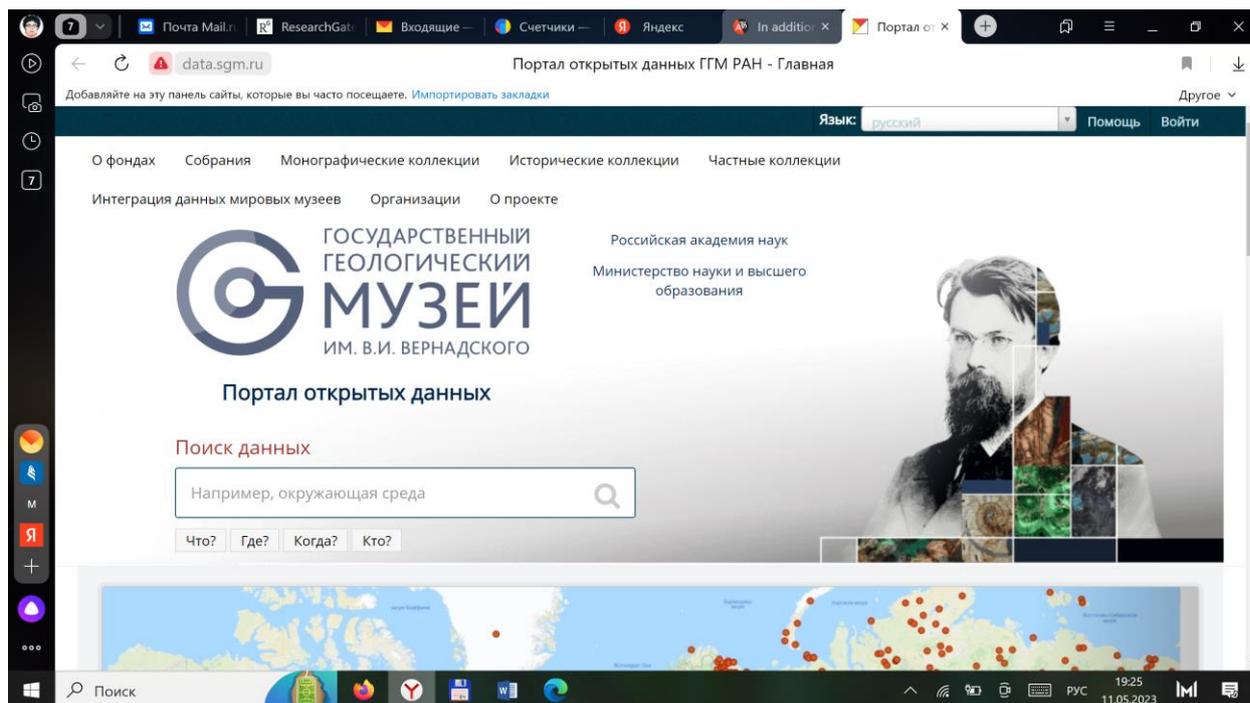


Рис. 1. Главная страница Портала открытых данных ГГМ РАН

Для представления географического расположения образцов на картах используется авторский модуль трансформации текстового названия географического положения образца в абсолютные координаты. Для этого применяется обращение к OpenStreetMap (<https://www.openstreetmap.org>). Данные OpenStreetMap распространяются по лицензии Open Data Commons Open Database License.

Встроенная система поиска системы организации и хранения данных SKAN обладает хорошими возможностями поиска по массивам данных. Проведенное исследование этой системы показало, что набор операций поиска и обработки данных SKAN фактически включает только операции прямого сравнения, поэтому нами предложено и реализовано решение для организации тематического поиска по всем коллекциям одновременно.

Портал данных предоставляет наборы данных пользователю в режиме онлайн. Это позволяет любому научному сотруднику или преподавателю/студенту Университета просматривать, искать, загружать и использовать данные для исследований или образовательных целей.

Наборы данных на Портале соответствуют основным коллекциям Музея (таблицы 1 и 2).

Таблица 1. Наборы данных на Портале открытых данных из собраний ГГМ РАН

Собрания	Экспонаты	Изображения	Публикации
Ископаемая флора	8671	165	8
Ископаемые беспозвоночные	13859	292	13
Ископаемые позвоночные	5977	81	20
Горные породы и руды	15380	167	8
Минералы	46807	657	15
Изделия из камня	1152	34	0
<b>Итого</b>	<b>91846</b>	<b>1396</b>	<b>64</b>

Таблица 2. Наборы данных на Портале открытых данных из монографических коллекций ГГМ РАН

Монографические коллекции		Экспонаты	Изображения	Публикации
<b>Неогеновая система</b>	Позвоночные	10	10	3
<b>Палеогеновая система</b>	Позвоночные	89	0	1
	Беспозвоночные	88	0	1
<b>Меловая система</b>	Позвоночные	63	39	10
	Беспозвоночные	312	6	22
	Флора	59	1	2
<b>Юрская система</b>	Беспозвоночные	1381	54	56

---

<b>Триасовая система</b>	Флора	20	0	1
	Позвоночные	1	1	1
<b>Пермская система</b>	Позвоночные	2	2	1
	Беспозвоночные	680	0	2
	Флора	255	0	4
<b>Девонская система</b>	Беспозвоночные	155	0	5
<b>Ордовикская система</b>	Беспозвоночные	15	15	2
<b>Итого</b>		3130	128	111

Авторы предполагают, что примененные при построении Портала открытых данных ГГМ РАН стандарты метаданных и протоколов будут в дальнейшем использованы для интеграции данных ГГМ РАН в мировую сеть естественнонаучных данных.

### **ЦИФРОВЫЕ СЕТИ ДАННЫХ ЕСТЕСТВЕННОНАУЧНЫХ МУЗЕЕВ МИРА**

Важной задачей является интеграция данных в единую цифровую сеть открытых данных естественнонаучных музеев.

Современные тенденции в информатизации музейной деятельности направлены на открытость данных, доступность в их получении и поиске, унификации и стандартизации описания экспонатов и коллекций и переход к новым объектам «цифровой экспонат» и «цифровая коллекция». Проект DISSCO (<https://www.dissco.eu/>) в Европе и проект iDigBio (<https://www.idigbio.org>) в Северной Америке предлагают решение по организации всестороннего полного доступа к информации естественнонаучных музеев мира. В сеть DISSCO входит 120 музеев из 20 стран. Сеть iDigBio включает 75 организаций.

В основе каждого проекта лежит создание научной инфраструктуры с использованием последних решений в области информатики: FAIR-принципы, единые политики доступа, стандарты и протоколы, создание технологической базы

по оцифровке и описанию экспонатов, единые словари (например, Encyclopedia of Life – <https://eol.org>), системы уникальной идентификация каждого образца.

**ПОРТАЛ ДОСТУПА К ДАННЫМ ПО РФ ГМ РАН ИЗ МИРОВЫХ МУЗЕЕВ – <http://museums.sgm.ru>**

Основная цель создания портала заключается в технологическом обеспечении возможности получения данных о каменном материале, отобранном на территории России, но хранящемся в разных, в т. ч. зарубежных музеях. С этой целью разработаны подходы, методы и технологии для интеграции данных по территории России из территориально распределенных музейных коллекций естественнонаучных музеев России и мира. В результате реализована единая точка входа к данным из музейных коллекций, обеспечивающая пользователям открытый доступ к данным по территории России из музейных коллекций российских и мировых естественнонаучных музеев.

Основные задачи, решенные при создании портала:

- Организация каталога музеев через систему мета описаний;
- Осуществление доступа ко всей возможной информации: описаниям экспонатов из соответствующих таблиц, изображениям, публикациям, картам, проектам и т.п.;
- Организация поиска информации по запросам: где (карта, текстовое значение из списка, текст, когда (возрастная шкала), что (текстовое значение из списка, текст), кто (текстовое значение из списка, текст);
- Организация поискового механизма, предусматривающая перевод текста с русского на английский и обратно;
- Вывод результатов, содержащий название музея, название коллекции и описания экспонатов из соответствующих таблиц, изображения, публикации, карты, проекты и т. п. (при их наличии);
- Визуализация результатов поиска на карте;
- Вывод информации на языке соответствующего музея (русский или английский);
- Организация хранения результатов и предоставления их пользователю.

В качестве решения использованы существующие протоколы доступа к музеям для получения выборки, связанной с Россией, и сохранением метаданных об экспонатах в виде наборов данных (Рис. 2).

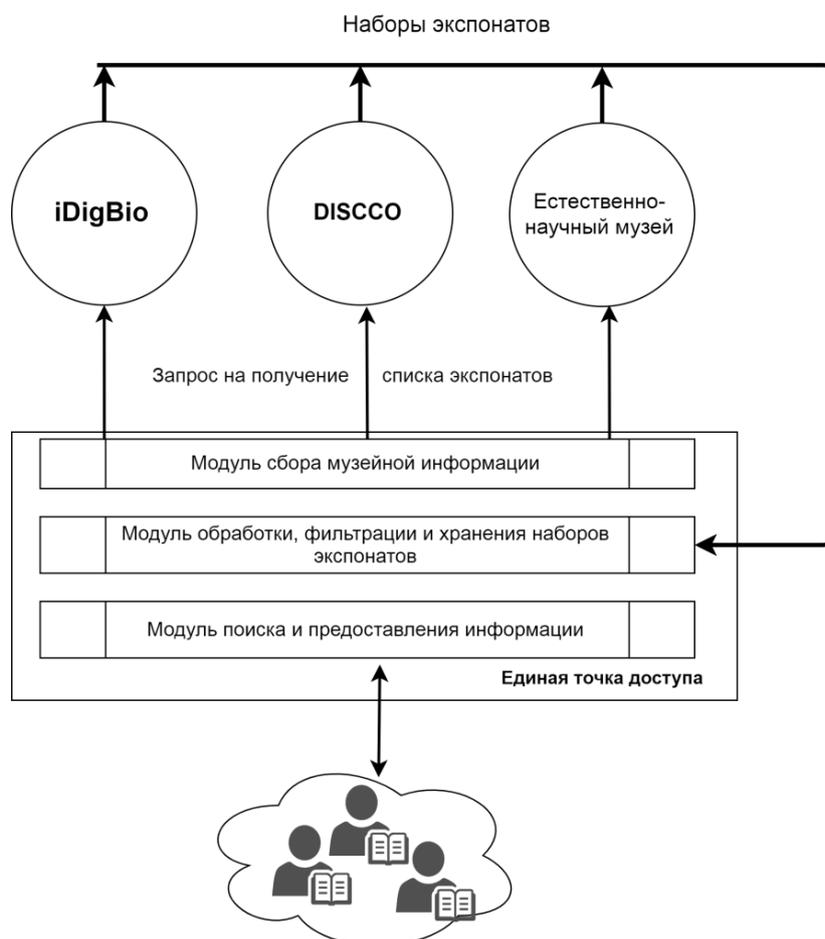


Рис. 2. Концептуальная схема единой точки доступа к музеям России и мира

Портал разработан на платформе SKAN. Данное ПО позволяет легко оперировать наборами данных и создавать дополнительные модули, например, единый поиск по всем коллекциям.

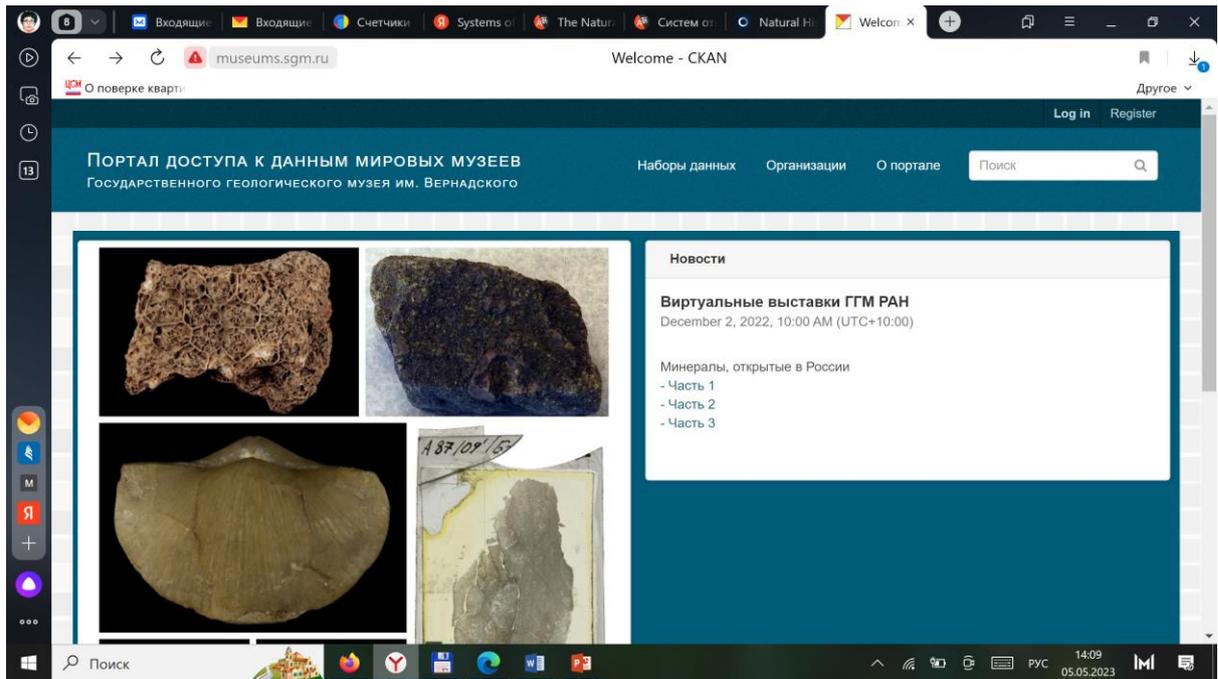


Рис. 3. Главная страница Портала доступа к данным мировых музеев ГГМ РАН

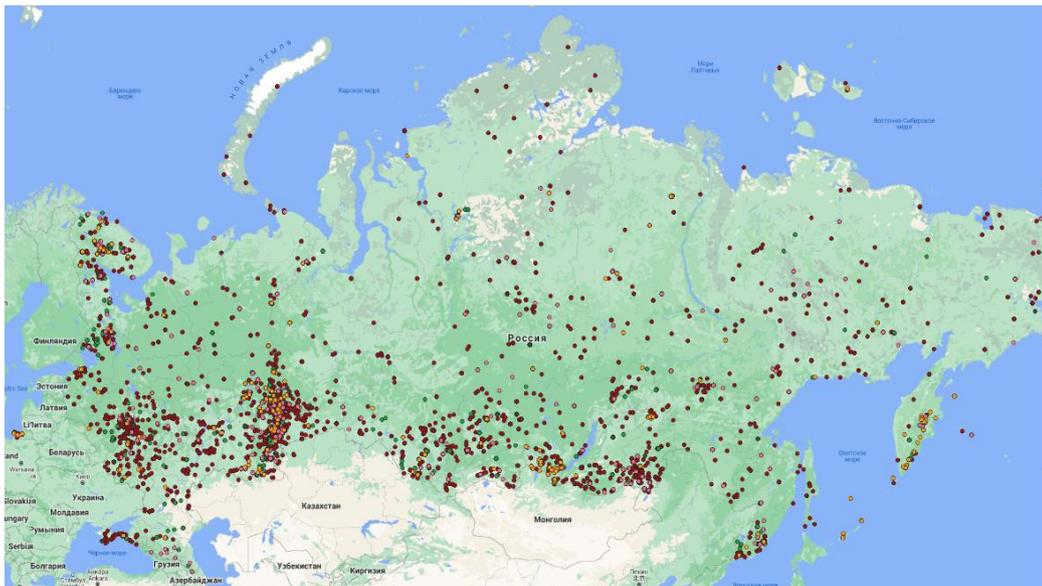


Рисунок 4. Географическая привязка образцов минералов из коллекций крупных естественнонаучных музеев мира

В настоящее время на Портале доступа к данным мировых музеев ГГМ РАН (Рис. 3) интегрированы данные по минералам из 9 музеев мира, включающие 32 538 образцов (Рис. 4, таблица 3).

Таблица 3. Перечень музеев и количество метаданных об образцах минералов, информация о которых интегрирована на Портал ГГМ РАН

Наименование музея	Количество записей
Государственный геологический музей им. В.И. Вернадского РАН, Россия	19426
Минералогический Музей им. А.Е. Ферсмана РАН, Россия	947
Музей естественной истории в Лондоне, Великобритания	5440
Музей естественной истории Оксфордского университета, Великобритания	728
Смитсоновский национальный музей естественной истории, США	4784
Музей минералов Университета Аризоны, США	244
Национальный музей естественной истории в Париже, Франция	648
Шведский музей естественной истории, Швеция	46
Музей Южной Австралии, Австралия	275
<b>Итого</b>	<b>32 538</b>

Кроме этого, 4494 записей об ископаемых животных России (позвоночных и беспозвоночных) импортировано из коллекций 6 мировых музеев (Табл. 4).

Таблица 4. Перечень музеев и количество метаданных об ископаемых животных России, информация о которых интегрирована на Портал ГГМ РАН

Наименование музея	Количество записей
Музей естественной истории в Лондоне (Великобритания)	2238
Смитсоновский национальный музей естественной истории (США)	372
Национальный музей естественной истории в Париже (Франция)	500
Музей палеонтологии университета Калифорнии (США)	170
Национальный исторический музей, Канзас (США)	11
Национальный музей природы и науки (Япония)	1203
<b>Итого</b>	<b>4494</b>

### ЦИФРОВОЕ ПРОСТРАНСТВО ГЕОЛОГИЧЕСКИХ ЗНАНИЙ

Под единым цифровым пространством знаний мы будем понимать компьютерную среду, обращаясь к которой пользователь должен получить ответы на вопросы, касающиеся тех или иных областей науки [6, 7]. Эта среда должна содержать достоверную фундаментальную и научно-популярную информацию; состоять из ряда подпространств, вообще говоря, связанных между собой и относящихся к отдельным научным направлениям; каждое подпространство должно иметь фундаментальную (статическую) основу, включающую базовые постулаты и результаты данного научного направления, и динамическую часть, включающую информацию о новейших достижениях в данной научной области.

Зачастую при исследовании конкретных объектов в геологии приходится использовать разнотипные данные, такие как геологические карты, количественные данные, спутниковые данные, музейные данные, текстовые данные и т. д.

Предполагается, что единое цифровое пространство научных геологических знаний поможет сформировать новые процессы генерации данных и знаний.

Авторами осуществлена разработка методических подходов и технологических решений для создания Единого цифрового пространства геологических научных знаний на базовой основе, созданной ранее авторами Информационно-аналитической среды для поддержки и сопровождения научных исследований в геологии, осуществляющей интеграцию территориально распределенной геологической информации с использованием специализированных служб её анализа и обработки. Используемые ранее методы и модульная архитектура Информационно-аналитической геологической среды (<http://geologyscience.ru>) позволили трансформировать ее в Единое цифровое пространство научных геологических знаний для решения широкого круга задач наук о Земле [8, 9] При этом решались следующие задачи:

а) Разработка основных принципов и методических подходов для создания информационной среды интеграции геологических и музейных данных, а также сервисов ее обработки и анализа для наук о Земле.

б) Разработка и адаптация методов и технологий для интеграции и совместного анализа территориально распределенных данных естественнонаучных музеев и геологической информации.

Новое решение сочетает в себе программное обеспечение (ПО) интеграции разнотипных данных из разнородных источников, технологию стандартизации, ПО организации структурированного хранения, поиска и визуализации геологических и музейных данных и набор сервисов и инструментов по анализу и обработке различных типов этих данных. Простота использования и гибкость персональной настройки под определённые задачи, достигнутая в процессе создания информационно-аналитической среды, обеспечили привлечение к использованию платформы широкого круга учёных-исследователей.

Современная разрабатываемая стадия проекта – это шаг вперед к цифровой среде будущего.

## БЛАГОДАРНОСТИ

Работы выполняются в рамках Государственного задания ГГМ РАН по теме № 0140-2019-0005 «Разработка информационной среды интеграции данных естественнонаучных музеев и сервисов их обработки для наук о Земле», а также темы государственного задания № 1021061009468-8-1.5.1 «Цифровая платформа интеграции и анализа геологических и музейных данных».

## СПИСОК ЛИТЕРАТУРЫ

1. *Rivals F., Belyaev R. I., Basova V. B., Prilepskaya N. E.* Hogs, hippos or bears? Paleodiet of European Oligocene anthracotheres and entelodonts // *Palaeogeography, Palaeoclimatology, Palaeoecology*. 2023. Vol. 611, 111363. P. 1–9.
2. *Felker A., Vasilenko D.* A new species of the ‘protozygopteran’ damselfly (Odonata: permagrionidae) from the lower-middle permian of Russia // *Palaeoentomology*. 2021. 004 (5). P. 462–467.
3. *Maslova N.P., Kodrul T.M., Kachkina V. V.* Leaves of *Ettingshausenia cuneifolia* (Bronn) Stiehler (Angiospermae) and associated carpels and stamens from the Turonian of Southern Kazakhstan // *Paleontological Journal*. 2021. Vol. 55, No. 10. P. 1193–1214.
4. *Scott B., Baker E., Woodburn M., Vincent S., Hardy H., Smith V. S.* The Natural History Museum Data Portal // Database. 2019. P. 1–14. <https://doi.org/10.1093/database/baz038>
5. *Черкасов С. В., Наумова В. В., Платонов К. А., Дьяков С. Е., Еременко В. С., Патук М. И., Стародубцева И. А., Басова В. Б.* Основные принципы разработки открытого доступа к фондовой информации Государственного геологического музея им. В.И. Вернадского РАН // *Информационные ресурсы России*. 2018. №4. С. 9–14.
6. *Антопольский А. Б., Каленов Н. Е., Серебряков В. А., Сотников А. Н.* О едином цифровом пространстве научных знаний // *Вестник РАН*. 2019. Т. 89, № 7. С. 728–735. <https://doi.org/10.31857/S0869-5873897728-735>
7. Единое цифровое пространство научных знаний: проблемы и решения: сборник научных трудов / под ред. Н. Е. Каленова, А. Н. Сотникова, 2021. <https://doi.org/10.23681/610687>

8. Наумова В. В., Ерёменко В. С., Еременко А. С., Загумённых А. А., Патук М. И. От информационно-аналитической среды для поддержки научных исследований в геологии к единому цифровому пространству геологических научных знаний // Электронные библиотеки. 2022. Т. 25(1), С. 15–41.

<https://doi.org/10.26907/1562-5419-2022-25-1-15-41>

9. Наумова В. В., Ерёменко В. С., Патук М. И. Разработка Единой цифровой системы научных геологических знаний // Геоинформатика. 2022. №4. С. 8–14.

<https://doi.org/10.47148/1609-364X-2022-4-8-13>

---

## **NATURAL SCIENCE MUSEUMS IN THE DIGITAL SPACE OF GEOLOGICAL KNOWLEDGE**

V. V. Naumova<sup>1</sup> [0000-0002-3001-1638], S. V. Cherkasov<sup>2</sup> [0000-0002-3360-0753],

V. S. Eremenko<sup>3</sup> [0000-0002-5250-5743], A. A. Zagumennov<sup>4</sup> [0000-0002-0501-5362]

<sup>1–3</sup> *State Geological Museum named after Vladimir Vernadsky, Moscow*

<sup>4</sup> *Institute of Automation and Control Processes FEB RAS, Vladivostok*

<sup>1</sup>naumova\_new@mail.ru, <sup>2</sup>s.cherkasov@sgm.ru, <sup>3</sup>vitaer@gmail.com,

<sup>4</sup>truepikvic@gmail.com

### ***Abstract***

The article describes the role of museum data in scientific geological research, as well as the integration of this data into the digital space of geological knowledge for more efficient use and analysis of distributed geological and museum resources and the possibility of building digital models.

**Keywords:** geological data, natural science museums, open data, integration of museum data, digital space of geological knowledge.

## REFERENCES

1. *Rivals F., Belyaev R.I., Basova V.B., Prilepskaya N.E.* Hogs, hippos or bears? Paleodiet of European Oligocene anthracotheres and entelodonts // *Palaeogeography, Palaeoclimatology, Palaeoecology*. 2023. Vol. 611, 111363. P. 1–9.
2. *Felker A., Vasilenko D.* A new species of the ‘protozygopteran’ damselfly (Odonata: permagrionidae) from the lower-middle permian of Russia» // *Palaeoentomology*. 2021. 004 (5). P. 462–467.
3. *Maslova N. P., Kodrul T. M., Kachkina V. V.* Leaves of *Ettingshausenia cuneifolia* (Bronn) Stiehler (Angiospermae) and associated carpels and stamens from the Turonian of Southern Kazakhstan // *Paleontological Journal*. 2021. Vol. 55, No. 10. P. 1193–1214.
4. *Scott B., Baker E., Woodburn M., Vincent S., Hardy H., Smith V. S.* *The Natural History Museum Data Portal // Database*. 2019. P. 1–14. <https://doi.org/10.1093/database/baz038>
5. *Cherkasov S. V., Naumova V. V., Platonov K. A., D'yakov S. E., Eremenko V. S., Patuk M. I., Starodubceva I. A., Basova V. B.* Osnovnye principy razrabotki otkrytogo dostupa k fondovoj informacii Gosudarstvennogo geologicheskogo muzeya im. V.I. Vernadskogo RAN // *Informacionnye resursy Rossii*. 2018. №4. S. 9–14.
6. *Antopol'skij A. B., Kalenov N. E., Serebryakov V. A., Sotnikov A. N.* O edinom cifrovom prostranstve nauchnyh znaniy // *Vestnik RAN*. 2019. T. 89, № 7. S. 728–735. <https://doi.org/10.31857/S0869-5873897728-735>
7. *Edinoe cifrovoe prostranstvo nauchnyh znaniy: problemy i resheniya: sbornik nauchnyh trudov / pod red. N.E. Kalenova, A.N. Sotnikova*, 2021. <https://doi.org/10.23681/610687>
8. *Naumova V. V., Eremenko V. S., Eremenko A. S., Zagumyonov A. A., Patuk M. I.* Ot informacionno-analiticheskoy sredy dlya podderzhki nauchnyh issledovanij v geologii k edinomu cifrovomu prostranstvu geologicheskikh nauchnyh znaniy // *Russian Digital Libraries Journal*. 2022. T.25(1). S. 15–41. <https://doi.org/10.26907/1562-5419-2022-25-1-15-41>
9. *Naumova V. V., Eremenko V. S., Patuk M. I.* Razrabotka Edinoj cifrovoj sistemy nauchnyh geologicheskikh znaniy // *Geoinformatika*. 2022. №4. S. 8–14. <https://doi.org/10.47148/1609-364X-2022-4-8-13>

## СВЕДЕНИЯ ОБ АВТОРАХ



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

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

Email: naumova\_new@mail.ru,

ORCID: 0000-0002-3001-1638

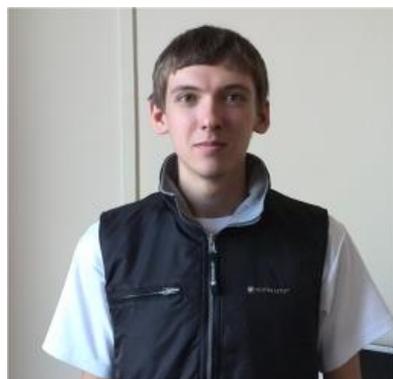


**ЧЕРКАСОВ Сергей Владимирович** – д. т. н., директор Государственного геологического музея им. В.И. Вернадского РАН, Москва.

**Sergey V. CHERKASOV** – Doctor of Technical Sciences, Director of the V.I. Vernadsky State Geological Museum of the Russian Academy of Sciences, Moscow.

Email: s.cherkasov@sgm.ru

ORCID: 0000-0002-3360-0753



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

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

Email: vitaer@gmail.com,

ORCID: 0000-0002-5250-5743



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

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

Email: truepikvic@gmail.com,

ORCID: 0000-0002-0501-5362

*Материал поступил в редакцию 26 мая 2023 года*

УДК 004.4

## СЕМАНТИЧЕСКИЕ СЕРВИСЫ ЦИФРОВОЙ ЭКОСИСТЕМЫ ONTOMATH ДЛЯ МАТЕМАТИЧЕСКОГО ОБРАЗОВАНИЯ

О. А. Невзорова<sup>1</sup> [0000-0001-8116-9446], Е. К. Липачёв<sup>2</sup> [0000-0001-7789-2332],  
К. С. Николаев<sup>3</sup> [0000-0003-3204-238X]

<sup>1-3</sup>Казанский (Приволжский) федеральный университет, ул. Кремлевская, 35,  
г. Казань, 420008

<sup>3</sup>Казанское отделение Межведомственного суперкомпьютерного центра Российской академии наук, ул. Лобачевского, 2, г. Казань, 420008

onevzoro@gmail.com, elipachev@gmail.com, konnikolaeff@yandex.ru

### **Аннотация**

Представлен набор семантических сервисов, разработанных с целью поддержки образовательного процесса в области математики. Функционал этих сервисов основан на использовании математических онтологий OntoMath<sup>Edu</sup> и OntoMath<sup>PRO</sup>. Онтология профессионального математического знания OntoMath<sup>PRO</sup> предназначена для классификации и систематизации математических понятий и включает несколько важнейших областей математики. Образовательная математическая онтология OntoMath<sup>Edu</sup> системно представляет знания по учебному курсу «Планиметрия». Для применения онтологий в образовательных приложениях разработан подход к проектированию пререквизитных отношений в названных онтологиях. Для поддержки математического образования разработаны сервисы семантического поиска по математическим формулам, семантического аннотирования учебных материалов, визуализации подграфов семантической сети онтологии OntoMath<sup>Edu</sup>, а также параллельный формальный/неформальный корпус математических утверждений и система автоматической генерации тестовых вопросов по математическим дисциплинам.

Приведены примеры успешного применения разработанных программных инструментов. Эти инструменты встроены в цифровую экосистему OntoMath, в рамках которой осуществляется взаимодействие семантических сервисов управления математическим знанием.

**Ключевые слова:** Цифровая экосистема, экосистема *OntoMath*, предметная онтология, математическая онтология *OntoMath<sup>PRO</sup>*, образовательная онтология *OntoMath<sup>Edu</sup>*, пререквизитное отношение, семантический сервис.

## ВВЕДЕНИЕ

В настоящее время информационно-коммуникационные технологии являются неотъемлемой частью образовательного процесса. Созданы и развиваются информационные платформы, поддерживающие как процесс обучения в общеобразовательной и вузовской системах образования, так и предоставляющие возможности индивидуального дополнительного обучения (см., например, [1–5]). Важное значение при разработке программных инструментов, ориентированных на образовательный процесс, имеет система организации знаний, например, в виде онтологий предметных областей (например, [6]). Формализация образовательных предметных областей в виде онтологий создаёт возможность разработки семантических сервисов, ориентированных на образовательный процесс. Применение семантических технологий в математических образовательных курсах может способствовать более глубокому усвоению математических знаний (см., например, [7–9]).

В статье представлены программные сервисы, разработанные с целью применения в математическом образовании. Эти сервисы включены в цифровую экосистему *OntoMath* и используют онтологии, входящие в её состав.

Образовательные сервисы цифровой экосистемы *OntoMath* поддерживают на текущем этапе системный подход к изучению основных положений математической теории, тем самым внося вклад в достижение цели, связанной с освоением теоретических положений на доказательном уровне. Успешное достижение этой цели возможно на пути учёта индивидуальных особенностей учащихся в восприятии достаточно сложных и абстрактных математических конструкций, а также в понимании и усвоении логики доказательства. Разработка системы пререквизитных отношений между классами математических онтологий обеспечивает возможность использования адаптивных технологий обучения, предлагающих, в частности, построение индивидуальных траекторий обучения.

## 1. ЦИФРОВАЯ ЭКОСИСТЕМА OntoMath: ОНТОЛОГИИ И СЕРВИСЫ

В рамках проекта построения цифровой математической библиотеки Lobachevskii-DML (<https://lobachevskii-dml.ru/>) с целью управления математическим знанием создана цифровая экосистема OntoMath (см., например, [10–12]). Эта экосистема спроектирована в виде комплекса онтологий и программных инструментов обработки научных документов.

Общая архитектура цифровой экосистемы OntoMath приведена в [13]. Семантическим ядром экосистемы являются онтология профессиональной математики OntoMath<sup>PRO</sup> и образовательная математическая онтология OntoMath<sup>Edu</sup> (см. [14, 15]), разработанные на основе методов онтологического инжиниринга. Также для представления метаданных и описания логической структуры математических документов используются внешние онтологии (подробнее см. [16, 17]).

На Рис. 1 приведена архитектура цифровой экосистемы OntoMath, в которой сделан акцент на образовательные приложения: компоненты экосистемы, предназначенные для применения в образовательном процессе, выделены рамками.



Рис. 1. Архитектура цифровой экосистемы OntoMath. Компоненты экосистемы распределены по уровням – компоненты, расположенные уровнем выше, используют функционал нижележащих компонент. Отдельный блок образуют компоненты экосистемы (выделены утолщенной внешней границей), предназначенные для образовательных целей

В цифровую экосистему OntoMath включены программные сервисы, поддерживающие образовательный процесс в области математики. На Рис. 2 приведена общая схема образовательной подсистемы, входящей в экосистему OntoMath.

Оранжевым цветом выделена онтология  $\text{OntoMath}^{\text{Edu}}$ , являющаяся источником понятий для всех остальных образовательных компонентов экосистемы. Серым цветом отмечены наборы данных, применяемые в соответствующих сервисах экосистемы. Зеленым цветом обозначен центральный компонент образовательной части экосистемы  $\text{OntoMath}$  – интеллектуальная цифровая образовательная платформа для школьной математики. Синим цветом обозначены программные сервисы на интеллектуальной цифровой образовательной платформе.



Рис. 2. Образовательная подсистема цифровой экосистемы  $\text{OntoMath}$ . Серым цветом выделены наборы данных, зеленым цветом обозначена интеллектуальная цифровая образовательная платформа для школьной математики, синим цветом обозначены программные сервисы

Отметим, что в математических документах существенная часть информации содержится в формулах, которые в поисковых сервисах общего назначения, как правило, не учитываются. Сервисы семантического поиска по математическим формулам, представленные в экосистеме  $\text{OntoMath}$ , позволяют по концептам, указанным в запросе, производить поиск формул, включающих формульные выражения этих концептов. Для выполнения разметки и аннотирования математических учебных материалов разработаны сервис семантической разметки и справочная база данных [18, 19]. Также создан ресурс, содержащий математические утвержде-

ния, представленные различными степенями формализации, – этот ресурс в экосистеме называется «Параллельный формальный/неформальный корпус математических утверждений» [20]. Для проверки математических знаний разработан сервис автоматической генерации тестовых вопросов [21]. Указанные сервисы подробнее описаны ниже в разделе 3.

## 2. ОБРАЗОВАТЕЛЬНЫЕ КОМПОНЕНТЫ ОНТОЛОГИЙ МАТЕМАТИЧЕСКОГО ЗНАНИЯ

### 2.1. Онтология профессиональной математики **OntoMath<sup>PRO</sup>**

Онтология профессионального математического знания **OntoMath<sup>PRO</sup>** разрабатывается с 2012 года, и первая ее версия представлена в [22]. В настоящей статье рассмотрены подход и некоторые частные решения по использованию онтологии в образовательном процессе.

Так, например, на основе онтологии **OntoMath<sup>PRO</sup>** разработаны тесты, по которым проводилась проверка испытуемых на знание применимости методов решения систем линейных алгебраических уравнений. В каждом тесте представлен фрагмент онтологии **OntoMath<sup>PRO</sup>**, содержащий иерархию задач и иерархию методов решения систем линейных уравнений, а также частично указаны связи между задачами и методами. Согласно условию задания, испытуемые различного уровня подготовки (студенты, бакалавры, магистранты) должны были на основе предъявленного списка методов и задач восстановить пропущенные исходные связи между классами иерархий. Для оценивания результатов применены стандартные метрики информационного поиска – точность, полнота и  $F$ -мера. При этом под полнотой понимается отношение числа связей, правильно указанных в ответе на тест, к числу связей во фрагменте онтологии, используемом в тесте, а под точностью – отношение числа связей, правильно указанных тестируемым, к общему числу связей, обозначенных в ответе тестируемого. Результаты выполнения задания показали, что установление связей в упрощенной таксономии задач было хорошо выполнено всеми участниками опроса, а самой сложной частью теста стало установление отношения «*решает*». Предложенный подход, по нашему мнению, позволяет не только выполнить проверку отдельных фактов по предмету, но и помочь установить содержательные связи между понятиями и, таким образом, сформировать целостное представление о данной области знаний.

В настоящее время разработана следующая версия онтологии OntoMath<sup>PRO</sup> (<https://github.com/CLLKazan/OntoMathPro>). Эта версия спроектирована на новой онтологической модели, которая также была применена при разработке образовательной онтологии OntoMath<sup>Edu</sup> [23, 24].

## 2.2. Образовательная математическая онтология OntoMath<sup>Edu</sup>

Образовательная математическая онтология OntoMath<sup>Edu</sup> используется как основной ресурс в математической образовательной платформе Казанского федерального университета и применяется в ряде семантических сервисов цифровой экосистемы OntoMath. Подробно онтология OntoMath<sup>Edu</sup> описана в [25]. Она организована в виде трех уровней: уровень предметной онтологии, лингвистический уровень и мета-онтологический уровень. Уровень предметной онтологии содержит независимые от языка концепты, относящиеся к разделу «Планиметрия» школьной математики. Лингвистический уровень включает русский, английский, татарский и испанский терминологические лексиконы, которые содержат информацию о морфологических и синтаксических характеристиках терминов [26].

## 2.3. Пререквизитные отношения в математических онтологиях

Важным применением онтологий является разработка семантических сервисов, поддерживающих процесс формирования образовательных траекторий. В образовательном процессе понятия предметной области изучаются в последовательности, заданной образовательной программой. В предметной онтологии последовательность изучения понятий можно обозначить с помощью пререквизитных отношений [27, 28]. Если, согласно образовательной программе, для изучения понятия  $B$  требуется предварительно изучить понятие  $A$ , то в предметной онтологии концепт  $A$  определяется как пререквизит для концепта  $B$ . Отметим, что пререквизитные отношения являются независимыми по отношению к таксономическим отношениям и образуют независимую иерархию онтологии.

В онтологии OntoMath<sup>Edu</sup> спроектированы пререквизитные отношения, связывающие понятия раздела «Планиметрия» в соответствии с общеобразовательными программами России и Великобритании. При создании пререквизитных отношений использованы прямой и опосредованный подходы. В соответствии с прямым подходом пререквизитные отношения устанавливаются между двумя концептами онтологии. В качестве примера прямого подхода создания пререквизитов

---

приведем последовательность пререквизитных отношений, начинающуюся с концепта «Прямая» и продолженную до концепта «Теорема Пифагора». В соответствии с образовательной программой эта последовательность имеет вид: «Прямая» → «Луч» → «Угол» → «Прямой угол» → «Прямоугольный треугольник» → «Теорема Пифагора». Каждый концепт может входить в несколько пререквизитных последовательностей, например, концепты «Луч» и «Угол» из приведенной ранее последовательности также входят в последовательность пререквизитных отношений «Точка» → «Луч» → «Угол» → «Равные углы» → «Биссектриса угла». При опосредованном подходе пререквизитные отношения устанавливаются между образовательными уровнями, содержащими понятия, включенные в образовательную программу в соответствии с определенным годом обучения и профилем подготовки [28].

В онтологии OntoMath<sup>PRO</sup> профессионального математического знания также разрабатывается система пререквизитных отношений. Эта система отражает последовательность изучения математических понятий в общих и специальных математических курсах университетов. В этом состоит основное отличие отношения «Пререквизит» от отношений «Класс–Подкласс» и «Определяется через», которые также используются при формировании образовательных траекторий в сервисах цифровой экосистемы OntoMath. Особенности указанных онтологических отношений проиллюстрированы на рисунках, приведенных далее. Покажем, как пример, взаимосвязи понятий онтологии, связанных с некоторыми концептами, использующими термин «производная». Отметим, что в профессиональной математике имеется большое число математических понятий, использующих термин «производная», и только часть из них изучается в университетах. На Рис. 3 и 4 показан фрагмент сети понятий онтологии OntoMath<sup>PRO</sup>, содержащий несколько классов онтологии, в названии которых присутствует термин «Производная».

- ▼ ○ отображение
  - $n$ -линейное отображение
  - E3144
  - Векторное произведение
  - Дисторсия
  - ▼ ○ Дифференцируемое отображение
    - ▼ ○ Производная отображения
      - Дифференциал Гато
      - Дифференциал Фреше
      - Производная Гато
      - Производная Фреше
  - ▶ ○ Ковариантное дифференцирование
  - Продолжение дифференциального оператора
  - Продолжение дифференциального уравнения

Рис. 3. Концепты «Производная Гато» и «Производная Фреше» в таксономии онтологии OntoMath<sup>PRO</sup>

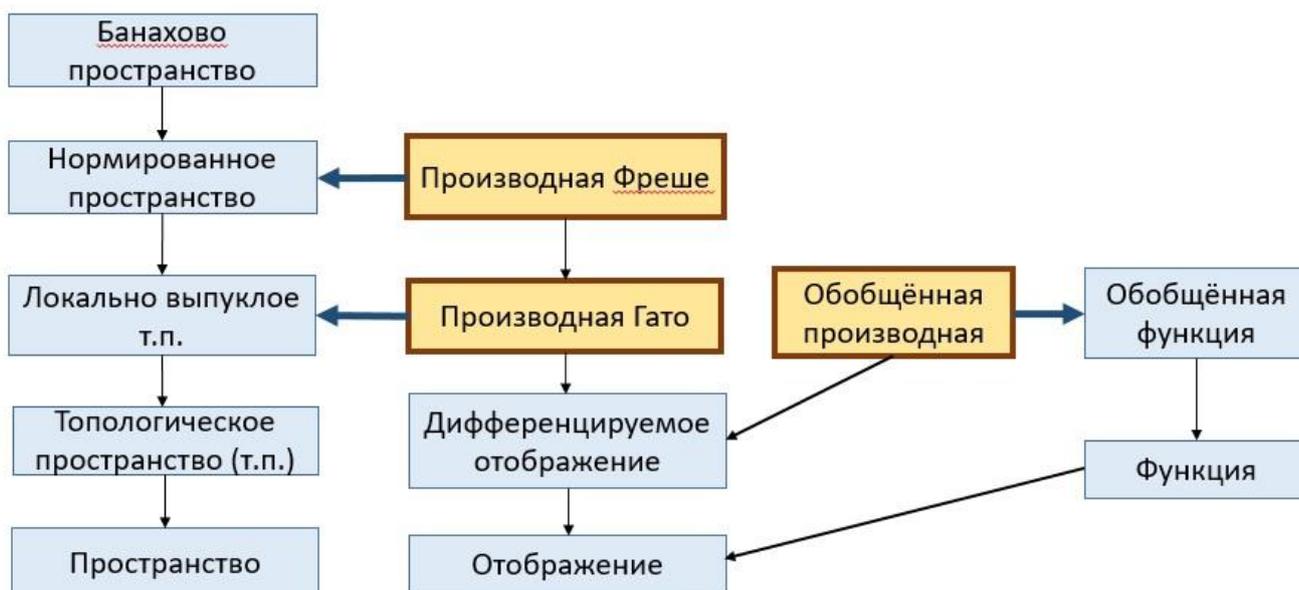


Рис. 4. Фрагмент сети понятий онтологии OntoMath<sup>PRO</sup>. Отношение «Класс–Подкласс» отображается в виде вертикальной или диагональной стрелки, направленной от подкласса к родительскому классу, а для отношения «Определяется через» используется горизонтальная стрелка. Утолщенной рамкой и цветом выделены классы, имеющие в названии термин «Производная»

В университетские образовательные программы по математическому анализу, функциональному анализу и теории функций включены разделы, предпола-

гающие изучение обобщенной производной, производных Фреше и Гато, производной дробного порядка и др. Производная Фреше определяется для нормированных пространств (см. [29]). Производная Гато является обобщением производной Фреше для отображений, определенных на локально выпуклых топологических пространствах [30]. В определении производной Фреше используется понятие нормированного пространства, а в определении производной Гато – более общее понятие локально выпуклого топологического пространства. В университетских курсах по функциональному анализу, как правило, производная Гато (слабая производная) и производная Фреше (сильная производная) изучаются одновременно и определяются в банаховых (см., например, [31]) или нормированных пространствах (см., например, [32]). На Рис. 5 показано, как последовательности изучения этих понятий определяются пререквизитными отношениями в математической онтологии. Первая цепочка пререквизитных отношений соответствует последовательности изучения указанных понятий согласно [31]. Вторая цепочка пререквизитов определяет последовательность изучения математических понятий согласно учебнику [32].



Рис. 5. Фрагмент сети пререквизитных отношений, определяющей последовательность изучения понятий «Производная Гато» и «Производная Фреше». Последовательность концептов связана пререквизитными отношениями, обозначенными стрелками.

### 3. СИСТЕМА ОБРАЗОВАТЕЛЬНЫХ СЕРВИСОВ ЭКОСИСТЕМЫ OntoMath

Приведем краткое описание семантических сервисов, входящих в образовательные компоненты цифровой экосистемы OntoMath (Рис. 2).

#### 3.1. Сервис MathEduSearch семантического поиска по математическим формулам

Поиск документов, включающих формулы с математическими понятиями, наиболее применим в научной среде. Для образовательных приложений предложена модификация сервиса семантического поиска, направленная на поиск формул по названию и понятиям, включенным в них. В качестве набора данных, лежащих в основе сервиса, выступает параллельный формальный/неформальный корпус математических утверждений. Функционал разработанного прототипа сервиса аналогичен функционалу сервиса Semantic Formula Search семантического поиска по математическим формулам [33, 34]. Особенностью прототипа сервиса является использование нового механизма выдачи документов, содержащих возвращаемые формулы. Сервис работает следующим образом.

В поисковой запрос передаются математические понятия или их обозначения, для которых необходимо найти содержащие их формулы. В качестве результата поисковый сервис выводит веб-страницу, содержащую список релевантных формул с визуализацией каждой из них. Для выполнения этой задачи на языке JavaScript с использованием библиотеки MathJax [35] разработан программный модуль, осуществляющий встраивание в веб-страницу формул в LaTeX-нотации, общепринятых названий формул, а также текстовые описания этих формул.

#### 3.2. Сервис семантического аннотирования учебных материалов и справочная база данных

Этот сервис выполняет поиск и аннотирование математических концептов из онтологии OntoMath<sup>Edu</sup> в образовательных документах. В настоящее время данный сервис встроен в процесс разработки в Казанском федеральном университете материалов курса дистанционного образования «Технология решения планиметрических задач» [2]. Основной задачей сервиса являются выделение концептов онтологии OntoMath<sup>Edu</sup> на страницах курса и обогащение страниц с этими концептами

ссылками на результаты работы сервиса визуализации подграфов семантической сети образовательной математической онтологии [19].

### 3.3. Сервис визуализации подграфов семантической сети образовательной математической онтологии

Сервис визуализации подграфов семантической сети принимает в качестве входных данных целевой концепт онтологии  $\text{OntoMath}^{\text{Edu}}$  и формирует подграф онтологии, состоящий из концептов, связанных с целевым концептом онтологическими отношениями [36]. В образовательном курсе в качестве целевых концептов выступают концепты, аннотированные сервисом семантического аннотирования. Информация о концептах и образовательных уровнях извлекается из онтологии  $\text{OntoMath}^{\text{Edu}}$  [28]. Сервис визуализации также предусматривает возможность фильтрации отображаемых понятий по образовательным уровням.

В качестве примера на Рис. 6 показана страница сервиса визуализации для концепта *Биссектриса треугольника*, в блоке «Определение» дано определение концепта из одного или нескольких учебников по геометрии, в блоке «Посмотреть на:» приведены ссылки на внешние ресурсы по текущему концепту, в блоке «Изучи сначала» – пререквизиты текущего понятия, а в блоке «Связанные понятия» – онтологически связанные концепты для текущего понятия.

**Биссектриса треугольника**

**Определение**  
Отрезок биссектрисы угла треугольника, соединяющий вершину треугольника с точкой противоположной стороны, называется биссектрисой треугольника. (Л.С. Атанасян)  
Отрезок биссектрисы угла треугольника от вершины до точки пересечения со стороной треугольника называется биссектрисой треугольника (И.Ф. Шарыгин)

**Посмотреть на:** [Mathvox](#), [Yaklass](#), [Wikipedia](#)

**Изучи сначала** [Биссектриса угла](#); [Треугольник](#)

**Связанные понятия** [Треугольник](#)

- ▲ [Чевиана треугольника](#)
- ▼ [Биссектриса внешнего угла треугольника](#); [Биссектриса внутреннего угла треугольника](#)

Рис. 6. Страница концепта *Биссектриса треугольника* в сервисе визуализации подграфов семантической сети.

Два нижних блока на Рис. 6 снабжены элементами управления в форме треугольников, которые позволяют производить навигацию по иерархии объектов онтологии (отражаются непосредственный надкласс текущего концепта и список его подклассов). Все концепты на странице сервиса представлены в виде гиперссылок, переводящих пользователя на страницы концептов.

### 3.4. Параллельный формальный/неформальный корпус математических утверждений

Данный ресурс содержит математические утверждения, представленные тремя способами: предложения на естественном языке, формулы в нотации LaTeX и формулы на языке OpenMath [20, 37].

Запись в базе данных корпуса содержит следующие поля:

1. математическое утверждение на естественном языке, извлеченное из учебных математических текстов;
2. представление этого утверждения в виде формулы в формате LaTeX;
3. формализация формулы в формате представления семантики математических объектов OpenMath (<https://openmath.org/>), где в качестве контентных словарей OpenMath (OpenMath content dictionaries) использована онтология OntoMath<sup>Edu</sup>.

На Рис. 7 представлен пример записи «Площадь прямоугольного треугольника равна половине произведения его катетов» из указанного корпуса.

<b>Утверждение на естественном языке</b>	Площадь прямоугольного треугольника равна половине произведения его катетов
<b>Визуализация утверждения в виде формулы</b>	$S_{ABC} = (a*b)/2$ , где ABC – прямоугольный треугольник, $S_{ABC}$ – площадь прямоугольного треугольника, a, b – катеты
<b>Представление утверждения в виде формулы (LaTeX-код)</b>	$S_{\{ABC\}} = \frac{a*b}{2} \text{ где } ABC - \text{прямоугольный треугольник, } S_{\{ABC\}} - \text{площадь прямоугольного треугольника, } a, b \text{ – катеты.}$

<p><b>Формализация в формате OpenMath (в визуальном виде)</b></p>	$\forall ABC_{\text{type:RightTriangle}}, a_{\text{type:LegOfTheRightTriangle}}, b_{\text{type:LegOfTheRightTriangle}}, S\_ABC_{\text{type:AreaOfTheRightTriangle}}$ $\text{isSideOfTriangle}(a, ABC) \wedge \text{isSideOfTriangle}(b, ABC) -$ $>$ $(a * b)/2 = S\_ABC$
---	--

Рис. 7. Пример записи из параллельного формального/неформального корпуса образовательных математических текстов.

В качестве исходных математических текстов были использованы учебные материалы по курсу планиметрии средней школы. Представления утверждений в форматах LaTeX и OpenMath построены вручную.

Сформированный корпус математических утверждений может быть использован в качестве тестовой коллекции при разработке методов автоматической формализации описаний формул на естественном языке в математических документах. На базе корпуса разработан сервис для автоматической генерации тестовых вопросов, описанный ниже. Также планируется применение корпуса математических утверждений в качестве источника данных для сервиса семантического поиска формул в образовательных приложениях.

### 3.5. Сервис автоматической генерации тестовых вопросов для проверки математических знаний

Данный сервис применяется для быстрого создания больших массивов заданий по геометрии общеобразовательного уровня. Тестовые задания с вводом ответа создаются на основе формализации OpenMath для формул из параллельного формально-неформального корпуса математических учебных материалов [21]. Формула на языке OpenMath представляется в виде дерева, включающего отношения между понятиями, включенными в формулу. Такая структура позволяет создавать задания и проверять ответы в автоматическом режиме.

Процесс генерации тестового задания состоит из следующих этапов:

1. Производится выбор случайной записи из корпуса, и из нее выделяется OpenMath-утверждение.
2. Строится дерево логической структуры утверждения, и в нем производится поиск элементов, соответствующих объектам формулы.

3. Из найденных элементов выбирается случайная искомая переменная, а для всех остальных элементов генерируется случайное значение в допустимых интервалах. Допустимые интервалы значений переменных определены экспертами и внедрены в OpenMath.
4. Формируется текст задания с указанием заданных и неизвестных значений объектов формулы. Обозначения переменных содержатся в формализации OpenMath.

Сгенерированное задание сохраняется в базе данных тестовых заданий для дальнейшего применения в образовательном процессе. На Рис. 8 приведен пример генерации задания. В строке «Формула на естественном языке» приведена текстовая формулировка формулы, на основе которой сформировано тестовое задание. Строка «Текст сгенерированного задания» содержит автоматически сгенерированный текст задания с указанием всех переменных, включенных в формулу, заданные значения, а также искомую переменную.

<b>Формула на естественном языке</b>	Площадь прямоугольного треугольника равна половине произведения его катетов
<b>Визуализация формулы</b>	$S_{ABC} = (a*b)/2$ , где ABC – прямоугольный треугольник, $S_{ABC}$ – площадь прямоугольного треугольника, a, b – катеты
<b>Текст сгенерированного задания</b>	Дано: Прямоугольный треугольник ABC; Катет a прямоугольного треугольника; Катет b прямоугольного треугольника; Площадь прямоугольного треугольника $S_{\triangle ABC}$ . $b = 10$ ; $S_{\triangle ABC} = 20$ . Найти: a.
<b>Формализация формулы в OpenMath (в визуальном виде)</b>	$\forall ABC_{\text{type:RightTriangle}}, a_{\text{type:LegOfTheRightTriangle}}, b_{\text{type:LegOfTheRightTriangle}}, S_{ABC_{\text{type:AreaOfTheRightTriangle}}}$ $\text{isSideOfTriangle}(a, ABC) \wedge \text{isSideOfTriangle}(b, ABC) \rightarrow$ $(a * b)/2 = S_{ABC}$

Рисунок 8. Пример генерации задания с помощью сервиса автоматической генерации тестовых вопросов для проверки математических знаний

## ЗАКЛЮЧЕНИЕ

Представлены результаты по разработке программных инструментов и сервисов, спроектированных с целью поддержки образовательного процесса в области математики. Эти программные инструменты являются семантическими, и их функционал основан на использовании онтологий, входящих в состав цифровой экосистемы OntoMath. Дальнейшее развитие образовательных сервисов связано с применением пререквизитных отношений онтологий OntoMath<sup>PRO</sup> и OntoMath<sup>Edu</sup> с целью формирования индивидуальных образовательных траекторий. Планируется разработка образовательной рекомендательной системы, предоставляющей доступ к разделам дополнительных образовательных источников (учебников, интернет-ресурсов) для уточнения или расширения описания понятия, обозначенного в запросе пользователя системы.

## Благодарности

Работа выполнена при финансовой поддержке Российского научного фонда (проект № 21-11-00105).

Авторы признательны А.М. Елизарову, А.В. Кирилловичу, Л.Р. Шакировой и М.В. Фалилеевой за обсуждение работы и замечания.

## СПИСОК ЛИТЕРАТУРЫ

1. Moodle – Open-source Learning Platform.  
URL: <https://moodle.org/?lang=ru>.
2. Технология решения планиметрических задач.  
<https://edu.kpfu.ru/course/view.php?id=2652>.
3. Муромцев Д. И. Модели и методы индивидуализации электронного обучения в контексте онтологического подхода // Онтологии проектирования. 2020. Т. 10. № 1. С. 34–49. <https://doi.org/10.18287/2223-9537-2020-10-1-34-49>.
4. Anderson J. Q. Individualisation of Higher Education: How Technological Evolution can Revolutionise Opportunities for Teaching and Learning // International Social Science Journal. 2013. V. 64. No. 213–214. P. 305–316.  
<https://doi.org/10.1111/issj.12046>.
5. Абрамский М. М., Батырова Э. Ф., Марданова А. Р., Ахметзянова Т.А. Генерация индивидуальных образовательных траекторий и расписания обучения

в парадигме индивидуализации образования // Электронные библиотеки. 2018. Т. 21. № 3–4. С. 129–145. URL: <https://rdl-journal.ru/article/view/466>.

6. *Jakus G., Milutinović V., Omerović S., Tomažič S.* Concepts, Ontologies, and Knowledge Representation. Series: SpringerBriefs in Computer Science. Springer New York, NY, 2013. <https://doi.org/10.1007/978-1-4614-7822-5>.

URL: <https://link.springer.com/book/10.1007/978-1-4614-7822-5>.

7. *Елизаров А. М., Кириллович А. В., Липачёв Е. К., Невзорова О. А., Шакурова Л. Р.* Семантические технологии в математическом образовании: онтологии и открытые связанные данные // Ученые записки ИСГЗ. 2018. Т. 16. № 1. С. 222–227.

8. *Tzoumpa D., Mitropoulos S.* Semantic Web Technologies for Ontologies Description: Case study in Geometry Education // In: 2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Corfu, Greece. 2020. P. 1–5.

<https://doi.org/10.1109/SEEDA-CECNSM49515.2020.9221781>.

9. *Tzoumpa D., Karvounidis T., Douligeris C.* Circle's Ontology Extended: Circumference and Surface Area of a Circle // Advances in Intelligent Systems and Computing. 2020. Vol. 916. P. 120–132.

10. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O.* Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management // L. Kalinichenko, S. Kuznetsov, and Y. Manolopoulos (Eds.). Revised Selected Papers of the XVIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016), Ershovo, Moscow, Russia, October 11–14, 2016. Communications in Computer and Information Science. Springer, 2017. Vol. 706. P. 33–46. [https://doi.org/10.1007/978-3-319-57135-5\\_3](https://doi.org/10.1007/978-3-319-57135-5_3).

11. *Elizarov A. M., Zhiltsov N. G., Kirillovich A. V., Lipachev E. K., Nevzorova O. A., Solovyev V. D.* The OntoMath ecosystem: Ontologies and applications for math knowledge management // Semantic Representation of Mathematical Knowledge Workshop, Fields Institute, Toronto, Canada, February 5, 2016.

URL: <https://video-archive.fields.utoronto.ca/view/4698>.

12. *Елизаров А. М., Липачев Е. К.* Цифровая библиотека Lobachevskii-DML в научном пространстве математических знаний // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2023. № 1. С. 32–37. <https://doi.org/10.36535/0548-0019-2023-01-3>.

13. *Елизаров А. М., Кириллович А. В., Липачёв Е. К., Невзорова О. А.* Цифровая экосистема OntoMath как подход к построению пространства математических знаний // *Электронные библиотеки*. 2023. Т. 26. № 2. С. 154–202.

<https://doi.org/10.26907/1562-5419-2023-26-2-154-202>.

URL: <https://rdl-journal.ru/article/view/771>.

14. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMath<sup>PRO</sup> Ontology: A Linked Data Hub for Mathematics // P. Klinov and D. Mouromstev (Eds.). Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014), Kazan, Russia, September 29 – October 1, 2014. Communications in Computer and Information Science. Springer, Cham, 2014. Vol. 468. P. 105–119.

[https://doi.org/10.1007/978-3-319-11716-4\\_9](https://doi.org/10.1007/978-3-319-11716-4_9).

15. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMath<sup>EDU</sup>: Towards an Educational Mathematical Ontology // E. Brady et al. (Eds.). Workshop Papers at 12th Conference on Intelligent Computer Mathematics (CICM-WS 2019), Prague, Czech Republic, 8–12 July 2019. CEUR Workshop Proceedings. CEUR-WS, 2020. Vol. 2634. URL: <https://ceur-ws.org/Vol-2634/WiP1.pdf>.

16. *Елизаров А. М., Кириллович А. В., Липачёв Е. К., Невзорова О. А.* Цифровая экосистема OntoMath: взаимодействие семантических сервисов и математических онтологий // *Научный сервис в сети Интернет: труды XXIV Всероссийской научной конференции (19–22 сентября 2022 г., онлайн)*. М.: ИПМ им. М.В. Келдыша, 2022. С. 167–178.

<https://doi.org/10.20948/abrau-2022-40>.

URL: <https://keldysh.ru/abrau/2022/theses/40.pdf>.

17. *Kirillovich A. V., Nevzorova O. A., Lipachev E. K.* OntoMathPRO 2.0 Ontology: Updates of Formal Model // *Lobachevskii J. Math.* 2022. Vol. 43. No 12. P. 3504–3514.

<https://doi.org/10.1134/S1995080222150136>.

URL: <https://link.springer.com/article/10.1134/S1995080222150136>.

18. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Dyupina A., Shakirova L.* Prerequisite Relationships of the OntoMath<sup>Edu</sup> Educational Mathematical Ontology // J.C. Figueroa-García et al. (Eds.). Proceedings of the 8th Workshop on Engineering Applications (WEA 2021), Medellín, Colombia, October 6–8, 2021. Communications in Computer and Information Science. Springer, 2021. Vol. 1431. P. 517–524.

19. *Николаев К. С., Невзорова О. А.* Метод автоматической семантической

разметки математических образовательных текстов // Информационные технологии в образовании и науке (ИТОН–2022) и II International Workshop “Digital Technologies for Teaching and Learning” (DTTL). Материалы III Международного форума по математическому образованию: Международной научно-практической конференции и II Международного научного семинара. Казань, 2022. С. 181–190.

20. *Kirillovich A., Nevzorova O., Nikolaev K., Galiaskarova K.* Towards a Parallel Informal-Formal Corpus of Educational Mathematical Texts in Russian // Zhengbing Hu et al. (Eds.). Proceedings of the 2019 International Symposium on Computer Science, Digital Economy and Intelligent Systems (CSDEIS 2019), Moscow, Russia, on 4–6 October 2019. Advances in Intelligent Systems and Computing. Springer, 2020. Vol. 1127. P. 325–334. [https://doi.org/10.1007/978-3-030-39216-1\\_29](https://doi.org/10.1007/978-3-030-39216-1_29).

21. *Nikolaev K., Kirillovich A., Nevzorova O.* A Corpus-Based Approach to Elementary Geometry Knowledge Test Generation // L. Gómez Chova et al. (Eds.). Proceedings of the 14th International Technology, Education and Development Conference (INTED 2020), Valencia, Spain, 2–4 March 2020. IATED, 2020. P. 6342–6348.

22. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMath<sup>PRO</sup> Ontology: A Linked Data Hub for Mathematics // P. Klinov and D. Mouromstev (Eds.). Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014), Kazan, Russia, September 29 – October 1, 2014. Communications in Computer and Information Science. Springer, Cham, 2014. Vol. 468. P. 105–119. [https://doi.org/10.1007/978-3-319-11716-4\\_9](https://doi.org/10.1007/978-3-319-11716-4_9).

23. *Елизаров А. М., Кириллович А. В., Липачёв Е. К., Невзорова О. А.* Онтология математического знания OntoMath<sup>PRO</sup> // Доклады Российской академии наук. Математика, информатика, процессы управления. 2022. Т. 507. № 1. С. 29–35. <https://doi.org/10.31857/S2686954322700011>.

24. *Nevzorova O. A., Falileeva M. V., Kirillovich A. V., Lipachev E. K., Shakirova L. R., Dyupina A. E.* OntoMathEdu Educational Ontology: Problems of Ontological Engineering // Pattern Recognition and Image Analysis. 2023. Vol. 33. No. 3. P. 460–466.

25. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMath<sup>EDU</sup>: Towards an Educational Mathematical Ontology // E. Brady et al. (Eds.). Workshop Papers at 12th Conference on Intelligent Computer Mathematics (CICM-WS 2019), Prague, Czech Republic, 8–12 July 2019. CEUR Workshop Proceedings. CEUR-WS, 2020. Vol. 2634. URL: <https://ceur-ws.org/Vol-2634/WiP1.pdf>.

26. Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L. OntoMath<sup>EDU</sup>: A Linguistically Grounded Educational Mathematical Ontology // C. Benz Müller and B. Miller (Eds.). Proceedings of the 13th International Conference on Intelligent Computer Mathematics (CICM 2020), Bertinoro, Italy, July 26–31, 2020. Lecture Notes in Computer Science. Springer, 2020. Vol. 12236. P. 157–172.

[https://doi.org/10.1007/978-3-030-53518-6\\_10](https://doi.org/10.1007/978-3-030-53518-6_10).

27. Loukachevitch N., Komissarov A., Dobrov B., Shternov S. Using Ontology for Natural Sciences and Technologies for Vacancies Analysis // CEUR Workshop Proceedings. CEUR-WS, 2021. Vol. 2910. P. 30–38.

URL: <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-2910/short2.pdf>.

28. Фалилеева М. В., Кириллович А. В., Невзорова О. А., Шакирова Л. Р., Липачёв Е. К., Дюпина А. Э. Системы образовательных проекций, уровней и пререквизитов математической онтологии OntoMath<sup>EDU</sup> // Электронные библиотеки. 2021. Т. 24. № 3. С. 505–530. <https://doi.org/10.26907/1562-5419-2021-24-3-505-530>. URL: <https://rdl-journal.ru/article/view/691>.

29. Fréchet M. Sur les fonctionelles continues // Acc. Ec. Norm. 1910. 27 (3). P. 193–216. URL: [http://www.numdam.org/article/ASENS\\_1910\\_3\\_27\\_\\_193\\_0.pdf](http://www.numdam.org/article/ASENS_1910_3_27__193_0.pdf).

30. Gateaux R. Sur les fonctionnelles continues et les fonctionnelles analytiques // Comptes Rendus Acad. Sci. Paris. Sér. I. Math. 1913. Vol. 157. P. 325–327. URL: <https://gallica.bnf.fr/ark:/12148/bpt6k31103/f325.image>.

31. Канторович Л. В., Акилов Г. П. Функциональный анализ. М.: Наука. Гл. ред. физ.-мат. лит., 1984. 752 с.

32. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. М.: Наука. Гл. ред. физ.-мат. лит., 1989. 624 с.

33. OntoMath Semantic Formula Search.

URL: <https://lobachevskii-dml.ru/mathsearch>.

34. Elizarov A., Kirillovich A., Lipachev E., Nevzorova O. Semantic Formula Search in Digital Mathematical Libraries // Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017), Vladivostok, Russia, 25–29 September, 2017. IEEE, 2017. P. 39–43.

<https://doi.org/10.1109/RPC.2017.8168063>.

35. MatJax. Beautiful and accessible math in all browsers.

URL: <https://www.mathjax.org/>.

36. Николаев К. С. Сервис генерации учебных карточек математических понятий для дистанционного курса по геометрии // Электронные библиотеки. 2023. Т. 26. № 3. С. 365–377. <https://doi.org/10.26907/1562-5419-2023-26-3-365-377>.

URL: <https://rdl-journal.ru/article/view/785>.

37. The OpenMath Standards. URL: <https://openmath.org/standard/>.

---

## SEMANTIC SERVICES OF THE DIGITAL ECOSYSTEM ONTOMATH FOR MATHEMATICAL EDUCATION

O. A. Nevzorova<sup>1</sup> [0000-0001-8116-9446], E. K. Lipachev<sup>2</sup> [0000-0001-7789-2332],

K. S. Nikolaev<sup>3</sup> [0000-0003-3204-238X]

<sup>1-3</sup> Kazan Federal University, ul. Kremlyovskaya, 35, Kazan, 420008

<sup>3</sup> Joint Supercomputer Center of the Russian Academy of Sciences, Kazan, Russia

<sup>1</sup>onevzoro@gmail.com, <sup>2</sup>elipachev@gmail.com, <sup>3</sup>konnikolaeff@yandex.ru,

### Abstract

We present a set of semantic services developed by us to support the educational process in mathematics. The functionality of these services is based on the use of mathematical ontologies OntoMath<sup>Edu</sup> and OntoMath<sup>PRO</sup>. The ontology of professional mathematical knowledge OntoMath<sup>PRO</sup> is designed to classify and systematize the concepts of professional mathematics and includes several important areas of mathematics. Educational mathematical ontology OntoMath<sup>Edu</sup> systematically represents knowledge on the training course “Planimetry”. For the use of ontologies in educational applications, an approach to the design of prerequisite relations in these ontologies has been developed. To support mathematical education, we have developed: a service for semantic search by mathematical formulas, a service for semantic annotation of educational materials, a service for visualizing subgraphs of the OntoMath<sup>Edu</sup> ontology semantic network, a parallel formal/informal corpus of mathematical statements, a system for automatically generating test questions in mathematical disciplines.

We provide examples of successful application of the developed software tools.

---

The created software tools are built into the OntoMath digital ecosystem. This ecosystem implements the interaction of semantic services for managing mathematical knowledge.

**Keywords:** *Digital ecosystem, OntoMath ecosystem, ontology, OntoMath<sup>PRO</sup> mathematical ontology, OntoMath<sup>Edu</sup> educational ontology, prerequisite relation, semantic service.*

## REFERENCES

1. Moodle – Open-source Learning Platform.  
URL: <https://moodle.org/?lang=ru>.
2. Tekhnologiya resheniya planimetriceskikh zadach.  
<https://edu.kpfu.ru/course/view.php?id=2652>.
3. *Mouromtsev D.* Models and methods of e-learning individualization in the context of ontological approach // *Ontology of Designing*. 2020. Vol. 10. No. 1. P. 34–49.  
<https://doi.org/10.18287/2223-9537-2020-10-1-34-49>.
4. *Anderson J. Q.* Individualisation of Higher Education: How Technological Evolution can Revolutionise Opportunities for Teaching and Learning // *International Social Science Journal*. 2013. V. 64. No. 213–214. P.305–316.  
<https://doi.org/10.1111/issj.12046>.
5. *Abramskiy M. M., Batyrova E. F., Mardanova A. R., Akhmetzyanova T. A.* Generation of Individual Educational Routes and Learning Schedule in Individualization Paradigm // *Russian Digital Libraries Journal*. 2018. V. 21. No. 3–4. P.129–145.  
URL: <https://rdl-journal.ru/article/view/466>.
6. *Jakus G., Milutinović V., Omerović S., Tomažič S.* Concepts, Ontologies, and Knowledge Representation. Series: SpringerBriefs in Computer Science. Springer New York, 2013. <https://doi.org/10.1007/978-1-4614-7822-5>.  
URL: <https://link.springer.com/book/10.1007/978-1-4614-7822-5>.
7. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O., Shakirova L.* Semanticheskie tekhnologii v matematicheskom obrazovanii: ontologii i otkrytye svyazannye dannye // *Uchenye zapiski ISGZ*. 2018. T. 16. № 1. S. 222–227.
8. *Tzoumpa D., Mitropoulos S.* Semantic Web Technologies for Ontologies Description: Case study in Geometry Education // In: 2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference

(SEEDA-CECNSM), Corfu, Greece. 2020. P. 1–5.

<https://doi.org/10.1109/SEEDA-CECNSM49515.2020.9221781>.

9. *Tzoumpa D., Karvounidis T., Douligeris C.* Circle's Ontology Extended: Circumference and Surface Area of a Circle // *Advances in Intelligent Systems and Computing*. 2020. Vol. 916. P. 120–132.

10. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O.* Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management // L. Kalinichenko, S. Kuznetsov, and Y. Manolopoulos (Eds.). *Revised Selected Papers of the XVIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016)*, Ershovo, Moscow, Russia, October 11–14, 2016. *Communications in Computer and Information Science*. Springer, 2017. Vol. 706. P. 33–46. [https://doi.org/10.1007/978-3-319-57135-5\\_3](https://doi.org/10.1007/978-3-319-57135-5_3).

11. *Elizarov A. M., Zhiltsov N. G., Kirillovich A. V., Lipachev E. K., Nevzorova O. A., Solovyev V. D.* The OntoMath ecosystem: Ontologies and applications for math knowledge management // *Semantic Representation of Mathematical Knowledge Workshop*, Fields Institute, Toronto, Canada, February 5, 2016.

URL: <https://video-archive.fields.utoronto.ca/view/4698>.

12. *Elizarov A. M., Lipachev E. K.* Lobachevskii Digital Library in the Scientific Space of Mathematical Knowledge // *Scientific and Technical Information Processing*. 2023. Vol. 50. No. 1. P. 35–39. <https://doi.org/10.3103/S0147688223010021>.

13. *Elizarov A. M., Kirillovich A. V., Lipachev E. K., Nevzorova O. A.* Digital Ecosystem OntoMath as an Approach to Building the Space of Mathematical Knowledge // *Russian Digital Libraries Journal*. 2023. Vol. 26. No 2. P. 154–202.

<https://doi.org/10.26907/1562-5419-2023-26-2-154-202>.

URL: <https://rdl-journal.ru/article/view/771>.

14. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMath<sup>PRO</sup> Ontology: A Linked Data Hub for Mathematics // P. Klinov and D. Mouromstev (Eds.). *Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014)*, Kazan, Russia, September 29 – October 1, 2014. *Communications in Computer and Information Science*. Springer, Cham, 2014. Vol. 468. P. 105–119.

[https://doi.org/10.1007/978-3-319-11716-4\\_9](https://doi.org/10.1007/978-3-319-11716-4_9).

15. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMath<sup>EDU</sup>: Towards an Educational Mathematical Ontology // E. Brady et al. (Eds.). *Workshop*

Papers at 12th Conference on Intelligent Computer Mathematics (CICM-WS 2019), Prague, Czech Republic, 8–12 July 2019. CEUR Workshop Proceedings. CEUR-WS, 2020. Vol. 2634. URL: <https://ceur-ws.org/Vol-2634/WiP1.pdf>.

16. *Elizarov A. M., Kirillovich A. V., Lipachev E. K., Nevzorova O. A.* OntoMath Digital Ecosystem: Semantic Service and Mathematical Ontologies // *Nauchnyj servis v seti Internet: trudy XXIV Vserossijskoj nauchnoj konferencii (19–22 sentyabrya 2022 g., onlajn)*. M.: IPM im. M.V.Keldysha, 2022. S. 167–178.

<https://doi.org/10.20948/abrau-2022-40>.

URL: <https://keldysh.ru/abrau/2022/theses/40.pdf>.

17. *Kirillovich A. V., Nevzorova O. A., Lipachev E. K.* OntoMathPRO 2.0 Ontology: Updates of Formal Model // *Lobachevskii J. Math.* 2022. Vol. 43. P. 3504–3514 (2022). <https://doi.org/10.1134/S1995080222150136>.

URL: <https://link.springer.com/article/10.1134/S1995080222150136>.

18. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Dyupina A., and Shakirova L.* Prerequisite Relationships of the OntoMath<sup>Edu</sup> Educational Mathematical Ontology // *J.C. Figueroa-García et al. (Eds.). Proceedings of the 8th Workshop on Engineering Applications (WEA 2021), Medellín, Colombia, October 6–8, 2021. Communications in Computer and Information Science. Springer, 2021. Vol. 1431. P. 517–524.*

19. *Nikolaev K. S., Nevzorova O. A.* Metod avtomaticheskoy semanticheskoy razmetki matematicheskikh obrazovatel'nyh tekstov // *Informacionnye tekhnologii v obrazovanii i nauke (ITON–2022) i II International Workshop “Digital Technologies for Teaching and Learning” (DTTL). Materialy III Mezhdunarodnogo foruma po matematicheskomu obrazovaniyu: Mezhdunarodnoj nauchno-prakticheskoy konferencii i II Mezhdunarodnogo nauchnogo seminara. Kazan', 2022. S. 181–190.*

20. *Kirillovich A., Nevzorova O., Nikolaev K., Galiaskarova K.* Towards a Parallel Informal-Formal Corpus of Educational Mathematical Texts in Russian // *Zhengbing Hu et al. (Eds.). Proceedings of the 2019 International Symposium on Computer Science, Digital Economy and Intelligent Systems (CSDEIS 2019), Moscow, Russia, on 4–6 October 2019. Advances in Intelligent Systems and Computing. Springer, 2020. Vol. 1127. P. 325–334. [https://doi.org/10.1007/978-3-030-39216-1\\_29](https://doi.org/10.1007/978-3-030-39216-1_29).*

21. *Nikolaev K., Kirillovich A., Nevzorova O.* A Corpus-Based Approach to Elementary Geometry Knowledge Test Generation // *L. Gómez Chova et al. (Eds.). Proceedings of the 14th International Technology, Education and Development Conference*

(INTED 2020), Valencia, Spain, 2–4 March 2020. IATED, 2020. P. 6342–6348.

22. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMath<sup>PRO</sup> Ontology: A Linked Data Hub for Mathematics // P. Klinov and D. Mouromstev (Eds.). Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014), Kazan, Russia, September 29 – October 1, 2014. Communications in Computer and Information Science. Springer, Cham, 2014. Vol. 468. P. 105–119.

[https://doi.org/10.1007/978-3-319-11716-4\\_9](https://doi.org/10.1007/978-3-319-11716-4_9).

23. *Elizarov A. M., Kirillovich A. V., Lipachev E. K., Nevzorova O. A.* OntoMath<sup>PRO</sup>: An Ontology of Mathematical Knowledge // Doklady Mathematics. 2022. Vol. 106. No. 3. S. 429–435. <https://doi.org/10.1134/S1064562422700016>.

24. *Nevzorova O. A., Falileeva M. V., Kirillovich A. V., Lipachev E. K., Shakirova L. R., Dyupina A. E.* OntoMathEdu Educational Ontology: Problems of Ontological Engineering // Pattern Recognition and Image Analysis. 2023. Vol. 33. No. 3. P. 460–466.

25. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMath<sup>EDU</sup>: Towards an Educational Mathematical Ontology // E. Brady et al. (Eds.). Workshop Papers at 12th Conference on Intelligent Computer Mathematics (CICM-WS 2019), Prague, Czech Republic, 8–12 July 2019. CEUR Workshop Proceedings. CEUR-WS, 2020. Vol. 2634. URL: <https://ceur-ws.org/Vol-2634/WiP1.pdf>.

26. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMath<sup>EDU</sup>: A Linguistically Grounded Educational Mathematical Ontology // C. Benzmüller and B. Miller (Eds.). Proceedings of the 13th International Conference on Intelligent Computer Mathematics (CICM 2020), Bertinoro, Italy, July 26–31, 2020. Lecture Notes in Computer Science. Springer, 2020. Vol. 12236. P. 157–172.

[https://doi.org/10.1007/978-3-030-53518-6\\_10](https://doi.org/10.1007/978-3-030-53518-6_10).

27. *Loukachevitch N., Komissarov A., Dobrov B., Shternov S.* Using Ontology for Natural Sciences and Technologies for Vacancies Analysis // CEUR Workshop Proceedings. CEUR-WS, 2021. Vol. 2910. P. 30–38. URL: <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-2910/short2.pdf>.

28. *Falileeva M. V., Kirillovich A. V., Nevzorova O. A., Shakirova L. R., Lipachev E. K., Dyupina A. E.* Educational Projection Systems, Levels and Prerequisites of Mathematical Ontology OntoMath<sup>EDU</sup> // Russian Digital Libraries Journal. 2021. Vol. 24. No. 3. P. 505–530. <https://doi.org/10.26907/1562-5419-2021-24-3-505-530>.

URL: <https://rdl-journal.ru/article/view/691>.

---

29. *Fréchet M.* Sur les fonctionelles continues // *Acc. Ec. Norm.* 1910. Vol. 27, No. 3. P. 193–216.

URL: [http://www.numdam.org/article/ASENS\\_1910\\_3\\_27\\_\\_193\\_0.pdf](http://www.numdam.org/article/ASENS_1910_3_27__193_0.pdf).

30. *Gateaux R.* Sur les fonctionnelles continues et les fonctionnelles analytiques // *C.R. Acad. Sci. Paris Sér. I Math.* 1913. Vol. 157. P. 325–327.

URL: <https://gallica.bnf.fr/ark:/12148/bpt6k31103/f325.image>.

31. *Kantorovich L. V., Akilov G. P.* *Funkcional'nyj analiz.* M.: Nauka. Gl. red. fiz.-mat. lit., 1984. 752 s.

32. *Kolmogorov A. N., Fomin S. V.* *Elementy teorii funkcij i funkcional'nogo analiza.* M.: Nauka. Gl. red. fiz.-mat. lit., 1989. 624 s.

33. *OntoMath Semantic Formula Search.*

URL: <https://lobachevskii-dml.ru/mathsearch>.

34. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O.* *Semantic Formula Search in Digital Mathematical Libraries* // *Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017), Vladivostok, Russia, 25–29 September, 2017.* IEEE, 2017. P. 39–43.

<https://doi.org/10.1109/RPC.2017.8168063>.

35. *MatJax.* Beautiful and accessible math in all browsers.

URL: <https://www.mathjax.org/>.

36. *Nikolaev K. S.* *Service for Generating Educational Cards of Mathematical Concepts for a Distance Learning Course in Geometry* // *Russian Digital Libraries Journal.* 2023. Vol. 26. No. 3. P. 365–377.

<https://doi.org/10.26907/1562-5419-2023-26-3-365-377>.

URL: <https://rdl-journal.ru/article/view/785>.

37. *The OpenMath Standards.* URL: <https://openmath.org/standard/>.

## СВЕДЕНИЯ ОБ АВТОРАХ



**НЕВЗОРОВА Ольга Авенировна** – кандидат технических наук, доцент кафедры информационных систем Института вычислительной математики и информационных технологий Казанского федерального университета.

Научные интересы: онтологическое моделирование, Semantic Web, лингвистические открытые связанные данные, управление математическим знанием.

**Olga NEVZOROVA** – Candidate of Technical Sciences, Associate Professor of the Information Systems Department of the Institute of Computational Mathematics and Information Technologies of the Kazan Federal University.

Research interests: ontology engineering, Semantic Web, Linguistic Linked Data, mathematical knowledge management.

email: onevzoro@gmail.com.

ORCID: 0000-0001-8116-9446.



**ЛИПАЧЁВ Евгений Константинович** – кандидат физико-математических наук, доцент кафедры цифровой аналитики и технологий искусственного интеллекта Института информационных технологий и интеллектуальных систем Казанского федерального университета.

Научные интересы: цифровые библиотеки, единое пространство научных математических знаний, интеллектуальный анализ данных, рекомендательные системы, облачные вычисления, технологии извлечения знаний.

**Evgeny Konstantinovich LIPACHEV** – Candidate of Physics and Mathematics, Associate Professor, Kazan Federal University.

Research interests: digital libraries, common space of scientific mathematical knowledge, data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: elipachev@gmail.com;

ORCID: 0000-0001-7789-2332



**НИКОЛАЕВ Константин Сергеевич** – ассистент кафедры системного анализа и информационных технологий Института вычислительной математики и информационных технологий Казанского федерального университета.

Основные направления научных исследований: обработка естественного языка, искусственный интеллект.

**Konstantin Sergeevich NIKOLAEV** – Assistant of the Department of System Analysis and Information Technologies of the Institute of Computational Mathematics and Information Technologies of Kazan Federal University.

Major fields of scientific research are Natural Language processing, artificial intelligence.

email: [konnikolaeff@yandex.ru](mailto:konnikolaeff@yandex.ru)

ORCID: 0000-0003-3204-238X

*Материал поступил в редакцию 25 февраля 2023 года*

*Переработанная версия – 16 мая 2023 года*

## «ВЫЧИСЛИТЕЛЬНОЕ МЫШЛЕНИЕ» КАК ВТОРАЯ ГРАМОТНОСТЬ В СОВРЕМЕННОМ ЦИФРОВОМ ОБРАЗОВАНИИ И ОБЩЕСТВЕ

Т. Р. Файзрахманов<sup>1</sup> [0000-0001-5013-4523]

<sup>1</sup> *Институт разработки программного обеспечения и инжиниринга,  
Университет Иннополис, ул. Университетская, 1, г. Иннополис, Республика  
Татарстан 420500*

<sup>1</sup> tim.fayzrakhmanov@gmail.com

### **Аннотация**

В ходе невероятного роста популярности цифровых технологий и цифрового образования появилось такое понятие, как «вычислительное мышление». Его резкий спрос и популяризация, как ответ на современные нужды цифрового мира, послужили реструктуризации образования и возникновению большого количества литературы на данную тему. Не всегда ясно, чем же является вычислительное мышление, какие аспекты оно включает и с чем его сравнивать. В данной работе мы рассмотрели смысл данного понятия, его значение в современном цифровом образовании и провели аналогию наличия вычислительного мышления с навыком написанием обычного текста.

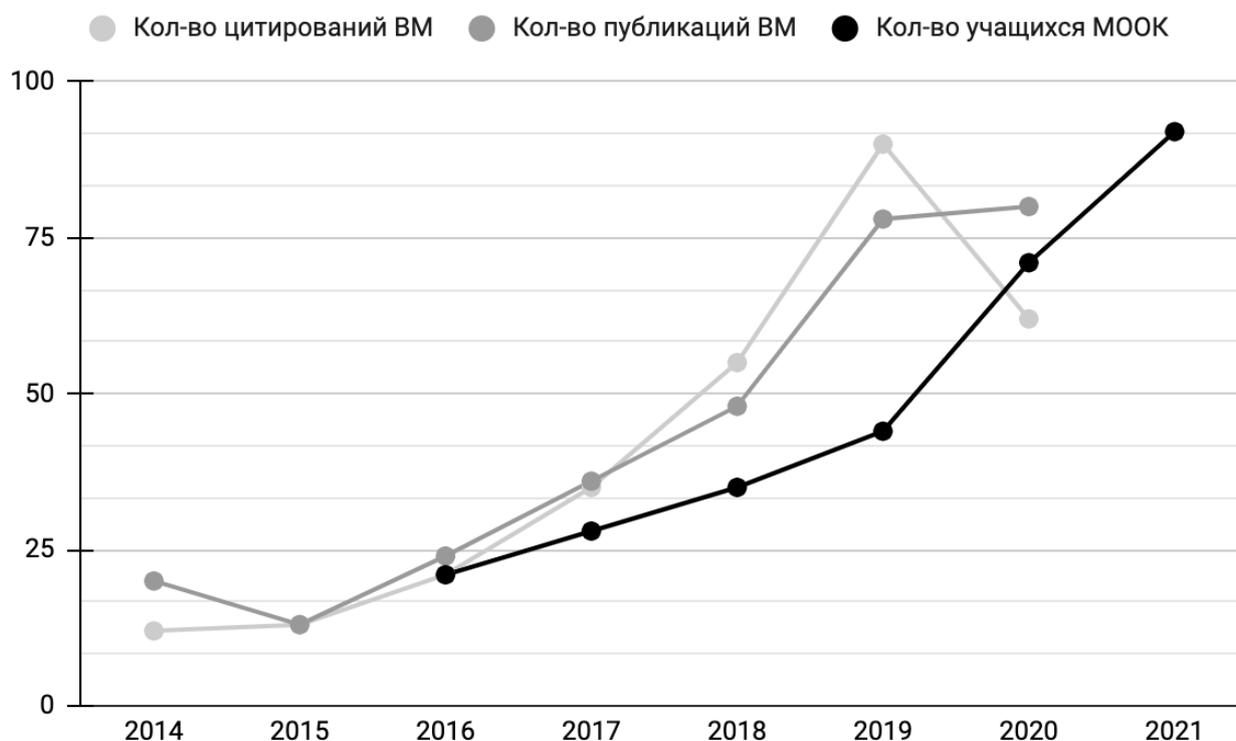
**Ключевые слова:** *вычислительное мышление, цифровое образование, кодинг, программирование, информатика, STEM.*

### **ВВЕДЕНИЕ**

Цифровое образование становится одним из ключевых факторов развития экономики из-за широкого распространения автоматизации, цифровых технологий, интернета и ограничений в связи с пандемией. По данным [1] и [2], с 2011 по 2021 годы число учащихся, охваченных массовыми открытыми онлайн-курсами (МООК), увеличилось с 58 тысяч до 170 миллионов и продолжает увеличиваться в среднем на 10–40 миллионов в год [3] внутри лишь одной платформы Coursera.

Параллельно с популярностью цифрового образования и развития «цифровизации» в целом наблюдается рост популярности другого, теперь уже смежного понятия – *вычислительное мышление*. На Рисунке 1, построенном по данным [3]

и [4], можно заметить, что экспоненциальный график роста популярности цифрового образования коррелирует с ростом количества научных статей и цитирований на данную тему.



Количество публикаций на тему Вычислительного мышления (ВМ): 100% = 100 статей;  
Количество цитирований статей на тему ВМ: 100% = 1500 цитирований;  
Количество учащихся в массовых открытых онлайн-курсах: 100% = 100 миллионов;  
Измерительные точки указаны приблизительно. Источники: [3], [4]

Рис. 1. Популярность цифрового образования и понятия «вычислительное мышление»

Вычислительное мышление является подмножеством (а иногда и самим множеством) более широкого понятия – *цифровая грамотность*. У цифровой грамотности имеется множество синонимов:

- Компьютерная грамотность (компетентность, навыки);
- Цифровая грамотность;
- Информационно-технологическая грамотность;
- Вычислительная грамотность;

- Вычислительное мышление (BM) или *Computational thinking* (CT).

В ежегодном докладе «Global Skills Report» от Coursera Inc. [5], на основе данных от более 100 миллионов учащихся в более чем 100 странах, отмечено, что цифровая грамотность («digital skills») стала одним из двух основных навыков *вне зависимости* от конкретной области знаний, где второй, комплементарной грамотностью, стала общая/человеческая грамотность («human skills»).

В связи с большой популярностью и значительным количеством пересекающихся по смыслам слов и понятий, представленных выше, мы будем придерживаться следующего разделения: цифровое обучение включает в себя цифровую грамотность, которая является синонимом компьютерной грамотности. В свою очередь, цифровая грамотность включает в себя (или является) *информационно-технологической грамотностью*, где технологическая часть включает в себя вычислительное мышление.

На основе обобщений в кратком глоссарии Coursera Inc. [5] и данных ChatGPT от OpenAI Inc. [6] мы построили простую *таксономию понятий современного цифрового образования*, где разделение понятий, перечисленных выше, можно увидеть наглядно (Рисунок 2).

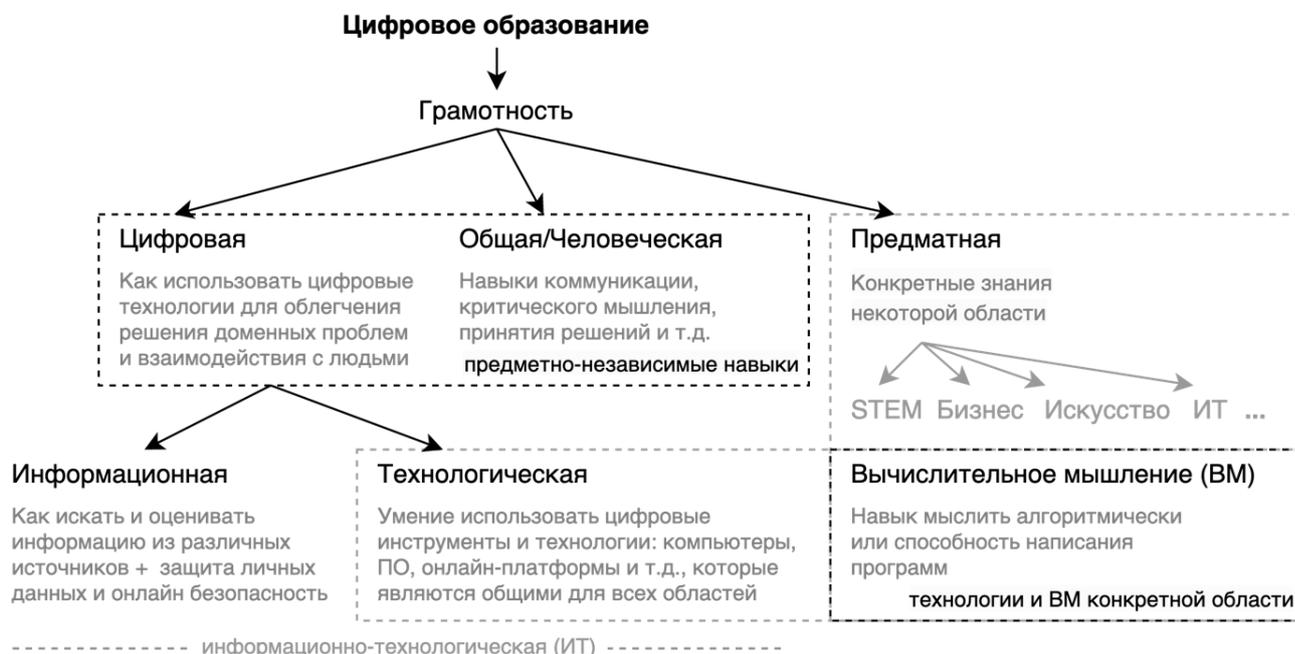


Рис. 2. Таксономия современного цифрового образования и грамотности

В данной таксономии мы видим, что цифровое образование включает в себя *доменную* и *доменно-независимую* грамотности. Доменная грамотность вполне очевидна – это навыки и терминология, необходимые для функционирования внутри некой области (естественные науки, математика, инженерия и т. д.). Доменно-независимая часть условно делится на «хард» и «софт скилы» – *человеческая грамотность* (например, «критическое мышление» или навыки коммуникации) и упомянутая ранее цифровая грамотность: *умение работать с информацией* (включая её безопасность и безопасность персональных данных) плюс работа с *цифровыми инструментами* (например, компьютером, текстовым редактором, браузером, онлайн сервисами и т. д.). Два последних объединяются в информационно-технологическую (ИТ) грамотность.

Мы хотим подчеркнуть, что «ИТ» как домен и «ИТ» как грамотность – разные вещи. Как домен, ИТ может включать в себя теоретическую информатику, профессиональную разработку ПО, машинное обучение и т. д. В то же время ИТ как грамотность – это скорее общие навыки «работы с компьютером» как для решения повседневных задач, так и для нужд автоматизации или быстрой работы с информацией внутри некой узкой области без перехода в индустриальную разработку ПО.

Фокусом данной работы является вычислительное мышление, которое мы поместили на Рисунке 2 в правый нижний угол. Мы сделали его частью технологической грамотности, т. к. появление и смысл этого понятия сложно представить в отрыве от цифровых технологий. В то же время, вычислительное мышление является и частью доменных компетенций, т. к. преподавание вычислительного мышления в отрыве от задач, решению которых оно способствует, также не имеет смысла.

Итак, на Рисунке 2 можно увидеть, что современная цифровая грамотность – это 25% навыков от всех имеющихся, а если допустить, что для коммуникации с людьми и навыками критического мышления мы овладеваем «по ходу жизни» и не проходим для этого отдельных курсов обучения, то цифровая грамотность и вовсе становится половиной современного набора компетенции.

Таким образом, можно заключить, что вычислительное мышление является ядром как информационно-технологической грамотности, так и тренда цифровой

---

грамотности в целом. Об этом, в том числе, свидетельствует повсеместная интеграция базовых понятий информатики под эгидой «Computational thinking» (Вычислительное мышление) в учебные программы обязательного школьного и дошкольного образования (K-12) западных стран [7], [8] (куда входят 29 европейских стран, США, Австралия, и Великобритания), в то время как преподавание вычислительного мышления в России и СССР давно известно под предметом «Информатика», который как правило преподается с 7-го по 11 классы. В современных же образовательных стандартах [9] присутствие информатики (именуемой как «Математика и информатика») предполагается с 1-го по 11 классы.

Итак, чем же является вычислительное мышление? На основе различных источников, куда вошли: анализ определений вычислительного мышления [10], [11], истории развития данного понятия [12], докладов Объединенного Исследовательского Центра Европейской Комиссии «European Commission’s Joint Research Centre» (JRC) [7], [13], руководств Британского сообщества для учителей информатики «Computing At School» (CAS) [14], [15] и международной образовательной программы по информатике для школьного образования «K–12 Computer Science Framework» [16] (под руководством бесчисленного количества ассоциаций, корпораций и образовательных организаций по всему миру [17]), мы выделили два простых определения. Первое, по нашему мнению, – наиболее часто встречающееся, совпадающее по смыслу с другими и просто лаконичное определение. Второе – неформальная версия первого:

1. **«Вычислительное мышление — это мыслительные процессы, связанные с выражением решений в виде вычислительных шагов или алгоритмов, которые могут быть выполнены компьютером»** [16, р. 68] или, как обобщение того, что вычисления доступны и человеку, «... выполнены человеком или машиной» [7, р. 24] (перевод).
2. **«Вычислительное мышление — это набор умственных навыков, которые преобразуют сложные, беспорядочные, частично определенные проблемы реального мира в форму, которую бездумный компьютер может решить без дополнительной помощи со стороны человека»** [15, р. 5] (перевод).

В ходе поисков были найдены и намного более обобщённые и, по нашему мнению, излишне размытые или «всеохватывающие» определения, которые, например, подчеркивают тот факт, что вычисления не являются исключительно человеческой конструкцией, но также присутствуют и в природе, например, в ДНК [13]:

*«Вычислительное мышление — это процесс распознавания аспектов вычислений в мире, который нас окружает, и применения инструментов и методов из компьютерных наук для раскрытия их сущности» [18, р. 29] (перевод).*

Также, совместно с попытками определить, что же такое вычислительное мышление, имеется и большое количество попыток определить его *аспекты* (также именуемые как «принципы», «концепции», «возможности» вычислительного мышления), включая использование различных способов их категоризации. Например, выделяют аспекты, связанные с методами работы с данными, моделированием и симуляцией, системным мышлением и т. д. [19, р. 10]. Дабы дать общее представление о наиболее часто встречающихся аспектах («ключевых словах»), ассоциируемых в литературе с вычислительным мышлением [13, р. 17], [20, р. 52], и [21, р. 14], приведем их в виде следующего списка:

1. **Абстракция.** Алгоритмы и данные как упрощённая версии явлений реального мира; решение сложных проблем путем отбрасывания ненужных деталей, сосредотачиваясь на сути.
2. **Алгоритмы.** Решение проблем через последовательность инструкций и промежуточных состояний, чтобы достигнуть конечное целевое состояние.
3. **Автоматизация.** Программы и код для выполнения задач с помощью компьютеров.
4. **Коллаборация, Параллелизм.** Разделение работы, задач и ролей с другими; использование готовых решений; написание программ с использованием возможностей параллельного вычисления.
5. **Креативность.** Использование «нестандартного мышления», творческого самовыражения и альтернативных подходов к решению проблем.
6. **Данные.** Поиск, сбор и обработка данных из различных источников; работа

- с типами данных; организация структур данных.
7. **Эффективность.** В алгоритмах нет лишних или ненужных шагов; разработанные решения просты в использовании.
  8. **Итерация.** Решения и совершенствование результатов путем нескольких/повторных подходов.
  9. **Логика, Анализ.** Аналитический склад ума и проверка гипотез / оптимальных решений перед началом написания программ.
  10. **Моделирование и дизайн.** Разработка систем, их внешнего вида и функциональности, которые впоследствии могут быть запрограммированы.
  11. **Паттерны и обобщение.** Поиск повторяющихся паттернов и формирование решений общего уровня, которые применимы к классу схожих проблем.
  12. **Декомпозиция проблемы.** Большие проблемы и артефакты распадаются на более мелкие и простые части, которые можно изучать по отдельности.
  13. **Тестирование и отладка.** Разработанные решения функционируют правильно, в соответствии с желаемым результатом; в обратном случае предполагается применение эвристики, формальных методов для выявления ошибок и их исправления.

Мы считаем, что данные аспекты излишне размыты, пересекаются друг с другом, с другими дисциплинами и вовсе относятся к *мышлению* в целом.

Склонность к всеохватываемости «всё есть вычисления» или «вычисления вокруг нас», а также попытки трактовать любые когнитивные проявления как часть вычислительного мышления объясняются невероятным всплеском (или даже «бумом») автоматизации, популярности разработки ПО, интеграции машинного обучения во всевозможные дисциплины и (о чем мы уже упомянули ранее) реструктуризации образовательных процессов в школах и ВУЗах для подготовки учащихся к неизбежно надвигающемуся спросу на навыки «компьютерных наук». Этот спрос можно легко увидеть, взглянув на данные популярного сервиса по поиску работы Glassdoor Inc. В его отчете «50 лучших рабочих мест в Америке на 2022 год» [22] 6 профессий из топ 10 требуют навыки *кодинга*.

В этом контексте не будет преувеличением сказать, что вычислительное мышление становится в некотором роде «*второй грамотностью*», как это было

предсказано академиком А.П. Ершовым в докладе «Программирование – вторая грамотность» еще в 1981 году [23]. В современном же мире подобное отношение проявляется в таких заголовках, как «Вычислительное мышление: Навык цифрового века для каждого» [24] или, например, сопровождается такими цитатами, как:

*«Независимо от того, хотят ли наши дети стать фермерами, врачами, учителями или предпринимателями, им будет легче осуществить свои мечты в цифровую эпоху, если они будут иметь некоторое образование в области информатики. Нам нужно, чтобы наши дети приобрели навыки 21-го века для мира 21-го века, а коддинг учит их творчеству и умению решать проблемы, которые необходимы для успеха»* (перевод)

— Джон Тун, Сенатор США, Южная Дакота [25].

Можно заметить, что до данного момента мы использовали вычислительное мышление, программирование и коддинг как понятия, *взаимно заменяющие друг друга*. Однако, в соответствии с Планом действий по цифровому образованию (2021–2027), политической инициативы Европейского союза [26, р. 95], рекомендуется следующее разделение понятий (перевод):

- *Программирование* – это деятельность по анализу проблемы, дизайна решения и его практической имплементации.
- *Коддинг* – имплементация решений на определенном языке программирования.
- *Вычислительное мышление* – мышление как у «программиста», или способность понимать базовые понятия и механизмы цифровых технологий для формулирования и решения проблем.

Посыл же нашей работы заключается в том, чтобы свести суть вычислительного мышления, которая представлена в формальных образовательных стандартах и рекомендациях как некий абстрактный способ мышления с бесчисленными

теоретическими аспектами, такими как «абстракция», «автоматизация», «декомпозиция» и т. д. [13, p. 18] (о которых сами применяющие данное мышление возможно не догадываются), к простой и соответствующей практике **способности написания программ**. Кодинг в данном случае используется лишь как синоним программирования (где программирование, снова, – это способность написания программ) или интерпретируется как его *упрощенная версия* (например, «скриптинг» или написание цепочки простых команд *в виде кода* внутри некоторой вычислительной среды, например, MATLAB [27]).

Мы считаем, что способность писать программы (например, посредством некоторого языка программирования, такого как Python [28] или его визуального аналога, такого как Scratch [29]) – это и есть то, что в кульминации приводит к так называемому «алгоритмическому мышлению» или *дисциплине мышления* в виде четкой и ясной последовательности действий, а не, соответственно, хаотичного и неструктурированного потока. Также это является тем самым «драйвером», что движет экономику и достижения различных областей наук и жизни, о чем многократно говорят их самые разные представители [25]. Иными словами, вычислительное мышление без умения писать код/программы (или строить блочные диаграммы в средах визуального программирования), чтобы делать что-то полезное посредством компьютера, вряд ли привлекло бы такое большое внимание в мире или, в частности, в образовательной среде.

В заключение подробнее рассмотрим, чем же являются навык, грамотность или способность написания программ, определенная выше как основа вычислительного мышления (и рядом с ним стоящих), на примере аналогии с написанием *обычного текста*.

**В первом приближении**, написание программы прямо аналогично написанию обычного текста и соответственно имеет те же свойства. Например, *письменный* вид (вместо устного) требуется для *масштабирования*. Предложение, сказанное вслух, не фиксируется в воздухе и требует повторения каждый раз, где данное предложение нужно применить, в то время как письменный вид можно дублировать, пересылать, перечитывать и комбинировать с другими неограниченное количество раз. Таким образом, можно сказать, что написание программы

требуется для воспроизведения смыслов, заложенных в ней, неограниченное количество раз.

**Во втором приближении**, написание программ, как и текста, требует знания языка и его грамматики. Естественный язык (например, русский или английский) — это система ассоциаций слов (идей) с ситуациями их применения. Например, когда мы видим небо, мы быстро понимаем, что это «небо». Мы увидели человека, который бежит, значит, мы понимаем, что это «человек» и что он «бежит». Несмотря на то, что в языках есть части речи, такие как глагол, существительное, прилагательное и т. д., в конечном итоге *все слова* сводятся к некоторым категориям, паттернам или *функциям*, которые несут те самые «смыслы», что мы хотим заложить в наши текст, речь или программу.

Грамматика же тут нужна для *быстрой коммуникации* или снижения когнитивных нагрузок *во время* коммуникации. Например, предложение «ждет у человек который окна» читается с трудом и оставляет неуверенность в том, что предложение было понято верно. Соответственно, грамматика создаёт некий «стандарт», по которому мы соглашаемся о порядке слов, правилах правописания и словосочетания (или синтаксисе), чтобы кодировать и декодировать информацию быстро и с наименьшими смысловыми потерями. Таким образом, можно сказать, что написание программы требует наличие языка и его грамматики для кодирования/декодирования смыслов.

**В третьем приближении**, написание программ или текста подразумевает наличие некоторого «агента», для которого данный текст/программа предназначена. В случае обычного текста это могут быть сам автор или люди, понимающие язык, на котором текст был записан. В случае программ ситуация точно такая же. Таким агентом могут являться сам автор программы (если он хорошо помнит семантику языка, чтобы выполнить программу в уме) или *другая* программа, которая будет его *вычислительным агентом*, платформой или интерпретатором (например, компилятор для исходного кода некоторого языка программирования или виртуальная машина/операционная система для финальной скомпилированной программы). Наличие агента является очевидным, но в тоже время и ключевым. Мы можем знать сколько угодно много языков, сколь угодно сложных, и уметь писать сколь угодно сложные программы, однако без стороны, которая

---

примет, скомпилирует и вычислит текст программы, весь смысл наличия вычислительного мышления становится под вопросом.

Таким образом, (полная версия) **написание программы – это текстовое кодирование смыслов с помощью языка программирования для их передачи и декодирования посредством вычислительного агента (человека или машины).** Рисунок 3 суммирует ключевые компоненты вычислительного мышления, описанные выше.

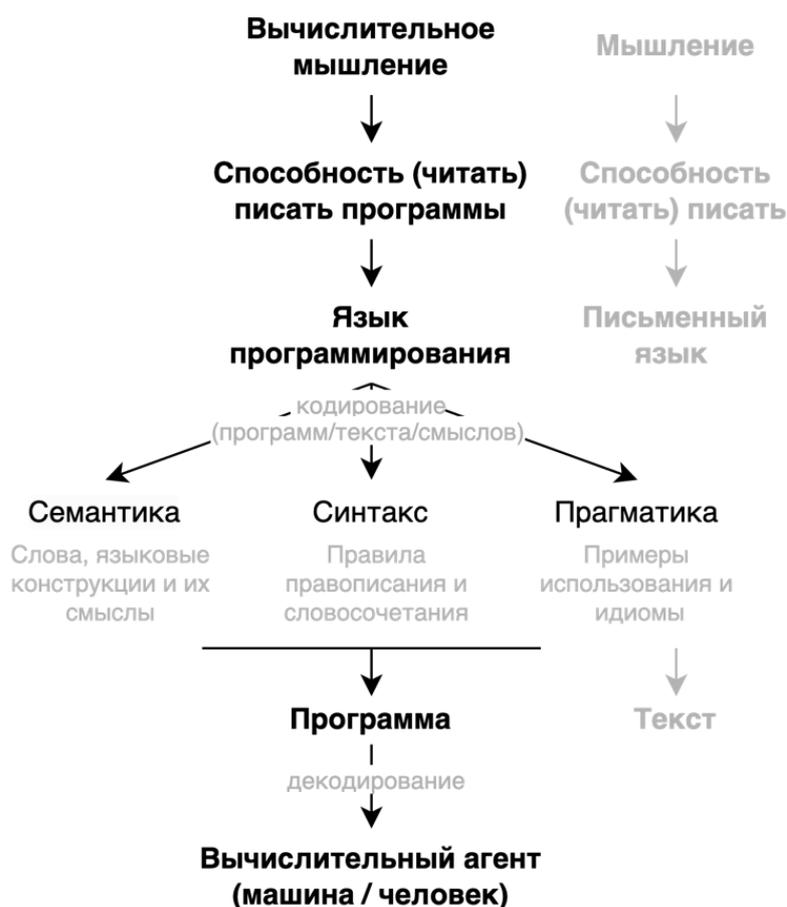


Рис. 3. Ключевые компоненты вычислительного мышления

На Рисунке 3 мы можем увидеть, что вычислительное мышление может быть рассмотрено как некая специализация обычного. Соответственно, мы можем заключить, что все те аспекты, которые затрагивают обычное мышление или использование естественных языков, легко переносимы (и применимы) в исполь-

зовании вычислительного мышления и его главного инструмента – языков программирования.

### **ЗАКЛЮЧЕНИЕ**

Мы рассмотрели смысл, значение и место «вычислительного мышления» в тренде современного цифрового образования и популяризации компьютерных наук. Мы определили, что вычислительное мышление есть ничто иное, как навык написания программ в терминах некоего языка программирования, схоже тому, как мы пишем текст в терминах естественного языка. В связи с невероятной популярностью данного понятия, всеобъемлющих определений и тренда «вычислительное мышление как вторая грамотность», мы рассмотрели простую аналогию написания программ с написанием обычного текста.

Мы считаем, что данная аналогия может послужить намного более надежным, понятным и легко применяемым в практике путеводителем разнообразных смыслов вычислительного мышления (и его основного инструмента – языка программирования), чем абстрактные определения с перечислением аспектов, которые пересекаются с друг другом, другими предметами и общими когнитивными навыками.

### **СПИСОК ЛИТЕРАТУРЫ**

1. *Goshtasbpour F.* JIME Virtual Special Collection – 2012 to 2022: The Decade of the MOOC: 1. Ubiquity Press, 2022. Vol. 2022, No. 1. P. 1.`
2. Capturing the Hype: Year of the MOOC Timeline Explained [Electronic resource] // The Report by Class Central. 2020.  
URL: <https://www.classcentral.com/report/mooc-hype-year-1/>.
3. Introducing the Coursera Impact Report 2021 [Electronic resource] // Coursera Blog. 2021. URL: <https://blog.coursera.org/coursera-impact-report-2021/>.
4. *Tekdal M.* Trends and development in research on computational thinking // Educ Inf Technol. 2021. Vol. 26, No. 5. P. 6499–6529.
5. Global Skills Reports | Coursera [Electronic resource] // Coursera's Skills Reports. URL: <https://www.coursera.org/skills-reports/global>.
6. ChatGPT. What are the main areas digital education teaching? What includes digital literacy and computational literacy? 2023.

7. European Commission. Joint Research Centre. Reviewing computational thinking in compulsory education: state of play and practices from computing education. LU: Publications Office, 2022.

8. (Computational Thinking) The international perspective [Electronic resource]. URL: <https://www.computingschool.org.uk/resource-library/2012/september/the-international-perspective>.

9. Федеральные государственные образовательные стандарты (ФГОС) [Electronic resource] // ФГОС. URL: <https://fgos.ru/>.

10. *Selby C. C. et al.* Computational Thinking: The Developing Definition // University of Southampton (E-prints). 2013.

11. *Denning P. J., Tedre M.* Computational Thinking: A Disciplinary Perspective // Informatics in Education. 2021.

12. *Tedre M., Denning P. J.* The long quest for computational thinking // Proceedings of the 16th Koli Calling International Conference on Computing Education Research. Koli Finland: ACM, 2016. P. 120–129.

13. European Commission. Joint Research Centre. Developing computational thinking in compulsory education: implications for policy and practice. LU: Publications Office, 2016.

14. Computing in the national curriculum - a guide for primary teachers [Electronic resource]. URL: <https://www.computingschool.org.uk/resource-library/2014/september/computing-in-the-national-curriculum-a-guide-for-primary-teachers>.

15. CAS computational thinking – A Guide for teachers [Electronic resource]. URL: <https://www.computingschool.org.uk/resource-library/2014/june/cas-computational-thinking-a-guide-for-teachers>.

16. K–12 Computer Science Framework [Electronic resource] // k12cs.org. URL: <http://k12cs.org>.

17. K–12 Computer Science Framework | Statements of Support [Electronic resource] // k12cs.org. URL: <https://k12cs.org/statements-of-support/>.

18. Shut down or restart? | Royal Society [Electronic resource]. URL: <https://royalsociety.org/topics-policy/projects/computing-in-schools/report/>.

19. *Shute V. J., Sun C., Asbell-Clarke J.* Demystifying computational thinking // Educational Research Review. 2017. Vol. 22. P. 142–158.

20. Barr V., Stephenson C. Bringing computational thinking to K-12: what is involved and what is the role of the computer science education community? // ACM Inroads. 2011. Vol. 2, No. 1. P. 48–54.

21. Fagerlund J. et al. Computational thinking in programming with Scratch in primary schools: A systematic review // Computer Applications in Engineering Education. 2021. Vol. 29, No. 1. P. 12–28.

22. Best Jobs in America | Glassdoor [Electronic resource].  
URL: [https://www.glassdoor.com/List/Best-Jobs-in-America-LST\\_KQ0,20.htm](https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm).

23. Ершов А. П. Программирование – вторая грамотность | Архив академика А.П. Ершова [Electronic resource].  
URL: [https://web.archive.org/web/20210429142621/http://ershov.iis.nsk.su/ru/second\\_literacy/pred](https://web.archive.org/web/20210429142621/http://ershov.iis.nsk.su/ru/second_literacy/pred).

24. Barr D., Harrison J., Conery L. Computational Thinking: A Digital Age Skill for Everyone // Learning & Leading with Technology. International Society for Technology in Education, 2011. Vol. 38, No. 6. P. 20–23.

25. Quotes on computer science – What leaders and trendsetters have to say [Electronic resource] // Code.org. URL: <https://code.org/quotes>.

26. Digital Education Action Plan (2021–2027) | see COMMISSION STAFF WORKING DOCUMENT [Electronic resource].  
URL: <https://education.ec.europa.eu/node/1518>.

27. Inc T.M. MATLAB. Natick, Massachusetts, United States: The MathWorks Inc., 2022.

28. G. van Rossum (Guido). Python reference manual // Department of Computer Science [CS]. CWI, 1995. No. R 9525.

29. Resnick M. et al. Scratch: programming for all // Commun. ACM. 2009. Vol. 52, No. 11. P. 60–67.

---

## “COMPUTATIONAL THINKING” AS A MAIN COMPETENCY IN MODERN DIGITAL EDUCATION AND SOCIETY

Timur Fayzrakhmanov<sup>1</sup> [0000-0001-5013-4523]

*Institute of software development and engineering, Innopolis University,*

---

*Universitetskaya St, 1, Innopolis, 420500*

<sup>1</sup>tim.fayzrakhmanov@gmail.com

### **Abstract**

The incredible growth in the popularity of digital technologies and education has led to the emergence of “computational thinking”. The dramatic demand and popularization of this notion ignited the redesign of educational standards and the upsurge of literature on this subject. During the extensive literature review, we found that it is not always clear what computational thinking really is, what aspects it entails, and how to think about it in simple yet robust terms. In this paper, we have examined the meaning of this concept, its importance in modern digital education, and drawn an analogy between computational thinking and the skills of writing ordinary text to facilitate its reliable understanding.

**Keywords:** *computational thinking, digital education, coding, programming, computer science, STEM*

### **REFERENCES**

1. *Goshtasbpour F.* JIME Virtual Special Collection – 2012 to 2022: The Decade of the MOOC: 1. Ubiquity Press, 2022. Vol. 2022, No. 1. P. 1.
2. Capturing the Hype: Year of the MOOC Timeline Explained [Electronic resource] // The Report by Class Central. 2020.  
URL: <https://www.classcentral.com/report/mooc-hype-year-1/>.
3. Introducing the Coursera Impact Report 2021 [Electronic resource] // Coursera Blog. 2021. URL: <https://blog.coursera.org/coursera-impact-report-2021/>.
4. *Tekdal M.* Trends and development in research on computational thinking // *Educ Inf Technol.* 2021. Vol. 26, No. 5. P. 6499–6529.
5. Global Skills Reports | Coursera [Electronic resource] // Coursera’s Skills Reports. URL: <https://www.coursera.org/skills-reports/global>.
6. ChatGPT. What are the main areas digital education teaching? What includes digital literacy and computational literacy? 2023.
7. European Commission. Joint Research Centre. Reviewing computational thinking in compulsory education: state of play and practices from computing education. LU: Publications Office, 2022.

8. (Computational Thinking) The international perspective [Electronic resource]. URL: <https://www.computingschool.org.uk/resource-library/2012/september/the-international-perspective>.

9. Federal State Educational Standards (FSES) [Electronic resource] // FGOS. URL: <https://fgos.ru/>.

10. *Selby C. C. et al.* Computational Thinking: The Developing Definition // University of Southampton (E-prints). 2013.

11. *Denning P. J., Tedre M.* Computational Thinking: A Disciplinary Perspective // Informatics in Education. 2021.

12. *Tedre M., Denning P. J.* The long quest for computational thinking // Proceedings of the 16th Koli Calling International Conference on Computing Education Research. Koli Finland: ACM, 2016. P. 120–129.

13. European Commission. Joint Research Centre. Developing computational thinking in compulsory education: implications for policy and practice. LU: Publications Office, 2016.

14. Computing in the national curriculum - a guide for primary teachers [Electronic resource]. URL: <https://www.computingschool.org.uk/resource-library/2014/september/computing-in-the-national-curriculum-a-guide-for-primary-teachers>.

15. CAS computational thinking – A Guide for teachers [Electronic resource]. URL: <https://www.computingschool.org.uk/resource-library/2014/june/cas-computational-thinking-a-guide-for-teachers>.

16. K–12 Computer Science Framework [Electronic resource] // k12cs.org. URL: <http://k12cs.org>.

17. K–12 Computer Science Framework | Statements of Support [Electronic resource] // k12cs.org. URL: <https://k12cs.org/statements-of-support/>.

18. Shut down or restart? | Royal Society [Electronic resource]. URL: <https://royalsociety.org/topics-policy/projects/computing-in-schools/report/>.

19. *Shute V. J., Sun C., Asbell-Clarke J.* Demystifying computational thinking // Educational Research Review. 2017. Vol. 22. P. 142–158.

20. *Barr V., Stephenson C.* Bringing computational thinking to K-12: what is involved and what is the role of the computer science education community? // ACM Inroads. 2011. Vol. 2, No. 1. P. 48–54.

21. *Fagerlund J. et al.* Computational thinking in programming with Scratch in

primary schools: A systematic review // *Computer Applications in Engineering Education*. 2021. Vol. 29, No. 1. P. 12–28.

22. Best Jobs in America | Glassdoor [Electronic resource].

URL: [https://www.glassdoor.com/List/Best-Jobs-in-America-LST\\_KQ0,20.htm](https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm).

23. *Ershov A. P.* Programming is the second literacy | Academician A.P. Ershov Archive [Electronic resource].

URL: [https://web.archive.org/web/20210429142621/http://ershov.iis.nsk.su/ru/second\\_literacy/pred](https://web.archive.org/web/20210429142621/http://ershov.iis.nsk.su/ru/second_literacy/pred).

24. *Barr D., Harrison J., Conery L.* Computational Thinking: A Digital Age Skill for Everyone // *Learning & Leading with Technology*. International Society for Technology in Education, 2011. Vol. 38, No. 6. P. 20–23.

25. Quotes on computer science – What leaders and trendsetters have to say [Electronic resource] // Code.org. URL: <https://code.org/quotes>.

26. Digital Education Action Plan (2021–2027) | see COMMISSION STAFF WORKING DOCUMENT [Electronic resource].

URL: <https://education.ec.europa.eu/node/1518>.

27. *Inc T. M.* MATLAB. Natick, Massachusetts, United States: The MathWorks Inc., 2022.

28. *G. van Rossum (Guido)*. Python reference manual // Department of Computer Science [CS]. CWI, 1995. No. R 9525.

29. *Resnick M. et al.* Scratch: programming for all // *Commun. ACM*. 2009. Vol. 52, No. 11. P. 60–67

---

## СВЕДЕНИЯ ОБ АВТОРЕ



**ФАЙЗРАХМАНОВ Тимур Расимович** – разработчик ПО, аспирант, исследователь. Сфера научных интересов: языки программирования, представление и обработка знаний, повторное использование знаний, формализация, обобщённое моделирование систем, цифровые библиотеки, графика, среды быстрого прототипирования, блочный подход.

**Timur Rasimovich FAYZRACHMANOV** – software developer, PhD student, researcher. Research interests: computational languages, knowledge representation and processing, knowledge reuse, formalization, generic systems modeling, digital libraries, graphics, low-code development platforms, block-based approach.

email: [tim.fayzrakhmanov@gmail.com](mailto:tim.fayzrakhmanov@gmail.com)

ORCID: 0000-0001-5013-4523

*Материал поступил в редакцию 3 марта 2023 года*