

## Применение трансформерных архитектур для анализа текстовых данных на примере автоматизации поиска типовых замечаний

О. В. Афанасьева<sup>1</sup>, И. М. Новожилов<sup>2</sup>, Д. А. Первухин<sup>1</sup>, П. А. Скуковский<sup>1</sup>✉

<sup>1</sup> Санкт-Петербургский горный университет императрицы Екатерины II, Санкт-Петербург, Россия

<sup>2</sup> Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина), Санкт-Петербург, Россия

✉ Pavel22122004@yandex.ru

**Аннотация.** Рассмотрены методы машинного обучения и обработки естественного языка, применяемые для автоматизации анализа текстовых данных. На примере экспертизы проектной документации показано, как использование модели BERT позволяет выявлять типовые замечания, ускорять обработку информации и снижать влияние человеческого фактора. Разработанный программный комплекс позволяет автоматизировать процесс поиска повторяющихся ошибок, облегчает работу экспертов и создает основу для оптимизации бизнес-процессов в условиях роста объемов данных.

**Ключевые слова:** обработка естественного языка (NLP), BERT, анализ текстовых данных, автоматизация обработки текста, трансформерные архитектуры

**Для цитирования:** Применение трансформерных архитектур для анализа текстовых данных на примере автоматизации поиска типовых замечаний / О. В. Афанасьева, И. М. Новожилов, Д. А. Первухин, П. А. Скуковский // Изв. СПбГЭТУ «ЛЭТИ». 2026. Т. 19, № 3. С. 43–51. doi: 10.32603/2071-8985-2026-19-3-43-51.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Original article

## Using Transformer Architectures to Analyze Text Data by Automating the Search for Common Comments

O. V. Afanasieva<sup>1</sup>, I. M. Novozhilov<sup>2</sup>, D. A. Pervukhin<sup>1</sup>, P. A. Skukovskii<sup>1</sup>✉

<sup>1</sup> Empress Catherine II Saint Petersburg Mining University, Saint Petersburg, Russia

<sup>2</sup> Saint Petersburg Electrotechnical University, Saint Petersburg, Russia

✉ Pavel22122004@yandex.ru

**Abstract.** The article discusses machine learning and natural language processing methods for automating the analysis of text data. Using the example of project documentation review, it shows how the use of the BERT model allows for the identification of typical comments, the acceleration of information processing, and the reduction of the human factor. The developed software package automates the process of finding repetitive errors, facilitates the work of experts, and provides a basis for optimizing business processes in the context of increasing data volumes.

**Keywords:** natural language processing (NLP), BERT, text data analysis, text processing automation, transformer architectures

**For citation:** Using Transformer Architectures to Analyze Text Data by Automating the Search for Common Comments / O. V. Afanasieva, I. M. Novozhilov, D. A. Pervukhin, P. A. Skukovskii // LETI Transactions on Electrical Engineering & Computer Science. 2026. Vol. 19, no. 3. P. 43–51. doi: 10.32603/2071-8985-2026-19-3-43-51.

**Conflict of interests.** The authors declare no conflicts of interests.

**Введение.** В современном мире с каждым днем растет количество текстовых данных, представляющих собой критически важный, но слабо структурированный ресурс. Ручной анализ большого объема данных аналитиком представляет собой значительную проблему, связанную с немалым количеством времени, необходимым аналитику на проведения анализа. Также при ручном анализе данных повышается количество ошибок, связанных с человеческим фактором, например усталостью или невнимательностью человека, выполняющего анализ данных [1], [2].

Для автоматизации задач анализа данных применяются методы машинного обучения и обработки естественного языка (Natural Language Processing, NLP), которые позволяют автоматизировать процесс анализа и извлечения полезных данных из больших массивов информации. Алгоритм анализа текстовых данных представлен на блок-схеме.



Рис. 1. Алгоритм анализа текстовых данных  
Fig. 1. Text data analysis algorithm

Анализ текстовых данных начинается с этапа предобработки и векторного представления текста, где неструктурированные данные трансформируются в числовые форматы, пригодные для обработки алгоритмами машинного обучения. Данный этап включает токенизацию, лемматизацию и стемминг, а также использование моделей числовых эмбедингов, для получения векторных представлений слов и передачи семантического контекста текста. Более продвинутым подходом служит применение трансформерных архитектур – таких, как BERT и GPT (Generative Pre-trained Model), которые генерируют контексту-

ализированные эмбединги на уровне предложений или целых документов, что значительно повышает качество последующего анализа [3], [4].

На следующем этапе происходит автоматизация и классификация, где на первый план выходят методы тематического моделирования, предназначенные для автоматического выявления тематических структур в больших объемах данных. При анализе данных важно извлечение информации, которое охватывает распознавание именованных сущностей для идентификации и категоризации объектов, извлечение связей для определения семантических отношений между сущностями, а также анализ тональности и распознавание аспектов для детального понимания мнений и оценок в пользовательских отзывах [5], [6].

Завершающим этапом служит подведение итогов, в котором создается краткий пересказ или резюме из большого текста, что позволяет быстро собрать основные выводы из большого объема данных. Для этого используются методы экстрактивной и абстрактивной суммаризации. Метод экстрактивной суммаризации заключается в выделении системной важной информации в виде готовых предложений и фраз, в то время как при абстрактивной суммаризации система анализирует смысл текста, после чего пересказывает его собственными словами, формулируя новые короткие предложения, сохраняя при этом все ключевые идеи.

**Постановка задачи.** В качестве объекта исследования для автоматизации процесса анализа текстовых данных был выбран процесс нахождения типовых замечаний в таблице данных результатов экспертиз проектной документации компании ООО «Газпром проектирование». Компания занимается разработкой проектной документации для строительства сооружений нефтегазовой отрасли. При разработке проектной документации обязательным этапом служит прохождение государственных экспертиз и экспертиз заказчиков, в результате которых эксперты выявляют недостатки в проектной документации, выносят замечания и отправляют документацию на доработку. Важный этап для оптимизации работы отделов компании – снижение количества получаемых замечаний по результатам экспертиз [7]–[9].

На данный момент анализ замечаний, полученных по результатам экспертиз, проходит вручную, что занимает большое количество времени и может приводить к возникновению ошибок при анализе, связанных с человеческим фактором. Входные данные для анализа представляют собой

таблицу Excel, где каждая запись – это замечание, полученное во время экспертизы проектной документации по какому-либо объекту. Одной из задач для работы с этими данными является поиск типовых замечаний – замечаний, которые встречаются несколько раз во время экспертизы разных объектов. Значимость нахождения таких замечаний заключается в оптимизации процесса разработки проектной документации, а именно – в снижении количества полученных замечаний посредством выработки мер по предотвращению появления типовых ошибок и их недопущению в будущем [10], [11].

**Алгоритм работы программы.** Для автоматизации процесса анализа данных был разработан алгоритм для поиска типовых записей в таблице данных результатов экспертиз проектной документации компании ООО «Газпром проектирование».

С этой целью была написана программа поиска типовых замечаний с использованием языковой модели BERT. BERT (Bidirectional Encoder Representations from Transformers) – одна из самых передовых и широко используемых моделей от компании Google для обработки естественного языка. Главная особенность BERT заключается в ее способности понимать контекст слов как слева, так и справа одновременно, в отличие от традиционных моделей, которые читают текст либо слева направо, либо справа налево. Это обеспечивает глубокое понимание смысла фраз, особенно когда речь идет о сложных формулировках, содержащих техническую терминологию или длинные предложения – как раз то, что характерно для проектной документации и замечаний к ней. Для реализации задачи была разработана программа на языке Python с использованием библиотеки sentence-transformers, которая позволяет работать с BERT и преобразовывать текст в векторы признаков (эмбединги) [12], [13].

Задача программы состоит в анализе полученных данных преобразованием текста содержания замечаний в эмбединги и их последующего сравнения между собой, а также проверки на удовлетворение условиям работы программы, а именно – количеству повторений, начиная с которого ошибка будет считаться типовой, а также проверка выполнения условия о том, что ошибка встречается хотя бы в двух разных объектах. Также имеется возможность выбрать данные, по которым будет происходить группировка, например при выборе группировки по отделу, разработа-

тывавшему проектную документацию, для каждого отдела программа будет искать типовые ошибки, которые были допущены каждым из отделов по отдельности [14].

Общий алгоритм работы программы следующий:

1. Предобработка данных. Загружается Excel-файл с замечаниями, после чего программа отфильтровывает только нужные строки (например, только принятые или только рассмотренные замечания – по выбору пользователя). Часть программы, отвечающая за обработку данных:

```
df = pd.read_excel(input_path)
```

```
if not selected_statuses:
```

```
    raise ValueError("Н выбран ни один статус")
```

```
df = df [df["Статус замечания"].isin(selected_statuses)].copy().reset_index(drop=True)
```

```
group_values = df [group_column].dropna().unique() if group_column else [None]
```

2. Преобразование текста в эмбединги. На данном этапе происходит преобразование текста замечания в эмбединги. Эмбединги – это способ представления слов, предложений, изображений и других типов данных в виде числовых векторов. Эти векторы позволяют анализировать информацию, преобразованную в них. Часть кода, отвечающая за преобразование текста в эмбединги:

```
model_path= os.path.join(os.path.dirname(os.path.abspath(_file_)), "local_model")
```

```
model = SentenceTransformer(model_path)
```

```
df [column_text] = df[column_text].fillna("").astype(str)
```

```
df [column_object] = df [column_object].fillna("").astype(str)
```

```
embeddings = model.encode(df[column_text].tolist(), show_progress_bar=True)
```

```
similarity_matrix = cosine_similarity(embeddings)
```

3. Сравнение векторов (поиск совпадений).

Для первого замечания в списке создается новая группа. Далее каждое последующее сравнивается с уже отнесенными к группам: если оно по смыслу похоже (по косинусному сходству эмбедингов), принадлежит к тому же типу (Принятые/Рассмотренные) и относится к другому объекту (по столбцу «Наименование объекта»), тогда оно считается типовым и включается в соответствующую группу [15]. Функция, отвечающая за сравнение эмбедингов:

```
def найти_группы(df_subset, indices) :
    nonlocal group_id_counter
    visited = set ()
    for i in range(len(indices)):
        idx_i = indices[i]
        if idx_i in visited:
            continue
        temp_group = []
        obj_i = df_subset.loc[idx_i, column_object].strip().lower()
        for j in range(i + 1, len(indices)) :
            idx_j = indices[j]
            if idx_j in visited:
                continue
            obj_j = df_subset.loc[idx_j, column_object].strip().lower()
            sim = similarity_matrix[idx_i][idx_j]
            if sim >= 0.99 and obj_i != obj_j:
                temp_group.append(idx_j)
        if temp_group:
            temp_group.append(idx_i)
            main_text = df_subset.loc[idx_i, column_text]
            full_group = [k for k in df_subset.index if df_subset.loc[k, column_text] == main_text]
            group = list(set(temp_group + full_group))
            if len(group) >= min_repeats:
                visited.update(group)
                global_visited.update(group)
                for idx in group:
                    группа_по_индексу[idx] = f"Группа {group_id_counter}"
                    group_id_counter += 1
```

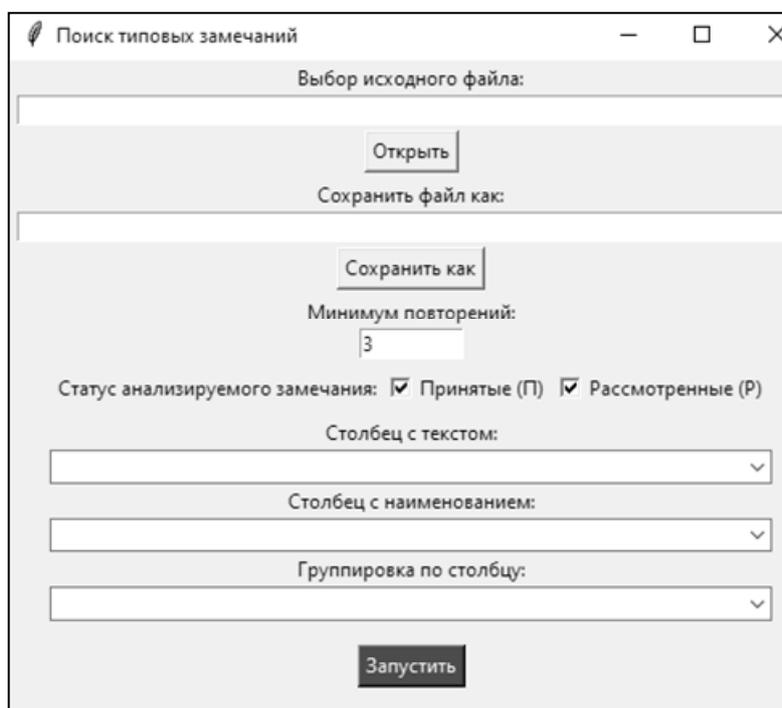


Рис. 2. Окно настройки программы  
Fig. 2. Program Settings window

Все дальнейшие совпадения с этим замечанием автоматически добавляются в группу без повторного сравнения, что ускоряет анализ при больших объемах данных.

Для удобства использования была реализована графическая оболочка на базе tkinter, благодаря которой процесс запуска программы стал максимально простым и доступным для пользователя без навыков программирования.

Интерфейс включает в себя следующие функции: выбор входного Excel-файла с замечаниями, указание пути и имени для сохранения результата, выбор, анализируемых замечаний: принятых, рассмотренных или обоих типов, ввод значения минимального количества повторений, начиная с которого замечание считается типовым, выбор столбца для группировки типовых ошибок (например, по филиалу, отделу, эксперту и др.), а также кнопка для запуска анализа. Пример окна интерфейса представлен на рис. 2.

**Процесс работы с программой и результаты.** Для начала работы с программой необходимо запустить саму программу. После запуска программы появится меню настройки программы, представленное на рис. 6.

В меню имеются следующие функции:

1. «Выбор исходного файла:» – для выбора файла, который необходимо проанализировать, необходимо нажать на кнопку «открыть», после чего появится окно для выбора файла с данными (файл с тестовыми данными находится в папке «Тестовые данные»).

2. «Сохранить файл как:» – для выбора названия файла, который будет создан после работы программы, и места его сохранения. После нажатия на кнопку «Сохранить как:» появится окно для выбора названия файла, который будет сохранен.

3. «Минимум повторений:» – данная настройка позволяет менять количество повторений, при котором ошибка будет считаться типовой. По умолчанию стоит значение «3».

4. «Статус анализируемого замечания:» – возможность выбора данных, которые будут рассматриваться программой в зависимости от их статуса. По умолчанию программа будет анализировать все данные в таблице. При снятии «галочки» с какого-либо из статусов программа не будет учитывать записи с этим статусом и будет анализировать только тот статус, который выбран (в случае, если не выбирать ни один из статусов, программа выдаст ошибку и предложит выбрать один из статусов для продолжения работы).

5. «Столбец с текстом:» и «Столбец с наименованием:» – вспомогательные настройки. После загрузки программы и выбора исходного файла программа автоматически проверяет загруженный файл, находит нужные столбцы, однако если название столбца в исходном файле записано неверно, программа не сможет его найти. В этом случае будет необходимо выбрать нужный столбец из выпадающего меню.

6. «Группировка по столбцу:» – позволяет выбрать столбец, по которому будут сгруппированы найденные данные. Также выбор любого столбца в данной настройке меняет логику работы программы, добавляя еще одно условие нахождения типовых ошибок, а именно – для каждого объекта группировки рассматриваются только записи, попадающие под конкретный объект группировки (т. е. при выборе группировки по столбцу ФИО для каждого объекта группировки будут рассматриваться только те записи, которые попадают в данную группировку, и последующий алгоритм ищет типичные ошибки для каждой из полученных групп по отдельности).

7. Кнопка «Запустить» запускает анализ данных. После завершения анализа будет выведено текстовое окно, сообщающее о завершении работы приложения. Полученный файл будет сохранен в место, выбранное при настройке программы. В результате работы программы формируется итоговая таблица со всеми найденными типовыми замечаниями, которая содержит информацию о каждой из найденных ошибок.

При выборе группировки по дополнительному столбцу анализ замечаний будет проходить отдельно для каждой из групп, полученных после группировки, следовательно, будут обнаруживаться типовые замечания, допущенные каким-либо конкретным экспертом или отделом. Пример результата работы с программой при использовании группировки по столбцу с ФИО эксперта представлен на рис. 3.

Созданный алгоритм был протестирован на реальных данных, полученных из отчетов экспертных проверок проектной документации. В ходе испытаний система продемонстрировала способность самостоятельно выявлять повторяющиеся и типовые замечания, возникающие в разных проектах, сопоставлять их между собой и группировать в соответствии с заданными условиями. Такой механизм автоматической кластеризации позволил не только идентифицировать сходные формулировки, но и установить взаимосвязи между замечаниями, относящимися к раз-

|    | A            | B           | C  | D  | E                   | F                                   | G                               | H      |
|----|--------------|-------------|----|--|---------------------|-------------------------------------|---------------------------------|--------|
| 1  |              |             | №  | Наименование объекта   | Отв.Филиал          | Содержание замечания                | Ф.И.О.эксперта,контактныеданные | Статус |
| 2  | Сергеев С.С. |             |    |  |                     |                                     |                                 |        |
| 3  |              | Группа 1    |    | В сборниках ведомостей объемов работ отсутствуют объемы работ в части КИТСО и ИБ |                     |                                     |                                 |        |
| 4  |              | Группа 1.7  | 7  | Объект 2   | Санкт-Петербургский | В сборниках ведомостей Сергеев С.С. |                                 | Р      |
| 5  |              | Группа 1.19 | 19 | Объект 5   | Нижегородский       | В сборниках ведомостей Сергеев С.С. |                                 | П      |
| 6  |              | Группа 1.20 | 20 | Объект 5   | Нижегородский       | В сборниках ведомостей Сергеев С.С. |                                 | П      |
| 7  | Петров П.П.  |             |    |  |                     |                                     |                                 |        |
| 8  |              | Группа 2    |    | В спецификации исключить информацию о поставщиках МТР                            |                     |                                     |                                 |        |
| 9  |              | Группа 2.24 | 24 | Объект 5   | Нижегородский       | В спецификации исключ Петров П.П.   |                                 | Р      |
| 10 |              | Группа 2.46 | 46 | Объект 9   | Нижегородский       | В спецификации исключ Петров П.П.   |                                 | П      |
| 11 |              | Группа 2.47 | 47 | Объект 10  | Уфимский            | В спецификации исключ Петров П.П.   |                                 | П      |
| 12 |              | Группа 2.48 | 48 | Объект 10  | Уфимский            | В спецификации исключ Петров П.П.   |                                 | П      |

Рис. 3. Пример сформированного отчета с группировкой по эксперту  
Fig. 3. Example of a generated report grouped by expert

личным объектам или видам работ. Применение разработанного подхода показало его высокую практическую эффективность: время, необходимое для анализа и систематизации информации, было значительно сокращено, а количество ошибок, связанных с субъективным восприятием и человеческим фактором, заметно уменьшилось. Полученные результаты упорядочиваются в структурированном и наглядном виде, что облегчает последующую интерпретацию, подготовку отчетных материалов и принятие решений. Дополнительно внедрение алгоритма открывает новые возможности для более глубокого анализа тенденций и закономерностей, характерных для процесса проектирования. Автоматизированная обработка данных позволяет выявлять системные проблемы, которые трудно обнаружить при традиционном ручном анализе, а также формировать предпосылки для разработки рекомендаций, направленных на повышение качества проектной документации и совершенствование внутренних процедур контроля.

**Заключение.** Предложенная в статье методика наглядно демонстрирует эффективность и практическую целесообразность применения инструментов обработки естественного языка для решения задач автоматизированного анализа текстовой информации. Использование модели семейства BERT позволило достичь высокой точности интерпретации сложных грамматических конструкций, терминов и устойчивых выражений, характерных для профессиональной речи экспертов в области проектной документации. Такой подход не только ускоряет процесс обработки данных, но и обеспечивает более объективную оценку содержания текстов, минимизируя влияние человеческого фактора. Кроме того, автоматизация анализа способствует оптимизации работы специалистов: эксперты могут сосредоточиться не на рутинном поиске повторяющихся ошибок, а на содержательной интерпретации резуль-

татов, выявлении скрытых закономерностей и формировании предложений по устранению системных недочетов. В дальнейшем представляется целесообразным развитие методики в направлении интеграции анализа текстовой информации с другими источниками данных – геотехническими, пространственными и статистическими. Это создаст предпосылки для более глубокого понимания контекста возникновения замечаний и обеспечит комплексный подход к оценке качества проектной документации. В результате можно будет формировать более точные прогнозы и повышать обоснованность принимаемых управленческих решений.

Применение современных методов машинного обучения, основанных на архитектурах типа «трансформер», позволяет автоматизировать рутинный процесс анализа данных. Программа для выявления типовых замечаний, разработанная в рамках исследования, подтвердила свою эффективность на практике, продемонстрировав способность существенно повысить качество контроля и сократить нагрузку на специалистов-экспертов. Полученные результаты свидетельствуют о потенциале дальнейшего развития данного направления. В перспективе предполагается создание комплексных аналитических систем, способных работать с мультимодальными источниками информации, включая текстовые, числовые и пространственные данные. Расширение функций алгоритма до уровня прогнозирования и предотвращения повторяющихся ошибок позволит не только своевременно выявлять проблемные зоны, но и формировать превентивные меры по повышению качества проектной документации. Таким образом, использование технологий искусственного интеллекта становится значимым инструментом повышения эффективности экспертной деятельности и совершенствования системы управления качеством в инженерно-технической сфере.

## Список литературы

1. Новоселов А. Л., Новоселова И. Ю., Желтенков А. В. Перспективы развития предприятия: анализ и моделирование // Вестн. Московского гос. обл. ун-та. Сер.: Экономика. 2021. № 4. С. 64–75. doi: 10.18384/2310-6646-2021-4-64-75.
2. Прогнозирование стоимости недвижимости на основе методов машинного обучения / С. И. Киселев, И. Ю. Мышкина, Л. Ю. Грудцына, Л. З. Гумерова, Г. Н. Аглямзянова // Науч.-техн. вестн. Поволжья. 2023. № 3. С. 60–62.
3. Салып Б. Ю., Смирнов А. А. Анализ модели BERT как инструмента определения меры смысловой близости предложений естественного языка // StudNet. 2022. Т. 5, № 5. С. 3509–3518.
4. Буравлев А. С., Демидова Д. Е., Ткачева Е. А. Анализ качества реконструкции бизнес-процессов с помощью языковой модели ChatGPT // Техника средств связи. 2025. № 1(169). С. 84–97. doi: 10.24412/2782-2141-2025-1-84-97.
5. Пимешков В. К., Шишаев М. Г. Методы извлечения знаний из естественно-языковых текстов // Тр. Кольского научного центра РАН. Сер.: Техн. науки. 2022. Т. 13, № 2. С. 31–45. doi: 10.37614/2949-1215.2022.13.2.003.
6. Катермина Т. С., Тагиров К. М., Тагиров Т. М. Элементы искусственного интеллекта в решении задач анализа текстов // Computational Nanotechnol. 2022. Т. 9, № 2. С. 35–44. doi: 10.33693/2313-223X-2022-9-2-35-44.
7. Загорская К. Р., Иванова И. В. Синергия RPA и искусственного интеллекта // Дневн. науки. 2023. № 11(83). С. 1–11. doi: 10.51691/2541-8327\_2023\_11\_8.
8. Корытов П. В. Сравнительный анализ методов извлечения ключевых слов без предварительного обучения // Изв. СПбГЭТУ «ЛЭТИ». 2025. Т. 18, № 7. С. 60–68. doi: 10.32603/2071-8985-2025-18-7-60-68.
9. Опалев А. С., Алексеева С. А. Методическое обоснование выбора оптимальных режимов работы оборудования схемы стадийного вывода концентрата при обогащении железных руд // Зап. Горного ин-та. 2022. Т. 256. С. 593–602. doi: 10.31897/PMI.2022.80.
10. Стоянова А. Д., Трофимец В. Я., Матрохина К. В. Методический подход к рейтинговой оценке компаний Арктической зоны на основе показателей устойчивого развития // Горный информ.-аналитич. бюл. (науч.-техн. журн.). 2024. № 6. С. 149–162. doi: 10.25018/0236\_1493\_2024\_6\_0\_149.
11. Нейрус С. К., Навацкая В. А., Афанасьев М. П. Исследование структурных подразделений промышленных предприятий с помощью систем массового обслуживания // Междунар. науч.-исслед. журн. 2022. № 12(126). С. 1–7. doi: 10.23670/IRJ.2022.126.32.
12. Ерошин С. Е. Формирование подходов к совершенствованию информационных систем поддержки принятия решений // Инновации и инвестиции. 2021. № 10. С. 194–202.
13. Дементьев В. Е., Киреев С. Х. Выбор алгоритмов машинного обучения для классификации текстовых документов // Техника средств связи. 2022. № 2(158). С. 22–52. doi: 10.24412/2782-2141-2022-2-22-52.
14. Макарова Е. А., Лагерев Д. Г. Модель обработки слабоструктурированных текстовых данных на русском языке для интеллектуальной поддержки информационного управления в динамических организационных системах // Модели, системы, сети в экономике, технике, природе и обществе. 2022. № 3. С. 104–125. doi: 10.21685/2227-8486-2022-3-7.
15. Квасов М. Н., Митряев Г. А., Прохоров М. А. Разработка математической модели семантического анализа и синтеза данных // Изв. ТулГУ. Техн. науки. 2023. № 11. С. 262–266. doi: 10.24412/2071-6168-2023-11-262-263.

## Информация об авторах

**Афанасьева Ольга Владимировна** – канд. техн. наук, доцент кафедры системного анализа и управления, Санкт-Петербургский горный университет императрицы Екатерины II, 21-я линия, д. 2, ВО, Санкт-Петербург, 199106, Россия.

E-mail: [afanaseva\\_ov@pers.spmi.ru](mailto:afanaseva_ov@pers.spmi.ru)

<https://orcid.org/0000-0003-3169-4781>

**Новожилов Игорь Михайлович** – канд. техн. наук, доцент кафедры автоматизации и процессов управления СПбГЭТУ «ЛЭТИ».

E-mail: [novozhilovim@list.ru](mailto:novozhilovim@list.ru)

<https://orcid.org/0000-0002-2056-3930>

**Первухин Дмитрий Анатольевич** – д-р техн. наук, профессор, зав. кафедрой системного анализа и управления, Санкт-Петербургский горный университет императрицы Екатерины II, 21-я линия, д. 2, ВО, Санкт-Петербург, Россия.

E-mail: [pervukhin\\_da@pers.spmi.ru](mailto:pervukhin_da@pers.spmi.ru)

<https://orcid.org/0000-0003-3547-2932>

**Скуковский Павел Алексеевич** – студент группы САМ-22, Санкт-Петербургский горный университет императрицы Екатерины II, 21-я линия, д. 2, ВО, Санкт-Петербург, 199106, Россия.

E-mail: Pavel22122004@yandex.ru

**Вклад авторов:**

Афанасьева О. В. – разработка алгоритма, проверка результатов на достоверность и воспроизводимость.

Новожилов И. М. – общее руководство.

Первухин Д. А. – разработка методологии проведения исследования, разработка концептуальной модели исследования.

Скуковский П. А. – программирование, написание кода и разработка ПО для исследования.

References

1. Novoselov A. L., Novoselova I. Yu., Zheltenkov A. V. Perspektivy razvitiya predpriyatija: analiz i modelirovanie // Vestn. Moskovskogo gos. obl. un-ta. Ser.: Ekonomika. 2021. № 4. S. 64–75. doi: 10.18384/2310-6646-2021-4-64-75. (In Russ.).
2. Prognozirovanie stoimosti nedvizhimosti na osnove metodov mashinnogo obuchenija / S. I. Kiselev, I. Yu. Myshkina, L. Yu. Grudtsyna, L. Z. Gumerova, G. N. Ag-ljamzjanova // Nauch.-tehn. vestn. Povolzhja. 2023. № 3. S. 60–62. (In Russ.).
3. Salyp B. Yu., Smirnov A. A. Analiz modeli BERT kak instrumenta opredelenija mery smyslovoj blizosti predlozhenij estestvennogo jazyka // StudNet. 2022. T. 5, № 5. S. 3509–3518. (In Russ.).
4. Buravlev A. S., Demidova D. E., Tkacheva E. A. Analiz kachestva rekonstruktsii biznes-protsessov s pomoshhju jazykovej modeli ChatGPT // Tekhnika sredstv svjazi. 2025. № 1(169). S. 84–97. doi: 10.24412/2782-2141-2025-1-84-97. (In Russ.).
5. Pimeshkov V. K., Shishaev M. G. Metody izvleshhenija znanij iz estestvenno-jazykovyh tekstov // Tr. Kolskogo nauch. tsentra RAN. Ser.: Tekhn. nauki. 2022. T. 13, № 2. S. 31–45. doi: 10.37614/2949-1215.2022.13.2.003. (In Russ.).
6. Katermina T. S., Tagirov K. M., Tagirov T. M. Elementy iskusstvennogo intellekta v reshenii zadach analiza tekstov // Computational Nanotechnol. 2022. T. 9, № 2. S. 35–44. doi: 10.33693/2313-223X-2022-9-2-35-44. (In Russ.).
7. Zagorskaya K. R., Ivanova I. V. Sinergija RPA i iskusstvennogo intellekta // Dnevnik nauki. 2023. № 11(83). S. 1–11. doi: 10.51691/2541-8327\_2023\_11\_8. (In Russ.).
8. Korytov P. V. Sravnitelnyj analiz metodov izvlechenija ključevyh slov bez predvaritelnogo obuchenija // Izv. SPbGETU «LETI». 2025. T. 18, № 7. S. 60–68. doi: 10.32603/2071-8985-2025-18-7-60-68. (In Russ.).
9. Opalev A. S., Alekseeva S. A. Metodicheskoe obosnovanie vybora optimalnyh rezhimov raboty oborudovaniya shemy stadialnogo vyvoda kontsentrata pri obogashhenii zheleznyh rud // Zap. Gornogo in-ta. 2022. T. 256. S. 593–602. doi: 10.31897/PMI.2022.80. (In Russ.).
10. Stojanova A. D., Trofimets V. Ya., Matrokhina K. V. Metodicheskij podhod k rejtingovoj otsenke kompanij Arkticheskoy zony na osnove pokazatelej ustojchivogo razvitiya // Gorn. informatsionno-analitich. byul. (nauch.-tehn. zhurn.). 2024. № 6. S. 149–162. doi: 10.25018/0236\_1493\_2024\_6\_0\_149. (In Russ.).
11. Neyrus S. K., Navatskaya V. A., Afanasev M. P. Issledovanie strukturnyh podrazdelenij promyshlennyh predpriyatij s pomoshhju sistem massovogo obsluzhivaniya // Mezhdunar. nauch.-issled. zhurn. 2022. № 12(126). S. 1–7. doi: 10.23670/IRJ.2022.126.32. (In Russ.).
12. Eroshin S. E. Formirovanie podkhodov k sovershenstvovaniju informatsionnyh sistem podderzhki prinjatija reshenij // Innovatsii i investitsii. 2021. № 10. S. 194–202. (In Russ.).
13. Dementev V. E., Kireev S. H. Vychor algoritmov mashinnogo obuchenija dlja klassifikatsii tekstovyh dokumentov // Tehn. sredstv svjazi. 2022. № 2(158). S. 22–52. doi: 10.24412/2782-2141-2022-2-22-52. (In Russ.).
14. Makarova E. A., Lagerev D. G. Model obrabotki slabostrukturirovannyh tekstovyh dannyh na rusском jazyke dlja intellektualnoj podderzhki informatsionnogo upravlenija v dinamicheskikh organizatsionnyh sistemah // Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshhestve. 2022. № 3. S. 104–125. doi: 10.21685/2227-8486-2022-3-7. (In Russ.).
15. Kvasov M. N., Mitryaev G. A., Prokhorov M. A. Razrabotka matematicheskoy modeli semanticheskogo analiza i sinteza dannyh // Izv. TulGU. Tekhn. nauki. 2023. № 11. S. 262–266. doi: 10.24412/2071-6168-2023-11-262-263. (In Russ.).

---

Information about the authors

**Olga V. Afanasieva** – Cand. Sci. (Eng.), Associate Professor of the Department of System Analysis and Management, Empress Catherine II Saint Petersburg Mining University, 21<sup>st</sup> line, 2, Vasilievsky Island, St. Petersburg, 199106, Russia.

E-mail: afanaseva\_ov@pers.spmi.ru

<https://orcid.org/0000-0003-3169-4781>

---

**Igor M. Novozhilov** – Cand. Sci. (Eng.), Associate Professor of the Department of Automation and Control Processes of Saint Petersburg Electrotechnical University.

E-mail: [novozhilovim@list.ru](mailto:novozhilovim@list.ru)

<https://orcid.org/0000-0002-2056-3930>

**Dmitry A. Pervukhin** – Dr Sci. (Eng.), Professor, Head of the Department of System Analysis and Management, Empress Catherine II Saint Petersburg Mining University, 21<sup>st</sup> line, 2, Vasilievsky Island, Saint Petersburg, 199106, Russia.

E-mail: [pervukhin\\_da@pers.spmi.ru](mailto:pervukhin_da@pers.spmi.ru)

<https://orcid.org/0000-0003-3547-2932>

**Pavel A. Skukovskii** – student of SAM-22 gr., Empress Catherine II Saint Petersburg Mining University, 21<sup>st</sup> line, 2, Vasilievsky Island, St. Petersburg, 199106, Russia.

E-mail: [Pavel22122004@yandex.ru](mailto:Pavel22122004@yandex.ru)

**Author contribution statement:**

Afanasieva O. V. – algorithm development, results validation and reproducibility testing.

Novozhilov I. M. – general management.

Pervukhin D. A. – research methodology development, conceptual model development.

Skukovskii P. A. – programming, coding, and software development for the research.

Статья поступила в редакцию 15.12.2025; принята к публикации после рецензирования 27.01.2026; опубликована онлайн 30.03.2026.

Submitted 15.12.2025; accepted 27.01.2026; published online 30.03.2026.

---