

DOI: 10.32517/0234-0453-2025-40-6-16-27



АВТОМАТИЗИРОВАННОЕ СРАВНЕНИЕ НАУЧНЫХ ИССЛЕДОВАНИЙ НА БАЗЕ АКАДЕМИЧЕСКОЙ ГЕНЕАЛОГИИ

А. Х. Мариносян¹ ✉, С. Г. Григорьев¹, И. М. Лернер^{2,3}, М. А. Аникьева⁴¹ *Московский городской педагогический университет, г. Москва, Россия*² *Казанский национальный исследовательский технический университет им. А. Н. Туполева-КАИ, г. Казань, Республика Татарстан, Россия*³ *Казанский (Приволжский) федеральный университет, г. Казань, Республика Татарстан, Россия*⁴ *Сибирский федеральный университет, г. Красноярск, Россия*✉ a.marinosityan@yandex.ru

Аннотация

Статья посвящена задаче анализа больших корпусов научных текстов. Отмечается, что традиционные наукометрические подходы, основанные на анализе цитирований, не всегда позволяют оценить содержательную близость идей и проследить интеллектуальную преемственность научных знаний. В работе предлагается алгоритм, сочетающий возможности больших языковых моделей (БЯМ) и количественных методов анализа. Подход включает два этапа. На первом этапе БЯМ используется для анализа содержания научной работы в соответствии с иерархическим классификатором и преобразования полученных данных в машиночитаемый формат в виде JSON-объекта. В результате каждой научной работе ставится в соответствие вектор ее тематического профиля. На втором этапе производится сравнение тематических профилей как отдельных работ, так и заранее заданных групп (научных школ) по расстояниям между центроидами и с применением методов кластерного анализа. Конечная цель подхода — перейти от формального анализа связей между исследователями (например, «научный руководитель — диссертант») к построению интеллектуальной генеалогии, отслеживающей эволюцию научных школ на основе содержательной близости их идей. Подчеркивается, что предложенный подход не заменяет экспертную оценку, а служит вспомогательным инструментом для навигации в научной литературе и выявления связей между исследованиями.

Ключевые слова: академическая генеалогия, наукометрия, сравнение научных работ, большие языковые модели, анализ содержания, графы знаний, диссертационные исследования, коэффициент силуэта.

Для цитирования:

Мариносян А. Х., Григорьев С. Г., Лернер И. М., Аникьева М. А. Автоматизированное сравнение научных исследований на базе академической генеалогии. *Информатика и образование*. 2025;40(6):16–27. DOI: 10.32517/0234-0453-2025-40-6-16-27.

AUTOMATED COMPARISON OF SCIENTIFIC RESEARCH BASED ON ACADEMIC GENEALOGY

A. Kh. Marinosyan¹ ✉, S. G. Grigoriev¹, I. M. Lerner^{2,3}, M. A. Anikieva⁴¹ *Moscow City University, Moscow, Russia*² *Kazan National Research Technical University named after A. N. Tupolev-KAI, Kazan, The Republic of Tatarstan, Russia*³ *Kazan (Volga Region) Federal University, Kazan, The Republic of Tatarstan, Russia*⁴ *Siberian Federal University, Krasnoyarsk, Russia*✉ a.marinosityan@yandex.ru

Abstract

The article addresses the problem of analyzing large corpora of scientific texts. It is noted that traditional scientometric approaches based on citation analysis do not always capture the substantive proximity of ideas or trace the intellectual continuity of scientific knowledge. The paper proposes an algorithm that combines the capabilities of large language models (LLMs) with quantitative analytical methods. The approach proceeds in two stages. At the first stage, LLM is used to automatically structure the text of a scientific paper and convert it into a machine-readable format. This yields, for each work, a vector representation of its thematic profile. At the second stage, thematic profiles of individual works and of pre-specified groups (schools of thought) are compared using centroid distances and clustering techniques. The ultimate goal is to move beyond formal relational analyses among researchers (e. g., “supervisor — dissertation candidate”) toward constructing an intellectual genealogy that tracks the evolution of schools of thought on the basis of the substantive proximity of their ideas. It is emphasized that the proposed approach does not replace expert assessment but serves as an auxiliary tool for navigating the scientific literature and identifying connections between studies.

© Мариносян А. Х., Григорьев С. Г., Лернер И. М., Аникьева М. А., 2025

Keywords: academic genealogy, scientometrics, comparison of scientific papers, large language models, content analysis, knowledge graphs, dissertation research, silhouette coefficient.

For citation:

Marinosyan A. Kh., Grigoriev S. G., Lerner I. M., Anikieva M. A. Automated comparison of scientific research based on academic genealogy. *Informatics and Education*. 2025;40(6):16–27. (In Russian.) DOI: 10.32517/0234-0453-2025-40-6-16-27.

1. Введение

В настоящее время в связи с переходом к информационному обществу наблюдается тенденция к быстрому росту объема научных публикаций, что ставит перед исследователями сложную задачу поиска нужных материалов в информационном потоке. Интегрированные распределенные информационные системы (ИРИС), используемые в научных базах данных [1], имеют фундаментальные ограничения методов и алгоритмов поиска требуемой научной информации. Наиболее значимыми ограничениями являются применяемые библиометрические и наукометрические подходы, основанные на анализе цитирований. С помощью ИРИС можно строить карты науки и оценивать влияние отдельных работ и авторов, но нельзя эффективно, без участия экспертов, разделять близкие тематические запросы, тем самым увеличивая «информационный шум». Анализ цитирований также не позволяет различить, является ссылка подерживающей или критической. Другое направление для ориентации в научных текстах — это **концепция академической генеалогии (АГ)**, которая исторически фокусировалась на отслеживании формальных связей «научный руководитель — диссертант» для построения деревьев научных школ [1–3]. Фундаментальное ограничение АГ в том, что формальная АГ слабо учитывает содержание самих научных работ и не отражает интеллектуальное влияние, которое часто выходит далеко за рамки прямого научного руководства и формирует «невидимые колледжи» исследователей [4]. Следовательно, стоит задача семантического анализа научных текстов для оценки содержательной близости научных идей и концепций, что позволит проследить интеллектуальную преемственность в развитии научных знаний. По сути, интеллектуальная преемственность отражает развитие научных школ, которые могут строиться по формальному признаку «учитель — ученик» или на основе содержательной близости исследований.

Переход от формальной к интеллектуальной академической генеалогии предполагает создание методов, которые позволят проводить глубокий, масштабный и, что особенно важно, автоматизированный анализ содержания научных текстов. Появление больших языковых моделей (БЯМ) предоставляет для этого уникальные возможности. Эти модели способны извлекать из текстов информацию, структурируя ее в соответствии с заранее заданными шаблонами, что позволяет преобразовывать неструктурированные данные в формат, доступный для машинного анализа и количественной обработки [1].

Цель представленного в данной статье исследования — разработка и апробация методики авто-

матизированного содержательного анализа текстов научных работ, позволяющей количественно оценивать и сравнивать тематические профили научных исследований для выявления интеллектуальной преемственности в развитии научных знаний.

Для достижения поставленной цели необходимо решить следующие **задачи**:

- 1) Разработать методику автоматизированного структурирования содержания научных текстов. Это включает создание алгоритма, который извлекает ключевую содержательную информацию из неструктурированного текста и преобразует ее в стандартизированный формат, отражающий ее тематический профиль. Данный подход должен быть масштабируемым, т. е. должна быть возможность быстрой обработки больших корпусов научных работ.
- 2) Создать и апробировать набор количественных метрик для сопоставления тематических профилей как отдельных научных работ, так и их кластеров (например, диссертаций, подготовленных в рамках одной научной школы).
- 3) Адаптировать разработанный инструментарий для решения задач интеллектуальной академической генеалогии. Необходимо продемонстрировать, как предложенная методика позволяет анализировать эволюцию научных школ, а именно выявлять и количественно оценивать процессы дивергенции (расхождения идей внутри одной школы) и конвергенции (тематического сближения идей между формально не связанными школами).

2. Обзор методов автоматизированного анализа научной информации

Для достижения поставленной цели проанализируем существующие методы автоматизированного анализа научной информации. Анализ работ [5–20] позволяет **классифицировать существующие методы и алгоритмы анализа научной литературы следующим образом**:

- *1-й класс*: алгоритмы и методы анализа на базе математической статистики;
- *2-й класс*: алгоритмы анализа, реализующие тематическое моделирование;
- *3-й класс*: алгоритмы анализа на базе извлечения структурированных данных и использования графов знаний;
- *4-й класс*: алгоритмы анализа, использующие результаты, полученные с помощью больших языковых моделей.

В основе **первого класса алгоритмов и методов анализа научной информации** — на базе матема-

тической статистики — лежит упрощающее допущение о тексте как о «мешке слов», т. е. порядок слов и синтаксические связи в исследуемой научной работе игнорируются. Классической реализацией такого подхода является анализ сопутствующих слов, который позволяет выявлять ключевые темы и связи между ними на основе частоты совместного появления терминов в документах. Идея заключается в том, что часто и устойчиво встречающиеся вместе термины, вероятно, описывают одну и ту же концептуальную область. Так, в работе [5] было показано, что применение указанного класса методов позволяет строить «социокогнитивные сети» научных направлений, визуализируя структуру и динамику исследовательских полей.

Однако эти методы сталкиваются с *фундаментальным ограничением*, а именно: невозможность учета синонимии и полисемии, игнорирование контекстуальных и грамматических связей.

В частности, это отражено в работах [6–10], в которых предпринимается попытка преодоления данного ограничения, однако это приводит к *дополнительным ограничениям*, являющимся следствием указанного выше ограничения:

- 1) время, в течение которого формируются правила классификации научной информации, возрастает при увеличении размера базы данных;
- 2) точность классификации научной информации с учетом потребностей пользователей зависит от степени подобия запросов пользователей;
- 3) проблема холодного старта;
- 4) отсутствие подходов к решению проблемы априорной оптимальной классификации данных;
- 5) необходимость большой репрезентативной выборки для определения значимой научной информации с высокой достоверностью.

Второй класс методов — тематическое моделирование — основан в большинстве случаев на методе скрытого распределения Дирихле (*англ.* Latent Dirichlet Allocation, LDA) [11]. LDA рассматривает каждый документ как уникальную «смесь» скрытых тем, а каждую тему — как распределение вероятностей на множестве слов. Этот подход позволяет автоматически, без предварительной разметки, выявлять тематические кластеры в больших корпусах текстов. В наукометрии LDA фактически является стандартным инструментом для проведения тематического оценочного анализа. Практическая реализация работы данного инструмента продемонстрирована в работах [12, 13]. Тематическое моделирование легло в основу «умных обзоров литературы», позволяя исследователям быстро получить общее представление о тематическом ландшафте незнакомой предметной области [14].

Тем не менее у *LDA* имеются *существенные ограничения*, которые не позволяют решить задачу глубокого семантического анализа:

- Во-первых, результат работы модели — это лишь набор тем, каждая из которых представ-

ляет собой список слов, ранжированных по вероятности их принадлежности к этой теме. Интерпретация этих тем осуществляется экспертами.

- Во-вторых, выбор количества тем является экспертной задачей, существенно влияющей на результат классификации документов по темам. Если задать слишком малое число тем, модель объединит разные по смыслу области в одну, сделав ее слишком общей и малополезной. Если же задать слишком большое число, одна и та же научная область будет искусственно раздроблена на множество мелких, пересекающихся подтем, что также делает тематическую классификацию малоэффективной.
- В-третьих, и это самое главное ограничение, стандартный LDA не выявляет явные логические связи между концепциями и не учитывает динамику изменения тематик во времени. Для преодоления этого ограничения были предложены расширения LDA, такие как динамические тематические модели, которые отслеживают эволюцию тем во времени [14]. Однако даже эти более сложные модели все еще не способны извлекать из текста направленные, структурированные факты.

К третьему классу методов — извлечение структурированных данных и использование графов знаний — относятся методы и подходы, которые реализуют направление, известное как извлечение информации. Происходит преобразование неструктурированного текста в структурированные данные, что означает автоматическое выявление упоминаний методов, материалов, метрик, задач и связей между ними. Для решения этой задачи могут, в частности, использоваться специализированные языковые модели, такие как SciBERT [15].

Процесс извлечения структурированных данных из текста многоэтапный:

- на первом этапе выполняется распознавание именованных сущностей (ИС): в тексте выделяются упоминания ключевых объектов (персоны, организации, термины и т. д.);
- на втором этапе осуществляется извлечение (выявление) отношений (ИО): между найденными сущностями устанавливаются семантические связи;
- на третьем этапе проводится связывание сущностей, т. е. ИС сопоставляются с идентификаторами в базах знаний, что позволяет агрегировать данные из разных источников.

Полученные таким образом структурированные данные подходят для представления в виде графов знаний. В таких графах узлами являются сущности (статьи, авторы, концепции, химические соединения и пр.), а типизированными ребрами — отношения между ними [16]. Графы знаний открывают возможность для выполнения сложных запросов, автоматического построения сравнительных таблиц и визуализации данных, например, в проекте Open Research

Knowledge Graph (ORKG) [17]. Аналогичные графы знаний активно строятся в узкоспециализированных областях, например, в биомедицине для моделирования взаимодействий между генами и лекарствами [18]. Практическая реализация графов знаний требует создания специализированных программных модулей, которые автоматизировали бы процессы наполнения и, что критически важно, обеспечивали бы контроль целостности вносимых данных [19].

Алгоритм построения графов знаний имеет серьезные *ограничения*:

- ошибки распознавания именованных сущностей на ранних этапах каскадом распространяются на последующие этапы;
- экспоненциальный рост требуемых вычислительных ресурсов при реализации систем анализа в рамках исследуемого научного направления, что наиболее критично при обучении каждой отдельной модели под конкретную предметную область и используемую ею онтологию.

К четвертому классу относятся **методы анализа научной информации, построенные на базе больших языковых моделей (БЯМ)**. Ключевым преимуществом БЯМ является их гибкость: одна и та же модель может быть адаптирована к разным предметным областям и типам задач путем изменения текстовой информации, обеспечивающей управление запросами к нейросети. Это приводит к исключению необходимости в создании и поддержке узкоспециализированных, дорогих в разработке моделей для каждой

научной дисциплины. Недавние исследования в данной области [20] показывают, что современные БЯМ успешно справляются с задачей извлечения с высокой точностью ключевых данных (таких, как цель, метод, результаты и выводы) из научных статей.

Однако *БЯМ обладает фундаментальными особенностями, которые существенно влияют на качество получаемых результатов*:

- Проблема «галлюцинаций» — генерации фактически неверной, но правдоподобной информации. В научном контексте этот риск особенно опасен, что было ярко продемонстрировано на примере модели Galactica, отозванной из публичного доступа из-за систематических ошибок и выдуманных ссылок [21]¹.
- Результаты работы системы, использующей БЯМ, критически зависят от формулировки текстовой информации, обеспечивающей управление запросами к нейросети [3]. Незначительные изменения в конструкции могут привести к совершенно разным результатам, что ставит под вопрос воспроизводимость и надежность получаемых результатов.

В таблице 1 приведен сравнительный анализ методов автоматизированного анализа текстов.

¹ Heaven W. D. Why Meta's latest large language model survived only three days online. *MIT Technology Review*. November 18, 2022. <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/>

Таблица 1 / Table 1

Сравнительный анализ методов автоматизированного анализа текстов

Comparative analysis of automated text analysis methods

№ п/п	Класс методов	Достоинства	Ключевые ограничения
1	Алгоритмы и методы анализа на базе математической статистики	<ul style="list-style-type: none"> • Простота использования. • Выявление устойчивых групп часто встречающихся понятий, вокруг которых организуется содержание 	<ul style="list-style-type: none"> • Невозможность обеспечения анализа узкоспециализированных областей знаний. • Невозможность проведения глубокого семантического анализа. • Отсутствие учета контекста
2	Алгоритмы анализа, реализующие тематическое моделирование	<ul style="list-style-type: none"> • Автоматическая кластеризация анализируемой научной информации. • Учет контекста и семантической близости. • Машиночитаемость получаемых результатов 	<ul style="list-style-type: none"> • Экспоненциальный рост требуемых вычислительных ресурсов при увеличении объема обрабатываемой научной информации
3	Алгоритмы анализа на базе извлечения структурированных данных и использования графов знаний	<ul style="list-style-type: none"> • Структурированное представление знаний. • Учет семантических связей и контекста 	<ul style="list-style-type: none"> • Каскадное распространение ошибок при многоэтапном процессе извлечения данных. • Требуется создание и поддержка жесткой, узкоспециализированной онтологии для каждой предметной области
4	Алгоритмы анализа, использующие результаты, полученные с помощью больших языковых моделей	<ul style="list-style-type: none"> • Гибкость. • Одноступенчатый анализ 	<ul style="list-style-type: none"> • Риск «галлюцинаций» (генерации правдоподобной, но ложной информации). • Нестабильность (результат сильно зависит от формулировки запроса, что снижает воспроизводимость). • Модель склонна давать один «обобщенный» ответ, не учитывающий нюансы

Анализ таблицы 1 позволяет сделать следующие выводы относительно формирования нового подхода к автоматизированному семантическому анализу научной информации:

1. Для преодоления существующих недостатков перечисленных в таблице 1 классов методов анализа научной информации необходимо использовать гибридную схему обработки научной информации.
2. На первичном этапе обработки научной информации целесообразно использовать БЯМ с применением справочника-классификатора для формирования тематических векторов, формализующих тематический профиль научной работы.
3. Вторичная обработка должна производиться с помощью методов математической статистики и теории информации, поскольку, согласно В. П. Беспалько [22], именно они должны обеспечить наибольший прорыв в области программированного образования.

3. Методика автоматизированного содержательного анализа текстов научных работ

Для достижения поставленной цели был разработан алгоритм, этапы работы которого представлены ниже.

3.1. Этап 1. Построение тематического классификатора научной работы

С учетом паспортов научных специальностей номенклатуры научных специальностей БЯМ производит построение тематического классификатора научной работы, структура которого является иерархической. Инструкция, которой следует БЯМ при построении классификатора, оформлена в форме запроса, основными компонентами которого являются следующие вопросы:

- Что исследовали?
- Как исследовали?
- Что получили?

1	<СПИСОК_КЛАССИФИКАТОРА>
2	1.1. Среда и предметная область
3	1.1.1. Уровень формального образования
4	1.1.1.1. Дошкольное образование
5	1.1.1.2. Школьное образование
6	1.1.1.2.1. Начальное общее образование
7	1.1.1.2.2. Основное общее образование
8	1.1.1.2.3. Среднее общее образование
9	1.1.1.3. Среднее профессиональное образование (СПО)
10	1.1.1.4. Высшее образование
11	1.1.1.4.1. Высшее образование – Бакалавриат
12	1.1.1.4.2. Высшее образование – Специалитет
13	1.1.1.4.3. Высшее образование – Магистратура
14	1.1.1.5. Подготовка кадров высшей квалификации
15	1.1.1.5.1. Подготовка кадров высшей квалификации – Аспирантура
16	1.1.1.5.2. Подготовка кадров высшей квалификации – Ординатура
17	1.1.1.6. Дополнительное профессиональное образование
18	1.1.1.6.1. Дополнительное профессиональное образование – Повышение • квалификации
19	1.1.1.6.2. Дополнительное профессиональное образование – Профессиональная • переподготовка
20	1.1.1.7. Дополнительное образование детей и взрослых
21	1.1.1.8. Профессиональное обучение
22	1.1.2. Предметный цикл / Область знания
23	1.1.2.1. Математические и естественнонаучные дисциплины
24	1.1.2.1.1. Математика
25	1.1.2.1.2. Информатика
26	1.1.2.1.3. Физика
27	1.1.2.1.4. Химия
28	1.1.2.1.5. Биология
29	1.1.2.1.6. Физическая география
30	1.1.2.1.7. Другие естественнонаучные дисциплины

Рис. 1. Пример реализации классификатора для формирования тематического профиля диссертации

Fig. 1. Example of hierarchical classifier for composing the thematic profile of a dissertation

Пример реализации такого классификатора представлен на рисунке 1.

Полученный классификатор позволяет декомпозировать сложный научный текст на независимые, но взаимосвязанные элементы и обеспечивает воспроизводимость извлекаемых из текста данных.

3.2. Этап 2. Формирование тематического профиля диссертации с помощью БЯМ

Под *тематическим профилем* понимается структурированное описание содержания диссертации по узлам предметного классификатора. Для каждого узла классификатора БЯМ выставляет оценку по разработанной 10-балльной шкале релевантности:

- 10 баллов присваивается, если предмет/тема/метод/качество/объект, заданный элементом классификатора, является центральным для исследования;
- 8–9 баллов — если элемент классификатора является одним из ключевых факторов исследования;
- 5–7 баллов — если элемент классификатора соответствует вспомогательному связующему звену исследования;

- 1–4 балла — когда элемент классификатора в исследовании имеет фоновое упоминание.

Пример запроса к БЯМ, обеспечивающий получение тематического профиля диссертации, представлен на рисунке 2.

Тематический профиль сохраняется в JSON-файле и помимо балла может содержать служебные поля для воспроизводимости (идентификатор документа, идентификатор узла и т. д.).

Для обеспечения возможности математического сопоставления различных работ этот профиль преобразуется в *тематический вектор* — числовой вектор фиксированной длины, где каждая позиция соответствует своему узлу во всем классификаторе. В него переносятся оценки из тематического профиля, а всем узлам, не упомянутым в профиле данной работы, присваивается нулевое значение.

Предлагаемый процесс стандартизации позволяет представить каждую диссертацию в едином многомерном пространстве и корректно применять к ней метрики для вычисления расстояний и последующей кластеризации.

```

1 Ты – экспертный AI-аналитик, специализирующийся на систематизации научных
• текстов в области педагогики. Твоя задача – провести исчерпывающий анализ
• предоставленного текста диссертации (или ее автореферата) и сгенерировать
• результат в виде единого, строго структурированного JSON-объекта.
2
3 **ЗАДАЧА:**
4 Проанализируй следующий текст автореферата диссертации для оценки
• релевантности каждого элемента из предоставленного классификатора.
5
6 **ИНСТРУКЦИЯ ПО АНАЛИЗУ:**
7 Проанализируй текст и для **каждого элементарного (конечного) пункта** из
• классификатора выше оцени степень его присутствия и важности в диссертации.
8 - Используй целочисленную шкалу от 1 до 10.
9 - **1:** Элемент подразумевается, но практически нерелевантен.
10 - **2-4:** Элемент упоминается вскользь, является фоновым.
11 - **5-7:** Элемент является важной, но не центральной частью исследования.
12 - **8-9:** Элемент является одной из ключевых тем исследования.
13 - **10:** Элемент является абсолютным ядром, центральной темой
• исследования.
14 Если элемент в принципе не имеет отношения к диссертации (как бы, если
• оценка была бы 0), то тогда не надо в принципе включать этот элемент в
• поле `basis_scores`.
15 Результат этого шага помести в JSON-поле `basis_scores`.
16
17 **ИНСТРУМЕНТАРИЙ (Классификационный базис):**
18 Ты ДОЛЖЕН использовать для анализа ИСКЛЮЧИТЕЛЬНО следующую иерархическую
• систему. Оцениваются только элементарные (самые глубокие) уровни иерархии.
19
20 ...
21 <СПИСОК_КЛАССИФИКАТОРА>
22 1.1. Среда и предметная область
23 1.1.1. Уровень формального образования

```

Рис. 2. Инструкция для БЯМ для формирования тематического профиля диссертации

Fig. 2. Prompt for LLM to form the thematic profile of a dissertation

3.3. Этап 3. Анализ динамики развития научных школ на соответствие двум сценариям развития

1. **Сценарий дивергенции** соответствует случаю, когда в рамках потомков одного научного руководителя со временем формируются два или более кластера диссертаций с отчетливо разными тематическими профилями, что позволяет говорить о зарождении новых, самостоятельных научных школ из одного корня.
2. **Сценарий конвергенции** соответствует случаю, когда научные работы, восходящие к разным основателям, демонстрируют высокую близость тематических профилей, что свидетельствует о формировании единого научного направления или парадигмы, которая преодолевает формальные границы школ. Таким образом, предлагаемый алгоритм позволяет увидеть, как формальная генеалогия преобразуется в интеллектуальную, и зафиксировать зарождения и слияния научных идей.

Оценка развития того или иного сценария производится с помощью двух видов анализа, примененных к полученным тематическим профилям диссертационных работ, а именно:

- 1) **пермутационного многомерного дисперсионного анализа (PERMANOVA)** — это непараметрический статистический тест, позволяющий проверить, различаются ли центроиды двух или более кластеров (групп) в многомерном пространстве. В его основе лежит идея разделения общей вариации данных на внутригрупповую и межгрупповую [23];
- 2) **оценки коэффициента силуэта** — метрики для оценки качества кластеризации на уровне отдельных объектов. Коэффициент силуэта показывает, насколько объект схож со своим собственным кластером (сплоченность) по сравнению с другими кластерами (обособленность). Это позволяет выявить как хорошо разделенные кластеры, так и объекты, которые находятся на границе или могут быть ошибочно классифицированы [24].

Таким образом, выбор этих видов анализа обусловлен тем, что PERMANOVA позволяет получить статистическую оценку степени различий между научными группами по их тематическим профилям, а коэффициент силуэта — метрическую оценку качества группировки, показывая соотношение внутригрупповой близости и межгруппового расстояния для отдельных работ и в среднем по группам.

В предлагаемом нами алгоритме выбранные методы анализа применяются к центроидному вектору тематических исследований научного руководителя/консультанта — среднему тематическому вектору по всем диссертациям, которые были защищены под руководством одного и того же научного руководителя/консультанта. Значение центроидного вектора определяется на основе тематических профилей диссертационных работ.

Параметр псевдо-F-статистики, рассчитываемый в ходе PERMANOVA, является показателем величины различий между сравниваемыми группами текстов. Значение F представляет собой отношение межгрупповой вариации (насколько центроиды кластеров отстоят друг от друга) к внутригрупповой вариации (насколько работы разбросаны внутри своих кластеров). Большое значение F (значительно больше 1) свидетельствует о том, что различия между кластерами превышают различия внутри них, что является признаком четкой тематической обособленности и хорошей разделенности групп. Малое значение F (близкое к 1 или меньше) указывает на то, что кластеры сильно перекрываются и различия между ними сопоставимы или даже меньше, чем внутреннее разнообразие каждой группы.

Параметр p-value — это условная вероятность (в случае истинности нулевой гипотезы) получить при случайном перераспределении принадлежность объектов по кластерам. В нашем случае нулевая гипотеза заключается в том, что тематические профили сравниваемых групп (кластеров, научных школ) не различаются и любые наблюдаемые различия в данных являются случайными. Таким образом, p-value отвечает на вопрос: «Если бы эти научные школы на самом деле были одинаковыми по тематике, каков шанс, что мы чисто случайно увидели бы в нашей выборке такое сильное различие между ними?» Теоретически p-value может принимать значения от 0 до 1. Если значение p-value ниже уровня значимости (например, $p < 0,05$), это позволяет отклонить нулевую гипотезу и сделать вывод о наличии статистически значимых различий между тематическими профилями сравниваемых научных школ. Если p-value превышает уровень значимости (например, $p > 0,05$), то у нас нет оснований отклонять нулевую гипотезу, т. е. статистически значимых различий между тематическими профилями кластеров не обнаружено.

Коэффициент силуэта (англ. silhouette score), в свою очередь, является метрикой, предназначенной для оценки качества и структурной целостности полученной кластеризации. Его расчет основан на сравнении двух величин для каждого объекта: среднего расстояния до объектов своего кластера (показатель компактности) и среднего расстояния до объектов ближайшего соседнего кластера (показатель обособленности).

Значение коэффициента силуэта для отдельного объекта варьируется от -1 до $+1$ и интерпретируется следующим образом:

- Значение, близкое к $+1$, указывает на то, что внутрикластерное расстояние для данного объекта значительно меньше, чем расстояние до ближайшего соседнего кластера. Это является признаком того, что объект уверенно принадлежит своему кластеру и хорошо отстоит от других.
- Значение, близкое к 0 , свидетельствует о том, что объект находится на границе двух кластеров и его принадлежность к одному из них неоднозначна.

- Отрицательное значение (близкое к -1) указывает на то, что объект в среднем ближе к членам соседнего кластера, чем к членам своего собственного. Это является прямым признаком возможной ошибочной классификации.

Помимо индивидуальных значений для каждого объекта рассчитывается среднеарифметическое значение коэффициента силуэта для всей выборки. Более высокое среднее значение указывает на более четкую и осмысленную кластерную структуру, где группы являются плотными и хорошо разделенными.

4. Апробация и анализ результатов

Апробация разработанного алгоритма была произведена путем анализа научной школы члена-корреспондента РАО, доктора технических наук, профессора, профессора МГПУ С. Г. Григорьева, состоящей на конец 2024 года из более чем 120 исследователей с ученой степенью. Выбор научной школы обусловлен репрезентативностью выборки, а также широким охватом научных интересов ее представителей.

С помощью применения разработанного программного обеспечения было построено дерево академической генеалогии научной школы профессора С. Г. Григорьева, которое показало, что в его научной школе имеется 17 научных руководителей¹, которые являются учениками ее основателя в первом или последующих поколениях академического родства и которые сами стали научными руководителями защитившихся диссертантов.

Программная реализация пермутационного многомерного дисперсионного анализа осуществлялась в среде программирования Python с помощью библиотеки scikit-bio². Реализация оценки коэффициента силуэта была произведена также в среде программирования Python с помощью применения библиотеки scikit-learn³. Настройки библиотек были заданы по умолчанию.

Полученные результаты анализа представлены в таблице 2.

Анализ данных, представленных в таблице 2, позволяет сделать следующие выводы:

1. Тематические профили диссертаций А. Г. Абросимова, Е. Ы. Бидайбекова, В. В. Гриншкун, В. Е. Жужжалова, О. Ю. Заславской, В. А. Ку-

динова и Н. И. Попова имеют большое перекрытие с профилем диссертаций, защищенных под руководством С. Г. Григорьева, что следует из того факта, что коэффициент силуэта принимает отрицательные значения. В частности, несмотря на то что В. В. Гриншкун является руководителем значительного числа диссертаций (16 работ), их тематический профиль практически не отличается от работ С. Г. Григорьева, что подтверждается отрицательным значением коэффициента силуэта ($-0,003668$) и большим значением параметра $p = 0,4$ и указывает на продолжение основного тематического направления школы. Таким образом, работы, защищенные под руководством В. В. Гриншкун, не образуют отдельного тематического кластера.

2. Кластер диссертаций, выполненных под руководством О. А. Козлова (28 работ) имеет малое, но положительное значение коэффициента силуэта ($0,017$), а значение параметра $p = 0,006$ указывает на то, что тематический профиль работ, выполненных под руководством О. А. Козлова, статистически отличается от тематики научной школы С. Г. Григорьева, но работы по своим тематикам значительно перекрываются с основным направлением исследований С. Г. Григорьева. Вышеказанное наглядно подтверждается силуэтным графиком, представленным на рисунке 3. Вертикальная пунктирная линия на графике показывает средний коэффициент силуэта для всех работ, равный $0,017$. В обоих кластерах присутствуют области с отрицательными значениями коэффициента силуэта. Это означает, что некоторые диссертации из одного кластера по своему тематическому профилю оказываются ближе к центроиду другого кластера, что является явным визуальным признаком их смешения и отсутствия четкой тематической границы.
3. Наибольшие значения коэффициента силуэта, а также наибольшие расстояния между центроидами наблюдаются у работ, выполненных под руководством О. В. Соколовой и Е. В. Яковлевой. Экспертный анализ подтверждает, что для диссертаций под руководством С. Г. Григорьева центральной темой являются информационные технологии в образовании, а работы Е. В. Яковлевой и О. В. Соколовой в используемой выборке этой теме не соответствуют. Однако, поскольку под их руководством в выборке имеется лишь по одной работе, делать вывод о формировании отдельных научных школ преждевременно.

По итогам апробации на других примерах был сделан вывод о том, что коэффициент силуэта предоставляет более релевантную информацию для определения тематической близости кластеров диссертаций и выявления дивергенции научных школ, чем результаты, получаемые с помощью пермутационного многомерного дисперсионного анализа.

¹ Следует отметить, что количество диссертаций, защищенных под руководством конкретного исследователя, и, соответственно, количество научных руководителей в научной школе рассчитывалось на основе собранной нами автоматизированными методами базы данных авторефератов диссертаций. Собранная нами база данных содержит практически все отсканированные РГБ и РНБ авторефераты по педагогическим наукам. Но в силу того, что далеко не все авторефераты были отсканированы, приведенные числа могут быть существенно ниже реальных.

² См. документацию метода PERMANOVA: <https://scikit.bio/docs/dev/generated/skbio.stats.distance.permanova.html>

³ См. документацию метода silhouette score: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html

Сравнение тематических профилей диссертаций, защищенных под непосредственным руководством С. Г. Григорьева, с тематическими профилями работ, защищенных под руководством ученых, принадлежащих к его научной школе в рамках академического генеалогического подхода

Comparison of thematic profiles of dissertations defended under the direct supervision of S. G. Grigoriev with thematic profiles of works defended under the supervision of scholars belonging to his scientific school within the framework of the academic genealogical approach

№ п/п	Научный руководитель	Количество диссертаций	Расстояние между центроидами	PERMANOVA (F)	PERMANOVA (p-value)	Коэффициент силуэта
1	Абросимов Александр Григорьевич	2	17,520525	0,623080	0,952	-0,095959
2	Жужжалов Валерий Евгеньевич	1	28,326409	0,830366	0,753	-0,043762
3	Заславская Ольга Юрьевна	3	14,854115	0,641750	0,952	-0,039485
4	Попов Николай Иванович	3	16,876173	0,837339	0,678	-0,035783
5	Бидайбеков Есен Ыкласович	3	17,705137	0,908826	0,554	-0,009272
6	Кудинов Виталий Алексеевич	8	11,523026	0,881385	0,590	-0,008507
7	Гриншкун Вадим Валерьевич	16	9,634963	1,045529	0,400	-0,003668
8	Макаров Сергей Иванович	4	20,353731	1,582866	0,055	0,012112
9	Козлов Олег Александрович	28	11,746444	2,032357	0,006	0,017492
10	Рудинский Игорь Давидович	6	17,051292	1,535580	0,051	0,026643
11	Левченко Ирина Витальевна	4	23,914369	2,163411	0,012	0,056011
12	Гейн Александр Георгиевич	4	22,336210	1,849012	0,028	0,062112
13	Лазарева Ирина Анатольевна	1	35,555594	1,308286	0,170	0,065958
14	Атанасян Сергей Леонович	2	30,202347	1,828316	0,016	0,088568
15	Поличка Анатолий Егорович	1	37,366755	1,444966	0,080	0,097195
16	Яковлева Елена Викторовна	1	40,892998	1,730552	0,033	0,144924
17	Соколова Ольга Владимировна	1	44,240675	2,025491	0,031	0,181243

Комментарии:

1) Низкое значение параметра p (менее 0,05) указывает на статистически значимое различие между тематическими профилями кластеров.

2) Значение коэффициента силуэта, близкое к +1, означает, что кластеры плотные и хорошо разделены; значение, стремящееся к 0, указывает на их сильное перекрытие; отрицательное значение коэффициента силуэта говорит о том, что объекты, вероятно, отнесены к неверному кластеру.

Коэффициент силуэта позволяет оценить не просто статистическую значимость различия центроидов, а реальную степень разделенности и обособленности тематических кластеров.

Дополнительно были получены результаты анализа научных школ, не связанных формальным родством. Апробация была произведена посредством сравнения тематических профилей диссертаций научной школы С. Г. Григорьева, специализирующейся на информатизации образования, с тематическими профилями диссертаций по педагогическим наукам научных школ А. Г. Мордковича и Л. А. Миловановой. Следует отметить, что если научные школы А. Г. Мордковича и С. Г. Григорьева, деятельность которых на протяжении многих лет связана с Институтом цифрового образования МГПУ, представляют смежные предметные области, то научная школа Л. А. Миловановой из Института иностранных язы-

ков МГПУ представляет иную предметную область (методика преподавания иностранных языков). Результаты количественного сравнения приведены в таблице 3.

Школа А. Г. Мордковича имеет малое значение коэффициента силуэта, равное 0,019622, что указывает на слабую тематическую разделенность со школой С. Г. Григорьева. Визуальный анализ силуэтного профиля (рис. 4, а) подтверждает этот вывод. Среднее значение коэффициента силуэта (пунктирная линия) находится близко к нулю, а значительная часть кластера школы С. Г. Григорьева (около половины) имеет отрицательные значения, что указывает на их высокую тематическую близость с работами, выполненными в научной школе А. Г. Мордковича. Иными словами, четкая тематическая граница между школами, работающими в смежных областях и в рамках одного института, отсутствует.

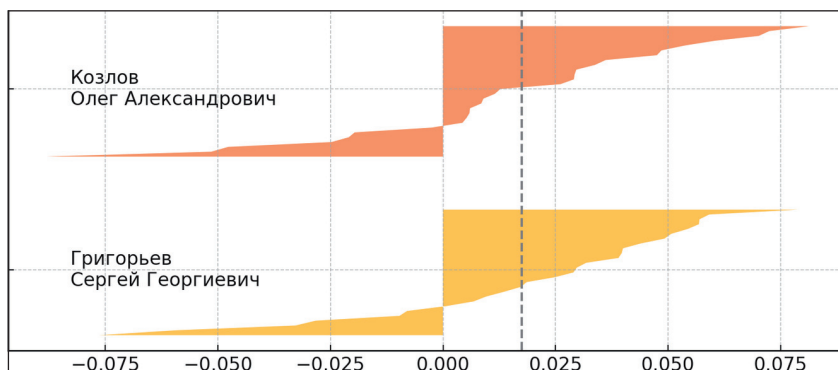


Рис. 3. Силуэтный график для кластеров диссертаций, защищенных под руководством С.Г. Григорьева и О.А. Козлова
 Fig. 3. Silhouette graph for clusters of dissertations defended under the supervision of S.G. Grigoriev and O.A. Kozlov

Таблица 3 / Table 3

Сравнение тематических профилей научной школы С. Г. Григорьева с другими научными школами
Comparison of thematic profiles of the scientific school of S. G. Grigoriev with other scientific schools

Научный руководитель	Количество диссертаций	Расстояние между центроидами	PERMANOVA (F)	PERMANOVA (p-value)	Коэффициент силуэта
Мордкович Александр Григорьевич	10	19,609013	3,815193	0,001	0,019622
Милованова Людмила Анатольевна	11	20,073797	4,246764	0,001	0,080422

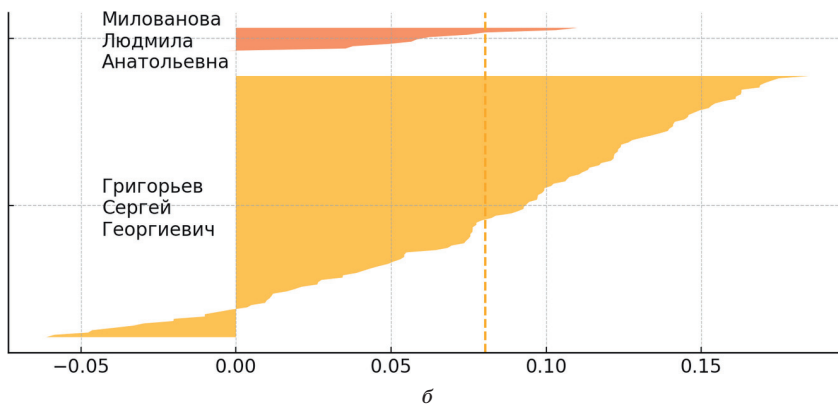
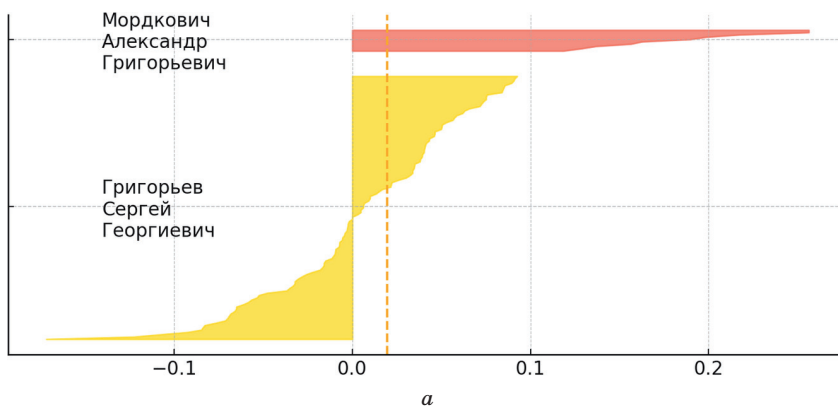


Рис. 4. Силуэтный график тематического сравнения:
 а — научных школ С. Г. Григорьева и А. Г. Мордковича; б — научных школ С. Г. Григорьева и Л. А. Миловановой

Fig. 4. Silhouette graph of thematic comparison:
 a — scientific schools of S. G. Grigoriev and A. G. Mordkovich; b — scientific schools of S. G. Grigoriev and L. A. Milovanova

Иная картина наблюдается при сравнении со школой Л. А. Миловановой, представляющей другую предметную область. Коэффициент силуэта равен 0,080422, что свидетельствует о большей тематической обособленности в сравнении с работами школы С. Г. Григорьева. Силуэтный профиль (рис. 4, б) демонстрирует, что оба кластера преимущественно находятся в положительной зоне, а их «лезвия» не имеют значительных провалов в отрицательную область. Это указывает на формирование двух четко различимых тематических ядер, соответствующих разным научным направлениям.

Полученные результаты позволяют говорить о том, что представленный в данной работе алгоритм позволяет получить результаты о тематической близости или удаленности научных школ и количественно оценить степень этой близости, что обладает значимостью для построения карт научного ландшафта, основанных на содержательном анализе текстов.

5. Заключение

В данной работе представлен и апробирован алгоритм создания интеллектуальной академической генеалогии, направленный на решение проблемы масштабного анализа и сравнения научных исследований. Предложенный подход, основанный на двухэтапном процессе — структурировании текстов с помощью больших языковых моделей и последующем применении метрик кластерного анализа, — позволяет перейти от анализа формальных связей к оценке содержательной близости и эволюции научных идей.

Было продемонстрировано, что на основе извлеченных многоаспектных данных можно количественно оценивать тематическую близость работ, а также структурную целостность и степень обособленности их кластеров, их уровень общности или специфичности, а также выявлять характер различий между ними. Совмещение этих инструментов с методами академической генеалогии позволяет строить карты научного ландшафта, выявлять интеллектуальные школы и отслеживать их конвергенцию или дивергенцию.

Предложенный алгоритм не заменит экспертное знание, но предоставит исследователям инструмент для навигации в растущих объемах научной информации. Дальнейшее развитие этого подхода в рамках человеко-машинных систем открывает перспективы для создания динамических, интерактивных карт науки, способных в реальном времени отражать структуру и развитие научного знания.

Финансирование

Исследование выполнено в рамках реализации гранта «Методика автоматизации создания академической генеалогии ученых МГПУ». Приказ Департамента образования и науки города Москвы от 18 апреля 2025 года № Пр-354.

Funding

The study was carried out within the framework of the grant “Methodology for automating the creation of academic genealogy of scientists of Moscow City University”. Order of the Department of Education and Science of the City of Moscow dated April 18, 2025 No. Pr-354.

Список источников / References

1. Григорьев С. Г., Лернер И. М., Мариносян А. Х., Арутюнова Н. К., Григорьева М. А. К вопросу подбора учебно-методической информации для реализации адаптивной электронной образовательной среды: алгоритм априорной классификации авторов. *Информатика и образование*. 2025;40(2):66–78. EDN: YDTFTG. DOI: 10.32517/0234-0453-2025-40-2-66-78.
2. Григорьев С. Г., Лернер И. М., Мариносян А. Х., Григорьева М. А. К вопросу подбора учебно-методической информации для реализации адаптивной электронной образовательной среды: алгоритм подбора авторов литературы с учетом эмоционально-психологических особенностей пользователей на базе идей академической генеалогии. *Информатика и образование*. 2025;40(3):69–79. EDN: DVPEUJ. DOI: 10.32517/0234-0453-2025-40-3-69-79.
3. Лернер И. М., Мариносян А. Х., Григорьев С. Г., Юсупов А. Р., Аникьева М. А., Гарифуллина Г. А. Подход к формированию интеллектуальной академической генеалогии с использованием больших языковых моделей. *Электромеханические волны и электронные системы*. 2024;29(4):108–120. EDN: BSNYZW. DOI: 10.18127/j5604128-202404-09.
4. Lerner I. M., Marinosyan A. Kh., Grigoriev S. G., Yusupov A. R., Anikieva M. A., Garifullina G. A. An approach to the formation of intellectual academic genealogy using large language models. *Electromagnetic Waves and Electronic Systems*. 2024;29(4):108–120. (In Russian.) EDN: BSNYZW. DOI: 10.18127/j5604128-202404-09.]
5. de Solla Price D. J., Beaver D. Collaboration in an invisible college. *American Psychologist*. 1966;21(11):1011–1018. DOI: 10.1037/h0024051.
6. Gallon M., Law J., Rip A. (Eds) Mapping the dynamics of science and technology: Sociology of Science in the Real World. Houndmills: Macmillan Press; 1986. 242 p.
7. Salton G., Wong A., Yang C. S. A vector space model for automatic indexing. *Communications of the ACM*. 1975;18(11):613–620. DOI: 10.1145/361219.361220.
8. Lu Yonghe, Li Yanfeng. Improvement of text feature weighting method based on TF-IDF algorithm. *Library and Information Service*. 2013;57(3):90–95. (In Chinese.)
9. Herlocker J. L., Konstan J. A., Terveen L. G., Riedl J. T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*. 2004;1(22):5–53. DOI: 10.1145/963770.963772.
10. Ji W., Liu S., Song Y., Qi J. Research of Intelligent recommendation system based on the user and association rules mining for books. *Proc. 2015 5th Int. Conf. on Computer Sciences and Automation Engineering*. Atlantis Press; 2016:294–299. DOI: 10.2991/iccsae-15.2016.56.
11. Yao Y. Library resource vertical search engine based on ontology. *2017 Int. Conf. on Smart Grid and Electrical Automation (ICSGEA)*. IEEE, Changsha, China; 2017:672–675. DOI: 10.1109/ICSGEA.2017.159.

11. Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet allocation. *Journal of Machine Learning Research*. 2003;3:993–1022.

12. Griffiths T. L., Steyvers M. Finding scientific topics. *Proc. of the National Academy of Sciences*. 2004;101(suppl_1):5228–5235. DOI: 10.1073/pnas.0307752101.

13. Asmussen C. B., Møller C. Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*. 2019;6(1):1–18. DOI: 10.1186/s40537-019-0255-7.

14. Blei D. M., Lafferty J. D. Dynamic topic models. *ICML '06: Proc. 23rd Int. Conf. on Machine Learning*. New York, Association for Computing Machinery; 2006:113–120. DOI: 10.1145/1143844.1143859.

15. Beltagy I., Lo K., Cohan A. SciBERT: A pretrained language model for scientific text. *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. on Natural Language Processing (EMNLP–IJCNLP)*. Kerrville, TX, Association for Computational Linguistics; 2019:3615–3620. DOI: 10.18653/v1/D19-1371.

16. Hogan A., Blomqvist E., Cochez M., d'Amato C., Melo G. D., Gutierrez C., Kirrane S., Gajo J. E. L., Navigli R., Neumaier S., Ngomo A. C. N. Knowledge graphs. *ACM Computing Surveys (CSUR)*. 2021;54(4):1–37. DOI: 10.1145/3447772.

17. Auer S., Oelen A., Haris M., Stocker M., D'Souza J., Farfar K. E., Vogt L., Prinz M., Wiens V., Jaradeh M. Y., Improving access to scientific literature with knowledge graphs. *Bibliothek Forschung und Praxis*. 2020;44(3):516–529. DOI: 10.1515/bfp-2020-2042.

18. Himmelstein D. S., Lizee A., Hessler C., Brueggeman L., Chen S. L., Hadley D., Green A., Khankhanian P., Baranzini S. E. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*. 2020;6(e26726). DOI: 10.7554/eLife.26726.

19. Гурьянов В. И., Елизаров А. М. Программный модуль формирования цифрового математического пространства на основе графов знаний. *Электронные библиотеки*. 2025; 28(3):622–639. EDN: PEVYAR. DOI: 10.26907/1562-5419-2025-28-3-622-639.

[Gurianov V. I., Elizarov A. M. Software module for forming digital mathematical space based on knowledge graphs. *Russian Digital Libraries Journal*. 2025;28(3):622–639. (In Russian.) EDN: PEVYAR. DOI: 10.26907/1562-5419-2025-28-3-622-639.

20. Shamsabadi S., D'Souza J., Auer S. Large language models for scientific information extraction: An empirical study for virology. *Findings of the Association for Computational Linguistics: EACL 2024*. Kerrville, TX, Association for Computational Linguistics; 2024:374–392.

21. Taylor R., Kardas M., Cucurull G., Scialom T., Hartshorn A., Saravia E., Poulton A., Kerkez V., Stojnic R. Galactica: A large language model for science. *arXiv preprint*. 2022:2211.09085.

22. Беспалько В. П. Программированное обучение. Дидактические основы. М.: Высшая школа; 1970. 296 с. EDN: ZHYULR.

[Bespalko V. P. Programmed learning. Didactic foundations. Moscow, Vysshaya shkola; 1970. 296 p. (In Russian.) EDN: ZHYULR.]

23. Anderson M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. 2001;26(1):32–46. DOI: 10.1111/j.1442-9993.2001.01070.pp.x.

24. Rousseeuw P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987;20:53–65. DOI: 10.1016/0377-0427(87)90125-7.

Информация об авторах

Мариносян Андреас Хачатурович, аспирант, департамент информатизации образования, Институт цифрового образования, Московский городской педагогический университет, г. Москва, Россия; *ORCID*: <https://orcid.org/0000-0003-0577-2360>; *e-mail*: a.marinosyan@yandex.ru

Григорьев Сергей Георгиевич, чл.-корр. РАО, доктор тех. наук, профессор, профессор департамента информатики, управления и технологий, Институт цифрового образования, Московский городской педагогический университет, г. Москва, Россия; *ORCID*: <https://orcid.org/0000-0002-0034-9224>; *e-mail*: grigorsg@yandex.ru

Лернер Илья Михайлович, доктор тех. наук, профессор кафедры нанотехнологий в электронике, Институт радиоэлектроники, фотоники и цифровых технологий, Казанский национальный исследовательский технический университет им. А. Н. Туполева-КАИ, г. Казань, Республика Татарстан, Россия; профессор кафедры радиофизики, профессор кафедры радиоастрономии, Институт физики, Казанский (Приволжский) федеральный университет, г. Казань, Республика Татарстан, Россия; *ORCID*: <https://orcid.org/0000-0001-5788-3113>; *e-mail*: aviap@mail.ru

Аникьева Марина Анатольевна, канд. тех. наук, доцент кафедры систем искусственного интеллекта, Институт космических и информационных технологий, Сибирский федеральный университет, г. Красноярск, Россия; *ORCID*: <https://orcid.org/0000-0001-8155-5902>; *e-mail*: MAnikieva@sfukras.ru

Information about the authors

Andreas Kh. Marinosyan, a postgraduate student at the Department of Education Informatisation, Institute of Digital Education, Moscow City University, Moscow, Russia; *ORCID*: <https://orcid.org/0000-0003-0577-2360>; *e-mail*: a.marinosyan@yandex.ru

Sergey G. Grigoriev, Corresponding Member of the RAE, Doctor of Sciences (Engineering), Professor, Professor at the Department of IT, Management and Technology, Institute of Digital Education, Moscow City University, Moscow, Russia; *ORCID*: <https://orcid.org/0000-0002-0034-9224>; *e-mail*: grigorsg@yandex.ru

Ilya M. Lerner, Doctor of Sciences (Engineering), Professor at the Department of Nanotechnology in Electronics, Institute of Radio Electronics, Photonics and Digital Technologies, Kazan National Research Technical University named after A. N. Tupolev-KAI, Kazan, The Republic of Tatarstan, Russia; Professor at the Department of Radio Physics, Professor at the Department of Radio Astronomy, Institute of Physics, Kazan (Volga Region) Federal University, Kazan, The Republic of Tatarstan, Russia; *ORCID*: <https://orcid.org/0000-0001-5788-3113>; *e-mail*: aviap@mail.ru

Marina A. Anikieva, Candidate of Sciences (Engineering), Associate Professor at the Department of Artificial Intelligence Systems, School of Space and Information Technology, Siberian Federal University, Krasnoyarsk, Russia; *ORCID*: <https://orcid.org/0000-0001-8155-5902>; *e-mail*: MAnikieva@sfukras.ru

Поступила в редакцию / Received: 07.09.25.

Поступила после рецензирования / Revised: 20.10.25.

Принята к печати / Accepted: 21.10.25.