

DOI: 10.32517/0234-0453-2025-40-2-66-78



К ВОПРОСУ ПОДБОРА УЧЕБНО-МЕТОДИЧЕСКОЙ ИНФОРМАЦИИ ДЛЯ РЕАЛИЗАЦИИ АДАПТИВНОЙ ЭЛЕКТРОННОЙ ОБРАЗОВАТЕЛЬНОЙ СРЕДЫ: АЛГОРИТМ АПРИОРНОЙ КЛАССИФИКАЦИИ АВТОРОВ

С. Г. Григорьев¹, И. М. Лернер^{2,3} ✉, А. Х. Мариносян¹, Н. К. Арутюнова³, М. А. Григорьева¹

¹ *Московский городской педагогический университет, г. Москва, Россия*

² *Башкирский государственный педагогический университет им. М. Акмуллы, г. Уфа, Республика Башкортостан, Россия*

³ *Казанский национальный исследовательский технический университет им. А. Н. Туполева-КАИ, г. Казань, Республика Татарстан, Россия*

✉ aviar@mail.ru

Аннотация

Статья посвящена формированию нового подхода к классификации отечественной учебно-методической информации, которая должна обеспечить реализуемость адаптивности электронных образовательных систем в рамках перехода общества к его информационной форме. Необходимость этого обусловлена тем, что учебно-методическая информация обычно формируется на базе новых научных результатов, полученных при выполнении диссертационных исследований. При этом объем научной информации в настоящее время растет по закону, близкому к экспоненциальному, и тем самым резко увеличиваются требования к интегрированным распределенным информационным системам (электронным библиотечным системам), обеспечивающим образовательный процесс.

Обоснование актуальности темы исследования и постановка решаемой задачи произведены на основании ретроспективного анализа существующих методов классификации документов, содержащих научную информацию, а также интегрированных распределенных информационных систем, широко представленных на мировом рынке. Результаты анализа указывают на то, что существующий рост публикуемой учебно-методической информации происходит по закону, близкому к экспоненциальному, что приводит к уменьшению глубины запроса или к существенному сужению области поиска и получения с высокой вероятностью информации, не являющейся релевантной, что в итоге снижает качество образовательного процесса.

Результатом данной работы является алгоритм априорной классификации авторов учебно-методической информации, базирующийся на основе анализа авторефератов, свойственных отечественной системе аттестации научно-педагогических кадров. Достижение результата получено посредством использования идей академической генеалогии, теории множеств и дискретной математики, теории графов.

Перечислены ограничения и рекомендации по практической реализации предложенного в работе подхода.

Ключевые слова: классификация информации, априорный алгоритм классификации авторов, академическая генеалогия, ретроспективный анализ, графовые БД.

Для цитирования:

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. DOI: 10.32517/0234-0453-2025-40-2-66-78.

ON THE ISSUE OF EDUCATIONAL AND METHODOLOGICAL INFORMATION SELECTION FOR IMPLEMENTING AN ADAPTIVE LEARNING MANAGEMENT SYSTEM: ALGORITHM OF A PRIORI AUTHORS CLASSIFICATION

S. G. Grigoriev¹, I. M. Lerner^{2,3} ✉, A. Kh. Marinosyan¹, N. K. Arutyunova³, M. A. Grigorieva¹

¹ *Moscow City Pedagogical University, Moscow, Russia*

² *Akmulla Bashkir State Pedagogical University, Ufa, The Republic of Bashkortostan, Russia*

³ *Kazan National Research Technical University named after A. N. Tupolev-KAI, Kazan, The Republic of Tatarstan, Russia*

✉ aviar@mail.ru

© Григорьев С. Г., Лернер И. М., Мариносян А. Х., Арутюнова Н. К., Григорьева М. А., 2025

Abstract

The article is devoted to the formation of a new approach to the classification of national educational and methodological information, which should ensure the feasibility of adaptability of electronic educational systems, within the framework of the transition of society to its information form. The need for this is due to the fact that educational and methodological information is usually formed on the basis of new scientific results obtained during the implementation of dissertation research. At the same time, the volume of scientific information is currently growing according to a law close to exponential, and thus the requirements for integrated distributed information systems (electronic library systems) that provide the educational process are sharply increasing. The justification for the relevance of the research topic and the formulation of the problem to be solved are based on a retrospective analysis of existing methods for classifying documents containing scientific information, as well as integrated distributed information systems, widely represented on the world market. The results of the analysis indicate that the existing growth of published scientific and methodological information occurs according to a law close to exponential, which leads to a decrease in the depth of the request or to a significant narrowing of the search area and obtaining information that is not relevant with a high probability, which ultimately reduces the quality of the educational process.

The result of this paper is an algorithm for the a priori classification of authors of educational and methodological information, based on the analysis of abstracts inherent in the domestic system of certification of scientific and pedagogical personnel. The result was achieved through the use of ideas from academic genealogy, set theory and discrete mathematics, graph theory.

The limitations and recommendations for the practical implementation of the approach proposed in the work are listed.

Keywords: information classification, a priori classification algorithm, academic genealogy, retrospective analysis, graph databases.

For citation:

Grigoriev S. G., Lerner I. M., Marinosyan A. Kh., Arutyunova N. K., Grigorieva M. A. On the issue of educational and methodological information selection for implementing an adaptive learning management system: Algorithm of a priori authors classification. *Informatics and Education*. 2025;40(2):66–78. (In Russian.) DOI: 10.32517/0234-0453-2025-40-2-66-78.

1. Введение

Одной из главных составляющих протекающего в настоящее время процесса, связанного с переходом к информационному обществу, является создание адаптивных электронных образовательных сред (АЭОС) [1–3]. Главной их чертой должно быть формирование индивидуальных траекторий обучения с учетом эмоционально-психологических особенностей каждого из обучающихся [там же]. Однако существующие на данный момент алгоритмы [1], которые лежат в основе построения индивидуальных образовательных траекторий, требуют достаточно больших вычислительных ресурсов при подборе образовательных объектов¹ и для удовлетворения крайне высоких нагрузок на интегрированных распределенных информационных системах (ИРИС) [4]. Это приводит к снижению возможностей по индивидуализации образовательного процесса или в ряде случаев к невозможности практической реализации АЭОС.

Это связано с тем, что образовательные объекты в большинстве случаев являются элементами научных публикаций и возрастание их числа в настоящее время происходит по закону, близкому к экспоненциальному. В качестве примера на рисунке 1 приведена статистика по числу научных документов, находящихся в отечественной базе данных eLIBRARY², с 2020 по 2024 год. При этом наиболее часто применяемый в ИРИС направленный поиск с использованием метаданных [5], хотя и должен уменьшить затраты на накладные расходы по передаче служебной информации внутри ИРИС при поиске информации, однако, из-за необходимости предварительного индексирования данных в системе требует весьма существенных объемов передаваемой информации внутри ее подсистем, особенно при возрастающей интенсивности запросов на индексацию,

обусловленных добавлением новых документов. Это в сочетании с ограниченной пропускной способностью межсистемных интерфейсов ввода-вывода ИРИС, обусловленных требованием обеспечения целостности сигнала [6, 7], приводит к невозможности получения релевантной информации при требуемой скорости считывания метаданных из базы данных, поскольку требуемая скорость считывания превосходит скорость обновления информации в базе данных [8].

Корректность вышеизложенного также подтверждается детальным анализом работ [4, 9–19], в которых отмечаются следующие факты:

- низкая релевантность информации, предоставляемой пользователю, при использовании широко распространенных стратегий поиска информации, приводящих к широкому охвату запроса на поиск информации в ИРИС. В частности, это достаточно отчетливо проявляется при использовании системы Online Public Access Catalogue (OPAC), выполняющей направленный поиск с использованием метаданных;
- необходимость использования довольно узкого поискового запроса и невозможность использования стратегий усечения запроса, что свойственно работе с поисковыми машинами (Google, Яндекс и т. д.);
- отсутствие взаимосвязей между ключевыми словами и предметными рубриками информационно-поисковых языков;
- принудительное ограничение возможностей библиотечной системы при ее развертывании;
- отсутствие учета объективной картины динамично изменяющихся потребностей пользователей, например, в поиске информации по редким книгам и рукописям;
- только 51,78 % пользователей отметили простоту в использовании автоматизированных информационных систем, и только 58,62 % пользователей удовлетворены скоростью поиска в этих системах.

¹ Образовательные объекты — это элементы текста, видеоматериалы, аудиозаписи и рисунки [1, 2].

² eLIBRARY.RU — научная электронная библиотека. https://www.elibrary.ru/stat_time_items.asp

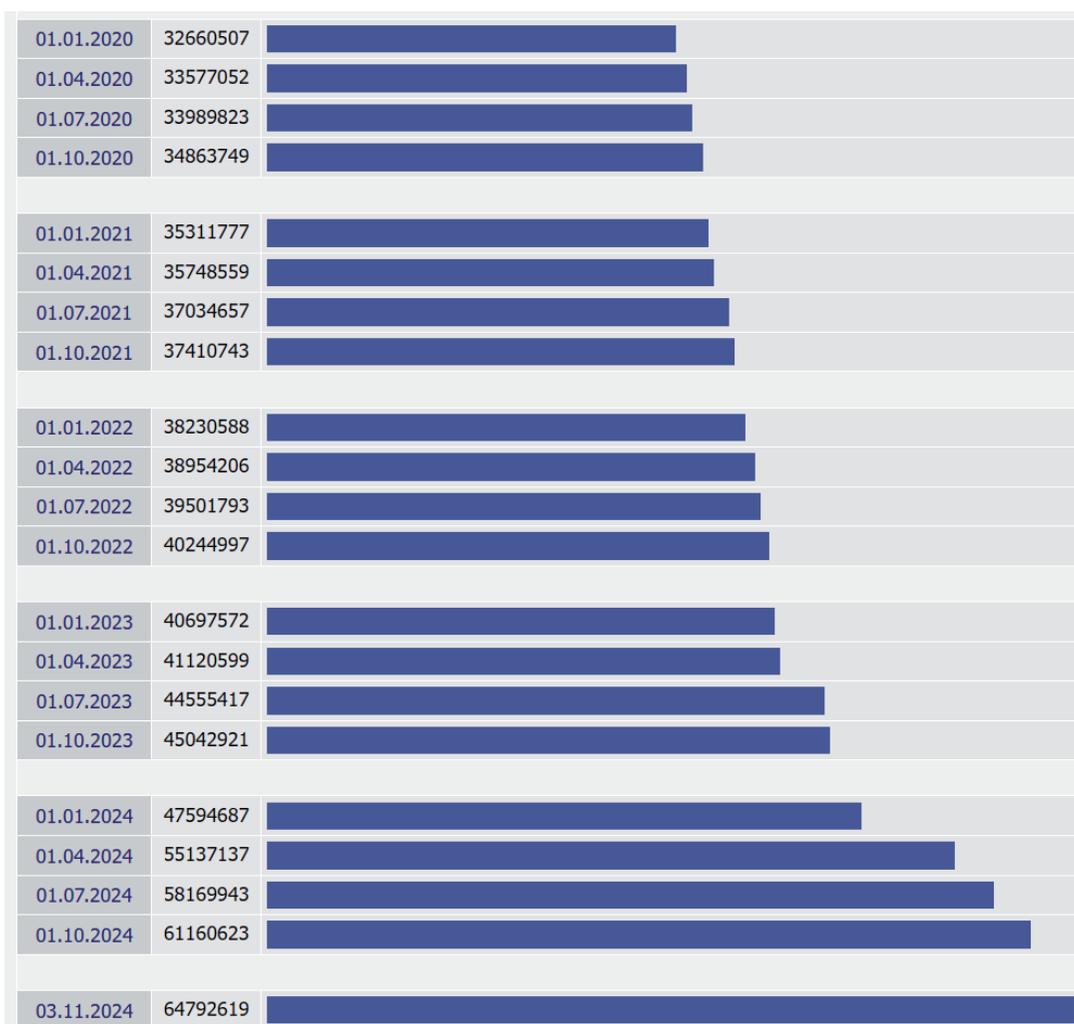


Рис. 1. Статистика по числу научных документов, находящихся в БД eLIBRARY, с 2020 по 2024 год

Fig. 1. Statistics on the number of scientific documents in the eLIBRARY database from 2020 to 2024

Попытки преодоления сложившейся ситуации можно обнаружить в работах [20–28]. Так, в работе [20] на основании использования многомерного векторного пространства и методов кластерного анализа представлен новый подход для автоматической индексации анализируемых документов, который основан на вычислении плотности распределения метрических оценок документов в пространстве. Там же представлено правило для оптимального выбора словаря индексации для коллекции документов, которое, в том числе, базируется на вычислении статистической меры TF-IDF¹. Однако детальный анализ работы [20] указывает на основные ограничения указанного подхода, которые свойственны и современным экспертным рекомендательным системам, ядром которых является данный подход:

- отсутствие алгоритма априорной классификации анализируемых документов на релевантные и нерелевантные по отношению к каждому из запросов пользователей. Это обусловлено тем, что оптимальную конфигурацию формируемой БД трудно создать при отсутствии априорного знания полной истории поиска для данной коллекции документов;
- высокая вычислительная сложность алгоритмов подсчета метрик для каждого документа коллекции, определяющих близость анализируемых документов, имеющих квадратичное время исполнения от числа документов в анализируемой коллекции;
- низкая производительность алгоритмов кластеризации для слабо определенных кластеров при малых расстояниях между их центроидами².

¹ В работе [28] указывается, что применение TF-IDF при анализе научной информации является достаточно хорошим средством при анализе отдельных документов, но не позволяет адекватно классифицировать информацию по книгам, поскольку названия книг обычно имеют малый размер.

² Это, в частности, объясняет низкую эффективность от использования метрики косинусного расстояния в задачах классификации большими языковыми моделями при обработке текстов с близким содержанием.

Практическая реализация некоторых подходов, связанных с анализом предпочтений пользователей, представлены в работах [21–28].

Так, в работе [21] рассмотрена экспертная система, которая осуществляет анализ взаимосвязи между wybranными пользователем книгами и другими литературными источниками и вырабатывает соответствующие рекомендации. Сам алгоритм функционирования системы состоит из двух крупных этапов. Сначала формируются наборы наиболее часто выбираемых пользователем книг по разным направлениям, а затем осуществляется выявление сильных ассоциативных правил и поиск книг, которые интересны читателю. К недостаткам данной системы можно отнести то, что процесс формирования ассоциативных правил зависит от размера базы данных и занимает достаточно большое время при ее значительных размерах.

В системах, описанных в [22, 23], получаемый пользователем результат формируется на основе алгоритма совместной фильтрации, совершаемой путем анализа взаимосвязей пользовательских предпочтений. Работа систем данной категории основывается на предположении о том, что пользователь воспользуется той же книгой, что и другие пользователи, при условии совпадения их интересов. К наиболее существенным недостаткам такого подхода относится зависимость его точности от того, с какой погрешностью определяется близость между пользователями. При этом с ростом БД и числа пользователей будут наблюдаться затруднения при работе самой системы. Немаловажной является проблема так называемого «холодного старта», когда для получения корректных результатов требуется предварительное наполнение БД информацией о взаимосвязях, что не всегда является возможным.

Преодоление проблемы «холодного старта» можно обнаружить в работе [24] — путем формирования и использования усредненного мнения за счет анализа социальной среды, в которой находится пользователь. Однако тогда данная система сталкивается с теми же проблемами производительности, которые свойственны системам, описанным в [22, 23].

Альтернативой к указанным в работах [20–23] подходам является создание специализированных систем анализа информации с вертикальным подходом поиска необходимой информации. Так, в работе [25] предложен новый алгоритм обработки текстовой информации, реализующий вертикальный поиск посредством использования онтологии с последующей тренировкой подсистем поисковой машины. Несмотря на перспективность такого подхода, он обладает низкой производительностью при увеличении размеров БД, в которой производится поиск, поскольку не решается проблема априорного оптимального размещения данных в ИРИС.

Интересное решение представлено в работе [26], в которой предлагается подход для повышения точности формирования запроса в ОПАС посредством

использования дополнительной системы уточнения запросов, реализованной в форме чат-бота, которая после обработки ответов пользователя позволяет сформировать более детализированный запрос для системы. Однако такой подход также не решает проблему оптимального хранения данных в системе.

Создание системы рекомендаций литературы отдельному пользователю на основе анализа рейтингов других пользователей относительно конкретного источника представлено в работе [27]. Главным недостатком такой системы является то, что для достижения поставленной цели должны быть промаркированы все документы с использованием оценок, что при большом библиотечном фонде является крайне трудоемкой задачей.

Одним из последних решений в области проектирования экспертных систем является система, рассмотренная в работе [28]. Она обеспечивает многоуровневую систему рекомендаций Upper and Lower Level Relations (ULLR), совмещенную с методом Hotspot Analysis Method (HAM), который использует Chinese Library Classification Method (CLCM) для построения иерархической структуры книг с учетом ссылок на другие книги. Недостатком данной системы является отсутствие решения проблемы «холодного старта».

На основе вышеизложенного становится очевидной актуальность поиска нового подхода к решению проблемы рационального хранения и обработки научной и учебно-методической информации, обладающей низкими значениями вычислительной сложности при обработке пользовательских запросов и расходами на передачу сервисной информации.

Для рекомендательных систем, к которым относятся адаптивные электронные образовательные системы, также актуальна задача учета индивидуальных особенностей обучающихся [29].

Поэтому целью планируемой серии работ является разработка нового единого подхода к хранению и обработке информации с учетом эмоционально-психологических особенностей ее пользователей, а задачей, решение которой будет описано в данной статье, — создание алгоритма классификации авторов для решения вопроса подбора учебно-методической информации для адаптивных электронных образовательных сред.

2. Алгоритм априорной классификации авторов учебно-методической информации для реализации адаптивных электронных образовательных сред

Непосредственно структурная схема алгоритма априорной классификации научной и образовательной информации представлена на рисунке 2.

Рассмотрим работу предлагаемого алгоритма, полагая, что БД не создана. В этом случае производится формирование БД научных школ в соответствии с алгоритмом, представленным в разделе 2.1. После

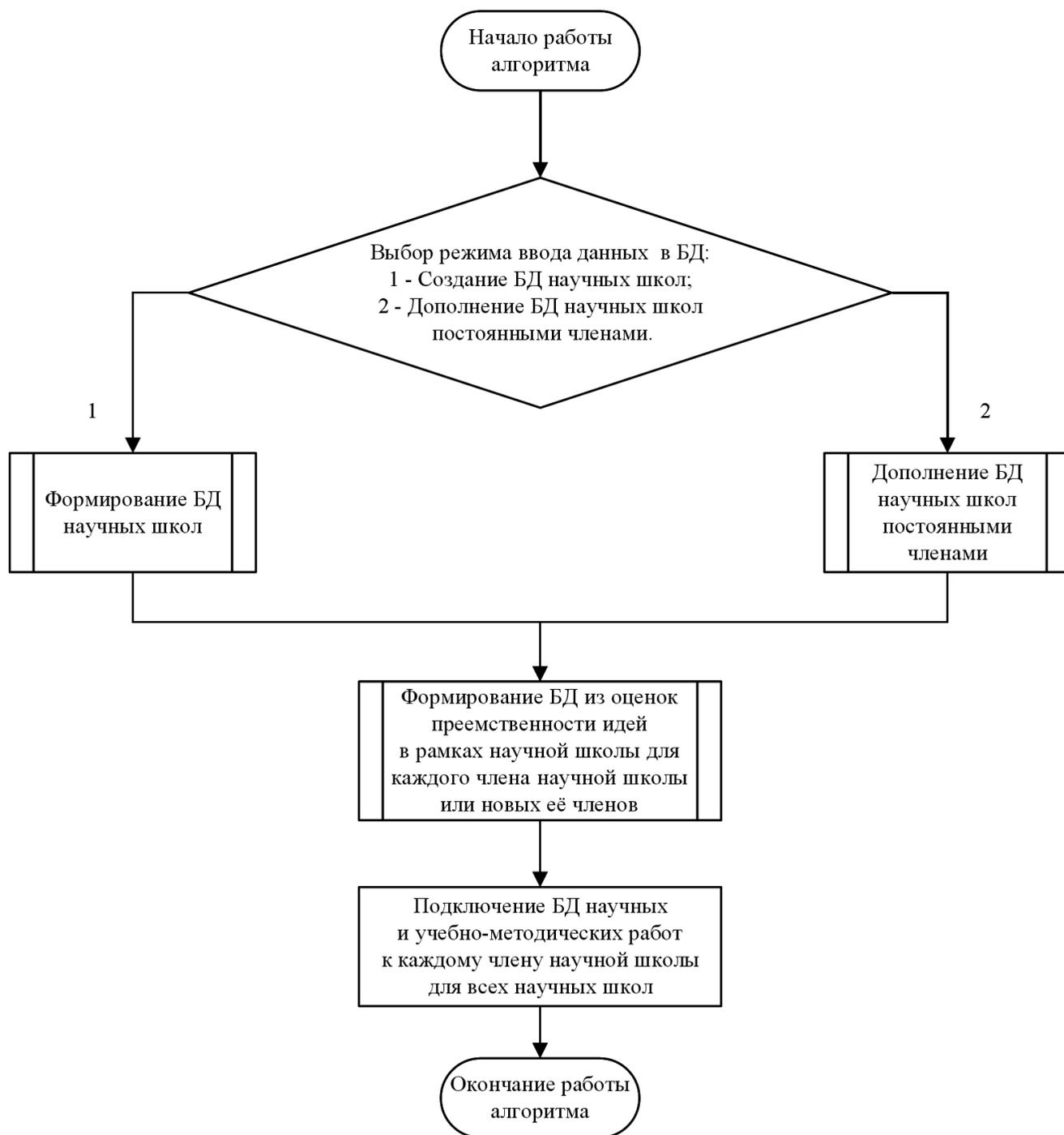


Рис. 2. Структурная схема алгоритма априорной классификации авторов учебно-методической информации

Fig. 2. Structural diagram of an algorithm for a priori classification of scientific and educational information

этого производится создание дополнительной БД для каждой научной школы, отражающей преемственность идей для каждого постоянного члена школы в соответствии с нормированными показателями, представленными в разделе 2.4. Далее производится подключение БД научных работ с помощью ссылок от каждого члена научной школы.

В том случае, если производится дополнение БД научных школ новыми ее членами, процесс аналогичен ранее изложенному, за исключением самого

процесса дополнения, описание которого представлено в разделе 2.2. Ограничения, возникающие при использовании разработанного алгоритма, указаны в разделе 2.3.

2.1. Алгоритм формирования базы данных постоянных научных школ

Алгоритм формирования базы данных для построения деревьев академической генеалогии для каждой научной школы состоит из трех этапов.

2.1.1. Этап 1. Подготовительный

Результатом его выполнения является формирование первичной базы данных в виде единой таблицы в формате CSV¹, содержащей следующие данные:

- По соискателю ученой степени:
 - фамилия, имя и отчество (ФИО) соискателя;
 - шифр и наименование научной специальности²;
 - ученая степень и отрасль знаний, по которой производилась защита диссертационной работы;
 - тема диссертационной работы;
 - организация³, в которой была выполнена диссертационная работа.
- По научному руководителю и/или консультанту⁴:
 - ФИО руководителя / консультанта;
 - его ученая степень и звание.
- По месту и времени, где проводилась защита диссертационной работы:
 - наименование организации, на базе которой создан диссертационный совет, в котором проводилась защита диссертации;
 - город, где она находится;
 - год защиты.
- По официальным оппонентам и ведущей организации:
 - ФИО оппонентов, их ученые степени и звания;
 - полное наименование ведущей организации, город, в котором она расположена.

Каждая i -я строка указанной таблицы содержит перечисленные выше данные для i -го автореферата, расположенные в различных полях (столбцах таблицы).

В качестве исходных данных при формировании первичной БД используются pdf-файлы авторефератов. Обработка pdf-файлов производится посредством авторского программного обеспечения⁵, реализованного в среде Python, который также обеспечивает орфографическую проверку и коррекцию посредством использования расстояния Дамерау—Левенштейна с использованием подхода, представленного в работе [30].

¹ Разделителем полей является «;» (точка с запятой).

² В случае, если защита производится по двум специальностям, на стыке двух областей знаний, указывается каждая из них.

³ При наличии нескольких организаций, на базе которых осуществлялась подготовка диссертационной работы, в базе данных перечисляются все из них.

⁴ При наличии у соискателя двух научных руководителей и/или консультантов в первичной БД отражаются данные о каждом из них.

⁵ Свидетельство о государственной регистрации программы для ЭВМ № 2024684406 Российская Федерация. Программа извлечения метаданных из авторефератов диссертаций: № 2024682059: заявлено 25.09.2024; опубликовано 17.10.2024 / Григорьев С. Г., Лернер И. М., Мариносян А. Х., Жуковская В. А., Аникьева М. А.; правообладатель ФГБОУ ВО «Башкирский государственный педагогический университет им. М. Акмуллы». EDN: GPZDFX.

Разделение обработанных данных по полям таблицы осуществляется посредством использования регулярных выражений, а также ключевых слов, таких как «автореферат», «наук», «профессор», «доцент», «кандидат» и «доктор», и словосочетаний, которые их содержат, с учетом знаков пунктуации.

2.1.2. Этап 2. Проверочный

Формируется множество индексов первичной базы данных $I_{\text{пт}}$, для которых «Тема диссертационной работы» является пустой, в соответствии со следующим выражением:

$$I_{\text{пт}} = \left\{ i_{\text{пт}_m} = iG(i) \mid m = \sum_{k=1}^i G(k) + (1 - \text{sgn}(G(i))), i = \overline{1, M} \right\}. \quad (1)$$

Здесь M — число записей в первичной БД, $\text{sgn}(\cdot)$ — функция знака;

$$G(k) = \begin{cases} 1, & \text{если } Z(k, T) = \text{NULL}; \\ 0, & \text{если } Z(k, T) \neq \text{NULL}, \end{cases} \quad (2)$$

где:

- $Z(k, T)$ — значение поля «Тема диссертационной работы» для k -го автореферата (k -й записи в первичной БД);
- T — порядковый номер поля «Тема диссертационной работы» в первичной БД.

Сама запись $Z(k, T)$, если она не является пустой, представляет собой совокупность слов, знаков пунктуации и разделяющих их пробелов (за исключением «;»).

После этого производится оценка корректности заполнения полей таблицы для каждой i -й записи первичной БД ($i \in J$), где $J = \{j \mid j = \overline{1, M}\} \setminus I_{\text{пт}}$ по следующим группам:

- *первая группа* включает в себя следующие поля:
 - ученую степень, на которую претендовал соискатель в рамках защиты диссертации;
 - шифр или шифры научных специальностей, по которым представлена диссертационная работа;
- *вторая группа* включает поля:
 - ФИО соискателя;
 - ФИО его научных руководителей/консультантов, официальных оппонентов;
- *третья группа* состоит из одного поля:
 - название города, где находится диссертационный совет, в котором состоялась защита диссертационной работы соискателем.

Анализ корректности значений для первой группы в первичной БД производится путем проверки наличия данных, не соответствующих полю «Тема диссертационной работы» для каждого i -го автореферата в следующем порядке:

1. Проверяется наличие шифров научных специальностей. Это происходит в соответствии со следующим правилом.

Вычисляется величина:

$$C_1(i) = \begin{cases} 0, & \text{если } Z_{\text{ш.авт}_i} = \emptyset, \\ 1, & \text{если } Z_{\text{ш.авт}_i} \neq \emptyset, \end{cases} \quad (3)$$

где:

- $Z_{\text{ш.авт}_i} = Z_{\text{ш}} \cap Z(i, T) = \{z_{i,k}\}_{k=1}^{k=N_i}$;
- $Z_{\text{ш}}$ — множество шифров научных специальностей, применяемых в Российской Федерации и СССР;
- N_i — число шифров научных специальностей, обнаруженных в $Z(k, T)$.

В случае если $C_1(i) = 0$, коррекция БД не производится.

Если же $C_1(i) = 1$, то осуществляется корректировка i -й записи первичной БД, что можно представить следующей последовательностью действий:

а) Формируется промежуточное значение в виде упорядоченного множества $Z_{\text{пр.3.1}} = Z(i, T) \setminus Z_{\text{ш.авт}_i}$, которое не содержит шифров научных специальностей.

б) Осуществляется переопределение исходного значения $Z(i, T) = Z_{\text{пр.3.1}}$.

в) Для i -го автореферата определяется наименьший порядковый номер пустого поля $D_A(i)$ среди полей, выделенных для хранения шифров специальностей в первичной БД, что можно представить следующим условием:

$$D_A(i) \in \overline{1, n_{\text{спец}}}: (Z(i, S_1 + D_A(i)) = NULL) \wedge (\forall D < D_A(i): Z(i, S_1 + D) \neq NULL), \quad (4)$$

где:

- $n_{\text{спец}}$ — число полей в БД, зарезервированных для шифров научных специальностей;
- S_1 — сдвиг по числу полей относительно начала индексации полей БД, определяющий расположение начального поля для шифра научной специальности, на которую претендовал соискатель.

г) Производится окончательное заполнение пустых полей шифров научных специальностей для i -го автореферата в соответствии со следующим правилом:

$$Z(i, D_A(i) + S_1 + k - 1) = Z_{\text{ш.авт}_i}(k) \mid k = \overline{1, N_i}. \quad (5)$$

2. Проверяется наличие записи об ученой степени, на которую претендовал соискатель, с использованием следующей величины:

$$C_2(i) = \begin{cases} 0, & \text{если } Z_{\text{уч.ст}_i} = \emptyset, \\ 1, & \text{если } Z_{\text{уч.ст}_i} \neq \emptyset. \end{cases} \quad (6)$$

Здесь:

- $Z_{\text{уч.ст}_i} = f_{\text{пс}}(Z_{\text{уч.ст}}, Z(i, T))$ — ученой степени, на которую претендует соискатель;
- $f_{\text{пс}}(\cdot)$ — функция поиска подстроки в строке, построенная на основе алгоритмов поиска подстроки в строке [31];
- $Z_{\text{уч.ст}}$ — множество, составленное из наименований ученых степеней, согласно классификации, принятой в России и СССР.

В случае, если $C_2(i) = 0$, коррекция БД не производится.

Если же $C_2(i) = 1$, то осуществляется корректировка i -й записи первичной БД, что можно представить следующей последовательностью действий:

а) Формируется промежуточное значение в форме упорядоченного множества $Z_{\text{пр.3.2}} = Z(i, T) \setminus Z_{\text{уч.ст}_i}$.

б) Осуществляется переопределение исходного значения $Z(i, T) = Z_{\text{пр.3.2}}$.

в) Производится заполнение в i -й записи пустого поля с порядковым номером D_c в первичной БД, в котором указывается ученая степень, на которую претендовал соискатель

$$Z(i, D_c) = Z_{\text{уч.ст}_i}. \quad (7)$$

Анализ корректности записей по значению второй группы полей в первичной БД производится путем их проверки: являются ли они пустыми или содержат цифры. Необходимо отметить, что если запись является некорректной, то коррекция первичной БД не производится. Результатом процедуры проверки корректности заполнения i -й записи является значение индикаторной функции $F_2(i)$, которое указывает на наличие ($F_2(i) = 1$) или отсутствие ($F_2(i) = 0$) ошибки для указанной группы полей. Сама функция определяется следующим образом:

$$F_2(i) = \text{sgn} \left(\sum_{l=1}^{N_{\text{ФИО}}} \delta_{i,l} \right), \quad (8)$$

где:

- $N_{\text{ФИО}} = 6$ — количество возможных ФИО¹ во второй группе — соответствует наибольшему количеству человек, исключая членов диссертационного совета и представителя ведущей организации, которые могут принимать участие в защите диссертации (см. табл. 1);
- l — порядковый номер человека (ФИО).

Величины $\delta_{i,l}$ определяются в соответствии со следующей формулой:

$$\delta_{i,l} = \begin{cases} h_l, & \text{если } (l \in \{1; 2; 4; 5; 6\}) \wedge \\ & \wedge (\exists k_{i,l} \in \{3(l-1)+1; 3(l-1)+2; 3(l-1)+3\}: \\ & (Z(i, S_2 + k_{i,l}) = NULL) \vee (Z(i, S_2 + k_{i,l}) = 0)); \\ 0, & \text{если } (l \in \{1; 2; 4; 5; 6\}) \wedge \\ & \wedge (\forall k_{i,l} \in \{3(l-1)+1; 3(l-1)+2; 3(l-1)+3\}: \\ & (Z(i, S_2 + k_{i,l}) \neq NULL) \wedge (Z(i, S_2 + k_{i,l}) \neq 0)); \\ 1, & \text{если } (l=3) \wedge (Z(i, S_2 + 4) \neq NULL) \wedge \\ & \wedge ((Z(i, S_2 + 5) = NULL) \wedge (Z(i, S_2 + 6) = NULL) \vee \\ & \vee (Z(i, S_2 + 5) = 0) \wedge (Z(i, S_2 + 6) = 0)); \\ 0, & \text{если } (l=3) \wedge (Z(i, S_2 + 4) \neq NULL) \wedge \\ & \wedge (Z(i, S_2 + 5) \neq NULL) \wedge (Z(i, S_2 + 6) \neq NULL) \wedge \\ & \wedge (Z(i, S_2 + 4) \neq 0) \wedge (Z(i, S_2 + 5) \neq 0) \wedge \\ & \wedge (Z(i, S_2 + 6) \neq 0). \end{cases} \quad (9)$$

¹ Следует отметить, что для хранения каждого из слов — фамилии, имени и отчества — выделяется три отдельных поля, соответственно

Здесь:

- h_l — индикаторная функция, которая определяется согласно таблице 1;
- S_2 — сдвиг по числу полей относительно начала индексации БД, определяющий расположение начального поля, в котором указывается фамилия соискателя.

Таблица 1 / Table 1

Таблица значений для индикаторной функции h_j Table of values for the indicator function h_j

j	Принадлежность ФИО	h_j	
		Кандидатская диссертация	Докторская диссертация
1	Соискатель	1	1
2	Научный руководитель / консультант № 1	1	1
3	Научный консультант № 2	не используется в выражении (9)	
4	Официальный оппонент № 1	1	1
5	Официальный оппонент № 2	1	1
6	Официальный оппонент № 3	0	1

Анализ и коррекция записей третьей группы полей производится с помощью метода, представленного в работе [30]. Если ошибка исправлена или отсутствует в i -й записи по данному полю, то индикаторная функция $F_3(i) = 0$, иначе $F_3(i) = 1$.

3. В заключительной стадии проверочного этапа формирования алгоритма базы данных постоянных научных школ производится формирование БД (табл. 2), указывающей на наличие ошибок в первичной БД, после ее корректировки.

Таблица 2 / Table 2

Пример БД, указывающей на наличие ошибок в первичной БД после ее коррекции

Example of DB indicating the presence of errors in the primary DB after its correction

i	$\Gamma(i)$	$G(i)$	$F_2(i)$	$F_3(i)$
1	1	0	0	1
2	1	0	1	1
⋮	⋯	⋯	⋯	⋯
M	1	1	0	0

Необходимо отметить, что значение функции $\Gamma(i)$ определяется следующим образом:

$$\Gamma(i) = \begin{cases} 1, & \text{если } \text{sgn}(F_2(i) + F_3(i)) + G(i) > 0; \\ 0, & \text{если } \text{sgn}(F_2(i) + F_3(i)) + G(i) = 0. \end{cases} \quad (10)$$

Наличие ошибки для i -й записи соответствует $\Gamma(i) = 0$, а ее отсутствию — значение $\Gamma(i) = 1$.

Множество номеров записей первичной БД для дальнейшего анализа можно определить следующим образом:

$$D_{\text{БД}} = \left\{ d_m = i\Gamma(i) \mid m = \sum_{k=1}^i \Gamma(k) + (1 - \text{sgn}(\Gamma(i))), i = \overline{1, M} \right\} \setminus \{0\}. \quad (11)$$

2.1.3. Этап 3. Создание БД генеалогического дерева научных школ

Алгоритм формирования БД можно представить как последовательность следующих действий.

1. Производится формирование множества $D_{\text{ст}}$, элементы которого являются упорядоченными парами индексов $(d_k, d_d)_r$. Так, первый из индексов r -й пары $(d_{k,r})$ определяет номер записи в первичной БД после ее коррекции, в которой содержатся сведения об автореферате диссертации, после защиты которой соискателю была присуждена ученая степень кандидата наук, а второй $(d_{d,r})$ — номер записи об автореферате диссертации, после защиты которой тому же соискателю была присуждена ученая степень доктора наук

$$D_{\text{ст}} = \left\{ d_{\text{ст}} = (d_k, d_d)_r \mid (Z(d_{k,r}, 1 + S_2) \cdot s_p \cdot Z(d_{k,r}, 2 + S_2) \cdot s_p \cdot Z(d_{k,r}, 3 + S_2) = Z(d_{d,r}, 1 + S_2) \cdot s_p \cdot Z(d_{d,r}, 2 + S_2) \cdot s_p \cdot Z(d_{d,r}, 3 + S_2)) \wedge (D' \in Z_{\text{у.авт}_{d_{k,r}}} \wedge (D'' \in Z_{\text{у.авт}_{d_{d,r}}}) \wedge (d_{k,r} < d_{d,r}; d_{k,r}, d_{d,r} \in D_{\text{БД}}) \right\}, \quad (12)$$

где:

- D' — "кандидат";
- D'' — "доктор";
- s_p — символ пробела;
- "." — операция конкатенации строк.

Далее производится анализ множества $D'_{\text{ст}}$ на наличие упорядоченных пар с одинаковыми значениями их первых элементов, что указывает на наличие ученых с одинаковыми ФИО. Данная операция производится с формированием следующего множества:

$$D_{\text{пов}} = \left\{ (d_k, d_d)_r \in D'_{\text{ст}} : \left(\exists r_1 \mid (r_1 \neq r) \wedge ((d_{k,r} = d_{k,r_1}) \vee (d_{d,r} = d_{d,r_1})) \right) \right\}. \quad (13)$$

После чего осуществляется анализ множества $D_{\text{пов}}$ с целью определения корректных упорядоченных пар данного множества. Данную процедуру можно представить следующим образом:

$$D_{\text{кор}} = \left\{ (d_k, d_d)_r \in D_{\text{пов}} \mid \left[(k_1, k_2 \in [1, 2]) \wedge (Z(d_{k,r}, 3k_1 + 1 + S_2) \cdot s_p \cdot Z(d_{k,r}, 3k_1 + 2 + S_2) \cdot s_p \cdot Z(d_{k,r}, 3k_1 + 3 + S_2) = Z(d_{d,r}, 3k_2 + 1 + S_2) \right. \right.$$

$$\cdot s_p \cdot Z(d_{д,р}, 3k_2 + 2 + S_2) \cdot s_p \cdot Z(d_{д,р}, 3k_2 + 3 + S_2) \Big] \vee \\ \vee (Z(d_{к,р}, \Gamma) = Z(d_{д,р}, \Gamma)) \vee (Z(d_{к,р}, O) = Z(d_{д,р}, O)) \Big\}, \quad (14)$$

где Γ, O — номера полей, в которых указываются город и наименование организации, где выполнялась диссертационная работа.

Определяется множество упорядоченных пар, которые требуют ручной проверки:

$$D_{р.п} = D_{пов} \setminus D_{кор}. \quad (15)$$

После ручной проверки $D_{р.п}$, которая заключается в экспертном контроле и сопоставлении наименований организаций и названий городов, обусловленных изменением уставов организаций и переименованием городов, в которых выполнялась работа соискателем, и его корректировки производится формирование скорректированного множества $D_{р.п.корр}$, его объединение со множеством $D_{кор}$ и создание итогового множества, указывающего на взаимосвязи между записями, когда соискатель последовательно стал кандидатом и доктором наук, т. е. $D_{скд} = D_{кор} \cup D_{р.п.корр}$.

2. Для каждого i -го соискателя производится формирование множества P_i , чьи элементы — это упорядоченные пары $(k, p_i)_l$, где k — номер записи о соискании ученой степени i -м научно-педагогическим работником (НПР), а p_i — номер записи, где этот же НПР выступает в роли научного руководителя или консультанта № 1 и № 2:

- для $i \in (D_{БД} \setminus \text{pr}_1 D_{р.п.корр}) \setminus \text{pr}_2 D_{р.п.корр}$:

$$P_i = \left\{ (i, p_i)_l \mid p_{i,l} \in D_{БД}, \right. \\ \left[(Z(i, S_2 + 1) \cdot s_p \cdot Z(i, S_2 + 2) \cdot s_p \cdot Z(i, S_2 + 3) = \right. \\ = Z(p_{i,l}, S_2 + 4) \cdot s_p \cdot Z(p_{i,l}, S_2 + 5) \cdot s_p \cdot \\ \cdot Z(p_{i,l}, S_2 + 6)) \wedge (Z(i, D_c) = Z(p_{i,l}, D_{с.рук.1})) \Big] \vee \\ \vee \left[(Z(i, S_2 + 1) \cdot s_p \cdot Z(i, S_2 + 2) \cdot s_p \cdot Z(i, S_2 + 3) = \right. \\ = Z(p_{i,l}, S_2 + 7) \cdot s_p \cdot Z(p_{i,l}, S_2 + 8) \cdot \\ \cdot s_p \cdot Z(p_{i,l}, S_2 + 9)) \wedge (Z(i, D_c) = Z(p_{i,l}, D_{с.рук.2})) \Big] \Big\};$$

- для $i \in \text{pr}_1 D_{р.п.корр} \cup \text{pr}_2 D_{р.п.корр}$:

$$P_i = \left\{ (i, p_i)_l \mid p_{i,l} \in D_{БД}, \right. \\ \left[(Z(i, S_2 + 1) \cdot s_p \cdot Z(i, S_2 + 2) \cdot s_p \cdot Z(i, S_2 + 3) = \right. \\ = Z(p_{i,l}, S_2 + 4) \cdot s_p \cdot Z(p_{i,l}, S_2 + 5) \cdot s_p \cdot \\ \cdot Z(p_{i,l}, S_2 + 6)) \wedge \left[(Z(i, D_c) = Z(p_{i,l}, D_{с.рук.1})) \vee \right. \\ \vee \exists i_1 \in \text{pr}_2 D_{скд} \left(((i; i_1) \in D_{скд}) \wedge (Z(i_1, D_c) = \right. \\ = Z(p_{i,l}, D_{с.рук.1})) \Big) \Big] \vee \left[(Z(i, S_2 + 1) \cdot s_p \cdot \right. \\ \cdot Z(i, S_2 + 2) \cdot s_p \cdot Z(i, S_2 + 3) = Z(p_{i,l}, S_2 + 7) \cdot s_p \cdot \\ \cdot Z(p_{i,l}, S_2 + 8) \cdot s_p \cdot Z(p_{i,l}, S_2 + 9)) \wedge \\ \wedge \left[(Z(i, D_c) = Z(p_{i,l}, D_{с.рук.2})) \vee \right. \\ \vee \exists i_1 \in \text{pr}_2 D_{скд} \left(((i; i_1) \in D_{скд}) \wedge \\ \wedge (Z(i_1, D_c) = Z(p_{i,l}, D_{с.рук.2})) \Big) \Big] \Big] \Big\}, \quad (17)$$

где $D_{с.рук.1}$ и $D_{с.рук.2}$ — номера полей, в которых указываются ученые степени научного руководителя / консультанта № 1 и научного консультанта № 2 соответственно.

3. Для каждой h -й научной школы ($h = \overline{1, SC}$)¹ производится формирование ориентированного графа $G_h = (V_h, X_h)$, где V_h — множество, элементы которого определяют вершины графа G_h ; X_h — множество упорядоченных пар, определяющее дуги графа G_h . Непосредственно сам процесс формирования графов научных школ представлен в форме структурной схемы алгоритма, отраженной на рисунке 3.

На рисунке 3 используются следующие обозначения:

- G_{d_m} — граф, определяющий научную школу для d_m -го соискателя;
- X — множество упорядоченных пар, определяющее дуги графа G_{d_m} ;
- V — множество, элементы которого определяют вершины графа G_{d_m} ;
- функция $\Gamma(i)$ определяется с помощью выражения (10).

2.2. Расширение базы данных научных школ

Расширение базы данных научных школ, содержащей информацию о каждой из научных школ ($h = \overline{1, SC}$) в форме двух множеств V_h и X_h , осуществляется посредством формирования промежуточных множеств:

$$X'_h = X_h \cup P_\omega; \\ V'_h = V_h \cup \text{Pr}_1 P_\omega \cup \text{Pr}_2 P_\omega \quad (18)$$

при выполнении следующего условия: $\text{Pr}_1 P_\omega \in V_h$.

После чего происходит обновление множеств, определяющих граф для h -й научной школы:

$$X_h = X'_h; \\ V_h = V'_h.$$

Формирование множеств P_ω , где $\omega = 1, 2, \dots, N_{н.з}$, $N_{н.з}$ — число множеств, определяющих новые записи в БД с привязкой к конкретному научному руководителю/консультанту, производится посредством анализа новых записей об автореферате с помощью формул (16), (17).

2.3. Ограничения, возникающие при использовании разработанного алгоритма

К таким ограничениям относятся:

1. Отсутствие сведений о руководителе/консультанте, официальных оппонентах, организации, где выполнялась диссертация, ведущей организации авторефератов диссертации, а также шифров научных специальностей в авторефератах до 1960-х годов.

2. Диссертации, защищаемые в Национальном исследовательском университете «Высшая школа экономики» (НИУ ВШЭ), сопровождаются не авторефератами, а резюме, которые не передаются в РГБ,

¹ SC — число научных школ.

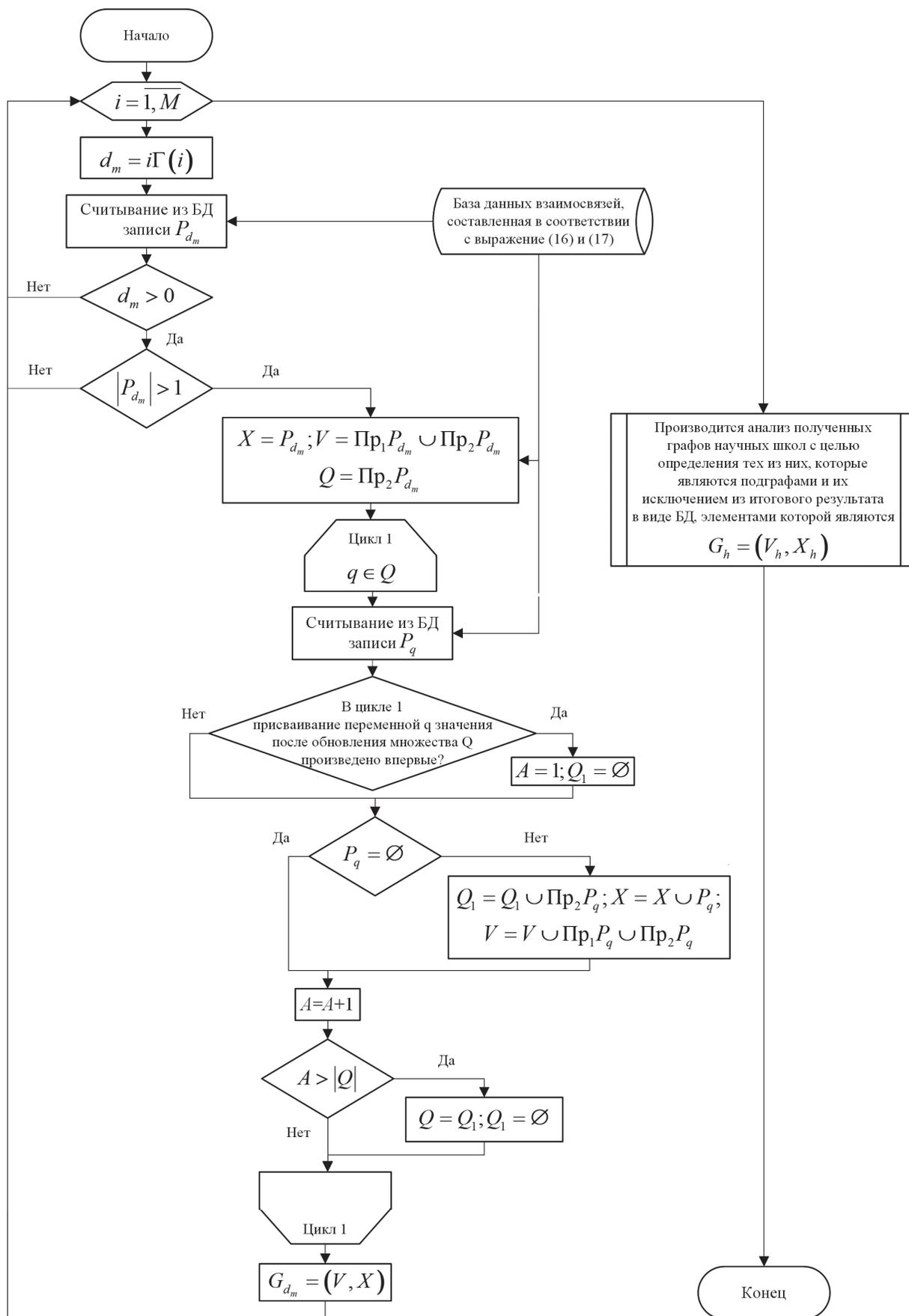


Рис. 3. Структурная схема алгоритма формирования базы данных графов для каждой из научных школ
 Fig. 3. Structural diagram of algorithm for forming a graph database for each of the scientific schools

а их структура в общем случае не предполагает наличия информации о научном руководителе/консультанте, указания данных о научной специальности, оппонентах и ведущей организации.

3. Задача поиска подграфов имеет субэкспоненциальную временную сложность алгоритма [32], что позволяет говорить о том, что предложенный алгоритм на данный момент позволяет обрабатывать относительно небольшие по размеру базы данных.

2.4. Нормированные показатели преемственности идей в рамках научной школы

В данном разделе представлены метрики, которые объективно позволяют проследить преемственность идей в рамках научной школы, тем самым обеспечив дополнительное снижение вычислительной сложности при поиске учебно-методической литературы, объективно уменьшив область поиска без снижения качества. Это обусловлено тем, что диссертация является законченной научной квалификационной работой, имеющей внутреннее единство, направленное на решение научной задачи или проблемы, и претерпевающей в процессе ее защиты соискателем достаточно обширное число процедур ее экспертной оценки, и оценка работ, на основе которых она построена, позволяет оценить преемственность идей основателя школы, научного руководителя/консультанта посредством анализа авторского коллектива в данных работах.

Вышесказанное можно оценить посредством следующих выражений:

1. Для j -го члена научной школы, в том случае, если основатель научной школы жив и не отошел от дел:

$$K_{\alpha_m}(j) = \frac{N_{\alpha_m}(j)}{N_{\text{раб}}(j)}, \quad (20)$$

где $N_{\alpha_m}(j)$ и $N_{\text{раб}}(j)$ — число научных работ j -го члена, в которых изложены основные положения, выносимые на защиту, в которых соавтором является основатель научной школы, и их общее количество, соответственно.

2. Влияние k -го члена научной школы на l -го члена научной школы ($l > k$, если члены школы из одной ветви):

$$K_{\text{вл}}(l, k) = \frac{1}{N_{\text{раб}}(l)} \sum_{q=1}^{N_{\text{раб}}(l)} \mathbf{1}(q) \left[(N_{\text{авт}_l}(q) - 1) \times \right. \\ \left. \times \text{sgn}(N_{\text{авт}_l}(q) - 1) + (1 - \text{sgn}(N_{\text{авт}_l}(q) - 1)) \right]^{-1}, \quad (21)$$

где:

- $\mathbf{1}(q)$ — индикаторная функция, равная 1, если соавтором q -й работы l -го члена научной школы, в которой опубликованы основные результаты его диссертационной работы, является k -й член научной школы, в обратном случае $\mathbf{1}(q) = 0$;
- $N_{\text{авт}_l}(q)$ — число авторов q -й работы l -го члена научной школы;

- $N_{\text{раб}}(l)$ — число научных работ l -го члена, в которых изложены основные выносимые на защиту положения.

3. Влияние не членов научной школы α_m на l -го ее члена:

$$K_{\text{вл.нш}}(l, \alpha_m) = \frac{1}{N_{\text{раб}}(l)} \times \\ \times \sum_{q=1}^{N_{\text{раб}}(l)} \frac{N_{\text{нчш}}(q)}{N_{\text{авт}_l}(q) - 1 + (1 - \text{sgn}(N_{\text{авт}_l}(q) - 1))}, \quad (22)$$

где $N_{\text{нчш}}(q)$ — число соавторов по q -й статье l -го члена научной школы α_m .

3. Заключение

Результатом представленной в данной статье работы является создание нового метода априорной классификации авторов научной и учебно-методической информации, реализующей принципы академической генеалогии, который позволяет снизить накладные расходы на передачу сервисной информации при индексации новой информации, что требуется для создания ИРИС, необходимой для реализации АЭОС. Представлены ограничения и рекомендации по практической реализации алгоритма.

Финансирование

Исследование выполнено в рамках выполнения государственного задания № 073-03-2024-046.

Funding

The study was carried out as part of state assignment No. 073-03-2024-046.

Список источников / References

1. Силкина Н. С., Соколинский Л. Б. Обзор адаптивных моделей электронного обучения. *Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика*. 2016;5(4):61–76. EDN: WZQWPJ. DOI: 10.14529/cmse160405. [Silkina N. S., Sokolinsky L. B. Survey of adaptive e-learning models. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2016;5(4):61–76. (In Russian.) EDN: WZQWPJ. DOI: 10.14529/cmse160405.]
2. Силкина Н. С., Соколинский Л. Б. Структурно-иерархическая дидактическая модель электронного обучения. *Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика*. 2019;8(4):56–83. EDN: GEWVSO. DOI: 10.14529/cmse190405. [Silkina N. S., Sokolinsky L. B. Structural-hierarchical didactic model of e-learning. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2019;8(4):56–83. (In Russian.) EDN: GEWVSO. DOI: 10.14529/cmse190405.]
3. Лернер И. М., Кадушкин В. В., Сычев А. С., Дымкова С. С., Кондратьев В. В., Вишнякова И. В. Модель адаптивной педагогической системы обучения студентов с ограниченными возможностями здоровья по слуху, осваивающих технические специальности. *Высшее образование сегодня*. 2021;(11-12):88–101. EDN: LTQYAQ. DOI: 10.18137/RNU.HET.21.11-12.P.088.

[Lerner I. M., Kadushkin V. V., Sychev A. S., Dymkova S. S., Kondratiev V. V., Vishnyakova I. V. Model of an adaptive pedagogical learning system students with hearing disabilities mastering technical specialties. *Higher Education Today*. 2021;(11-12):88–101. (In Russian.) EDN: LTQYAQ. DOI: 10.18137/RNU.HET.21.11-12.P.088.]

4. Стукалова А. А. Функциональность электронного каталога: требования российских и зарубежных пользователей. *Научные и технические библиотеки*. 2020;(8):63–84. EDN: AYMLQJ. DOI: 10.33186/1027-3689-2020-8-63-84.

[Stukalova A. A. OPAC functionality: Russian and foreign users' requirements. *Scientific and Technical Libraries*. 2020;(8):63–84. EDN: AYMLQJ. DOI: 10.33186/1027-3689-2020-8-63-84.]

5. Галиев Т. Э. Методы ранжирования поисковой информации в корпоративных поисковых системах. *Открытое образование*. 2012;(1):46–51. EDN: PLQHGJ.

[Galiev T. E. Methods of ranking of searching information in corporate searching systems. *Open Education*. 2012;(1):46–51. (In Russian.) EDN: PLQHGJ.]

6. Лернер И. М., Файзуллин Р. Р., Хайруллин А. Н., Шушпанов Д. В., Ильин В. И., Рябов И. В. Повышение удельной пропускной способности как фундаментальная проблема теории связи. Стратегия развития в постшанновскую эпоху. Часть 1. Ретроспективный обзор методов приема и обработки сигналов в частотно-селективных каналах связи при скоростях передачи информации выше скорости Найквиста. *Успехи современной радиоэлектроники*. 2023;77(1):37–50. EDN: OYDZCG. DOI: 10.18127/j20700784-202301-02.

[Lerner I. M., Fayzullin R. R., Khairullin A. N., Shushpanov D. V., Il'in V. I., Ryabov I. V. Specify capacity increasing as a fundamental problem of communication theory. Strategy development in the post-Shannon era. Part 1. Retrospective review of methods for receiving and processing signals in frequency-selective communication channels at data transfer rates faster than Nyquist rate. *Achievements of Modern Radioelectronics*. 2023;77(1):37–50. (In Russian.) EDN: OYDZCG. DOI: 10.18127/j20700784-202301-02.]

7. Лернер И. М., Файзуллин Р. Р., Хайруллин А. Н., Шушпанов Д. В., Ильин В. И., Рябов И. В. Повышение удельной пропускной способности как фундаментальная проблема теории связи. Стратегия развития в постшанновскую эпоху. Часть 2. Ретроспективный обзор методов приема и обработки сигналов в частотно-селективных каналах связи при наличии межсимвольных искажений. *Успехи современной радиоэлектроники*. 2023;77(2):16–33. EDN: TIQOYC. DOI: 10.18127/j20700784-202302-02.

[Lerner I. M., Fayzullin R. R., Khairullin A. N., Shushpanov D. V., Il'in V. I., Ryabov I. V. Specify capacity increasing as a fundamental problem of communication theory. Strategy development in the post-Shannon era. Part 2. Retrospective review of methods for receiving and processing signals in frequency-selective communication channels in the presence of ISI. *Achievements of Modern Radioelectronics*. 2023;77(2):16–33. (In Russian.) EDN: TIQOYC. DOI: 10.18127/j20700784-202302-02.]

8. Шалагин С. В. Оценка вероятности своевременной актуализации информационной модели динамически меняющейся предметной области. *Вестник НЦБЖД*. 2024;(3(61)):117–122. EDN: FYHKYL.

[Shalagin S. V. Assessment of the probability of timely updated information model dynamically changing subject domain. *Vestnik NTsBZhD*. 2024;(3(61)):117–122. (In Russian.) EDN: FYHKYL.]

9. Lombardo S. V., Condie K. S. Empowering users with a new online catalog. *Library Hi Tech*. 2000;18(2):130–141. DOI:10.1108/0737883001033346.

10. Borgman C. L. Why are online catalogs still hard to use? *Journal of the American Society for Information*

Science. 1996;47(7):493–503. DOI: 10.1002/(sici)1097-4571(199607)47:7<493::aid-asi3>3.0.co;2-p.

11. Ortiz-Repiso V. How researchers are using the OPAC of the Spanish Council for Scientific Research Library Network. *Electronic Library*. 2006;24(2):190–211. DOI: 10.1108/02640470610660378.

12. Меркулова А. Ш. Оценка качества тематического поиска информации в электронном каталоге библиотеки. Книга: Сибирь — Евразия: Труды I Международного научного конгресса. Т. 3. Новосибирск; 2016:257–264. EDN: ZHCKRX.

[Merkulova A. Sh. Assessment of the quality of thematic information search in the electronic library catalog. *Book: Siberia — Eurasia: Proc. of the 1st Int. Scientific Congress. Vol. 3*. Novosibirsk; 2016:257–264. (In Russian.) EDN: ZHCKRX.]

13. Скарук Г. А. Читательский поиск в электронном каталоге: итоги анализа поисковой статистики ГПНТБ СО РАН. *Научные и технические библиотеки*. 2017;(12):63–72. EDN: ZWTHPL. DOI:10.33186/1027-3689-2017-12-63-72.

[Skaruk G. A. Users' search in OPAC: Findings of search statistics analysis at RAS SB State Public Scientific and Technological Library. *Scientific and Technical Libraries*. 2017;(12):63–72. (In Russian.) EDN: ZWTHPL. DOI:10.33186/1027-3689-2017-12-63-72.]

14. Зайцева Е. М. Проблемы организации предметного поиска в специализированных тематических электронных каталогах. *Научные и технические библиотеки*. 2017;(2):5–13. EDN: XVNSOZ. DOI: 10.33186/1027-3689-2017-2-5-13.

[Zaytseva E. M. Subject search in special subject-related electronic catalogs. *Scientific and Technical Libraries*. 2017;(2):5–13. (In Russian.) EDN: XVNSOZ. DOI: 10.33186/1027-3689-2017-2-5-13.]

15. Селиванова Ю. Г., Масхулия Т. Л. Электронный каталог: формирование и поиск. Тенденции современной каталогизации. *Библиотечное дело*. 2004;(8):20–21.

[Selivanova Yu. G., Mashuliya T. L. Electronic catalogue: Formation and search. Trends of modern cataloguing. *Bibliotechnoe delo*. 2004;(8):20–21. (In Russian.)]

16. Сукиасян Э. Р. Два шага назад. О поиске в наших электронных каталогах. *Информационное обслуживание библиотек в меняющейся социальной среде: тенденции, новации, перспективы: Моргенштерновские чтения — 2018. Материалы всероссийской научно-практической конференции (Челябинск, 20–21 сентября 2018 г.)*. Челябинск: Челябинский Дом печати; 2018:120–126. EDN: YNMHNJ.

[Sukiasyan E. R. Two steps back. On searching our electronic catalogues. *Information Services for Libraries in a Changing Social Environment: Trends, Innovations, Prospects: Morgenstern Readings—2018. Proc. of the All-Russian Scientific and Practical Conf. (Chelyabinsk, September 20–21, 2018)*. Chelyabinsk, Chelyabinsk House of Printing; 2018:120–126. (In Russian.) EDN: YNMHNJ.]

17. Сукиасян Э. Р. Система каталогов. Как мы ее долго строили и как быстро ликвидировали. *Научные и технические библиотеки*. 2019;(6):28–40. EDN: BLGQBF. DOI: 10.33186/1027-3689-2019-6-28-40.

[Sukiasyan E. R. Catalog system. How we spent a long time building it and how quickly we liquidated it. *Scientific and Technical Libraries*. 2019;(6):28–40. (In Russian.) EDN: BLGQBF. DOI: 10.33186/1027-3689-2019-6-28-40.]

18. Скарук Г. А. Электронный каталог как объект оценки качества. *Библиосфера*. 2018;(2):85–92. EDN: XQVVOH. DOI: 10.20913/1815-3186-2018-2-85-92.

[Skaruk G. A. Electronic catalog as an object of quality assessment. *Bibliosfera*. 2018;(2):85–92. (In Russian.) EDN: XQVVOH. DOI: 10.20913/1815-3186-2018-2-85-92.]

19. Скарук Г. А. Средства и методы помощи пользователям электронного каталога в самостоятельном поиске. *Труды ГПНТБ СО РАН*. 2015;(8):275–282. EDN: UNWVRV.

[Skaruk G. A. Means and methods of electronic catalog users' assistance in their independent search. *Trudy GPNTB SO RAN*. 2015;(8):275–282. (In Russian.) EDN: UNWVRV.]

20. 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.

21. Ding Xue. Books intelligent recommendation system based on association rules. *Library and Information Forum*. 2010;1(4):48–52. (In Chinese.)

22. Sun Sh., Wang W. Collaborative filtering based on user clustering personalized book recommendation system. *Modern Library and Information Technology*. 2011;1(11):139–142. (In Chinese.)

23. 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.

24. Ji W., Liu S., Song Y., Qi J. Research of Intelligent recommendation system based on the user and association rules mining for books. *Proc. of the 2015 5th International Conference on Computer Sciences and Automation Engineering*. Atlantis Press; 2016;294–299. DOI: 10.2991/iccsae-15.2016.56.

25. Yao Y. Library resource vertical search engine based on ontology. *2017 International Conf. on Smart Grid and Electrical Automation (ICSGEA)*. IEEE, Changsha, China; 2017:672–675. DOI: 10.1109/ICSGEA.2017.159.

26. Kim J. H., Lee J. H., Lee K. J. A Study on the Issues Related to Building a Library Information System Based on Deep Learning. *2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*. Ho Chi Minh City, Vietnam, IEEE; 2021:287–289. DOI: 10.1109/SNPDWinter52325.2021.00076.

27. Trivedi Y., Kansaria K., Vora D. Smart book recommendation system for library books: LibX. *IOSR Journal of Engineering (IOSRJEN)*. 2018;8(7):01–05.

28. Zhang H., Xiao Y., Bu Z. Personalized Book Recommender System Based on Chinese Library Classification. *14th Web Information Systems and Applications Conf. (WISA)*. Liuzhou, China; 2017:127–131. DOI: 10.1109/WISA.2017.42.

29. Anikieva M. A. The choice of customization strategies in training: An overview of parameters and their systematization. *Australasian Journal of Educational Technology*. 2021;37(3):170–186. EDN: SFXOOH. DOI 10.14742/ajet.6185.

30. Zhao C., Sahni S. String correction using the Damerau–Levenshtein distance. *BMC Bioinformatics*. 2019;20(Suppl 11):277. DOI: 0.1186/s12859-019-2819-0.

31. Окулов С. М. Алгоритмы обработки строк. 5-е изд., электрон. М.: Лаборатория знаний; 2024. 258 с.

[Okulov S. M. String processing algorithms. 5th ed., electronic. Moscow, Laboratoriya znaniy; 2024. 258 p. (In Russian.)]

32. Черемисинов Д. И., Черемисинова Л. Д. Поиск часто встречающихся подграфов. *Big Data and Advanced Analytics*. 2018;(4):173–178. EDN: YZJFKX.

[Cheremisinov D. I., Cheremisinova L. D. Frequent subgraph mining. *Big Data and Advanced Analytics*. 2018;(4):173–178. (In Russian.) EDN: YZJFKX.]

Информация об авторах

Григорьев Сергей Георгиевич, чл.-корр. РАО, доктор тех. наук, профессор, профессор департамента информатики, управления и технологий, Институт цифрового образования, Московский городской педагогический университет, г. Мо-

сква, Россия; ORCID: <https://orcid.org/0000-0002-0034-9224>; e-mail: grigorsg@yandex.ru

Лернер Илья Михайлович, доктор тех. наук, старший научный сотрудник, Башкирский государственный педагогический университет им. М. Акмуллы, г. Уфа, Республика Башкортостан, Россия; 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; ORCID: <https://orcid.org/0000-0001-5788-3113>; e-mail: aviarp@mail.ru

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

Арутюнова Наталья Константиновна, канд. физ.-мат. наук, доцент кафедры прикладной математики и информатики, Институт компьютерных технологий и защиты информации, Казанский национальный исследовательский технический университет им. А. Н. Туполева-КАИ, Республика Татарстан, г. Казань, Россия; ORCID: <https://orcid.org/0000-0001-8450-7535>; e-mail: arut_natalie@mail.ru

Григорьева Марина Александровна, канд. пед. наук, доцент, доцент департамента информатики, управления и технологий, Институт цифрового образования, Московский городской педагогический университет, г. Москва, Россия; ORCID: <https://orcid.org/0009-0006-7296-3107>; e-mail: grigorevama@mgpu.ru

Information about the authors

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), Senior Research Fellow, Akmulla Bashkir State Pedagogical University, Ufa, The Republic of Bashkortostan, Russia; 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; ORCID: <https://orcid.org/0000-0001-5788-3113>; e-mail: aviarp@mail.ru

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

Natalya K. Arutyunova, Candidate of Sciences (Physics and Mathematics), Associate Professor at the Department of Applied Mathematics and Informatics, Institute for Computer Technologies and Information Protection, Kazan National Research Technical University named after A. N. Tupolev-KAI, Kazan, The Republic of Tatarstan, Russia; ORCID: <https://orcid.org/0000-0001-8450-7535>; e-mail: arut_natalie@mail.ru

Marina A. Grigorieva, Candidate of Sciences (Education), Docent, Associate Professor at the Department of IT, Management and Technology, Institute of Digital Education, Moscow City University, Moscow, Russia; ORCID: <https://orcid.org/0009-0006-7296-3107>; e-mail: grigorevama@mgpu.ru

Поступила в редакцию / Received: 02.12.24.

Поступила после рецензирования / Revised: 17.01.25.

Принята к печати / Accepted: 22.01.25.