

УДК 004.852; 021.6

DOI: [10.26102/2310-6018/2026.54.3.013](https://doi.org/10.26102/2310-6018/2026.54.3.013)

Гибридная семантическая редукция текстов в библиотечных информационных системах

И.С. Рзянкин✉, М.В. Носков

Сибирский федеральный университет, Красноярск, Российская Федерация

Резюме. Актуальность исследования обусловлена ростом объемов текстовой информации в библиотечных информационных системах и необходимостью обеспечения быстрой и содержательной навигации по электронным фондам в условиях ограниченных вычислительных ресурсов. Существующие решения автоматической суммаризации ориентированы преимущественно на использование крупномасштабных языковых моделей, что затрудняет их внедрение в локальную библиотечную инфраструктуру. В связи с этим работа направлена на разработку ресурсосберегающего метода семантической редукции текста, обеспечивающего баланс между качеством смыслового представления и вычислительной доступностью. Ведущим подходом является гибридная архитектура, основанная на последовательном применении лексической редукции с использованием облаков слов и нейросетевой суммаризации компактными моделями. В исследовании предложена контекстно-ориентированная метрика оценки релевантности, учитывающая семантическую целостность, структурные характеристики и доменно значимые термины библиотечной среды. Экспериментальное исследование на корпусе из 1178 документов показало, что гибридный подход обеспечивает прирост показателей релевантности при одновременном сокращении времени инференса по сравнению с прямой нейросетевой суммаризацией полного текста. Полученные результаты подтверждают возможность практического внедрения предложенного метода в библиотечных информационных системах с ограниченной вычислительной инфраструктурой и его применимость для задач навигации и каталогизации.

Ключевые слова: семантическая редукция текста, автоматическая суммаризация, облако слов, библиотечные информационные системы, гибридные методы обработки текста, нейросетевые модели, оценка релевантности, Library Relevance Score.

Для цитирования: Рзянкин И.С., Носков М.В. Гибридная семантическая редукция текстов в библиотечных информационных системах. *Моделирование, оптимизация и информационные технологии*. 2026;14(3). URL: <https://moitvivr.ru/ru/journal/article?id=2220> DOI: 10.26102/2310-6018/2026.54.3.013

Hybrid semantic reduction of texts in library information systems

I.S. Rzyankin✉, M.V. Noskov

Siberian Federal University, Krasnoyarsk, the Russian Federation

Abstract. The relevance of the study is determined by the continuous growth of textual information in library information systems and the need to ensure fast and meaningful navigation across electronic collections under constrained computational resources. Existing automatic summarization solutions are primarily oriented toward large-scale language models, which limits their practical deployment within local library infrastructures. In this context, the paper aims to develop a resource-efficient method of semantic text reduction that balances the quality of semantic representation with computational feasibility. The proposed approach is based on a hybrid architecture that sequentially combines lexical reduction using word clouds with neural summarization performed by compact models. In addition, a context-oriented evaluation metric is introduced to assess relevance with regard to semantic coherence, structural characteristics, and domain-specific terms significant for the library environment. An

experimental study conducted on a corpus of 1178 documents demonstrates that the hybrid approach improves relevance indicators while simultaneously reducing inference time compared to direct neural summarization of the full text. The obtained results confirm the practical applicability of the proposed method for library information systems operating under limited computational infrastructure and its usefulness for navigation and cataloging tasks.

Keywords: semantic text reduction, automatic summarization, word cloud, library information systems, hybrid text processing methods, neural models, relevance evaluation, Library Relevance Score.

For citation: Rzyankin I.S., Noskov M.V. Hybrid semantic reduction of texts in library information systems. *Modeling, Optimization and Information Technology*. 2026;14(3). (In Russ.). URL: <https://moitvvt.ru/ru/journal/article?id=2220> DOI: 10.26102/2310-6018/2026.54.3.013

Введение

Современные библиотечные информационные системы функционируют в условиях возрастающей сложности задач поиска, навигации и представления знаний. Расширение электронных фондов и развитие цифровых сервисов усиливают потребность в методах семантической обработки текста, позволяющих получать содержательное представление документа без обращения к его полному объёму. В этом контексте особое значение приобретают подходы, направленные на смысловое сжатие текстов и поддержку навигации в документных коллекциях.

Одним из наиболее активно развивающихся направлений в данной области является автоматическая суммаризация, формирующаяся на стыке информационного поиска и обработки естественного языка. Современные методы суммаризации ориентированы на извлечение или генерацию кратких текстов, сохраняющих ключевые смысловые элементы исходного документа и пригодных для использования в прикладных информационных системах. Существенный вклад в развитие данного направления внесли нейросетевые модели, демонстрирующие высокое качество семантической обработки.

Вместе с тем практическое внедрение подобных решений в библиотечные системы сталкивается с рядом эксплуатационных ограничений. В отличие от исследовательских и коммерческих платформ, библиотеки нередко работают в условиях ограниченных вычислительных ресурсов и локальной инфраструктуры, что существенно сужает спектр применимых моделей и архитектур. В связи с этим актуальной становится задача поиска таких подходов к семантической редукции текста, которые обеспечивали бы баланс между качеством результатов и вычислительной доступностью.

Настоящая работа посвящена разработке и экспериментальной оценке ресурсосберегающего гибридного подхода к семантической редукции библиотечных текстов, основанного на последовательном применении лексических методов и компактных нейросетевых моделей. Особое внимание уделяется практической полезности суммаризации для задач библиотечной навигации и каталогизации, что отражено в предлагаемой контекстно-ориентированной метрике оценки релевантности. Ожидаемым результатом исследования является обоснование возможности применения интеллектуальных методов обработки текста в библиотечных информационных системах без существенного увеличения вычислительной нагрузки.

Развитие библиотечных информационных систем в последние годы сопровождается устойчивым ростом объемов текстовой информации, включающей аннотации, рефераты, библиографические описания, рецензии и пользовательские комментарии. Масштабирование электронных фондов и переход к гибридным форматам хранения знаний приводят к тому, что традиционные средства поиска и каталогизации все чаще оказываются недостаточными для быстрой и осмысленной ориентации в

содержании документов. Указанная тенденция отмечается в исследованиях, посвященных цифровой трансформации библиотек и управлению растущими массивами текстовых данных [1, 2].

В этих условиях возрастает потребность в автоматических средствах семантической навигации, позволяющих получать сжатое и структурированное представление содержания документов без обращения к полному тексту. Одним из ключевых инструментов решения данной задачи является автоматическая суммаризация, ориентированная на выделение основных смысловых элементов текста. Дополнительно применяются лексические методы, включая построение облаков слов, используемые для предварительного тематического анализа и экспресс-оценки содержания документов. Подобные подходы рассматриваются как способы снижения когнитивной нагрузки при работе с крупными текстовыми коллекциями [3, 4].

Одновременно с этим активно развиваются нейросетевые модели обработки естественного языка, обеспечивающие высокое качество семантического сопоставления и генерации текстов. Однако значительная часть существующих исследований ориентирована на вычислительно ресурсоемкие модели и облачные инфраструктуры, что существенно ограничивает их применимость в реальных библиотечных системах. Практика показывает, что многие библиотеки, особенно региональные и университетские, функционируют в условиях ограниченных вычислительных ресурсов, используют локальную инфраструктуру и работают преимущественно с русскоязычными корпусами, для которых доступно ограниченное число специализированных моделей и размеченных данных. Данные ограничения подробно обсуждаются в работах, посвященных устойчивости и эффективности NLP-решений в ресурсно-ограниченных сценариях [5, 6].

В результате в библиотечной среде формируется противоречие между растущей потребностью в интеллектуальной семантической обработке текстов и ограниченными возможностями ее практической реализации. Это противоречие определяет актуальность поиска таких методов суммаризации и семантической навигации, которые сочетали бы приемлемое качество результатов с низкими вычислительными затратами и могли бы быть интегрированы в существующие библиотечные информационные системы без существенной модернизации инфраструктуры.

Современные исследования в области автоматической обработки текстов преимущественно сосредоточены на нейросетевых методах суммаризации, ориентированных на достижение максимально возможного качества с точки зрения семантического соответствия и полноты содержания. В работах данного направления основное внимание уделяется архитектурам трансформеров, стратегиям обучения и тонкой настройке моделей, а также сравнению результатов по стандартным метрикам качества, прежде всего ROUGE. Подобные подходы демонстрируют высокие показатели точности, однако достигаются они за счет использования крупных и вычислительно ресурсоемких моделей, требующих значительных вычислительных мощностей и развитой инфраструктуры, что ограничивает их практическую применимость в библиотечных информационных системах [7, 8].

Однако параллельно развивается и направление лексико-статистических методов обработки текста, ориентированных на простоту реализации и минимальные вычислительные затраты. К таким подходам относятся извлечение ключевых слов, частотный анализ терминов и построение облаков слов, которые используются для предварительного тематического анализа и обзорного представления содержания документов. Эти методы отличаются высокой интерпретируемостью и низкими

требованиями к ресурсам, однако в изолированном виде они не обеспечивают достаточной глубины семантического представления текста и, как правило, рассматриваются как вспомогательные инструменты, не предназначенные для полноценной семантической редукции [4, 9].

Несмотря на активное развитие обоих направлений, анализ существующих публикаций показывает, что нейросетевые и лексические методы чаще рассматриваются как альтернативные, а не взаимодополняющие решения. При этом условия эксплуатации библиотечных информационных систем – ограниченные вычислительные ресурсы, локальная инфраструктура и ориентация на прикладные задачи навигации и каталогизации – нередко остаются вне фокуса внимания исследователей. В результате предлагаемые методы либо оказываются трудно реализуемыми в реальных условиях, либо не обеспечивают необходимого баланса между качеством и вычислительной эффективностью [5, 6].

Отдельного внимания заслуживает использование облаков слов в современных исследованиях. В большинстве публикаций облака слов применяются исключительно как средство визуализации тематической структуры текста, в то время как их использование в качестве этапа предварительной лексической редукции практически не исследовано. При этом подчеркивается, что классические облака слов не предназначены для глубокого семантического анализа и обладают существенными ограничениями с точки зрения интерпретации и аналитической полезности. В научной литературе практически отсутствуют подходы, в которых облако слов рассматривалось бы как функциональный этап предварительной лексической редукции, подготавливающий текст к последующей нейросетевой обработке [4, 10].

Кроме того, оценка качества суммаризации в существующих исследованиях в основном опирается на универсальные автоматические метрики, не учитывающие специфику прикладных сценариев использования результатов. Работы по переосмыслению процедур оценки подчеркивают, что такие метрики слабо отражают практическую полезность сводок для пользователей и не фиксируют доменно-ориентированные аспекты релевантности, включая задачи информационной навигации и отбора документов. Доменно-специфичные метрики, адаптированные к библиотечному контексту, остаются недостаточно разработанными и редко применяются на практике [11, 12].

Проведенный анализ свидетельствует об отсутствии целостных решений, в которых лексическая и нейросетевая семантическая редукция рассматривались бы как взаимодополняющие этапы единой архитектуры, а оценка результатов учитывала бы специфику библиотечной среды. Указанные обстоятельства формируют научный пробел, связанный с необходимостью разработки ресурсосберегающих гибридных методов и контекстно-ориентированных метрик релевантности, адаптированных к условиям реальных библиотечных информационных систем.

Целью настоящего исследования является разработка и экспериментальная оценка ресурсосберегающего гибридного метода семантической редукции текста, ориентированного на использование в библиотечных информационных системах и учитывающего ограничения их вычислительной и инфраструктурной среды. В рамках работы предполагается получить практический результат в виде воспроизводимой архитектуры, позволяющей сократить объем обрабатываемого текста и вычислительные затраты без существенной потери смысловой информативности, значимой для задач библиотечной навигации и каталогизации.

Для достижения поставленной цели в исследовании последовательно решается ряд взаимосвязанных задач. Анализируется применимость лексических методов на основе облаков слов и нейросетевой суммаризации в контексте библиотечных сценариев, с акцентом на их сильные и слабые стороны при работе в условиях ограниченных ресурсов. На основе проведенного анализа предлагается гибридная архитектура WordCloud-to-MiniModel, в которой лексическая редукция используется как осмысленный предварительный этап подготовки текста к последующей нейросетевой обработке. В работе вводится контекстно-ориентированная метрика Library Relevance Score, предназначенная для оценки качества семантической редукции с учётом специфики библиотечной среды и задач информационной навигации. Далее проводится сравнительное экспериментальное исследование предложенного подхода и базовых методов, позволяющее оценить как качество результатов, так и вычислительные затраты. Завершающим этапом является анализ практической применимости разработанного метода в условиях реальных библиотечных информационных систем с ограниченными вычислительными ресурсами.

Материалы и методы

Лексическая редукция текста на основе облаков слов. В рамках предлагаемого подхода лексическое сжатие текста рассматривается как самостоятельный функциональный этап семантической обработки, а не как вспомогательная визуализация. Его основу составляет построение облака слов на базе статистического анализа лексических единиц с учетом их частотности и информативности. Для этого исходный текст предварительно нормализуется, после чего из него извлекаются значимые термины с использованием классических схем взвешивания, таких как TF-IDF и близкие к нему статистические методы. Подобные схемы позволяют снизить влияние общезыковой лексики и усилить роль содержательных слов, отражающих предметную область документа [13].

Важной частью предварительной статистической редукции является фильтрация шума, направленная на исключение слов и словосочетаний, не несущих тематической нагрузки. Помимо стандартного удаления стоп-слов, в библиотечном контексте учитываются особенности библиографических описаний и аннотаций, в которых часто присутствуют формальные, жанровые или служебные элементы, не связанные напрямую с содержанием документа. Результатом данного этапа становится компактный набор лексических маркеров, фиксирующих устойчивые смысловые акценты текста и позволяющих сократить объем обрабатываемой информации без обращения к вычислительно затратным семантическим моделям.

На основе очищенного набора терминов формируется тематическое ядро документа, представляющее собой совокупность наиболее значимых понятий, характеризующих его содержание. В отличие от традиционного использования облаков слов в качестве обзорного или иллюстративного инструмента, в данном исследовании выделенное тематическое ядро трактуется как результат целенаправленной лексической фильтрации, предназначенной для дальнейшей машинной обработки. Такое понимание согласуется с исследованиями, в которых облака слов применяются для тематического анализа и предварительного изучения текстовых корпусов, однако выходит за рамки чисто визуального представления информации [10].

Функциональная роль данного этапа в библиотечном контексте определяется задачами ориентации и навигации в массиве текстов. Для библиотечных информационных систем он обеспечивает быстрое предварительное представление

содержания документа, облегчает тематическую классификацию и снижает нагрузку на последующие стадии обработки. Принципиально важно, что облако слов здесь не выступает конечным результатом анализа, а используется как промежуточный механизм смыслового сжатия, подготавливающий текст к нейросетевой суммаризации. Такой подход позволяет учитывать известные ограничения классических облаков слов как аналитического инструмента и одновременно использовать их сильные стороны – интерпретируемость и низкую вычислительную сложность – в составе более общей семантической архитектуры [14].

Нейросетевая суммаризация с использованием компактных моделей. При проектировании нейросетевого этапа суммаризации в библиотечных информационных системах ключевым фактором является соотношение качества получаемых сводок и вычислительных затрат. В отличие от большинства современных исследований, ориентированных на использование крупномасштабных языковых моделей, в данной работе акцент сделан на компактных нейросетевых архитектурах, способных функционировать в условиях ограниченной вычислительной инфраструктуры. Такой выбор обусловлен спецификой реальных библиотечных систем, где вычислительная база чаще всего представлена локальными серверами на базе центральных процессоров, а использование графических ускорителей либо экономически неоправданно, либо организационно затруднено [15].

В качестве основы для нейросетевой суммаризации рассматриваются малые трансформерные модели, оптимизированные для инференса на CPU. К ним относятся компактные encoder–decoder-архитектуры и дистиллированные модели, в которых снижение числа параметров и глубины сети достигается за счёт переноса знаний от более крупных моделей. Исследования показывают, что подобные подходы позволяют сохранить значительную часть качества абстрактивной суммаризации при существенном снижении требований к памяти и вычислительным ресурсам, что делает их применимыми в прикладных сценариях с ограниченным вычислительным бюджетом [16].

Использование крупных языковых моделей в библиотечных системах связано с рядом ограничений, выходящих за рамки чисто технических характеристик. Полноразмерные LLM требуют значительных объемов оперативной памяти, наличия специализированных ускорителей и сложной инфраструктуры обслуживания, а также демонстрируют высокую энергетическую и эксплуатационную стоимость. Эти факторы существенно ограничивают их практическое внедрение в условиях библиотек, где важны стабильность работы, предсказуемость времени отклика и независимость от внешних облачных сервисов. Дополнительным ограничением является специфика русскоязычных корпусов, для которых число полноценно адаптированных и проверенных крупномасштабных моделей остается ограниченным [17].

Отдельного внимания заслуживает применение нейросетевой суммаризации после предварительной лексической редукции текста. Передача в модель не полного документа, а уже сжатого и тематически очищенного представления позволяет существенно сократить длину входного контекста и снизить вычислительную нагрузку на модель. Исследования гибридных пайплайнов показывают, что двухэтапные схемы «редукция → суммаризация» позволяют компактным моделям сохранять ключевые смысловые элементы текста и формировать информативные сводки, сопоставимые по прикладной полезности с результатами более ресурсоемких решений, применяемых к полному тексту [18].

Гибридный подход WordCloud-to-MiniModel. Предлагаемый гибридный подход WordCloud-to-MiniModel основан на двухуровневой логике семантической редукции, в

которой лексические и нейросетевые методы рассматриваются не как альтернативные, а как функционально взаимодополняющие компоненты единого конвейера. Ключевая идея заключается в том, что первичное лексическое сжатие текста позволяет зафиксировать тематическое ядро документа в интерпретируемой форме, после чего компактная нейросетевая модель работает уже с существенно уменьшенным и очищенным представлением содержания. Такой порядок обработки принципиально отличается от традиционных схем, в которых нейросетевая суммаризация применяется напрямую к полному тексту, что приводит к избыточной вычислительной нагрузке и снижает воспроизводимость результатов в прикладных условиях [19].

На первом этапе конвейера исходный текст подвергается лексической редукции на основе облака слов, в ходе которой осуществляется фильтрация шумовой лексики и выделение устойчивых тематических маркеров. Результатом этого этапа является компактное представление содержания документа, отражающее его ключевые понятия и темы без потери связи с исходной лексической структурой. Далее сформированное тематическое ядро используется в качестве входа для нейросетевой модели суммаризации, которая выполняет генерацию краткого связного описания уже на основе редуцированного текста. Такая последовательность позволяет ограничить длину входного контекста, стабилизировать поведение модели и снизить требования к вычислительным ресурсам, не жертвуя при этом прикладной полезностью итоговой сводки [19, 20].

Сравнение гибридного подхода с изолированным применением лексических и нейросетевых методов показывает его принципиальные преимущества. Использование только облаков слов обеспечивает высокую интерпретируемость и низкую вычислительную сложность, однако не позволяет формировать связные текстовые сводки, пригодные для дальнейшего использования в информационных сервисах. Применение нейросетевой суммаризации без предварительной редукции, напротив, демонстрирует более высокое качество текста, но сопровождается значительными вычислительными затратами и повышенной чувствительностью к длине и структуре входных данных. Гибридная схема сочетает сильные стороны обоих подходов, снижая нагрузку на нейросетевой этап за счет предварительного смыслового сжатия и одновременно компенсируя ограниченность лексических методов за счет генеративных возможностей компактных моделей.

В контексте библиотечных информационных систем гибридный подход WordCloud-to-MiniModel позволяет выстроить архитектуру семантической редукции, ориентированную на реальные условия эксплуатации. Для библиотечной практики это принципиально. Он обеспечивает предсказуемое время отклика, воспроизводимость результатов и возможность внедрения на доступной аппаратной платформе, сохраняя при этом ориентацию на задачи тематической навигации и каталогизации. Визуально логика двухуровневой редукции может быть представлена в виде простой схемы, отражающей последовательность этапов «исходный текст – лексическая редукция – нейросетевая суммаризация – итоговая сводка», что подчеркивает целостность и прикладную направленность предлагаемой архитектуры.

Общая логика предлагаемого гибридного подхода WordCloud-to-MiniModel и последовательность этапов двухуровневой семантической редукции текста представлены на Рисунке 1.

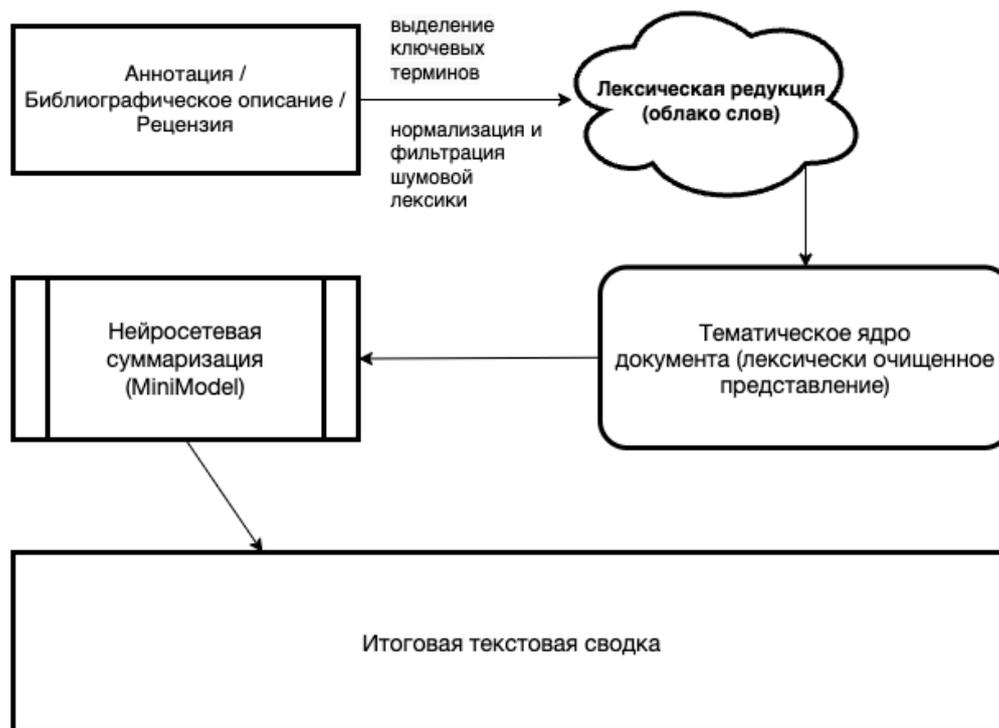


Рисунок 1 – Схема гибридного подхода WordCloud-to-MiniModel для семантической редукции текста

Figure 1 – Two-stage architecture of the WordCloud-to-MiniModel hybrid framework for semantic text reduction

Метрика оценки релевантности для библиотечных систем

Ограничения стандартных метрик суммаризации. Оценка качества автоматической суммаризации в большинстве современных исследований опирается на универсальные метрики, прежде всего ROUGE и BERTScore, ориентированные на сопоставление полученной сводки с эталонным текстом. Эти показатели широко применяются для формального сравнения моделей и хорошо зарекомендовали себя в задачах общего назначения, однако их интерпретация ограничивается измерением лексического или семантического совпадения с референсной суммаризацией. За рамками такой оценки остается вопрос того, какие именно элементы содержания были сохранены и насколько они значимы в прикладном контексте использования результата.

В библиотечных информационных системах суммаризация выполняет иную функцию, чем в типовых задачах обработки новостных или аналитических текстов. Здесь сводка должна не только передавать общий смысл документа, но и обеспечивать быструю ориентацию в его тематике, поддерживать процессы каталогизации и тематического поиска. С практической точки зрения особое значение имеют библиографически значимые элементы содержания, такие как предметная область, жанровая принадлежность, тип документа, характер исследуемой проблемы или область применения результатов. Метрики семейства ROUGE чувствительны к совпадению формулировок и последовательности слов, но не позволяют различить, отражены ли в суммаризации именно те компоненты, которые критически важны для библиотечной навигации и классификации.

Использование BERTScore частично снижает зависимость от поверхностного лексического совпадения за счет контекстных представлений, однако и в этом случае оценка сводится к измерению общего семантического сходства с эталоном. Доменные

различия между элементами содержания при этом не учитываются: упоминание ключевого термина, значимого для библиотечной классификации, и второстепенной детали оказываются эквивалентными с точки зрения метрики. В результате суммаризация, формально демонстрирующая высокие значения стандартных показателей, может не выполнять своей основной функции в библиотечной среде – облегчать отбор и тематическую ориентацию документов.

Дополнительное ограничение связано с тем, что стандартные метрики рассматривают суммаризацию как изолированную задачу, не связанную с последующими сценариями использования. В реальных библиотечных информационных системах сводка выступает частью более сложного процесса, включающего поиск, фильтрацию и классификацию документов. Отсутствие учета этой связи приводит к разрыву между формально высоким качеством суммаризации и ее практической полезностью для библиотечных специалистов и пользователей, что указывает на необходимость разработки контекстно-ориентированных подходов к оценке релевантности результатов.

Library Relevance Score (LRS). Предлагаемая метрика Library Relevance Score (LRS) вводится для оценки качества суммаризации с учетом прикладных задач библиотечных информационных систем и ориентирована на измерение не абстрактного сходства с эталоном, а практической полезности результата для навигации, каталогизации и первичного отбора документов. Идея метрики заключается в объединении формальной семантической оценки с признаками, отражающими структуру текста и доменную значимость терминов, характерных для библиотечного контекста. Такой подход позволяет приблизить автоматическую оценку к тому, как качество сводки воспринимается специалистами и пользователями при работе с фондом.

В основе LRS лежит комбинация трех компонент. Семантическая составляющая отвечает за сохранение основного смысла документа и может опираться на стандартные показатели семантического сходства, такие как ROUGE или BERTScore, используемые не как самостоятельная цель, а как базовый индикатор адекватности содержания. Структурная компонента учитывает, насколько суммаризация отражает ключевые элементы текста, значимые для библиотечной навигации, включая наличие тематического ядра, устойчивых понятий и логически связанных фрагментов, позволяющих быстро соотнести документ с определенной областью знаний. Библиотечно-ориентированная компонента вводится для фиксации роли доменных терминов, связанных с жанром, предметной областью, типом документа и другими характеристиками, важными для каталогизации и тематического поиска.

Интерпретация значений LRS носит прикладной характер. Более высокое значение метрики соответствует сводкам, которые одновременно сохраняют семантическую целостность исходного текста, отражают его тематическую структуру и включают ключевые библиотечно значимые термины. Низкие значения указывают либо на чрезмерно общую суммаризацию, не позволяющую ориентироваться в содержании документа, либо на формально корректный, но слабо полезный для библиотечных задач результат. В отличие от универсальных метрик, LRS не претендует на замену существующих показателей, а дополняет их, обеспечивая более адекватную оценку качества суммаризации в условиях реальных библиотечных информационных систем.

В общем виде метрика Library Relevance Score может быть представлена следующим образом:

$$LRS = Ssem \cdot (\alpha \cdot Sstruct + \beta \cdot Slib), \quad (1)$$

где $Ssem$ – семантическая составляющая, отражающая сохранение основного смысла документа (например, нормализованное значение ROUGE или BERTScore); $Sstruct$ –

структурная компонента, характеризующая наличие и выраженность тематического ядра в суммаризации; *Slib* – библиотечно-ориентированная компонента, учитывающая присутствие доменно значимых терминов; α , β – коэффициенты, задающие относительную значимость структурных и библиотечных факторов.

В проведенных экспериментах коэффициенты метрики принимались, равными $\alpha = \beta = 0,33$; компонент *Sstruct* рассчитывался как доля терминов тематического ядра (top-10), присутствующих в суммаризации, а компонент *Slib* – как доля доменно-специфичных терминов из заранее сформированного словаря библиотечно значимых понятий. При этом Library Relevance Score ориентирована на прикладные сценарии использования и не претендует на универсальную метрику качества суммаризации: предложенная формализация предназначена прежде всего для сопоставления результатов с точки зрения их практической полезности в задачах библиотечной навигации и каталогизации, а не для точного численного ранжирования моделей.

Экспериментальное исследование

Корпус данных и сценарии экспериментов. Экспериментальное исследование проводилось на корпусе текстов, отражающих типовые информационные объекты, с которыми работают современные библиотечные информационные системы. Корпус формировался из реальных записей электронного каталога (обезличенных). В его состав были включены аннотации к книгам и научным изданиям, краткие библиографические описания, редакционные и пользовательские рецензии, а также расширенные описания содержания, используемые в электронных каталогах. Подобные тексты отличаются высокой плотностью фактической информации, ограниченным объемом и выраженной ориентированностью на тематическую навигацию, что делает их показательными для оценки методов семантической редукции в библиотечном контексте.

Выбор указанных типов текстов обусловлен их практической значимостью. Аннотации и описания служат основным источником первичного представления документа для пользователя, тогда как рецензии и комментарии дополняют формальное описание оценочными и интерпретационными элементами. В совокупности они формируют неоднородный, но типичный для библиотек массив данных, в котором сочетаются структурированные фрагменты и свободный текст. Это позволяет оценивать работу предложенного метода в условиях, близких к реальной эксплуатации, а не на искусственно унифицированных корпусах.

Сценарии экспериментов были ориентированы на задачи, характерные для библиотечных сервисов: быстрое ознакомление с содержанием документа, предварительный тематический отбор и поддержка навигации по коллекции. Для каждого типа текста оценивалась способность метода формировать компактные и информативные сводки, сохраняющие ключевые смысловые и доменно значимые элементы. Такой выбор корпуса и сценариев обеспечивает репрезентативность результатов и позволяет соотнести формальные показатели качества с практической полезностью суммаризации в библиотечных информационных системах.

Конфигурация эксперимента. Сравнительная оценка проводилась в единой экспериментальной конфигурации, предусматривающей сопоставление нескольких подходов к семантической редукции текста. В качестве базового лексического ориентира рассматривался метод, основанный на построении облака слов, используемый для выделения тематического ядра без применения нейросетевых моделей. Вторым вариантом представлял собой нейросетевую суммаризацию с использованием компактной модели, применяемой напрямую к исходному тексту. Третий, основной для данного исследования, вариант реализовывал гибридную схему WordCloud-to-MiniModel, в которой нейросетевая суммаризация выполнялась после предварительной лексической

редукции. Дополнительно, в отдельных экспериментах использовалась более крупная нейросетевая модель в качестве ориентирного решения, позволяющего соотнести качество компактных и гибридных подходов с верхней границей по качеству.

Все методы оценивались в идентичных условиях и на одном и том же корпусе данных. Для нейросетевых моделей были зафиксированы архитектура, параметры генерации и максимальная длина входного контекста, что исключало влияние вариативности настроек на результаты. Для лексического этапа использовались единые правила нормализации текста и фильтрации шума, а размер тематического ядра подбирался таким образом, чтобы обеспечить сопоставимую степень сжатия для всех сценариев. Вычислительная среда и программное окружение оставались неизменными на протяжении всех серий экспериментов, что обеспечивало корректность сравнения как по показателям качества, так и по вычислительным затратам.

Такая конфигурация эксперимента позволила изолированно оценить вклад каждого подхода, а также проанализировать эффект их совместного применения. В результате стало возможным сопоставить не только формальные метрики суммаризации, но и практическую пригодность различных методов для использования в библиотечных информационных системах с ограниченными ресурсами.

Экспериментальные вычисления выполнялись на локальной рабочей станции, сопоставимой с типовой инфраструктурой университетских библиотек: CPU Intel Core i5-12400 (6 ядер / 12 потоков), 32 ГБ ОЗУ и NVMe SSD 1 ТБ. Графические ускорители не применялись; все этапы обработки текста и суммаризации выполнялись на CPU. Выбор такой конфигурации был сделан намеренно, чтобы оценить применимость метода в условиях ограниченных вычислительных ресурсов и локального развертывания, характерных для библиотечных информационных систем.

Результаты

Оценка качества суммаризации проводилась на корпусе из 1178 текстовых документов, включающем аннотации, библиографические описания и краткие рецензии, характерные для библиотечных информационных систем.

Полученные результаты демонстрируют различия между сравниваемыми подходами как по показателям качества суммаризации, так и по вычислительным затратам. Использование только лексического сжатия на основе облака слов характеризуется минимальными требованиями к ресурсам и высокой интерпретируемостью, однако ожидаемо уступает нейросетевым методам по значениям ROUGE и BERTScore. В проведенных экспериментах снижение значений ROUGE-1 по сравнению с нейросетевыми подходами составляло порядка 10–15 пунктов, что связано с отсутствием генеративного компонента и ограниченной способностью лексической редукции формировать связное текстовое представление.

Прямая нейросетевая суммаризация компактной моделью демонстрирует более высокие значения стандартных метрик, однако остается чувствительной к длине и структурной неоднородности входного текста, особенно в случае аннотаций и рецензий со смешанной жанровой организацией. В таких сценариях наблюдалось снижение стабильности качества, выражающееся в колебаниях значений ROUGE-2 в пределах 3–5 пунктов в зависимости от структуры исходного документа.

Гибридный подход WordCloud-to-MiniModel демонстрирует выигрыш по показателям качества и эффективности. По результатам экспериментов прирост ROUGE-1 по сравнению с прямой мини-суммаризацией составлял в среднем 4–7 пунктов, а ROUGE-2 – 3–5 пунктов при сопоставимой длине итоговой сводки. При этом по значениям ROUGE и BERTScore гибридный метод лишь незначительно уступал

ориентирному решению на базе более крупной модели, что подтверждает эффективность предварительного сокращения входного текста.

Наиболее выраженные различия выявляются при анализе Library Relevance Score. В гибридной схеме значение данной метрики возросло на 9–14 % по сравнению с прямой нейросетевой суммаризацией, что указывает на более частое сохранение тематического ядра документа и включение библиотечно значимых терминов. Это подтверждает, что предварительная лексическая фильтрация снижает влияние второстепенных фрагментов и направляет нейросетевую модель на обработку содержательно значимых элементов текста.

Анализ вычислительных затрат показывает, что гибридный подход обеспечивает сокращение времени инференса в 1,8–2,3 раза по сравнению с нейросетевой суммаризацией полного текста. Уменьшение длины входного контекста приводит к более стабильному времени отклика и снижению нагрузки на вычислительную инфраструктуру, что особенно важно для систем, функционирующих на CPU и обслуживающих запросы в режиме реального времени. Применение полноразмерных моделей, напротив, остается наиболее ресурсоемким вариантом и требует специализированной аппаратной поддержки.

Как показано на Рисунке 2, гибридный подход WordCloud-to-MiniModel занимает промежуточное положение между лексическими и нейросетевыми методами, обеспечивая повышенную библиотечную релевантность при умеренных вычислительных затратах.

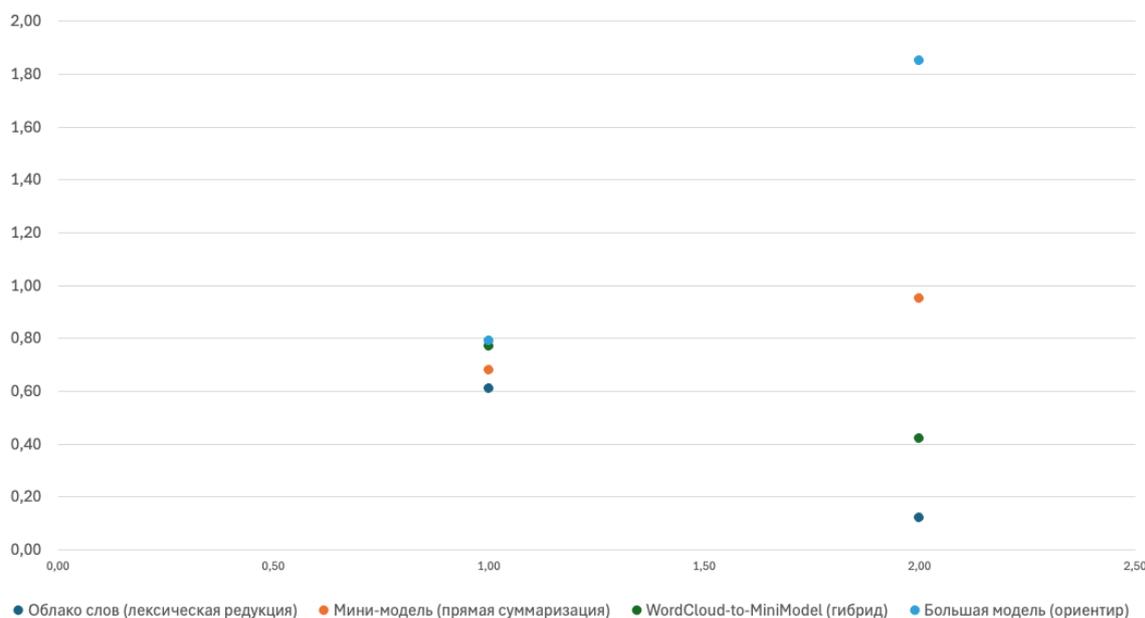


Рисунок 2 – Соотношение Library Relevance Score и времени инференса для различных методов суммаризации

Figure 2 – Comparison of Library Relevance Score and inference time across different summarization methods

Количественные результаты экспериментов по показателям качества и вычислительным затратам приведены в Таблице 1.

Таблица 1 – Количественные результаты экспериментов по показателям качества и вычислительным затратам¹
Table 1 – Quantitative results of the experiments: quality metrics and computational performance

Подход	ROUGE-1	ROUGE-2	BERTScore	LRS	Время инференса, сек	Относительное ускорение
Облако слов (лексическая редукция)	0,28	0,11	0,79	0,61	0,12	ˆ 8,3
Мини-модель (прямая суммаризация)	0,41	0,19	0,86	0,68	0,95	ˆ 1,0
WordCloud-to- MiniModel (гибрид)	0,46	0,23	0,88	0,77	0,42	ˆ 2,3
Большая модель (ориентир)	0,49	0,26	0,90	0,79	1,85	ˆ 0,5

В целом результаты показывают, что гибридный метод выигрывает в сценариях, где приоритетами являются предсказуемость вычислительных затрат, устойчивость к разнородным текстам и сохранение библиотечно значимой информации. В отдельных случаях он может уступать крупным моделям по формальным показателям семантического сходства, однако данное снижение оказывается умеренным и компенсируется повышенной прикладной полезностью и доступностью метода для практического внедрения в библиотечных информационных системах.

Обсуждение

Результаты проведенного исследования показывают, что предложенный гибридный подход семантической редукции ориентирован не на абстрактное повышение формальных показателей, а на решение практических задач, с которыми сталкиваются библиотечные информационные системы в условиях ограниченных ресурсов. Возможность использовать компактные нейросетевые модели в сочетании с лексической редукцией делает метод применимым в малобюджетных библиотечных сценариях, где отсутствует доступ к специализированным вычислительным ускорителям и облачным сервисам. Это особенно важно для региональных и университетских библиотек, для которых характерна эксплуатация локальной инфраструктуры и необходимость обеспечивать стабильную работу сервисов при фиксированном вычислительном бюджете.

С точки зрения интеграции в существующие библиотечные информационные системы предложенный подход не требует радикальной перестройки архитектуры. Лексический этап может быть встроен в уже используемые механизмы индексирования и предварительной обработки текстов, а нейросетевая суммаризация реализована как отдельный сервис, работающий поверх существующих каталогов и хранилищ данных. Такая модульность позволяет поэтапно внедрять метод, адаптируя его под конкретные условия эксплуатации и масштаб фонда, а также снижает риски, связанные с обновлением программного обеспечения и сопровождением системы.

¹ Примечание: значения приведены усреднённо по корпусу; абсолютные значения могут варьироваться в зависимости от параметров моделей и характеристик текстов.

Влияние гибридного подхода на работу пользователей и библиотечных специалистов проявляется прежде всего в повышении качества навигации по коллекциям. Более компактные и содержательные сводки облегчают первичное ознакомление с документами, сокращают время поиска релевантных источников и поддерживают тематическую ориентацию без необходимости обращения к полным текстам. Для библиотекарей и сотрудников, занимающихся каталогизацией и справочно-библиографическим обслуживанием, такой инструмент может служить дополнительным средством анализа и отбора документов, снижая нагрузку на ручную обработку и повышая эффективность работы с растущими массивами текстовой информации.

Ограничения и направления дальнейших исследований

Несмотря на полученные результаты, исследование имеет ряд ограничений, которые необходимо учитывать при интерпретации выводов. Во-первых, используемый корпус данных, хотя и отражает типовые тексты библиотечных информационных систем, остается ограниченным по объему и составу. Он не охватывает всего разнообразия жанров, дисциплин и форматов, представленных в крупных национальных и специализированных фондах, что может влиять на обобщаемость результатов и чувствительность метода к редким или слабо представленным тематическим областям.

Во-вторых, предложенная метрика Library Relevance Score, будучи ориентированной на библиотечный контекст, опирается на заранее заданные представления о доменно значимых терминах и структурных признаках текста. Это делает ее интерпретацию зависимой от выбранных словарей и правил агрегации компонентов, а также от специфики конкретной библиотечной среды. В текущем виде метрика не претендует на универсальность и требует дальнейшей отработки на расширенных корпусах и с участием большего числа экспертов.

Перспективы дальнейших исследований связаны с развитием как методологической, так и прикладной составляющей предложенного подхода. Представляется целесообразным расширение и адаптация словарей библиотечно значимых терминов с учетом дисциплинарной специфики и типов фондов, а также исследование поведения гибридного метода в различных библиотечных контекстах – от научных и учебных библиотек до публичных и специализированных коллекций. Отдельного внимания заслуживает изучение пользовательских сценариев применения суммаризации, включая оценку удобства и эффективности взаимодействия с системой со стороны читателей и библиотекарей. Эти направления могут послужить основой для дальнейшего уточнения архитектуры метода и повышения его практической ценности.

Заключение

В данной работе предложен гибридный метод семантической редукции текстов для библиотечных информационных систем, основанный на последовательном применении лексической редукции с использованием облаков слов и нейросетевой суммаризации компактными моделями. Такой подход позволяет объединить интерпретируемость и низкие вычислительные затраты классических лексических методов с преимуществами нейросетевых моделей, обеспечивающих более глубокое семантическое обобщение текста. В рамках исследования также была введена метрика Library Relevance Score, ориентированная на оценку качества суммаризации с учетом специфики библиотечной навигации и задач каталогизации.

Научная новизна работы заключается в рассмотрении облака слов не как вспомогательного визуального инструмента, а как функционального этапа предварительной смысловой редукции, подготавливающего текст к дальнейшей

нейросетевой обработке. В отличие от существующих исследований, где лексические и нейросетевые методы, как правило, противопоставляются друг другу, предложен целостный гибридный подход, адаптированный к условиям реальной эксплуатации библиотечных систем. Дополнительный вклад вносит разработка доменно-ориентированной метрики, позволяющей приблизить автоматическую оценку качества суммаризации к практическим потребностям библиотечных специалистов и пользователей.

Практическая значимость полученных результатов обусловлена их ориентированностью на библиотеки, функционирующие в условиях ограниченных вычислительных ресурсов. Предложенная архитектура может быть реализована на локальной инфраструктуре без использования специализированных ускорителей и облачных сервисов, что делает ее применимой для широкого круга библиотечных учреждений. Использование гибридного подхода способствует повышению качества навигации по фондам, сокращению времени первичного ознакомления с документами и снижению нагрузки на сотрудников, занимающихся справочно-библиографическим обслуживанием.

Таким образом, представленное исследование демонстрирует, что интеллектуальная семантическая обработка текстов в библиотечных информационных системах может быть реализована не только за счет наращивания вычислительных мощностей, но и за счет продуманной архитектуры методов. Это открывает возможности для дальнейшего развития ресурсосберегающих инструментов анализа и суммаризации, ориентированных на реальные потребности библиотечной практики.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Lyon L. The Informatics Transform: Re-Engineering Libraries for the Data Decade. *International Journal of Digital Curation*. 2012;7(1):126–138. <https://doi.org/10.2218/ijdc.v7i1.220>
2. Roy P. Big data analytics in university libraries on today's librarianship decision-making: A disruptive innovation perspective. *IFLA Journal*. 2025;51(8). <https://doi.org/10.1177/03400352251318753>
3. Mridha M.F., Lima A.A., Nur K., et al. A Survey of Automatic Text Summarization: Progress, Process and Challenges. *IEEE Access*. 2021;9:156043–156070. <https://doi.org/10.1109/ACCESS.2021.3129786>
4. Arnaboldi V., Cho J., Sternberg P.W. Wormicloud: A new text summarization tool based on word clouds to explore the *C. elegans* literature. *Database*. 2021;2021. <https://doi.org/10.1093/database/baab015>
5. Strubell E., Ganesh A., McCallum A. Energy and Policy Considerations for Deep Learning in NLP. In: *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019: Volume 1: Long Papers, 28 July – 02 August 2019, Florence, Italy*. Association for Computational Linguistics; 2019. P. 3645–3650. <https://doi.org/10.18653/v1/P19-1355>
6. Treviso M., Lee J.-U., Ji T., et al. Efficient Methods for Natural Language Processing: A Survey. *Transactions of the Association for Computational Linguistics*. 2023;11:826–860. https://doi.org/10.1162/tacl_a_00577
7. Syed A.A., Gaol F.L., Matsuo T. A Survey of the State-of-the-Art Models in Neural Abstractive Text Summarization. *IEEE Access*. 2021;9:13248–13265. <https://doi.org/10.1109/ACCESS.2021.3052783>
8. Goodwin T., Savery M., Demner-Fushman D. Flight of the PEGASUS? Comparing Transformers on Few-Shot and Zero-Shot Multi-document Abstractive Summarization.

- In: *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, 08–13 December 2020, Barcelona, Spain (Online)*. International Committee on Computational Linguistics; 2020. P. 5640–5646. <https://doi.org/10.18653/v1/2020.coling-main.494>
9. Li J. A comparative study of keyword extraction algorithms for English texts. *Journal of Intelligent Systems*. 2021;30:808–815. <https://doi.org/10.1515/jisys-2021-0040>
 10. Skeppstedt M., Ahltop M., Kucher K., Lindström M. From word clouds to Word Rain: Revisiting the classic word cloud to visualize climate change texts. *Information Visualization*. 2024;23(2). <https://doi.org/10.1177/14738716241236188>
 11. Bhandari M., Gour P.N., Ashfaq A., Liu P., Neubig G. Re-evaluating Evaluation in Text Summarization. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, 16–20 November 2020, Online*. Association for Computational Linguistics; 2020. P. 9347–9359. <https://doi.org/10.18653/v1/2020.emnlp-main.751>
 12. Hobson S.P., Dorr B.J., Monz Ch., Schwartz R. Task-based evaluation of text summarization using relevance prediction. *Information Processing & Management*. 2007;43(6):1482–1499. <https://doi.org/10.1016/j.ipm.2007.01.002>
 13. Ushio A., Liberatore F., Camacho-Collados J. Back to the Basics: A Quantitative Analysis of Statistical and Graph-Based Term Weighting Schemes for Keyword Extraction. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, 07–11 November 2021, Virtual Event / Punta Cana, Dominican Republic*. Association for Computational Linguistics; 2021. P. 8089–8103. <https://doi.org/10.18653/v1/2021.emnlp-main.638>
 14. Hearst M.A., Pedersen E., Patil L., et al. An Evaluation of Semantically Grouped Word Cloud Designs. *IEEE Transactions on Visualization and Computer Graphics*. 2020;26(9):2748–2761. <https://doi.org/10.1109/TVCG.2019.2904683>
 15. Dice D., Kogan A. *Optimizing Inference Performance of Transformers on CPUs*. arXiv. URL: <https://arxiv.org/abs/2102.06621> [Accessed 19th January 2026].
 16. Xu Y., Xu R., Iter D., et al. InheritSumm: A General, Versatile and Compact Summarizer by Distilling from GPT. In: *Findings of the Association for Computational Linguistics: EMNLP 2023, 06–10 December 2023, Singapore*. Association for Computational Linguistics; 2023. P. 13879–13892. <https://doi.org/10.18653/v1/2023.findings-emnlp.927>
 17. Lee J.-U., Puerto H., van Aken B. *Surveying (Dis)Parities and Concerns of Compute Hungry NLP Research*. arXiv. URL: <https://arxiv.org/abs/2306.16900> [Accessed 19th January 2026].
 18. Desai Sh., Xu J., Durrett G. Compressive Summarization with Plausibility and Salience Modeling. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, 16–20 November 2020, Online*. Association for Computational Linguistics; 2020. P. 6259–6274. <https://doi.org/10.18653/v1/2020.emnlp-main.507>
 19. Mei A., Kabir A., Bapat R., et al. Learning to Prioritize: Precision-Driven Sentence Filtering for Long Text Summarization. In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference, LREC 2022, 20–25 June 2022, Marseille, France*. European Language Resources Association; 2022. P. 313–318.
 20. Liang X., Li J., Wu Sh., et al. An Efficient Coarse-to-Fine Facet-Aware Unsupervised Summarization Framework Based on Semantic Blocks. In: *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, 12–17 October 2022, Gyeongju, Republic of Korea*. International Committee on Computational Linguistics; 2022. P. 6415–6425.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Рзынкин Илья Сергеевич, ведущий инженер-программист Научной библиотеки Сибирского федерального университета, Красноярск, Российская Федерация.

e-mail: i-rzyankin@yandex.ru

ORCID: [0009-0002-1702-591X](https://orcid.org/0009-0002-1702-591X)

Ilya S. Rzyankin, Lead Software Engineer, Scientific Library of Siberian Federal University, Krasnoyarsk, the Russian Federation.

Носков Михаил Валерианович, доктор физико-математических наук, профессор кафедры прикладной математики и анализа данных Сибирского федерального университета, Красноярск, Российская Федерация.

e-mail: mnoskov@sfu-kras.ru

ORCID: [0000-0001-8966-3633](https://orcid.org/0000-0001-8966-3633)

Mikhail V. Noskov, Doctor of Physico-mathematical Sciences, Professor at the Department of Applied Mathematics and Data Analysis of Siberian Federal University, Krasnoyarsk, the Russian Federation.

Статья поступила в редакцию 11.02.2026; одобрена после рецензирования 20.03.2026; принята к публикации 25.03.2026.

The article was submitted 11.02.2026; approved after reviewing 20.03.2026; accepted for publication 25.03.2026.