DOI: 10.6060/snt.20258101.0009
УДК 004.651.4

МЕТОДОЛОГИЯ СОЗДАНИЯ МНОГОСВЯЗНЫХ СТРУКТУР ДАННЫХ С ПРИМЕНЕНИЕМ LLM В РАБОЧИХ ПРОЕКТАХ

Зимнуров М.Ф., Астраханцева И.А.

Зимнуров Марат Фаридович (ORCID 0000-0002-3115-0912),
Астраханцева Ирина Александровна (ORCID 0000-0003-2841-8639)
Ивановский государственный химико-технологический университет,
г. Иваново, Россия. 153000, Ивановская область, г. Иваново, пр. Шереметевский, д. 7.
E-mail: zimtir@mail.ru, i.astrakhantseva@mail.ru

В данной работе предлагается методология создания многосвязных структур данных на основе использования больших языковых моделей (LLMs) для решения задач поиска, анализа и систематизации текстовых данных. Основное внимание уделено разработке динамической информационной среды, позволяющей преобразовывать разрозненные текстовые данные в многосвязные структуры, такие как графы знаний и сети. Методология обеспечивает удобный доступ к данным и улучшенную визуализацию, что способствует повышению эффективности взаимодействия пользователей с информацией. Новизна подхода заключается в интеграции LLMs для создания адаптивных информационных систем в различных рабочих сценариях, включая разработку ПО и анализ данных.

Ключевые слова: LLM, NLP, анализ текстовых данных, инновации в области обработки текстовой информации, векторное представление данных, эффективность использования языковых моделей в рабочих проектах, графовые структуры, графы знаний

USAGE OF LARGE LANGUAGE MODELS FOR BUILDING MULTI-VECTORED AND MULTI-LINKED DATA MODEL OF WORK PROCESS

Zimnurov M.F., Astrakhantseva I.A.

Zimnurov Marat Faridovich (ORCID 0000-0002-3115-0912),
Astrakhantseva Irina Aleksandrovna (ORCID 0000-0003-2841-8639)
Ivanovo State University of Chemical Technology,
Ivanovo, Russia. 153000, Ivanovo region, Ivanovo, Sheremetevsky Ave., 7.
E-mail: zimtir@mail.ru, i.astrakhantseva@mail.ru

This paper presents a methodology for creating multi-linked data structures using large language models (LLMs) to address the challenges of information retrieval, analysis, and organization of textual data. The focus is on the development of a dynamic information environment that transforms scattered textual data into interconnected structures, such as knowledge graphs and linked networks. The proposed methodology ensures convenient data accessibility and improved visualization, enhancing user interaction with complex information. The novelty of the approach lies in the integration of LLMs for building adaptive information systems applicable to various work scenarios, including software development and data analysis.

Keywords: LLM, NLP, text data analysis, innovations in text processing, vector data representation, efficiency of language models in work projects, graph structures, knowledge graphs

ВВЕДЕНИЕ

В современном мире обработки и анализа данных, взаимодействие с текстовой информацией становится все более значимым. Natural Language Processing (NLP) играет ключевую роль, позволяя эффективно работать с большими объемами текстовых данных и извлекать из них полезную информацию. Одним из мощных инструментов в области NLP является использование больших языковых моделей (Large Language Models, LLMs), которые способны не только анализировать и генерировать текст, но и служить основой для построения сложных структур данных [1].

Объектом исследования является процесс построения многовекторной и многосвязной модели данных на основе текстовой информации, полученной из различных источников. Предмет исследования - применение больших языковых моделей для анализа и структурирования текстовых данных в виде сложных графов, а также разработка API для интерактивного взаимодействия с этими данными. Целью работы является разработка и обоснование подхода к созданию многовекторной и многосвязной модели данных, использующей возможности LLM для обработки текстовой информации и её последующего представления в виде сложных структур, таких как графы. Задачи, которые были сформированы для достижения цели:

- анализ существующих методов применения LLM для обработки текстовых данных;
- создание API для получения и обработки пользовательских запросов;
- формирование структуры данных на основе полученной текстовой информации и её визуализация;
- разработка валидационных механизмов для проверки корректности структуры данных.

Оригинальность подхода заключается в интеграции возможностей LLM для формирования гибких и сложных структур данных на основе текстовой информации, что позволяет улучшить процессы обработки данных и взаимодействия с ними. Разработанная модель может быть использована для создания интерактивных систем анализа и визуализации данных, что открывает новые перспективы в области автоматизации обработки текстовой информации и разработки интеллектуальных систем.

Важность данного исследования обусловлена потребностью в новых методах и инструментах для обработки и представления текстовых данных в условиях их растущего объема и сложности. Полученные результаты могут найти применение в различных областях, включая информационные технологии, бизнес-аналитику, образовательные платформы и другие сферы, где необходима эффективная работа с текстовой информацией.

МАТЕРИАЛЫ И МЕТОДЫ

В данном исследовании рассматриваются методы обработки и анализа сырых данных с применением больших языковых моделей (LLM) для создания многоузловой структуры данных. Под сырыми данными понимаются неструктурированные текстовые материалы, которые возникают в рабочих проектах и коммуникациях.

Это могут быть переписки в мессенджерах (Slack, Telegram), комментарии к задачам в системах управления проектами, обсуждения в корпоративных Wiki и других платформах. Сырые данные характерны тем, что они не упорядочены, не классифицированы и содержат информацию в различном формате и контексте, часто с множеством ненормализованных элементов, таких как неформальные упоминания, сленг или сокращения.

Для эффективного использования сырых данных, таких как переписки, комментарии и другие текстовые источники, сначала их необходимо преобразовать в машиночитаемый вид.

Этот процесс начинается с токенизации, которая разбивает текст на отдельные единицы (токены) – слова. В ходе этого процесса важно выделить ключевые параметры для каждой записи, а именно, уникальный идентификатор сообщения, сам токен, дату отправки сообщения, источник данных и связи между токенами, которые представляют контекст их появления. Такая модель позволяет представить данные в виде графовой структуры, где узлы - это токены, а связи между ними указывают на контекст их использования. Это создает основу для построения динамической модели текста, где сохраняются контекстуальные зависимости между словами.

Ключевым компонентом такой архитектуры является сервис для получения сообщений из различных потенциальных источников, который реализуется с помощью Message Broker. Message Broker выступает в качестве промежуточного уровня, отвечающего за прием и передачу сообщений между различными источниками данных и системой хранения. Он обеспечивает стандартизированный способ взаимодействия с такими системами, как мессенджеры (например, Slack и Telegram), системы управления проектами и корпоративные Wiki. Каждое полученное сообщение помещается в первую таблицу базы данных, предназначенную для хранения исходных текстов. Это позволяет гарантировать, что все сообщения, поступающие в систему, сохраняются в их оригинальном формате и с соответствующими метаданными, такими как дата и время получения, источник сообщения и, при наличии, информация о пользователе [2].

Таким образом, использование Message Broker обеспечивает не только унифицированный интерфейс для взаимодействия с разными источниками, но и позволяет обрабатывать входящие сообщения в реальном времени, минимизируя риски потери данных. Регулярная и надежная передача сообщений между источниками и системой

хранения гарантирует высокую степень адаптивности системы, что позволяет ей легко масштабироваться и интегрировать новые источники данных. Важным нюансом является определение от кого и кому было адресовано сообщение и для того чтобы определить, кому направлено сообщение и кто его отправил, можно использовать LLM для анализа упоминаний и идентификации пользователей в неструктурированных данных. Так как упоминания могут быть представлены в виде ников, фамилий, должностей, ролей или социальных и гендерных признаков, то нужно распознавать эти паттерны и нормализовать данные, превращая неформальные упоминания в стандартизированные идентификаторы [3]. Важно также учитывать, что некоторые сообщения могут быть адресованы не конкретному человеку, а группе, и в этом случае система должна уметь распознавать такие паттерны. В случае групповых сообщений возможна классификация таких сообщений на уровне групповой принадлежности.

Если все пользователи по умолчанию считаются анонимными, тогда их можно деанонимизировать, анализируя уникальные паттерны общения, такие как стиль письма, частота использования определенных выражений или особенности взаимодействия. Однако в случае появления сообщений, имитирующих чужие стили общения (симулякр), система может столкнуться с трудностями, так как возникнет путаница в идентификации отправителей. В таких ситуациях целесообразно относить подобные сообщения к категории "от толпы" или "к толпе", чтобы избежать сбоев в работе системы и сохранить целостность анализа [4-6].

Определение морфологических признаков токенов может быть успешно реализовано с помощью LLM. Однако это не всегда гарантирует корректное установление связей между несколькими сообщениями одного и того же пользователя. Для решения проблемы можно учитывать временные задержки между сообщениями. Если сообщения следуют друг за другом с минимальным интервалом (например, в пределах одной минуты), есть высокая вероятность, что они принадлежат одному и тому же агенту и должны рассматриваться как связанные.

После токенизации важным этапом становится создание связей между токенами. Эти связи могут существовать как внутри одного сообщения, так и между сообщениями, исходя из их контекста. Связи между токенами можно устанавливать внутри одного сообщения, где все токены, участвующие в этом сообщении, автоматически получают связь. Каждая такая связь внутри одного сообщения будет отражать их непосредственную близость и контекстуальную зависимость в пределах одного высказывания [7].

Помимо установления связей между токенами на основе их прямого появления в одном сообщении или в различных сообщениях, важным аспектом является обработка различных форм одного и того же слова. Используя возможности LLM, можно определить, что разные морфологические формы одного и того же слова фактически представляют один и тот же токен, но в разных грамматических формах.

В таких случаях связи между разными формами одного слова также могут быть учтены в модели. Вес связи между токенами, представляющими разные формы одного и того же слова, может быть установлен на уровне 0.5, так как, хотя они и отражают одно и то же понятие, их использование в тексте может различаться в зависимости от контекста. Это помогает учесть не только прямые совпадения, но и разнообразные варианты морфологических форм, что значительно расширяет возможности анализа и обеспечивает большую точность при обработке естественного языка [8, 9]. Далее для повышения уровня обработки сообщений и создания контекстов вводится сервис, использующий LLM в качестве компаратора. Данный сервис работает по принципу сопоставления каждого нового сообщения с уже существующими в базе данных, что позволяет выявить общие контексты. Если LLM определяет, что новое сообщение связано с уже имеющимся сообщением общим контекстом, то оно ищет этот контекст среди существующих. В случае его нахождения создается новая запись в таблице связей с весом 1, что указывает на полное соответствие между сообщениями.

Если сообщения частично связаны общим контекстом, система ищет подходящий контекст среди существующих записей и создает запись в таблице связей с весом 0,5. Это значение отражает то, что сообщения имеют определённые взаимосвязи, но не полностью соответствуют друг другу по смыслу. Такой подход позволяет учитывать более тонкие градации взаимосвязей и контекстов, что особенно важно при работе с неструктурированными данными, где значение слов может меняться в зависимости от контекста.

В случаях, когда сообщения не имеют общих контекстов, итерация пропускается, система переходит к следующему сообщению. Это минимизирует затраты ресурсов и время на обработку и сосредотачивает внимание на значимых взаимосвязях.

При добавлении нового сообщения в базу данных алгоритм обработки и анализа сообщений необходимо воспроизвести для этого сообщения. Это гарантирует, что каждая новая запись будет учитывать контексты, уже существующие в системе и, в случае необходимости, будет создан новый контекст, что позволит постоянно обновлять и улучшать структуру данных. Такой подход обеспечивает высокую степень точности и полноты анализа, позволяет системе эффективно работать с большими объемами неструктурированных текстовых данных.

Для представления связей между токенами используется структура, которая включает идентификаторы сообщений и токенов, а также вес связи. Это записывается в формате: ID сообщения - ID токена - ID соседнего токена - вес связи, где вес отражает силу связи между токенами в зависимости от контекста и частоты их совместного появления [10].

В процессе обработки и анализа текстовых данных с помощью токенов и LLM важно учитывать, как данные структурируются и управляются на разных уровнях. Помимо токенизации и построения связей между токенами, существенную роль играет то, как исходные данные хранятся и обрабатываются для обеспечения полноты и качества анализа. Стратегии хранения исходных сообщений и оптимизации таблиц токенов могут существенно повлиять на эффективность работы системы, особенно при работе с большими объемами текстов и множественными контекстами.

Хранение исходных сообщений и их использование в качестве родительских для таблиц токенов может иметь значительный смысл с точки зрения структурирования данных. Сохранение исходного текста позволяет не только вернуться к первоначальному контексту при необходимости, но и помогает организовать агрегацию данных на более высоком уровне. Это полезно, например, для категоризации контекста, поскольку одно и то же слово или токен может приобретать разные значения в зависимости от того, где и как оно используется. Хранение исходных сообщений позволяет отслеживать, как конкретные токены связаны с темами, обсуждениями, проектами или авторами сообщений. Это может быть особенно важно для анализа больших объемов текстовых данных, где агрегация контекстов становится ключевым инструментом для извлечения смысла из данных. Таким образом, хранение исходных сообщений позволяет не только сохранять полноту информации, но и улучшить возможности для классификации, тематического анализа и создания более сложных и гибких структур данных. Что касается необходимости оптимизации таблиц токенов, это также представляет собой важный шаг в управлении данными [11]. Со временем в таблице могут появляться дублирующиеся или схожие токены, особенно если речь идет о разных морфологических формах слов, упоминаниях, которые могут быть нормализованы, или ошибочных токенах, возникших в результате неверной токенизации. Оптимизация позволяет избежать избыточности и повысить качество данных. Это включает нормализацию токенов, объединение разных форм одного и того же слова, удаление шумовых данных и дублированных записей. Регулярная очистка таблицы токенов позволяет поддерживать систему в актуальном состоянии. При это минимизируются ошибки и улучшается эффективность работы модели. В результате таблица токенов становится более точной и информативной, что напрямую влияет на качество последующего анализа и использования данных.

Таким образом, стоит определиться с вариантом хранения данных. Хранение токенов в базе данных организовано с использованием трех основных таблиц: таблицы сообщений, таблицы контекстов и таблицы для весов и связей сообщений и контекстов. Такая структура позволяет эффективно управлять данными, обеспечивает хранение исходных сообщений, построение иерархии связей между токенами, что упрощает анализ и позволяет динамически расширять модель.

Таблица сообщений содержит все исходные сообщения в необработанном виде. Каждое сообщение получает уникальный идентификатор (ID), который связывает его с остальными элементами базы данных. Помимо самого текста, таблица сообщений может содержать дополнительные поля, такие как дата и время отправки сообщения, источник (мессенджер, система управления проектами, корпоративная Wiki и т.д.), а также информацию о пользователе, если она доступна (например, ID отправителя или группы пользователей). Хранение сообщений в отдельной таблице позволяет в любой момент вернуться к исходным данным, что важно для уточнения контекста и верификации связей, построенных на уровне токенов.

В таблице контекстов хранятся токены, извлеченные из сообщений, и их контексты. Каждый токен связан с уникальным ID сообщения, из которого он был извлечен. Контексты могут включать как непосредственные соседние токены (слова), так и более широкие фрагменты текста, окружающие токен, которые помогают уточнить его смысл в конкретном случае.

Например, если слово встречается в нескольких предложениях, контексты могут помочь различить его различные значения в зависимости от ситуации. Эта таблица является ключевой для анализа, так как в ней хранятся данные, необходимые для дальнейшего построения графов связей между токенами. Нормализация токенов и приведение их к единой форме (например, с учетом разных морфологических вариантов) также реализуются на этом уровне.

Таблица весов и связей сообщений и контекстов служит для хранения всех связей между токенами и сообщений, а также весов этих связей. Каждая запись в таблице содержит ID сообщения, ID токена и ID другого токена, с которым он связан. Помимо этого, указывается вес связи, который показывает степень близости или контекстуальной зависимости токенов. Например, связи между токенами внутри одного сообщения будут иметь вес 1, тогда как связи между токенами в разных сообщениях или разных контекстах могут иметь меньший вес (например, 0.5).

Эти данные используются для построения графов связей, которые визуализируются и анализируются с помощью модели.

Для обеспечения взаимодействия с моделью и визуализации данных необходимо построить WEB-API, которое будет отвечать за получение всех токенов и их весов. API дает возможность получать структурированные данные, представляющие собой граф, где токены - это узлы, а связи между ними - это рёбра с определённым весом. На основе этих данных можно визуализировать граф, который показывает, как токены взаимосвязаны между собой. Это позволяет пользователям не только видеть связи между словами или выражениями, но и оценивать силу этих связей, анализируя веса, которые были присвоены во время обработки данных. Таким образом, графовая структура дает наглядное представление об отношениях внутри текста и о том, как различные элементы взаимодействуют друг с другом [12-15].

Кроме того, важным аспектом работы с токенами является постоянная тренировка LLM на основе обновляемой таблицы токенов. Каждый раз, когда добавляется новый токен или обновляются существующие данные, модель должна адаптироваться к этим изменениям, чтобы сохранять актуальность анализа и предсказаний. Постепенное добавление новых токенов позволяет модели улучшать свою способность распознавать и связывать новые термины с уже известными, а также корректировать существующие связи между токенами.

Такой подход гарантирует, что модель всегда учитывает актуальные данные и более точно предсказывает возможные связи между токенами в будущем.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ДИСКУССИИ

Важным нюансом является определение от кого и кому направлены сообщения в рамках анализа неструктурированных данных. Одной из ключевых задач при работе с комментариями, перепиской и другими текстовыми источниками является установление отправителя и получателя, особенно в тех случаях, когда контекст общения не очевиден. Это важно не только для точной интерпретации данных, но и для создания корректных связей между токенами в графовой модели. В данном исследовании предложена методология, которая позволяет определить авторов и адресатов сообщений, даже если они не указаны явно, например, через упоминания ников, имен, должностей или других характеристик.

Для решения этой задачи применяются большие языковые модели (LLM), которые анализируют текст с учетом множества факторов: морфологические признаки, паттерны общения, контекстные связи и временные интервалы между сообщениями. Это позволяет с высокой степенью точности определить, кто является отправителем (FROM) и получателем (TO) сообщения, даже в тех случаях, когда общение ведется в групповых чатах, и непосредственные адресаты могут не быть явно упомянуты. Также, если сообщение адресовано группе пользователей или не имеет четкого получателя, оно может быть классифицировано как сообщение «к толпе» или «от толпы».

Один из результатов исследования показывает, что анализ паттернов общения, таких как временная последовательность сообщений, использование местоимений и гендерных маркеров, помогает улучшить качество идентификации участников диалога. Для этого используется модель предсказания, которая анализирует задержку между сообщениями. Если сообщения следуют друг за другом с малым временным интервалом, они могут быть частью одной беседы, что упрощает определение их взаимосвязи. Это особенно полезно в переписках, где упоминания отправителей и получателей могут отсутствовать.

Еще одним важным результатом является создание графовой структуры, которая позволяет учитывать не только прямые связи между сообщениями, но и их контекстуальные зависимости. Исследование показало, что LLM эффективно идентифицируют общие контексты между сооб-

щениями, даже если они частично различаются по форме или стилю. Например, морфологические различия между формами одного и того же слова могут быть учтены в модели, что расширяет возможности анализа. Контекстные связи между сообщениями играют ключевую роль в построении многоузловой структуры данных. Вес связи в графе может варьироваться в зависимости от степени взаимосвязанности сообщений. Полные совпадения оцениваются весом 1, частичные - 0.5. Такой подход позволяет учитывать не только явные совпадения, но и тонкие смысловые связи, что делает модель более гибкой и точной при анализе больших объемов неструктурированных данных.

Результаты исследования показывают, что такая методология позволяет эффективно работать с неструктурированными данными, включая сложные переписки и комментарии.

Применение LLM для определения участников общения и установления контекстуальных связей между сообщениями открывает возможности для построения более точных и многоуровневых моделей взаимодействия.

Для визуализации архитектуры и взаимодействия компонентов системы применены диаграммы, выполненные в нотации C4 [13-19]. Нотация C4 используется для представления архитектуры программного обеспечения на разных уровнях абстракции, включая контекст, контейнеры, компоненты и код.

Это позволяет детально отобразить, как взаимодействуют различные части системы, начиная от общего контекста и заканчивая конкретными техническими деталями (рис. 1-3).

Рис. 1. Диаграмма в нотации C4 на уровне контекста
Fig. 1. C4 notation diagram at the context level

Рис. 2. Диаграмма в нотации C4 на уровне контейнеров
Fig. 2. C4 notation diagram at the container level

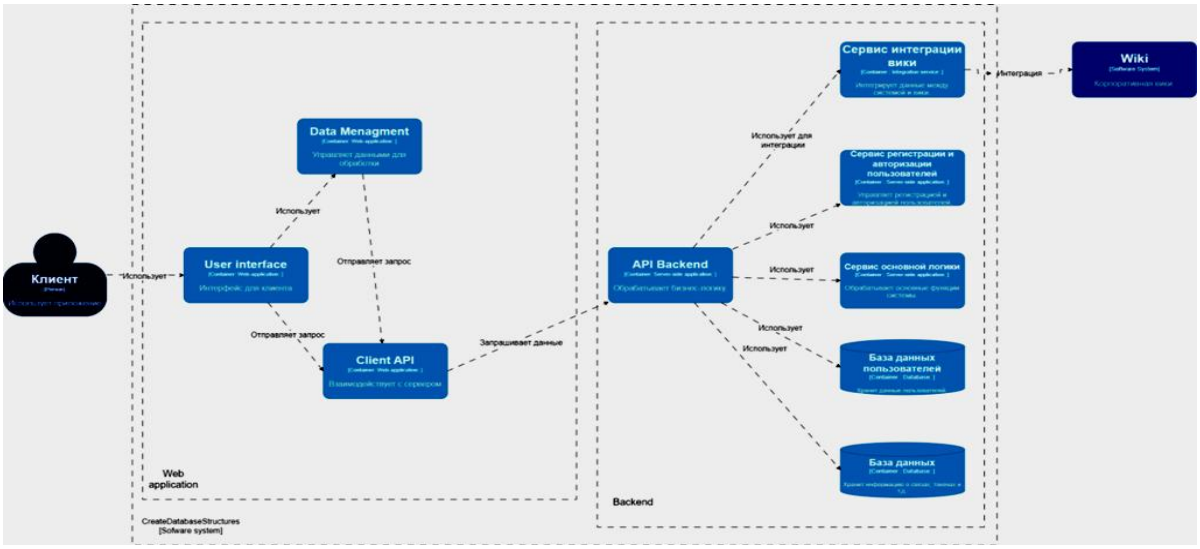

Рис. 3. Диаграмма в нотации C4 на уровне компонентов
Fig. 3. C4 notation diagram at the component level

Однако важной задачей остается оптимизация процесса создания новых контекстов. В некоторых случаях большое количество контекстов может усложнить структуру данных, что требует дополнительных механизмов агрегации и категоризации. Введение системы динамического обновления контекстов на основе новых данных также может стать следующим шагом в развитии методологии.

Таким образом, предложенная система анализа текстовых данных с применением LLM показывает высокую эффективность, но требует дальнейших исследований и оптимизации для работы с большими объемами данных и более сложными структурами взаимодействия.

Авторы заявляют об отсутствии конфликта интересов, требующего раскрытия в данной статье.

The authors declare the absence a conflict of interest warranting disclosure in this article.

ЛИТЕРАТУРА

1. **Прохоренко А.В., Сергеев С.И., Куркин Я.И.** Применение и дообучение современных больших языковых моделей (LLM). Экспериментальные и теоретические исследования в современной науке: сборник статей по материалам CI международной научно-практической конференции, Новосибирск, 27 мая 2024 года; под ред. ООО "Сибирская академическая книга". Новосибирск: ООО "Сибирская академическая книга", 2024. С. 70–81.
2. FastText. Википедия: свободная энциклопедия. https:// en. wikipedia.org/wiki/FastText.
3. **Краснов Ф.В.** Исследование влияния текстового представления товара посредством моделей с использованием искусственных нейронных сетей глубокого обучения на релевантность поиска товаров на электронной торговой интернет-площадке. *International Journal of Open Information Technologies.* 2023. № 11. https:// cyberleninka. ru/article n/issledovanie-vliyaniya.
4. Актуальное членение предложения и текста. 4Brain: блог об обучении и саморазвитии. https://4brain.ru/ blog/ актуальное-членение
5. Лемма о разрастании. Википедия: свободная энциклопедия. https://ru.wikipedia.org/wiki/Лемма_о_разрастании
6. Темы и ремы. Quest-book: форум. https://quest-book.ru/ forum/topic/1138.
7. Абстрактное синтаксическое дерево. Википедия: свободная энциклопедия. https://ru.wikipedia. org/wiki/ Абстрактное_синтаксическое_дерево.
8. Нечеткая логика. Википедия: свободная энциклопедия. https://ru.wikipedia.org/wiki/Нечёткая_логика.
9. **Грибова В.В., Переволоцкий В.С.** Разработка графов знаний на основе больших языковых моделей для поддержки принятия решений в медицине. *Программная инженерия.* 2024. Т. 15, № 6. С. 308–321. УДК 004.89. ISSN 2220-3397.
10. **Маркин Е.И., Зупарова В.В.** Интеграция языковых моделей с базами знаний и внешними источниками данных. *XXI век: итоги прошлого и проблемы настоящего плюс.* 2024. Т. 13, № 2 (66). С. 25–31. УДК 004.89. ISSN 2221-951X. Учредители: Пензенский государственный технологический университет. Дата поступления в редакцию: 17.05.2024. Дата принятия к печати: 17.06.2024.
11. **Миронов А.И., Мунерман В.И.** Создание частичного индексирования таблицы для оптимизации поисковых запросов. *Современные информационные технологии и ИТ-образование.* 2022. Т. 18, № 3. С. 558–565. УДК 004.22:004.27. ISSN 2411-1473. Учредители: Фонд содействия развитию интернет-медиа, ИТ-образования, человеческого потенциала Лига интернет-медиа.

REFERENECES

1. **Prokhorenko A.V., Sergeev S.I., Kurkin Y.I.** Application and Fine-Tuning of Modern Large Language Models (LLM). Experimental and Theoretical Research in Modern Science: Proceedings of the CI International Scientific and Practical Conference, Novosibirsk, May 27, 2024; edited by "Siberian Academic Book" LLC. Novosibirsk: "Siberian Academic Book" LLC, 2024. P. 70–81.
2. FastText. Wikipedia: The Free Encyclopedia. https://en. wikipedia.org/wiki/FastText.
3. **Krasnov F.V.** Study of the Influence of Product Text Representation Using Deep Learning Artificial Neural Network Models on Product Search Relevance in an E-commerce Platform. I*nternational Journal of Open Information Technologies.* 2023. N 11. https:// cyberleninka.ru/article n/ issledovanie-vliyaniya.
4. The Actual Division of Sentence and Text. 4Brain: Blog on Learning and Self-Development. https://4brain.ru/blog/ актуальное-членение.
5. Lemma Expansion. Wikipedia: The Free Encyclopedia. https://ru.wikipedia.org/wiki/Лемма_о_разрастании
6. Themes and Remas. Quest-book: Forum. https://quest-book.ru/ forum/topic/1138.
7. Abstract Syntax Tree. Wikipedia: The Free Encyclopedia.: https:// ru.wikipedia.org/wiki/Абстрактное_синтаксическое_дерево
8. Fuzzy Logic. Wikipedia: The Free Encyclopedia. https://ru.wikipedia.org/wiki/Нечёткая_логика.
9. **Gribova V.V., Perevolotsky V.S.** Development of Knowledge Graphs Based on Large Language Models for Decision Support in Medicine. *Software Engineering.* 2024. V. 15, N 6. P. 308–321. UDC 004.89. ISSN 2220-3397.
10. **Markin E.I., Zuparova V.V.** Integration of Language Models with Knowledge Bases and External Data Sources. *XXI Century: Results of the Past and Problems of the Present Plus.* 2024. V. 13, N 2 (66). P. 25–31. UDC 004.89. ISSN 2221-951X. Publisher: Penza State Technological University. Submitted: 17.05.2024. Accepted for publication: 17.06.2024.
11. **Mironov A.I., Munerman V.I.** Creating Partial Indexing of a Table for Optimizing Search Queries. *Modern Information Technologies and IT-Education.* 2022. V. 18, N 3. P. 558–565. UDC 004.22:004.27. ISSN 2411-1473. Publisher: Internet Media Development Fund, IT Education, Human Potential "Internet Media League."
12. **Álvarez P.F., Quevedo S.** Academic Query Assistant: Integrating LLM API into an Academic Assistant Using a Microservices Architecture. *Journal of Computing Science and Engineering.* 2024. V. 18, N 2. P. 125–133. ISSN 1976-4677. eISSN 2093-8020. Publisher: Korean Institute of Information Scientists and Engineers. Published: 30.06.2024.

12. **Álvarez P.F., Quevedo S.** Academic Query Assistant: Integrating LLM API into an Academic Assistant Using a Microservices Architecture. *Journal of Computing Science and Engineering.* 2024. Vol. 18, N 2. P. 125–133. ISSN 1976-4677. eISSN 2093-8020. Publisher: Korean Institute of Information Scientists and Engineers. Published: 30.06.2024.
13. C4 Model for Visualising Software Architecture. https://c4model.com/ (дата обращения: 21.01.2024).
14. **Астраханцева И.А., Котенев Т.Е., Горев С.В. [и др.].** Интеграция методов линейной регрессии и случайного леса в системы интеллектуальной поддержки управления криогенными транспортными системами. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2024. № 3(61). С. 81-90. DOI 10.6060/ivecofin.2024613.692.
15. **Астраханцева И.А., Котенев Т.Е., Горев С.В. [и др.].** Метод градиентного бустинга при прогнозировании управленческих решений в многослойной криогенной системе. *Современные наукоемкие технологии. Региональное приложение.* 2024. № 2(78). С. 50-58. DOI 10.6060/snt.20247802.0007. EDN CDNSYQ.
16. **Astrakhantseva I.A., Astrakhantsev R.G., Usoltsev S.D. [et al.].** K-means analysis of spectral properties of BODIPY dye during compression on air - water interface: construction of dataset for effective clustering. *Liquid Crystals and their Application.* 2024. V. 24, N 2. P. 43-53. DOI 10.18083/LCAppl. 2024.2.43. EDN CXAIQV.
17. **Зимнуров М.Ф., Астраханцева И.А.** Построение API для систем отслеживания задач. *Современные наукоемкие технологии. Региональное приложение.* 2024. № 4(80). С. 97-103. DOI 10.6060/snt.20248004.00013.
18. **Зимнуров М.Ф.** Разработка интерфейса по метрике загруженности сотрудников. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2024. № 4(62). С. 82-87. DOI 10.6060/ivecofin.2024624.705. EDN OETJTW.
19. **Макшанова А.О., Зимнуров М.Ф.** Мемоизация и агрегация в визуализации построения химических соединений. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством [Ивэкофин]*. 2022. № 2(52). С. 106-111. DOI 10.6060/ ivecofin. 2022522.607. EDN YSSNPT.
13. C4 Model for Visualising Software Architecture https://c4model.com.
14. **Astrakhantseva I.A., Kotenev T.E., Gorev S.V. [et al.].** Integration of Linear Regression and Random Forest Methods into Intelligent Management Support Systems for Cryogenic Transport Systems. *Ivecofin.* 2024. N 3 (61). P. 81-90. DOI 10.6060/ivecofin.2024613.692.
15. **Astrakhantseva I.A., Kotenev T.E. , Gorev S.V. [et al.].** Gradient Boosting Method for Forecasting Management Decisions in a Multilayer Cryogenic System. *Modern high technology. Regional application.* 2024. N 2 (78). P. 50-58. DOI 10.6060/snt.20247802.0007. EDN CDNSYQ.
16. **Astrakhantseva I.A., Astrakhantsev R.G., Usoltsev S.D. [et al.].** K-means analysis of spectral properties of BODIPY dye during compression on air - water interface: construction of dataset for effective clustering. *Liquid Crystals and their Application.* 2024. V. 24, N 2. P. 43-53. DOI 10.18083/LCAppl.2024.2.43. EDN CXAIQV.
17. **Zimnurov M.F., Astrakhantseva I.A.** Construction of API for task tracking systems. *Modern high technology. Regional application.* 2024. N 4(80). P. 97-103. DOI 10.6060/snt.20248004.00013.
18. **Zimnurov M.F.** Development of an interface based on the employee workload metric. *Ivecofin.* 2024. N 4(62). P. 82-87. DOI 10.6060/ivecofin.2024624.705. EDN OETJTW.
19. **Makshanova A.O., Zimnurov M.F.** Memoization and aggregation in visualization of chemical compound construction. *Ivecofin.* 2022. N 2(52). P. 106-111. DOI 10.6060/ivecofin.2022522.607. EDN YSSNPT.

Поступила в редакцию(Received) 16.01.2024
Принята к опубликованию (Accepted) 10.02.2025