



## Графовый метод расширенного поиска знаний в специализированных текстах, содержащих координаты, географические названия и аббревиатуры

А. Б. Менисов<sup>а</sup>, доктор техн. наук, старший преподаватель, [orcid.org/0000-0002-9955-2694](https://orcid.org/0000-0002-9955-2694), [vka@mil.ru](mailto:vka@mil.ru)

Д. С. Нагорный<sup>а</sup>, преподаватель, [orcid.org/0009-0008-5867-0011](https://orcid.org/0009-0008-5867-0011)

А. Б. Шелест<sup>а</sup>, канд. техн. наук, доцент, [orcid.org/0009-0009-1944-0229](https://orcid.org/0009-0009-1944-0229)

<sup>а</sup>Военно-космическая академия им. А. Ф. Можайского, Ждановская наб., 13, Санкт-Петербург, 197198, РФ

**Введение:** современные методы расширенного поиска знаний демонстрируют ограниченную эффективность при обработке специализированных текстов, содержащих координаты, топонимы и аббревиатуры, поскольку такие данные требуют одновременного учета семантических и пространственно-структурных зависимостей. Традиционные векторные подходы ориентированы преимущественно на обобщенную семантическую близость и слабо отражают явные пространственные связи между сущностями, что снижает точность поиска в специализированных корпусах. **Цель:** разработать метод семантического поиска в специализированных текстах, содержащих координаты, топонимы и аббревиатуры, за счет интеграции пространственно-структурной информации в процесс расширенного поиска знаний. **Результаты:** предложена гибридная архитектура, объединяющая извлечение и нормализацию сущностей, построение графа знаний на основе пространственной и семантической близости, а также гибридный алгоритм поиска, сочетающий графовый и векторный компоненты с настраиваемым весовым коэффициентом. Экспериментальная проверка метода проведена на открытом наборе геопро пространственных данных, преобразованном в текстовый корпус, отражающий типовые сценарии пространственного поиска. Результаты показали, что предложенный гибридный подход демонстрирует улучшение по метрике MAP по сравнению с векторным поиском и значительное превосходство над графовым методом по всем метрикам. При этом значение MRR уступает векторному поиску, что указывает на компромисс между точностью первой позиции и устойчивостью ранжирования в целом. Установлено, что оптимальное соотношение вкладов графового и векторного компонентов позволяет повысить устойчивость ранжирования и полноту ответов при обработке запросов, содержащих координаты и географические наименования. **Практическая значимость:** предложенный метод может быть использован в системах геоинформационного анализа, интеллектуального поиска и автоматической обработки технической документации. Метод обеспечивает интерпретируемость результатов за счет явного представления знаний в виде графа, устойчивость к разнородным данным и адаптивность к различным типам пользовательских запросов. **Обсуждение:** перспективы дальнейших исследований связаны с применением методов обучения без учителя для снижения зависимости от качества разметки, а также с оптимизацией вычислительной эффективности и масштабируемости графовых моделей знаний.

**Ключевые слова** — гибридный поиск, граф знаний, Retrieval-Augmented Generation, GraphRAG, семантический анализ, пространственные данные, геоинформационные системы, координаты, топонимы, генеративные модели искусственного интеллекта.

**Для цитирования:** Менисов А. Б., Нагорный Д. С., Шелест А. Б. Графовый метод расширенного поиска знаний в специализированных текстах, содержащих координаты, географические названия и аббревиатуры. *Информационно-управляющие системы*, 2026, № 2, с. 2–15. doi:10.31799/1684-8853-2026-2-2-15, EDN: WDEXZF

**For citation:** Menisov A. B., Nagorny D. S., Shelest A. B. A graph-based method for extended knowledge retrieval from specialized texts containing coordinates, geographical names, and abbreviations. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 2, pp. 2–15 (In Russian). doi:10.31799/1684-8853-2026-2-2-15, EDN: WDEXZF

### Введение

В последние годы модели генеративного искусственного интеллекта достигли значительных успехов в решении задач, связанных с извлечением, интерпретацией и интеграцией знаний из текстовых данных [1]. Их применение распространено в таких областях, как автоматическое аннотирование документов, анализ смысловых связей между фразами и построение интеллектуальных поисковых систем [2]. Однако при работе с текстами, содержащими специализированные типы информации, традиционные методы анализа демонстрируют ряд ограничений [3].

Классические подходы расширенного поиска (Retrieval-Augmented Generation, RAG), объединяющие механизмы работы с базами данных и генерацию ответов языковой моделью, показывают высокую результативность при обработке общих текстов и недоменных запросов [4]. Тем не менее при применении RAG к специализированным корпусам, например текстам, включающим географические координаты, картографические обозначения, топонимы, технические термины или доменные аббревиатуры, возникают сложности и снижение качества обработки [5]. Такие данные часто представлены в структурированной или полуструктурированной форме, что делает их трудными для корректного индекси-

рования и семантической интерпретации традиционными векторными методами поиска. В контексте данной работы под знанием понимается структурированное представление сущностей предметной области, их свойств и взаимосвязей, формируемое в виде графа знаний и используемое для синтеза интерпретируемых ответов на пользовательские запросы.

В ответ на эти вызовы развивается новый класс подходов — GraphRAG (Graph-based Retrieval Augmented Generation) [6]. Их ключевая идея заключается в использовании графов знаний для представления информации в виде узлов (сущностей) и ребер (связей между ними). Такой подход обеспечивает возможность более гибкого и точного поиска, позволяя языковым моделям не только находить нужные фрагменты данных, но и понимать контекст их взаимосвязей.

### Анализ научно-технических работ

GraphRAG-система строится на принципах интеграции графовой структуры знаний с возможностями генеративных языковых моделей, что обеспечивает надежное извлечение и интерпретацию сложной информации, особенно пространственно-временных данных и доменных обозначений [7].

Работа системы начинается с этапа получения информации из различных внешних источников (таблиц, текстов, баз данных), которые могут содержать пространственно-временные параметры (координаты, временные метки, топонимы, технические аббревиатуры).

Далее специализированные процессы (например, преобразователи типа LLMGraphTransformer) анализируют эти данные, выделяют сущности (узлы) и их связи (ребра), например объекты, события, их местоположение и время. Разделение информации по уровням (семантическому, символическому, контекстному) позволяет точно индексировать сложные фрагменты, структурированные записи или неоднозначные обозначения.

Для поиска релевантной информации используются как классические векторные методы [8], так и графовые [9, 10], которые могут динамически извлекать подграфы, соответствующие запросу. Графы знаний предоставляют богатый контекст, например, для поиска информации о событиях в определенный временной период или для выявления связей между аббревиатурами и их расшифровками.

Найденный фрагмент подграфа либо извлеченная последовательность событий или объектов передается в языковую модель, которая ге-

нерирует объяснение, аналитический отчет или структурированный ответ для пользователя.

Можно привести следующие примеры применения:

- в юриспруденции — анализ связей между нормативными документами, судебными прецедентами, законами для поиска и объяснения сложных юридических взаимосвязей, автоматическое составление отчетов по юридическим вопросам [11];

- в бизнес-аналитике — исследование сложных сетей связей между компаниями, технологиями и рынками, включающее временные аспекты трендов и событий, взаимодействие между структурированными (финансовыми) и неструктурированными (новостными) данными [12];

- в медицине — построение графов клинических испытаний, взаимодействий между препаратами и симптомами, анализ пространственно-временных моделей распространения заболеваний или генетических связей [13];

- в образовании и науке — ускорение обработки научных публикаций, поиск скрытых связей между концепциями и исследованиями с учетом времени публикации и географии исследований [14].

Таким образом, GraphRAG эффективно решает задачи поиска и анализа информации в сложных специализированных корпусах, обеспечивая точность, полноту и объяснимость ответа посредством интеграции графовых моделей знаний и генеративных языковых систем. Дополнительно стоит отметить, что для задач, где важно учитывать изменения, например отслеживание хронологии событий или обновлений нормативных актов, используются специальные уровни графов с временной иерархией, что позволяет корректно отличать и объединять факты, относящиеся к разным периодам.

Современные системы информационного поиска и извлечения знаний используют различные уровни обработки текстовой информации, включая архитектурные подходы, модели представления текста и алгоритмы сопоставления. К архитектурным решениям относятся классические системы полнотекстового поиска, а также модели расширенной генерации ответов (RAG) [15], в которых поиск релевантных фрагментов осуществляется с помощью внешнего поискового механизма, а генерация ответа — языковой моделью. Несмотря на высокую эффективность RAG-подходов, их качество существенно зависит от используемого механизма извлечения и ограничено возможностями базовых моделей поиска.

На уровне представления текста широко применяются векторные и псевдовекторные модели, включая TF-IDF [16], эмбединги и др. К числу наиболее эффективных решений относятся мо-

■ **Таблица 1.** Типы данных в специализированных текстах

■ **Table 1.** Data types in specialized texts

| Тип данных                | Пример                 | Особенности                          | Примечание   |
|---------------------------|------------------------|--------------------------------------|--|
| Координаты                | 55.7558° N, 37.6176° E | Требует парсинга и геокодирования    | Модели не интерпретируют координаты как объект реального мира, не сопоставляют точки |
| Топонимы                  | Москва, Лужники        | Полисемия, синонимия                 | Требуется нормализация, разрешение неоднозначности, поиск синонимов                  |
| Аббревиатуры              | ГИС, ЦОД, ВПП          | Требует онтологического словаря      | Одно сокращение может иметь 5–10 значений, модели без доменной онтологии их путают   |
| Пространственные описания | В 3 км к северу        | Необходима контекстная интерпретация | Нужна интерпретация пространства, а не только текста                                 |

дели, использующие механизм токен-уровневого сопоставления [17] и сочетающие преимущества разреженных представлений и нейросетевого обучения.

Однако даже данные модели ориентированы преимущественно на оценку семантической близости текстовых фрагментов и не обеспечивают явного представления структурных и пространственных связей между объектами. В условиях специализированных корпусов, содержащих координаты, географические наименования и доменные аббревиатуры, это приводит к потере пространственной и топологической информации и снижению качества аналитических запросов. Особенностью специализированных текстов является присутствие устойчивых форматов данных, представленных в табл. 1.

Обработка специализированных текстов, содержащих координаты, географические названия и аббревиатуры, представляет собой комплексную задачу, выходящую за рамки чисто лингвистического анализа. Таким образом, проблема носит не лингвистический, а семантико-пространственный характер, что обуславливает необходимость графового подхода, актуальность которого по сравнению с другими подходами определена рядом факторов.

Во-первых, традиционные методы обработки естественного языка демонстрируют ограниченную эффективность при работе с такими данными [18], поскольку:

- координаты требуют точной математической интерпретации и привязки к системе отсчета;

- географические названия образуют сложные иерархические структуры (страна/регион/город) [19];

- аббревиатуры обладают контекстно-зависимой семантикой [20].

Во-вторых, используемые в классических RAG-системах векторные представления (эмбеддинги) формируются на основе анализа совмест-

ной встречаемости терминов, фраз или предложений в контекстном окружении и тем самым аппроксимируют статистико-контекстную близость между текстовыми фрагментами. Такие представления эффективно захватывают обобщенные смысловые ассоциации, однако не кодируют явно типы отношений между объектами, такие как пространственная вложенность, направленность, причинно-следственные или метрические зависимости. В частности, расстояние между векторами отражает вероятность совместного употребления сущностей в сходных контекстах, но не сохраняет информацию о конкретной природе связи, например: «расположен в», «находится на расстоянии», «является частью» или «следует после во времени» [21]. Это ограничение принципиально для задач, связанных с геоинформационным анализом, где пространственные отношения и метрика расстояний должны интерпретироваться детерминированно, а не вероятностно [22].

В-третьих, специализированные тексты формируют сложные системы связей, которые могут быть представлены в виде многоуровневой структуры (табл. 2).

■ **Таблица 2.** Типы связей в специализированных текстах

■ **Table 2.** Types of connections in specialized texts

| Тип сущности                  | Тип связи          | Характер связи                     |
|-------------------------------|--------------------|------------------------------------|
| Объекты и геолокации          | Пространственные   | Метрические отношения, вложенность |
| Термины и аббревиатуры        | Онтологические     | Иерархические, синонимические      |
| События и временные интервалы | Причинно-временные | Последовательности, зависимости    |
| Названия и координаты         | Сопоставительные   | Тождество, эквивалентность         |

В-четвертых, стандартный полнотекстовый поиск не реконструирует такие связи, что ведет к потере смысловых цепочек и контекстных зависимостей и пропуску критически важных фактов, а также некорректным ответам на аналитические запросы [23].

Графовая модель позволяет строить семантические и топологические контуры знаний, что особенно важно для систем аналитической обработки данных и обеспечивает:

- возможность проследить путь от исходных данных к результату;
- проверяемость фактов, которые могут быть верифицированы по исходным данным;
- соответствие стандартам и процедурам принятия решений.

Таким образом, для информационных систем эти характеристики являются обязательными требованиями, а не опциональными преимуществами. GraphRAG-подход является не просто альтернативой, а наиболее рациональным решением для интеграции геоданных, терминологических цепочек и семантического анализа текстов. Это обусловлено фундаментальными ограничениями векторных представлений [24] и необходимостью сохранения структурной целостности пространственно-семантических данных [25].

### Формализованная постановка задачи

Пусть задан корпус текстов  $D = \{d_1, \dots, d_n\}$ , относящихся к специализированной предметной области (например, геодезия, логистика, топография). Каждый документ  $d_i$  содержит неструктурированную текстовую информацию, включающую упоминания географических координат, названий, аббревиатур и иных именованных сущностей.

Задача заключается в построении семантико-пространственной модели знаний в виде графа  $G = (V, E)$ , где  $E$  – множество ребер, отражающих связи между сущностями. Ребро  $e_{ij}$  между вершинами  $v_i$  и  $v_j$  существует, если выполняются условия, определяемые функциями близости;  $V$  – множество вершин, представляющих собой извлеченные и нормализованные сущности. Каждая сущность  $v \in V$  определяется как результат применения последовательности функций извлечения к корпусу  $D$ :

$$v = f(d) = \text{NER}(d) \cup \text{GeoParse}(d) \cup \text{AbbrevResolve}(d), \quad (1)$$

где  $\text{NER}(d)$  – множество стандартных именованных сущностей (персоны, организации, локация);  $\text{GeoParse}(d)$  – множество географических объектов, извлеченных из текстовых описаний

и координат;  $\text{AbbrevResolve}(d)$  – множество расшифрованных аббревиатур и терминов.

Последовательности функций извлечения и нормализации сущностей включают несколько этапов, реализуемых с использованием стандартных и открыто доступных методов обработки естественного языка и геоданных.

Извлечение стандартных именованных сущностей (множество  $\text{NER}(d)$ ) проводится с применением моделей распознавания именованных сущностей на основе трансформерных архитектур и дообученных на специализированных данных. В экспериментальной реализации использовались модели класса BERT/roBERTa, доступные в библиотеке spaCy.

Извлечение географических объектов (множество  $\text{GeoParse}(d)$ ) выполняется посредством комбинации правил регулярных выражений для распознавания координатных форматов и процедур геокодирования, обеспечивающих приведение координат и топонимов к унифицированному представлению.

Расшифровка аббревиатур и терминов (множество  $\text{AbbrevResolve}(d)$ ) реализуется на основе словаря, формируемого автоматически из корпуса текстов и дополняемого внешними источниками. Каждая аббревиатура нормализуется путем сопоставления с ее полной формой, после чего добавляется в граф знаний как отдельная сущность.

Все используемые функции извлечения и нормализации реализуются в виде модульных компонентов и могут быть воспроизведены с использованием открытых программных библиотек, что обеспечивает воспроизводимость предложенного подхода.

На основе построенного графа  $G$  необходимо синтезировать механизм поиска ответов  $A(q, G)$  на пользовательский запрос  $q$ , принадлежащий некоторому распределению запросов  $Q$ . Целью является максимизация релевантности ответа механизма эталонному ответу  $y_q$ :

$$E(\text{Rel}(A(q, G), y_q)) \rightarrow \max, \quad (2)$$

где  $\text{Rel}$  – выбранная метрика релевантности (например, BLEU, ROUGE, Mean Reciprocal Rank, MRR); пользовательский запрос  $q \in Q$  представляет собой текстовый запрос на естественном языке, содержащий указание на сущности предметной области и их взаимосвязи.

Запрос может включать пространственные условия (радиус, направление, расстояние), указание на конкретные объекты или контрольные точки.

Примерами пользовательских запросов являются: «Найти объекты в радиусе 2 км от КП 45», «Связь между КП 44 и выходами ПГС в районе р. Быстрая».

## Описание метода

Предлагаемый метод расширенного поиска знаний состоит из трех последовательных этапов.

Этап 1. *Извлечение и нормализация сущностей.*

Извлечение и нормализация сущностей выполняются с использованием комбинации открытых лингвистических моделей, доменных словарей и правил постобработки. На этом этапе происходит пополнение множества вершин графа  $V$ :

– для извлечения стандартных именованных сущностей (персоны, организации, географические наименования) применяются предварительно обученные модели распознавания именованных сущностей (NER), доступные в открытых библиотеках обработки естественного языка (например, spaCy). В качестве базовых моделей используются трансформерные архитектуры, дообученные на специализированных текстах предметной области;

– географические объекты и координатная информация извлекаются с применением процедур геопарсинга, включающих распознавание координатных форматов, топонимов и пространственных указаний, с последующей нормализацией к унифицированному представлению (для этих целей используются открытые геоинформационные словари и библиотеки геокодирования, такие как GeoNames и GeoPy). Дополнительно производится геокодирование координат в унифицированное представление:

$$g = \text{GeoNorm}(c, s, r), \quad (3)$$

где  $g$  – нормализованное географическое представление;  $c$  – исходное географическое представление, извлеченное из текста (координаты, топонимы, текстовые пространственные описания);  $s$  – система координат и единицы измерения;  $r$  – пространственное разрешение (точность) нормализации;

– нормализация терминов и расшифровка аббревиатур осуществляются на основе доменных словарей, формируемых из корпуса документов и дополняемых внешними открытыми источниками. В процессе нормализации различные лексические формы, синонимы и сокращения приводятся к каноническому виду, после чего соответствующие сущности добавляются в множество вершин графа.

Этап 2. *Формирование графа знаний.* Ребра графа  $E$  формируются на основе гибридного подхода, комбинирующего пространственную и семантическую близость:

$$e_{ij} = \begin{cases} 1, & \text{если } \text{dist}(v_i, v_j) < \varepsilon - \\ & \text{пространственный критерий;} \\ 1, & \text{если } \text{sim}(h_i, h_j) > \tau - \text{семантический} \\ & \text{критерий;} \\ 0 & \text{в других случаях,} \end{cases} \quad (4)$$

где  $\text{dist}(v_i, v_j)$  – геодезическое расстояние между географическими координатами сущностей  $v_i$  и  $v_j$ , вычисляемое с использованием формулы гаверсинусов для сферической модели Земли;  $h_i$  и  $h_j$  – векторные представления сущностей, полученные с помощью предварительно обученной языковой модели (например, Sentence-BERT [26]);  $\text{sim}(h_i, h_j)$  – косинусное сходство соответствующих векторов;  $\varepsilon$  – порог пространственной близости, выраженный в метрах;  $\tau$  – порог семантической близости [27].

Ребро между вершинами графа создается только в том случае, если выполняются условия пространственной и семантической близости, что позволяет исключить семантически сходные, но географически удаленные сущности, а также пространственно близкие, но тематически несвязанные объекты [28].

При формировании ребер графа каждому ребру сопоставляется тип близости (пространственный, семантический или гибридный). Данная информация сохраняется в графе и используется на этапе поиска и интерпретации результатов. На рис. 1 представлен пример формирования графа по следующим данным:

```
[ {
  «id»: «doc_001»,
  «text»: «Установка КП 45 в точке 55.755831,
37.617673. Рельеф: всхолмленная равнина с ПГС.
Смещение относительно КП 44 – 125 м по азимуту 45°.»
},
{
  «id»: «doc_002»,
  «text»: «Обследование района р. Быстрая.
Координаты базового лагеря: 54.1234, 38.5678.
Обнаружены выходы ПГС в 2 км к северу от
КП 45.»
},
{
  «id»: «doc_003»,
  «text»: «Маршрутная съемка от КП 44
до КП 45. Зафиксированы точки с координатами
55.750000, 37.620000. Ландшафт: холмистая
местность с преобладанием ПГС.»
},
{
  «id»: «doc_004»,
```

```

    «text»: «Геодезические работы в районе
перевала Северный. Установлен КП 46 на высоте
1250 м. Координаты: 55.800000, 37.600000.»
  },
  {
    «id»: «doc_005»,
    «text»: «В районе р. Быстрая обнаружен
КП. Координаты: 54.0000, 38.0000»
  }

```

В процессе нормализации пространственных описаний дополнительно формируются служебные пространственно-семантические сущности (например, «в районе»), используемые для привязки событий и объектов к географическому контексту. Такие узлы маркируются префиксом *geo\_* и отличаются от именованных географических объектов (река, перевал, населенный пункт).

Следует отметить, что на рис. 1 представлен упрощенный фрагмент графа знаний, сфокусированный на координатных и контрольных точках. Географические объекты, такие как река Быстрая и перевал Северный, извлекаются системой и присутствуют в полном графе, однако опущены на рисунке для повышения наглядности. Пространственные параметры, выраженные в текстах в виде расстояний, направлений и высот (например, «125 м по азимуту 45°», «в 2 км к северу», «на высоте 1250 м»), используются при формировании пространственных связей между

сущностями и сохраняются в виде атрибутов ребер графа.

На рис. 2 представлен результат анализа сформированного графа знаний с использованием нормализованных метрик центральности вершин (степени, по посредничеству) и важности узла (по аналогу с метрикой PageRank [29]), позволяющих количественно оценить структурную роль и значимость сущностей в графе [30]. Анализ центральности используется как инструмент интерпретации структуры графа знаний, выявления ключевых сущностей и оценки их влияния на распространение информации в графе.

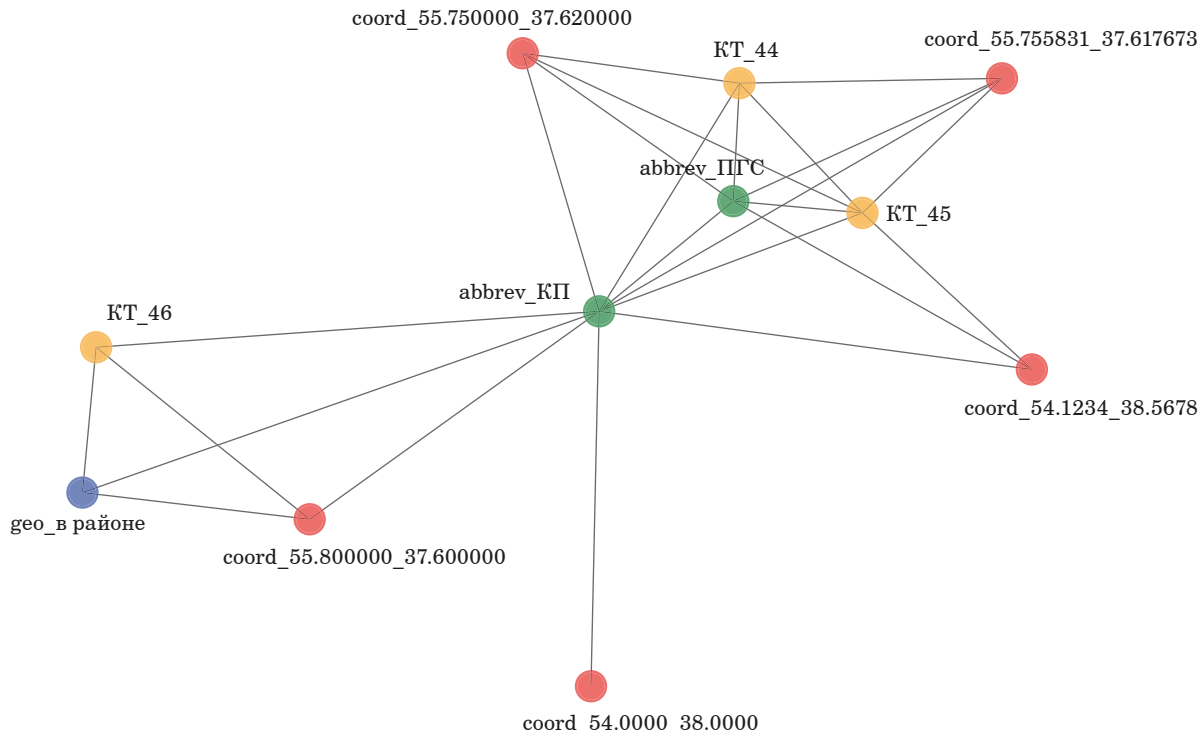
В рамках анализа использовались стандартные метрики центральности графов. В частности, степенная центральность (degree) вершины  $v_i$  определяется как

$$C_d(v_i) = \frac{\deg(v_i)}{V-1}, \quad (5)$$

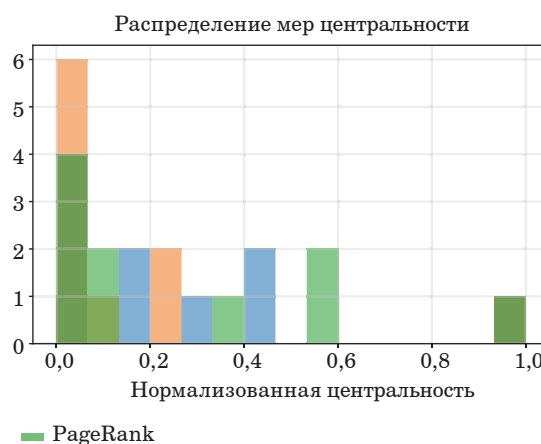
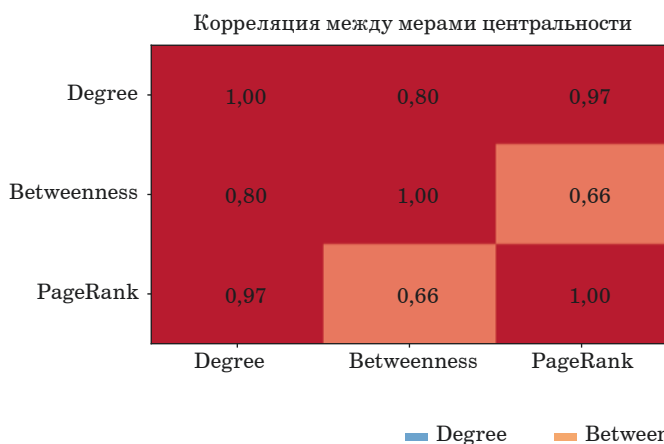
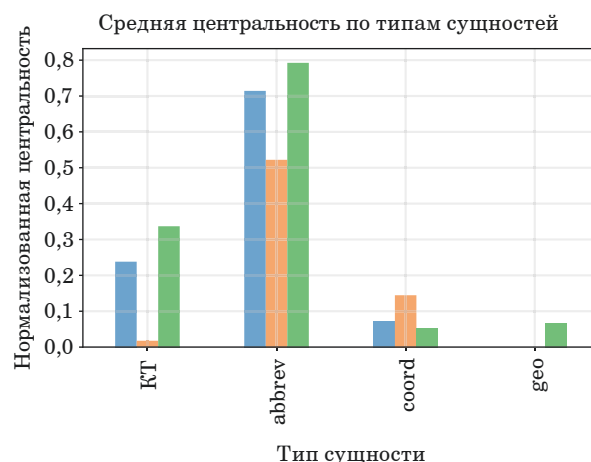
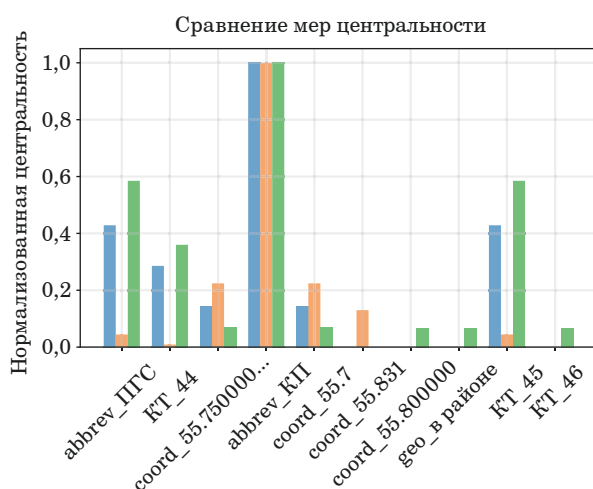
где  $\deg(v_i)$  – количество ребер вершины  $v_i$ ;  $V$  – общее число вершин в графе. Нормализация делением на  $V - 1$  обеспечивает диапазон значений  $C_d \in [0, 1]$ .

Метрика центральности по посредничеству (betweenness) измеряется как

$$C_b(v_i) = \frac{1}{(V-1)(V-2)} \sum_{v_j \in V, j \neq i \neq k} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}}, \quad (6)$$



■ **Рис. 1.** Пример формирования графа знаний: *КТ* – контрольная точка  
 ■ **Fig. 1.** An example of knowledge graph construction: *КТ* – checkpoint



■ **Рис. 2.** Анализ графа знаний  
 ■ **Fig. 2.** Knowledge graph analysis

где  $\sigma_{jk}$  — количество кратчайших путей между вершинами  $v_k$  и  $v_i$  в графе;  $\sigma_{jk}(v_i)$  — количество таких путей, проходящих через вершину  $v_i$ . Нормализация  $(V - 1)(V - 2)$  позволяет масштабировать значения в диапазон  $[0, 1]$ .

Важность узла графа:

$$C_{pr}(v_i) = (1 - d) + d \sum_{v_j \in N(v_i)} \frac{1}{\deg(v_j)}, \quad (7)$$

где  $d$  — дампинг-фактор;  $N(v_i)$  — множество соседних вершин;  $\deg(v_j)$  — степень вершины  $v_j$ . Результаты нормализуются так, чтобы сумма всех  $C_{pr}(v_i)$  равнялась 1.

Этап 3. Гибридный механизм поиска. Для обработки запроса  $q$  используется комбинированная стратегия, объединяющая структурные возможности графа знаний и семантические свойства векторных представлений. Данный этап обеспечивает ранжирование документов и сущ-

ностей с учетом их смысловой близости к запросу и их структурной значимости в графе знаний.

Пользовательский запрос  $q$  приводится к унифицированному виду с использованием процедур предварительной обработки текста, аналогичных применяемым для корпуса документов. В результате формируются множество сущностей запроса  $V_q \subset V$  и векторное представление запроса  $q \in R^2$ .

Графовый компонент поиска основан на анализе структуры графа знаний  $G = (V, E)$ . Для каждой сущности  $v_i \in V_q$  выполняется поиск релевантных вершин графа, связанных с запросом через кратчайшие пути.

При этом для оценки структурной роли и значимости сущностей используются нормализованные метрики центральности:

- степенная центральность (5), характеризующая локальную связанность сущности;
- центральность по посредничеству (6), отражающая роль сущности как структурного посредника;

– метрика важности узла (7), учитывающая глобальное положение сущности в графе.

На их основе для каждой вершины вычисляется интегральная структурная центральность

$$C_s(v_i) = \alpha_1 \cdot C_d(v_i) + \alpha_2 \cdot C_b(v_i) + \alpha_3 \cdot C_{pr}(v_i) \quad (8)$$

при  $\alpha_1, \alpha_2, \alpha_3 \geq 0$  и  $\alpha_1 + \alpha_2 + \alpha_3 = 1$ .

Для информации, ассоциированной с вершиной  $v_i$ , графовая оценка релевантности запросу GraphSearch определяется как

$$\text{GraphSearch}(q) = \frac{1}{|V|} \sum_{v_i \in V} \max_{v_i \in V} \left( \frac{C_s(v_i)}{1 + \text{dist}(v_i, v_q)} \right), \quad (9)$$

где  $\text{dist}(v_i, v_q)$  – длина кратчайшего пути между сущностями документа  $v_i$  и сущностями запроса  $v_q$ .

Таким образом, больший вклад в итоговую оценку получают сущности, которые одновременно структурно значимы в графе и находятся на малом расстоянии от сущностей запроса.

Параллельно выполняется векторный поиск в семантическом пространстве. Для оценки семантической близости между запросом и документами используется метод плотных векторных представлений текста [31]. Каждый текстовый запрос  $q$  и документ  $d$  кодируются в векторное пространство фиксированной размерности с использованием модели трансформерных эмбедингов. Для каждой информации вычисляется мера семантического сходства

$$\text{VectorSearch}(q) = \text{sim}(E(q), E(d)), \quad (10)$$

где  $E(q)$  – эмбединг запроса;  $E(d)$  – эмбединг документа;  $\text{sim}(\ )$  – косинусная мера сходства между векторами.

Полученные значения используются для ранжирования документов по убыванию семантической близости к запросу [32]. Поиск ближайших соседей выполняется посредством алгоритма k-ближайших соседей. Для каждого запроса вычисляется косинусное сходство между эмбедингом запроса и эмбедингами всех документов коллекции, после чего выбираются  $K$  документов с максимальным значением меры сходства.

Итоговая оценка определяется как взвешенная комбинация графовой и векторной оценок:

$$A(q) = \lambda \cdot \text{GraphSearch}(q) + (1 - \lambda) \cdot \text{VectorSearch}(q), \quad (11)$$

где параметр  $\lambda \in [0, 1]$  регулирует вклад каждого из типов поиска.

Таким образом, VectorSearch выявляет скрытые семантические связи между запросом и документами, а GraphSearch позволяет учитывать

явные структурные, пространственные и топологические связи между сущностями. Гибридный механизм поиска объединяет преимущества обоих подходов, усиливая релевантность для запросов, содержащих как семантические, так и структурно-координатные зависимости.

## Экспериментальное исследование

Для оценки качества предложенного метода использовался корпус текстов с географическими сущностями и привязками к координатам Tesla Supercharge Locations Globally (<https://www.kaggle.com/datasets/omarsobhy14/supercharge-locations>), содержащий около 6000 записей, каждая из которых соответствует одной точке сети зарядной инфраструктуры. Каждый объект описывается координатами (широта, долгота), топонимической привязкой (город, регион, страна), а также дополнительными атрибутами, такими как идентификатор станции и тип локации. Несмотря на то, что этот корпус текстов изначально не предназначен для задач информационного поиска или ранжирования документов, его структура позволяет сформировать коллекцию текстовых описаний объектов инфраструктуры, которые могут использоваться как документы в задаче семантического поиска.

В рамках эксперимента рассмотрены следующие методы поиска:

1) векторный поиск (Vector) – семантический поиск в векторном пространстве на основе современных эмбедингов (модель E5, <https://huggingface.co/intfloat/multilingual-e5-base>) с использованием косинусной меры сходства;

2) графовый поиск (Graph) – поиск по графу знаний, построенному в соответствии с формулой (9), с учетом пространственных связей и структурной важности вершин, определяемой нормализованными метриками центральности (коэффициенты подбирались эмпирически на валидационной выборке путем перебора с шагом 0,1 с максимизацией значения метрики MAP);

3) гибридный поиск (Hybrid) – предложенный метод, реализующий линейную комбинацию результатов графового и векторного поиска согласно формуле (11).

Все методы использовали единый корпус документов и одинаковый набор запросов, что обеспечивало корректность сравнительного анализа. Для оценивания качества метода был сформирован набор запросов на русском языке, отражающих типовые сценарии взаимодействия с системой поиска информации о зарядных станциях Tesla, включая поиск объектов в заданной географической области, поиск ближайших объектов к указанным координатам и запросы, со-

держащие топонимы и пространственные ограничения в неявной форме. Запросы моделируют реальные пользовательские формулировки и используют различные лексические формы, географические указания и синонимы (например, «найти зарядные станции Tesla в Германии», «суперчарджеры рядом с Берлином», «сеть зарядных станций Tesla в Калифорнии»).

Поскольку корпус текстов содержит структурированные географические атрибуты, эталонная релевантность документов определялась автоматически на основе этих полей. Для каждого тестового запроса формировалось множество документов, удовлетворяющих соответствующему условию. Например, для запроса «найти зарядные станции Tesla в Германии» релевантными считаются все записи, у которых поле Country = Germany. Таким образом создается тестовая выборка, проверяющая способность системы находить релевантные документы.

Качество поиска оценивалось с использованием стандартных метрик ранжирования [33, 34]:

- средний обратный ранг первого релевантного ответа (Mean Reciprocal Rank, MRR);

- доля релевантных документов среди первых  $k$  результатов (в эксперименте  $k = 5$  – Precision@k);

- средняя точность всех ответов (Mean Average Precision, MAP).

Данные метрики позволяют оценить как позицию первого релевантного документа, так и общее качество ранжирования поисковой выдачи. Дополнительно добавлены доверительные интервалы, рассчитанные методом бутстрап (<https://wiki.loginom.ru/articles/bootstrap.html>) с 1000 повторов на уровне доверия 95 %. Такой подход позволяет оценить устойчивость результатов и отражает статистическую вариативность метрик, приведенных в табл. 3.

Как видно из таблицы, гибридный метод демонстрирует улучшение качества ранжирования

по сравнению с отдельным графовым и векторным поиском, при этом доверительные интервалы показывают, что различия статистически устойчивы.

Анализ полученных результатов выявил, что эффективность графового поиска ограничена, что обусловлено неполнотой и разреженностью связей. Векторный поиск обеспечивает высокое качество за счет семантической близости текстов, однако игнорирует явные пространственные и структурные зависимости между сущностями.

Доверительные интервалы показывают высокую вариативность из-за небольшого числа тестовых запросов, но общий тренд иллюстрирует, что графовый поиск добавляет улучшение в ранжировании топ-кандидатов, а векторная модель обеспечивает базовую семантическую релевантность.

Предложенный метод сочетает преимущества обоих подходов и представляет наилучшие значения метрики MAP, что указывает на более устойчивое качество ранжирования по всему списку результатов. Повышение Precision@5 по сравнению с графовым методом подтверждает эффективность интеграции семантического компонента, а рост MAP относительно векторного поиска свидетельствует о положительном вкладе графовой структурной информации.

Проведенные экспериментальные исследования позволили выявить особенности применения векторного, графового и гибридного методов поиска при обработке специализированных геопрограммных запросов на русскоязычном корпусе.

## Обсуждение

Результаты показали, что чисто графовый подход демонстрирует наименьшие значения метрик качества (MRR = 0,186; Precision@5 = 0,198; MAP = 0,013). Это объясняется ограниченной структурной насыщенностью графа знаний, сформированного на основе рассматриваемого корпуса, так как большинство сущностей обладают малой степенью связности, а расстояния между ними слабо дифференцированы, что снижает способность структурных метрик центральности и путей.

Векторный поиск на основе модели E5 обеспечивает высокие показатели ранжирования (MRR = 0,59; Precision@5 = 0,32; MAP = 0,11), выступая в роли сильного базового решения. Это подтверждает, что современные мультиязычные бинкодеры эффективно определяют семантическое сходство даже в специализированных текстах, содержащих координаты и топонимы, при условии относительно простой структуры запросов.

Наиболее показательные результаты получены для гибридного подхода, объединяющего

■ **Таблица 3.** Сравнительные результаты методов поиска по различным метрикам

■ **Table 3.** Comparative retrieval performance across metrics

| Метод                       | MRR                        | Precision@5                | MAP                        |
|-----------------------------|----------------------------|----------------------------|----------------------------|
| Векторный поиск (Vector)    | 0,590<br>[0,200;<br>1,000] | 0,320<br>[0,040;<br>0,680] | 0,110<br>[0,002;<br>0,297] |
| Графовый поиск (Graph)      | 0,186<br>[0,000;<br>0,600] | 0,198<br>[0,000;<br>0,600] | 0,013<br>[0,000;<br>0,037] |
| Предложенный метод (Hybrid) | 0,464<br>[0,067;<br>0,867] | 0,324<br>[0,040;<br>0,680] | 0,129<br>[0,000;<br>0,357] |

векторный и графовый поиск с адаптивным весовым коэффициентом. Несмотря на некоторое снижение MRR по сравнению с чисто векторным методом (0,464 против 0,59), гибридная модель существенно превосходит векторный поиск по метрике MAP (0,129 против 0,110), что свидетельствует о более качественном ранжировании релевантных документов на протяжении всей выдачи, а не только в первой позиции. При этом значение Precision@5 для гибридного метода (0,32) соответствует векторному базису, что указывает на отсутствие деградации качества верхней части списка результатов.

Данный эффект объясняется тем, что графовый компонент в гибридной архитектуре выполняет роль структурного регуляризатора, усиливая вклад документов, содержащих сущности, топологически близкие к сущностям запроса и обладающие высокой структурной значимостью в графе знаний. Использование нормализованных метрик центральности вершин (степени, посредничества и аналогов PageRank) позволило количественно учитывать роль сущностей в общей структуре знаний, что особенно важно при обработке пространственно ориентированных запросов.

Дополнительный анализ показал, что ключевым фактором успешности гибридного подхода является адаптивный характер интеграции графового и векторного компонентов. Введение весового коэффициента, зависящего от уверенности векторного поиска и наличия пространственных маркеров в запросе, позволило избежать негативного влияния графа в случаях, когда структурные связи выражены слабо, и, напротив, усилить его вклад для запросов, требующих пространственно-структурной интерпретации. Оптимальное значение весового коэффициента, определяющего вклад графового поиска в итоговую функцию релевантности (11), составляет 0,6. Это указывает на доминирующую роль структурно-пространственных связей при обработке запросов, содержащих координаты и топонимы, при сохранении значимого вклада семантического поиска.

Таким образом, полученные результаты подтверждают, что гибридные методы поиска не предназначены для прямой замены сильных векторных моделей на слабо структурированных корпусах, однако они обеспечивают более устойчивое и интерпретируемое ранжирование, особенно в сценариях, где релевантность определяется не только семантическим сходством, но и пространственными и реляционными связями между сущностями. Это делает предложенный подход перспективным для применения в системах геоинформационного анализа и интеллектуального поиска по специализированным техническим документам.

Несмотря на полученные положительные результаты, проведенное исследование имеет два направления ограничений, которые необходимо учитывать при интерпретации выводов.

Первое направление ограничений заключается в следующем: качество гибридного поиска в значительной степени зависит от точности этапов извлечения и нормализации сущностей, а также от корректности геокодирования. Ошибки на этих этапах приводят к искажению структуры графа знаний и, как следствие, к снижению эффективности графового компонента. В работе данные процессы рассматривались как заданные, без отдельного анализа их устойчивости к шуму и пропущенным сущностям.

Для второго направления выделяется, что подбор параметров гибридной модели, включая весовой коэффициент интеграции графового и векторного поиска, осуществлялся эмпирически и не опирался на формальную процедуру оптимизации. Это оставляет пространство для дальнейшего улучшения качества за счет применения методов автоматического подбора гиперпараметров или обучения параметров интеграции в рамках единой оптимизационной задачи.

В качестве направлений дальнейших исследований представляется перспективным расширение корпуса за счет реальных пользовательских запросов и документов из различных прикладных областей, что позволит более полно оценить устойчивость предложенного метода.

## Заключение

В работе предложен и экспериментально проверен гибридный графовый метод для расширенного поиска знаний в специализированных текстах, обогащенных координатами, географическими названиями и аббревиатурами. Разработанный метод сочетает извлечение и нормализацию сущностей, построение графа на основе пространственных и семантических связей и гибридный алгоритм поиска, комбинирующий графовые и векторные техники. Такое сочетание позволяет повысить качество ответа больших языковых моделей при работе с узкоспециализированными корпусами, сохраняя при этом преимущества генеративных архитектур искусственного интеллекта.

Проведенные эксперименты показали, что предложенный гибридный подход эффективно сочетает преимущества графовых и векторных методов, обеспечивая высокие показатели качества при сохранении вычислительной эффективности. Полученные результаты подтверждают жизнеспособность и практическую применимость метода в задачах семантического поиска и анали-

за пространственно ориентированных данных. Кроме того, выявленные зависимости и ограничения формируют основу для дальнейшего совершенствования архитектуры, направленного на повышение точности и адаптивности системы.

Основные преимущества метода: интерпретируемость результатов, устойчивость к разнородным данным и адаптивность — делают его перспективным для внедрения в системах анализа документов для геодезии, картографии, логистики и оборонного комплекса. Интерпретируемость результатов обеспечивается возможностью явной трассировки путей в графе знаний от исходных сущностей к итоговому ответу. Устойчивость к разнородным данным достигается за счет раз-

дельной обработки текстовых, координатных и терминологических компонентов. Адаптивность метода обеспечивается гибридной схемой поиска с настраиваемым вкладом графовых и векторных механизмов.

Направлениями для будущих исследований являются снижение зависимости от качества разметки за счет применения методов обучения без учителя; оптимизация вычислительной эффективности для работы с эксабайтными корпусами данных; реализация механизма инкрементального обновления графа, позволяющего снижать вычислительные затраты при добавлении новых данных.

## Литература

1. Grace K., Sandkühler J. F., Stewart H., Weinstein-Raun B., Thomas S., Stein-Perlman Z., Salvatier J., Brauner J., Korzekwa R. C. Thousands of AI authors on the future of AI. *Journal of Artificial Intelligence Research*, 2025, vol. 84, Article 9. doi:10.1613/jair.1.19087
2. Radanliev P. Artificial intelligence: Reflecting on the past and looking towards the next paradigm shift. *Journal of Experimental & Theoretical Artificial Intelligence*, 2024, vol. 37, no. 7, pp. 1045–1062. doi:10.1080/0952813X.2024.2323042
3. Vijayakumaran S., Vimaladevi M., Thangamani R., Vibeesh N., Easwaramoorthy S. V. Revolutionizing legal access: An AI-Driven RAG chatbot for real-time judicial insights. *2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)*, IEEE, 2025, pp. 1–6. doi:10.1109/AIMLA36882.2025.11040439
4. Kim J., Hur M., Min M. From RAG to QA-RAG: Integrating generative AI for pharmaceutical regulatory compliance process. *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*, 2025, pp. 1293–1295. doi:10.1145/3672608.3707749
5. Peng Z., Kuai X., Ke S., Dong X., Guo R. Enhancing geodatabases operability: Advanced human-computer interaction through RAG and Multi-Agent Systems. *Big Earth Data*, 2025, vol. 9, no. 2, pp. 217–242. doi:10.1080/20964471.2025.2483541
6. Barry M., Caillaut G., Halftermeyer P., Qader R., Mouayad M., Le Deit F., Cariolaro D., Gesnouin J. GraphRAG: Leveraging graph-based efficiency to minimize hallucinations in LLM-driven RAG for finance data. *Proceedings of the Workshop on Generative AI and Knowledge Graphs (GenAIK)*, 2025, pp. 54–65.
7. Yang R., Xue H., Razzak I., Hacid H., Salim F. D. KG-IRAG: A knowledge graph-based iterative retrieval-augmented generation framework for temporal reasoning. *arXiv preprint*, 2025. doi:10.48550/arXiv.2503.14234
8. Шалагин Н. Д. Обзор алгоритмов семантического поиска по текстовым документам. *International Journal of Open Information Technologies*, 2024, vol. 12, no. 9, pp. 11–21.
9. Procko T. T., Ochoa O. Graph retrieval-augmented generation for large language models: A survey. *2024 Conference on AI, Science, Engineering, and Technology (AIxSET)*, Laguna Hills, CA, USA, 2024, pp. 166–169. doi:10.1109/AIxSET62544.2024.00030
10. Никольский Д. Р., Барабанов В. Ф., Гребенникова Н. И., Коваленко С. А., Нужный А. М. Анализ графовых систем управления базами данных. *Вестник Воронежского государственного технического университета*, 2023, т. 19, № 6, с. 13–20. doi:10.36622/VSTU.2023.19.6.002, EDN: RFOQQE
11. Dang D. V., Nguyen H., Le T. T., Do H., Nguyen H., Ngo H. Q., Nguyen H. D. Knowledge graph-based legal query system with LLM and retrieval augmented generation. *Asian Conference on Intelligent Information and Database Systems: Proceedings*, Springer, Singapore, 2025, pp. 161–172. doi:10.1007/978-981-96-5881-7\_13
12. Pattabiraman K., Sachin A., Lokeshkumar R. Enhancing business intelligence through NLP and contextual AI synergy. *Engineering Research Express*, 2025, vol. 7, no. 2, Article 025279. doi:10.1088/2631-8695/add9e6
13. Wu J., Zhu J., Qi Y., Chen J., Xu M., Menolascina F., Jin Y., Grau V. Medical graph RAG: Evidence-based medical large language model via graph retrieval-augmented generation. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2025, pp. 28443–28467. doi:10.18653/v1/2025.acl-long.1381
14. Yang R., Yang B., Zhao X., Gao F., Feng A., Ouyang S., Blum M., She T., Jiang Y., Lecue F., Lu J., Li I. Graphusion: A RAG framework for scientific knowledge graph construction with a global perspective. *Companion Proceedings of the ACM on Web*

- Conference 2025*, 2025, pp. 2579–2588. doi:10.1145/3701716.3717821
15. **Arslan M., Ghanem H., Munawar S., Cruz C.** A survey on RAG with LLMs. *Procedia Computer Science*, 2024, vol. 246, pp. 3781–3790. doi:10.1016/j.procs.2024.09.178
  16. **Zhou H.** Research of text classification based on TF-IDF and CNN-LSTM. *Journal of Physics: Conference Series*, IOP Publishing, 2022, vol. 2171, no. 1, Article 012021. doi:10.1088/1742-6596/2171/1/012021
  17. **Dai B., Shen X., Wang J.** Embedding learning. *Journal of the American Statistical Association*, 2022, vol. 117, no. 537, pp. 307–319. doi:10.1080/01621459.2020.1775614
  18. **Вицентий А. В., Диковицкий В. В., Шишаев М. Г.** Технология извлечения и визуализации пространственных данных, полученных при анализе текстов. *Труды Кольского научного центра РАН*, 2020, т. 11, № 8–11, с. 115–119. doi:10.37614/2307-5252.2020.8.11.012, EDN: YLXXHN
  19. **Syed M. A., Arsevska E., Roche M., Teisseire M.** GeospatRE: Extraction and geocoding of spatial relation entities in textual documents. *Cartography and Geographic Information Science*, 2023, vol. 52, no. 3, pp. 221–236. doi:10.1080/15230406.2023.2264753
  20. **Рязанова В. А.** Множественная трактовка сложных слов в эквивалентных словосочетаниях. *Вестник Московского городского педагогического университета. Серия: Филология. Теория языка. Языковое образование*, 2021, № 3 (43), с. 143–148. doi:10.25688/2076-913X.2021.43.3.15
  21. **Klemmer K., Rolf E., Robinson C., Mackey L., Rußwurm M.** Satclip: Global, general-purpose location embeddings with satellite imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, vol. 39, no. 4, pp. 4347–4355. doi:10.1609/aaai.v39i4.32457
  22. **Cao K., Zhou C., Church R., Li X., Li W.** Revisiting spatial optimization in the era of geospatial big data and GeoAI. *International Journal of Applied Earth Observation and Geoinformation*, 2024, vol. 129, Article 103832. doi:10.1016/j.jag.2024.103832
  23. **Zhang J., Fang J., Zhang C., Zhang W., Ren H., Xu L.** Geographic named entity matching and evaluation recommendation using multi-objective tasks: A study integrating a large language model (LLM) and retrieval-augmented generation (RAG). *ISPRS International Journal of Geo-Information*, 2025, vol. 14, no. 3, p. 95. doi:10.3390/ijgi14030095
  24. **Coelho J., Magalhães J., Martins B.** Improving neural models for the retrieval of relevant passages to geographical queries *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, 2021, pp. 268–277. doi:10.1145/3474717.3483960
  25. **Абрамович Р. К., Добрынин В. Ю., Платонов А. В.** Объединение глубоких моделей и разреженных представлений в информационном поиске: обзор и анализ современных подходов. *Информационные и математические технологии в науке и управлении*, 2025, № 2 (38), с. 5–17. doi:10.25729/ESI.2025.38.2.001, EDN: HRDWSX
  26. **Chu Y., Cao H., Diao Y., Lin H.** Refined SBERT: Representing sentence BERT in manifold space. *Neurocomputing*, 2023, vol. 555, Article 126453. doi:10.1016/j.neucom.2023.126453
  27. <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2> (дата обращения: 15.11.2025).
  28. **Казакова Е. Д.** Анализ точности актуальных алгоритмов определения сообществ в сетях при работе с графовыми базами данных. *Computational Nanotechnology*, 2023, т. 10, № 1, с. 49–59. doi:10.33693/2658-3329-2023-10-1-49-59, EDN: WWHRVH
  29. **Yang M., Yang M., Wang H., Wei Z., Wang S., Wen J. R.** Efficient algorithms for personalized PageRank computation: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2024, vol. 36, no. 9, pp. 4582–4602. doi:10.1109/TKDE.2024.3376000
  30. **Lu X., Wu J., Yuan J.** Optimizing reciprocal rank with bayesian average for improved next item recommendation. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 2236–2240. doi:10.1145/3539618.3592033
  31. **Citarella A. A., Barbella M., Ciobanu M. G., De Marco F., Di Biasi L., Tortora G.** Assessing the effectiveness of ROUGE as unbiased metric in Extractive vs. Abstractive summarization techniques. *Journal of Computational Science*, 2025, vol. 87, Article 102571. doi:10.1016/j.jocs.2025.102571
  32. **Sumithmon K. S., Shajan R.** Evaluating similarity measures in collaborative filtering: Insights into accuracy, precision, and computational performance. *Georgian Education Mine*, St. George's College Aruvithura, 2024, vol. 4, no. 1, pp. 99–108.
  33. **Bellogin A., Castells P., Cantador I.** Precision-oriented evaluation of recommender systems: An algorithmic comparison. *Proceedings of the Fifth ACM Conference on Recommender Systems*, 2011, pp. 333–336. doi:10.1145/2043932.2043996
  34. **Jeunen O., Potapov I., Ustimenko A.** On (normalised) discounted cumulative gain as an off-policy evaluation metric for top-n recommendation. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 1222–1233. doi:doi.org/10.1145/3637528.36716

UDC 004.891

doi:10.31799/1684-8853-2026-2-2-15

EDN: WDEXZF

**A graph-based method for extended knowledge retrieval from specialized texts containing coordinates, geographical names, and abbreviations**A. B. Menisov<sup>a</sup>, Dr. Sc., Tech., Senior Lecturer, orcid.org/0000-0002-9955-2694, vka@mil.ruD. S. Nagorny<sup>a</sup>, Lecturer, orcid.org/0009-0008-5867-0011A. B. Shelest<sup>a</sup>, PhD, Tech., Associate Professor, orcid.org/0009-0009-1944-0229<sup>a</sup>A. F. Mozhaikskiy Military Space Academy, 13, Zhdanovskaia Emb., 197198, Saint-Petersburg, Russian Federation

**Introduction:** Current methods of advanced knowledge retrieval demonstrate limited effectiveness when processing specialized texts containing coordinates, toponyms, and abbreviations, as such data requires simultaneous consideration of semantic and spatial-structural dependencies. Traditional vector approaches focus primarily on generalized semantic proximity and poorly capture explicit spatial relationships between entities, which reduces the accuracy of search in specialized corpora. **Purpose:** To develop a method for semantic search in specialized texts containing coordinates, toponyms, and abbreviations by integrating spatial-structural information into the process of advanced knowledge retrieval. **Results:** We propose a hybrid architecture that combines entity extraction and normalization, knowledge graph construction based on spatial and semantic proximity, and a hybrid search algorithm combining graph and vector components with a configurable weighting coefficient. We have conducted experimental validation of the method on an open geospatial dataset converted into a text corpus reflecting typical spatial search scenarios. The research proves that the proposed hybrid approach demonstrates an improvement in the MAP metric as compared to vector search and significant superiority over the graph method for all metrics. However, the MRR value is inferior to vector search, indicating a trade-off between the accuracy of the first position and the stability of the overall ranking. We have established that the optimal ratio of the contributions of the graph and vector components allows increased ranking stability and response completeness when processing queries containing coordinates and geographic names. **Practical relevance:** The proposed method can be used in geoinformation analysis systems, intelligent search, and automatic processing of technical documentation. The method ensures interpretability of results due to the explicit representation of knowledge in the form of a graph, resilience to heterogeneous data, and adaptability to various types of user queries. **Discussion:** Prospects for further research are related to the use of unsupervised learning methods to reduce dependence on the quality of labeling, as well as to optimizing the computational efficiency and scalability of graph knowledge models.

**Keywords** – hybrid search, knowledge graph, Retrieval-Augmented Generation, GraphRAG, semantic analysis, spatial data, geographic information systems, coordinates, toponyms, generative AI models.

**For citation:** Menisov A. B., Nagorny D. S., Shelest A. B. A graph-based method for extended knowledge retrieval from specialized texts containing coordinates, geographical names, and abbreviations. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 2, pp. 2–15 (In Russian). doi:10.31799/1684-8853-2026-2-2-15, EDN: WDEXZF

**References**

- Grace K., Sandkühler J. F., Stewart H., Weinstein-Raun B., Thomas S., Stein-Perlman Z., Salvatier J., Brauner J., Korzekwa R. C. Thousands of AI authors on the future of AI. *Journal of Artificial Intelligence Research*, 2025, vol. 84, Article 9. doi:10.1613/jair.1.19087
- Radanliev P. Artificial intelligence: Reflecting on the past and looking towards the next paradigm shift. *Journal of Experimental & Theoretical Artificial Intelligence*, 2024, vol. 37, no. 7, pp. 1045–1062. doi:10.1080/0952813X.2024.2323042
- Vijayakumaran S., Vimaladevi M., Thangamani R., Vibeesh N., Easwaramoorthy S. V. Revolutionizing legal access: An AI-Driven RAG chatbot for real-time judicial insights. *2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)*, IEEE, 2025, pp. 1–6. doi:10.1109/AIMLA36882.2025.11040439
- Kim J., Hur M., Min M. From RAG to QA-RAG: Integrating generative AI for pharmaceutical regulatory compliance process. *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*, 2025, pp. 1293–1295. doi:10.1145/3672608.3707749
- Peng Z., Kuai X., Ke S., Dong X., Guo R. Enhancing geodatabases operability: Advanced human-computer interaction through RAG and Multi-Agent Systems. *Big Earth Data*, 2025, vol. 9, no. 2, pp. 217–242. doi:10.1080/20964471.2025.2483541
- Barry M., Caillaut G., Halftermeyer P., Qader R., Mouayad M., Le Deit F., Cariolaro D., Gesnoin J. GraphRAG: Leveraging graph-based efficiency to minimize hallucinations in LLM-driven RAG for finance data. *Proceedings of the Workshop on Generative AI and Knowledge Graphs (GenAIK)*, 2025, pp. 54–65.
- Yang R., Xue H., Razzak I., Hacid H., Salim, F. D. KG-IRAG: A knowledge graph-based iterative retrieval-augmented generation framework for temporal reasoning. *arXiv preprint*, 2025. doi:10.48550/arXiv.2503.14234
- Shalagin N. D. A survey on natural language semantic search algorithms. *International Journal of Open Information Technologies*, 2024, vol. 12, no. 9, pp. 11–21 (In Russian).
- Procko T. T., Ochoa O. Graph retrieval-augmented generation for large language models: A survey. *2024 Conference on AI, Science, Engineering, and Technology (AIxSET)*, Laguna Hills, CA, USA, 2024, pp. 166–169. doi:10.1109/AIxSET62544.2024.00030
- Nikolsky D. R., Barabanov V. F., Grebennikova N. I., Kovalenko S. A., Nuzhny A. M. Analysis of graph database management systems. *Bulletin of Voronezh State Technical University*, 2023, vol. 19, no. 6, pp. 13–20 (In Russian). doi:10.36622/VSTU.2023.19.6.002, EDN: RFOQQE
- Dang D. V., Nguyen H., Le T. T., Do H., Nguyen H., Ngo H. Q., Nguyen H. D. Knowledge graph-based legal query system with LLM and retrieval augmented generation. *Asian Conference on Intelligent Information and Database Systems: Proceedings*. Springer, Singapore, 2025, pp. 161–172. doi:10.1007/978-981-96-5881-7\_13
- Pattabiraman K., Sachin A., Lokeshkumar R. Enhancing business intelligence through NLP and contextual AI synergy. *Engineering Research Express*, 2025, vol. 7, no. 2, Article 025279. doi:10.1088/2631-8695/add9e6
- Wu J., Zhu J., Qi Y., Chen J., Xu M., Menolascina F., Jin Y., Grau V. Medical graph RAG: Evidence-based medical large language model via graph retrieval-augmented generation. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2025, pp. 28443–28467. doi:10.18653/v1/2025.acl-long.1381
- Yang R., Yang B., Zhao X., Gao F., Feng A., Ouyang S., Blum M., She T., Jiang Y., Lecue F., Lu J., Li I. Graphusion: A RAG framework for scientific knowledge graph construction with a global perspective. *Companion Proceedings of the ACM on Web Conference 2025*, 2025, pp. 2579–2588. doi:10.1145/3701716.3717821
- Arslan M., Ghanem H., Munawar S., Cruz C. A survey on RAG with LLMs. *Procedia Computer Science*, 2024, vol. 246, pp. 3781–3790. doi:10.1016/j.procs.2024.09.178
- Zhou H. Research of text classification based on TF-IDF and CNN-LSTM. *Journal of Physics: Conference Series*, IOP Publishing, 2022, vol. 2171, no. 1, Article 012021. doi:10.1088/1742-6596/2171/1/012021

17. Dai B., Shen X., Wang J. Embedding learning. *Journal of the American Statistical Association*, 2022, vol. 117, no. 537, pp. 307–319. doi:10.1080/01621459.2020.1775614
18. Vicentiy A. V., Dikovitsky V. V., Shishaev M. G. The technology of extraction and visualization of spatial data obtained by texts analysis. *Trudy Kol'skogo Nauchnogo Centra RAN*, 2020, vol. 11, no. 8–11, pp. 115–119 (In Russian). doi:10.37614/2307-5252.2020.8.11.012, EDN: YLXXHN
19. Syed M. A., Arsevska E., Roche M., Teisseire M. GeospatRE: Extraction and geocoding of spatial relation entities in textual documents. *Cartography and Geographic Information Science*, 2023, vol. 52, no. 3, pp. 221–236. doi:10.1080/15230406.2023.2264753
20. Ryazanova V. A. Multiple interpretations of complex words in equivalent word combinations. *MCU Journal of Philology. Theory of Linguistics. Linguistic Education*, 2021, no. 3 (43), pp. 143–148 (In Russian). doi:10.25688/2076-913X.2021.43.3.15
21. Klemmer K., Rolf E., Robinson C., Mackey L., Rußwurm M. Satclip: Global, general-purpose location embeddings with satellite imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, vol. 39, no. 4, pp. 4347–4355. doi:10.1609/aaai.v39i4.32457
22. Cao K., Zhou C., Church R., Li X., Li W. Revisiting spatial optimization in the era of geospatial big data and GeoAI. *International Journal of Applied Earth Observation and Geoinformation*, 2024, vol. 129, Article 103832. doi:10.1016/j.jag.2024.103832
23. Zhang J., Fang J., Zhang C., Zhang W., Ren H., Xu L. Geographic named entity matching and evaluation recommendation using multi-objective tasks: A study integrating a large language model (LLM) and retrieval-augmented generation (RAG). *ISPRS International Journal of Geo-Information*, 2025, vol. 14, no. 3, p. 95. doi:10.3390/ijgi14030095
24. Coelho J., Magalhães J., Martins B. Improving neural models for the retrieval of relevant passages to geographical queries. *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, 2021, pp. 268–277. doi:10.1145/3474717.3483960
25. Abramovich R. K., Dobrynin V. Y., Platonov A. V. Combining deep language models and sparse vector representations in information retrieval: A review and analysis of modern approaches. *Information and Mathematical Technologies in Science and Management*, 2025, no. 2 (38), pp. 5–17 (In Russian). doi:10.25729/ESI.2025.38.2.001, EDN: HRDWSX
26. Chu Y., Cao H., Diao Y., Lin H. Refined SBERT: Representing sentence BERT in manifold space. *Neurocomputing*, 2023, vol. 555, Article 126453. doi:10.1016/j.neucom.2023.126453
27. Hugging Face. *Sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2*. Available at: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2> (accessed 15 November 2025).
28. Kazakova E. D. Analysis of the modern algorithms' accuracy for communities identification on networks when working with graph databases. *Computational Nanotechnology*, 2023, vol. 10, no. 1, pp. 49–59 (In Russian). doi:10.33693/2658-3329-2023-10-1-49-59, EDN: WWHRVH
29. Yang M., Yang M., Wang H., Wei Z., Wang S., Wen J. R. Efficient algorithms for personalized PageRank computation: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2024, vol. 36, no. 9, pp. 4582–4602. doi:10.1109/TKDE.2024.3376000
30. Lu X., Wu J., Yuan J. Optimizing reciprocal rank with bayesian average for improved next item recommendation. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 2236–2240. doi:10.1145/3539618.3592033
31. Citarella A. A., Barbella M., Ciobanu M. G., De Marco F., Di Biasi L., Tortora G. Assessing the effectiveness of ROUGE as unbiased metric in Extractive vs. Abstractive summarization techniques. *Journal of Computational Science*, 2025, vol. 87, Article 102571. doi:10.1016/j.jocs.2025.102571
32. Sumithmon K. S., Shajan R. Evaluating similarity measures in collaborative filtering: Insights into accuracy, precision, and computational performance. *Georgian Education Mine*, St. George's College Aruvithura, 2024, vol. 4, no. 1, pp. 99–108.
33. Bellogin A., Castells P., Cantador I. Precision-oriented evaluation of recommender systems: An algorithmic comparison. *Proceedings of the Fifth ACM Conference on Recommender Systems*, 2011, pp. 333–336. doi:10.1145/2043932.2043996
34. Jeunen O., Potapov I., Ustimenko A. On (normalised) discounted cumulative gain as an off-policy evaluation metric for top-n recommendation. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 1222–1233. doi:doi.org/10.1145/3637528.36716