

Исследовательский проект

Тема исследования:	Разработка средств анализа ответов в конвейере RAG с применением графов
Студент	Гладышев Виталий Владимирович Науки о данных

Аннотация

Большие языковые модели (LLM) – важный технологический прорыв современности. Для практического применения этих моделей на практике используется технология RAG. Качественная реализация RAG на продуктивном уровне – сложная задача требующая точного представления о влиянии изменений на результат.

Метрики оценки конвейера RAG имеют ограничения:

- Часто применяется подход использования LLM как судьи. Соответственно оценки непрозрачны и сложны для интерпретации
- Традиционные метрики (ROUGE, BLEU) не оценивают структурную согласованность ответов
- Ошибки RAG-систем в критичных доменах

Исследование заполняет пробел между теоретическими работами по графам и прикладными задачами RAG, предлагая метод, который одновременно улучшает точность и прозрачность оценки. Графы, в отличие от векторов, сохраняют структурные и семантические связи между текстом, изображениями и знаниями.

Применение GraphRAG, метрик структурного анализа и методов валидации позволяет не только оценить качество RAG-систем, но и выявить конкретные источники ошибок на этапах извлечения и генерации. Это напрямую способствует достижению цели исследования — повышению прозрачности и точности оценки конвейеров RAG через семантическое моделирование.

Ключевые слова: граф, семантика, RAG (retrieval-augmented generation), генерация, интерпретируемость, метрика, контекст

Общее описание исследования

Цель и задачи, исследовательский вопрос

Цель исследования - реализовать средство анализа ответов в конвейере RAG с применением графов для обеспечения интерпретируемости и прозрачности определения влияния семантики ответа на количественные характеристики

Задачи исследования

1. Интеграция RAG с графовыми подходами – Изучить, как комбинированные методы, использующие графы и RAG, могут улучшить качество генерации и понимания текстов в различных доменах.
2. Применить и развить подход реализованный в библиотеке GraphRAG применительно к конвейеру RAG для численного анализа семантических признаков вопроса пользователя и контекста получаемого семантическим поиском

Объект исследования: Конвейер RAG (retrieval-augmented generation).

Предмет исследования: Семантическое моделирование с помощью графов для оценки качества RAG.

Исследовательский вопрос: Как графовые представления влияют на объяснимость и точность оценки этапов извлечения и генерации в RAG?

Проблема

Недостаточная объяснимость традиционных методов оценки RAG, основанных на векторных представлениях, затрудняет анализ ошибок в retrieval и генерации. В литературе обсуждаются: преимущества графов для структурирования контекста [30], важность интерпретируемости в NLP [11]. Однако не исследовано, как графы улучшают прозрачность оценки RAG.

Новизна исследования

Исследование источников выявило наличие теоретических лакун. Отсутствие унифицированных метрик для оценки графов в контексте RAG. Большинство работ [1, 2] предлагают общие метрики сравнения графов, но не адаптируют их для оценки качества извлечения и генерации в RAG.

Выявлен недостаток исследований по интеграции графовых нейронных сетей (GNN) с RAG. Хотя GNN активно применяются в NLP [11], их роль в улучшении объяснимости оценки RAG остаётся неисследованной. Обнаружена слабая связь между структурой графов и семантикой контекста. Работы по текстовым графам [14] фокусируются на классификации, но не на интерпретации связей для оценки качества генерации.

Обнаружен недостаток средств анализа и визуализации - практические лакуны. Отсутствие инструментов для визуализации и анализа графов в RAG. Методы вроде graphkernels [6] не интегрированы в конвейеры RAG, что затрудняет их применение для объяснимой оценки. Ограниченное использование графов для улучшения фазы извлечения. Исследования RAG [30] акцентируют генерацию, игнорируя оптимизацию retrieval через графовые представления. Нет стандартизированных датасетов для тестирования графовых подходов в RAG. Существующие бенчмарки [22] не включают аннотированные графовые структуры, что усложняет валидацию.

Данные лакуны должны быть закрыты в ходе моего исследования. Предусмотрена разработка специализированного аналитического средства оценки графов для RAG, сочетающих структурный и семантический анализ включая создание инструмента визуализации, связывающего графовые представления контекста и вопроса пользователя.

Актуальность исследования

Контекст исследования

Рост использования RAG-систем в критически важных областях (медицина, финансы, юридический анализ) требует повышения доверия к их результатам. Однако, как отмечают [28], большинство пользователей сомневаются в обоснованности решений, сгенерированных LLM, из-за низкого уровня интерпретируемости. В научной сфере актуальность подтверждается увеличением числа публикаций по RAG кратно за 2023–2024 гг. (по данным Semantic Scholar), а также ростом интереса к графовым методам в NLP.

Своевременность исследования

Исследование заполняет пробел между теоретическими работами по графам [1] и прикладными задачами RAG [30], предлагая метод, который одновременно улучшает точность и прозрачность оценки. Появление

мультимодальных LLM (например, GPT-4o) требует новых методов оценки, способных работать с гетерогенными данными. Графы, в отличие от векторов, сохраняют структурные и семантические связи между текстом, изображениями и знаниями [23].

Обзор литературы

При исследовании источников выявлено несколько ключевых направлений для проведения исследования. Важным аспектом является изучение метрик и методов для сравнения графов, что служит основой для дальнейшего анализа в разных областях их применения. Графовые нейронные сети играют важную роль в обработке текстов и их анализе, существенно повышая качество извлечения информации и выполнения классификации. Также анализ структуры графов способствует более глубокому пониманию сложных взаимосвязей в данных, что критично для задач, связанных с RAG. Исследование комплексных графовых моделей, включая взвешенные и ориентированные графы, открывает новые возможности для их адаптации в задачах генерации текста и извлечения информации. Наконец, интеграция методов RAG с подходами, основанными на графах, способна повысить качество генерации и понимания текстов во множестве доменных областей.

Анализ источников

Metrics for Graph Comparison: A Practitioner's Guide (Wills & Meyer, 2020). Основная идея статьи заключается в представлении различных метрик для сравнения графов, что позволяет исследователям и практикам выбирать подходящие методы для их исследований. Эта работа важна для моего исследования, поскольку правильная оценка графов имеет критическое значение для оценки их эффективности.

Efficient Graphlet Kernels for Large Graph Comparison (Shervashidze et al., 2009). Авторы описывают эффективные графлетные ядра, которые применяются для сравнительного анализа больших графов и предлагают алгоритмы, способные эффективно обрабатывать большие объемы данных. Статья будет полезна для моего исследования, так как позволяет интегрировать методы сравнения графов.

A Fair Comparison of Graph Neural Networks for Graph Classification (Errica et al., 2019). Статья сравнивает различные графовые нейронные сети, оценивая их эффективность для задач классификации графов. Важность этого исследования для меня заключается в том, что понимание различных архитектур нейронных сетей поможет в выборе оптимального подхода для интеграции в систему RAG.

Graph Comparison Using Fine Structure Analysis (Macindoe & Richards, 2010). В работе предложен метод сравнения графов с использованием анализа их тонкой структуры, что может выявить скрытые паттерны и различия. Эта статья имеет значение для моего проекта, так как методы тонкой структуры могут улучшить качество извлечения информации.

Weighted Graph Comparison Techniques for Brain Connectivity Analysis (Alper et al., 2013). Работа посвящена техникам сравнения взвешенных графов, применяемым к анализу связей в мозге, что демонстрирует методологию, полезную для анализа сложных систем. Статья актуальна для моего исследования, так как она иллюстрирует концепцию применения графов, которая может быть адаптирована для текстовых данных.

Graph Neural Networks for Natural Language Processing: A Survey (Wu et al., 2023). Авторы рассматривают применение графовых нейронных сетей в областях обработки естественного языка, делая акцент на эффективных методах. Этот обзор полезен, поскольку включает в себя передовые подходы, которые могут применяться к задачам RAG и анализа текстов.

G-Retriever: Retrieval-Augmented Generation for Textual Graph Understanding and Question Answering (He et al., 2024). В этом исследовании предлагается метод G-Retriever, который использует RAG для понимания текстов на основе графов и отвечает на вопросы. Статья критически важна, так как она непосредственно связана с исследуемыми мною методами и может помочь разработать более эффективные подходы.

Гипотеза исследования

Использование семантических графов для моделирования контекста в RAG-конвейере реализовать сравнение семантической структуры вопроса пользователя и контекста с привязкой к отдельным элементам вопроса и контекста в отличие от существующих метрик таких как Faithfulness или Answer

Assurasy. Соответственно можно связать количественные параметры со конкретными элементами семантической структуры.

Методология исследования

Разработан детальный план проверки гипотезы о влиянии семантических графов на объяснимость и точность оценки этапов извлечения и генерации в RAG-конвейере. Определены методы сбора и анализа данных, описан процесс эксперимента, а также обоснована связь выбранных подходов с решением поставленных исследовательских задач. В рамках исследования будут использованы смешанные методы, сочетающие количественные метрики структурного анализа графов и качественную интерпретацию их семантических связей.

Методы исследования

Основным методом исследования станет адаптация подхода GraphRAG (Edge et al., 2024), который предполагает автоматическое построение графов знаний из неструктурированного текста с использованием больших языковых моделей (LLM). Этот метод был выбран по нескольким причинам. Во-первых, он позволяет преобразовывать текстовый контекст и ответы RAG в семантические графы, узлы которых соответствуют сущностям (например, «приказ на отпуск»), а рёбра — связям между ними (например, «оформить»). Во-вторых, GraphRAG позволяет выполнить формирование графа знаний в удобной форме (сущности, связи, комьюнити и другие элементы доступны в виде parquet-файлов) с высокой степенью автоматизации. Формирование необходимых параметров также может быть автоматизировано.

Для количественной оценки различий между графами будет использована метрика Graph Edit Distance (GED), измеряющая минимальное количество операций (добавление, удаление, замена узлов/рёбер), необходимых для преобразования одного графа в другой. Эта метрика позволяет объективно оценить, насколько структура ответа RAG соответствует эталонному контексту. Дополнительно будет проведён анализ совпадения узлов и рёбер, что поможет выявить семантические расхождения (например, пропущенные сущности).

Кроме того, для валидации результатов будет применяться метод проверки подграфа: если граф ответа является подграфом эталонного контекста, это

свидетельствует о структурной согласованности. Данный подход дополняет традиционные метрики (ROUGE, BLEU), которые фокусируются на поверхностном сходстве текстов, игнорируя семантику.

Применение этих методов позволит закрыть пробел между теоретическими работами по графам и прикладными задачами RAG. Например, в отличие от исследований, где графы используются для классификации текстов, данный проект адаптирует их для оценки качества генерации.

Данные для исследования

Для исследования будут использоваться два вида данных:

1. Реальные данные из датасета Vikhrmodels/Grounded-RAG-RU-v2, содержащего вопросы на русском языке, контексты и эталонные ответы. Этот датасет охватывает разнообразные домены, что соответствует актуальности исследования для различных областей.
2. Синтетические данные, сгенерированные с помощью LLM (например, GPT-4o) для моделирования контролируемых аномалий: отсутствие ключевых сущностей, некорректные связи между узлами, избыточные элементы. Такие данные помогут проверить чувствительность метрик к структурным ошибкам.

Для обработки данных потребуются следующие ресурсы:

- Большие языковые модели (Qwen, Gemma, Saiga, Vikhr) — для извлечения сущностей и построения графов.
- Библиотеки NLP (SpaCy, Natasha) — для токенизации и лемматизации текста.
- Инструменты анализа графов (NetworkX, PyTorch Geometric) — для расчёта GED и визуализации структур.

Ожидаемые результаты

Количественный анализ будет направлен на установление корреляции между структурными метриками графов и субъективными оценками качества ответов. Например, если высокое значение GED соответствует низким оценкам точности (по данным экспертов), это подтвердит гипотезу о связи структурных различий с качеством генерации.

Качественный анализ сосредоточится на интерпретации конкретных ошибок. Визуализация графов позволит выявить паттерны: например, отсутствие узла «заявление» в тестовом графе может объяснить, почему RAG-система неверно ответила на вопрос о процедуре оформления отпуска. Такая интерпретация недоступна при использовании традиционных метрик вроде Answer Accuracy, которые лишь констатируют факт ошибки, но не объясняют её причин.

По итогам исследования будут разработаны:

1. Библиотека для анализа графов в RAG, поддерживающая расчёт GED, визуализацию и экспорт результатов в форматы JSON и CSV.
2. Аннотированные датасеты, включающие эталонные графы для контекстов и ответов RAG, что упростит воспроизведение экспериментов.
3. Методические рекомендации по интеграции графовых метрик в существующие конвейеры RAG.

План реализации

Эксперимент будет проводиться в четыре этапа:

1. Построение эталонных графов. Исходные контексты из датасета преобразуются в графы с помощью GraphRAG. Например, текст «Сотрудник должен подать заявление за две недели до отпуска» будет представлен как граф с узлами «сотрудник», «заявление», «отпуск» и рёбро «подать».
2. Генерация тестовых графов. Ответы, полученные от RAG-системы, аналогичным образом преобразуются в графы. На этом этапе также будут созданы синтетические данные с преднамеренными ошибками для тестирования устойчивости метода.
3. Расчёт метрик. Для каждой пары «эталонный граф — тестовый граф» вычисляются GED, процент совпадения узлов/рёбер и проводится проверка на подграф.
4. Оптимизация. Графовые алгоритмы будут адаптированы для потоковой обработки данных, что повысит скорость анализа.

Инструментальная часть эксперимента включает написание Python-скриптов для автоматизации построения графов и расчёта метрик, а также использование Jupyter Notebook или Streamlit/Gradio UI для интерактивной визуализации результатов.

Этапы реализации

Этап	Период	Задачи	Ожидаемые результаты
1. Подготовительный	август 2025	Анализ литературы, настройка среды разработки, получение доступа к LLM API	Техническое задание, обзор литературы
2. Адаптация GraphRAG	сентябрь 2025	Адаптация GraphRAG для русского языка, тестирование извлечения сущностей	Модуль построения графов
3. Подсчёт метрик	октябрь 2025	Реализация GED, метрик совпадения, валидация на тестовых данных	Библиотека для вычисления метрик
4. Разработка инструмента	ноябрь - декабрь 2025	UI для визуализации, интеграция компонентов, тестирование	Готовый аналитический инструмент
5. Экспериментальная валидация	январь - февраль 2026	Сбор экспертных оценок, статистический анализ, тестирование гипотез	Экспериментальные результаты
6. Финализация	март - апрель 2026	Документация, методические рекомендации, подготовка к публикации	Итоговые материалы исследования

Контрольные точки:

- Сентябрь: Генерация графов для 50 примеров
- Октябрь: Расчёт метрик примеров
- Декабрь: Рабочий прототип инструмента
- Февраль: Подтверждение гипотезы

Список литературы

1. Wills P., Meyer F. G. Metrics for graph comparison: a practitioner's guide //Plos one. – 2020. – Т. 15. – №. 2. – С. e0228728.
2. Shervashidze N. et al. Efficient graphlet kernels for large graph comparison //Artificial intelligence and statistics. – PMLR, 2009. – С. 488-495.
3. Errica F. et al. A fair comparison of graph neural networks for graph classification //arXiv preprint arXiv:1912.09893. – 2019.
4. Macindoe O., Richards W. Graph comparison using fine structure analysis //2010 IEEE Second International Conference on Social Computing. – IEEE, 2010. – С. 193-200.
5. Alper B. et al. Weighted graph comparison techniques for brain connectivity analysis //Proceedings of the SIGCHI conference on human factors in computing systems. – 2013. – С. 483-492.
6. Sugiyama M. et al. graphkernels: R and Python packages for graph comparison //Bioinformatics. – 2018. – Т. 34. – №. 3. – С. 530-532.
7. Ray S. S. Graph theory with algorithms and its applications: in applied science and technology. – New Delhi : Springer, 2013.
8. Lovász L., Plummer M. D. Matching Theory, vol. 29 //Annals of Discrete Mathematics, North-Holland, Amsterdam. – 1986.
9. Cormen T. H. et al. Introduction to Algorithms, chapter 26. – 2001..
10. Tantardini M. et al. Comparing methods for comparing networks //Scientific reports. – 2019. – Т. 9. – №. 1. – С. 17557.
11. Wu L. et al. Graph neural networks for natural language processing: A survey //Foundations and Trends® in Machine Learning. – 2023. – Т. 16. – №. 2. – С. 119-328.
12. Wang K., Ding Y., Han S. C. Graph neural networks for text classification: A survey //Artificial Intelligence Review. – 2024. – Т. 57. – №. 8. – С. 190.

13. Sonawane, S.S., Mahalle, P.N., & Ghotkar, A. (2022). Text Analytics Using Graph Theory. Studies in Big Data.
14. Huang L. et al. Text level graph neural network for text classification //arXiv preprint arXiv:1910.02356. – 2019.
15. Das N. et al. Comparison of different graph distance metrics for semantic text based classification //Polibits. – 2014. – T. 49. – C. 51-57.
16. Castillo E., Cervantes O., Vilarino D. Text analysis using different graph-based representations //Computación y Sistemas. – 2017. – T. 21. – №. 4. – C. 581-599.
17. Petersen C. et al. Entropy and graph based modelling of document coherence using discourse entities: An application to IR //Proceedings of the 2015 International Conference on The Theory of Information Retrieval. – 2015. – C. 191-200.
18. Rousseau F., Kiagias E., Vazirgiannis M. Text categorization as a graph classification problem //Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). – 2015. – C. 1702-1712.
19. Koncel-Kedziorski R. et al. Text generation from knowledge graphs with graph transformers //arXiv preprint arXiv:1904.02342. – 2019.
20. Zhang Y. et al. Every document owns its structure: Inductive text classification via graph neural networks //arXiv preprint arXiv:2004.13826. – 2020.
21. Hu Y. et al. CG-RAG: Research Question Answering by Citation Graph Retrieval-Augmented LLMs //arXiv preprint arXiv:2501.15067. – 2025.
22. Auer S. et al. The sciq scientific question answering benchmark for scholarly knowledge //Scientific Reports. – 2023. – T. 13. – №. 1. – C. 7240.
23. Peng C. et al. Knowledge graphs: Opportunities and challenges //Artificial Intelligence Review. – 2023. – T. 56. – №. 11. – C. 13071-13102.
24. Huang X. et al. Knowledge graph embedding based question answering //Proceedings of the twelfth ACM international conference on web search and data mining. – 2019. – C. 105-113.
25. He X. et al. G-retriever: Retrieval-augmented generation for textual graph understanding and question answering //Advances in Neural Information Processing Systems. – 2024. – T. 37. – C. 132876-132907.

26. Yasunaga M. et al. QA-GNN: Reasoning with language models and knowledge graphs for question answering //arXiv preprint arXiv:2104.06378. – 2021.
27. Zheng W. et al. Question answering over knowledge graphs: question understanding via template decomposition //Proceedings of the VLDB Endowment. – 2018. – T. 11. – №. 11. – C. 1373-1386.
28. Sui Y. et al. Can knowledge graphs make large language models more trustworthy? an empirical study over open-ended question answering //arXiv preprint arXiv:2410.08085. – 2024.
29. Ibrahim N. et al. A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): models, evaluation metrics, benchmarks, and challenges //Discover Artificial Intelligence. – 2024. – T. 4. – №. 1. – C. 76.
30. Edge D. et al. From local to global: A graph rag approach to query-focused summarization //arXiv preprint arXiv:2404.16130. – 2024.