

Домашнее задание 2. Валидация списка источников

Гладышев Виталий Владимирович

1 Список баз данных, профильных журналов, сборников, научных издательств

1. Arxiv.org - Платформа для размещения препринтов научных статей в различных областях науки, включая математику, физику, информатику и многое другое;
2. Google Scholar - Поисковая система, обеспечивающая доступ к научным статьям, книгам, диссертациям и другим академическим материалам;
3. ResearchGate - Социальная сеть для ученых и исследователей, где они могут делиться своими работами, задавать вопросы и находить коллег по интересам;
4. PLOS One - Журнал, публикующий результаты исследований в общем и прикладном направлениях наук;
5. Journal of Machine Learning Research (JMLR) - Профильный журнал по машинному обучению;
6. Scientific Reports - Мультидисциплинарный журнал, публикующий результаты оригинальных исследований;
7. Artificial Intelligence Review - Журнал, специализирующийся на теоретических и прикладных аспектах искусственного интеллекта;
8. IEEE Xplore - База данных для публикаций по электронике и компьютерным наукам;
9. SpringerLink - Издательство научной литературы, охватывающее различные области науки и техники;
10. Papers with Code - Платформа, объединяющая научные публикации и соответствующий код для воспроизводимости исследований;
11. Semantic Scholar - Инструмент для поиска научной информации, использующий искусственный интеллект для улучшения поиска и анализа статей.

2 Список ключевых слов

Для темы исследования "Применение графов для оценки конвейера RAG retrieval augmented generation" можно выделить следующие ключевые слова:

1. Графы

- Графовая структура;
- Графовые представления;
- Графовая модель.

2. RAG (Retrieval-Augmented Generation)

- Фаза извлечения;
- Обогащение контекста;
- Фаза генерации.

3. Метрики

- Сравнение;
- Оценивание;
- Рейтинг.

4. Извлечение информации

- Извлечение знаний;
- Информационный поиск.

5. Текстовая классификация

- Классификация текста;
- Анализ текста;
- Классификация документов.

6. Нейронные сети

- Графовые нейронные сети;
- Нейронные модели.

3. Список источников (публикаций)

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. – Т. 49. – С. 51-57.
16. Castillo E., Cervantes O., Vilarino D. Text analysis using different graph-based representations //Computación y Sistemas. – 2017. – Т. 21. – №. 4. – С. 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. – С. 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 sciqa 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.

4. Классификация публикаций по типам

4.1 Теоретические исследования

- Alper, B. et al., Weighted graph comparison techniques for brain connectivity analysis;
- Ray, S. S., Graph Theory with Algorithms and its Applications.

4.2 Эмпирические данные / исследования

- Wills, P. et al., Metrics for graph comparison: A practitioner's guide;
- Hu, Y. et al., CG-RAG: Research Question Answering by Citation Graph Retrieval-Augmented LLMs.

4.3 Метаанализы

- Wu, L. et al., Graph neural networks for natural language processing: A survey;
- Wang, K., et al. Graph neural networks for text classification: A survey.

4.4 Разработка и использование методик и исследовательского инструментария

- Sugiyama, M. et al., graphkernels: R and Python packages for graph comparison;
- Edge, D. et al. G-Retriever: Retrieval-Augmented Generation for Textual Graph Understanding and Question Answering.

5. Краткий обзор семи ключевых источников

5.1 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.

5.2 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.

5.3 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.

5.4 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.

5.5 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.

5.6 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.

5.7 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.

6. Ключевые теоретические концепты

На основе краткого обзора источников выделяются следующие теоретические концепты, которые будут развиваться в рамках моего исследования:

1. Метрики и методы сравнения графов - Изучение различных метрик, необходимых для оценки характеристик графов, является основой для последующих анализов в различных приложениях;
2. Графовые нейронные сети - Использование графовых нейронных сетей для обработки и анализа текстов, что позволяет значительно улучшить качество извлечения информации и классификации;
3. Анализ структуры графов - Применение этой концепции может привести к лучшему пониманию сложных взаимосвязей в данных, что важно для задач RAG;
4. Комплексные графовые модели - Исследование взвешенных и ориентированных графов, которые могут быть адаптированы для различных прикладных задач, включая генерацию текста и извлечение информации;
5. Интеграция RAG с графовыми подходами - Изучение того, как комбинированные методы, использующие графы и RAG, могут улучшить качество генерации и понимания текстов в различных доменах.