

ПОВЫШЕНИЕ КАЧЕСТВА GRAPH RAG ЗА СЧЕТ ИСПОЛЬЗОВАНИЯ ОНТОЛОГИЙ

Рассматривается роль онтологий в системах Graph RAG. Обосновано их положительное влияние на качество индексации и точность поиска ответов. Приведена архитектура интеграции онтологий в процесс обработки документов, а также примеры успешных реализаций подхода.

Введение

Большие языковые модели (LLM) эффективны при анализе текстов, однако подвержены галлюцинациям — генерации выдуманных фактов. Они часто упускают детали в объемных текстах и теряют контекст при превышении лимита токенов. При этом классический RAG плохо подходит для ответов на глобальные вопросы, требующие агрегации информации сразу из нескольких документов. Решением выступает Graph RAG — гибридный подход, объединяющий граф знаний и LLM.

Использование онтологий в Graph RAG крайне актуально, поскольку они задают строгую схему предметной области. Это позволяет перейти от извлечения разрозненных троек (субъект-предикат-объект) к построению надежной, семантически связанной базы знаний.

I. Архитектура системы и интеграция онтологий

Для реализации онтологического Graph RAG предлагается архитектура, в которой базовым источником правил выступает расширяемая онтология предметной области. Первичная схема создается экспертом в среде визуального проектирования Protégé. В форматах OWL она загружается в систему, проходит валидацию на отсутствие конфликтов (например, циклических зависимостей) и сохраняется в графовой СУБД. Архитектура программно предусматривает возможность регулярного расширения онтологии, реализуемого с помощью алгоритма слияния.

На этапе индексации документов онтология выступает жестким фильтром. Добавление новой информации в граф происходит после устранения дублирования синонимических сущностей. Опираясь на классы и свойства онтологии, переданные в промпте, LLM генерирует первичный вариант сущностей и отношений на основе текста. Затем предварительный набор кандидатов уточняется путем поиска совпадений в локальном реестре документа, после чего выполняется векторный поиск семантически близких узлов по всей СУБД и итоговое слияние синонимов.

На этапе формирования ответа онтологии строго ограничивают типы извлекаемых сущностей, а для определения релевантных узлов применяется алгоритм Personalized PageRank. Граф со встроенной онтологической схемой также позволяет осуществлять формальный логический вывод, извлекая скрытые связи за счет транзитивности и иерархии классов.

Для генерации качественных ответов на общие вопросы в системе применяется алгоритм Лейдена. Он выделяет сообщества в графе, для каждого из которых LLM формирует описание, используемое при поиске. Анализ сообществ позволяет системе определить, какие новые обобщенные классы полезно добавить в онтологию при её последующем расширении.

II. Выводы

Эффективность онтологий подтверждается на практике. В системе OG-RAG [1] использование домен-специфических схем повысило полноту (recall) ответов на 55 % и их фактическую точность на 27 %. Система Wikontic RAG [2] позволяет строить граф, расходуя в 10–20 раз меньше токенов за счет опоры на онтологию Wikidata. Активно развивается фреймворк KAG [3], использующий иерархию знаний для многошаговых рассуждений.

Внедрение расширяемых онтологий выводит Graph RAG на новый качественный уровень. Интеграция символического и нейросетевого подходов повышает точность извлечения структурированных данных и значительно снижает вероятность галлюцинаций.

Список литературы

1. Шарма, К. OG-RAG: Ontology-Grounded Retrieval-Augmented Generation For Large Language Models / К. Шарма [и др.] // arXiv preprint arXiv:2412.15235. – 2024.
2. Чепурова, А. Wikontic: Constructing Wikidata-Aligned, Ontology-Aware Knowledge Graphs with Large Language Models / А. Чепурова [и др.] // arXiv preprint arXiv:2512.00590. – 2025.
3. Лян, Л. KAG: Boosting LLMs in Professional Domains via Knowledge Augmented Generation / Л. Лян [и др.] // arXiv preprint arXiv:2409.13731. – 2024.

Дичковский Владимир Андреевич, студент кафедры интеллектуальных информационных технологий БГУИР, volodya.dichkovski@gmail.com.

Научный руководитель: Ковалёв Михаил Владимирович, магистр технических наук, старший преподаватель кафедры интеллектуальных информационных технологий БГУИР, michail.kovalev7@gmail.com.