

Д31 «Тема, цель, исследовательский вопрос»

Гладышев В.В.

1 Тема и проблема исследования

1.1 Тема исследования

Темой исследования является повышение эффективности применения генеративных моделей посредством интеграции графовых структур данных в генерации кода, запросах к реляционным базам и Retrieval Augmented Generation (RAG).

1.2 Проблема исследования

Современные методы генерации текста, кода и запросов к базам данных сталкиваются с ограничениями при работе со сложными контекстами, требующими учета многоуровневых связей между сущностями. Традиционные подходы к структурированию информации (например, векторные представления или промпт инжиниринг) недостаточно эффективно организуют контекстную информацию в задачах со сложной структурой данных, например:

1. Генерация кода для крупных проектов с множеством взаимозависимостей;
2. Формирование нетривиальных запросов к объемным реляционным базам данных;
3. Подготовка релевантных данных для RAG-систем, где критична точность извлечения и агрегации контекста.

Это приводит к ошибкам в генерации (галюцинациям), таким как нарушение логических связей, неучтенные зависимости или избыточность информации.

1.2.1 Аргументация актуальности и значимости

Современные LLM (Large Language Models) активно внедряются в разработку ПО, анализ данных и диалоговые системы, но их эффективность снижается при работе с данными значительно превышающими размер окна контекста модели. Графовые структуры, способные явно моделировать связи между объектами, предлагают потенциал для решения этой проблемы.

Системы типа GraphRAG демонстрируют, что графы повышают точность поиска контекста за счет учета семантических и структурных связей. Однако отсутствуют универсальные методики их интеграции в различные генеративные задачи.

Практическая значимость:

1. В генерации кода графы могут улучшить понимание архитектуры проекта, сократив ошибки в автоматизированной разработке.
2. В запросах к базам данных графовое представление схемы позволяет точнее формулировать операции объединения и агрегации.
3. В RAG графы способны оптимизировать этап retrieval, обеспечивая более релевантный контекст для генерации ответов.

1.2.2 Научная новизна

Научная новизна исследования заключается в определении значимости вклада графовой организации структуры информации для повышения качества решения задач генерации кода, генерации запросов к базам данных, решения задачи RAG.

2 Цель исследования

Разработать методологию оценки вклада графовых структур данных в генеративные модели для повышения качества генерации кода, SQL-запросов и контекста в RAG-системах.

3 Задачи исследования

1. Провести сравнительный анализ существующих подходов к организации контекстной информации в генеративных моделях (векторные представления, графы, промпт инжиниринг) и выявить их ограничения в задачах с высокой структурной сложностью.
2. Определить доступные средства преобразования контекстной информации в графовое представление, для трёх целевых областей: генерация кода, SQL-запросов и RAG.
3. Реализовать конвейер оценки систем для с применением графовых структур для формирования контекста и аналогичных систем без применения графовых структур.

4. Выполнить оценку вклада применения графовых структур для решения задач генерации кода, сложных SQL запросов и для решения задачи RAG.
5. Провести анализ и предложить пути повышения эффективности организации контекста больших языковых моделей с применением графовых структур данных.

4 Объект и предмет исследования

4.1 Объект исследования

Объектом исследования является процесс генерации структурированных данных (код, SQL-запросы, контекст для RAG) с использованием языковых моделей.

4.2 Предмет исследования

Предметом исследования являются методы организации контекстной информации на основе графовых структур для оптимизации генеративных моделей.

5 Исследовательский вопрос

Формулировка исследовательского вопроса: как интеграция графовых структур данных в этапы контекстного представления влияет на качество генерации в задачах с высокой структурной сложностью (код, SQL, RAG), и какие типы графовых моделей (семантические, синтаксические, онтологические) наиболее эффективны для различных доменов?

Вопрос фокусируется на установлении причинно-следственных связей между выбором графовой модели (причина) и метриками качества генерации (следствие).