

РАЗРАБОТКА МЕТОДА ОЦЕНКИ СЕМАНТИЧЕСКОЙ ПОЛНОТЫ ГРАФОВ ЗНАНИЙ НА ОСНОВЕ ДОМЕНА НАУЧНОЙ ЛИТЕРАТУРЫ

Гинева Д.С.¹

Научный руководитель: Томилов А.А.²

¹Университет ИТМО, ²ООО «ЦРТ»

forestscout.95@gmail.com

Введение

Существующие подходы к оценке полноты графов знаний преимущественно ориентированы на структурные характеристики графа и не всегда учитывают особенности представления семантики предметной области. В связи с этим возникает необходимость разработки методов, позволяющих оценивать именно семантическую полноту графов знаний [1, 2]. В данной работе предлагается разработка метода оценки семантической полноты графов знаний для домена научной литературы с использованием графовой СУБД Neo4j. Целью исследования является повышение семантической полноты графов знаний, построенных на основе корпуса научной литературы.

Основная часть

В работе предлагается метод оценки семантической полноты графов знаний в домене научной литературы, ориентированный на анализ представленности ключевых семантических структур предметной области [3]. Такие структуры отражают характерные паттерны представления научного знания, включающие взаимосвязи между публикациями, авторами и научными понятиями. Неполное представление данных структур в графе интерпретируется как проявление семантической неполноты. Метод предполагает использование графовой базы знаний, построенной на основе научных текстов с применением графовой СУБД Neo4j, и включает анализ распределения сущностей и отношений в графе. Предлагаемый метод направлен на выявление и устранение пропусков в типовых семантических структурах научного знания.

Для повышения устойчивости оценки вводится процедура калибровки на основе семантических инвариантов. Под семантическими инвариантами понимаются устойчивые структурно-содержательные элементы научного текста, которые должны сохраняться независимо от вариаций формулировок. Использование таких инвариантов позволяет снизить влияние лингвистических различий в текстах на результаты оценки полноты. Анализ графов знаний, построенных на основе исходных текстов и их парафразированных версий, позволяет выявлять устойчивые семантические элементы и использовать их в качестве калибровочных ориентиров для оценки полноты.

Для оценки качества предложенного метода был проведён сравнительный анализ с использованием метода MINE-1, применяемого для проверки сохранения фактов исходного текста в графе знаний [4]. Полученные результаты показали, что использование разработанного метода позволяет выявлять области семантической неполноты графа, которые не обнаруживаются при использовании только методов анализа сохранения фактов. Проверка качества оценки с использованием MINE-1 показала повышение показателей сохранения информации в графе знаний после применения предложенного подхода к анализу и корректировке структуры графовой базы знаний.

Выводы

В работе разработан метод оценки семантической полноты графов знаний в домене научной литературы, основанный на анализе семантических структур графа и их

устойчивости к лингвистическим вариациям текстов. Ключевой особенностью метода является использование семантических инвариантов, позволяющих калибровать оценку полноты и снижать влияние различий в формулировках исходных текстов. Проведённый анализ показал, что применение предложенного метода позволяет более эффективно выявлять области семантической неполноты графов знаний и улучшать структуру графовых баз знаний. Проверка результатов с использованием метода MINE-1 показала повышение показателей семантической полноты, что подтверждает эффективность разработанного подхода.

Литература

1. Chen H., Zhang R., Wang X. et al. Graph Edit Distance Computation and Its Applications in Automatic Summarization // Information. — 2026. — Т. 17, № 1. — С. 35.
2. Pan S. J., et al. Unifying Large Language Models and Knowledge Graphs: A Roadmap // IEEE Transactions on Knowledge and Data Engineering. — 2024.
3. Tavarez B et al. Ontology Learning and Knowledge Graph Construction: A Comparison of Approaches and Their Impact on RAG Performance, 2025, URL: <https://arxiv.org/html/2511.05991v1>
4. Mo B., Yu K., Kazdan J et al. KGen: Extracting Knowledge Graphs from PlainText with Language Models, 2025, URL: <https://arxiv.org/pdf/2502.09956>