УДК 004.912

Исследование квантовоподобной модели триадных зависимостей слов на основе теста Белла для задач семантического ранжирования

Саблин Д. П.[1]
Научный руководитель – доктор техн. наук, доцент Бессмертный И.А.[1],
канд. техн. наук Авдюшина А. Е.[1]
[1]Университет ИТМО
sablin.dmitrii2024@gmail.com

Введение

В современном мире системы поиска занимают важное место в жизни каждого человека. Хотя IR(information Retriever) существует с середины прошлого века, наблюдается тенденция к поиску лучшего способа осуществлять поиск с затратой меньшего количество времени, памяти и более качественной выдачей.

В рамках задачи ранжирования активно развиваются смежные направления, связанные с использованием структурированного и контекстно-обогащенного поиска. Все чаще стало использоваться технология больших языковых моделей для составления матриц документа и осуществляться поиск над ними. Но данные методы также обладают проблемами с пониманием текста. Так в новых исследованиях, как [1] показывается, что даже современные LLM-based подходы имеют проблемы с пониманием и соответственно качеством поиска.

На данный момент имеется различные системы поиска. В нашей стране особенно активно исследуются подходы bi-encoders. Так в работах [2], [3] наблюдается активное исследование технологии и различные вариации для лучшей работы в рамках русского языка. Стоит отметить что большинство описываемых методов работают над улучшением за счет различных новых подходов работы ANN, Contextual IR или заморозок мультиязычных моделей.

В мире наблюдается более широкий спектр возможных подходов. Исследуются методы от модификаций трансформерных слоев с добавлением «квантовых интерференций» до работ SPAR-RAG и Агентных поисковых систем [4, 5]. При этом присутствует большое количество различных датасетов для оценки качества ранжирования от классических до относительно новых таких как MTEB(Massive Text Embedding Benchmark) [6].

Ввиду описанных выше причин становится крайне важно отметить, наличие систем квантового поиска в небольшом количестве ввиду отсутствия построенной системы и качественного анализа с существующими подходами.

Основная часть

В рамках данной работы рассматривается способы лучшего представления триад слов для систем ранжирования. Как известно на данный момент большинство исследуемых методов квантового поиска базируется на парах или триадах слов.

Предлагается проверить следующие методы по работе с предложениями:

1. Представление предложений как матрицы квантовых состояний - составление большой гильбертовом матрицы слов объединенных простой конкатенацией
2. Тензорная сборка предложений - составление эмбединга слов по окружению через энкодеры или матрицу Холла и дальнейшая работа как с Tensor-Train (TT) преобразованием.

Таким образом предлагаемый способ ранжирования и работе с текстовыми данными будет обладать следующим видом:

Первоначально планируется получить эмбединги слов через различные виды эмбедеров, далее на основе квантовоподобного теста Белла проверить наличие квантовой связи между триадами слов в предложении.

В зависимости от наличия или отсутствия взаимосвязей, нарушению теста Белла между словами необходимо построить представление либо с наличием слова, либо без него. Данная операция необходима ввиду наличия связности и необходимости «осведомленности» слов между собой. И затем строится представление через SVD-сжатие.

Таким образом полученное представление должно обладать дополнительной информацией между связанными словами. Особенностью данной работы является использование нарушения неравенства Белла как критерия для динамического установления TT-рангов в разложении.

Предложенная парадигма должна помочь системам ранжирования учитывать взаимосвязь между словами лучше и помогать ускорить процесс поиска.

Выводы

При проведении экспериментов было исследовано время работы системы поиска основанной на разных моделях энкодинга и с помощью разной системы построения триад слов. Методология разбиения текста показала ускорение работы.

В рамках дальнейших работ, планируются провести оценку данного подхода на разных датасетах и собрать фреймворк для поисковой системы. После сбора большей информации также будет осуществлено сравнение метода для предложений.

Литература

1. Weller O. On the Theoretical Limitations of Embedding-Based Retrieval [Электронный ресурс] / O. Weller, M. Boratko, I. Naim, J. Lee // arXiv. 2025. Режим доступа: https://doi.org/10.48550/arXiv.2508.21038 (Дата обращения: 17.02.2026).
2. Khrylchenko K. Scaling Recommender Transformers to One Billion Parameters [Электронный ресурс] / K. Khrylchenko, A. Matveev, S. Makeev, V. Baikalov // arXiv. 2025. Режим доступа: https://doi.org/10.48550/arXiv.2507.15994 (Дата обращения: 17.02.2026).
3. Kovalev G. Wikipedia-based Datasets in Russian Information Retrieval Benchmark RusBEIR [Электронный ресурс] / G. Kovalev, N. Loukachevitch, M. Tikhomirov [et al.] // arXiv. 2025. Режим доступа: https://doi.org/10.48550/arXiv.2511.05079 (Дата обращения: 17.02.2026).
4. Yang Y. SPARC-RAG: Adaptive Sequential-Parallel Scaling with Context Management for Retrieval-Augmented Generation [Электронный ресурс] / Y. Yang, G. Deng, O. F. Akgül [et al.] // arXiv. 2026. Режим доступа: https://doi.org/10.48550/arXiv.2602.00083 (Дата обращения: 17.02.2026).
5. Ning J. Agentic Search in the Wild: Intents and Trajectory Dynamics from 14M+ Real Search Requests [Электронный ресурс] / J. Ning, J. Coelho, Y. Kong [et al.] // arXiv. 2026. Режим доступа: https://doi.org/10.48550/arXiv.2601.17617 (Дата обращения: 17.02.2026).
6. Muennighoff N. MTEB: Massive Text Embedding Benchmark [Электронный ресурс] / N. Muennighoff, N. Tazi, L. Magne, N. Reimers // arXiv. 2023. Режим доступа: https://doi.org/10.48550/arXiv.2210.07316 (Дата обращения: 17.02.2026).