

УДК 004.89, 002.513.5

Подход к поиску научных статей с применением LLM

Ларин А. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Волчек Д.Г.
(ИТМО)

Введение. Задача поиска статей является важным аспектом для ученых и исследователей в различных областях знаний. Они постоянно нуждаются в доступе к новейшей информации и результатам исследований. Занимаясь подбором литературы, перед исследователем возникает задача отсмотра огромного количества источников и работ, прежде чем найдётся та, из которой можно будет извлечь пользу. В целом, поиск статей является важным аспектом работы ученых, но может стать сложной задачей из-за множества баз данных и журналов, а также проблемы отбора подходящих статей [1]. В последнее время с развитием больших языковых моделей (LLM), возможности автоматизации поиска и отбора статей значительно расширились [2]. Таким образом часть задач связанных с поиском научной литературы может решена с их помощью.

Основная часть. Представим типичный алгоритм работы исследователя при поиске научной литературы по теме. На входе имеем тематику для поиска: некоторый образ статьи, которая была бы полена исследователю, выражаемая в поисковом запросе. Таким образом процесс примерно следующий:

1) Формирование запроса.

Исходя из образа конечного результата исследователь формулирует запрос, который будет послан в поисковую систему.

2) Использование выбранной поисковой системы.

Вносим в базу поисковый запрос, получаем список статей.

3) Ручной отсев.

Это тот этап, где нужно понять отвечает ли статья запросу. Как правило этот процесс происходит итеративно. На выходе получается список отложенных статей на прочтение. Требуется система с поведением похожим на распространённые сервисы для поиска статей. На вход приходит поисковый запрос, поисковая выдача содержит статьи с дополнительной информацией в виде суммаризации, поясняющей как статья отвечает на запрос.

С точки зрения пайплайна, изначально решается задача трансформации запроса в соответствие с полезными практиками для поиска. На выходе получается несколько запросов, расширяющие изначальный.

Затем получившиеся запросы отправляются в поиск при помощи утилиты Gesck проекта Nexus STC. Получается массив из полнотекстовых статей и метаданных к ним.

Затем массив статей проходит две стадии обработки - фильтрацию и суммаризацию с дополнительным запросом [3]. В итоге остаются статьи релевантные запросу, и к ним пояснение об их релевантности.

Данная информация отправляется на выход, в выдачу пользователю.

Выводы. В данной работе рассмотрено решение задачи поиска научных статей. Обосновано применение LLM для оптимизации данного процесса. Предложена процедура для последовательной обработки запроса для получения поисковой выдачи по пользовательскому запросу.

Список использованных источников:

1. Источники исследовательского поиска: где искать информацию для научных исследований // Научные Статьи.Ру — портал для студентов и аспирантов. — URL <https://nauchniestati.ru/spravka/istochniki-issledovatel'skogo-poiska/>

2. Hambarde K. A., Proenca H. Information Retrieval: Recent Advances and Beyond //arXiv preprint arXiv:2301.08801. – 2023.

3. Applied prompting: Overview // Learn Prompting. — URL: https://learnprompting.org/docs/applied_prompting/overview

Ларин А. (автор)

Волчек Д.Г. (научный руководитель)