

1 СЛАЙД

Добрый день, уважаемые председатели и члены комиссии. Вашему вниманию представляется выпускная квалификационная работа на тему "программный комплекс интеллектуального поиска в корпоративных документах".

За последние несколько десятилетий произошел значительный рост объемов информации, который требует тщательного анализа поступающих данных, что существенно замедляет рабочий процесс в корпорациях. В этих условиях традиционные методы работы с информацией становятся неэффективными и нуждаются в значительных ресурсах для поддержания актуальности и доступности информации.

2 СЛАЙД

В связи с этим были сформулированы цель и задачи

Но сперва проведем краткий экскурс в предметную область.

3 СЛАЙД

В качестве обработки естественного языка были использованы языковые модели.

Чтобы языковая модель могла искать необходимую информацию среди документов их необходимо добавить в виде контекста. Для этого существует метод RAG. Запрос, который был создан векторизуется, далее ищется необходимый файл и контекст в векторной БД путем косинусного сходства. Контекст и запрос отправляют в языковую модель, которая ищет релевантную информацию среди контекста и выдает ответ.

4 СЛАЙД

Переходя к видам языковых моделей, было выделено 2 вида, трансформер и mamba, которые обрабатывают большие последовательности быстро и эффективно, с их описанием можно ознакомиться в пояснительной записке

5 СЛАЙД

Далее был проведен обзор аналогов. Несмотря на наличие обширного списка аналогов их использование невозможно в связи с санкциями российским пользователям, несоответствием требований и необходимостью мощного ПО

6 СЛАЙД

Связи с этим были выбраны следующие модели. Llama и GigaChat. GigaChat мощная модель, но к ней можно подключиться и использовать API. Llama имеет несколько моделей, для которых не требуется мощное ПО

7 СЛАЙД

В качестве стека технологий был выбран набор: streamlit, ollama, langchain. Вся разработка производилась на языке python.

8 СЛАЙД

Проведя анализ, были сформулированы функциональные требования, на основе которых происходили этапы проектирования. На 1м этапе была спроектирована диаграмма прецедентов. Сотруднику необходимо сформировать свой запрос используя дополнительные настройки: выбрать языковую модель, определить область поиска. Так же сотрудник может загружать свои документы в хранилища и при необходимости удалять

9 СЛАЙД

На 2м этапе была спроектирована диаграммы последовательности.

На 1м рисунке показано, как сотрудник попадает в систему, происходит получение токена доступа для взаимодействия с API GigaChat и получение документов как из корпоративной базы, так и из личного хранилища сотрудника. После инициализации отображается чат с необходимыми настройками, и сотрудник может вводить свой запрос.

На 2м рисунке показано, как сотрудник работает с документами, ему необходимо их загрузить в хранилище. После он может посмотреть их наличие и удалить при необходимости актуализации документа или устаревания

10 СЛАЙД

Завершая этап проектирования, были составлены макеты интерфейса. При входе в комплекс сотрудник попадает на главную страницу, где будет размещена информация о данном комплексе. На станице Чат сотрудник будет задавать запрос. В разделе Документы сотрудник будет загружать, просматривать и удалять документы

11 СЛАЙД

Переходя к этапу реализации, была составлена диаграмма пакетов, которая отображает зависимость пакетов иллюстрируя функциональность программного комплекса

12 СЛАЙД

На этапе тестирования были проверены ответы языковых моделей в зоне поиска определенного документа и среди базы документов. В качестве примера был использован документ ОС ТУСУР. На 1м рисунке показано, как данный файл выбирается и в нем ищется ответ. На 2м рисунке, как система по аналогичному запросу находит данный файл, а потом получает из него ответ.

13 СЛАЙД

Llama3.1:8b

14 СЛАЙД

GigaChat

15 СЛАЙД

Всего было сделано 30 запросов к моделям. Корректность ответа определялась вручную

16 СЛАЙД

В заключении, был разработан программный комплекс интеллектуального поиска в корпоративных документах. За период обучения была опубликована 1 статья, которая написана по теме ВКР.

Продолжение работы планируется в рамках магистерской диссертации.

Спасибо за внимание.