**СИСТЕМА ИНТЕЛЛЕКТУАЛЬНОГО ПОИСКА В КОРПОРАТИВНЫХ БАЗАХ ЗНАНИЙ**

***Е.П. Бекиш, студент каф. АСУ***

*г. Томск, ТУСУР,* egorbeckish@mail.ru

*Научный руководитель:* *А.А. Захарова, д-р технических наук, профессор каф. АСУ*

В современном мире объемы информации с каждым днем растут все больше и больше, поэтому анализировать информацию становится труднее, как и выбрать более корректную. Особенно организациям, у которых есть своя база знаний, из которой они хотят так же получать точную информацию, избегая анализа лишнего материала. Следовательно, обоснована актуальность разработки системы, которая обеспечивает по заданному запросу пользователя в корпоративной базе знаний быстрый и корректный ответ для дальнейших действий.

**Ключевые слова:** языковые модели, эмбеддинг, машинное обучение.

Современные корпорации сталкиваются с экспоненциальным ростом объемов данных и знаний, что делает управление информацией все более сложным процессом. В этих условиях традиционные методы работы с базами знаний становятся неэффективными и требуют значительных ресурсов для поддержания актуальности и доступности информации. Языковые модели (LLM), такие как ChatGPT, предлагают инновационные решения для автоматизации обработки и анализа данных, улучшения поиска и доступа к знаниям для сотрудников компаний, а также поддержки принятия решений [1]. Однако, несмотря на очевидные преимущества, многие компании еще не полностью осознали потенциал LLM и сталкиваются с вызовами при их внедрении. Важно понять, почему интеграция LLM в корпоративную базу является ключевым фактором успеха в условиях быстро меняющегося цифрового ландшафта и как преодолеть барьеры на пути к успешной реализации таких проектов.

Целью проекта является создание системы, которая позволяет сотруднику упростить поиск и анализ информации на основе запроса.

Функции приложения:

* авторизация;
* получение токена пользователя;
* обработка запроса языковой моделью;
* формирование ответа пользователю;
* хранение документов.

Входная информация:

* запрос, интересующий пользователя;
* файлы с документами.

Выходная информация:

* ответ на запрос.

Из существующих аналогов разрабатываемого приложения, являются [2]:

* IBM Watson — это мощная платформа искусственного интеллекта, которая включает в себя множество инструментов для обработки естественного языка, машинного обучения и аналитики данных. Она широко используется в корпоративных средах для автоматизации бизнес-процессов, улучшения клиентского обслуживания и оптимизации принятия решений. Watson может интегрироваться с корпоративной базой данных для создания чат-ботов, анализа текстов, прогнозирования и других задач, связанных с обработкой больших объемов данных.
* Azure Cognitive Services от Microsoft предлагает набор API-интерфейсов для интеграции функций искусственного интеллекта в приложения и системы. Эти сервисы включают обработку текста, речи, изображений и видео. С помощью Azure Cognitive Services можно создавать интеллектуальные приложения, которые взаимодействуют с корпоративными данными, анализируют тексты, распознают речь и многое другое.
* Google Cloud AI предоставляет инструменты и сервисы для разработки и развертывания моделей машинного обучения, включая Natural Language Processing (NLP). Платформа поддерживает создание и обучение собственных моделей, а также использование готовых решений. Google Cloud AI может быть использована для анализа текстов из корпоративных баз данных, автоматического ответа на запросы клиентов, классификации документов и других NLP-задач.
* Amazon Lex — это сервис для создания голосовых и текстовых интерфейсов на основе технологии Alexa. Он позволяет разрабатывать чат-боты и другие интерактивные интерфейсы, которые могут взаимодействовать с пользователями через текстовые сообщения или голосовые команды. Lex может быть интегрирован с корпоративной системой для автоматизации взаимодействия с клиентами, сотрудников или партнеров, обеспечивая быстрый доступ к необходимой информации.
* OpenAI GPT — это семейство языковых моделей, разработанных компанией OpenAI. Они используются для генерации текста, перевода, обобщения и других сложных задач обработки естественного языка. Хотя GPT изначально был создан для открытых систем, он может быть адаптирован и использован в корпоративных приложениях для различных целей, таких как автоматическое составление отчетов, ответы на вопросы пользователей и поддержка принятия решений на основе данных.

Следует отметить, что несмотря на наличие аналогов их использование невозможно в связи с санкциями и ограничением доступа российским пользователям.

Таким образом, актуальна разработка системы интеллектуального поиска, которую можно внедрять в российские компании.

Для данного проекта был выбран следующий стек технологий:

1. Python – это высокоуровневый язык программирования общего назначения, который широко используется для разработки программного обеспечения, веб-приложений, научных вычислений и анализа данных.
2. Adobe Acrobat — это программное обеспечение, предназначенное для работы с документами формата PDF. Оно было создано компанией Adobe Systems и впервые выпущено в 1993 году. Сегодня Adobe Acrobat является одним из самых популярных и мощных инструментов для создания, редактирования, защиты и совместного использования PDF-документов.
3. Microsoft Office Word (обычно просто Word) — это текстовый процессор, входящий в состав офисного пакета Microsoft Office. Он позволяет создавать, редактировать и форматировать текстовые документы, включая отчеты, письма, статьи и многое другое. Word предоставляет инструменты для работы с текстом, таблицами, изображениями и другими элементами, а также возможности для совместного редактирования и рецензирования документов.
4. GigaChat — это нейросетевая модель искусственного интеллекта, созданная компанией Сбер. Она предназначена для помощи пользователям в выполнении различных задач, таких как анализ текста, ответы на вопросы, программирование и многое другое.

На данном этапе разработке уже реализованы: чат, в котором пользователь может ввести запрос и получить на него ответ.

Частично реализован процесс поиска необходимого файла, из которого с помощью языковой модели извлекается ответ на запрос пользователя.

Реализованный прототип представлен на рисунке 1. Тестирование проведено на примере стандарта ОС ТУСУР 01-2021.

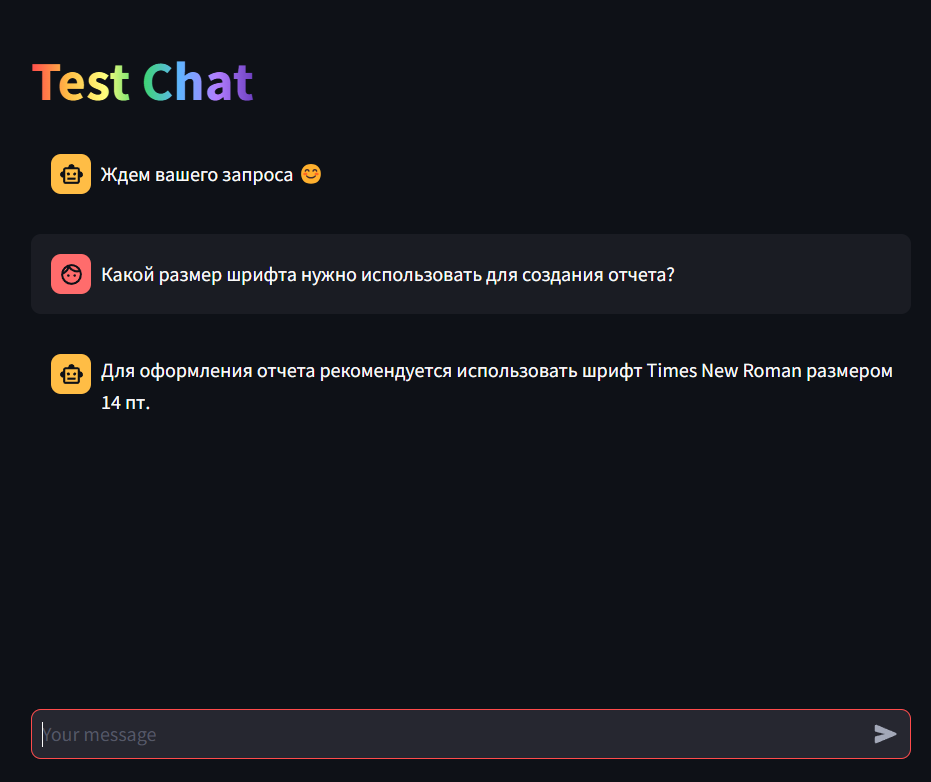


Рис. 1 – Экран приложения

Таким образом, был разработан начальный прототип системы интеллектуального поиска. Дальнейшее развитие проекта будет заключаться в доработке алгоритма авторизации, регистрации, а также направлено на работу с файлами. Эта система сможет помочь большинству организациям ускорить рабочий процесс.

**ЛИТЕРАТУРА:**

1. Раскрывая секреты LLM: руководство по основным понятиям больших языковых моделей [Электронный ресурс]. – Режим доступ: <https://habr.com/ru/articles/768844/> (дата обращения: 09.09.2024).
2. GigaChat Alternatives [Электронный ресурс]. – Режим доступ: <https://sourceforge.net/software/product/GigaChat/alternatives?ysclid=m2wz78922j338546976> (дата обращения: 21.10.2024).