

Федеральное государственное автономное образовательное учреждение высшего образования ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

Кафедра автоматизированных систем управления (АСУ)

Программный комплекс интеллектуального поиска в корпоративных документах

Выполнил: студент гр. 431-3 Бекиш Егор Павлович

Руководитель: доцент кафедры АСУ, к.т.н. Суханов Александр Яковлевич

Цель и задачи

Цель работы

разработать программный комплекс, который сокращает время доступа сотрудников к корпоративной информации за счет обработки запросов к внутренним документам предприятия, формата txt, pdf, docx, doc, pptx, ppt, на естественном языке и средствами ИИ.

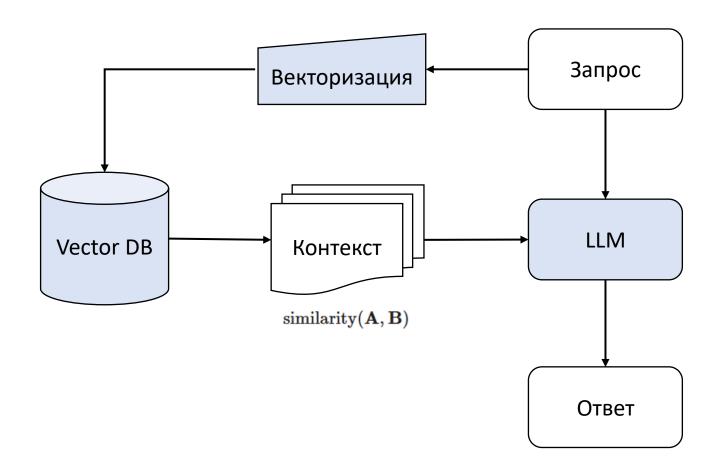
Задачи:

- 1. Провести анализ необходимых требований;
- 2. Определить спецификации и реализовать функциональные возможности;
- 3. Выполнить проектирование интерфейса и архитектуры;
- 4. Реализация программного комплекса;
- 5. Тестирование программного комплекса.

Обзор предметной области



- Выборка необходимых документов и данных
- Виды языковых моделей



Обзор предметной области



- Выборка необходимых документов и данных
- Виды языковых моделей

Модель	Вычислительная сложность	Длина контекста L	Пропускная способность
Трансформер	L^2	10^{3}	x
Mamba	L	10^{6}	5 <i>x</i>

Обзор аналогов



- IBM Watson
- Azure Cognitive Services or Microsoft
- Google Cloud AI
- Amazon Lex
- OpenAI GPT
- Hugging Face Chat
- DeepSeek

Функциональные требования

TYCYP
TUSUR UNIVERSITY

Языковые модели



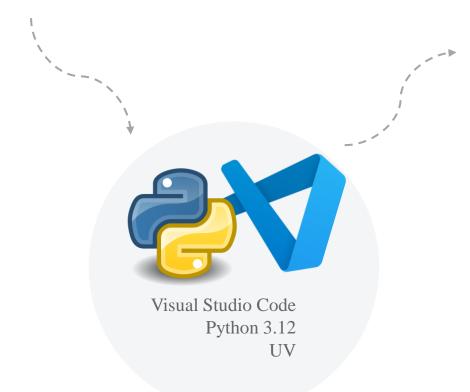


Функциональные требования

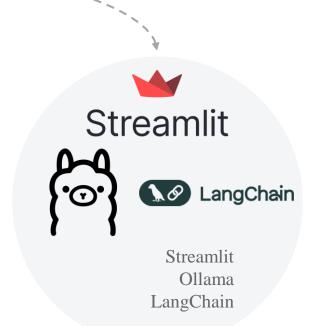
TYCYP TUSUR UNIVERSITY

Инструменты реализации

1. IDE, ЯП, Менеджер зависимостей



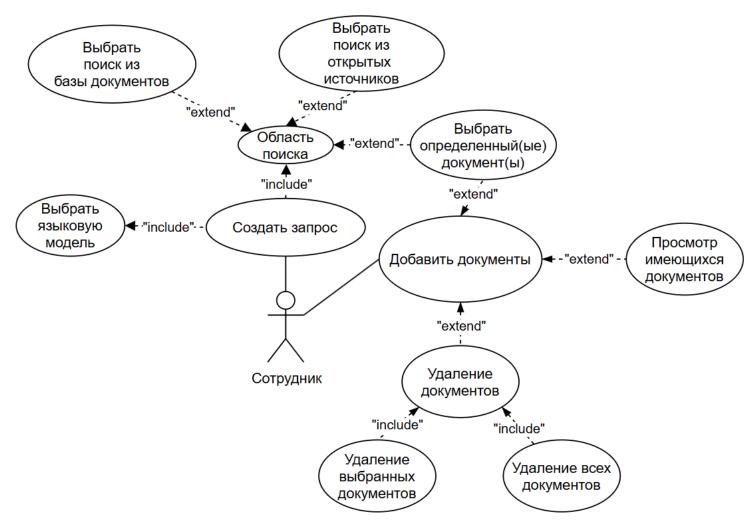
2. Библиотеки



Проектирование

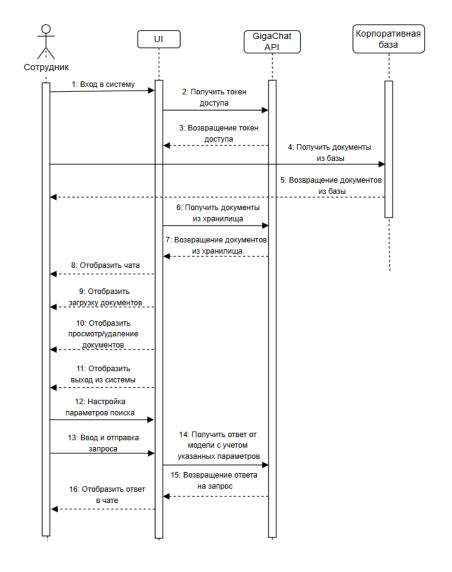
Диаграмма прецендентов



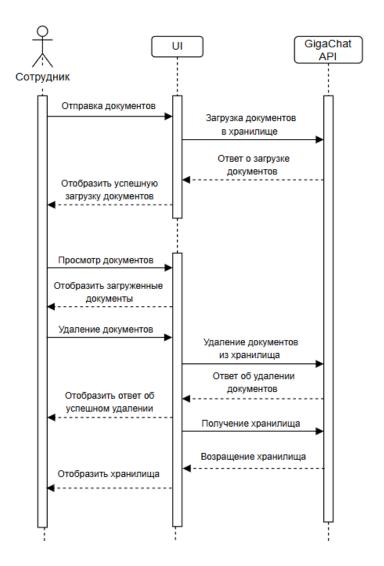


Проектирование

Диаграмма последовательности



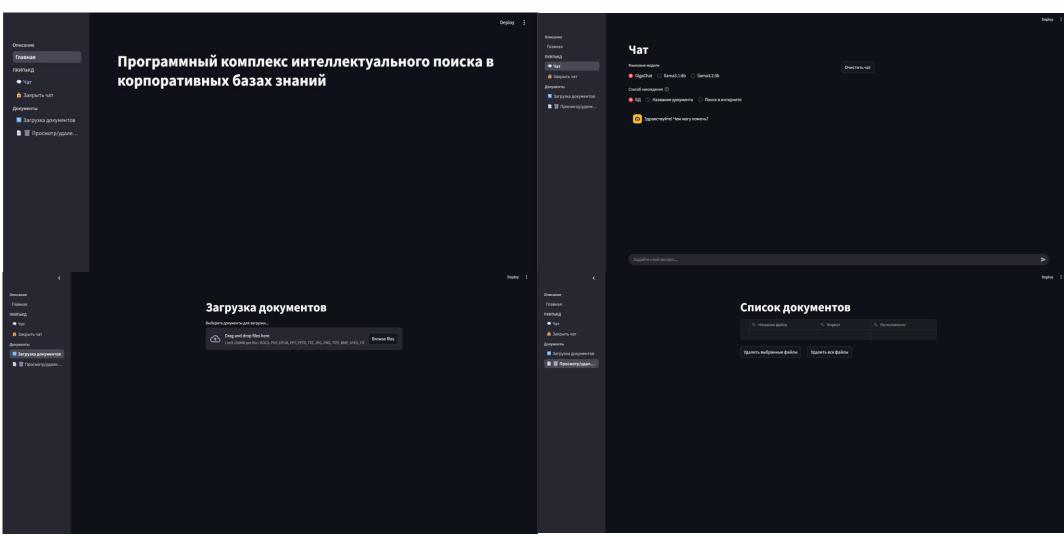




Проектирование

Макеты интерфейса

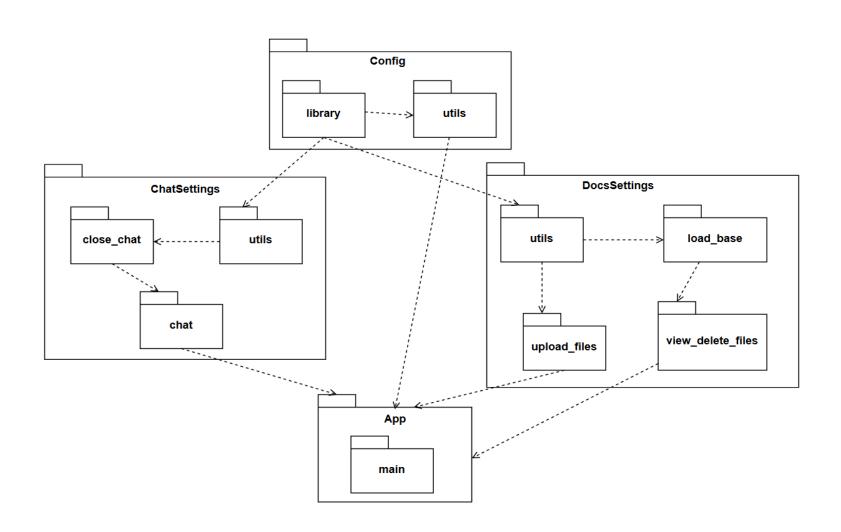




Реализация

Диаграмма пакетов



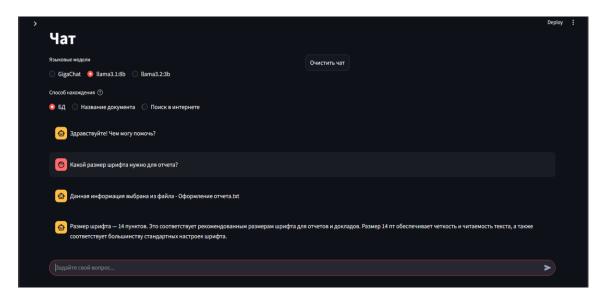




		Deploy
Чат		
Языковые модели	Очистить чат	
◯ GigaChat ◯ llama3.1:8b 🧕 llama3.2:3b		
Способ нахождения ③		
🔾 БД 🔞 Название документа 🤍 Поиск в интернете		
Список документов		
Оформление от х		
6 Какой размер шрифта нужно для отчета?		
 Да, по описанию отчета recommends использовать шрифт Times New Ro- четко читаемым и соответствующим стандартам Туроgraphy. 	man в черном цвете и размером 14 пт (пикселей) для всего текста отчета. Эт	о помогает сделать отчет четким,
Задайте свой вопрос		Deple
Вадайте свой вопрос Чат Языковые модели GigaChat ○ Ilama3.1:8b ● Ilama3.2:3b	Очистить чат	
Чат Языкопые модели ○ GigaChat ○ Ilama3.1:8b Ilama3.2:3b	Очистить чат	
Чат Языковые модели	Очистить чат	
Чат Языковые модели ○ GigaChat ○ Ilama3.1:8b • Ilama3.2:3b Способ нахождения ⑦	Очистить чат	
Чат Языковые модели GigaChat	Очистить чат	
Чат Языковые модели	Очистить чат	
Чат Языковые модели	Очистить чат	Deploy

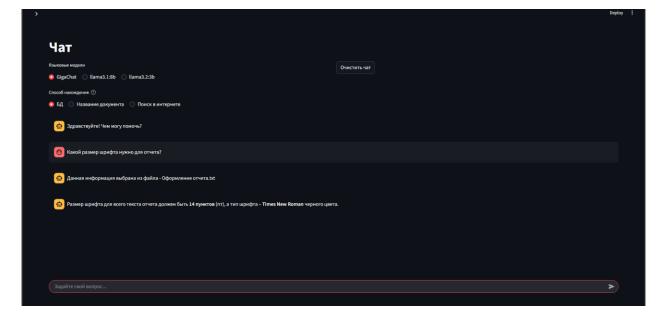


>		Deploy :
Чат		
Языковые модели GigaChat 🐧 llama3.1:8b 🦳 llama3.2:3b	Очистить чат	
Способ нахождения ③ БД Название документа Поиск в интернете		
Список документов		
Оформление от ×		0 v
 Какой размер шрифта нужно для отчета?		
<mark>⊕</mark> Размер: 14 pt		
Задайте свой вопрос		>



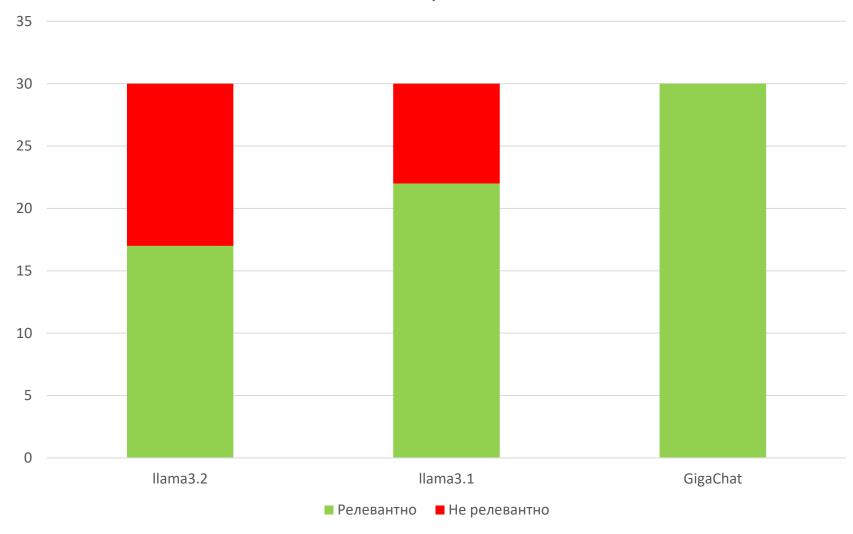


		Deploy :
Чат		
Языковые модели	Очистить чат	
⑤ GigaChat ○ Ilama3.1:8b ○ Ilama3.2:3b		
Способ нахождения ①		
○ БД 🔞 Название документа ○ Поиск в интернете		
Список документов		
Оформление от х		
Какой размер шрифта нужно для отчета?		
(a) Для отчета рекомендуется использовать шрифт Times New Roman размера 14 пунктов и черного цвета.		
		>
	<u> </u>	









Заключение



В ходе выполнения работы был разработан программный комплекс интеллектуального поиска в корпоративных документах, который позволяет повысить эффективность работы с информационными данными, что поможет ускорить бизнес-процессы.

Дальнейшее развитие программного комплекса планируется в рамках магистерской диссертации.

По теме бакалаврской работы была подготовлена статья:

• Бекиш Е.П. Система интеллектуального поиска в корпоративных базах знаний // XIII Региональная научно-практическая конференция «НАУКА И ПРАКТИКА: ПРОЕКТНАЯ ДЕЯТЕЛЬНОСТЬ – ОТ ИДЕИ ДО ВНЕДРЕНИЯ – 2024». Принято к публикации.