



Министерство науки и высшего образования Российской Федерации

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 базах знаний

Презентация по результатам
бакалаврской работы

Выполнил: студента гр. 431-3
Бекиш Егор
Павлович

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

Томск 2025

Цель и задачи

Цель работы

разработать программный комплекс, который сокращает время доступа сотрудников к корпоративной информации за счет обработки запросов к внутренним документам предприятия на естественном языке и их обработка средствами ИИ.

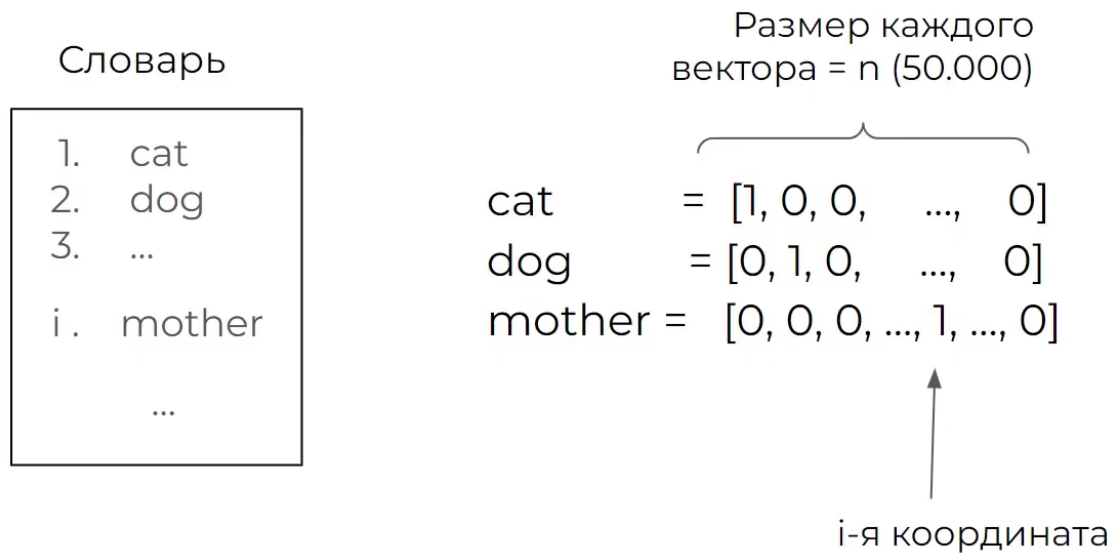
Задачи:

1. Анализ необходимых требований;
2. Реализация функциональных возможностей и выработка спецификаций;
3. Проектирование;
4. Реализация системы;
5. Тестирование.

Обзор предметной области

- Развитие векторного представления слов
- Виды языковых моделей
- Выборка необходимых документов и данных

Например, $n=50.000$ слов.



$$x_{ij} = \begin{cases} 1, & X_i = c_j \\ 0, & X_i \neq c_j \end{cases}$$

где x_{ij} – вектор слов;
 X_i – текущее слово;
 c_j – слово из словаря.

Обзор предметной области

- Развитие векторного представления слов
- Виды языковых моделей
- Выборка необходимых документов и данных

1. a dog eats meat
2. a dog hunts cat

TF-IDF векторы для этих предложений:

1.	[0, ..., 0, ... 0.075, ...	0, ...	0, ...	0, ...]
2.	[0, ..., 0, ... 0, ...	0.075, ...	0.075, ...	0.075, ...]
	↑ ↑ ↑	↑	↑	↑
	a dog eats	hunts	cat	meat

$$tf-idf(t, d, D) = tf(t, d) \times idf(t, D)$$

где t – входное слово;

d – текущий документ;

D – коллекция документов;

$tf(t, d)$ – относительная частота встречаемого слова в документе;

$idf(t, D)$ – обратная частота встречаемого слова в наборе документов.

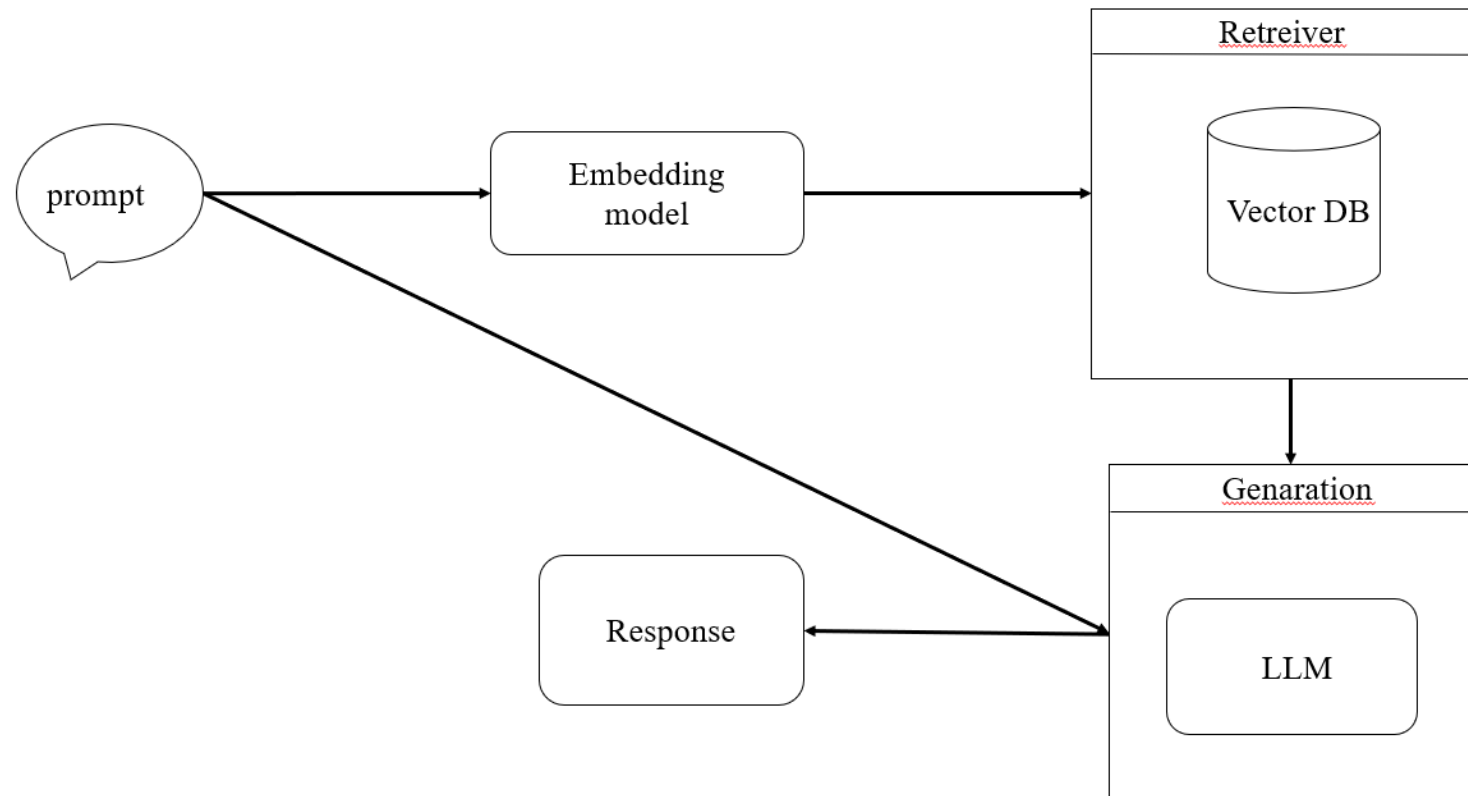
Обзор предметной области

- Развитие векторного представления слов
- **Виды языковых моделей**
- Выборка необходимых документов и данных

Модель	Вычислительная сложность	Длина контекста L	Пропускная способность
Трансформер	L^2	10^3	x
Mamba	L	10^6	$5x$

Обзор предметной области

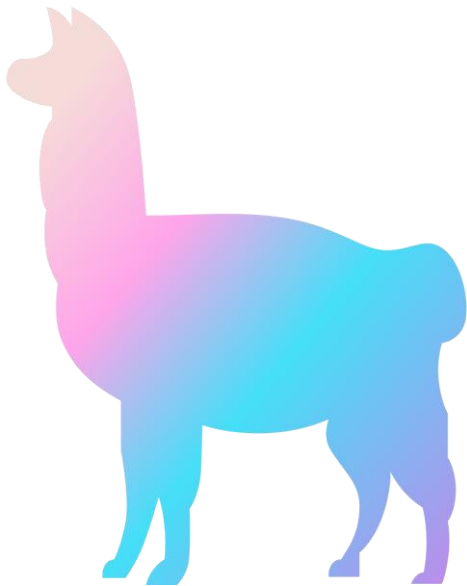
- Развитие векторного представления слов
- Виды языковых моделей
- **Выборка необходимых документов и данных**



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

Функциональные требования

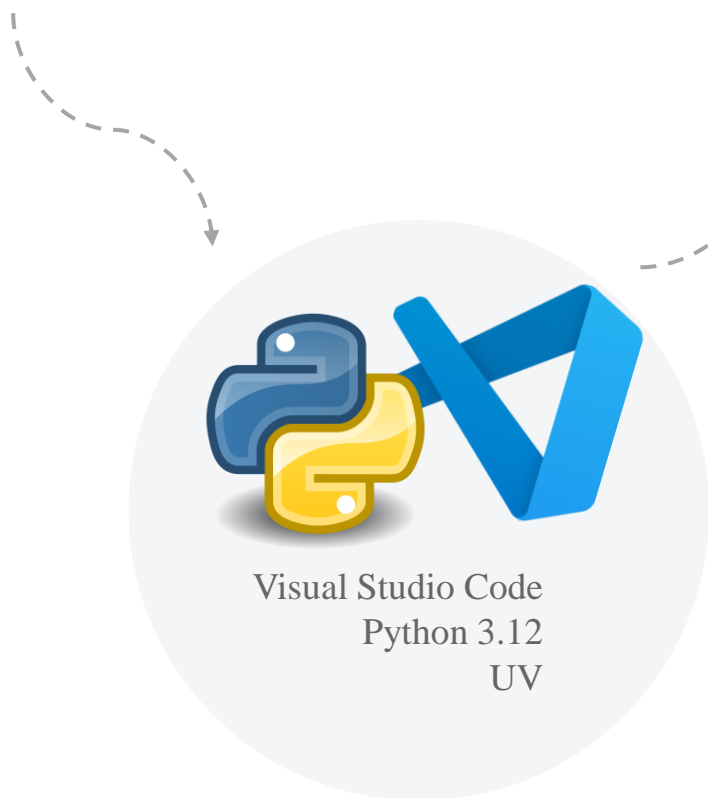
Языковые модели



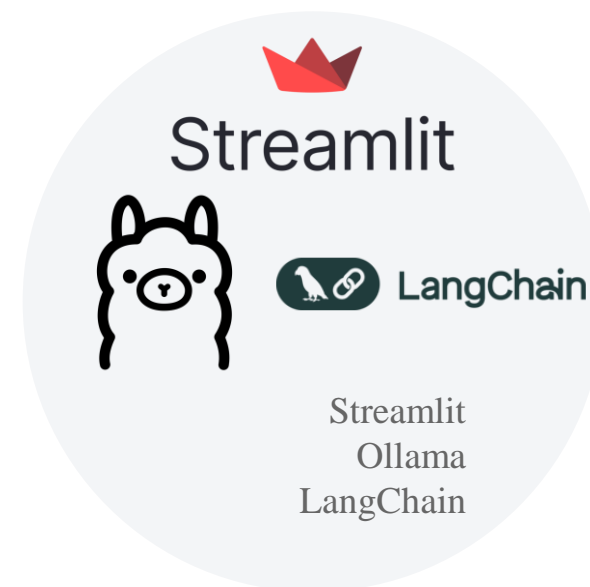
Функциональные требования

Инструменты реализации

1. IDE, ЯП, Менеджер зависимостей

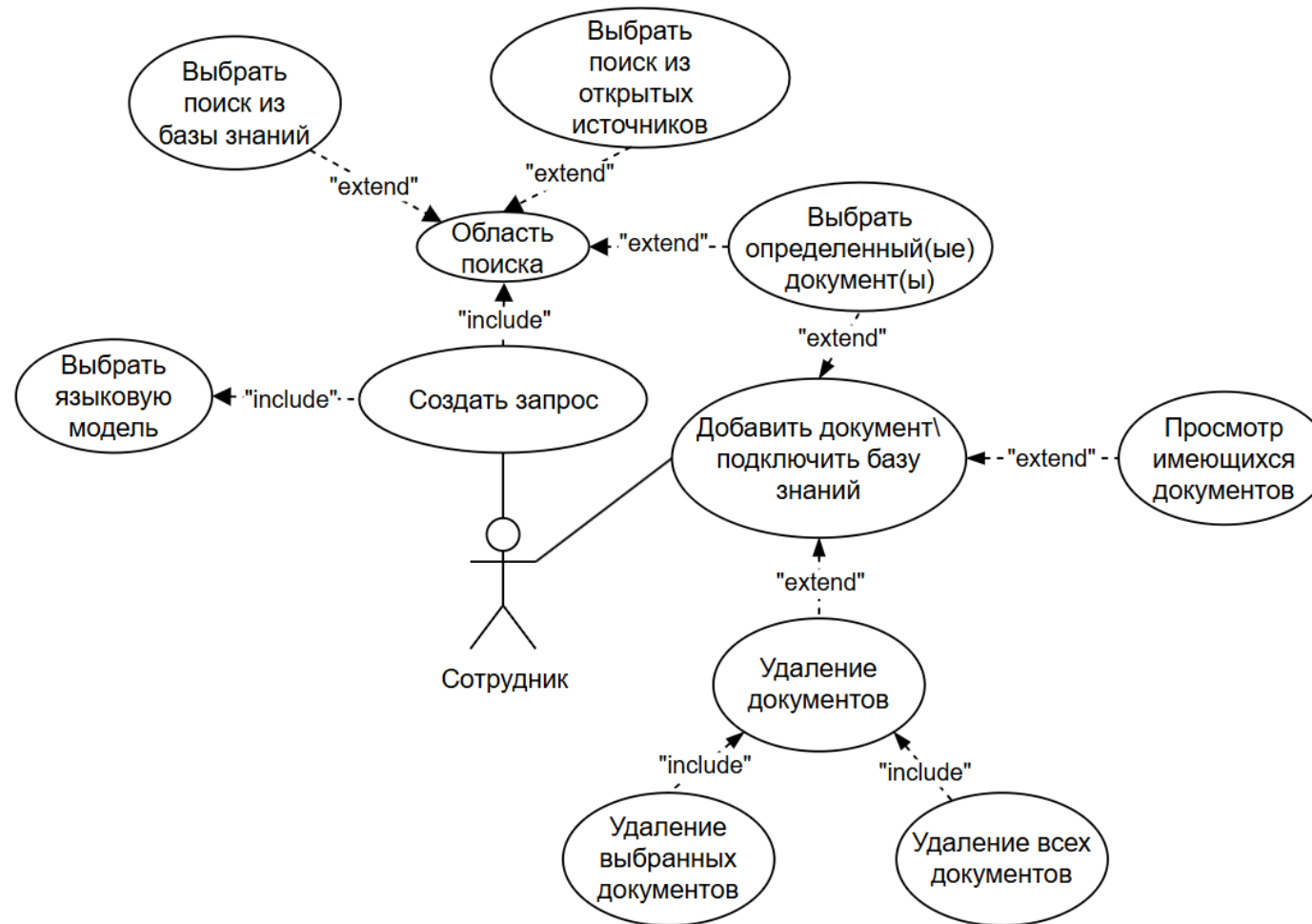


2. Библиотеки



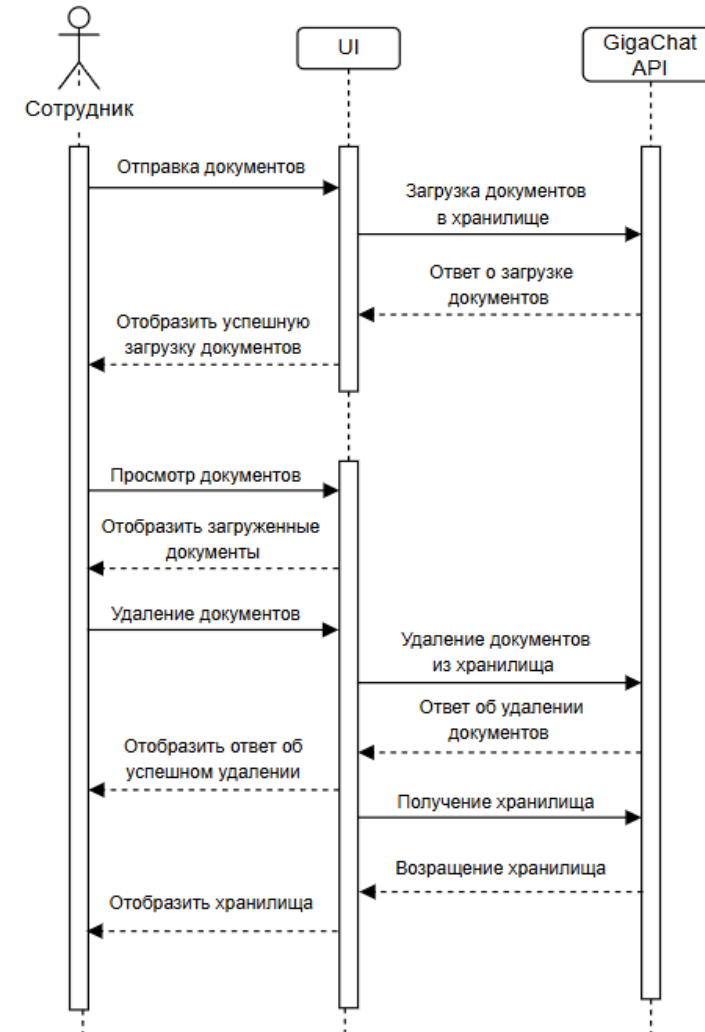
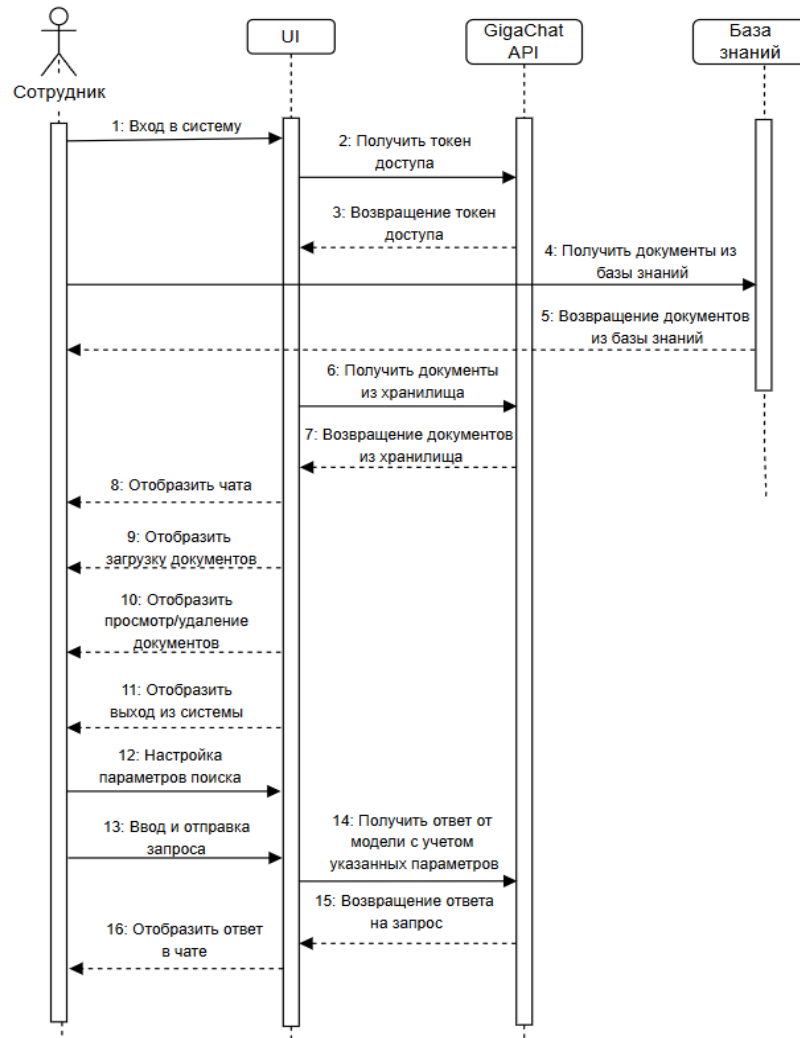
Проектирование

Диаграмма прецедентов



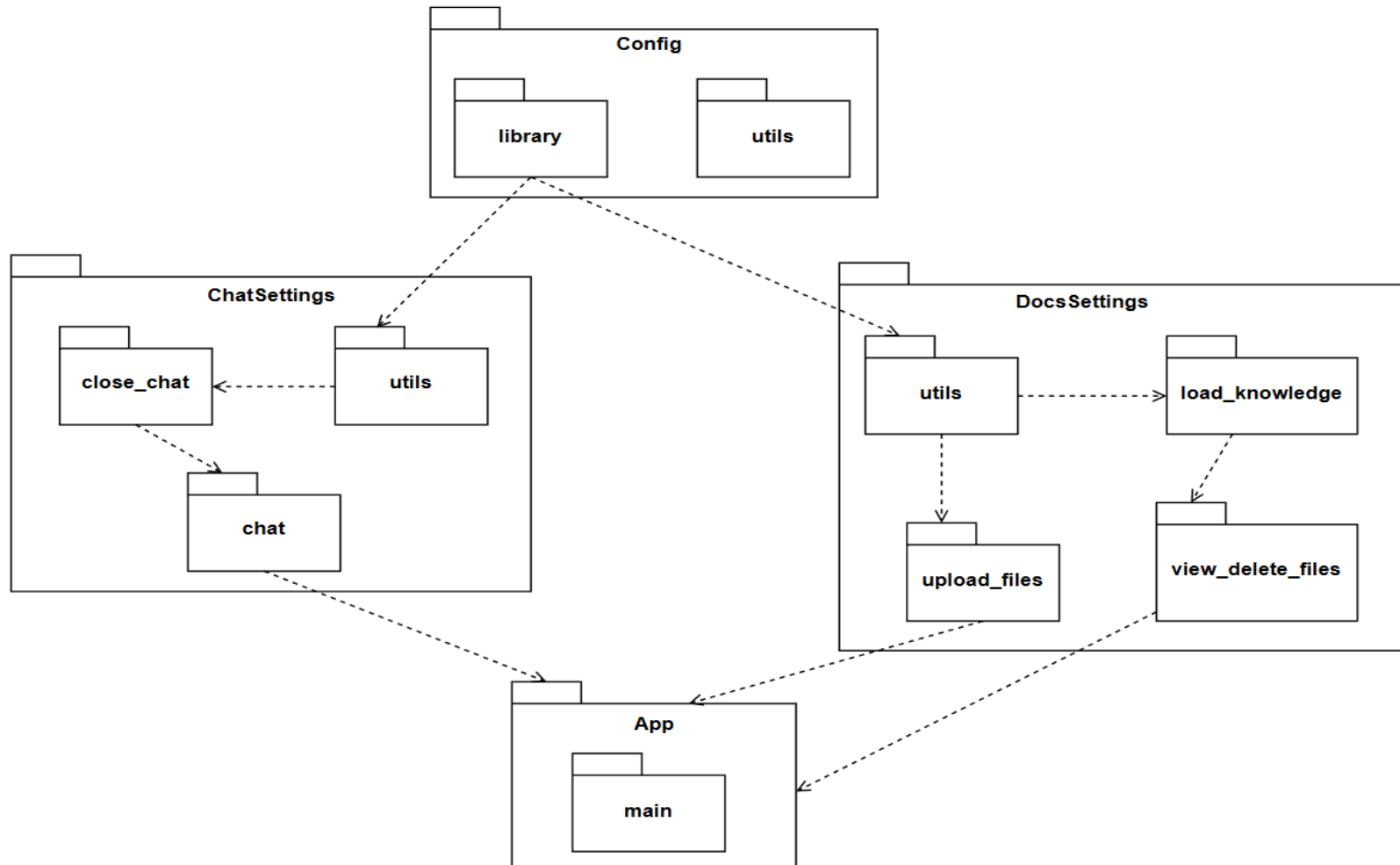
Проектирование

Диаграмма последовательности



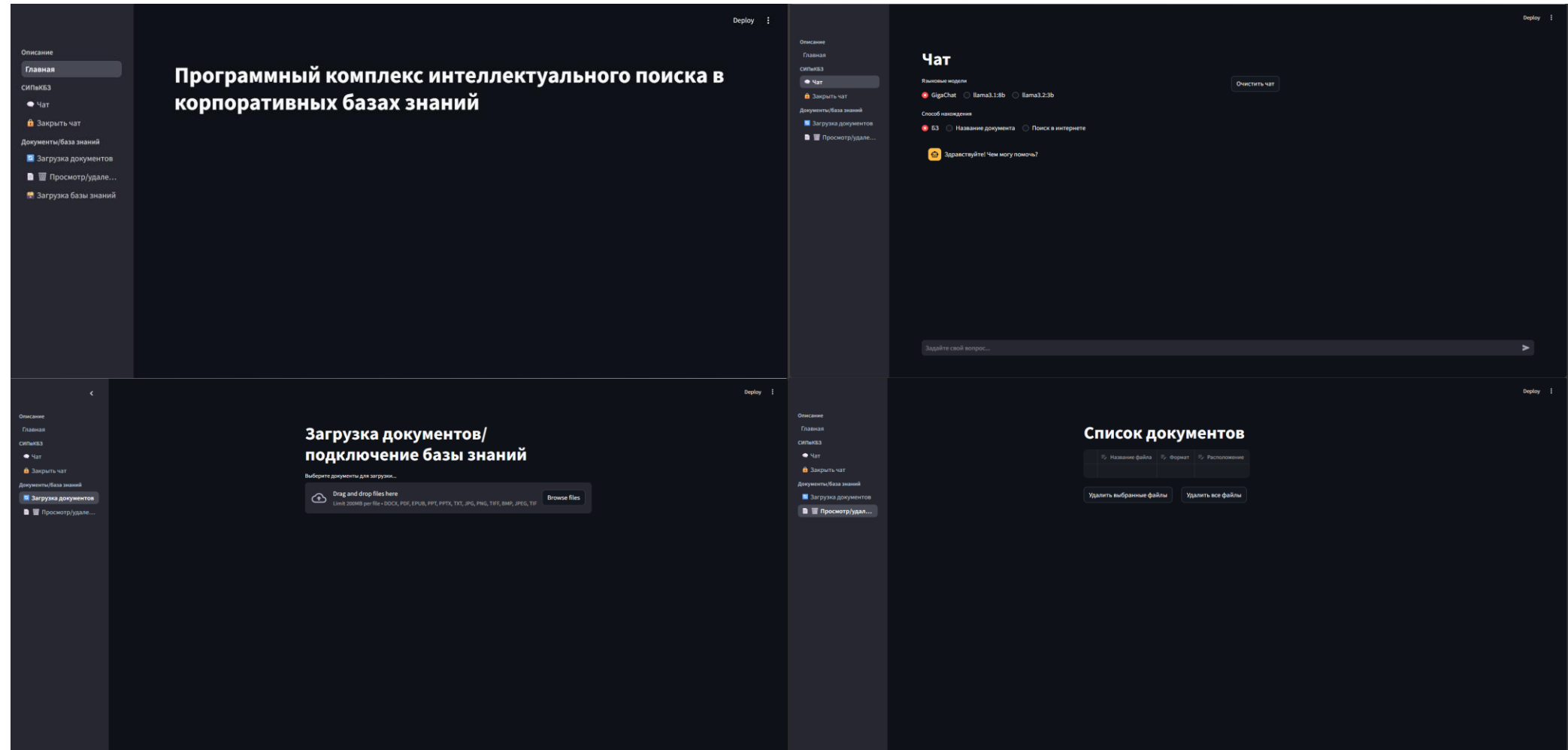
Проектирование

Диаграмма пакетов

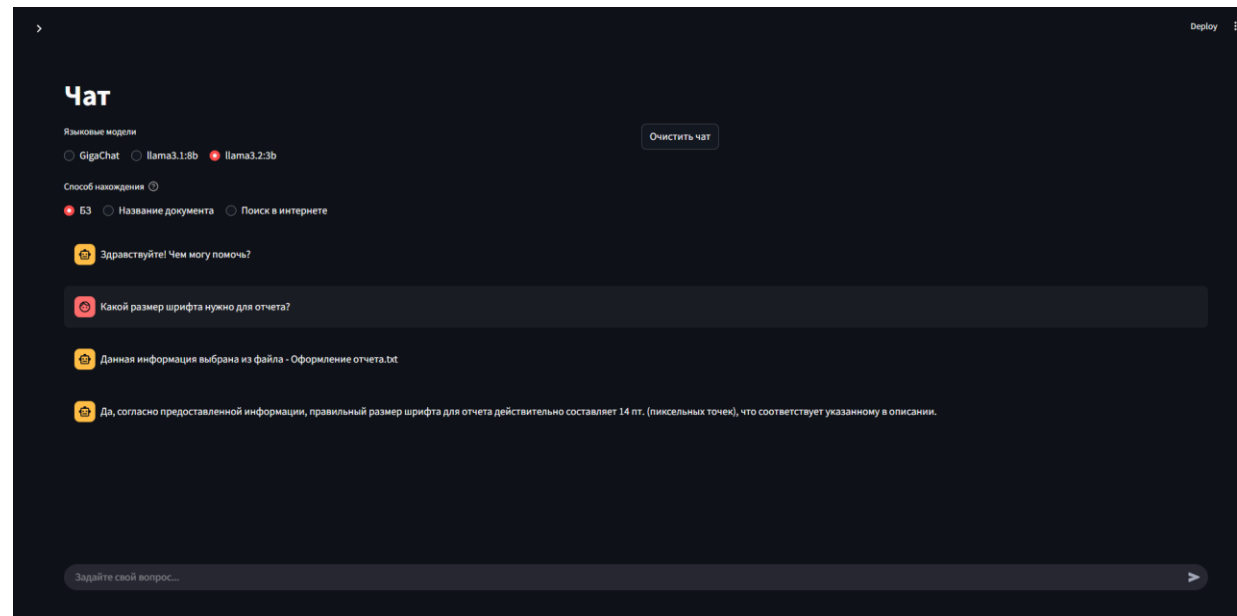
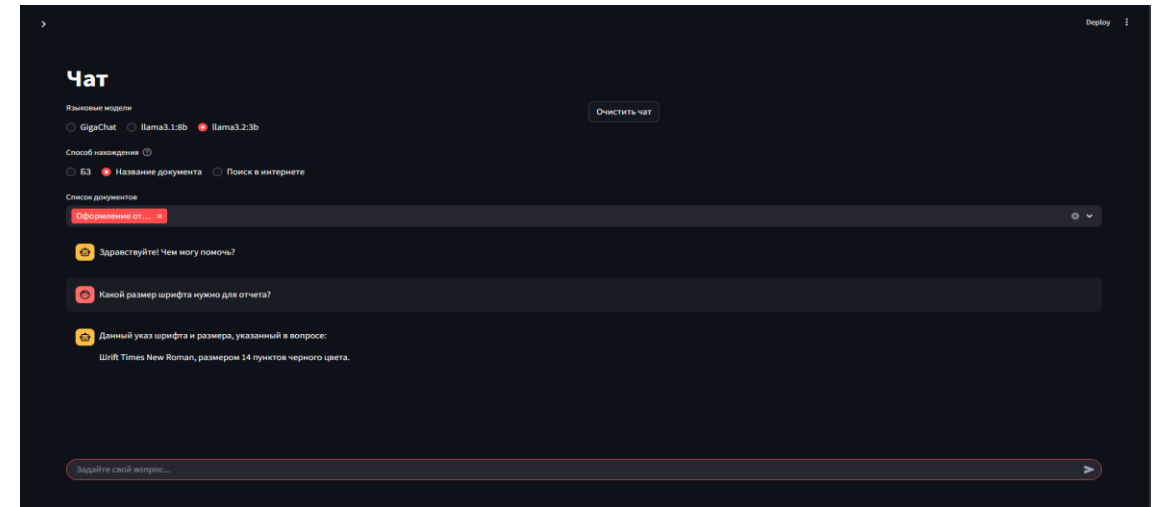
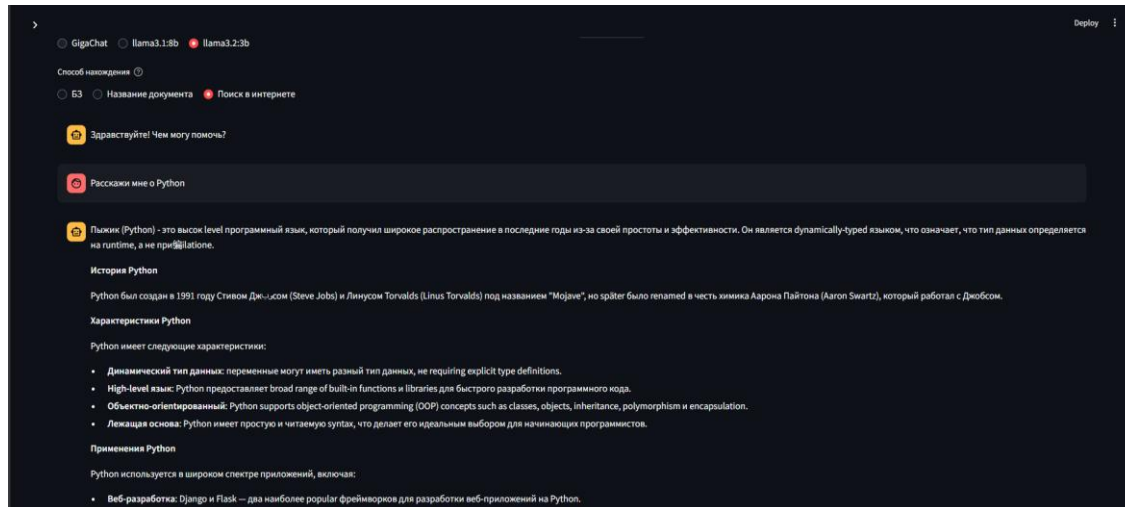


Проектирование

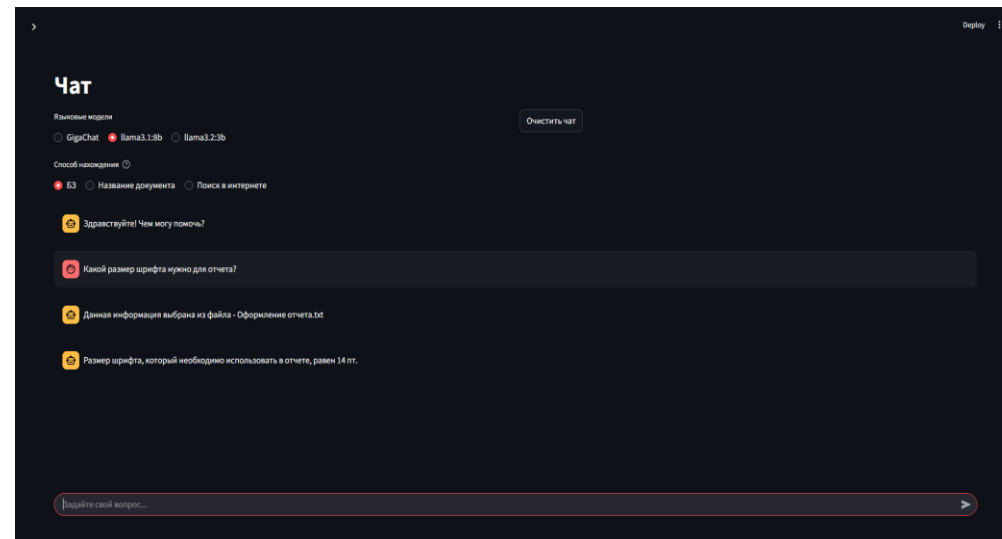
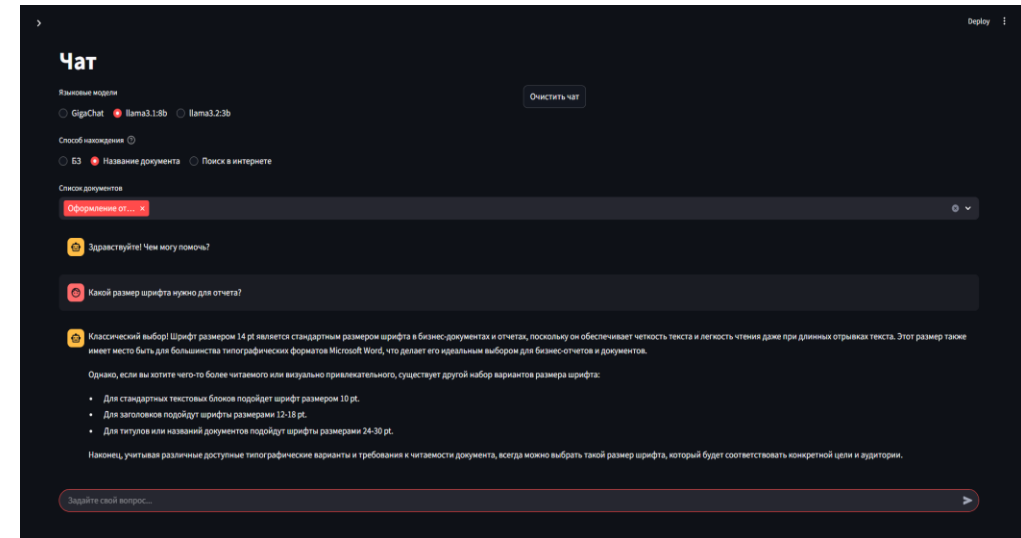
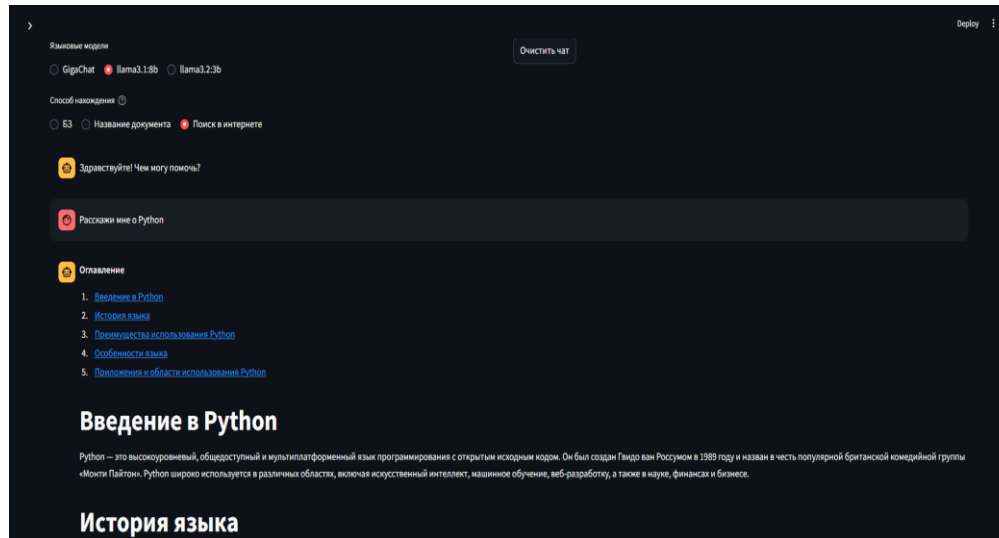
Макеты интерфейса



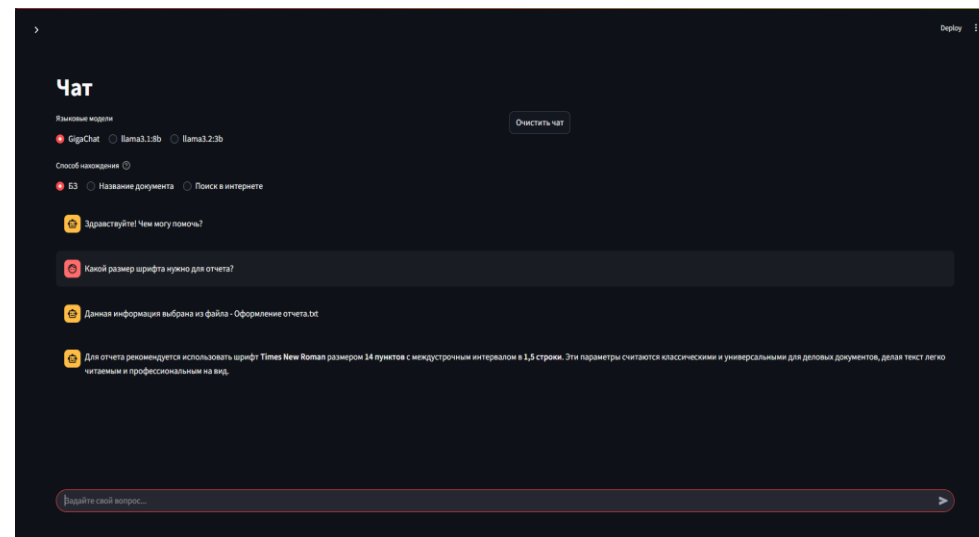
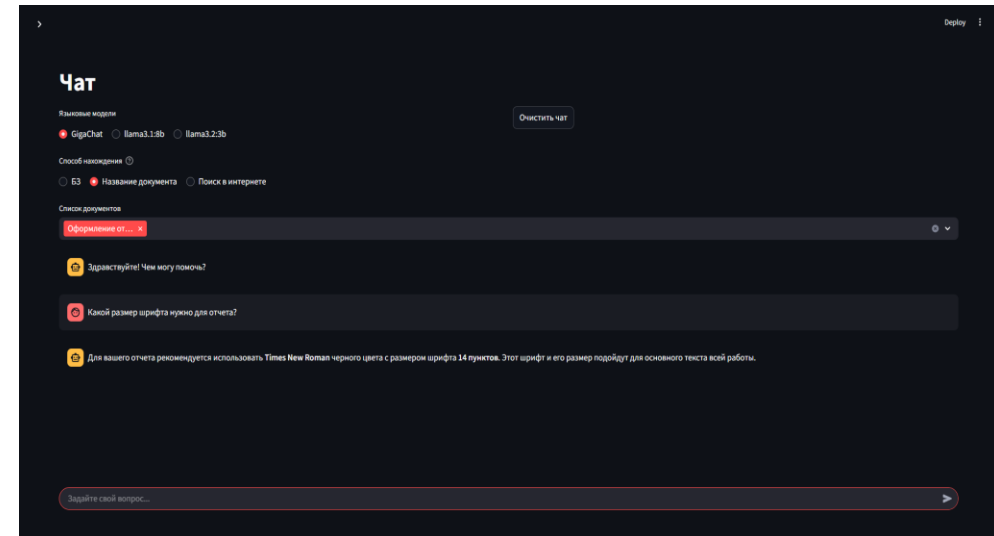
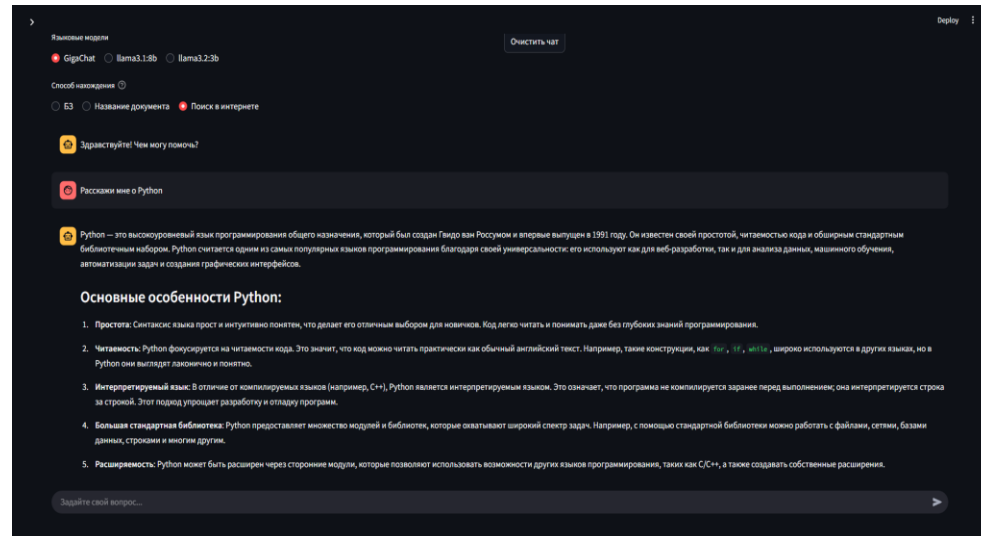
Тестирование



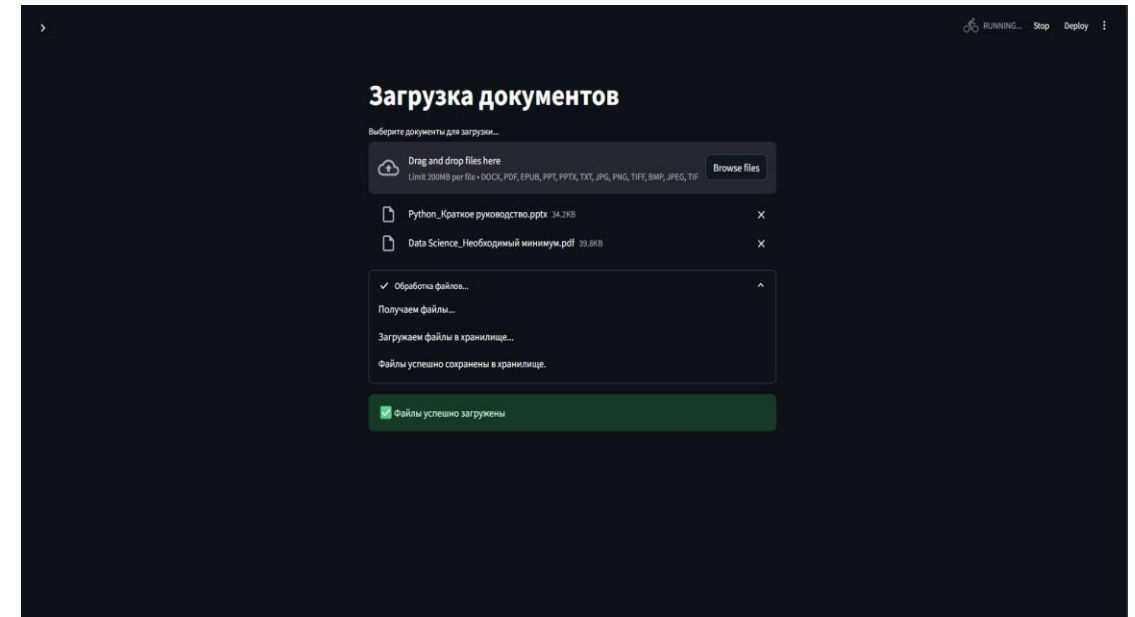
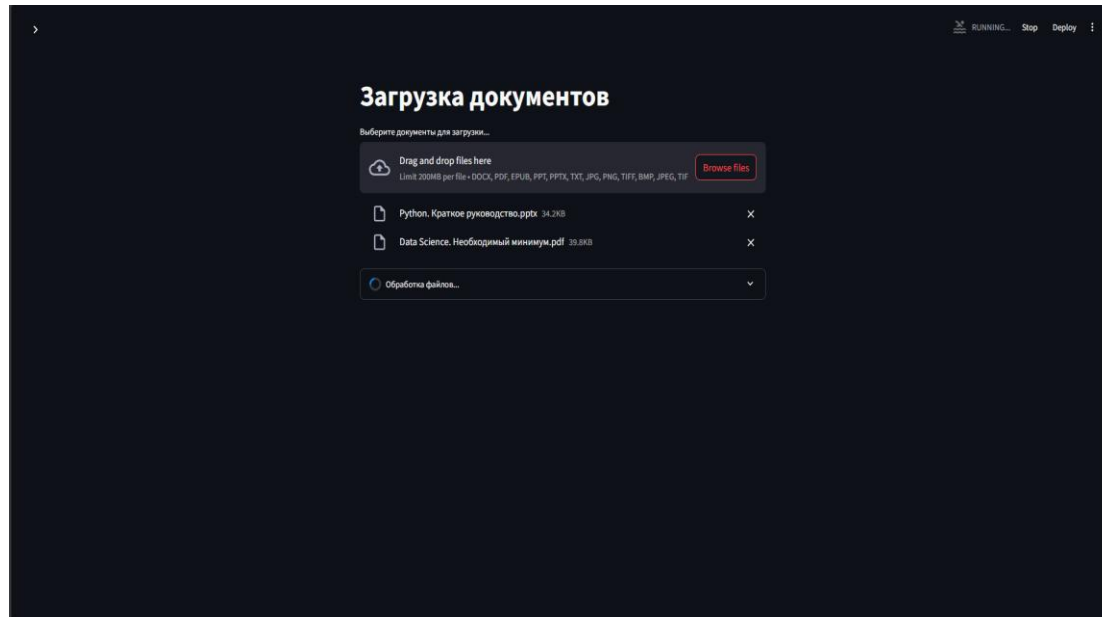
Тестирование



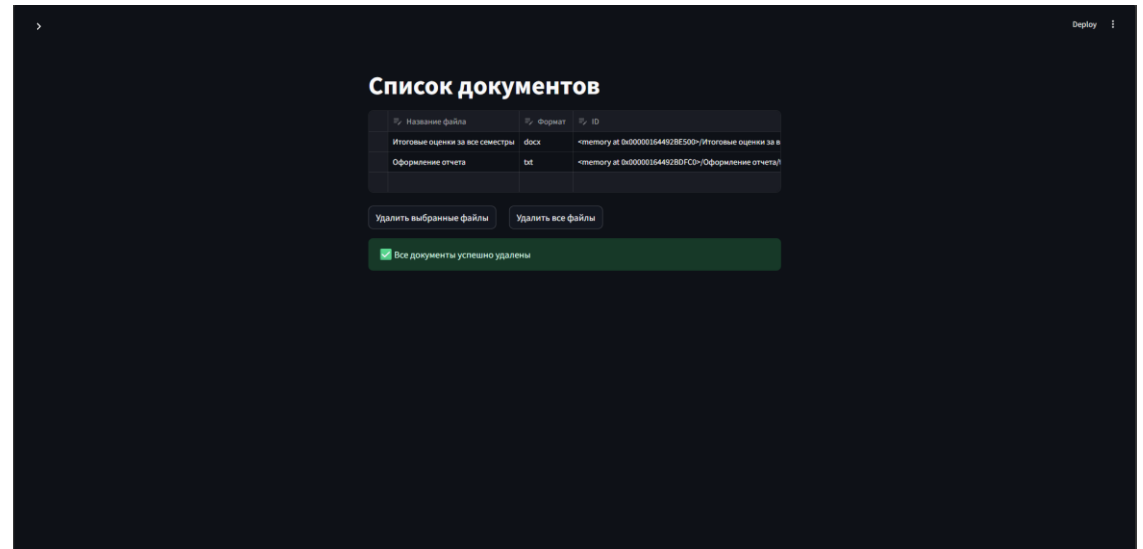
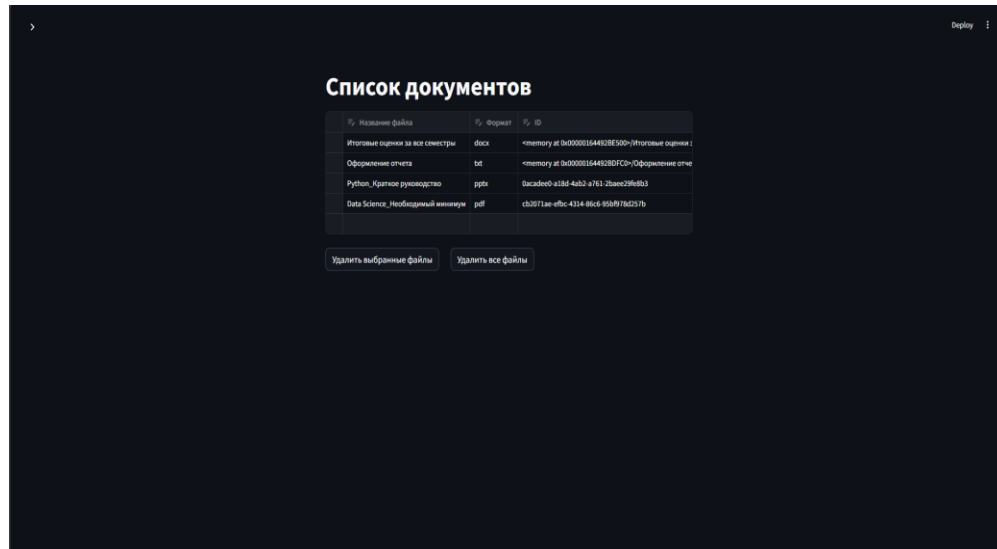
Тестирование



Тестирование



Тестирование



Заключение

В ходе выполнения работы был разработан программный комплекс интеллектуального поиска в корпоративных базах знаний, 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.