

KONSPECTO

LLM Агент для работы с конспектами

Неронов Роман Михайлович
Фазлыев Альберт Айратович

План доклада

1. Введение
2. Функциональные возможности
3. Архитектура системы
4. Технологический стек
5. Демонстрация
6. Список источников



01

ВВЕДЕНИЕ



Задача

*Эффективный поиск и
получение ответов из
больших объёмов конспектов*

Решение

*Создать LLM-агента для
семантического поиска и
генерации ответов на вопросы*



02

**ФУНКЦИОНАЛЬНЫЕ
ВОЗМОЖНОСТИ**

Функциональные ВОЗМОЖНОСТИ

Предоставление
структурированных ответов на
вопросы по образовательным
материалам, опираясь на
имеющуюся базу знаний.

Поиск по конспектам с
предоставлением точных
результатов и возможностью
просмотра исходных материалов.

Преобразование видео-лекций в
удобный формат конспекта с
автоматическим извлечением
ключевых кадров.

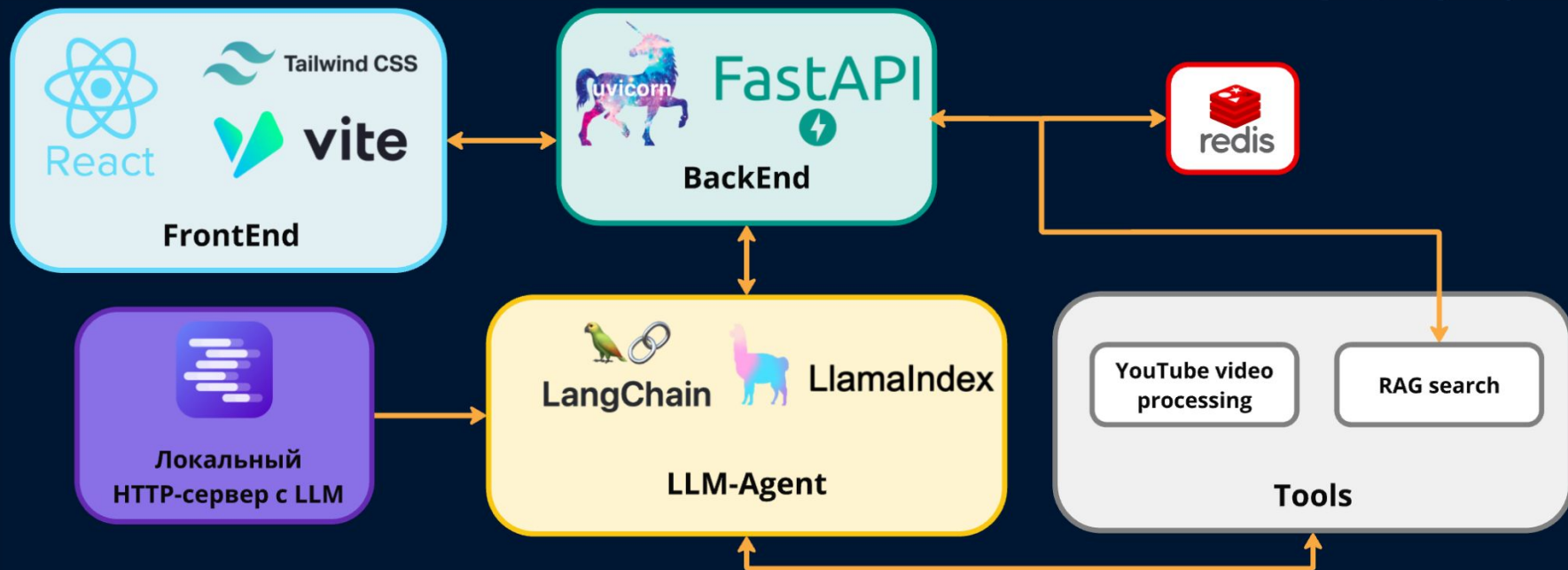
Обеспечение возможности
голосового ввода запросов на
русском языке.



03

**АРХИТЕКТУРА
СИСТЕМЫ**

Архитектура



Retrieval Augmented Generation

Конфигурация

- Vector Store – Redis
- Embedding model – intfloat/multilingual-e5-large
- Similarity top – 1
- Документы - пользовательские конспекты из Google Drive

LLM-агент: ReACT

Шаблон подсказки ReAct

```

Thought: ...

Action: ...

Observation: ...

... (Repeated many times)

```

Tools

RAGSearch: Поиск терминов и концепций по базе знаний

YouTubeToDocx: Создание документа со слайдами из YouTube видео

LLM - IlyaGusev/saiga_nemo_12b_gguf/saiga_nemo_12b.Q8_0.gguf



04

**ТЕХНОЛОГИЧЕСКИЙ
СТЕК**

Технологический стек

BACKEND

- FastAPI
- Redis
- Faster Whisper
- OpenCV
- Python-DOCX
- PyTubeFix

LLM-AGENT

- LangChain
- LlamaIndex
- Sentence Transformers
- Redis Vector Store

FRONTEND

- React
- Vite
- TailwindCSS
- Jest



05

Демонстрация

Ссылка на демонстрационное видео

<https://drive.google.com/file/d/1w7kXH8nV2uqvnqkOwmO5zURi39p-mr1T/view?usp=sharing>

(для лучшего качества скачайте видео)



06

**СПИСОК
ИСТОЧНИКОВ**

1. GitHub репозиторий KONSPECTO.

URL: <https://github.com/RomiconEZ/KONSPECTO-LLM>

2. docs.llamaindex.ai | Building a Live RAG Pipeline over Google Drive Files.

URL: https://docs.llamaindex.ai/en/v0.10.34/examples/ingestion/ingestion_gdrive

3. llamahub.ai | LlamaIndex Integration: Google Readers.

URL: <https://llamahub.ai/l/readers/llama-index-readers-google?from=readers>

4. docs.llamaindex.ai | Redis Docstore+Index Store Demo.

URL: <https://docs.llamaindex.ai/en/v0.10.34/examples/docstore/RedisDocstoreIndexStoreDemo>

Спасибо за внимание

Неронов Роман

e-mail: roman.nieronov@mail.ru