



Факультет компьютерных наук

БКНАД **232**

Москва **2025**

Сервис-помощник по поиску информации в видеолекциях

Курсовая работа | Программный проект

Выполнил студент группы БКНАД **232**
Романенко Никита Дмитриевич

Руководитель проекта
Сотрудник ФКН
Потапов Иван Андреевич



Описание предметной области

В работе представлен подход к автоматизированному поиску информации в видеолекциях на основе современных методов обработки аудиоданных и текстового поиска. Система предназначена для студентов, которым необходимо быстро находить нужные фрагменты в собственных видеозаписях лекций или семинаров. Решение включает инструменты для преобразования аудиозаписей в текст и эффективного поиска по временным промежуткам, что позволяет оптимизировать процесс подготовки к экзаменам и повторения учебного материала.



Актуальность работы

Существующие решения для поиска информации в видео не позволяют загружать собственные записи. Ручной просмотр лекций требует значительных временных затрат и не обеспечивает быстрого доступа к нужной информации. Дополнительную сложность создаёт объём и разнообразие учебных материалов, а также отсутствие удобных инструментов для поиска по содержимому видео. Всё это обуславливает необходимость разработки собственной системы, находить релевантные фрагменты по запросу пользователя с помощью современных методов обработки данных.



Цель и задачи работы

Цель:

Разработка сервиса, позволяющего студентам быстро находить нужные фрагменты в видеолекциях на основе текстового запроса.

Задачи:

- Исследовать и выбрать подходящую модель транскрипции аудиоданных.
- Реализовать быстрое преобразование аудио лекции в текст с сохранением временных меток.
- Построить эффективный механизм поиска релевантных фрагментов
- Внедрить систему ранжирования ответов на основе **LLM** и оптимизации промптов.
- Создать простой пользовательский сервис с возможностью быстрой развертки.



Анализ существующих решений



Преимущества:

- Предложение n количества видео по запросу
- Дополнительный поиск по выбранному видео
- Видимая транскрипция видео и выделение слов, которые говорятся

Недостатки:

- Недостаточная база видео
- Нельзя подгрузить собственное видео



Функциональные требования

- Загрузка и обработка пользовательских видеолекций, либо транскрипций
- Транскрипция аудио с привязкой текста ко времени
- Ввод текстового запроса и получение **3** релевантных временных фрагментов
- Быстрый отклик (около **5** минут на лекцию до **1:40** часов)
- Удобный интерфейс, работающий из браузера
- Создание ранжирующей системы
- Подготовка к развертыванию на сервере



Анализ существующих подходов

Модель транскрипции:

- GigaAM RNNT(Сбербанк)
- GigaAM CTC(Сбербанк)

Метод обработки аудио для транскрипции:

- Voice Activity Detection
- Chunk-Based Transcription

Поиск релевантных промежутков

- IVF
- PQ
- IndexFlatL2
- HNSW

Оптимизация промптов:

- DSPy: OPRO
- DSPy: MIPROv2



Выбор методов и технологий

• Модель транскрипции – **GigaAM CTC**(Сбербанк)

Для преобразования аудио в текст используется модель **GigaAM** с архитектурой **Connectionist Temporal Classification (CTC)**. Эта модель эффективно справляется с задачей распознавания речи, обеспечивает высокую скорость транскрипции и хорошо подходит для обработки длинных лекционных записей.

• Метод обработки аудио для транскрипции – **Chunk-Based Transcription**

Аудиофайл разбивается на небольшие равные по длине фрагменты (чанки) с небольшим перекрытием между ними. Каждый чанк обрабатывается отдельно, что позволяет ускорить транскрипцию и избежать обрезания слов или предложений на границах сегментов.

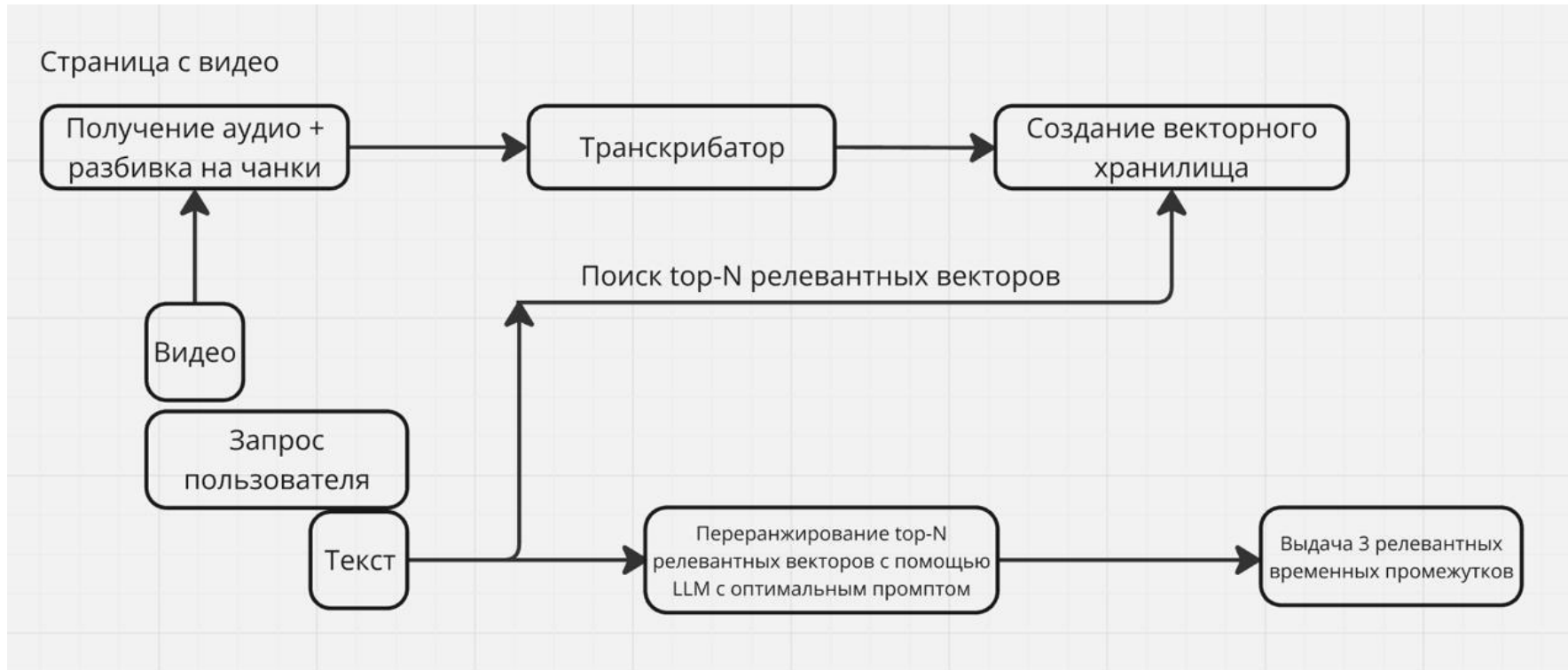
• Поиск релевантных промежутков – **IndexFlatL2**

Все фрагменты текста представляются в виде векторов, и при поиске выбираются те, которые имеют минимальное евклидово расстояние (**L2**) до вектора пользовательского запроса. Это обеспечивает быструю и точную выдачу наиболее подходящих временных промежутков.

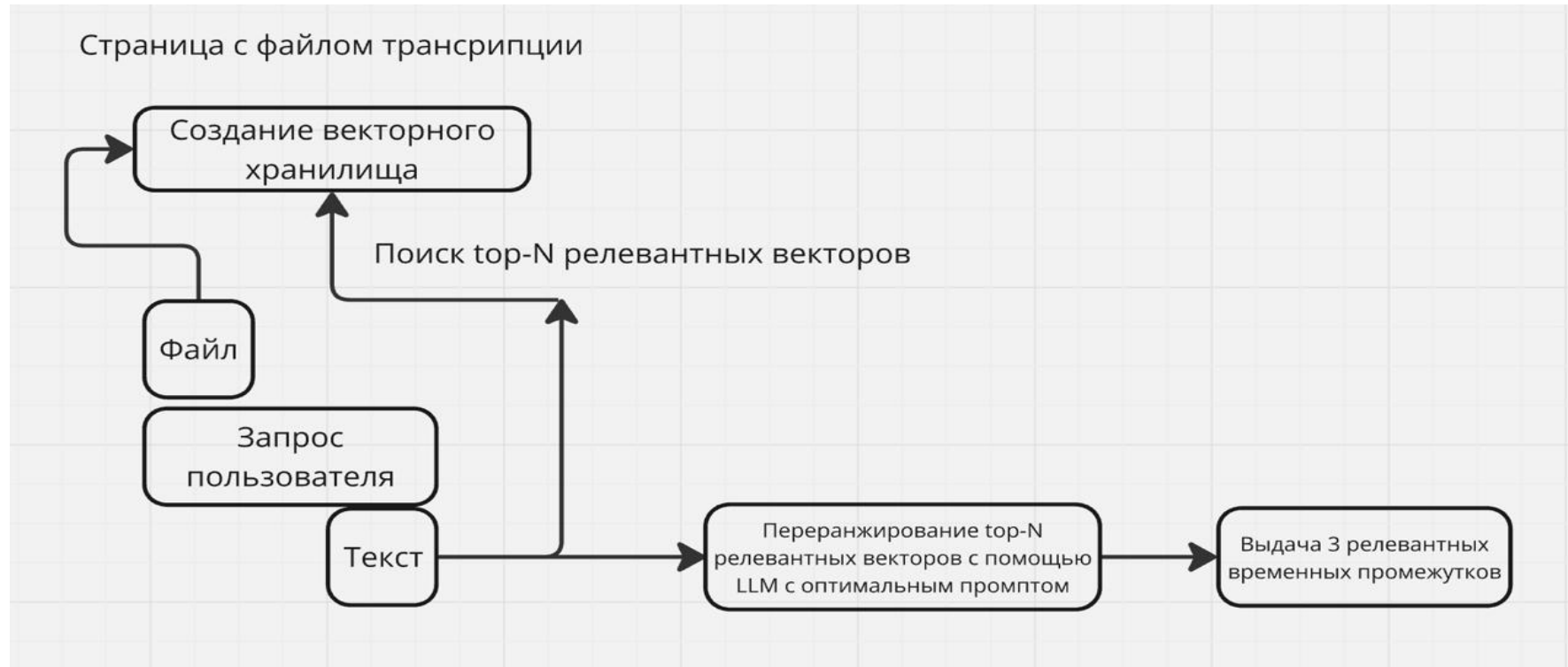
• Оптимизация промптов – **DSPy: MIPROv2**

Эта система позволяет подобрать наилучшие инструкции и примеры для взаимодействия с языковой моделью, что существенно увеличивает точность поиска и релевантность выдаваемых ответов.

Архитектура приложения



Архитектура приложения





Эксперимент по выбору метода для обработки аудио

	VAD (Речевая активность)	Разбиение на чанки
Получение временных промежутков, с	52	
Разбиение на wav файлы, с	95	45
Получение всех транскрипций, с	697	270
Общее время, с	844	315



Эксперимент по выбору системы ранжирования

Система	асс@1 до	асс@1 после	асс@3 до	асс@3 после
С перефразированием	0.38	0.786	0.548	0.881
Без перефразирования	0.333	0.762	0.595	0.857



Основные результаты и выводы

- Реализован сервис с поиском **3** релевантных временных промежутков по видеолекциям
- Внедрение модели **GigaAM-CTC** от «Сбербанка» и получение транскрипции за **3** минуты
- Внедрение и улучшение модели ранжирования временных промежутков
- Сервис готов к развертыванию на сервере



Направление дальнейшей работы

- Реализация поиска по нескольким видеолекциям одновременно
- Дальнейшая оптимизация промптов для повышения качества результатов
- Расширение и автоматизация сбора обучающих данных
- Поиск новых подходов для сокращения времени обработки и расшифровки видеозаписей

Список используемой литературы

- **Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazare, Maria Lomeli, Lucas Hosseini и Herve Jegou. “THE FAISS LIBRARY”. В: arXiv preprint (2024).**
- **Alex Graves. “Sequence Transduction with Recurrent Neural Networks”. В: arXiv preprint (2012).**
- **Alex Graves, Santiago Fern, Faustino Gomez и Jürgen Schmidhuber. “Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks”. В: ICML (2006).**
- **Krista Opsahl-Ong, Michael J Ryan, David Broman Josh Purtell, Christopher Potts, Matei Zaharia и Omar Khattab. “Optimizing Instructions and Demonstrations for Multi-Stage Language Model Programs”. В: arXiv preprint (2024).**



Демонстрация работы сервиса

Демонстрация страницы с загрузкой видео

Помощник по видео

Загрузите видео лекцию или любой файл в формате `.mp4`.

Система автоматически создаст текстовую расшифровку. После этого вы сможете задавать сколько угодно вопросов по содержанию видео.

В ответ вы получите до **трёх временных промежутков**, где, скорее всего, встречается нужная информация. Первый вариант — самый точный, следующие — менее уверенные.

Возможна погрешность до **двух минут** относительно точного момента в видео.

Длительность обработки видео **около 5 минут**

Загрузите видеофайл лекции

Выберите видеофайл



Drag and drop file here

Limit 200MB per file • MP4, MPEG4

Browse files



линал_лекция_24.mp4 151.9MB



51% обработано

Система автоматически создаст текстовую расшифровку. После этого вы сможете задавать сколько угодно вопросов по содержанию видео.

В ответ вы получите до **трёх временных промежутков**, где, скорее всего, встречается нужная информация. Первый вариант — самый точный, следующие — менее уверенные.

Возможна погрешность до **двух минут** относительно точного момента в видео.

Длительность обработки видео **около 5 минут**

Загрузите видеофайл лекции

Выберите видеофайл



Drag and drop file here

Limit 200MB per file • MP4, MPEG4

Browse files



линал_лекция_24.mp4 151.9MB



Введите интересующий вас запрос по видео:

Неравенство Коши Буняковского

Найденные временные промежутки:

1. Время: 19:05

2. Время: 26:35

3. Время: 19:55

Демонстрация страницы с загрузкой транскрипции


ответ вы получите три временных промежутка, в которых с высокой вероятностью встречается информация, связанная с вашим запросом.

- Первый интервал — наиболее релевантный;
- Второй и третий — возможные дополнения, с чуть меньшей точностью;
- Возможна погрешность до 2 минут относительно точного упоминания.


Этот режим особенно удобен для тестирования или работы с уже расшифрованными видео.

Загрузите файл для работы

Выберите файл с транскрипцией

 Drag and drop file here
Limit 200MB per file • JSON

Browse files

 линал_лекция_24.json 111.2KB ×

Введите интересующий вас запрос:

Неравенство Коши Буняковского

Найденные временные промежутки:

1. Время: 19:05
2. Время: 26:35
3. Время: 21:35



Видео демонстрации





ФКН

БКНАД **232**

Москва **2025**

20

Спасибо за внимание!

