



Нейросетевой помощник “Нейротьютор”

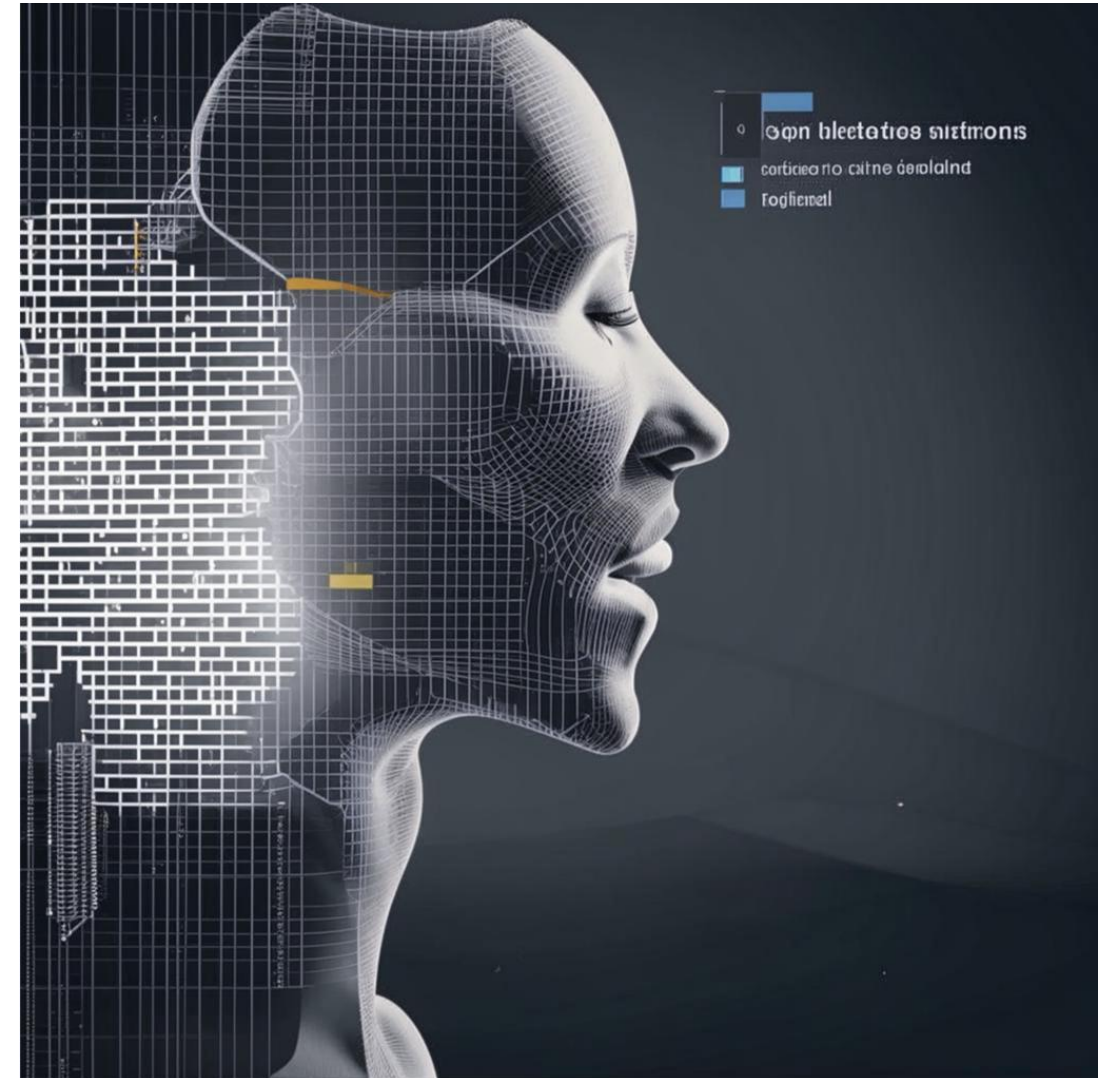
Постановка задачи

Разработать нейросетевой помощник Нейротьютор для автоматической обработки образовательного контента с платформ Dzen и VK.

Система должна обеспечивать высокую точность распознавания речи и сохранение смысловой нагрузки при обработке лекций.

Требуется реализовать функционал пересказа лекций разной степени подробности.

Ограничения включают работу только с русскоязычным контентом и зависимость качества от исходного аудио.





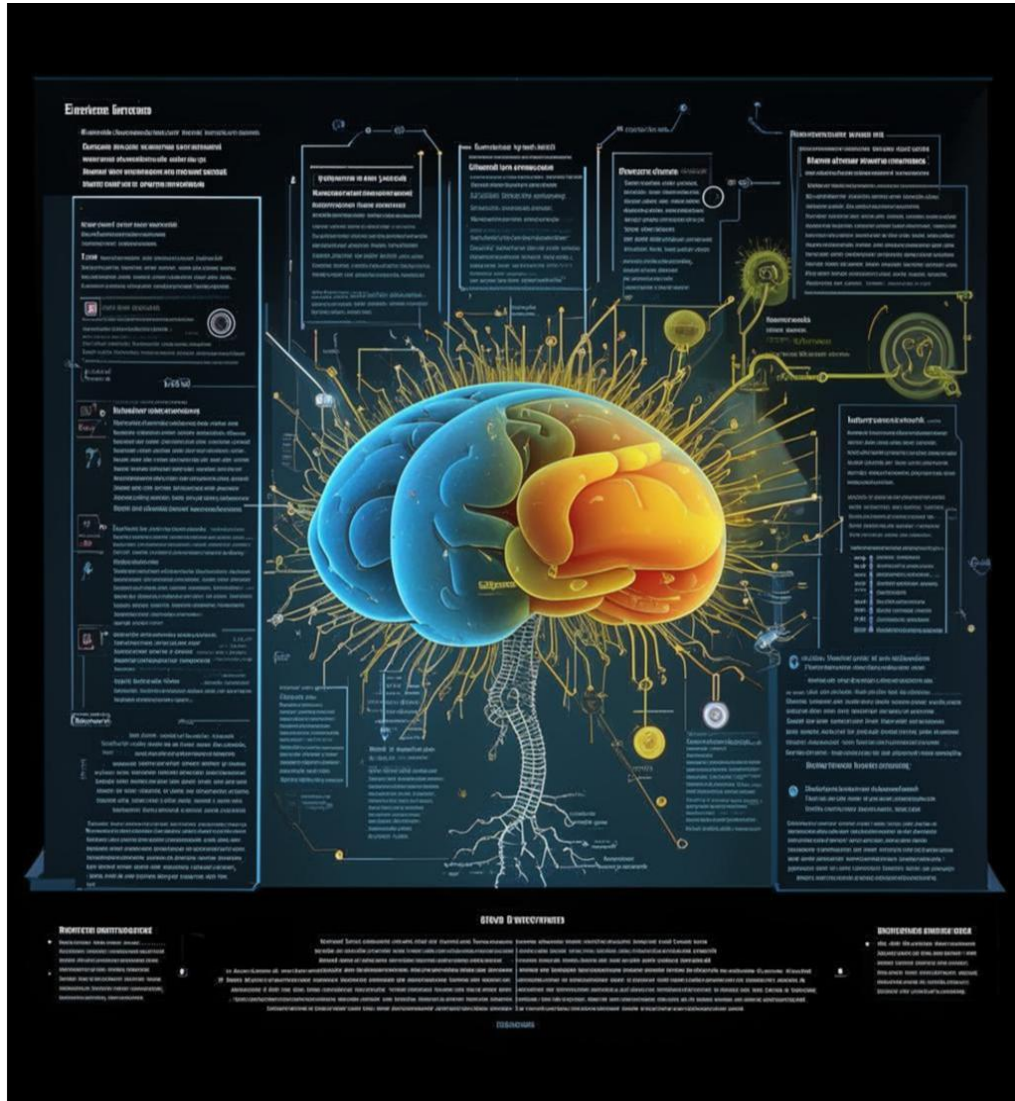
Цель

Основная цель проекта “Нейротьютора” - создание интеллектуальной системы для автоматической обработки и анализа образовательного контента с платформ Dzen и VK.

Система направлена на повышение доступности и эффективности работы с лекционными материалами за счет их автоматической транскрибации, структуризации и формирования базы знаний.

Задачи

- Транскрибация аудио:
 - Преобразование аудиодорожек видеолекций в текст
 - Обеспечение точности и качества транскрибации.
- Формирование базы знаний:
 - Создание структурированной базы знаний на основе транскрибированных текстов.
 - Организация данных по дисциплинам, темам и лекциям для удобства поиска и анализа.
- Генерация ответов:
 - Использование языковой модели GPT для формирования развернутых ответов на экзаменационные вопросы.
 - Создание пересказов лекций разной длины на основе анализа релевантных фрагментов текста.
- Интерфейс взаимодействия:
 - Разработка интерфейса для добавления новых данных, запросов и получения ответов.



Обучающая база



Объем базы
37 видеолекции

Источник сбора данных
Платформы Dzen и VK

Инструменты для сбора данных
Библиотеки requests, BeautifulSoup, requests-html и yt-dlp для извлечения аудиодорожек из онлайн-лекций. Для обработки и транскрипции аудио применяется модель Faster Whisper

Возникшие трудности
при сборе базы
Прямые ссылки на аудио (s3.dzeninfra.ru) больше не доступны
Разработан новый метод получения аудиоданных через HLS-поток

Возможности системы

1. Распознавание аудио и видео:
 - Транскрибация лекций с использованием Faster Whisper.
 - Поддержка платформ Dzen и VK.
2. Формирование базы знаний:
 - Разбиение текста на чанки (фрагменты).
 - Создание эмбеддингов с использованием FAISS.
3. Генерация ответов:
 - Использование GPT для формирования развернутых ответов на вопросы.
 - Пересказ лекций в различных объемах (очень короткий, краткий, подробный).
4. Интеграция:
 - Возможность добавления новых лекций через форму.
 - Хранение данных на Google Drive.

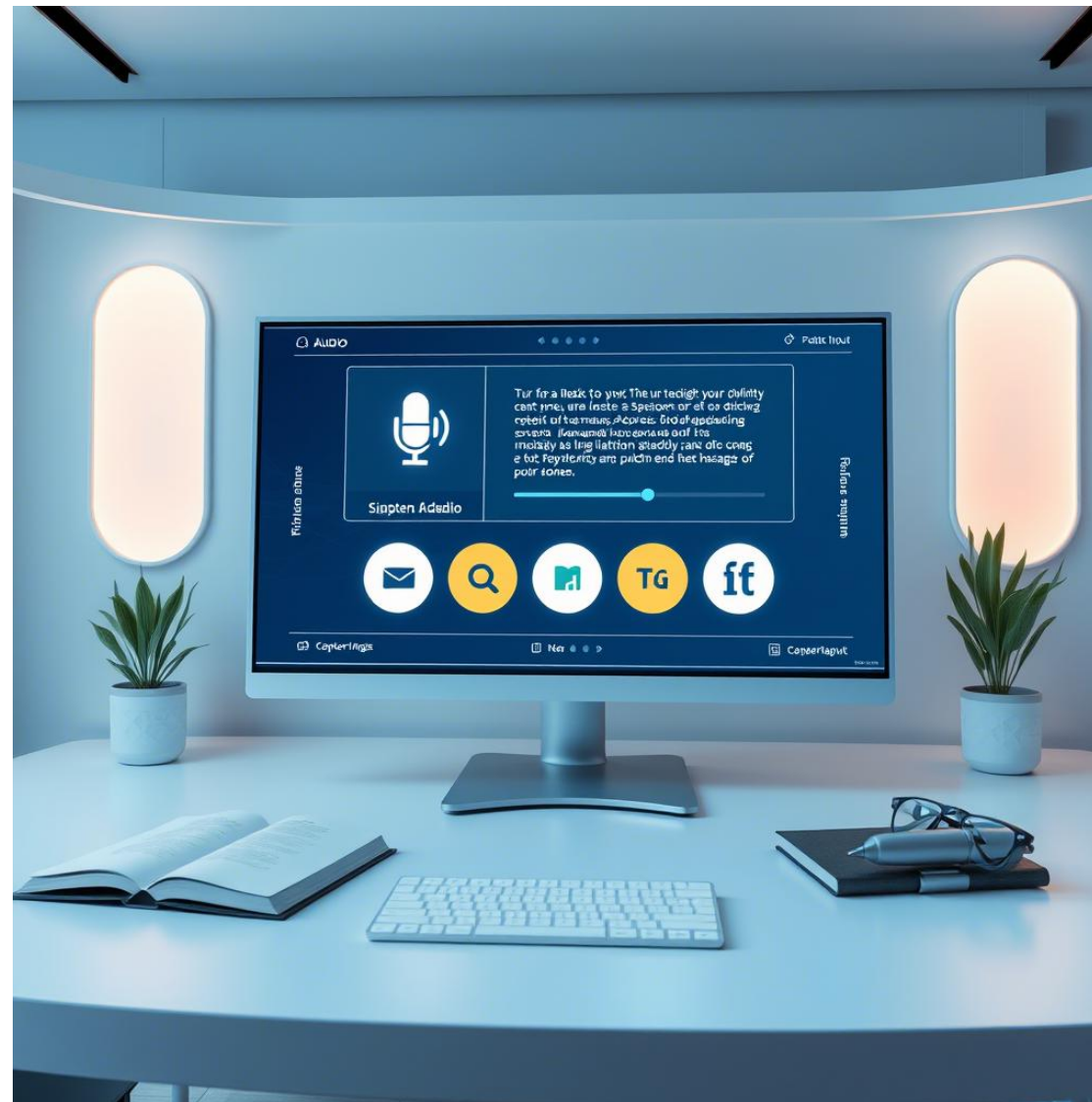
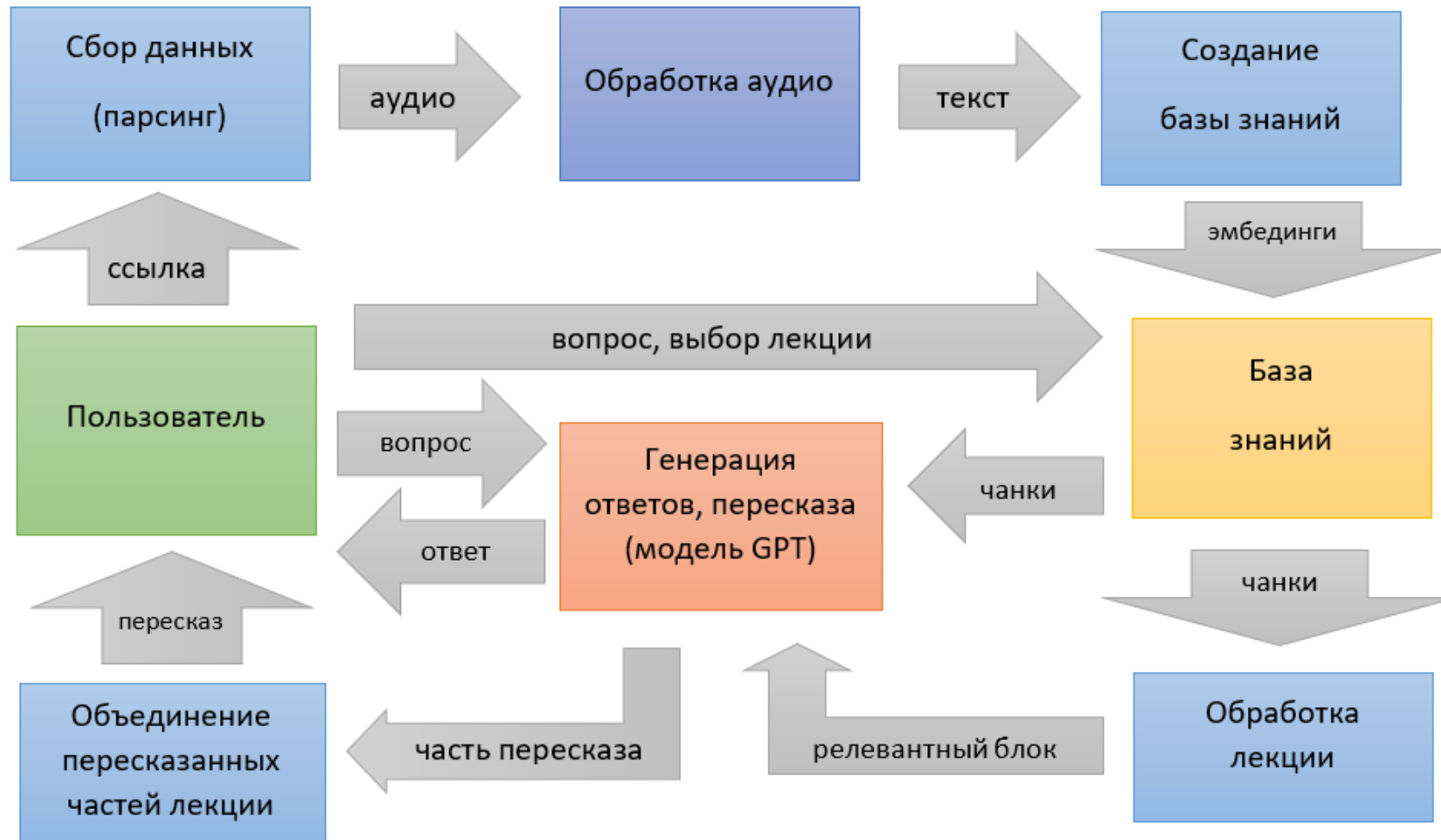


Схема работы интеллектуальной системы



Пользовательский интерфейс

Технологический стек интерфейса:

- Jupyter Notebook
- IPython widgets
- HTML/CSS



IPYWIDGETS



> Конфигурация HC

1:

pk1_url:

faiss_url:

1:

model_gpt:

temperature:

system:

k_chunks:

1:

model_whisper:

use_gpu: ☐

> Форма для добавления лекции в базу знаний

[5] [Показать код](#)

Дисциплина:

Тема лекц...:

Ссылка на...:

```

Директория создана: /content/faiss_data
Файлы не найдены. Начинаю загрузку...
Downloading...
From: https://drive.google.com/uc?id=1-0mfXxZkUhEWIYq7cDXPgWikdnZ3tG0n
To: /content/faiss_data/index.pkl
100% 2.11M/2.11M [00:00<00:00, 76.6MB/s]
Downloading...
From: https://drive.google.com/uc?id=1-DOm6s5Rp_qYi0rrotTVVc9ec0R-BkTC
To: /content/faiss_data/index.faiss
100% 6.33M/6.33M [00:00<00:00, 124MB/s]
Файлы успешно загружены.
База знаний успешно загружена.
База знаний готова к использованию!
Добавьте лекцию в базу знаний!
  
```


Пользовательский интерфейс

Компоненты интерфейса:

- Формы ввода данных (Text, Textarea)
- Выпадающие списки (Dropdown)
- Кнопки управления (Button)
- Слайдеры (FloatSlider)
- Области вывода информации (Output)

IPYWIDGETS

</>
HTML & CSS

Google
colab

> Форма для ответа на вопрос пользователя

Показать код



Вопрос: Как влияет длительное подавление негативных эмоций (таких как гнев или тревога) на развитие психологических механизмов защиты у личности, и может ли это привести к формированию соматических расстройств через дисфункцию вегетативной нервной системы?

Отправить вопрос

Ответ GPT:

Длительное подавление негативных эмоций, таких как гнев или тревога, может оказать значительное влияние на психологическое состояние личности и привести к различным защитным механизмам. Когда человек не позволяет себе выражать свои эмоции, он может начать использовать различные психологические защиты, чтобы справиться с внутренним конфликтом. Эти механизмы могут включать отрицание, подавление, проекцию и рационализацию.

Подавление эмоций, в частности, может привести к накоплению внутреннего напряжения и стресса. Как упоминалось в лекциях, стресс является неспецифическим ответом организма на требования окружающей среды. Если человек не справляется с этими требованиями и не может выразить свои эмоции, это может вызвать состояние эмоционального выгорания и привести к снижению уровня психологического благополучия. В результате, человек может начать ощущать себя неэффективным в своей профессиональной деятельности, что усугубляет его эмоциональное состояние.

Кроме того, длительное подавление негативных эмоций может вызвать дисфункцию вегетативной нервной системы. Вегетативная нервная система отвечает за регуляцию многих автоматических функций организма, таких как сердечный ритм, дыхание и пищеварение. Когда человек находится в состоянии постоянного стресса и подавления эмоций, это может привести к нарушениям в работе этой системы, что, в свою очередь, может вызвать соматические расстройства. Например, это может проявляться в виде головных болей, проблем с пищеварением, повышенного артериального давления и других физических симптомов.

> Формы для пересказа лекции - новая

[8] Показать код



Лекция: 1371! Афанасьев В.В. Страте

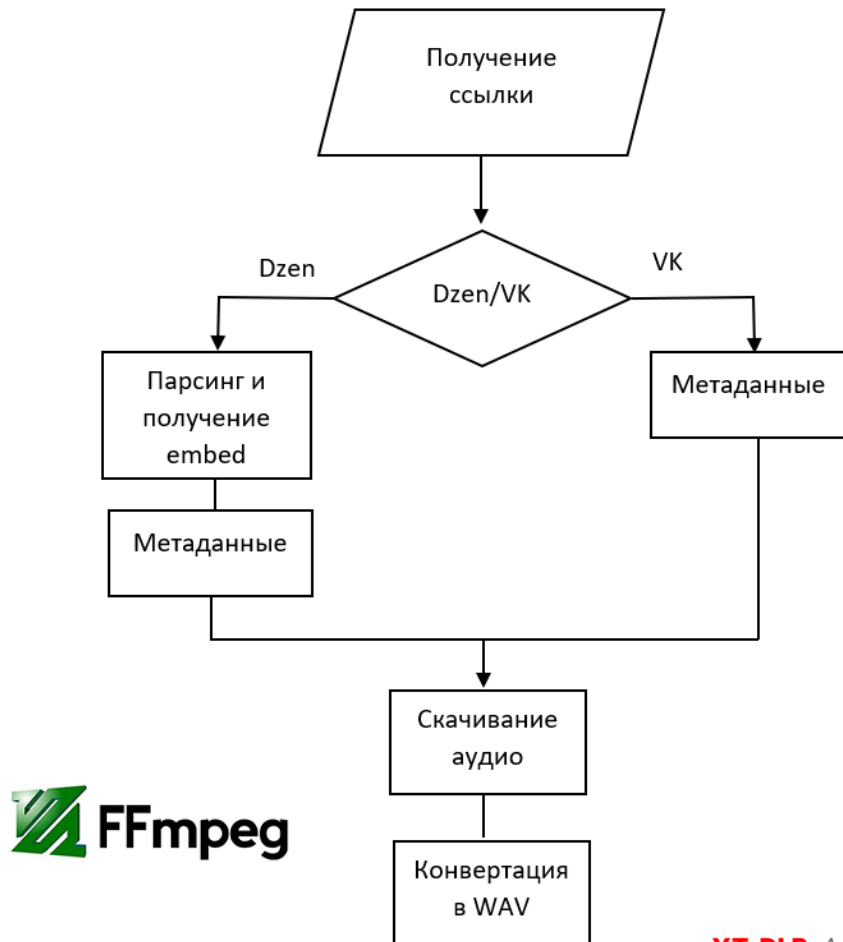
Длина пер... Очень короткий

Температу... 0.10

База знаний готова к использованию!
Выберите лекцию и запустите пересказ!

Сгенерировать пе...

Модуль сбора данных (парсинг)



Используются библиотеки Requests-HTML и BeautifulSoup, которые извлекают ссылки на медиафайлы с платформ Dzen, а также получают метаданные о контенте. Скачивание аудио осуществляется через yt-dlp и ffmpeg, где видеофайлы конвертируются в аудиодорожки в формате WAV перед дальнейшей обработкой.

Requests-HTML

a beautiful soup

Технологии и инструменты парсинга и извлечения данных:

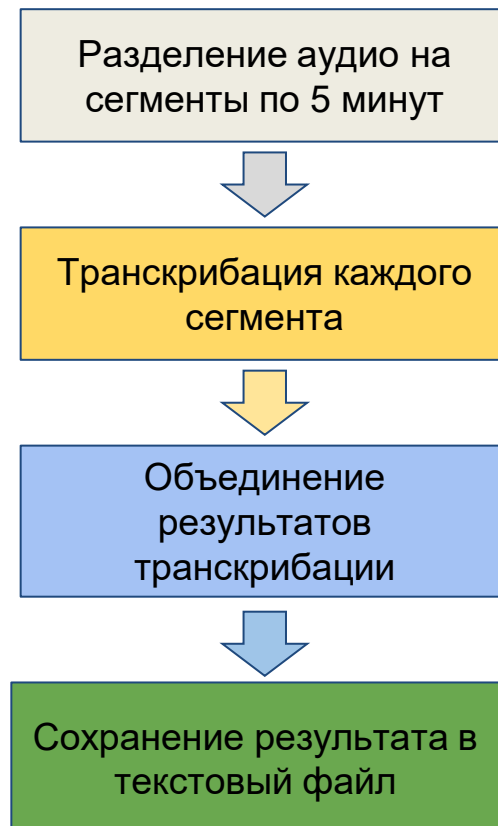
- Requests-HTML (для работы с HTML)
- BeautifulSoup (для парсинга)
- yt-dlp (для работы с медиаконтентом)
- AsyncHTMLSession (для асинхронных запросов)
- FFmpeg для преобразования аудио в WAV с помощью

> **dlp**

YT-DLP A youtube-dl fork with additional features and fixes

Asycio IN Python

Модуль обработки аудио



Используемые технологии и инструменты:

1. Обработка аудио:
 - pydub: для работы с аудиофайлами
 - subprocess: для вызова FFmpeg команд
 - tempfile/io: для временного хранения сегментов
2. Транскрибация:
 - faster-whisper: модель для распознавания речи
 - ctranslate2: для оптимизации работы модели
 - int8: вычисления для повышения производительности
3. Обработка текста:
 - re: для работы с регулярными выражениями
 - json: для работы со структурированными данными
4. Вспомогательные инструменты:
 - tqdm: для отображения прогресса
 - os/shutil: для работы с файловой системой
 - numpy: для численных операций
 - scipy: для различных вычислений



NumPy

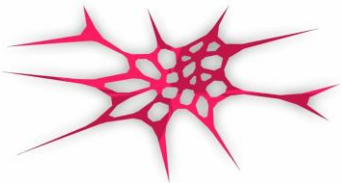
Модуль создания базы знаний



LangChain

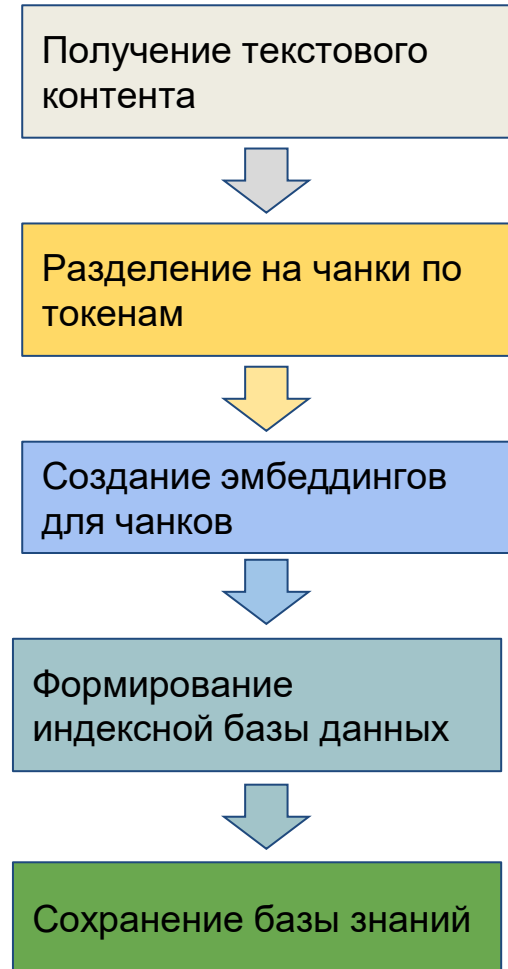
FAISS

Scalable Search With Facebook AI

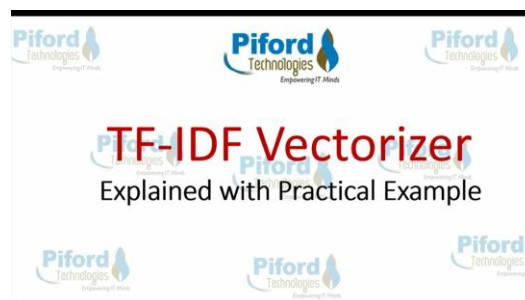
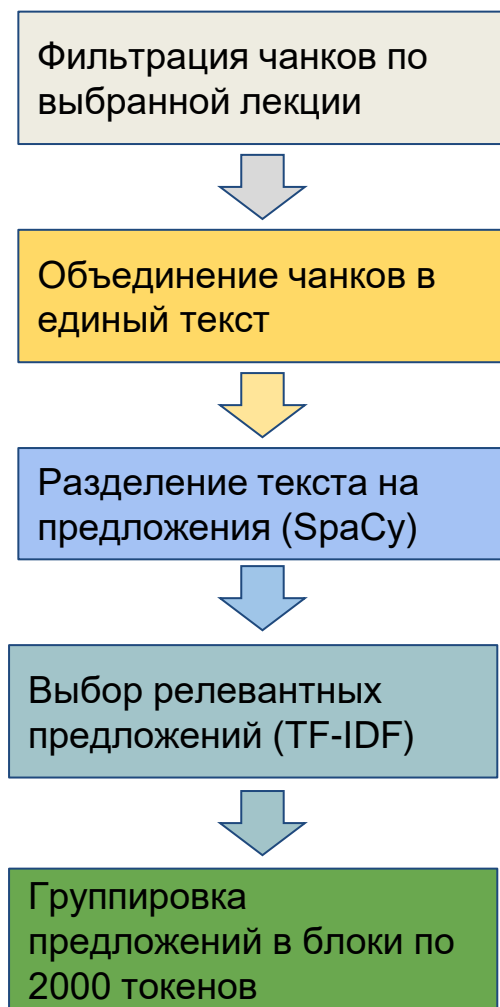


Используемые технологии и инструменты:

1. Обработка текста:
 - MarkdownHeaderTextSplitter: для разделения текста по заголовкам
 - RecursiveCharacterTextSplitter: для разбиения текста на чанки
 - tiktoken: для подсчета токенов
2. Создание эмбеддингов:
 - OpenAIEmbeddings: для генерации векторных представлений текста
 - FAISS (Facebook AI Similarity Search): для создания индексной базы данных
3. Хранение данных:
 - FAISS: для хранения векторной базы данных
 - pickle (.pkl): для сохранения метаданных
 - Google Drive: для хранения файлов базы знаний
4. Вспомогательные инструменты:
 - os: для работы с файловой системой
 - re: для работы с регулярными выражениями
 - json: для работы со структурированными данными
 - tqdm: для отображения прогресса



Модуль обработки лекций для пересказа



spaCy

FAISS
Scalable Search With Facebook AI



Используемые технологии и инструменты

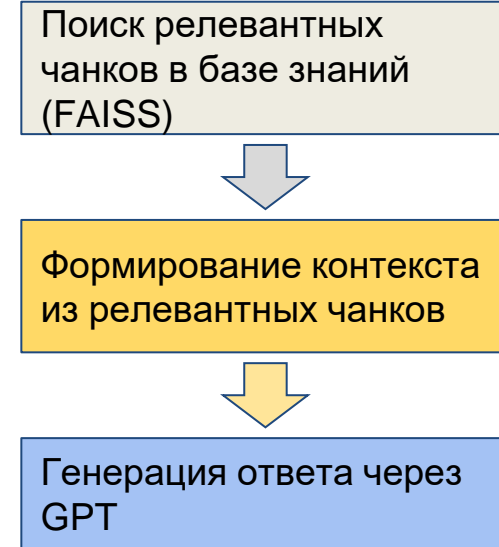
1. Обработка текста:
 - SpaCy - Используется для разделения текста на предложения.
 - TfidfVectorizer (sklearn) - Выбор наиболее релевантных предложений на основе TF-IDF.
 - FAISS - База знаний используется для фильтрации чанков по выбранной лекции.
2. Вспомогательные инструменты:
 - time - Для задержки между запросами к API.
 - re - Для очистки текста от лишних символов.



Модуль GPT

Используемые технологии и инструменты

1. Поиск релевантных чанков:
 - FAISS (Facebook AI Similarity Search) :
 - База данных для быстрого поиска наиболее релевантных фрагментов текста.
 - Используется для нахождения k ближайших чанков к запросу пользователя.
2. Генерация текста:
 - OpenAI API :
 - Модель GPT (gpt-4o-mini, gpt-4o, gpt-4) используется для генерации ответов.
 - Настройка параметров, таких как температура, максимальное количество токенов.
 - tiktoken :
 - Подсчет количества токенов для контроля размера входных данных.



Эксперимент №1 — Сравнение моделей Faster Whisper

Таблица результатов сравнения моделей

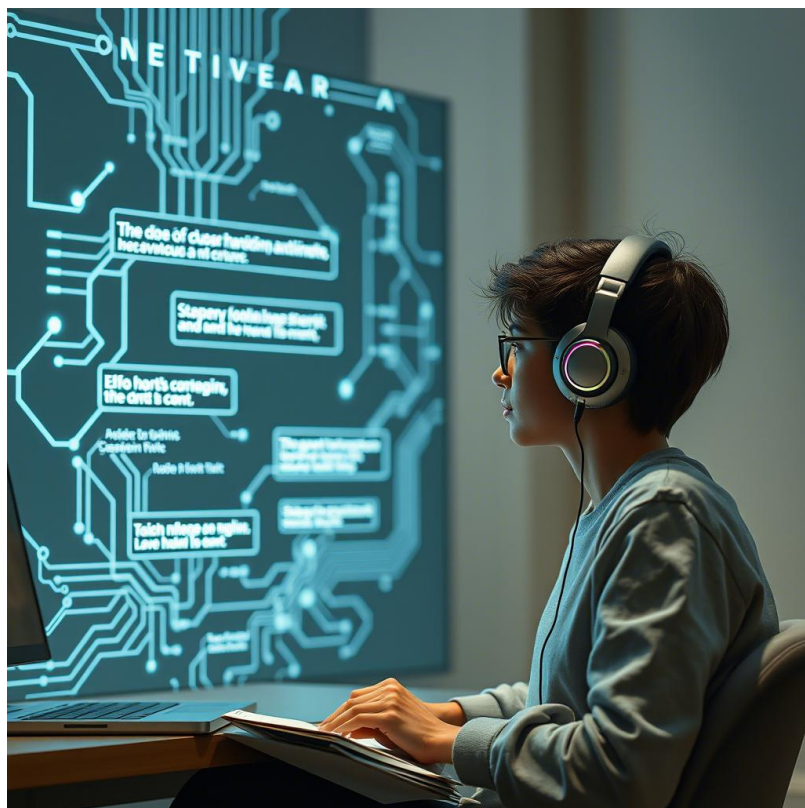
Модель small vs medium	WER: 40.90%	CER: 8.37%
Модель small vs large-v3	WER: 45.13%	CER: 8.04%
Модель medium vs large-v3	WER: 22.24%	CER: 5.71%

Основные выводы:

- Модели medium и large-v3 наиболее схожи
- Small значительно отличается от остальных
- Увеличение размера модели не всегда дает большой прирост качества



Эксперимент №2 — Функциональность системы



Методы:

- Скачивание аудио из VK.
- Распознавание текста с помощью Faster Whisper.
- Добавление распознанной лекции в существующую базу.

Результаты:

- Успешное скачивание и обработка короткой лекции (105 секунд).
- Добавление лекции в базу подтверждено.

Вывод:

Компоненты системы работают корректно.

Эксперимент №3 — Анализ влияния параметров на ответы GPT



Параметры:

- $k=7$, $temperature=0.1$: Подробный, структурированный ответ.
- $k=3$, $temperature=0.9$: Короткий, креативный ответ.

Результаты:

- Высокое значение k увеличивает объем и детализацию ответа.
- Высокая $temperature$ делает стиль более креативным, но менее стабильным.

Вывод:

Для научных запросов рекомендуется $k=7$, $temperature=0.1$.

Для творческих задач — $k=3$, $temperature=0.9$.

Эксперимент №4 — Пересказ лекций

Параметры:

- Подробный пересказ:
 - $k_{\text{recap}} = 0.7$, temperature = 0.9.
 - Количество предложений: 74.
 - Количество токенов: 3031.
- Краткий пересказ:
 - $k_{\text{recap}} = 0.1$, temperature = 0.1.
 - Количество предложений: 10.
 - Количество токенов: 1310.

Результаты:

- Подробный пересказ охватывает все ключевые темы, включая второстепенные аспекты.
- Краткий пересказ фокусируется только на основных идеях.

Вывод:

Для глубокого анализа рекомендуется Подробный пересказ .

Для быстрого обзора — Краткий пересказ .

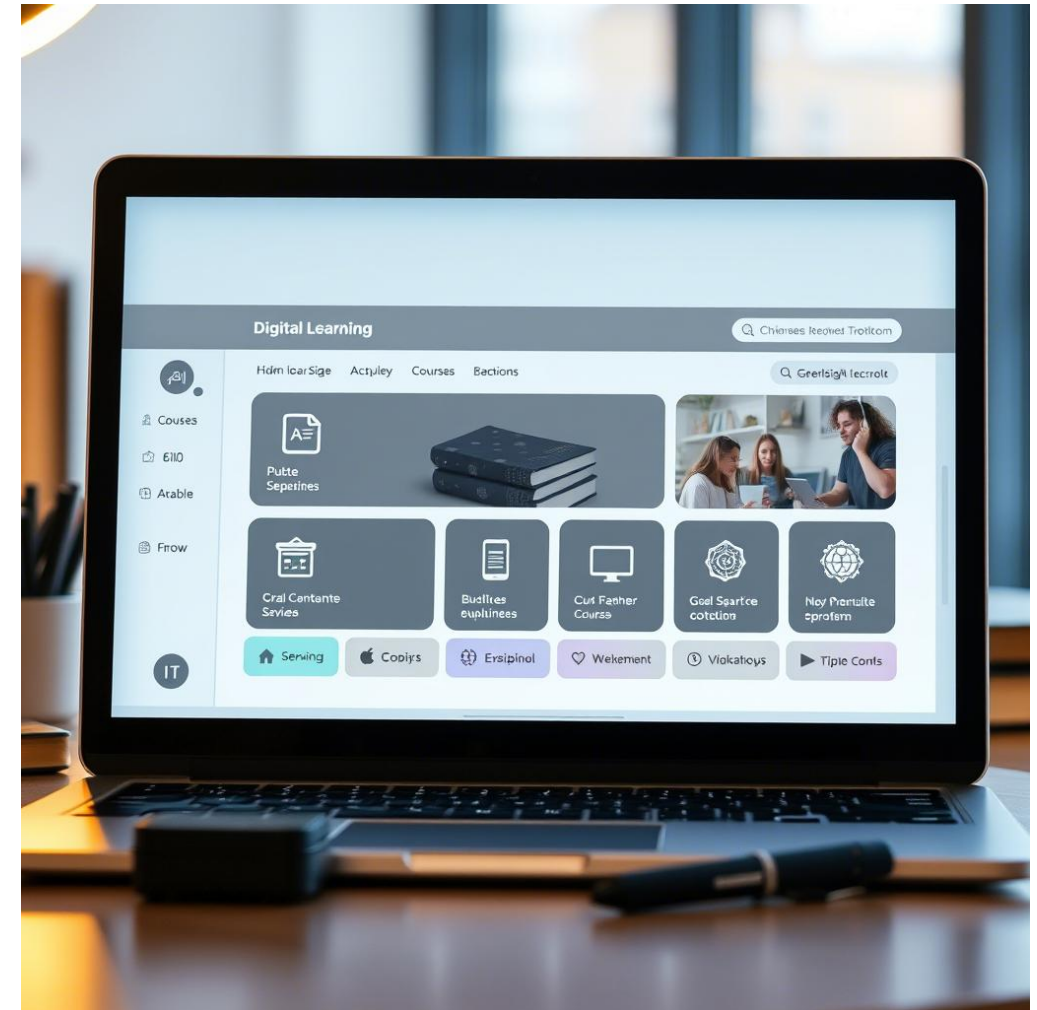


Вывод

Нейротьютор представляет собой комплексное решение для обработки образовательного контента, включающее автоматический сбор данных с платформ Dzen и VK, их транскрибацию с помощью Faster Whisper и формирование структурированной базы знаний.

Внедрение новых методов работы с HLS-потоками значительно повысило надежность сбора аудиодорожек, а оптимизированные алгоритмы обработки текста обеспечивают качественное формирование пересказов различной степени подробности.

Система успешно демонстрирует свою эффективность при работе с базой из 37 видеолекций, сохраняя высокую точность распознавания речи и качество генерируемых ответов





Заключение

Нейротьютор представляет собой эффективную систему для работы с образовательным контентом, демонстрирующую высокую точность транскрибации и качество формирования базы знаний.

Внедрение новых методов обработки данных, таких как работа с HLS-потокom, значительно повысило стабильность и надежность системы.

Дальнейшее развитие проекта может включать расширение функционала пересказов и оптимизацию алгоритмов обработки данных.

Спасибо за внимание!