



Хакатон «Транснефть — Технологии» по искусственному интеллекту

Кейс: ASR

Проект: Voice Transcriber Pro

Команда:

Федоров Максим

Федоров Дмитрий

Полубарьев Михаил

Самсоненко Станислав

ПОСТАНОВКА ЦЕЛИ И ЗАДАЧ

Цель проекта:

- Создать веб-приложение для автоматического распознавания речи и интеллектуальный тайм-трекер, которые преобразуют аудиофайлы в текст и фиксируют трудозатраты по голосовым командам.

Задачи проекта:

- Разработать веб-приложение для преобразования аудиофайлов в текст с поддержкой диаризации и транскрипции в реальном времени.
- Создать интеллектуальный тайм-трекер для фиксации трудозатрат сотрудников на основе голосовых команд, с выводом данных в структурированном формате.

ВЫБОР ТЕХНОЛОГИЙ РАЗРАБОТКИ



OpenAI Whisper



Perplexity AI



Python



Node.js

Whisper используется для автоматического распознавания речи и преобразования аудио в текст. Python обеспечивает реализацию серверной логики, включая обработку данных и интеграцию моделей. Node.js используется для создания веб-сервера и управления пользовательским интерфейсом. Perplexity применяется для анализа транскрипции и структурирования данных в удобном формате.

АРХИТЕКТУРА ПРИЛОЖЕНИЯ

Архитектура приложения состоит из следующих компонентов:

- **Веб-интерфейс** реализован на HTML, CSS и JavaScript, обеспечивает загрузку аудиофайлов, отображение результатов транскрипции и формализованных данных.
- **Серверная часть** - Flask (Python) обрабатывает запросы от веб-интерфейса, управляет транскрипцией аудио с использованием модели Whisper и анализирует данные через Perplexity API.
- **Модели и обработка данных:** Whisper используется для преобразования аудио в текст, Perplexity API — для структурирования текста в формате JSON.
- Результаты транскрипции и анализа сохраняются в **JSON-файлах** на сервере для дальнейшего использования.

```
> data
> results
v src
  > __pycache__
  📄 app.py
  📄 json_analyzer.py
  📄 openai_model.py
  📄 perplexity_model.py
v web
  > .vscode
  > node_modules
  > public
  > server
  {} package-lock.json
  {} package.json
  ⚙ .env
  ≡ .gitignore
  ⓘ README.md
  ≡ requirements.txt
```

СЕРВЕРНАЯ ЧАСТЬ ПРИЛОЖЕНИЯ

```
127.0.0.1 - - [13/Dec/2024 15:24:46] "OPTIONS /transcribe-audio HTTP/1.1" 200
Используемая модель:
Filename: ..\data\2.mp3
Model_name: Whisper large
Company: OpenAI
Architecture: Transformer кодировщик-декодер
Готово!
Результат транскрипции сохранен в файл: ..\results\transcription.json
```

POST localhost:5000/transcribe-audio

Params Authorization Headers (8) **Body** Scripts Settings

☐ none ☐ form-data ☐ x-www-form-urlencoded ☒ raw ☐ binary ☐ GraphQL **JSON** ▾

```
1 {
2   "filename": "1.mp3"
3 }
```

Body Cookies Headers (6) Test Results ↻

Pretty Raw Preview Visualize **JSON** ▾ ⌵


```
1 {
2   "formalized_data": {
3     "duration": "2 hours 30 minutes",
4     "end_time": "12:30",
5     "start_time": "10:00",
6     "task": "worked on a programming project"
7   },
8   "transcription": {
9     "audio_file_name": "1.mp3",
10    "language": "ru",
11    "model_info": {
12      "architecture": "Transformer кодировщик-декодер",
13      "company": "OpenAI",
14      "filename": "..\\data\\1.mp3",
15      "model_name": "Whisper large"
16    },
17    "segments": [
18      {
19        "avg_logprob": -0.09201018583206903,
```

```
{
  "text": " Я проводил обучение сотрудников 5 часа дня до половины третьего.",
  "segments": [
    {
      "id": 0,
      "seek": 0,
      "start": 0.0,
      "end": 3.66,
      "text": " Я проводил обучение сотрудников 5 часа дня до половины третьего.",
      "tokens": [ ...
    ],
    "temperature": 0.0,
    "avg_logprob": -0.03095232021240961,
    "compression_ratio": 1.2688172043010753,
    "no_speech_prob": 0.05505261942744255
  ]
},
"language": "ru",
"model_info": {
  "filename": "..\\data\\5.mp3",
  "model_name": "Whisper large",
  "company": "OpenAI",
  "architecture": "Transformer кодировщик-декодер"
},
"audio_file_name": "5.mp3"
```

ВЕБ-ИНТЕРФЕЙС ПРИЛОЖЕНИЯ

Voice Transcriber Pro



Преобразуйте голос в текст с помощью ИИ




Перетащите аудиофайл сюда или

Выберите файл

Обработать файл



 4.mp3 

Обработка... 


Voice Transcriber Pro

Преобразуйте голос в текст с помощью ИИ


Обработка завершена успешно

 2.mp3 

Обработать файл

 Результаты обработки

Информация о файле

 2.mp3

Транскрибированный текст

С 2 до 3.30 я занимался дизайном нового макета.

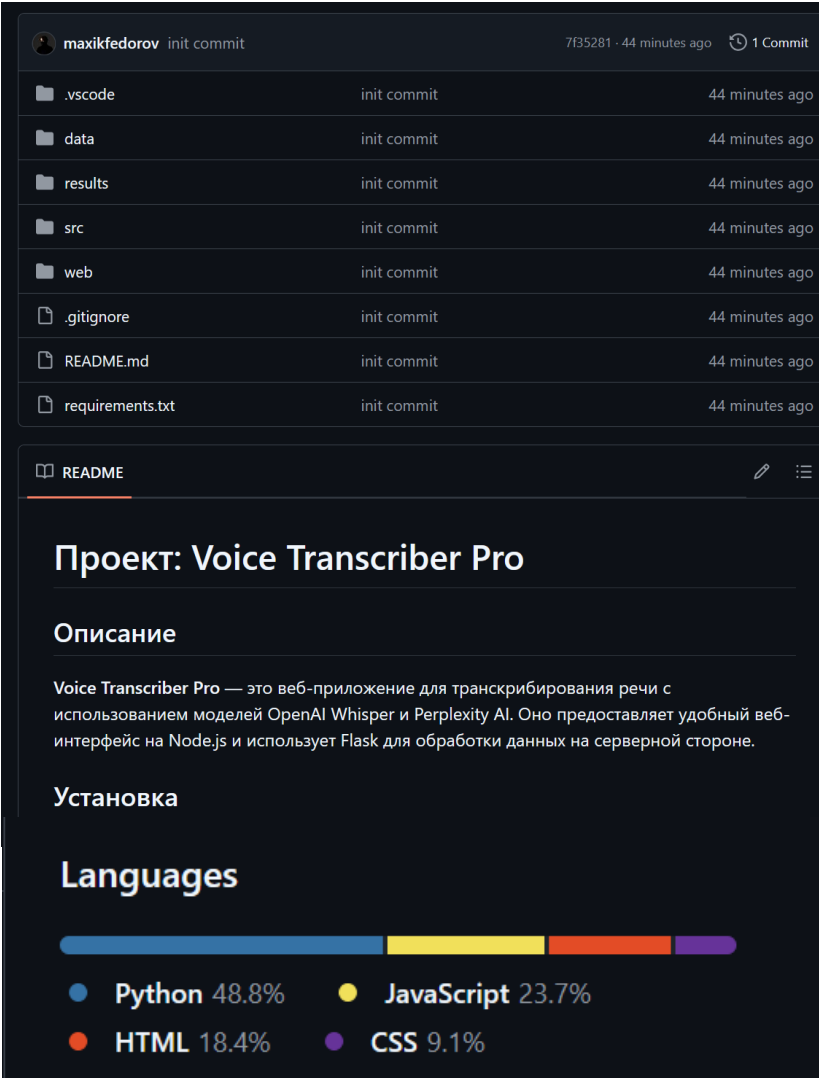
Формализованные данные

Task	design of a new layout
Start Time	02:00
End Time	03:30
Duration	1 hour 30 minutes

ИТОГИ РАЗРАБОТКИ

- Создано веб-приложение для автоматического распознавания речи с использованием модели Whisper от OpenAI, реализованное на Flask и Node.js.
- Реализован интерфейс для загрузки аудиофайлов и отображения результатов транскрипции и анализа.
- Интегрирован анализатор для структурирования данных транскрипции через Perplexity API.
- Организовано хранение данных в формате JSON, включая результаты транскрипции и формализованные данные.
- Оформлено решение в виде GitHub-репозитория с подробной документацией и инструкциями по использованию.

<https://github.com/maxikfedorov/HACK-TRANSNEFT-ASR>



The screenshot shows a GitHub repository page for user 'maxikfedorov'. The top section displays the commit history for the 'init commit', which was made 44 minutes ago. The files listed in the commit include .vscode, data, results, src, web, .gitignore, README.md, and requirements.txt. Below the commit history, the README file is open, showing the project title 'Проект: Voice Transcriber Pro'. The README content describes the project as a web application for speech transcription using OpenAI Whisper and Perplexity AI models, built with Node.js and Flask. It also includes a section for 'Установка' (Installation) and a 'Languages' section showing the code language distribution: Python (48.8%), JavaScript (23.7%), HTML (18.4%), and CSS (9.1%).

File	Commit	Time
.vscode	init commit	44 minutes ago
data	init commit	44 minutes ago
results	init commit	44 minutes ago
src	init commit	44 minutes ago
web	init commit	44 minutes ago
.gitignore	init commit	44 minutes ago
README.md	init commit	44 minutes ago
requirements.txt	init commit	44 minutes ago

Проект: Voice Transcriber Pro

Описание

Voice Transcriber Pro — это веб-приложение для транскрибирования речи с использованием моделей OpenAI Whisper и Perplexity AI. Оно предоставляет удобный веб-интерфейс на Node.js и использует Flask для обработки данных на серверной стороне.

Установка

Languages

Language	Percentage
Python	48.8%
JavaScript	23.7%
HTML	18.4%
CSS	9.1%

**СПАСИБО ЗА
ВНИМАНИЕ!**