ML-FIESTA SYNERGY'24 PROJECT DOCUMENTATION

BY CODE WIZARDS

PROJECT NAME: SandalQuest

Team Members

Shani Sinojiya (Team Lead / AI/ML & Backend Developer)

Mohammad Anas Africawala (AI/ML Engineer)

Tisha Patel (Full Stack Developer)

Project Overview

The goal is to utilize AI and machine learning to develop a pipeline that processes and understands audio content related to sandalwood cultivation. We are focusing on creating an Automatic Speech Recognition (ASR) model for the Kannada language and a speech-based question-answering system that will help users easily access information from the audio data.

Project Objectives:

- Develop an ASR model that can accurately recognize colloquial Kannada speech.
- Create a searchable audio database using the ASR output.
- Implement a question-answering system allowing users to ask questions via speech input.

Problem Statement

Karnataka is a key region for sandalwood, which holds significant cultural, religious, and economic value in India. However, much of the traditional knowledge around sandalwood cultivation is conveyed informally and captured in audio recordings. These resources are not easily accessible, and there's a need to digitize and preserve this indigenous knowledge. The main challenge is handling colloquial Kannada speech with background noise, as it differs from standard formal language.

Challenges:

- Limited digital information on sandalwood cultivation.
- Audio recordings often contain noise and informal language.
- Standard ASR models struggle with colloquial language.

Scope

The project includes:

- Building a Kannada ASR model for colloquial language recognition.
- Creating a searchable database by transcribing audio files.
- Developing a speech-based question-answering system to query the audio corpus.
- Fine-tuning the ASR model using both provided and publicly available Kannada datasets.

Out of Scope:

- Processing audio in languages other than Kannada.
- Real-time transcription of live streams.
- Handling complex multi-turn dialogues.

Dataset Description

Source: Audio files scraped from YouTube.

Content Type: Informal Kannada speech, possibly with background noise.

Format: Common audio formats like MP3.

Dataset Challenges

Noisy recordings made in public spaces.

Informal and colloquial language use.

Variations in pronunciation and dialects.

Functional Requirements

Task 1: Speech Recognition

- Develop an ASR model for Kannada speech from audio files.
- Handle informal and colloquial speech.
- Support model fine-tuning with additional datasets.

Key Features:

- Kannada speech transcription to text.
- Noise reduction and speech enhancement.
- Accommodate dialect variations.

**Code Explanation is in GitHub Repository.*

Functional Requirements

Task 2: Speech-based Question-Answering System

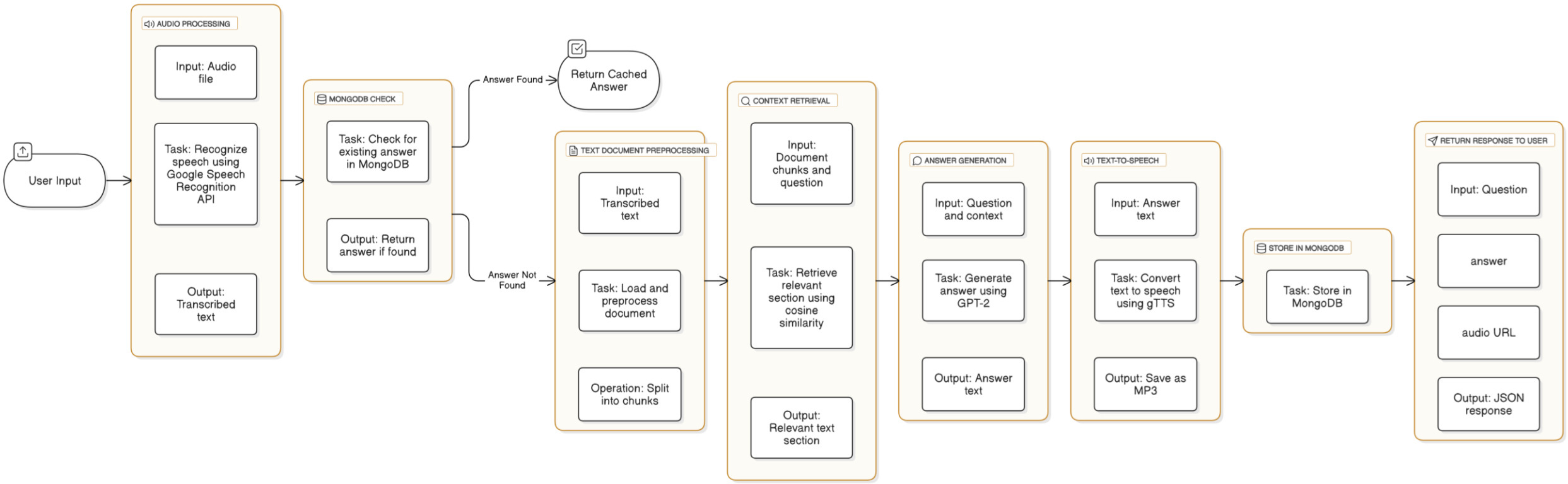
- Allow users to ask questions via speech input.
- Convert the spoken question to text using the ASR model.
- Search the transcribed audio data for relevant answers.
- Return the most relevant audio segment as the answer.

Key Features:

- Accurate answer retrieval from speech queries.
- Efficient search and indexing of the transcribed corpus.
- User-friendly query and response interface.

**Code Explanation is in GitHub Repository.*

SandalQuest - Pipeline Architecture



Technical Requirements

Languages: Python.

Libraries & Frameworks: PyTorch, Whisper (ASR), Hugging Face Transformers.

Database: MongoDB

Deployment: Google Cloud Platform (GCP) or AWS for scalable processing.

Hardware: GPU servers for training, cloud instances for deployment.

Tools: Google Colab, Jupyter Notebooks, GitHub.

Risks & Mitigation

Risks

- Low-quality audio data
- Poor ASR accuracy on colloquial speech
- High query processing latency

Mitigation

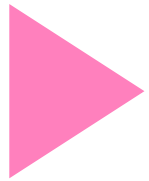
- Use noise reduction and data augmentation
- Fine-tune model with additional colloquial data
- Optimize search algorithms and indexing

Conclusion

This project aims to document and provide access to indigenous knowledge of sandalwood cultivation through advanced ASR and NLP technologies. By building a robust pipeline, we will not only aid conservation but also provide valuable insights for users interested in sandalwood cultivation.

Project Links

Github Repository:
Click me



Youtube Video:
Click me

References

Whisper: A Speech Recognition Framework by OpenAI.

Hugging Face Transformers for NLP models.

Google Colab for collaborative coding.

MongoDB for database management.

PyTorch for deep learning models.

GitHub for version control and collaboration.