



AudioMitra

Author: Team_11
[Vilal, Shriom, Madan, Hanuma]

1. Introduction:

AudioMitra is a document and content transformation platform to provide authentic and trustworthy transformations using the computational power of artificial intelligence and cognitive power of human in the loop.

2. Problem Statement:

“Next-Gen Multilingual AudioContent Converter Application”

Description – AudioMitra:

AudioMitra stands as an innovative solution within the realm of language processing, aiming to streamline human endeavors through technological advancement. Functioning as a document and content transformation system, **AudioMitra** facilitates the transition from digital images into text through Optical Character Recognition (OCR), and then converting text into audio using a text-to-speech (TTS) system.

Throughout the transformation process, **AudioMitra** allows for the optional collaboration between human intervention and machine operation. This collaborative approach ensures the authenticity and reliability of the transformations performed. The degree of human and machine involvement is tailored to meet specific workflow requirements, with the balance dictated by the desired level of accuracy and the nature of the transformations needed.

This includes:

- **Optical Character Recognition (OCR):** Extracting text from digital images or scanned documents to create editable text files.
- **Text-to-Speech (TTS) Conversion:** Transforming text content into audio formats, enabling the creation of audiobooks in various languages.

Stakeholders:

The primary users of **AudioMitra** are individuals and organizations that require the conversion of documents and content into different formats for various purposes. Schools, colleges, and universities that need to digitize textbooks, lecture notes, and other educational materials for online access in different formats.

3. Abstract Overview of the Proposed Solution

The proposed solution offers a comprehensive approach to text extraction, content validation, and text-to-speech/audio conversion. Leveraging open-source APIs for OCR and TTS, it ensures flexibility and scalability. Text extraction from images is facilitated through an Open-Source API, with options for using other OCR APIs. The system validates OCR'd content for accuracy, allowing human intervention for editing and refinement if necessary.

Output includes validated text content and audio files generated from the extracted and validated text. Accessible via web browsers with user authentication, the multiuser system supports various roles like validators, authors, and voice-over contributors, ensuring accessibility, flexibility, and collaborative efficiency in document transformation processes.

4. Technology Stack Overview for AudioMitra

Backend: Flask (Python), Node.js, MySQL, MongoDB, OCR and TTS

Frontend: React Framework, Bootstrap and Tailwind

5. Project Domain:

The domain of the **AudioMitra** project can be best categorized under **Education**. This classification is based on the project's focus on transforming documents and content from Image to Text and Text Speech, specifically aiming to enhance the learning experiences and accessibility of educational materials.

6. Key Functionalities

Core Features

- **Image to Text Transformation:** Utilizes the Open-Source API for OCR, converting images into editable text.
- **Text Editing and Review:** Provides tools for text editing and review, ensuring the accuracy of the extracted text.
- **Text to Audio Transformation:** Converts text into audio format using the Open-Source API, facilitating the creation of audiobooks.
- **Browser-Based Access:** Users interact with the system through a web interface, providing a user-friendly and accessible platform for document conversion. Support for various roles, including validators, authors, and voice-over artists.

Architectural Tactics

- Microservices Architecture, API Gateway, Event-Driven Architecture, Security

Design Patterns

- In this project, the Factory Pattern creates diverse objects seamlessly, the Observer Pattern tracks state changes efficiently, and the Singleton Pattern ensures singular resource instances, collectively enhancing flexibility and reliability.

7. Expected Time to Build a Prototype

Project's complexity and the need to develop a working prototype of the **AudioMitra** system, a realistic timeline can be outlined as follows. This timeline assumes a team of 4 members, with each member potentially taking on multiple roles to ensure the project's success.

- Research and Planning (1 weeks)
- Design (2 weeks)
- Development (4 weeks)
- Testing (1 weeks)
- Refinement (1 weeks)

Total Estimated Time: 9 weeks

8. Conclusion

AudioMitra utilizes OCR and TTS to convert content into various formats, enhancing accessibility and efficiency. Its multilingual support and collaborative approach ensure transformation authenticity and reliability. With a robust tech stack, it's set to impact education, libraries, and content creation, advancing global knowledge dissemination.