Versatile Text: An Integrated Solution For Text-To Speech, PDF And Image Text Extraction With Smart OCR And Text-To-Speech System For PDF And Image Text Recognition

Santhosh D
Dept of Computer Applications
Dr. M.G.R Educational and Research Institute
Chennai, India

Abstract- *The increasing use of digital documents has created a demand for efficient text extraction and accessibility solutions. Traditional methods of extracting text from images and PDF files are time-consuming and often inaccessible to visually impaired users. This project proposes a web-based system that integrates Optical Character Recognition (OCR) and Text-to-Speech (TTS) technologies to extract text from PDF documents and images and convert it into speech. The system provides features such as text customization, copy/download functionality, and theme management. Developed using HTML, CSS, and JavaScript, the solution improves accessibility, reduces manual effort, and enhances productivity for students, researchers, professionals, and visually impaired users.*

I. INTRODUCTION

The rapid growth of digital information has increased the use of PDF documents and image-based content. Extracting information from such sources manually is inefficient and error-prone. OCR technology enables automatic text recognition from images and scanned documents, while TTS technology converts extracted text into speech. Integrating these technologies into a single platform improves accessibility and productivity. The proposed system provides a user-friendly web application for text extraction, speech conversion, and document management.

II. EXISTING SYSTEM

ch functionalities as separate applications. Users must extract text using one application and convert it into speech using another. This process increases complexity and reduces efficiency. Most systems also lack customization features and require software installation.

I.imitations

- OCR and TTS available separately
- Manual transfer of extracted text
- Limited accessibility features
- Increased processing time
- Dependency on multiple applications
- Lack of integrated workflow

III. PROPOSED SYSTEM

The proposed system integrates OCR and TTS functionalities into a single web-based platform. Users can upload PDF or image files, extract text automatically, customize text appearance, and convert the extracted content into speech. Additional features include text copying, downloading, font customization, and dark/light themes.

Advantages

- Automated text extraction
- Integrated speech generation
- Enhanced accessibility
- User-friendly interface
- Cross-platform compatibility
- Improved productivity and efficiency

IV. METHODOLOGY

The proposed methodology consists of the following stages:

1. User uploads PDF or image files.
2. File validation and preprocessing are performed.
3. OCR engine extracts text from uploaded content.
4. Extracted text is displayed and managed.
5. Users can customize text appearance.
6. TTS engine converts text into speech.
7. Output can be copied or downloaded.

Technologies Used

- HTML
- CSS
- JavaScript
- OCR Technology
- Text-to-Speech API

V. RESULTS AND DISCUSSION

The system was tested using various PDF documents and image files. OCR successfully extracted text from academic PDFs, scanned documents, and images. Text-to-Speech conversion generated accurate audio output. Testing confirmed successful implementation of PDF upload, image upload, OCR extraction, text customization, speech generation, copy functionality, and theme management. The integrated approach reduced manual effort and improved accessibility compared to existing systems.

VI. CONCLUSION

The project successfully developed an integrated OCR and Text-to-Speech platform for extracting and converting text from PDF documents and images. The system improves accessibility, enhances productivity, and provides a userfriendly environment for document processing. The objectives of accurate text extraction, speech generation, and customization were achieved successfully. Future enhancements include multilingual OCR support, AI-based text correction, cloud storage integration, mobile applications, and advanced speech synthesis features.

REFERENCES

[1] World Wide Web Consortium. HTML5 Specification.
[2] Mozilla Foundation. JavaScript Documentation.
[3] Google. Speech Synthesis Documentation.
[4] Optical Character Recognition Research Papers and Documentation.
[5] Text-to-Speech Research Papers and Documentation.
[6] Visual Studio Code Documentation.
[7] Research articles on OCR and Speech Processing from IEEE, Springer, and Elsevier publications.