

Voice Controlled Platform to Access Audio e-Books for Visually Impaired People

Rudragoud Patil¹, Darshankumar Madihalli², Sucheta Kulkarni³ and Girish Rayamane⁴

¹⁻⁴ KLS Gogte Institute of Technology / Department of Computer Science, Belagavi, India
Email: rspatil@git.edu, {darshanmadihalli, ksucheta49, girishrayamane}@gmail.com

Abstract—People who are blind or visually impaired rely mostly on braille books. However, translating regular books to braille takes a long time. Only 1% of books are translated into braille at the moment. Our platform allows them to read English books of their choice. The book is used as the input, which might be in the form of images or a pdf file. Optical Character Recognition (OCR) is used to extract text from images, while the Pypdf2 library from Python is used to extract text from PDF files. The text may then be modified to make any necessary modifications. The text is then passed into the Google-Text-to-Speech (GTTS) engine, which produces an audio file. Apart from this, they can use their voice to look for other books. They may also use their voice to control the audio by saying things like play, pause, replay, stop, forward, backward, mute, unmute, and so on. And even these audiobooks are subsidiary to people who are slothful to read.

Index Terms— Audio Books, Voice control, blind people, visually impaired.

I. INTRODUCTION

Books hold a wealth of information and wisdom. The desire to read the books will be universal. Visually impaired or blind people are unable to read books in the same manner that others can. For them to understand the subject, you must sit and narrate the entire book to them. If the book is too long, the narrator will feel tired. There are numerous products on the market, both in terms of software and hardware. For English language OCR (Optical Character Recognition), products like ABBYY Fine Reader [8], Readiris [9], and others are available on the market, while products like Natural Reader [10], AT & T Natural Voices [11], and others fulfil the purpose of TTS (Text to Speech) systems in the form of software. However, they are prohibitively expensive for the visually impaired people for whom the work is designed, with prices ranging from \$100 to \$500. Other free versions are also available on the internet; however, they lack sufficient precision in English OCR or clear perceptibility in English TTS.

There is another group of products on the market, such as the KNFB Reader [12], Optelec ClearReader+ [13], and others, that provide a complete English OCR & TTS solution in the form of hardware using a stand-alone camera and processing unit. However, these devices are extremely costly, ranging from \$2500 to \$3500. Such technologies are out of reach for visually impaired persons, at least in India. Only Braille books can be read by blind persons. However, only a few of the books are available in Braille. Furthermore, Braille books are both long and heavy to handle. Some systems, like screen readers and e-book readers, allow people to read a book but do not allow them to control the audio with their voice. Keeping all of the above-mentioned aspects in mind we

designed a platform or website, especially for visually impaired people. The major goal is to close the accessibility gap between the ordinary user and those who are visually impaired. Using the Google-Text-to-Speech (GTTS) engine, the platform facilitates the conversion of books to audiobooks. For regular people, the platform is straightforward to use. However, visually impaired people may find it difficult to navigate the platform. As a consequence, we developed a virtual assistant that can do tasks such as searching for books, playing audio, stopping audio, and so on.

II. THEORY OF SUPPORT

Many researchers have developed remedies for the visually handicapped and blind in recent decades. Printed book to audiobook converter for visually impaired, DAISY (Digital Accessible Information System), Virtual assistant for the visually impaired, Digital Audio Book are just a few of them. In the United States, Congress established the Talking Book Program in 1931, which was handled by the Library of Congress's National Library Service for the Blind and Physically Handicapped (NLS). As technology advances, the trend is to move away from cassette talking books towards digital talking books. Digital talking books offer improved sound quality, the ability to read a whole book without having to manipulate the equipment, and the ability to skip sections and input bookmarks. Leas, Persoon, Soiffer, and Zacherle (2008) created a Digital Accessible Information System (DAISY) that enables visually impaired people to interact with books. A reader will need a hardware or software playback mechanism to read these books. [1] Depending on the needs of a blind user, a digital file can be quickly converted to Braille, audio, or large print. A blind person can use screen readers to read documents to them. The blind can now access the book's electronic material without difficulty. Unfortunately, nearly all books are never made available in accessible formats like Braille, large print, or analog or digital audio. Pescod (2009) Domale, Padalkar, Parekh, and Joshi (2013) created a printed book to audiobook converter for the blind people that provides audiobooks in English and Marathi [2], but no control over the audio over voice. Iyer, Shah, Sheth, and Devadkar(2020) created a Virtual Assistant for the Visually Impaired that helps visually impaired people browse the internet and automates any webpage using text to speech and speech to text modules[3].

III. PROPOSED SOLUTION

The project is classified into two sections: the front end and the back end. We used HTML (HyperText Markup Language), CSS (Cascading Style Sheets), Bootstrap (a framework), Javascript, and Ajax to build the front end of the platform. We utilized Flask (a Python framework) for the back end and SQL for the database (to store users details and books).

IV. SYSTEM OVERVIEW AND DESIGN

The flowchart illustrates how the platform works. There are two categories of users: those who are visually impaired and normal. The user will be on the Index page when the platform first loads. Normal users can use services such as image to speech, PDF to speech, and book viewing by logging in or signing up. Users who are blind or visually impaired can listen to the audiobooks without having to register to our platform.

A. Normal user

The following are the steps for a normal user:

Step 1: Users can log in/signup by entering their credentials. Fig. 1. Working of the Platform

Step 2: If the user's credentials are right, they will be able to use services such as image to speech, PDF to speech, and view books.

Step 3: They can upload the images or pdf of the book.

Step 4: They have the option to crop and delete the images that have been uploaded.

Step 5: Using OCR or Pypdf2 library, the text from the uploaded book is extracted.

Step 6: They also have the option to make adjustments to the extracted text.

Step 7: Using GTTS, the captured text will be transformed to audio. Then the audio along with the uploaded images/pdf will be saved in the database.

B. Visually impaired user

The following are the steps for a Visually impaired user:

Step 1: To activate the voice assistant, they need to press the button.

Step 2: After the beep, say the title of the book.

Step 3: The audio will begin to play if the book is present in the database.

Step 4: They may use their voice to control the audio by saying things like pause, increase/decrease volume, stop, forward, backward, and so on.

Step 5: To end the process and close the voice assistant, say "exit."

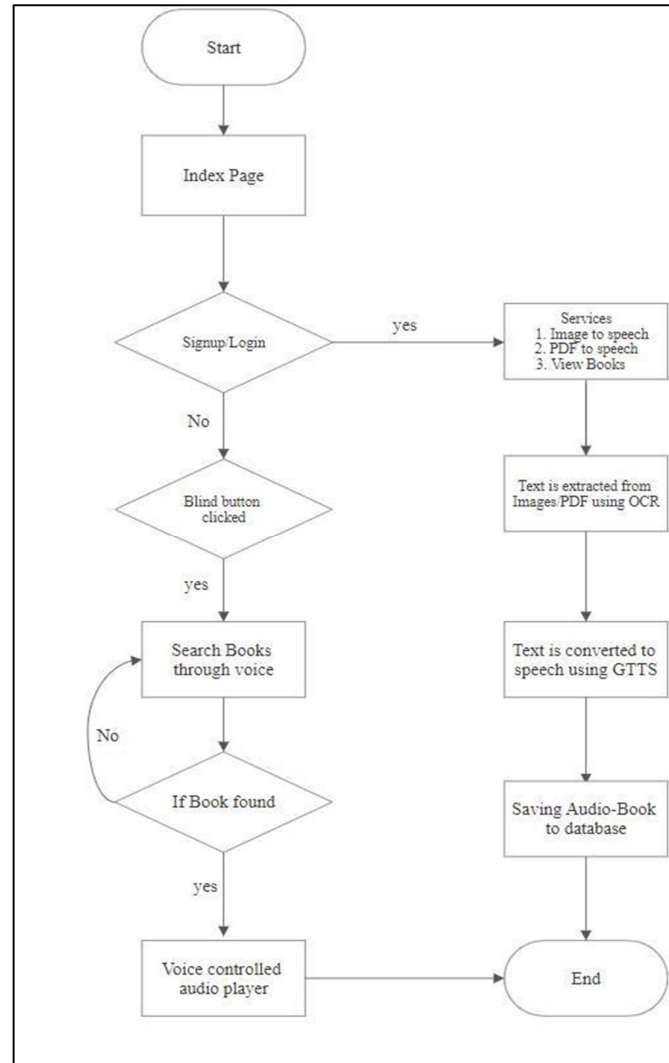


Fig 1. Working of the Platform

V. LIBRARIES USED IN DEVELOPING THE PLATFORM

The platform was developed using Flask, a Python framework. Optical Character Recognition (OCR), Google Text-to-Speech (GTTS), and Speech Recognition are some of the important libraries we used.

A. Optical Character Recognition (OCR)

OCR is used to extract text from images. The user can input images into the OCR. OCR verifies if the picture is properly aligned during pre-processing, and if it isn't, it aligns it before converting it to a binary picture (separating text from the background) [4]. The picture or script's language is then identified. A tesseract is used to detect text from an image in the text-recognition stage. Tesseract employs its lexicon to affect the character

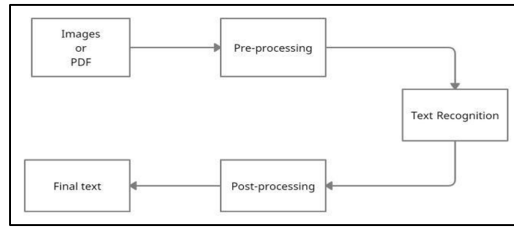


Fig. 2. Working of OCR

segmentation phase in the post-processing stage, which improves accuracy [5]. Finally, the text is printed onto the site so that any necessary changes may be made[20,21].

B. Google Text to Speech (GTTS)

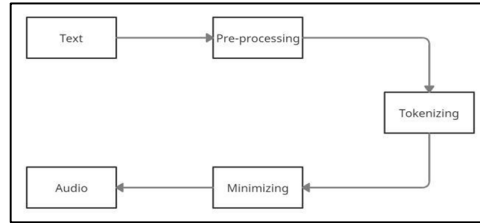


Fig. 3. Working of GTTS

The Google-Text-to-Speech is used to transform words into vocal audio form. The Python programming language is used to implement GTTS. It makes use of the Text-to-Speech API from Google Translate. The key reason for adopting GTTS API is that it has outstanding voice quality and is simple to integrate with Python. The text is sent into the GTTS algorithm. The Google Translate host's top-level domain (tld) is configured to "google.cn." It will be 'google.com' by default. In pre-processing, it verifies that the letters are pronounced correctly and the space between them is right. Tokenizer separates the pre-processed text into tokens (strings). Tokenizer contributes to the naturalization of sound. A maximum of 100 characters is allowed in the GTTS API. If any of the tokens exceed 100 characters, they will be split into two during the minimization step. The text will then be converted to an audio file and stored in the database [6].

C. Speech Recognition

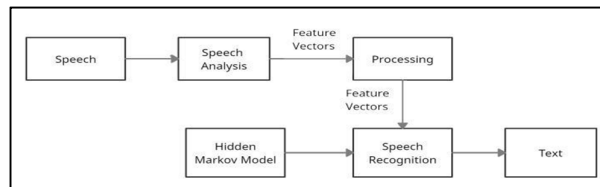


Fig. 4. Working of Speech Recognition

Speech Recognition is used to transform the user's speech into text. There are a variety of speech recognition engines available. As a result, we have chosen Google Speech Recognition as our platform. The speech will be used as a source of information. Vectors are derived from a time series of short-duration voice signals in the speech analysis section. The removal of noise from the speech is crucial. So, the noise is reduced with the use of vectors in robust processing. Using test utterances, the speech recognition section identifies the best match. As a reference model, the Hidden Markov Model is employed. The system does not have a training function because it will be pre-trained [7]. The user is given a beep at all phases of speech recognition for effective speech recognition, after which he can talk freely.

Other Libraries used in building the platform are as follows:

- 1) Pypdf2: It is a Python PDF package capable of splitting, merging, cropping, and modifying the pages of PDF files. It can extract text from PDF files.
- 2) Werkzeug: It is a library for web applications. We used this library to improve the security of our platform. We are converting and creating hashed passwords with the aid of this library.

- 3) PIL: It is a Python imaging package that allows you to access, manipulate, and save several image formats (jpg, png, etc.)
- 4) PyAudio: It is a cross-platform audio I/O library that is used to capture user's voices, which are then processed by the Speech Recognition module.
- 5) Playsound: The audio files are played using this library. We must provide an argument for this, like the location of the audio file.
- 6) SQLAlchemy: It is a package that allows Python programs to interact with databases.
- 7) Shutil: This is useful for copying or deleting files or directories from the file system. This library is used to delete the books that have been uploaded.

VI. RESULTS AND DISCUSSION

Number of words in the image: 127 Incorrect words from the extracted text: 1 Accuracy:99% An OCR for the English language is 99% accurate. For different images, the accuracy may vary. It is dependent on the image quality and clarity. The voice quality of Google-Text-to-Speech is outstanding. Python's speech recognition system recognized the voice with 95% accuracy
Result of an OCR:

The first and most widespread transmission mode is called *unicast*. In the unicast transmission mode, information is sent by one sender to one receiver. Most of today's Internet applications rely on the unicast transmission mode. The example below shows a network with two types of devices: hosts (drawn as computers) and intermediate nodes (drawn as cubes). Hosts exchange information via the intermediate nodes. In the example below, when host 5 uses unicast to send information, it sends it via three intermediate nodes. Each of these nodes receives the information from its upstream node or host, then processes and forwards it to its downstream node or host. This is called *store and forward* and we will see later that this concept is key in computer networks.

The first and most widespread transmission mode is called *unicast*. In the unicast transmission mode, information is sent by one sender to one receiver. Most of today's Internet applications rely on the unicast transmission mode. The example below shows a network with two types of devices: hosts (drawn as computers) and intermediate nodes (drawn as cubes). Hosts exchange information via the intermediate nodes. In the example below, when host 5 uses unicast to send information, it sends it via three intermediate nodes. Each of these nodes receives the information from its upstream node or host, then processes and forwards it to its downstream node or host. This is called *store and forward* and we will see later that this concept is key in computer networks.

Fig. 5. Input image to OCR

Fig. 6. Text extracted from the image using OCR

VII. CONCLUSION

The platform has been developed specifically for persons who are blind or visually impaired. Even the platform can be helpful for people who are unable to read. It provides the experience of reading books/novels. If the images are of excellent quality, the OCR produces 99% accurate results. In a short amount of time, the book will be converted to audio. The audio will have an Indian accent with clear word pronunciations. The voice assistant does not rely on the blind user's keyboard inputs, it provides a simple approach for visually impaired people to use the platform. Blind people can use a voice assistant to get a list of books present in a database and perform tasks such as searching for a book, playing a book, pausing a book, and so on.

FUTURE ENHANCEMENT

Currently, the platform only transforms books into audiobooks in English. Also, the voice assistant only accepts commands in English. We intend to make this platform available in all of India's native languages and to accept commands given in any language. At the moment, our platform/website may be accessed via Chrome or Mozilla Firefox. However, in the future, we can create an application that will allow it to be accessed via mobile phones.

REFERENCES

- [1] Dennis Leas, Emilia Persoon, Neil Soiffer, Michael Zacherle, "A Standard for Accessible Multimedia Books," IEEE MultiMedia, Vol 15, 4, 2008, 1070-986X/08,doi:10.1109/MMUL.2008.84
- [2] A. Domale, B. Padalkar, R. Parekh and M. A. Joshi, "Printed Book to Audio Book Converter for Visually Impaired," 2013 Texas Instruments India Educators' Conference, 2013, pp. 114-120, doi:10.1109/TIIEC.2013.27.
- [3] V. Iyer, K. Shah, S. Sheth and K. Devadkar, "Virtual assistant for the visually impaired," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 1057-1062, doi: 10.1109/ICCES48766.2020.9137874.
- [4] Sezgin, Mehmet; Sankur, Bulent (2004). "Survey over image thresholding techniques and quantitative performance evaluation" (PDF). Journal of Electronic Imaging. 13 (1): 146. Bibcode:2004JEL....13..146S. doi:10.1117/1.1631315. Archived from the original (PDF) on October 16, 2015. Retrieved May 2, 2015.
- [5] Ray Smith (2007). "An Overview of the Tesseract OCR Engine" (PDF). Archived from the original (PDF) on September 28, 2010. Retrieved May 23, 2013.
- [6] <https://gtts.readthedocs.io/en/latest/tokenizer.html>
- [7] Yoshizawa, Shingo Wada, N. Hayasaka, N. Miyanaga, Yoshikazu. (2006). Scalable architecture for word HMM-based speech recognition and VLSI implementation in complete system. Circuits and Systems I: Regular Papers, IEEE Transactions on. 53. 70 - 77. 10.1109/TCSI.2005.854408.
- [8] <http://finereader.abbyy.com/>

- [9] <http://www.irislink.com/c2-2115-189/Readiris-14-OCR-Software-Scan-Convert-Manage-your-Documents-.aspx>
- [10] <http://www.naturalreaders.com/>
- [11] <http://www.naturalvoices.att.com/>
- [12] <http://www.knfbreader.com/>
- [13] <http://www.optelec.com/en-US/optelec-clear-reader>
- [14] <https://moam.info/>
- [15] <http://www.statelibrary.sc.gov/>
- [16] National Federation of the Blind, "The Braille Literacy Crisis in America", Jernigan Institute, March 26, 2009
- [17] UNESCO, "World Braille Usage", National Library Service for the Blind and Physically Handicapped, Library of Congress, Washington, D.C., USA, 1990, ISBN 0-8444-0676-7
- [18] S. M. Felix, S. Kumar and A. Veeramuthu, "A Smart Personal AI Assistant for Visually Impaired People," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018, pp. 1245-1250, doi: 10.1109/ICOEI.2018.8553750.
- [19] R. Chen, Z. Tian, H. Liu, F. Zhao, S. Zhang and H. Liu, "Construction of a voice driven life assistant system for visually impaired people," 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 2018, pp. 87-92, doi: 10.1109/ICAIBD.2018.8396172.
- [20] Suleman. J. Nadaf et.al (2016), "Cloud Based Privacy Preserving Health Data Storage and Retrieval System", International Conference on Inventive Computation Technologies.
- [21] Rudragoud Patil and R. H. Goudar "Implementation of User-Centric SSE for Privacy-Preserving and Enhancing Searching Efficiency in Cloud Environment", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019.