A Mini Project Report on

AI Audio Book

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Engineering

In

Computer Engineering

By

Soumojyoti Dutta (19102014)

Vedang Gore (19102065)

Aarya Totey (19102070)

Sumati Hans (18102028)

Under the Guidance of

Prof. Sofiya Mujawar

Department of Computer Engineering

A.P. Shah Institute of Technology

G.B.Road, Kasarvadavli, Thane(W),

Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year 2021-2022

Approval Sheet

This Mini Project Report entitled “AI AudioBook” Submitted By “Soumojyoti Dutta” (19102014), “Vedang Gore” (19102065), “Aarya Totey” (19102070) “Sumati Hans” (18102028) is approved by partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Science** from **University of Mumbai.**

Prof. Sachin H. Malave

Head of Department

(Computer Engineering)

Place: A. P. Shah Institute of Technology, Thane

Date: 29.10.2021

CERTIFICATE

This is to certify that the mini project entitled "**AI Audio Books**" submitted by "Soumojyoti Dutta" (19102014), "Vedang Gore" (19102065), "Aarya Totey" (19102070), "Sumati Hans" (18102028) for the partial fulfillment of the requirement for award of a degree **Bachelor of Engineering** in **Computer Science**, to the University of Mumbai, is a bonafide work carried out during academic year 2020-2021.

Prof. Sachin H. Malave
Head of Department
(Computer Engineering)

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place: A. P. Shah Institute of Technology, Thane

Date: 29.10.2021

Declaration

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will cause disciplinary action by the Institute can also evoke penal action from the sources which have thus not been properly cited or from whom properly cited or from whom proper permission has not been taken when needed.

Signature

Soumojyoti Dutta (19102014)

Vedang Gore (19102065)

Aarya Totey (19102070)

Sumati Hans (18102028)

Abstract

We came across the fact that reading books in the traditional way was getting very difficult with the advent of technological advancements. We thought listening to audiobooks would be a very good alternative. However, currently the apps that provide such services have a limited catalogue of books and we can not guarantee that the book that we want would be present in their collection. So we attempt to make an app that would convert the pdfs or the pictures of the book pages and convert them into audiobooks so that we can listen as much as we want and whenever we want. Our application will first take a pdf or an image file from the user using the front-end . Then by applying OCR we will convert the file to text and then we will use Machine Learning to read out the contents from the file to the user.

Contents

Chapter 1

Introduction

Books are the source of knowledge. Ancient Sumerians wrote their data in clay tablets. Egyptians used papyrus for the same reason. In ancient India knowledge of Vedas and other scriptures were passed from the gurus to disciples through the process of listening. Gurus recited the mantras and the disciples had to remember it by listening. We may call it the proto version of audiobooks. Later people wrote these down and gave them the shape of a book made of “Bhurja patra”(tree skin) or palm leaf. After several centuries paper was invented by the Chinese and spread throughout the world by the Arabians. Then people started to store their knowledge in those papers. We use these paper books till nowadays. But this paper has a serious problem regarding its production. Many trees are cut down to make paper which is not eco-friendly at all. Secondly, heavy books are hard to carry around. Nowadays, students have to carry many books in their bags, which causes problems related to backbones. So, with the advancement of technology, especially in the field of GUI, we have invented an e-book system, portable document files(PDF) to read books using the phone, tablet, or the computer. It helps us to carry books wherever we can. As all books are stored in the device, and the device being light-weighted no problem of backbone occurs and above all, it is eco-friendly as no trees are cut to make these e-books. The usage and demand for e-books has risen significantly. But continuous usage of e-books leads to eye-related problems. Here comes the audiobooks. You can not read books in any environment, especially in a crowd. But you can listen to an audiobook easily in any environment if you have a compatible electronic device and headphones. Besides this, humans have a longer duration of auditory memory than of visual memory. Now let's consider a good student or employee who loves to read books(story, philosophical, etc) but you don’t have the time to read them or your economic condition is not sufficient to buy your favourite books then you can easily listen to the audiobook versions and fulfill your wish. It is not that audiobooks are a very new concept. Audiobooks were used to teach blind students. There are also many youtube channels and apps available on the internet where we can find many audiobooks. However, currently, the apps that provide such services

have a limited catalogue of books and we can not guarantee that the book that we want would be present in their collection. So we thought, what if we make an app that would convert the pdfs or the pictures of the book pages and convert them into audiobooks so that we can listen as much as we want and whenever we want. Our application will first take a pdf or an image file from the user using the front-end. Then by applying OCR we will convert the file to text and then we will use Machine Learning to make a reader that will read out the contents from the file.

1.1 Problem Definition:

Designing an AI book reader application that will read whatever the input file is given be it a pdf or an image using OCR. The app will convert pdf or pages scanned into audiobooks.

1.2 Objectives:

- Any people can gain knowledge by listening to audiobooks in any situation
- People can listen their favourite books just by uploading them
- People can listen to books of their choice if they have an image or pdf of the book.
- With this knowledge can be imparted in a convenient way.
- Audio books offer blind students to study more independently wherever or whenever they want without other people's help.

1.3 Scope:

- Our application will first take a pdf or an image file from the user
- If the file is a pdf then the app will convert the pdf into image and then the image to text and then that text to audio respectively.
- Else the image file is directly converted to text and then the text is converted to audio.
- The audio can then be listened to properly.

1.4 Existing System /Project :

- Narrative Science is a data storytelling company, founded in 2010 in Chicago. The company's technology relies on AI to transform statistics into stories and numbers into knowledge. Customers include Deloitte, MasterCard, USAA, and members of the US intelligence community.
- Charisma.ai for instance offers "a new dimension of storytelling: re-playable, interactive conversations with crafted characters." Components include a "hyper-advanced" story editor and a chat engine. Characters have emotions, memories and voices.

Chapter 2

Technology Stack

- **Pytesseract:** Python-tesseract is an optical character recognition (OCR) tool for python. That is, it will recognize and "read" the text embedded in images. Python-tesseract is a wrapper for Google's Tesseract-OCR Engine. It is also useful as a stand-alone invocation script to tesseract, as it can read all image types supported by the Pillow and Leptonica imaging libraries, including jpeg, png, gif, bmp, tiff, and others. Additionally, if used as a script, Python-tesseract will print the recognized text instead of writing it to a file.

- **OpenCV:** OpenCV is the huge open-source library for computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even handwriting of a human. When integrated with various libraries, such as NumPy, python is capable of processing the OpenCV array structure for analysis. .

- **Pyttsx3:** pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline and is compatible with both Python 2 and 3. An application invokes the pyttsx3.init() factory function to get a reference to pyttsx3. Engine instance. It is a very easy to use tool which converts the entered text into speech.

- **Visual Studio Code:** Visual Studio Code is an Integrated Development Environment (IDE) made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. Users can change the theme, keyboard shortcuts, preferences, and install extensions that add additional functionality.

- **Jupyter Notebooks:** Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating notebook documents.A Jupyter Notebook document is a browser-based REPL containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media. Underneath the interface, a notebook is a JSON document, following a versioned schema, usually ending with the ".ipynb" extension.

Chapter 3

Benefits and Applications:

3.1 Benefits for society:

- Continuous usage of e-books leads to eye-related problems whereas audiobooks are a much better alternative.
- It is easy to use and can be accessed whenever required, especially in a crowd. you can listen to an audiobook easily in any environment
- No need to carry and store heavy books as a single device can hold many of them!

3.2 Benefits for environment:

- Firstly, the use of e-book helps us save the environment as there is a significant reduction in the usage of pages required.
- Trees are saved and that helps in reducing the global warming
- people will get knowledge and become more aware about the conservation of the environment.

3.3 Applications:

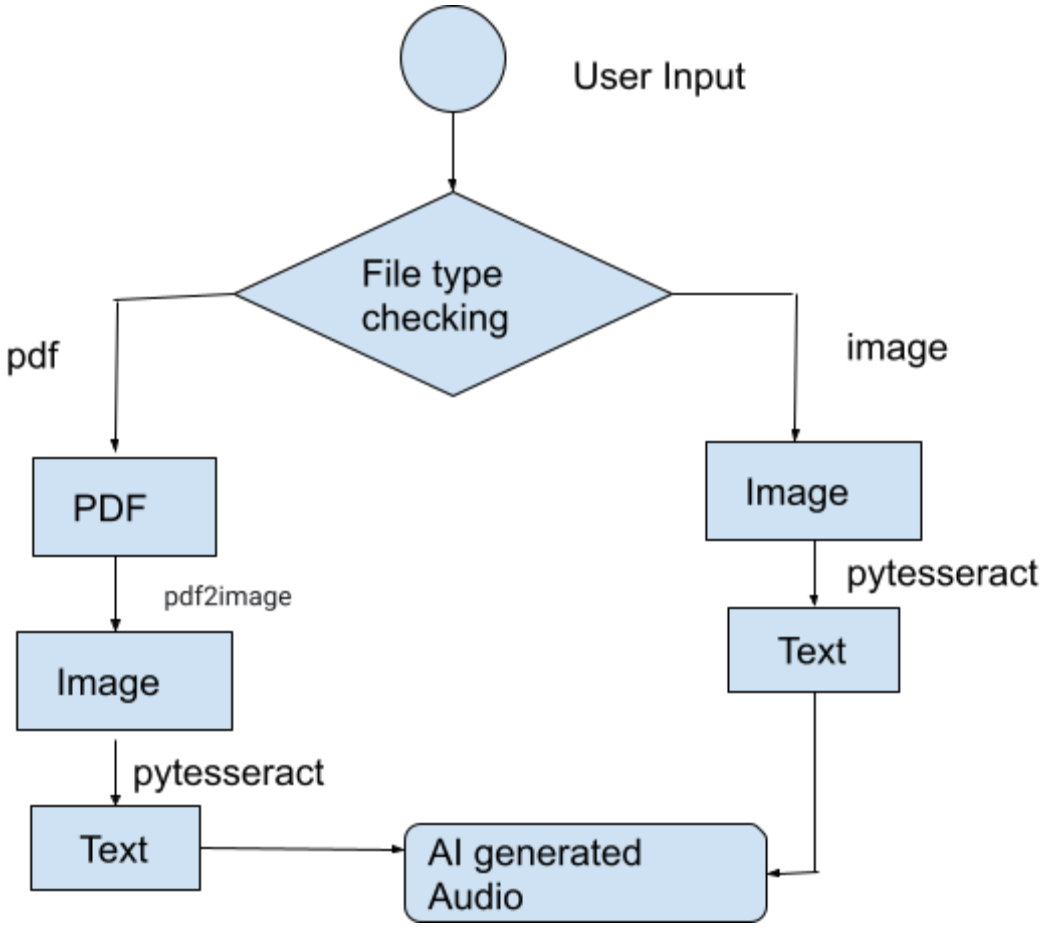
- Anyone can gain knowledge by listening to audiobooks in any situation by uploading images/pdf of their desired book.
- To help those who cannot read to listen to the books thereby can be used to provide education and it can help sustain literacy.
- People with eye related problems can use this system.

Project Design

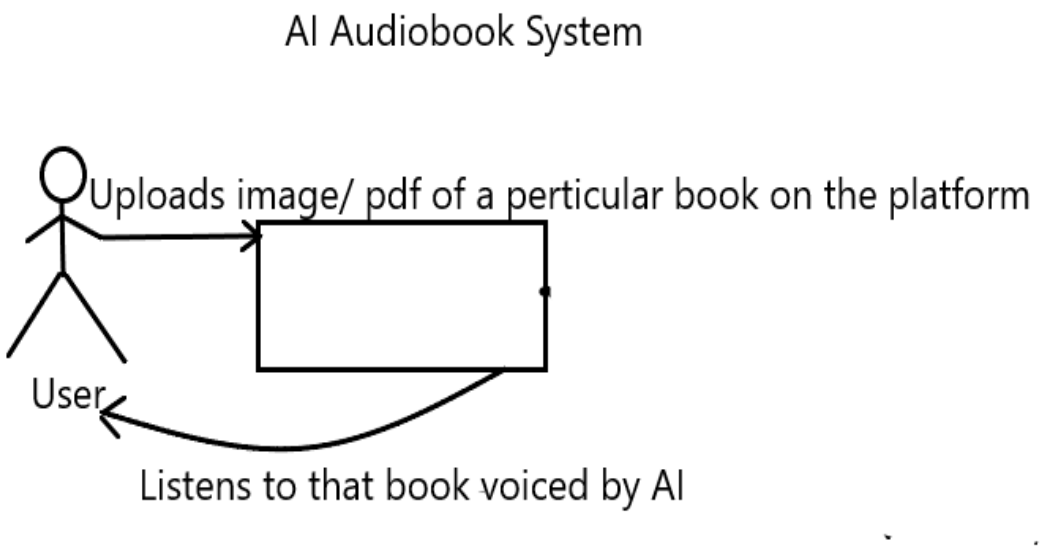
4.1 Proposed System

- Our Audiobook enables users to get a personalized experience. The issue with the traditional audiobooks is that an artist records a certain set of stories and the user has to choose from them. With Audiobook, readers have an advantage to upload and listen to a book of their choice.
- Any book in the form of image of PDF can be uploaded
- With OCR the text within the media is recognized and stored.While text processing is enabled using the OpenCV module.
- The next step is converting the text to audio file.
- Once the audio file is generated it is available for the users to access at any time
- Text to Audio is converted by using Pyttsx3 python module

4.2 Flow of Modules



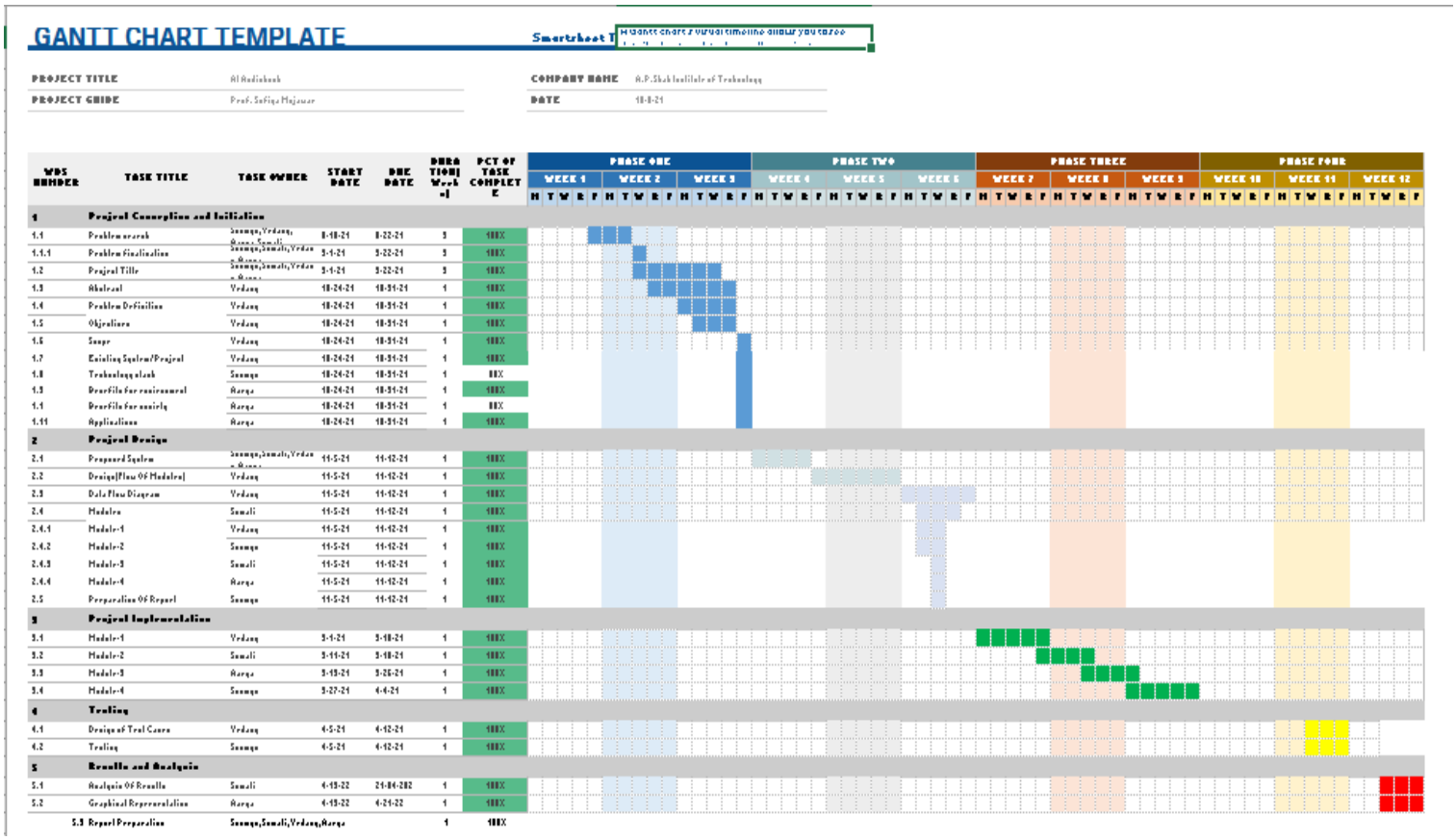
4.3 Data Flow Diagram



Chapter 5

Annexure A

5.1 Gantt Chart



6 Conclusion:

Thus, this report concludes how this project is initiated and how all the modules work accordingly after one another. This project takes into consideration the problem every reader faces and with the help of AI Audiobook we provide a better alternative to reading.

7 Bibliography:

Robert Keefer and Nikolaos Bourbakis ,"Interaction with a Mobile Reader for the Visually Impaired", Assistive Technologies Research Center, Wright State University, 21st IEEE International Conference on Tools With Artificial Intelligence.

Dongnnil Song, Arafeh Karemi, Paul Kim, "Toward designing mobile games for visually challenged children", Stanford University. 2011 International conference on e-education, entertainment, e-management.

S. Leuthold, J.A.Bargas-Avila, K. Opwis, "Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users", Int. J. Human-Computer Studies

J. Kjeldskov, J. Stage, "New Techniques for usability evaluation of mobile systems" on Int. J.Human-Computer Studies, vol 60, 2004

Acknowledgement

We have great pleasure in presenting the mini project report on **AI Audiobook**. We take this opportunity to express our sincere thanks towards our guide **Prof. Sofiya Mujawar** Department of Computer Engineering, APSIT Thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of the project.
We thank **Prof.Sachin Malave** Head of Department, Computer Engineering, APSIT for his encouragement during the progress meeting and providing guidelines to write this report.
We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement

Student Name1: Soumyojyoti Dutta
Student ID1: 19102014
Student Name2: Vedang Gore
Student ID2: 19102065
Student Name3: Aarya Totey
Student ID3: 19102070
Student Name4: Sumati Hans
Student ID4: 18102028