VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Mini Project Report on

AUDIOBOOK SYSTEM

Under the subject: Human Machine Interaction(HMI)

Submitted by

Jayesh Samtani D12A 57

Sagar Sidhwa D12A 62

Somesh Tiwari D12A 71

Under the guidance of

Subject Lab Teacher

Dr. Mrs. Gresha S Bhatia

Mrs. Ashvini Abhijit Gaikwad

(2020-2021)

Index

Title **Page No**

Chapter 1: Problem Definition and Scope of Project ……....2

1.1 Problem definition …… ..2

1.2 scope of project ……....2

1.3 Users of the System ……... 2

1.4 User Requirements (functional and nonfunctional(usability)) …….....3

1.5 Technologies used …….....3

Chapter 2: Requirement Gathering and Persona Creation ……...4

Chapter 3: High and Low Fidelity Prototypes and their explanation ……..5

Chapter 4: UI design and Implementation. ……..11
Screenshots of webpages of Website/App

Chapter 5: UI Evaluation ……..15

Chapter 6: Conclusion and Future Enhancement ……..16

References ……...17

Chapter - 1 : Problem Definition and Scope of Project

1.1 Problem definition

There are about 45 million blind people and 135 million visually impaired people worldwide. Disability of visual text reading has a huge impact on the quality of life for visually disabled people. We are going to propose an approach to create an portable text to speech converter.This system can help the visually impaired people or any person to learn from audio read-back of any scanned text, by converting the uploaded pdf to image,extracting the text from image,and converting the text to audio as mp3 file.

Therefore we need a low cost system that will be able to automatically locate and read the text aloud to visually impaired persons. The handwritten text reader is needed to help the visually impaired listen to an audio read-back of printed text.

1.2 scope of project

The e-book industry is starting to flourish due, in part, to the availability of affordable and user-friendly e-book readers. As users are increasingly moving from traditional paper books to e-books, there is an opportunity to reinvent and enhance their reading experience, for example, by leveraging the multimedia capabilities of these devices in order to turn the act of reading into a real multimedia experience. Here we focus on the augmentation of the written text with its associated audiobook, so that users can listen to the book they are (currently) reading. We propose an audiobook-to-ebook alignment system by applying a Text-to-Speech(TTS)-based text to audio alignment algorithm, and enhance it with a silence filtering algorithm to cope with the difference on reading style between the TTS output and the speakers in the ebook environment.An end-to-end speech-to-text translation on a corpus of audiobooks specifically augmented for this task.In this System, a single model is trained to decode source speech into target text in a single pass. Experimental results show that it is possible to train compact and efficient end-to-end speech translation models in this setup

1.3 Users of the System

- Person Between Age Group 21-65
    - Student
    - Teachers
    - Employees
    - Retired Person
- Visually Challenged can use with the help of others

1.4 User Requirements (functional and nonfunctional(usability))

- PDF to mp3 file converter
- Can listen from a particular page number

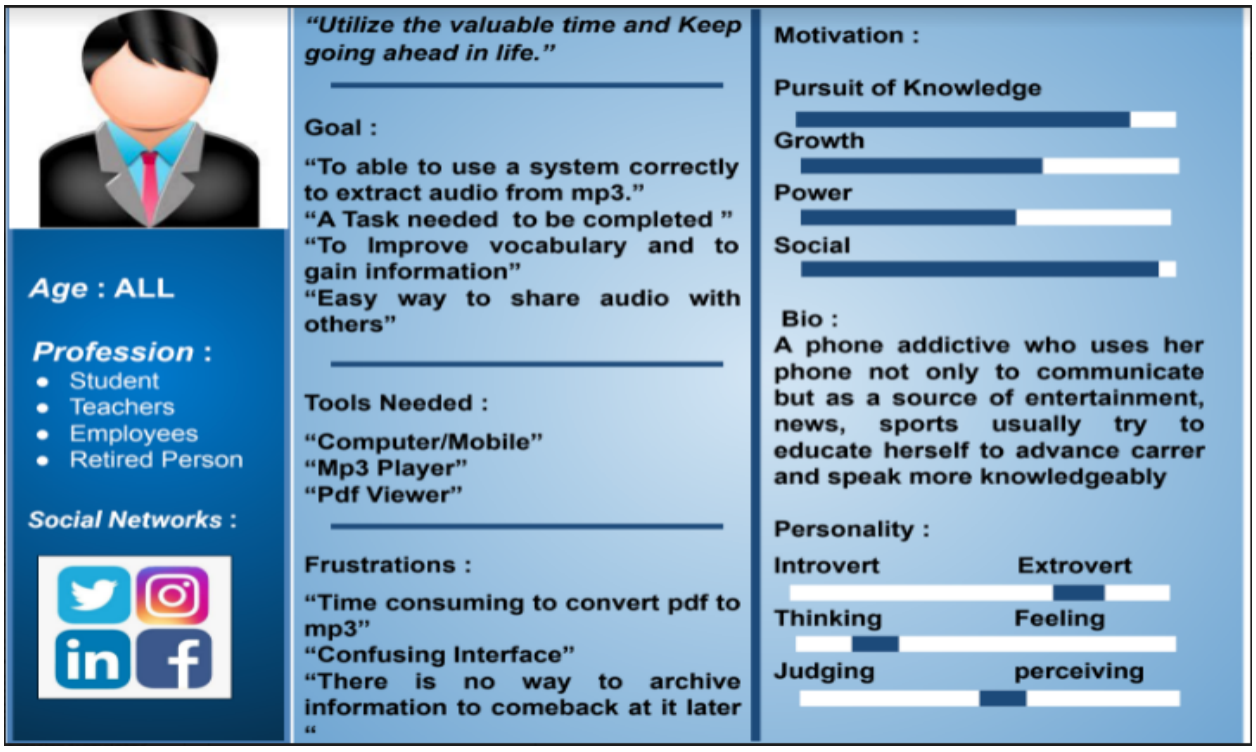
1.5 Technologies used

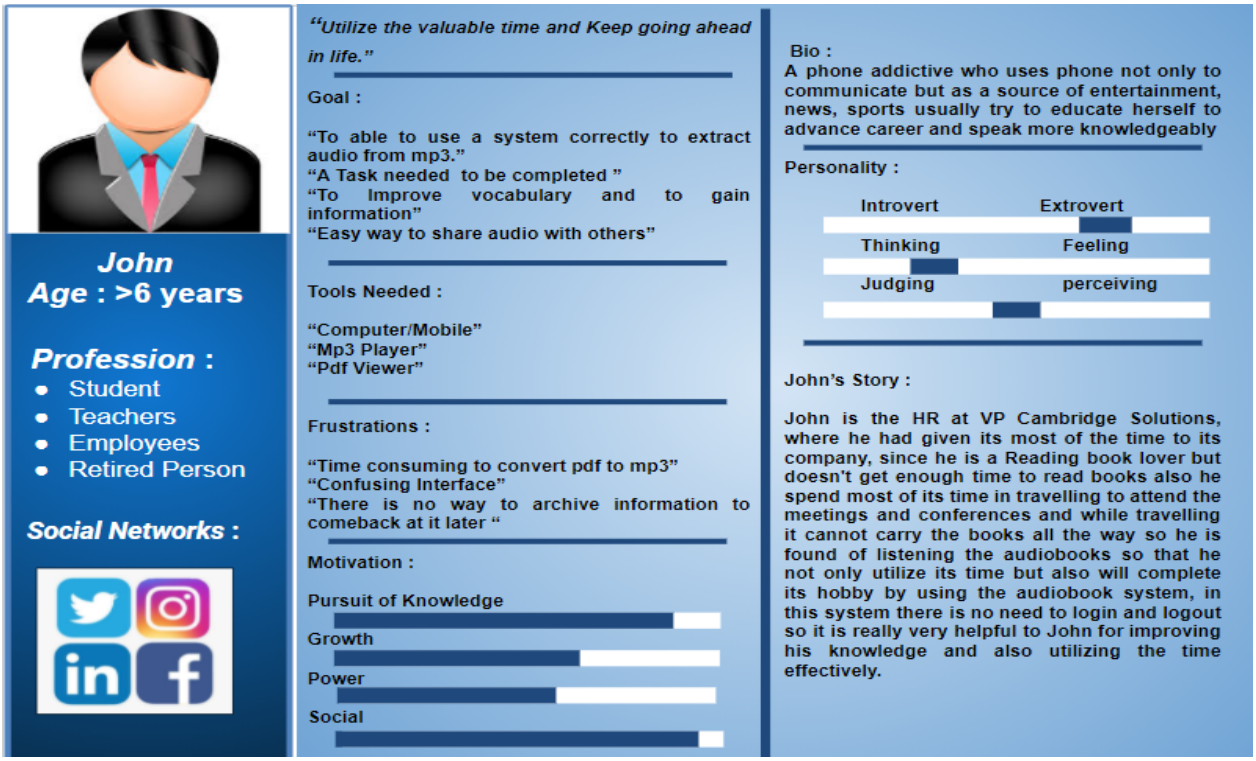
Our goal is to convert a given text image into a string of text, saving it to a file and to hear what is written in the image through audio.
For this, we need to import some Libraries

1. Pytesseract(Python-tesseract) : It is an optical character recognition (OCR) tool for python sponsored by google Open source OCR module Tesseract is used as a basis for the implementation of a text reading system.
2. pyttsx3 : It is an offline cross-platform Text-to-Speech library.When the OCR process is complete it produces a string of text which is displayed on the user interface screen,
3. Python Imaging Library (PIL) :Pillow is built on top of PIL (Python Image Library). PIL is one of the important modules for image processing in Python. It adds image processing capabilities to your Python interpreter
4. Django : It is a python web framework that encourages rapid development and clean, pragmatic design.

Optical Character Recognition involves the detection of text content on images and translation of the images to encoded text that the computer can easily understand. An image containing text is scanned and analyzed in order to identify the characters in it. Upon identification, the character is converted to machine-encoded text.
The Uploaded Pdf is first converted to image and then the image is first scanned and the text and graphics elements are converted into a bitmap, which is essentially a matrix of black and white dots. The image is then pre-processed where the brightness and contrast are adjusted to enhance the accuracy of the process.
The image is now split into zones identifying the areas of interest such as where the images or text are and this helps kickoff the extraction process. The areas containing text can now be broken down further into lines and words and characters and now the software is able to match the characters through comparison and various detection algorithms. The final result is the text in the image that we're given.

Chapter - 2 : Requirement Gathering and Persona Creation





Chapter - 3 : High and Low Fidelity Prototypes and their explanation

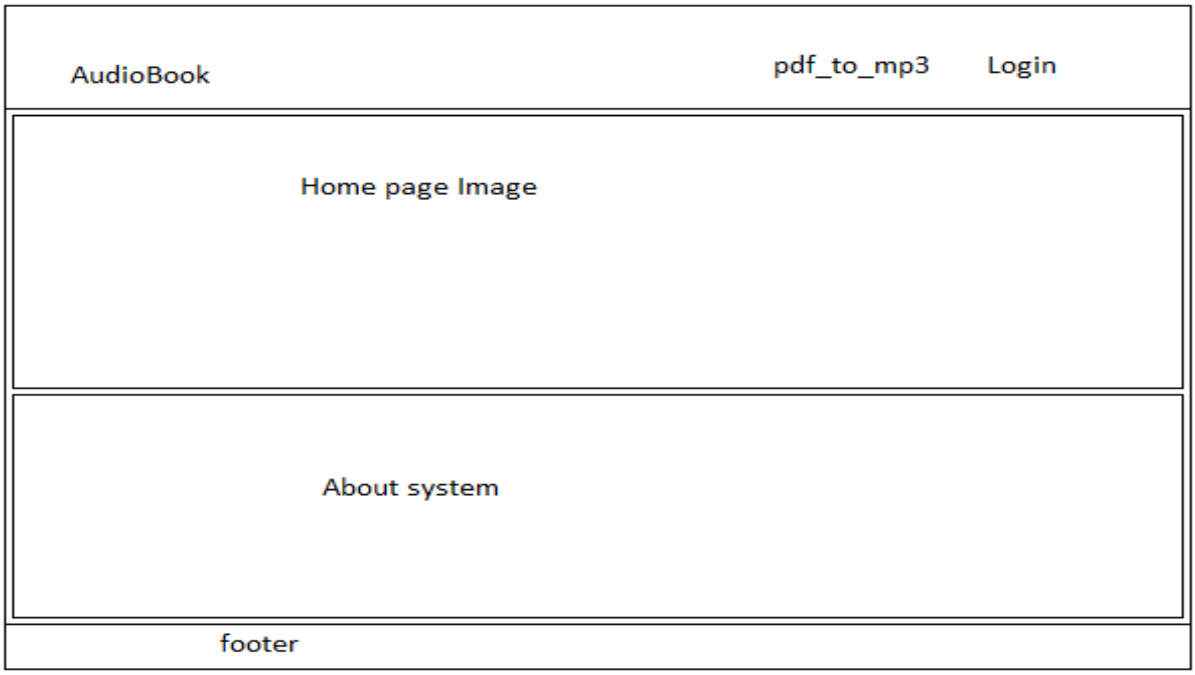
Prototyping

A prototype is a draft version of a product that allows you to explore your ideas and show the intention behind a feature or the overall design concept to users before investing time and money into development. A prototype can be anything from paper drawings (low-fidelity) to something that allows click-through of a few pieces of content to a fully functioning site (high-fidelity).

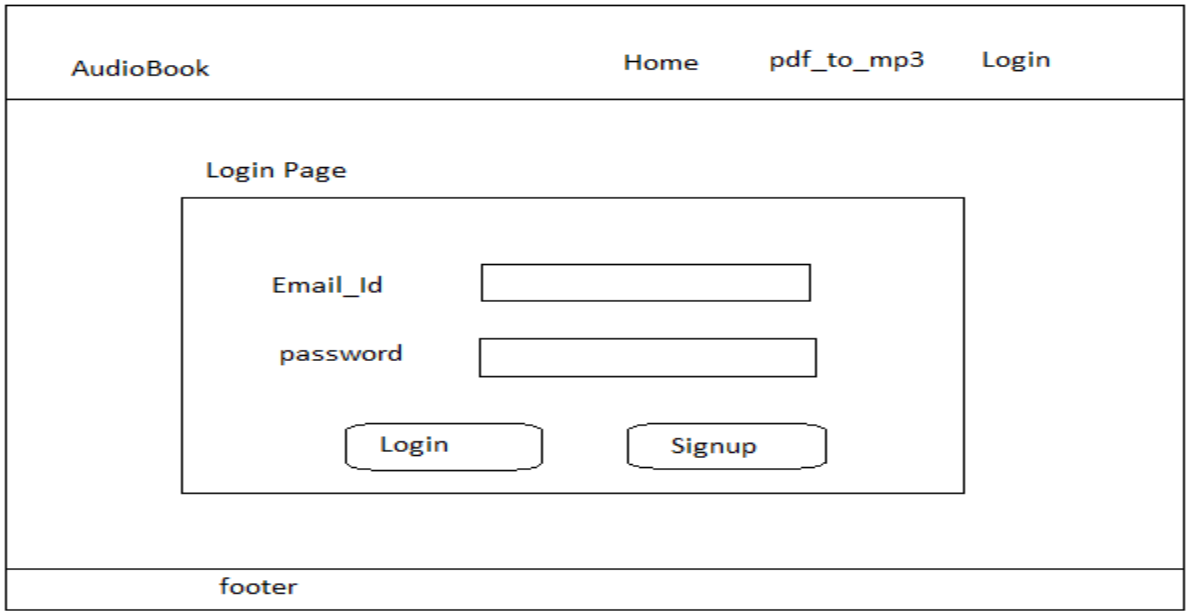
Low fidelity :

Low-fidelity prototyping — known as low-tech, is a simple and easy translation of the product and design concepts. It's used to turn the design ideas into testable and tangible artifacts, collecting and analyzing the user demands at the early stage.

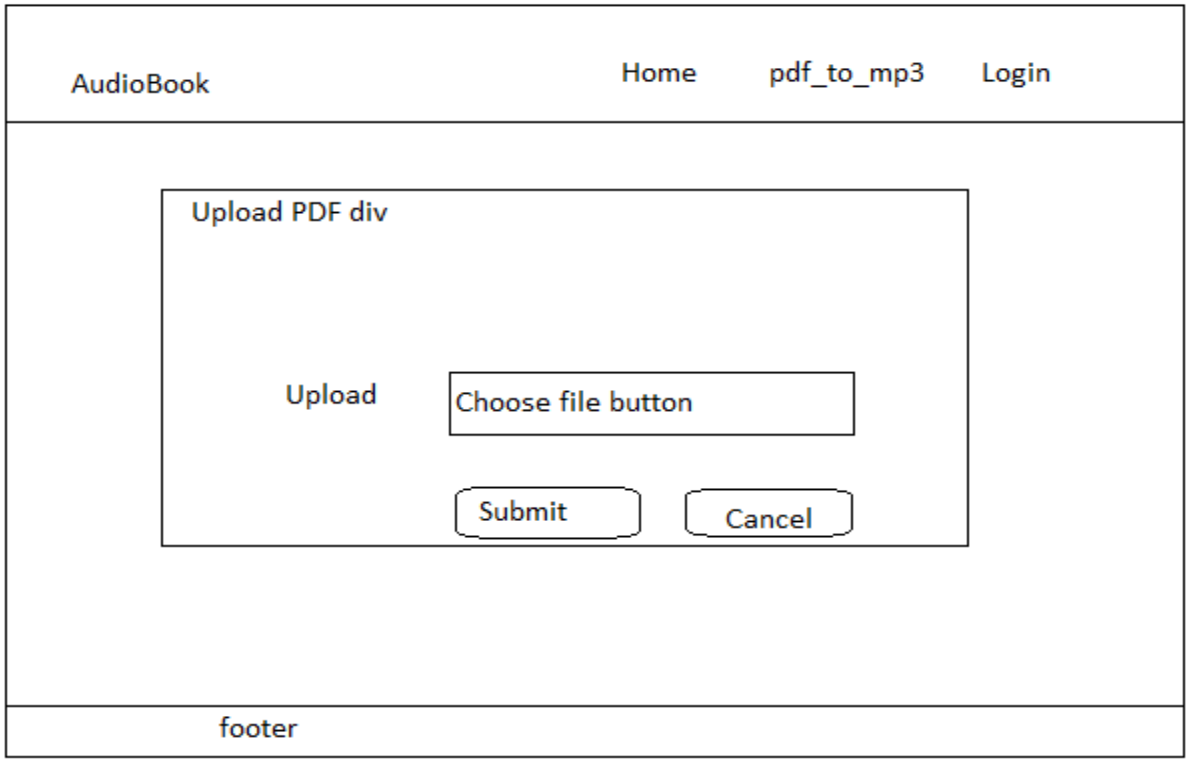
Homepage:



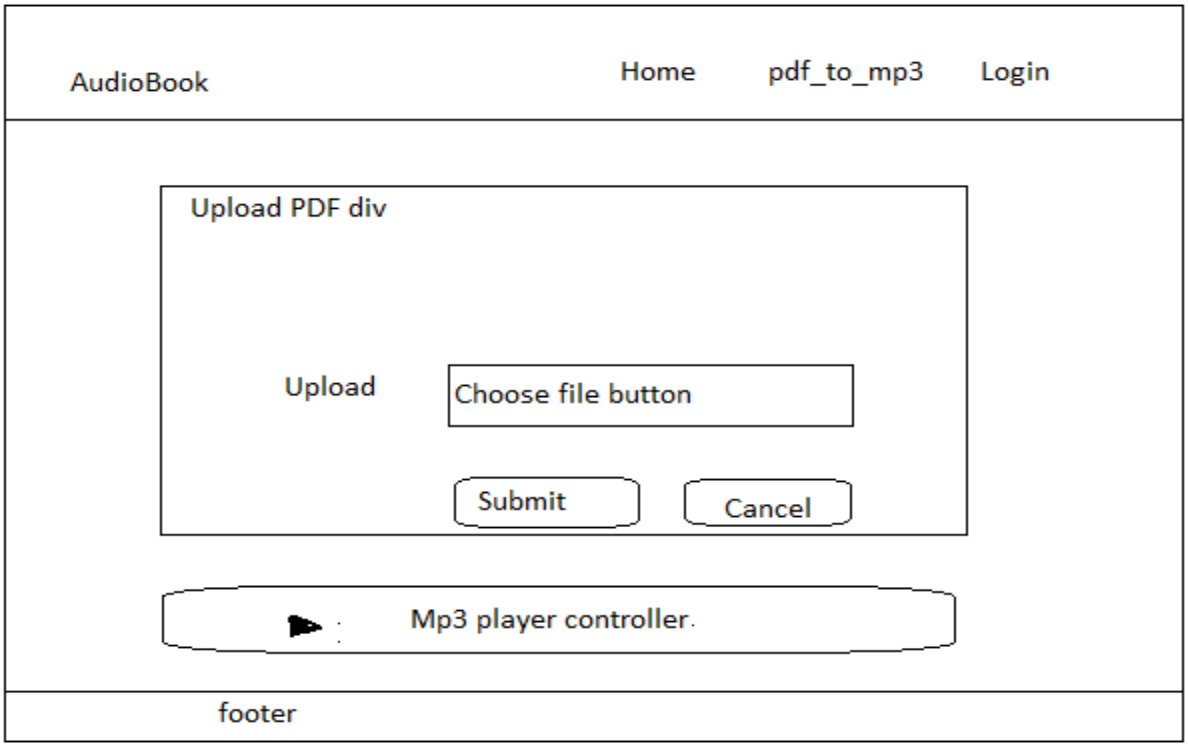
Login and signup page :



Pdf upload page:



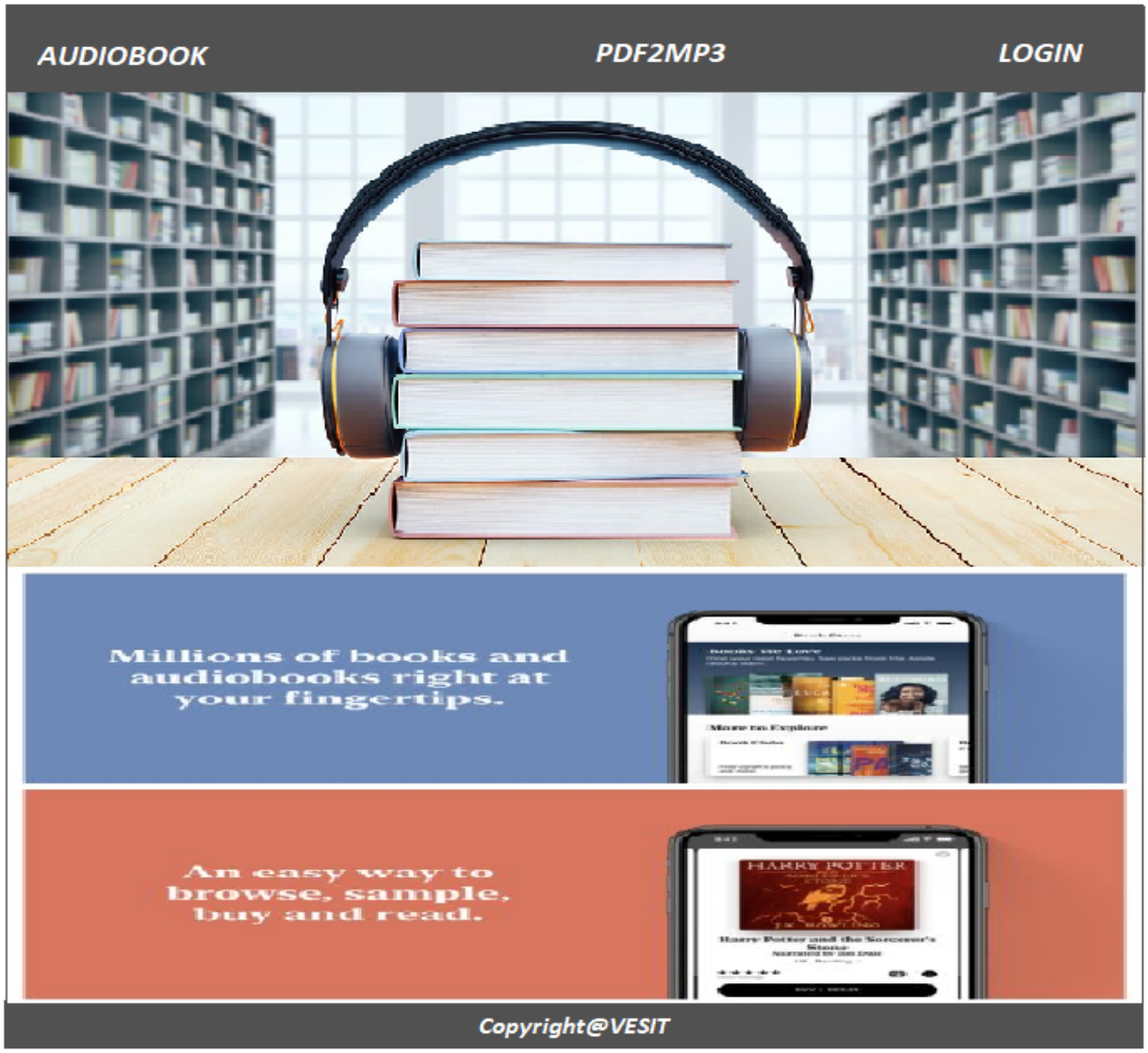
Showing mp3 controller :



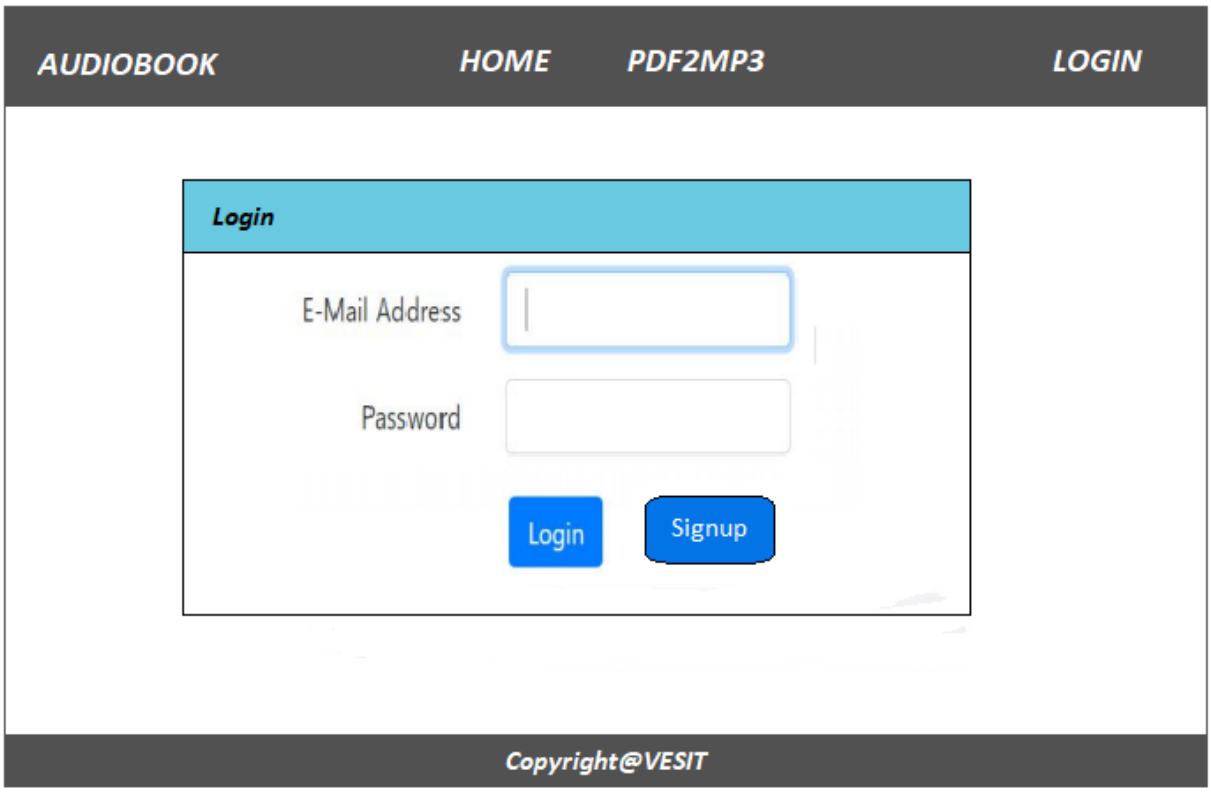
High Fidelity :

The fidelity of the prototype refers to the level of details and functionality built into a prototype. In this sense, a high-fidelity (sometimes referred as high-fi or hi-fi) prototype is a computer-based interactive representation of the product in its closest resemblance to the final design in terms of details and functionality. The “high” in high-fidelity refers to the level of comprehensiveness that allows you to examine usability questions in detail and make conclusions about the user behaviour.

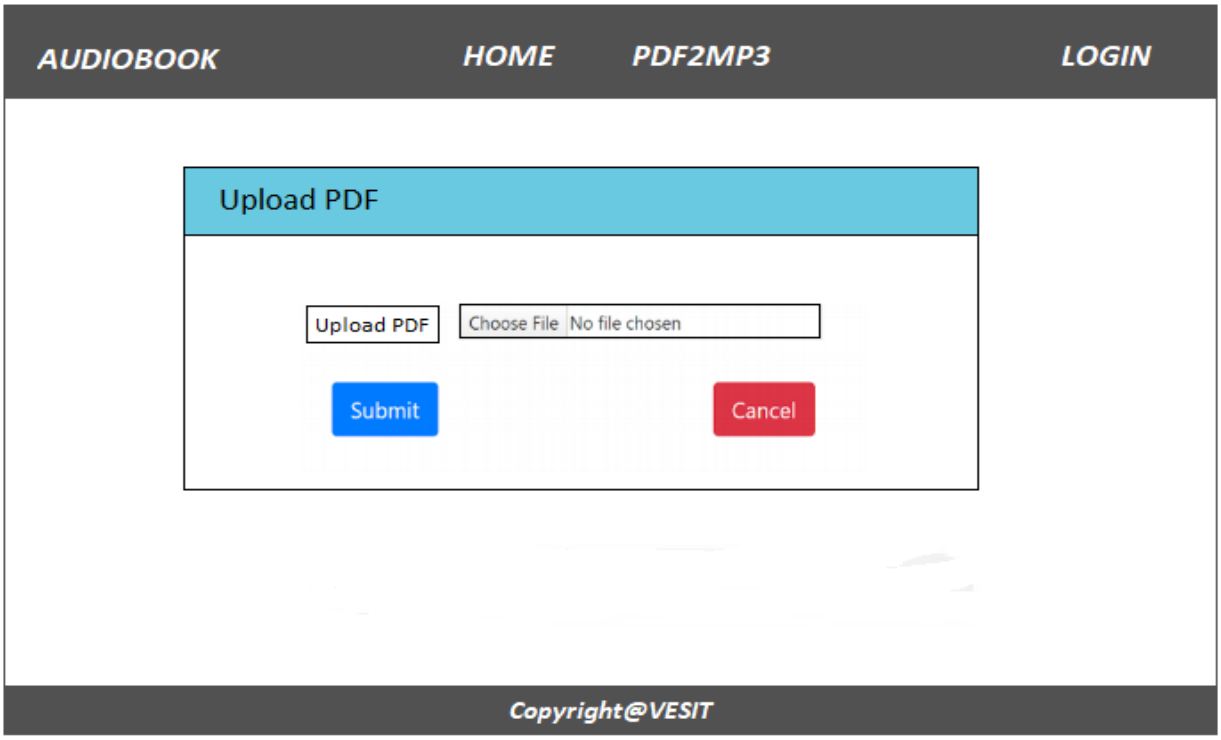
Homepage:


AUDIOBOOK
PDF2MP3
LOGIN
Millions of books and audiobooks right at your fingertips.
An easy way to browse, sample, buy and read.
HARRY POTTER
Copyright@VESIT

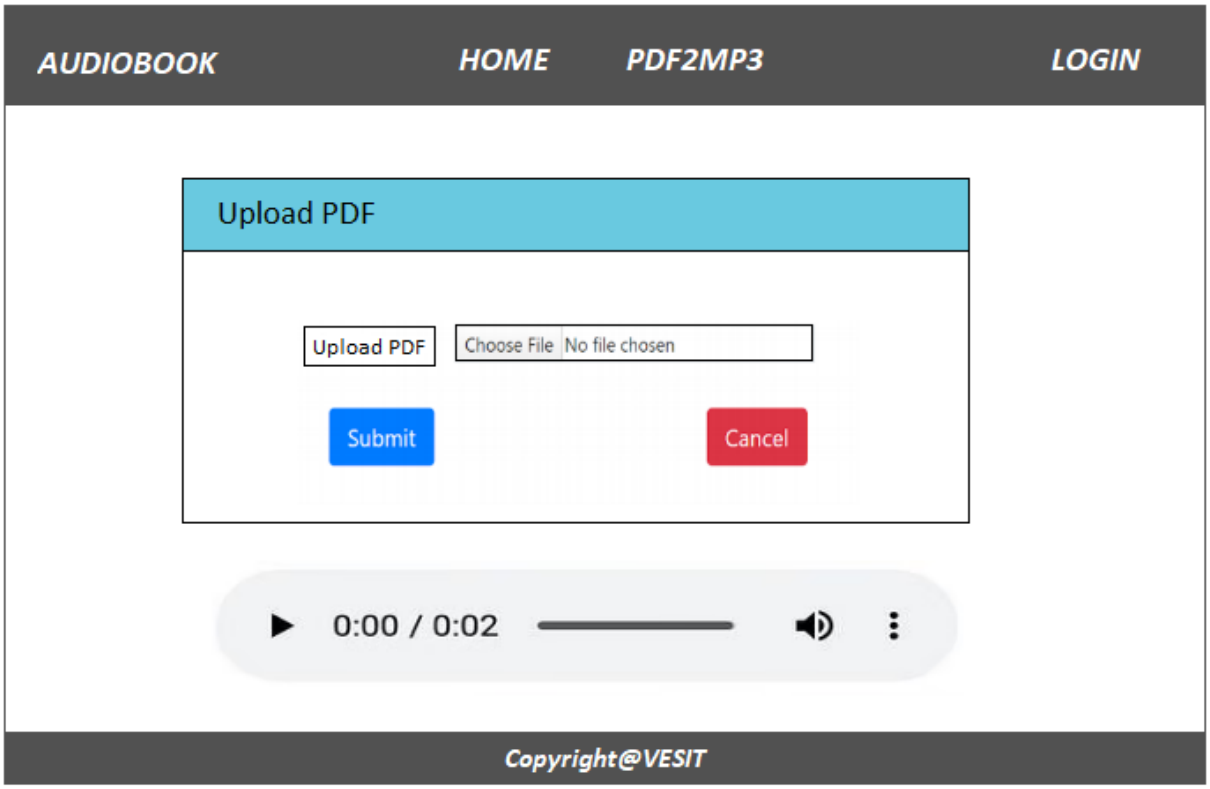
Login and signup page :



Pdf upload page:

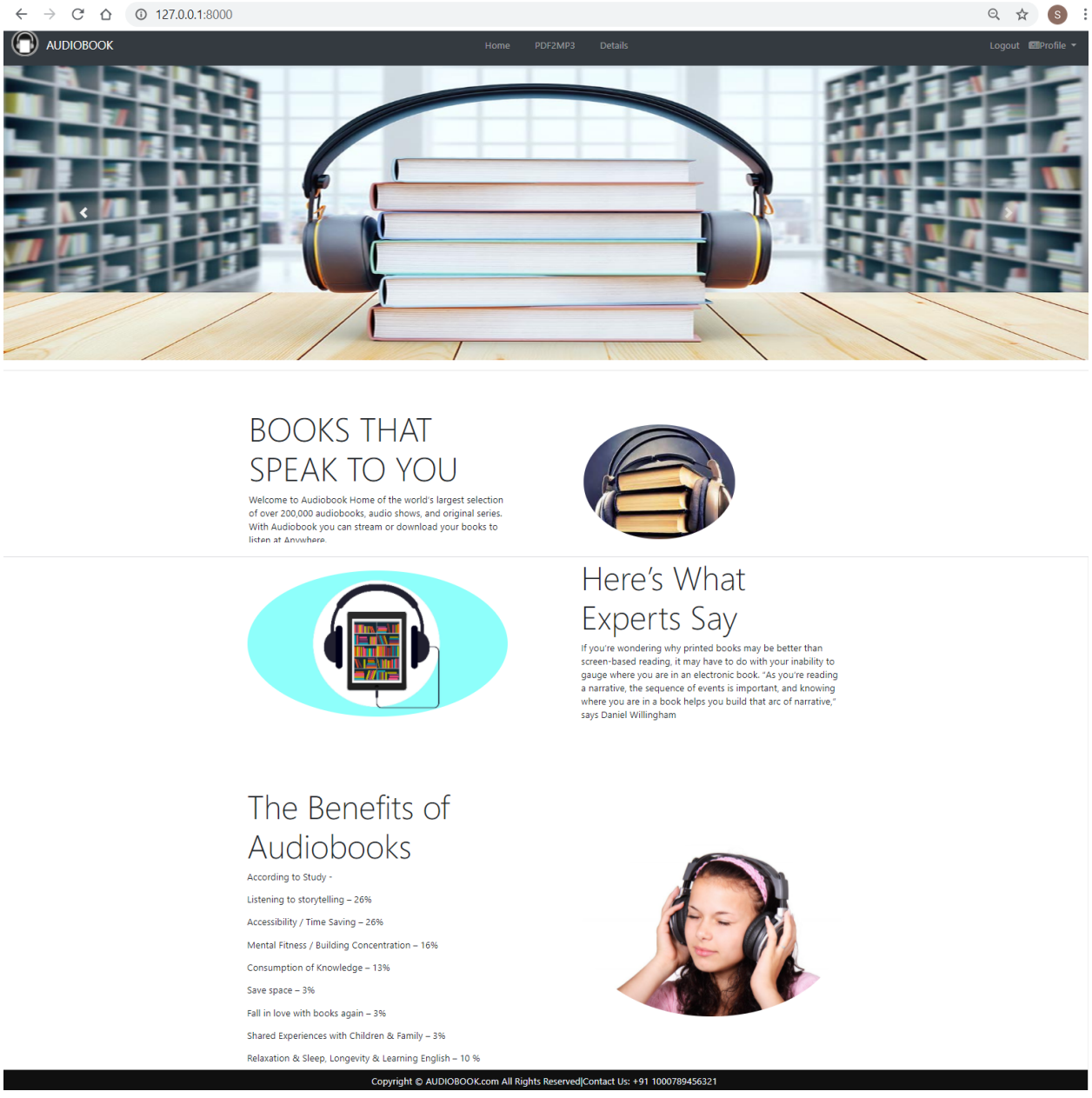


Showing mp3 controller :

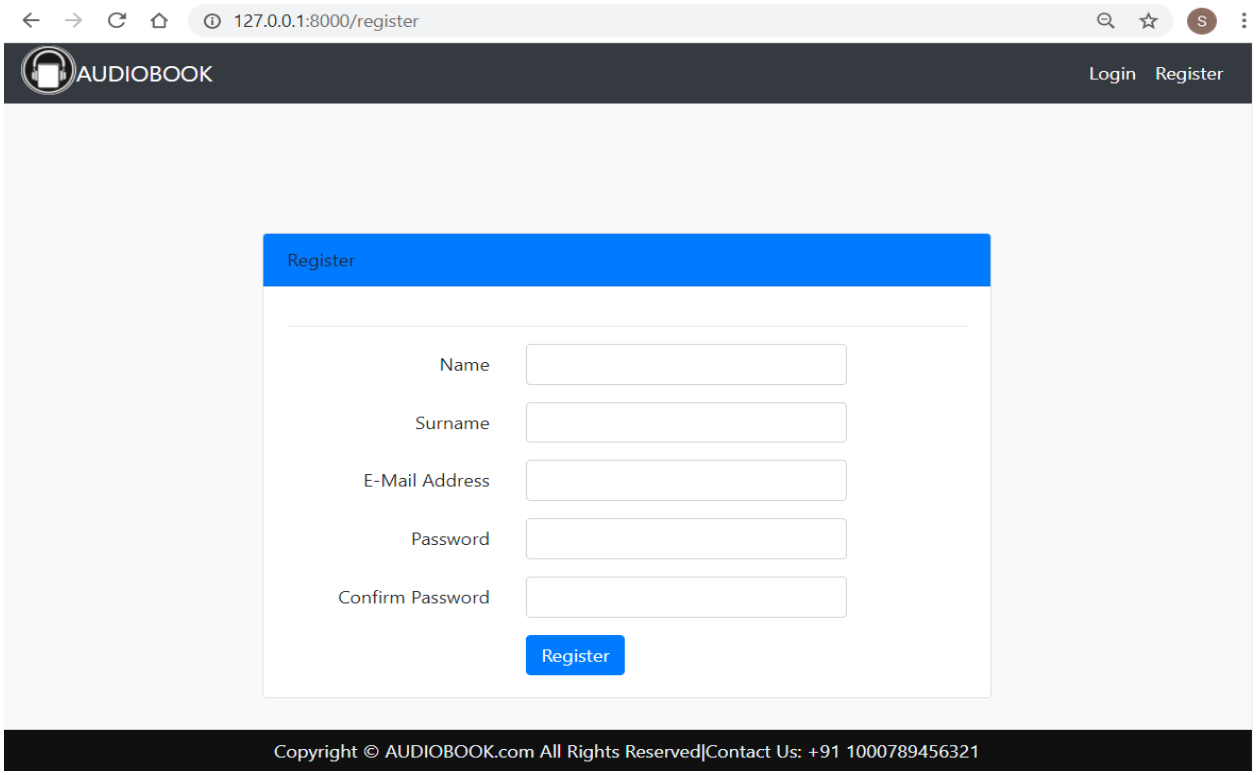


Chapter - 4 UI design and Implementation. Screenshots of webpages of Website/App

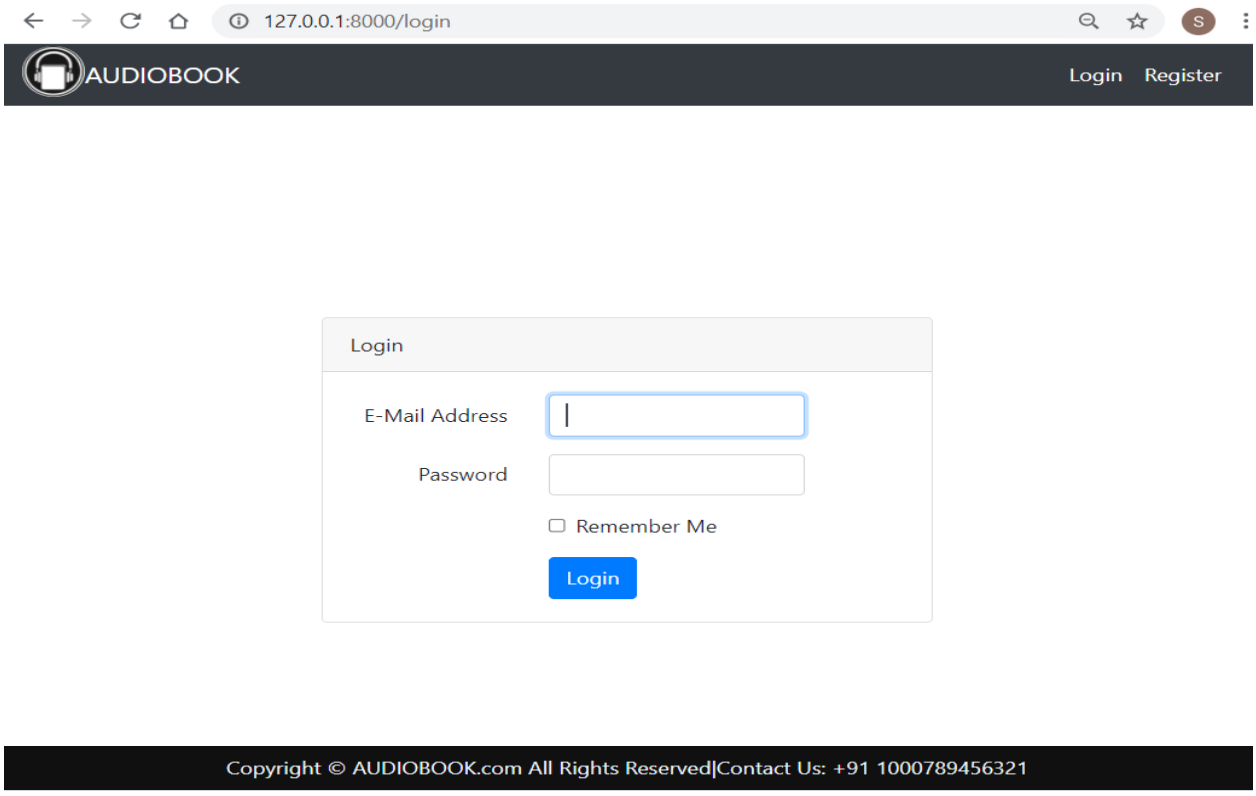
//Home Page



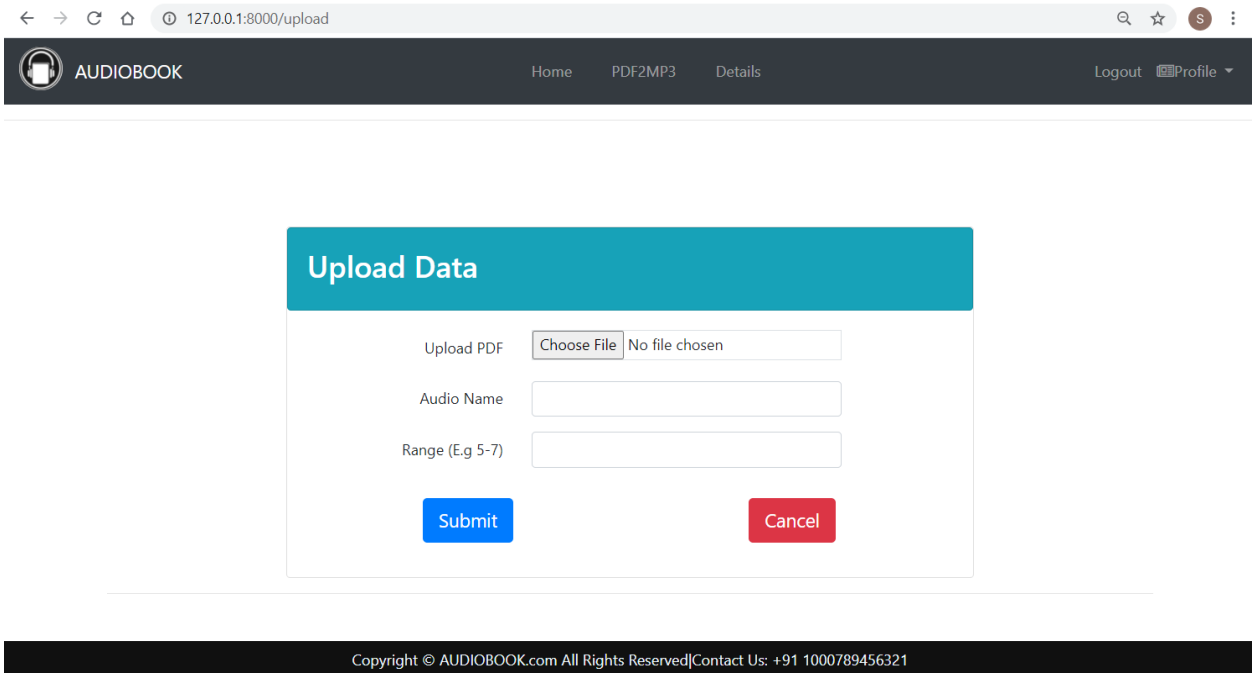
//Register Page



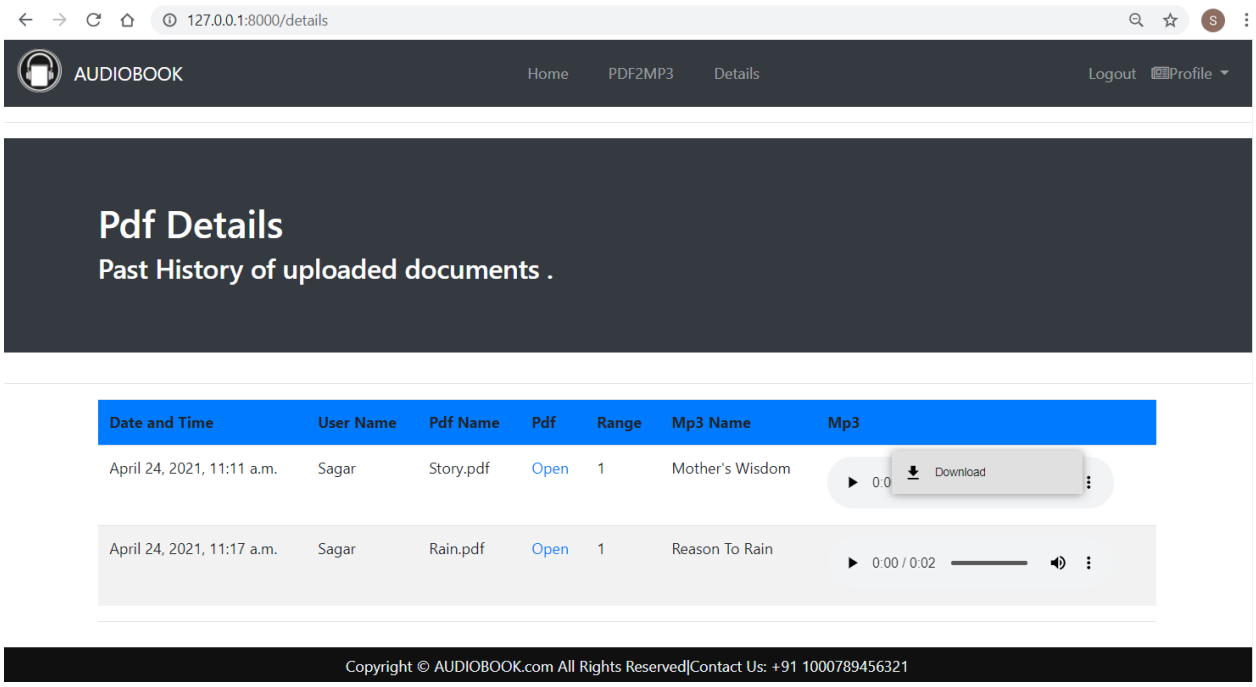
//Login Page



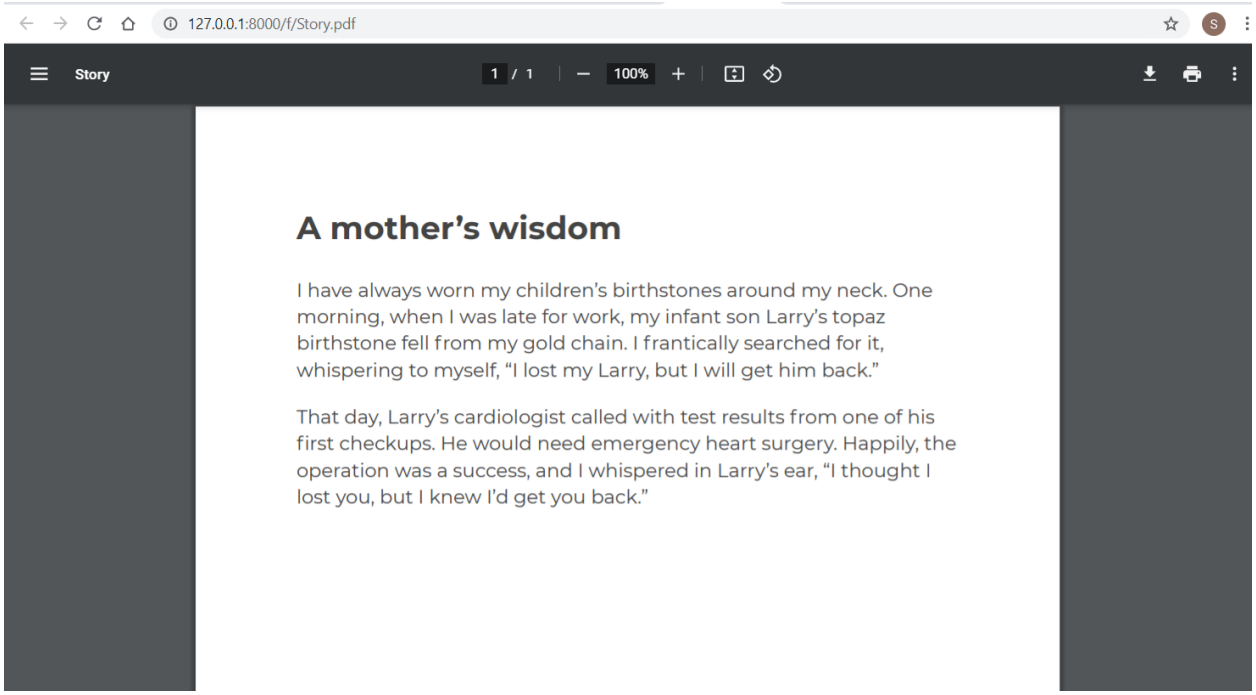
//PDF2MP3 Page



//Details Page



//Story PDF (Example)



Chapter - 5 : UI Evaluation

Sr No	Group No	Roll No	Name	Parameters to be tested (rate out of 5, 1 being lowest and 5 being highest)										Any suggestions
				Visibility of system status	Match between system and real world	User control and freedom	Consistency and standards	Error Prevention	Recognition rather than recall	Flexibility and efficiency of use	Aesthetic and minimalist design	Helps user recognise, diagnose and recover from	Help and Documentation	
1	13	27	Karan Idnani	4	5	4	5	4	4	4	4	4	4	Help manual can be provided
2		34	Mohit Khiani	5	5	4	4	4	4	5	4	4	4	Download option should be given
3		42	Sourav Mantri	4	4	5	4	4	4	5	4	5	4	Pop-Ups can be included for opening the PDFs
4	14	26	Aishwarya Goythale	4	5	5	4	4	4	5	4	4	5	There should be multiple voice options for users to choose
5		49	Janhvi Patil	5	5	4	4	4	4	5	4	4	4	audiobook recommendations can be given to users
6		52	Purav Rathod	5	4	5	4	4	4	5	4	5	4	hyperlinks should also be used for document uploading

Chapter - 6 **:** Conclusion and Future Enhancement

CONCLUSION

Hence, we proposed an approach to create an portable text to speech converter.This system can help the visually impaired people or any person to learn from audio read-back of any scanned text, by converting the uploaded pdf to image,extracting the text from image,and converting the text to audio as mp3 file.

The application developed will be user friendly, cost effective and applicable in the real time. This can save time by allowing the user to listen to background materials while performing other tasks. System can also be used to make information browsing for people who do not have the ability to read or write. This approach can be used in part as well. If we want only text conversion then it is possible and if we want only text to speech conversion then it is also possible easily. People with poor vision or visual dyslexia or totally blindness can use this approach for reading the documents and books.People with speech loss or totally dumb person can utilize this approach to turn typed words into vocalization.

FUTURE ENHANCEMENT

We can recommend audio books to users based on the genre of the books. We can also use collaborative filtering technique to recommend books to the users.
For now our system accepts only pdf as a input but in future or portal should also accepts docs , txt , rtf formats as well as we can also provide an option where user can input link of a specific document and our system should web scrape all the text and should convert it into audio format.

References :

[1] S. A. Stevelink, E. M. Malcolm, and N. T. Fear, “Visual impairment, coping strategies and impact on daily life: a qualitative study among working-age uk ex-service personnel,” BMC public health, vol. 15, no. 1, p. 1118, 2015.

[2] P. Baranyi, A. Csapo, and G. Sallai, Cognitive Infocommunications (CogInfoCom). Springer, 2015

[3] Erdamar, Bengul (1992). Radio Programming. Istanbul, Der Publishing. Ozgur, Aydin Ziya (2017).

[4] “Radio As a Medium Supports Learning at Distance Learning and Open Education Faculty Applications”, BITE, Ministry of Education: FRTEB, Ankara. Race, Phil (2014).

[5] 500 Tips for Open and Flexible Learning, London, Kogan Page. Rowntree, Derek (1994). Teaching with Audio in Open and Distance Education, London, Kogan Page.