Image To Speech Conversion Website

Sanket Munot [1], Akshay Patil [2], Utkarsha Kandale [3], Prof. Supriya S Ambarkar [4]
[1, 2, 3, 4] Dept of Computer Science and Engineering
[1, 2, 3, 4] Walchand Institute of Technology, Solapur

Abstract- *The traditional way of reading a book is through a printed copy. It's not possible to carry a physical copy of a book everywhere, that's when e-books came in handy. Printed books as well as e-books affect the eyes. Visual impairment is one of the biggest limitation for humanity, especially in this day and age when information is received by reading a lot of books (electronic and paper based). A study performed on "Listening or Reading: what's more effective?" shows us that people tend to understand more while listening to book rather than reading it.*

Considering the result our website converts hardcopy of a book into mp3 audible format. On this platform user just needs to upload the image of pages which will be converted into an audible form. The end user will not have to worry about eye problems. This web based platform converts the book into a saga which you can listen and thus carry the book anywhere by a few clicks.

I. INTRODUCTION

The purpose behind this project is to provide a web-based platform which converts any textual snippet into an .mp3 audible format, which the end user can listen to whenever they want .The main objective is take an input image from the user in the web interface and process it to extract text, which then shall be converted into speech. Our platform also provides a Mail Service where audio format that is generated can be sent to the mail address provided by the user.

Reading books and e-books have effects on your health. Most of the problems is caused in eye i.e. eye straining. Elderly people have a problem with the eyesight. Reading books is not possible for them. As a solution to this problem our platform provides an audio of the book to the user so that they can listen to book instead of reading.

Visually impaired people require special type of books which tend to be costlier than the normal one. As we observe these students lack in receiving proper education due to their disability and we live in a society where spoon feeding is preferred more henceforth there is not much of effort put to help these students. So providing them with the audio of the book freely is one of the objective.

II. BACKGROUND

There are many online web-based platforms which do convert from text to speech but they do have some limitations.

1) www.naturalreaders.com:

Is an online site which provides the same facility here you can upload different formats such as docx, txt, jpeg etc. But for .png or .jpg format i.e. an image they as the user to upgrade the since OCR is required to extract text from an image.

Figure 2.1 shows us that the site is asking the user to upgrade i.e. pay charges for providing the facility of Optical character Recognition.

The monthly subscription around 1100rs which is way too expensive if the user doesn't use it often.

And it's not possible for the end user to have a word file of every book they want to listen.

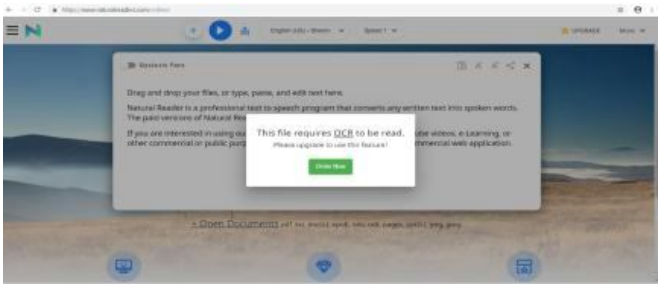


Figure 2.1

2)More platforms like ttsreader.com , fromtexttospeech.com and many more offer to the service of text to speech .The problem here is that the user needs to enter the text in the text box they provide shown in Figure 2.2.

After entering the text it is converted to speech. Again the problem is same that no image is taken and the user needs to have a text file or enter the text separately which is a lot of work to do.

Figure 2.2

We do have individual platforms which perform separate conversions such as:

3) Platform called www.onlineocr.net offers service to perform conversion from image to pdf, which consumes more time to provide an output.

After this user can copy the text from the pdf or doc file into one of the site providing text to speech conversion

All of these process needs a lot of time uploading the .jpg format then converting it into the docx file then the user can copy the text from the output i.e the pdf and paste that text into textbox. But it's a crisscross way for every page of the book the end user has to repeat the same process over and over again.

III. TECHNOLOGIES USED

1. HTML

HyperText Markup Language is used to create the website it also provides two forms out of which user can submit query in any one.

2. CSS

Cascading Style Sheet is the collection of all the styles used to develop the website.

3. FLASK

Flask is micro framework used to render web applications and websites using python. Flask is light weight module and requires no additional libraries.

4. OPENCV

Open Computer Vision is a library of functions to perform real time computer vision. OpenCV is used to read and write the images provided by user and to convert the image to grey scale.

5. TESSERACT-OCR

Tesseract Optical Character Recognition engine is considered to be the most accurate engine for character recognition. Tesseract OCR is the heart of the project it converts the image to a readable text.

6. PYTHON

Python framework is used for building the entire backed and the flask server. Python is the base programming language used for the project.

IV. WORKING

The website works solely on python framework, for the project python 3.6.7 is used. As the project idea describes that an uploaded image is converted into audible mp3 format. So the process of is divided into six parts:

1. Getting the image from user in website.
2. Reading the image.
3. Process Image.
4. Image to Text.
5. Text to speech.
6. Return audio file -
    a. Play on Browser.
    b. Mail the audio.

1) Getting the image from user in web app –:

A web based platform is provided for user to interact with Image to Speech API. User need to upload an image in the web form. 'POST' method is used to send the image to python script.

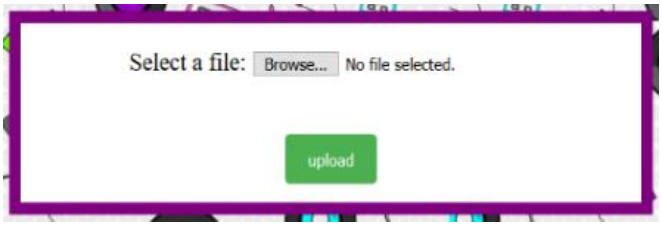


Figure 4.1 Upload image to get audio instantly

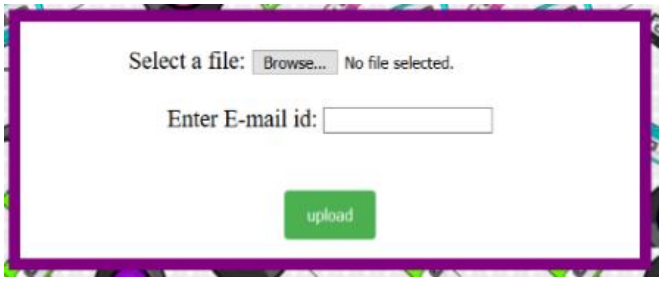


Figure 4.2 Upload image and email id to get audio in mailbox

2) Reading the Image -:

Once the received by python script Open-CV's cv2 class is used to read the image and convert it into grey scale image of 1's and 0's.

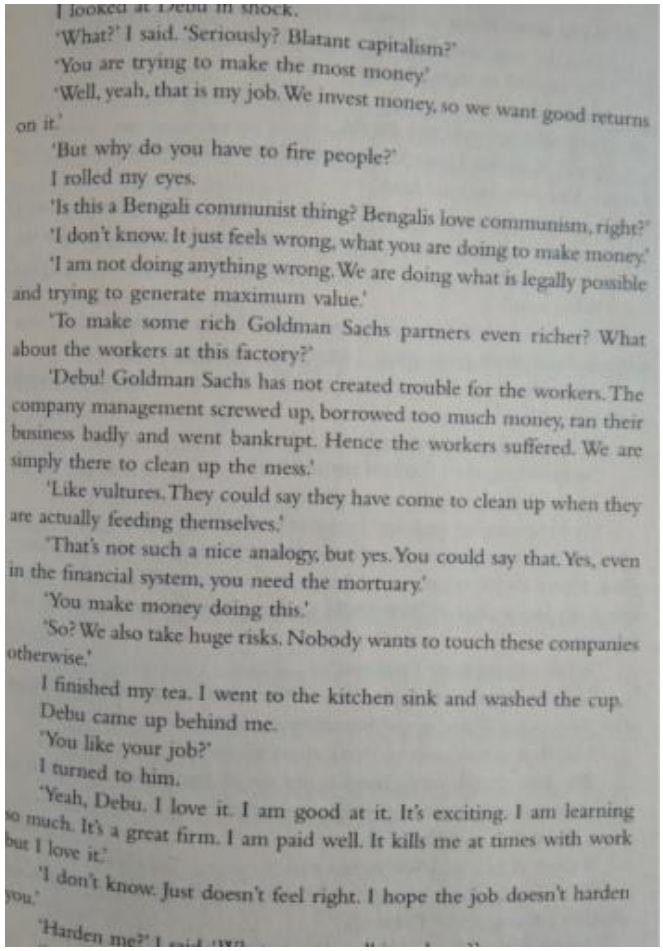
3) Process Image -:

The array of image is further dilated and eroded to make the image noise free and the pixels of image are increased and it is converted into gray scale so that it can be ready to pass to OCR Engine.

4) Image to Text

The noise free clear image is passed to Tesseract-OCR engine. The OCR engine converts the typed data in the image to string format.

The string's accuracy depends on the clarity of the image, however in some cases there are mistakes in conversion of punctuation symbols like (",',!).



I looked at Debu in shock.
'What?' I said. 'Seriously? Blatant capitalism?'
'You are trying to make the most money.'
'Well, yeah, that is my job. We invest money, so we want good returns on it.'
'But why do you have to fire people?'
I rolled my eyes.
'Is this a Bengali communist thing? Bengalis love communism, right?'
'I don't know. It just feels wrong, what you are doing to make money.'
'I am not doing anything wrong. We are doing what is legally possible and trying to generate maximum value.'
'To make some rich Goldman Sachs partners even richer? What about the workers at this factory?'
'Debu! Goldman Sachs has not created trouble for the workers. The company management screwed up, borrowed too much money, ran their business badly and went bankrupt. Hence the workers suffered. We are simply there to clean up the mess.'
'Like vultures. They could say they have come to clean up when they are actually feeding themselves.'
'That's not such a nice analogy, but yes. You could say that. Yes, even in the financial system, you need the mortuary.'
'You make money doing this.'
'So? We also take huge risks. Nobody wants to touch these companies otherwise.'
I finished my tea. I went to the kitchen sink and washed the cup.
Debu came up behind me.
'You like your job?'
I turned to him.
'Yeah, Debu. I love it. I am good at it. It's exciting. I am learning so much. It's a great firm. I am paid well. It kills me at times with work but I love it.'
'I don't know. Just doesn't feel right. I hope the job doesn't harden you.'
'Harden me?'

Figure 4.3 Input to the application.

```
| looked at Debu in shock.
'What?' I said. "Seriously? Blatant Capitalism?
'You are trying to make the mogt money,

'Well, yeah, that is my job. We invest money, so we want good returns

it.
m 'But why do you have to fire people?'

I rolled my eyes.

'Is this a Bengali communist thing? Bengalis love communism, right?"

'| don't know. It just feels wrong, what you are doing to make money?

'Tam not doing anything wrong. We are doing what is legally possible
and trying to generate maximum value,

'To make some rich Goldman Sachs partners even richer? What
about the workers at this factory?'

'Debu! Goldman Sachs has not created trouble for the workers, The
company Management screwed up, borrowed too much money, ran their
business badly and went bankrupt. Hence the workers suffered. We are
simply there to clean up the mess"

'Like vultures. They could say they have come to clean up when they
are actually feeding themselves.

'That's not such a nice analogy, but yes. You could say that. Yes, even
in the financial system, you need the mortuary'

'You make money doing this.
```

Figure 4.4 Textual input at the server

5) Text to Speech

Google Text to Speech API gTTS is used to convert the generated string from previous phase to audio format. GTTS also gives option to slower the voice of speech but for ease this is kept false in the program.

6) Return Audio File

a) Play on Browser

The audio is returned using the return statement as flask environment is used it allows to directly return the file on the browser and the support to play media on the browser.

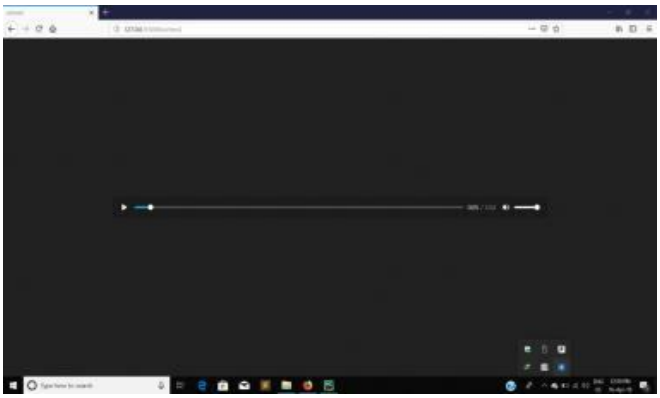


Figure 4.5 playing the audio file on browser.

b) Mail the Audio

The other or more private way to save the media file is using the mail service. For this Yagmail API of python is used, which provides unlimited mails in a day and allows to integrate G-Mail and attachment.

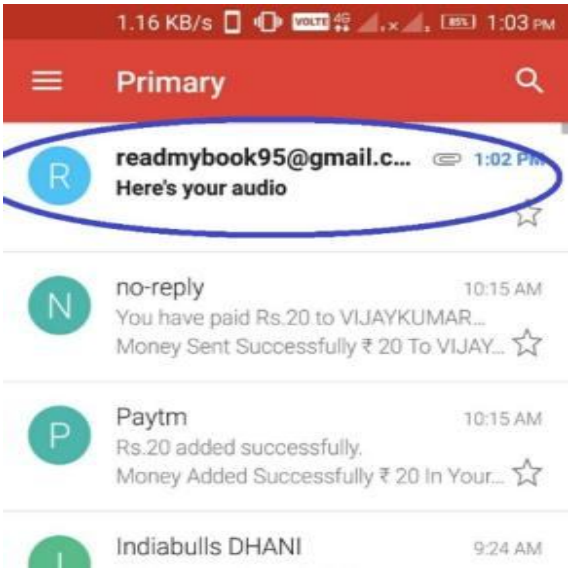


Figure 4.6 Audio as an attachment in the mail box

V. SNAPSHOTS

1. Website home page

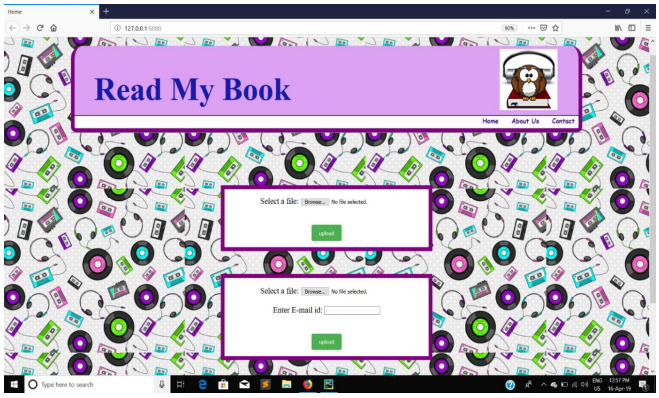


Figure 5.1

2. Server end

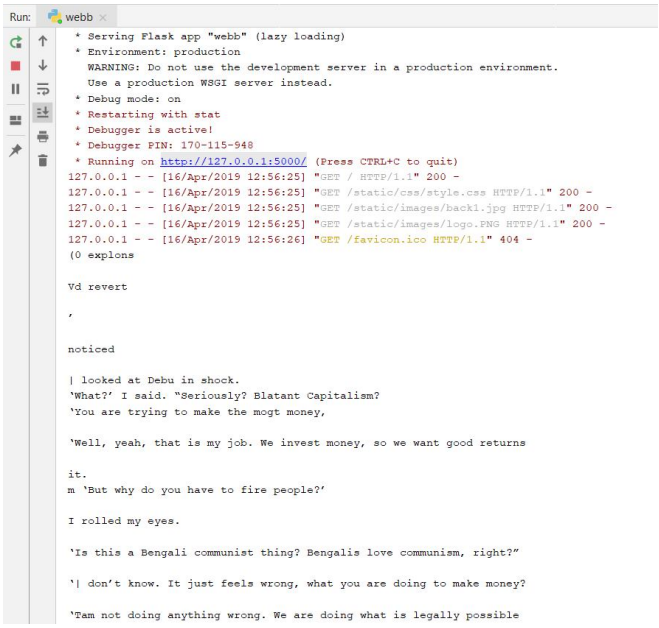


Figure 5.2

VI. ADVANTAGES

1) The mp3 format of the book is helpful for visually impaired persons.
2) As the output is in audible format of book user can listen it while doing some other task's this way time is utilized.
3) Since the book is audio user don't actually need to carry the book.
4) Books are mainly used by students and often while reading students intend to sleep with our audio format they can listen to books a live audio lecture.

5) Several research(s) say that listening improves the power of imagination, so the end users can simulate the scenes occurring in book faster than reading.

VII. FUTURE SCOPE

1) The product provides an Optical Character Recognizer tool so for instance it is only able to capture textual data, due to this various images throughout the book are left uncaptured.Hence the audiobook doesn't involve any diagrams or images present in the book.
2) Only one image can be uploaded, so if there are more pages the end user needs to upload them one by one which consumes more time.
3) The whole text to speech conversion depends on internet connectivity, with slow internet connection the speech conversion might take long to generate or may even end up breaking the code.
4) GTTS will need an internet connection to convert the text into an audio. So it can be slow then other offline APIs.

Right now only one image can be uploaded at a time, work is to be done so that user can upload multiple images at a time. The audio generated is for per page, so if you're having multiple pages you'd get many mp3 format, so considering the sequence in which the image is uploaded the audio files can be combined into a single .mp3 which will be a more convenient way for user to download.

Currently the program is limited to English language however, user might upload an image in some other language where the program fails. Tesseract-OCR is capable to recognize various languages. Hence for future scope this might be a goal to implement various languages in project.

The string generated by OCR contains various punctuation symbols, some of the symbols like (",',!) Are not recognized by Text to speech service therefore they are spelled (double inverted comma etc.). (String optimization)

VIII. CONCLUSION

Our website provides the user with an audio file of the book. It is a web based platform which allows the user to upload the image of the book pages and converts them into an mp3 format. Facilities of downloading the mp3 file, playing it on the browser or getting it mailed are there .This way the user gets many ways to receive an audio file.

So by letting go of the traditional way of reading a book this platform offers a new way i.e. listening to a book.

REFERENCES

[1] Dafe, S. G., and S. S. Chavhan. "Optical Character Recognition Using Image Processing." *International Research Journal of Engineering and Technology* 5.3 (2018): 962-964.
[2] Mizan, C. M., Chakraborty, T., & Karmakar, S. (2017). Text Recognition using Image Processing. *International Journal of Advanced Research in Computer Science*, *8*(5)