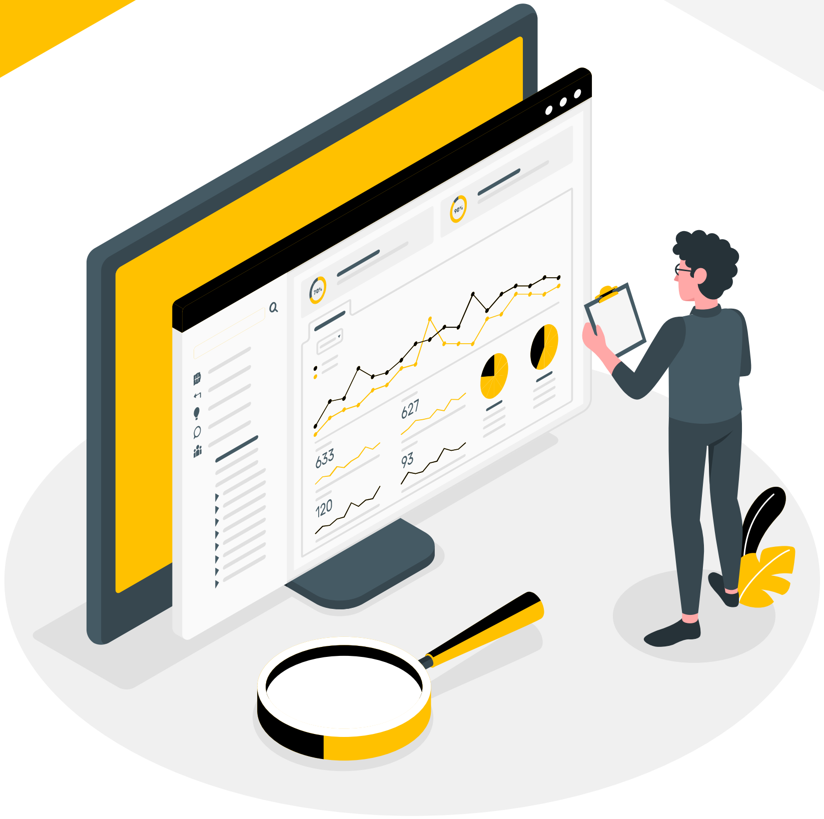
ReadIt: A Multilingual Text to Speech Tool for OCR and Text Summarization Models

Srinidhi Shukla,
Nithin Teja Reddy Gottam,
Sandeep Srinivas Guthula,
Divyaj Reddy Nalla



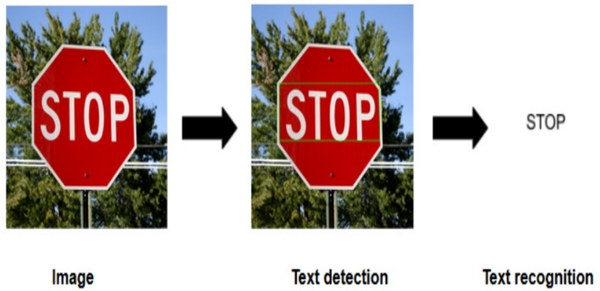
Contents

Introduction

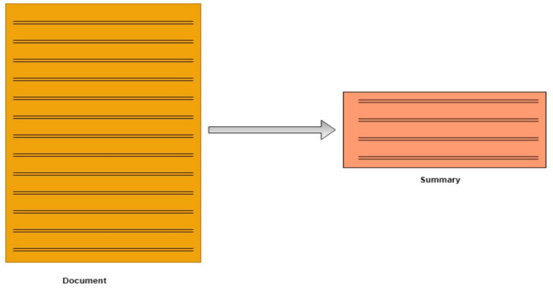
- Textual data is booming in today's world and it's not practical for everyone to read the entire text and comprehend it.
- So, what we need is a method to condense the text while maintaining the essence of the source text.
- That Method should work even in the case of visually impaired and those with color blindness who are unable to interpret the text present in an image.

Introduction

Image-Text-Speech Conversion



Text Summarization and Speech Conversion

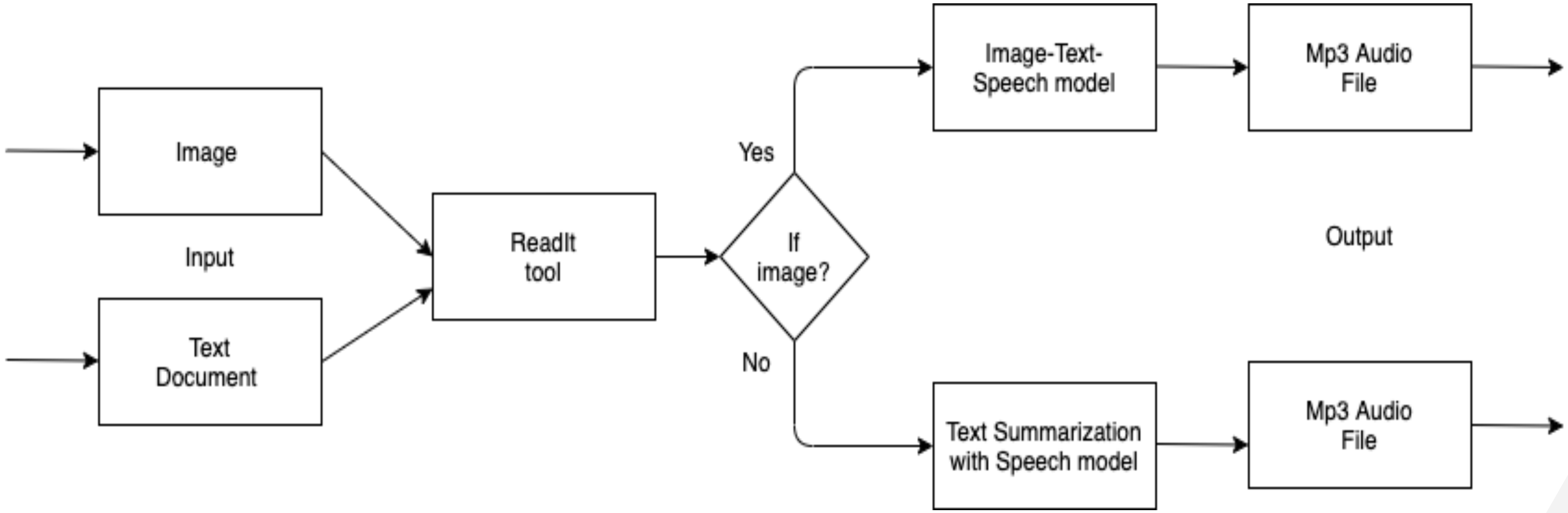


Related Work

- Image Processing Based Scene-Text Detection and Recognition with Tesseract by Ebin Zacharias, Martin Teuchler and Bénédicte Bernier. The aim of this project is to detect and recognize words in natural photographs. The intended challenge is considerably more difficult than translating text from scanned papers. Because of the limited number of photographs available, the use case in focus makes it easier to detect the text field in natural scenes with greater precision.

- Smart Summarizer For Blind People by Mona Teja K , Mohan Sai.S, H S S S Raviteja ,Sai Kushagra. This paper discusses an effective technique for condensing news into relevant keywords in order to avoid having to read the whole text each time. Many APIs and modules, such as tesseract and GTTS, are explored and introduced in depth in this article, as are many algorithms, such as Luhn's Algorithm, Latent Semantic Analysis Algorithm, and Text Ranking Algorithm.

Architecture



Proposed Model

Readit tool consisting of the following two models

Image-Text-Speech Conversion

- Optical Character Recognition (OCR)
- Translated into an audio file (mp3) using GTT

Text Summarization and Speech Conversion

- TextRank Algorithm
- Translated into an audio file (mp3) using GTT

Image-Text-Speech Conversion

- We used Optical Character Recognition (OCR) to enable listening to the text written on the image. It is a technique for extracting text from images and translating it to an electronic format.
- OCR is typically split into many sub-processes.
  - a) Preprocessing of the Image
  - b) Text Localization
  - c) Character Segmentation
  - d) Character Recognition
  - e) Post Processing

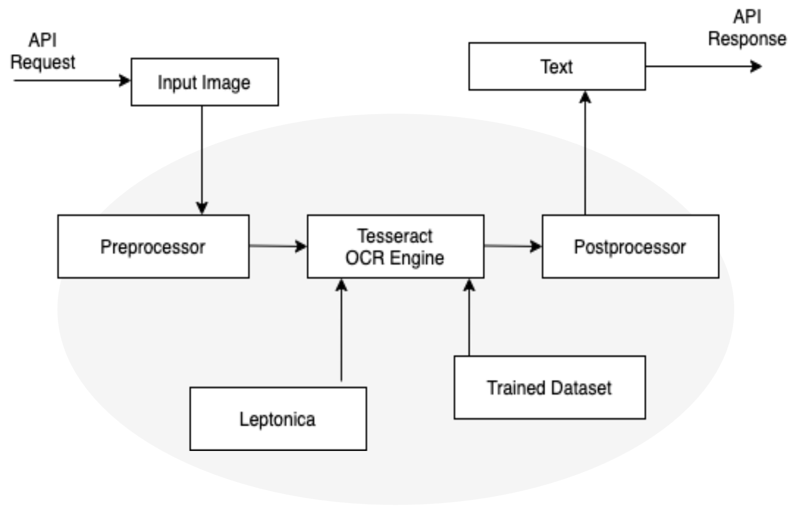


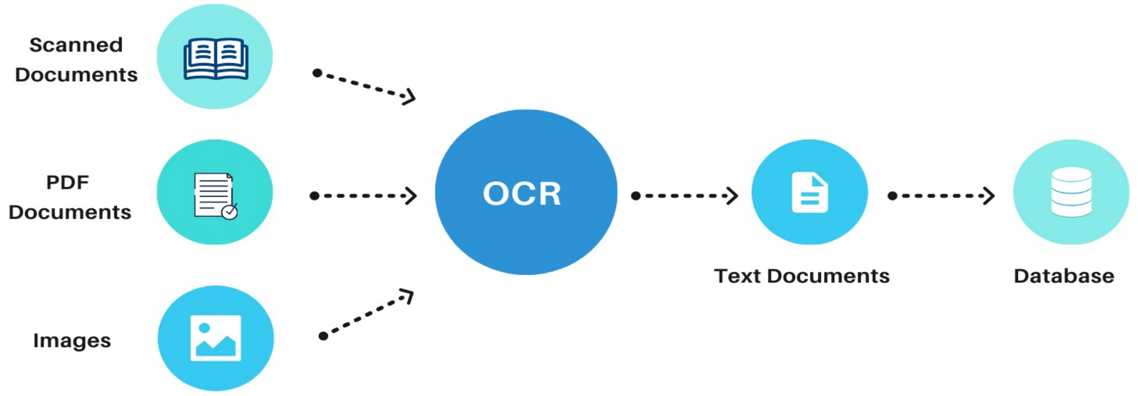
Image-Text-Speech Conversion

- The system's module is the TESSERACT GOOGLE API, which is an optical character recognition (OCR) engine that supports Unicode and can identify several languages out of the box

- It is adaptable and can be learned to recognize other languages as well.

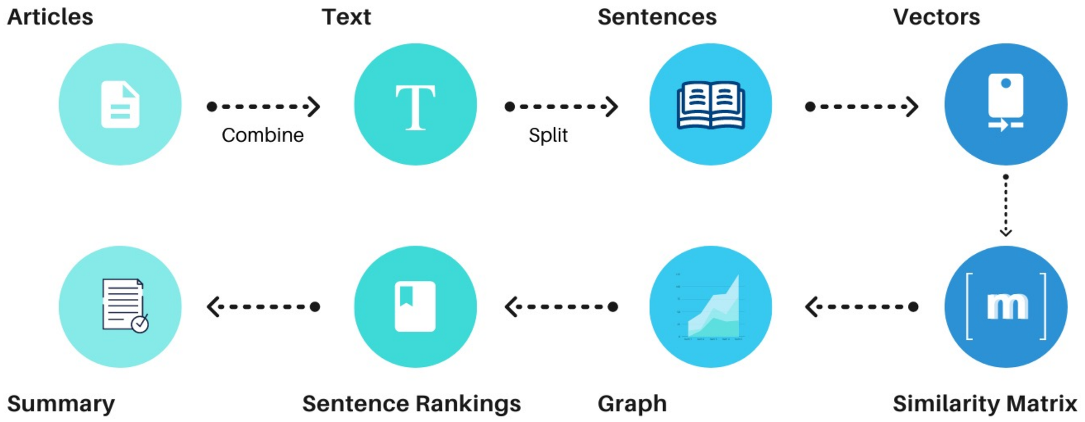
- The Tesseract can detect over 100 languages, and Google has boosted it with a neural network framework based on LSTM starting with version 4. (long short-term memory).

- The Libraries we used in this model are Pytesseract, PIL, Nltk, Numpy, OS, GTT’s.



Text Summarization and Speech Conversion

- TextRank is a graph-based text processing ranking model that can be used to find the most relevant sentences in a text as well as keywords.
- Tasks involved in summarization
  a) Relevant sentences identification
  b) Relevant keywords identification



Text Summarization and Speech Conversion

- The PageRank algorithm is applied to the resulting network to determine the importance of each expression.
- Libraries used: Nltk, NetworkX, Langdetect

Training

Because of the libraries we used, the whole model does not require any data to be trained, and the results can be obtained simply by uploading an image or file and making a selection on the front-end interface.

The Front-end Execution

- In the final stage, we use Tkinter, the only one that is used in the standard library, to create a GUI that incorporates all of these models and makes it simpler for the user.

- Tkinter programs appear like they belong to the platform they're running on because visual elements are made using native operating system elements.

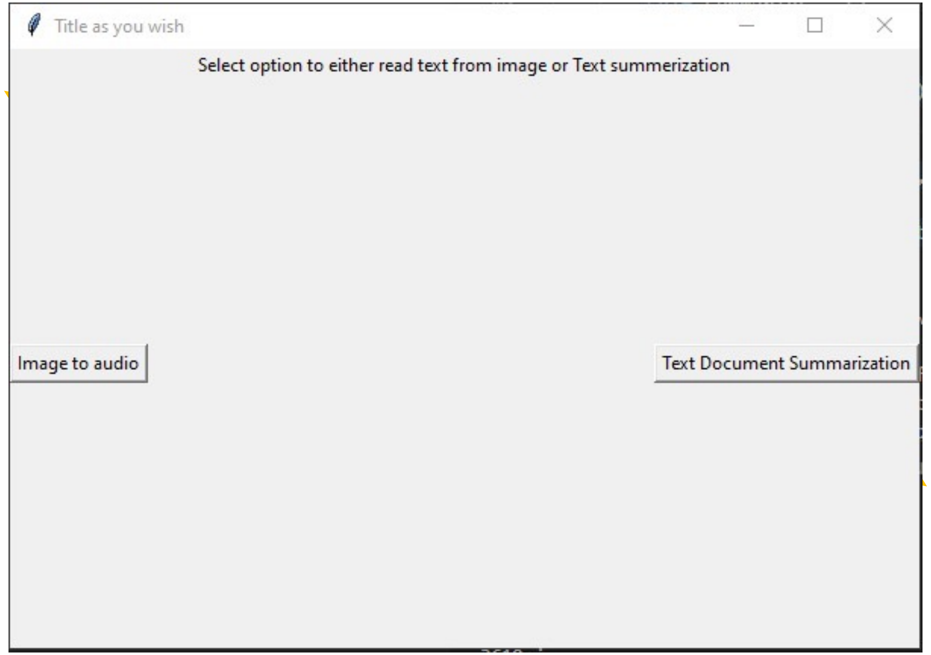
- Tkinter makes it easy to make a graphical user interface using the steps below.

a) Import the Tkinter module to the project.
b) Build the main window for the GUI application.
c) In the GUI framework, add one or more of the above-mentioned widgets.
d) Enter the main event loop to respond to each user-triggered event.

Results and Evaluation

Results and Evaluation

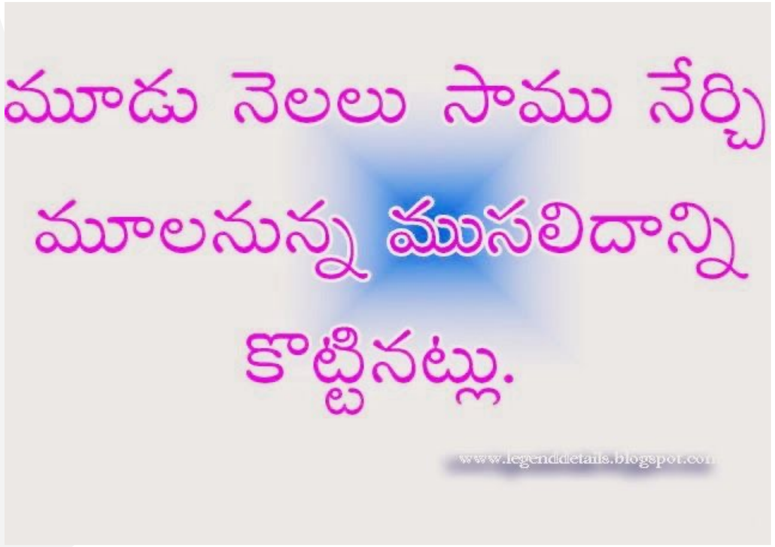
The output of the code is shown below. It shows a front-end interface with two choices for the user to choose from:


Title as you wish
Select option to either read text from image or Text summerization
Image to audio
Text Document Summarization

Results and Evaluation

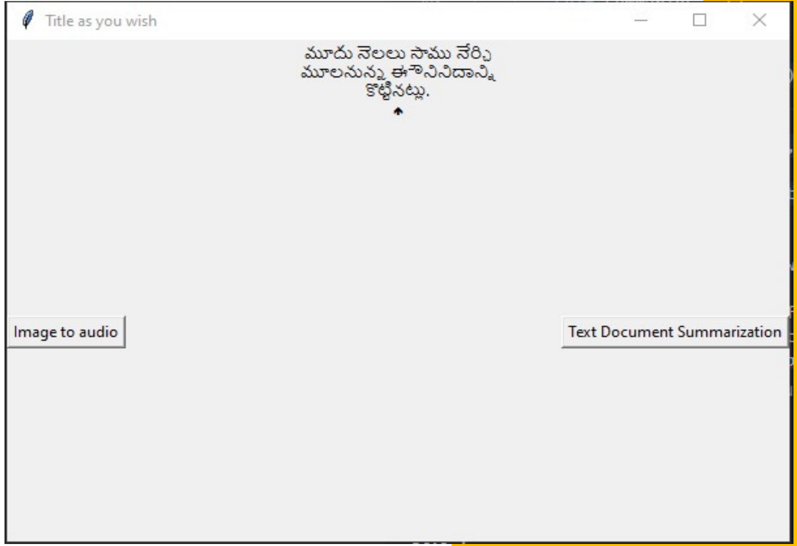
Execution of Image-Text-Speech model:

Step 1: Uploading the images and converting of the text present on the image and displaying it



మూడు నెలలు సాము నేర్చి
మూలనున్న ముసలిదాన్ని
కొట్టినట్లు.

Input



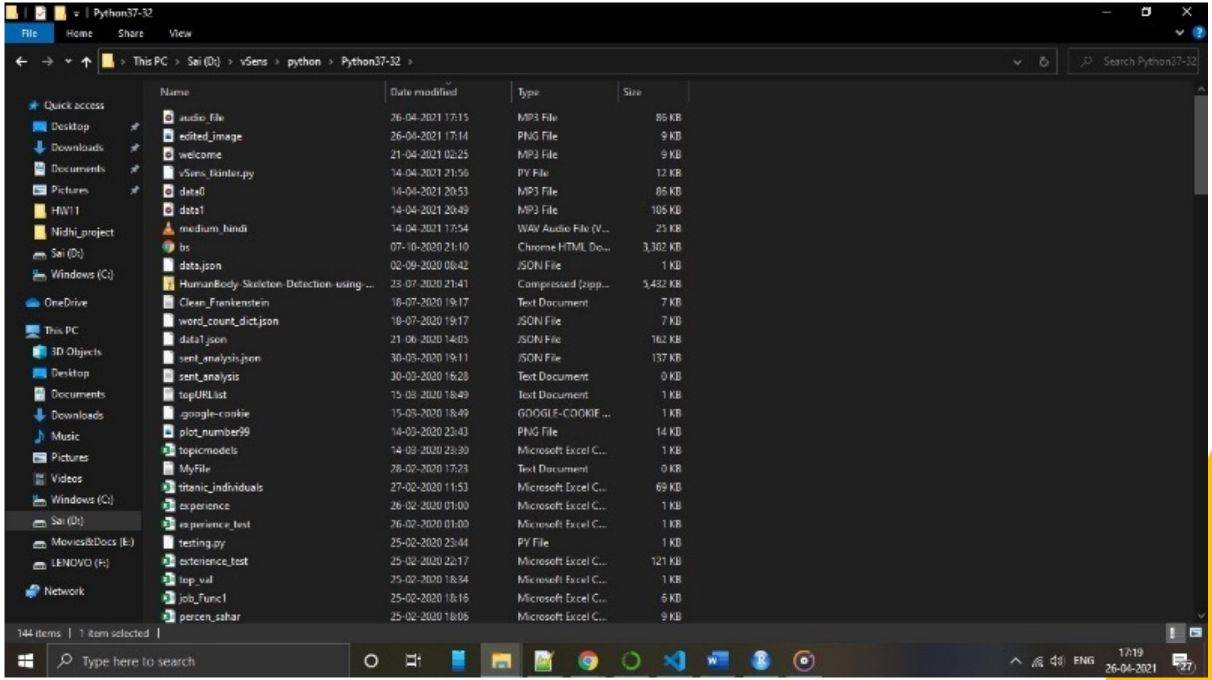
మూడు నెలలు సాము నేర్చి
మూలనున్న ఈానినిదాన్ని
కొట్టినట్లు.

Output

Results and Evaluation

Execution of Image-Text-Speech model:

Step 2: Saving the converted text into audio file to play.



This PC › Sai (D:) › vSens › python › Python37-32

Name	Date modified	Type	Size
audio_file	26-04-2021 17:15	MP3 File	86 KB
edited_image	26-04-2021 17:14	PNG File	9 KB
welcome	21-04-2021 02:25	MP3 File	9 KB
vSens_tkinter.py	14-04-2021 21:56	PY File	12 KB
data0	14-04-2021 20:53	MP3 File	86 KB
data1	14-04-2021 20:49	MP3 File	106 KB
medium_hindi	14-04-2021 17:54	WAV Audio File (V...	25 KB
bs	07-10-2020 21:10	Chrome HTML Do...	3,302 KB
data.json	02-09-2020 08:42	JSON File	1 KB
HumanBody-Skeleton-Detection-using-...	23-07-2020 21:41	Compressed (zipp...	5,432 KB
Clean_Frankenstein	18-07-2020 19:17	Text Document	7 KB
word_count_dict.json	18-07-2020 19:17	JSON File	7 KB
data1.json	21-06-2020 14:05	JSON File	162 KB
sent_analysis.json	30-03-2020 19:11	JSON File	137 KB
sent_analysis	30-03-2020 16:28	Text Document	0 KB
topURLlist	15-03-2020 18:49	Text Document	1 KB
.google-cookie	15-03-2020 18:49	GOOGLE-COOKIE ...	1 KB
plot_number99	14-03-2020 23:43	PNG File	14 KB
topicmodels	14-03-2020 23:30	Microsoft Excel C...	1 KB
MyFile	28-02-2020 17:23	Text Document	0 KB
titanic_individuals	27-02-2020 11:53	Microsoft Excel C...	69 KB
experience	26-02-2020 01:00	Microsoft Excel C...	1 KB
experience_test	26-02-2020 01:00	Microsoft Excel C...	1 KB
testing.py	25-02-2020 23:44	PY File	1 KB
exterience_test	25-02-2020 22:17	Microsoft Excel C...	121 KB
top_val	25-02-2020 18:34	Microsoft Excel C...	1 KB
job_Func1	25-02-2020 18:16	Microsoft Excel C...	6 KB
percen_sahar	25-02-2020 18:06	Microsoft Excel C...	9 KB

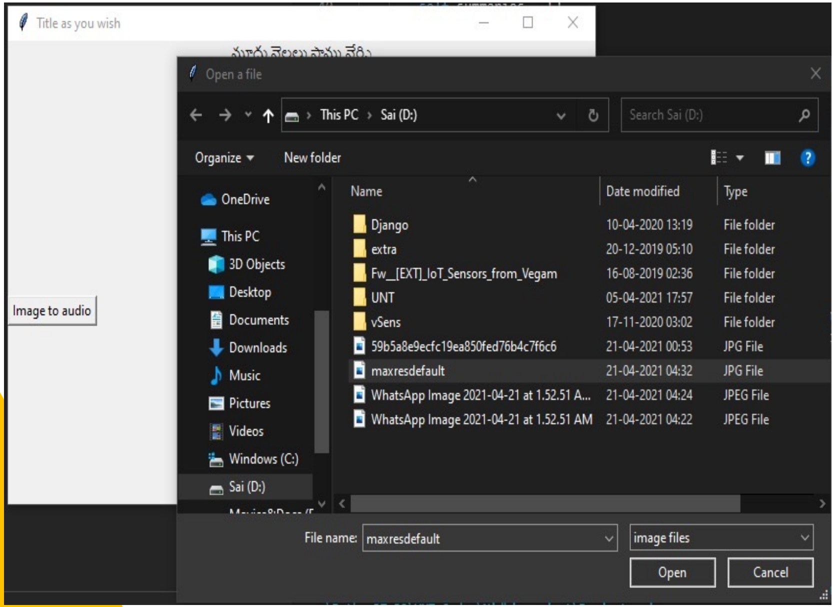
144 items | 1 item selected |

Output 2

Results and Evaluation

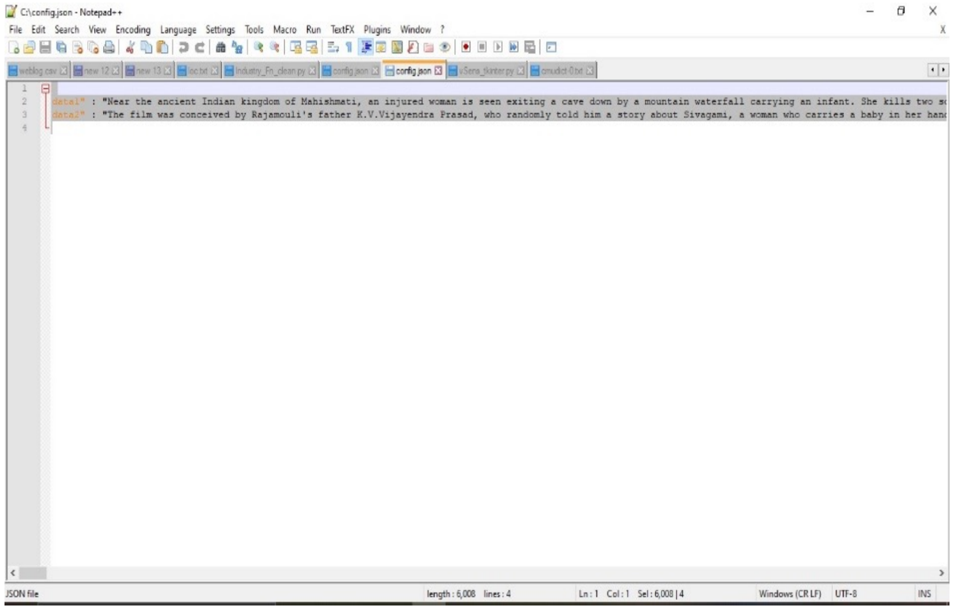
Execution of Text DOCUMENT SUMMARIZATIOIN MODEL:

Step 1: Uploading the document and then getting the summarized text



Name	Date modified	Type
Django	10-04-2020 13:19	File folder
extra	20-12-2019 05:10	File folder
Fw__[EXT]_IoT_Sensors_from_Vegam	16-08-2019 02:36	File folder
UNT	05-04-2021 17:57	File folder
vSens	17-11-2020 03:02	File folder
59b5a8e9ecfc19ea850fed76b4c7f6c6	21-04-2021 00:53	JPG File
maxresdefault	21-04-2021 04:32	JPG File
WhatsApp Image 2021-04-21 at 1.52.51 A...	21-04-2021 04:24	JPEG File
WhatsApp Image 2021-04-21 at 1.52.51 AM	21-04-2021 04:22	JPEG File

File name: maxresdefault
image files



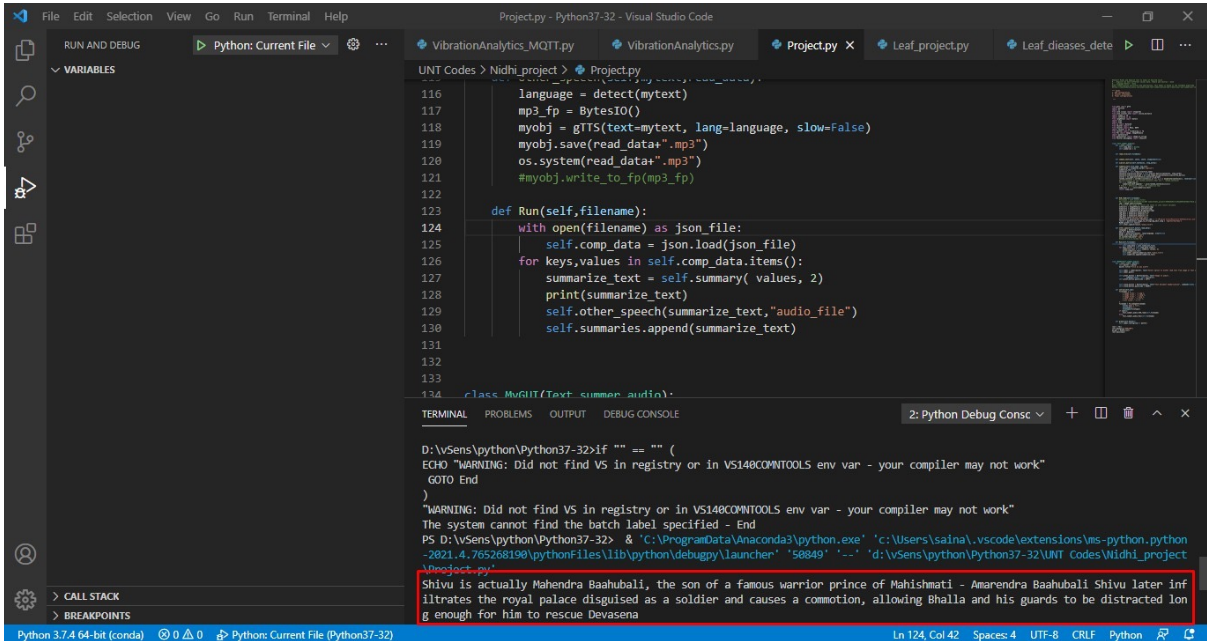
```
data1" : "Near the ancient Indian kingdom of Mahishmati, an injured woman is seen exiting a cave down by a mountain waterfall carrying an infant. She kills two s
data2" : "The film was conceived by Rajamouli's father K.V.Vijayendra Prasad, who randomly told him a story about Sivagami, a woman who carries a baby in her han
```

Input

Results and Evaluation

Execution of Text DOCUMENT SUMMARIZATIOIN MODEL:

Step 1: Uploading the document and then getting the summarized text



```python
        language = detect(mytext)
        mp3_fp = BytesIO()
        myobj = gTTS(text=mytext, lang=language, slow=False)
        myobj.save(read_data+".mp3")
        os.system(read_data+".mp3")
        #myobj.write_to_fp(mp3_fp)

    def Run(self,filename):
        with open(filename) as json_file:
            self.comp_data = json.load(json_file)
        for keys,values in self.comp_data.items():
            summarize_text = self.summary( values, 2)
            print(summarize_text)
            self.other_speech(summarize_text,"audio_file")
            self.summaries.append(summarize_text)



class MyGUI(Text_summer_audio):
```

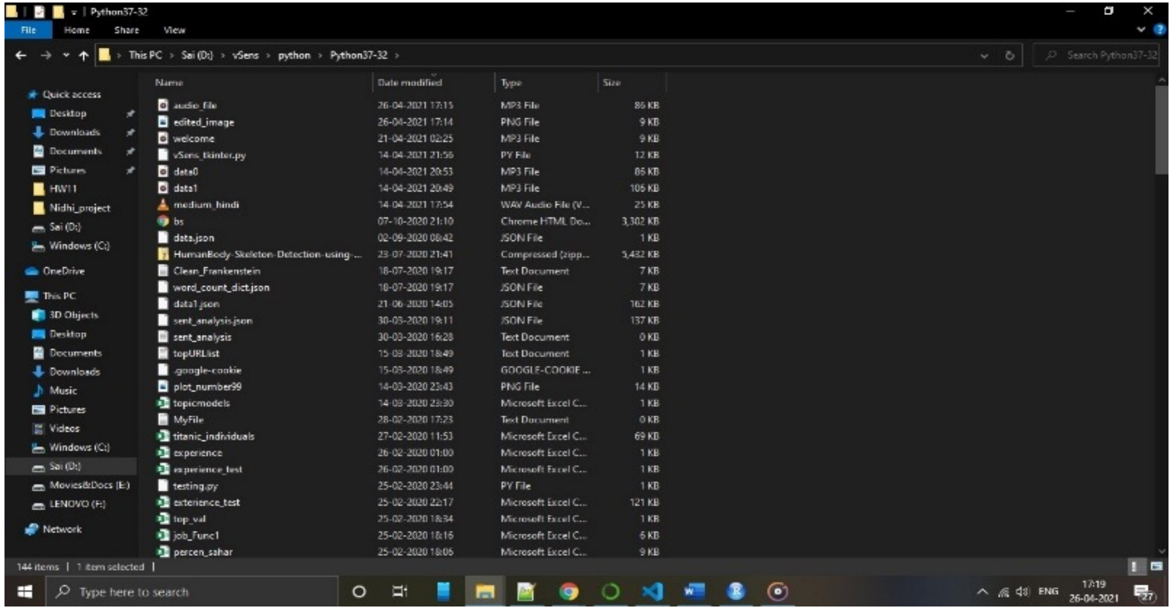
```
D:\vSens\python\Python37-32>if "" == "" (
ECHO "WARNING: Did not find VS in registry or in VS140COMNTOOLS env var - your compiler may not work"
 GOTO End
)
"WARNING: Did not find VS in registry or in VS140COMNTOOLS env var - your compiler may not work"
The system cannot find the batch label specified - End
PS D:\vSens\python\Python37-32>  & 'C:\ProgramData\Anaconda3\python.exe' 'c:\Users\saina\.vscode\extensions\ms-python.python
-2021.4.765268190\pythonFiles\lib\python\debugpy\launcher' '50849' '--' 'd:\vSens\python\Python37-32\UNT Codes\Nidhi_project
\Project.py'
Shivu is actually Mahendra Baahubali, the son of a famous warrior prince of Mahishmati - Amarendra Baahubali Shivu later inf
iltrates the royal palace disguised as a soldier and causes a commotion, allowing Bhalla and his guards to be distracted lon
g enough for him to rescue Devasena
```

Text Summarization Output

Results and Evaluation

Execution of Text DOCUMENT SUMMARIZATIOIN MODEL:

Step 2: Storing the file into audio Mp3 format to play it out:



This PC › Sai (D:) › vSens › python › Python37-32 ›

Name	Date modified	Type	Size
audio_file	26-04-2021 17:15	MP3 File	86 KB
edited_image	26-04-2021 17:14	PNG File	9 KB
welcome	21-04-2021 02:25	MP3 File	9 KB
vSens_tkinter.py	14-04-2021 21:56	PY File	12 KB
data0	14-04-2021 20:53	MP3 File	86 KB
data1	14-04-2021 20:49	MP3 File	106 KB
medium_hindi	14-04-2021 17:54	WAV Audio File (V...	25 KB
bs	07-10-2020 21:10	Chrome HTML Do...	3,302 KB
data.json	02-09-2020 08:42	JSON File	1 KB
HumanBody-Skeleton-Detection-using-...	23-07-2020 21:41	Compressed (zipp...	5,432 KB
Clean_Frankenstein	18-07-2020 19:17	Text Document	7 KB
word_count_dict.json	18-07-2020 19:17	JSON File	7 KB
data1.json	21-06-2020 14:05	JSON File	162 KB
sent_analysis.json	30-03-2020 19:11	JSON File	137 KB
sent_analysis	30-03-2020 16:28	Text Document	0 KB
topURLlist	15-03-2020 18:49	Text Document	1 KB
.google-cookie	15-03-2020 18:49	GOOGLE-COOKIE ...	1 KB
plot_number99	14-03-2020 23:43	PNG File	14 KB
topicmodels	14-03-2020 23:30	Microsoft Excel C...	1 KB
MyFile	28-02-2020 17:23	Text Document	0 KB
titanic_individuals	27-02-2020 11:53	Microsoft Excel C...	69 KB
experience	26-02-2020 01:00	Microsoft Excel C...	1 KB
experience_test	26-02-2020 01:00	Microsoft Excel C...	1 KB
testing.py	25-02-2020 23:44	PY File	1 KB
exterience_test	25-02-2020 22:17	Microsoft Excel C...	121 KB
top_val	25-02-2020 18:34	Microsoft Excel C...	1 KB
job_Func1	25-02-2020 18:16	Microsoft Excel C...	6 KB
percen_sahar	25-02-2020 18:06	Microsoft Excel C...	9 KB

144 items | 1 item selected |

Audio Output

Conclusion

- Successfully developed a front-end interface that allows us to incorporate Image-text-speech and text summarization and speech conversion depending on the user's request.

- This concept was mostly aimed at making it simple for everyone, even blind people, to convert text or image data to audio, and we plan to add more features in the future to make it more practical and reliable.

- For fast conversions and convenient access for the disabled, this tool may also be used as a plug-in or add-on for applications or smart devices (such as a smart watch or Alexa).

References

1]Zacharias, E., Teuchler, M., and Bernier, B., “Image Processing Based Scene-Text Detection and Recognition with Tesseract”, 2020.

[2]Teja K, M., Sai. S, M., Raviteja D, H. S. S. S., and V, S. K. P., “Smart Summarizer for Blind People”, 2020.

[3]Bharath B, M., Gowtham B, A., and M, A., “Neural Abstractive Text Summarizer for Telugu Language”, 2021.

[4]Galatolo, F. A., Cimino, M. G. C. A., and Vaglini, G., “Generating images from caption and vice versa via CLIP-Guided Generative Latent Space Search”, 2021.

Thanks

Do you have
any questions?

Srinidhi Shukla,
Nithin Teja Reddy Gottam,
Sandeep Srinivas Guthula,
Divyaj Reddy Nalla