

Computer Science Project



Multilingual Audiobook Generator

Prepared By

Vikash P

Tharun Kumar M.S

Index :

1. Certificate - Page 3
2. Acknowledgement - Page 4
3. Project Description - Page 5
4. Python Libraries - Page 6 to 8
5. Software Requirements - Page 9
6. Coding - Page 10 to 17
7. Output - Page 18 to 28
8. Bibliography - Page 29

Certificate

Acknowledgement

I would like to express my special thanks and gratitude to my teacher (Mrs Chandi Priya), our principal (Mrs Dolly Mohan), Vice principal (Mr. Anand Thyagarajan) who gave me the golden opportunity to do this wonderful project on this topic (Multilingual Audiobook Generator), which also helped me in doing a lot of research and came to know about so many new things. I am really thankful to them.

Secondly I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame.

Project Description

The Multilingual Audiobook Generator is used to convert PDF to speech which suits best for Novels and story books. Coming to its functionality, the user feeds in the PDF which is converted into text where the text is changed to the desired language and gets converted as an audio file which is saved .

This program provides the users to translate the text into 107 different languages. Thus using this MAG helps different people in different ways. For the blind people, by just uploading the book with someone's help, they can listen to the entire book. People need not buy audiobooks from the market rather generate their own ones with just a softcopy of the book.

Python Libraries

1) PyPDF2

This library is used for performing tasks on PDF files .

2) pdfplumber

It is used to find text in PDF, return page number, extract text etc.

3) deep_translator

A tool to translate between different languages in a simple way using multiple translators.

4) gTTS

Google Text-to-Speech is a Python library and CLI tool used to interface with Google Translate's text-to-speech API.

5) Tkinter

This library is used to generate the necessary GUI needed for the client-side portal.

6) pygame

Pygame is a module which includes computer graphics and sound libraries, designed to be used with the Python programming language.

7) Tabulate

Tabulate is an open-source python module which is used to print tabular data in nicely formatted tables.

Software Requirements

-> Operating System

- Windows, macOS, or Linux

-> Program

- Python IDLE 3.X (or Newer)
- File Manager

Coding

```
1 # Importing the modules
2
3 import PyPDF2
4 import pdfplumber
5 from deep_translator import GoogleTranslator
6 from gtts import gTTS
7
8 from tkinter import *
9 import pygame
10 from tkinter import filedialog
11 from pygame import mixer
12 from tabulate import tabulate
13
14 #
15 print("")
16 print("-----")
17 print(" Hey user! Welcome to the application of Multilingual Audiobooks Generator!!")
18 print("-----")
19 print("")
20 print("")
21
22 # Getting PDF from user
23
24 pdfname = input(" Please enter the Name of the PDF : ")
25 pdf = pdfname+".pdf"
26 print("")
27 pdflen = len(pdf)
28
29 # Reading the data in PDF
30
31 pdfFileObj = open(pdf, "rb")
32 pdfReader = PyPDF2.PdfFileReader(pdfFileObj)
33 pages=pdfReader.numPages
34
```

```

35 hyphencount1 = (36 + pdfllen) * "-"
36
37 # Printing the total number of pages
38
39 print(hyphencount1)
40 print(" The total number of pages in",'','',pdf,'','', "is : " ,pages)
41 print(hyphencount1)
42 print("")
43 pdfFileObj.close()
44
45 mytext = ""
46
47 # Extracting the text from the PDF
48
49 with pdfplumber.open(pdf) as pdf:
50     for i in range(0,pages):
51         page=pdf.pages[i]
52         mytext += page.extract_text()
53
54 # Printing the text
55
56 print("-----")
57 print(" The text in this PDF is :- ")
58 print("-----")
59 print("")
60 print(mytext)
61 print("")
62 print("")
63
64 # Listing the Languages available for translation
65
66 print("-----")
67 print(" The list of languages for translation includes :- " )
68 print("-----")

```

```

69 print("")
70
71 llist = [[1,"Afrikaans"],[2," Albanian"],[3," Amharic"], [4,"Arabic"],[5," Armenian"],
72         [6," Azerbaijani"], [7,"Basque"],[8,"Belarusian"],[9," Bengali"],[10,"Bosnian"],
73         [11,"Bulgarian"],[12," Catalan"],[13," Cebuano"],[14," Chinese"],[15,"Corsican"],
74         [16," Croatian"],[17," Czech"],[18," Danish"],[19," Dutch"],[20," English"],
75         [21,"Esperanto"],[22,"Estonian"],[23," Finnish"],[24," French"],[25," Frisian"],
76         [26," Galician"],[27," Georgian"],[28," German"],[29,"Greek"],[30," Gujarati"],
77         [31," Hausa"],[32,"Haitian Creole"],[33," Hawaiian"],[34," Hebrew"],[35," Hindi"],
78         [36,"Hmong"],[37," Hungarian"],[38," Icelandic"],[39," Igbo"],[40," Indonesian"],
79         [41,"Irish"],[42,"Italian"],[43," Japanese"],[44," Javanese"],[45," Kannada"],
80         [46," Kazakh"],[47," Khmer"],[48," Kinyarwanda"],[49," Korean"],[50," Kurdish"],
81         [51,"Kyrgyz"],[52," Lao"],[53,"Latvian"],[54," Lithuanian"],[55," Luxembourgish"],
82         [56," Macedonian"],[57,"Malagasy"],[58," Malay"],[59," Malayalam"],[60," Maltese"],
83         [61,"Maori"],[62," Marathi"],[63,"Mongolian"],[64," Myanmar"],[65," Nepali"],
84         [66," Norwegian"],[67," Nyanja"],[68," Odia"],[69," Pashto"],[70," Persian"],
85         [71,"Polish"],[72," Portuguese"],[73," Punjabi"],[74,"Romanian"],[75," Russian"],
86         [76," Samoan"],[77," Scots Gaelic"],[78," Serbian"],[79," Sesotho"],[80," Shona"],
87         [81,"Sinhalese"],[82," Slovenian"],[83," Spanish"],[84," Swedish"],[85," Tagalog"],
88         [86,"Tamil"],[87,"Telugu"],[88,"Thai"],[89," Turkish"],[90," Ukrainian"],
89         [91,"Urdu"],[92," Vietnamese"],[93,"Welsh"],[94,"Xhosa"],[95,"Yiddish"],
90         [96,"Yoruba"],[97,"Zulu"],[98,"Afrikaans"],[99,"Albanian"],[100,"Arabic"],
91         [101,"Armenian"],[102,"Azerbaijani"],[103,"Basque"],[104,"Belarusian"],[105,"Bengali"],
92         [106,"Bosnian"],[107,"Bulgarian"],[108,"Catalan"],[109,"Cebuano"],[110,"Chinese"],
93         [111,"Corsican"],[112,"Croatian"],[113,"Czech"],[114,"Danish"],[115,"Dutch"],[116,"English"],
94         [117,"Esperanto"],[118,"Estonian"],[119,"Finnish"],[120,"French"],[121,"Frisian"],
95         [122,"Galician"],[123,"Georgian"],[124,"German"],[125,"Greek"],[126,"Gujarati"],
96         [127,"Hausa"],[128,"Haitian Creole"],[129,"Hawaiian"],[130,"Hebrew"],[131,"Hindi"],
97         [132,"Hmong"],[133,"Hungarian"],[134,"Icelandic"],[135,"Igbo"],[136,"Indonesian"],
98         [137,"Irish"],[138,"Italian"],[139,"Japanese"],[140,"Javanese"],[141,"Kannada"],
99         [142,"Kazakh"],[143,"Khmer"],[144,"Kinyarwanda"],[145,"Korean"],[146,"Kurdish"],
100        [147,"Kyrgyz"],[148,"Lao"],[149,"Latvian"],[150,"Lithuanian"],[151,"Luxembourgish"],
101        [152,"Macedonian"],[153,"Malagasy"],[154,"Malay"],[155,"Malayalam"],[156,"Maltese"],
102        [157,"Maori"],[158,"Marathi"],[159,"Mongolian"],[160,"Myanmar"],[161,"Nepali"],

```

```

103     [81,"Sindhi"],[82," Sinhala"],[83," Slovak"],[84," Slovenian"],[85,"Somali"],
104
105     [86," Spanish"],[87," Sundanese"],[88," Swahili"],[89," Swedish"],[90," Tagalog"],
106
107     [91,"Tajik"],[92," Tamil"],[93," Tatar"],[94," Telugu"],[95," Thai"],[96," Turkish"],
108
109     [97,"Turkmen"],[98," Ukrainian"],[99," Urdu"],[100," Uyghur"],[101,"Uzbek"],
110
111     [102," Vietnamese"],[103," Welsh"],[104," Xhosa"],[105," Yiddish"],[106," Yoruba"],[107," Zulu"]]
112
113 # Arranging the languages in tabular form
114
115 print(tabulate(l1ist,headers=["S.No","Languages"],tablefmt="fancy_grid"))
116 print("")
117 print("-----")
118
119 # Getting user's choice for translation language
120
121 lang = input("To which language do you want to TRANSLATE this text : ")
122 print("-----")
123 print("")
124 print("")
125
126 # Translating the text from English to user's preferred language
127
128 to_translate = mytext
129 translated = GoogleTranslator(source="auto",target=lang).translate(to_translate)
130
131 # Printing the translated text
132
133 print("-----")
134 print(" The translated text from this PDF : ")
135 print("-----")
136 print("")

```

```

137 print(translated)
138 print("")
139 print("")
140
141 # Getting the language code from the language chosen by the user
142
143 languages= {"Afrikaans": "af", "Albanian" : "sq", "Amharic" : "am", "Arabic" : "ar",
144
145             "Armenian": "hy", "Azerbaijani": "az", "Basque": "eu", "Belarusian": "be",
146
147             "Bengali": "bn", "Bosnian": "bs", "Bulgarian": "bg", "Catalan": "ca",
148
149             "Cebuano": "ceb", "Chinese": "zh", "Corsican": "co", "Croatian": "hr",
150
151             "Czech": "cs", "Danish": "da", "Dutch": "nl", "English": "en", "Esperanto": "eo",
152
153             "Estonian": "et", "Finnish": "fi", "French": "fr", "Galician": "gl", "Georgian": "ka",
154
155             "German": "de", "Greek": "el", "Gujarati": "gu", "Haitian Creole": "ht", "Hausa": "ha",
156
157             "Hawaiian": "haw", "Hebrew": "he", "Hindi": "hi", "Hmong": "hmn", "Hungarian": "hu",
158
159             "Icelandic": "is", "Igbo": "ig", "Indonesian": "id", "Irish": "ga", "Italian": "it",
160
161             "Japanese": "ja", "Javanese": "jv", "Kannada": "kn", "Kazakh": "kk", "Khmer": "km",
162
163             "Kinyarwanda": "rw", "Korean": "ko", "Kurdish": "ku", "Kyrgyz": "ky", "Lao": "lo",
164
165             "Latvian": "lv", "Lithuanian": "lt", "Luxembourgish": "lb", "Macedonian": "mk",
166
167             "Malagasy": "mg", "Malay": "ms", "Malayalam": "ml", "Maltese": "mt", "Maori": "mi",
168
169             "Marathi": "mr", "Mongolian": "mn", "Myanmar": "my", "Nepali": "ne", "Norwegian": "no",
170 |

```

```

171         "Nyanja": "ny", "Odia": "or", "Pashto": "ps", "Persian": "fa", "Polish": "pl",
172
173         "Portuguese": "pt", "Punjabi": "pa", "Romanian": "ro", "Russian": "ru", "Samoan": "sm",
174
175         "Scots Gaelic": "gd", "Serbian": "sr", "Sesotho": "st", "Shona": "sn", "Sindhi": "sd",
176
177         "Sinhala": "si", "Slovak": "sk", "Slovenian": "sl", "Somali": "so", "Spanish": "es",
178
179         "Sundanese": "su", "Swahili": "sw", "Swedish": "sv", "Tagalog": "tl", "Tajik": "tg",
180
181         "Tamil": "ta", "Tatar": "tt", "Telugu": "te", "Thai": "th", "Turkish": "tr", "Turkmen": "tk",
182
183         "Ukrainian": "uk", "Urdu": "ur", "Uyghur": "ug", "Uzbek": "ug", "Vietnamese": "vi",
184
185         "Welsh": "cy", "Xhosa": "xh", "Yiddish": "yi", "Yoruba": "yo", "Zulu": "zu"}
186
187 langcode = languages.get(lang.title())
188
189 # Getting the file name for saving the audio file
190
191 print("-" * 56)
192 taudio = input(" Enter the Name in which the audio file should be saved : ")
193 tfilename = taudio+".mp3"
194 print("-" * 56)
195 print("")
196 print(" Your Audio file is downloading, please wait till the audio player pops up")
197 print("-----")
198
199 # converting the translated text to speech and saving it
200
201 text_speech = gTTS(text=translated, lang=langcode)
202 text_speech.save(tfilename)
203 print("")

```

```

204
205 print("NOTE :")
206 print("")
207 print("    Click on load and play ")
208 print("-----")
209 print("")
210 print("")
211 print("Enjoy the Audiobook!!!")
212
213
214 # The below space is for the audio player
215
216 print("""
217
218
219
220
221
222
223 """)
224 print("-----")
225 print(" Thank you! Visit Again!! keeping gaining knowledge!!!")
226 print("-----")
227
228 # Building an audio player for playing the saved file
229
230 class MusicPlayer:
231
232 # Defining the geometry of the Music Player
233
234     def __init__(self, window ):
235         window.geometry('430x80'); window.title(tfilename); window.resizable(0,0)
236         Load = Button(window, text = 'Load',  width = 10, font = ('Times', 10), command = self.load)
237         Play = Button(window, text = 'Play',  width = 10,font = ('Times', 10), command = self.play)|

```



```

238     Pause = Button(window,text = 'Pause', width = 10, font = ('Times', 10), command = self.pause)
239     Stop = Button(window ,text = 'Stop', width = 10, font = ('Times', 10), command = self.stop)
240     Load.place(x=0,y=20);Play.place(x=110,y=20);Pause.place(x=220,y=20);Stop.place(x=330,y=20)
241     self.music_file = False
242     self.playing_state = False
243
244 # Defining the functionality of the Music player
245
246 def load(self):
247     self.music_file = filedialog.askopenfilename()
248 def play(self):
249     if self.music_file:
250         mixer.init()
251         mixer.music.load(self.music_file)
252         mixer.music.play()
253 def pause(self):
254     if not self.playing_state:
255         mixer.music.pause()
256         self.playing_state=True
257     else:
258         mixer.music.unpause()
259         self.playing_state = False
260 def stop(self):
261     mixer.music.stop()
262
263 # Calling the function
264
265 root = Tk()
266 app= MusicPlayer(root)
267 root.mainloop()
268
269 #
270

```

Output



IDLE Shell 3.8.10

File Edit Shell Debug Options Window Help

```
-----  
Hey user! Welcome to the application of Multilingual Audiobooks Generator!!  
-----
```

```
Please enter the Name of the PDF : the-lion-and-the-rabbit
```

```
-----  
The total number of pages in " the-lion-and-the-rabbit.pdf "is :  2  
-----
```

```
-----  
The text in this PDF is :-  
-----
```

The Lion and the Rabbit

Once there was a Lion in the jungle who used to kill 2-3 animals daily for his meal. All animals went to him to tell, that daily one of them will come to him for his meal.

So, the Lion agreed and this started going for many days. One day, it was Rabbit's turn. When he was on his way he saw a well.

Now he plans to kill the lion and save himself.

He went to lion and told him that, there is another lion who claims to be more powerful than him. Then lion asks the rabbit to take him to that lion. The rabbit takes him to the well and said he lives here. When the lion looked in the well he saw his own reflection and jumped in the well and dies.

Moral: Wisdom wins might.

The list of languages for translation includes :-

S.No	Languages
1	Afrikaans
2	Albanian
3	Amharic
4	Arabic
5	Armenian
6	Azerbaijani
7	Basque
8	Belarusian
9	Bengali
10	Bosnian
11	Bulgarian
12	Catalan
13	Cebuano
14	Chinese

15	Corsican
16	Croatian
17	Czech
18	Danish
19	Dutch >
20	English
21	Esperanto
22	Estonian
23	Finnish
24	French
25	Frisian
26	Galician
27	Georgian
28	German
29	Greek
30	Gujarati
31	Hausa

32	Haitian Creole
33	Hawaiian
34	Hebrew
35	Hindi
36	Hmong
37	Hungarian
38	Icelandic
39	Igbo
40	Indonesian
41	Irish
42	Italian
43	Japanese
44	Javanese
45	Kannada
46	Kazakh
47	Khmer
48	Kinyarwanda

49	Korean	
50	Kurdish	
51	Kyrgyz	
52	Lao	
53	Latvian	
54	Lithuanian	
55	Luxembourgish	
56	Macedonian	
57	Malagasy	
58	Malay	
59	Malayalam	
60	Maltese	
61	Maori	
62	Marathi	
63	Mongolian	
64	Myanmar	
65	Nepali	

66	Norwegian
67	Nyanja
68	Odia
69	Pashto
70	Persian
71	Polish
72	Portuguese
73	Punjabi
74	Romanian
75	Russian
76	Samoan
77	Scots Gaelic
78	Serbian
79	Sesotho
80	Shona
81	Sindhi
82	Sinhala

83	Slovak
84	Slovenian
85	Somali
86	Spanish
87	Sundanese
88	Swahili
89	Swedish
90	Tagalog
91	Tajik
92	Tamil
93	Tatar
94	Telugu
95	Thai
96	Turkish
97	Turkmen
98	Ukrainian
99	Urdu

100	Uyghur
101	Uzbek
102	Vietnamese
103	Welsh
104	Xhosa
105	Yiddish
106	Yoruba
107	Zulu

To which language do you want to TRANSLATE this text : `tamil`

The translated text from this PDF :

சிங்கம் மற்றும் மயல்

ஒருமூறை காட்டில் ஒரு சிங்கம் இருந்தது, அது தனது உணவிற்காக தினமும் 2-3 விலங்குகளை கொன்றது. அனாதை விலங்குகள் அவர்களில் ஒருவர் தினமும் அவரிடம் உணவாக்காக வருவார் என்று சொல்ல அவரிடம் சென்றார்.

எனவே, சிங்கம் ஒப்புக்கொண்டது, இது பல நாட்கள் நடக்கத் தொடங்கியது. ஒரு நாள், மயலின் மூறை வந்தது. எப்பொழுது அவர் வழியில் ஒரு கிணற்றைக் கண்டார்.

இப்போது சிங்கத்தைக் கொண்டு தன்னைக் காப்பாற்றிக் கொள்ளத் திட்டமிடுகிறான்.

அவர் சிங்கத்திடம் சென்று, தன்னை விட சக்தி வாய்ந்ததாகக் கூறும் மற்றொரு சிங்கம் இருப்பதாகச் சொன்னார்.

அப்போது சிங்கம் மயலை அந்த சிங்கத்திடம் அழைத்துச் செல்லும்படி கேட்கிறது. மயல் அவனை கிணற்றுக்கு அழைத்துச் சென்று சொன்னது இங்கு வாழ்கிறார். சிங்கம் கிணற்றில் பார்த்தபோது தன் பிரதிபலிப்பைக் கண்டு கிணற்றில் காதித்தது இறக்கிறார்.

ஒழுக்கம்: ஞானம் பலத்தை வல்லும்.

.

Enter the Name in which the audio file should be saved : tamil_audi

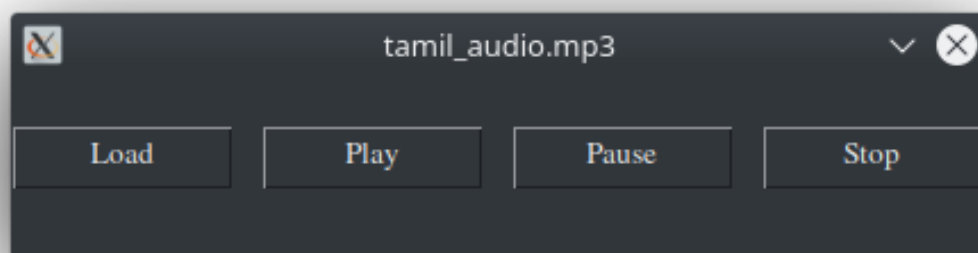
Enter the Name in which the audio file should be saved : tamil_audio

Your Audio file is downloading, please wait till the audio player pops up

NOTE :

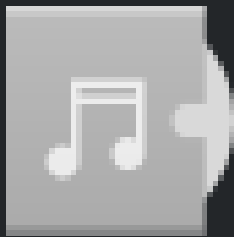
Click on load and play

Enjoy the Audiobook!!!



Thank you! Visit Again!! keeping gaining knowledge!!!

The Audio file will
be saved in your
File Manager



tamil_audio.mp3

245.5 KiB 09/03/22 12:11 AM

Bibliography

-> Textbooks

- Computer Science with Python
Class XII

-> Websites for reference

- <https://cloud.google.com/speech-to-text/docs/languages>
- <https://www.geeksforgeeks.org/language-translator-using-google-api-in-python/>
- <https://dev.to/kalebu/make-your-own-python-music-player-44n>

