Indian Institute of Information and Technology Allahabad

Project on

Translator of Indian local languages

Under guidance

Dr.k.p. Singh

By

Shivarajkumar G

MIT2021042

Contents

1) Introduction

Language Translation is the process of conveying a written source language text in a target language in a clear, comprehensive, accurate, and acceptable manner. Information can be transferred across languages, increasing its accessibility. Translation, editing, and proofreading are all part of the translation process.

Translation is a product, a service, and a method. Translation is a difficult job that necessitates careful reading of a text in the source language, comprehension of its meaning, and creation of an equivalent text in the target language. The term" translation" also refers to the final target language text that will be published or distributed as a result of this operation.

Although we will discuss these characteristics of translation in this document, we will concentrate on the process of making a translation using this definition. Translation is now more prevalent and accessible than it has ever been. Organizations with greater resources may choose to hire a translation business or freelance professional translators to handle all of their translation needs.

In this project I am going to Implement language translator where it can accept audio or text or image as input of one language and convert to other language in the form of text or audio. i am trying for Indian languages Telugu to Kannada and vice versa where these languages have less resources. later i want to extend this to other Indian languages.

2) Problem statement

Translation of one language contents to another language in the form of different formats like audio, text and image.so these translators will helpful to learn languages and it remove the dependency on language. Mainly in this project concentrating on taking input of source language as audio and produce translated output in audio format.

3) Objective

The language is a major communication barrier. there are many languages in India. and due less resource constraint and domination of English these languages are losing its identity and speakers. so, if we are removing a dependency on one language like English by language translators where these can convert any language to any other language. the speaker will not feel discomfort to talk in his native language and if needed he convert to any other language. so, it will help to save our Indian languages.

4) Methodology

Algorithm

Step-1: - Accepting the source language audio input.

Step-2: - Convert the audio format to text format of source language.

Step-3: - Translate text of source language to intermediate language English. Because English has huge resources related to any other language in the world.

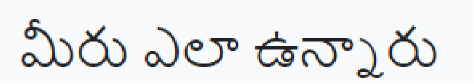
Step-4: - Translate English text to destination language text.

Step-5: - Translate destination language text to destination language audio file.

Example: -

Audio input of kannada

-→ ಹೇಗಿದ್ದೀರ? → how are you? →



మీరు ఎలా ఉన్నారు

Fig: Translating kannada to Telugu.

5) Technical Implementation

1) Audio Input

In this project i am taking input from mic. because i am planning this project work like real time application. where it can accept voice of someone when they are speaking and it can translate that voice to destination language.

To accept audio input i am using following python libraries.

1)Pyaudio

2)Speech Recognition

2) text output

source language: - where the input of a project is audio so at beginning i am converting audio file to text string then using google translator API translating text string to desired source language then i am printing the source language in the form of text in the same format of the source language.

destination language: - the text output will be printing for destination language. for this I am using source language text and then with the help of translator function in google API I am translating source language to destination language.

3) Audio output

The main objective of this project is to generate audio output of destination language. where the speaker is speaking in his native language that will be converted to chosen destination language. and the output will be produced in the form of audio.

python packages used to get audio output.

1)win sound

2)gtts

3)os

4) google API

Google Translate is a multilingual neural machine translation service developed by Google to translate text.an API that helps developers build browser extensions and software applications. The google translate API will be convert from one language to another language with the help of English as intermediate language. so, at the first it will convert source language to English then it will convert English to destination language.

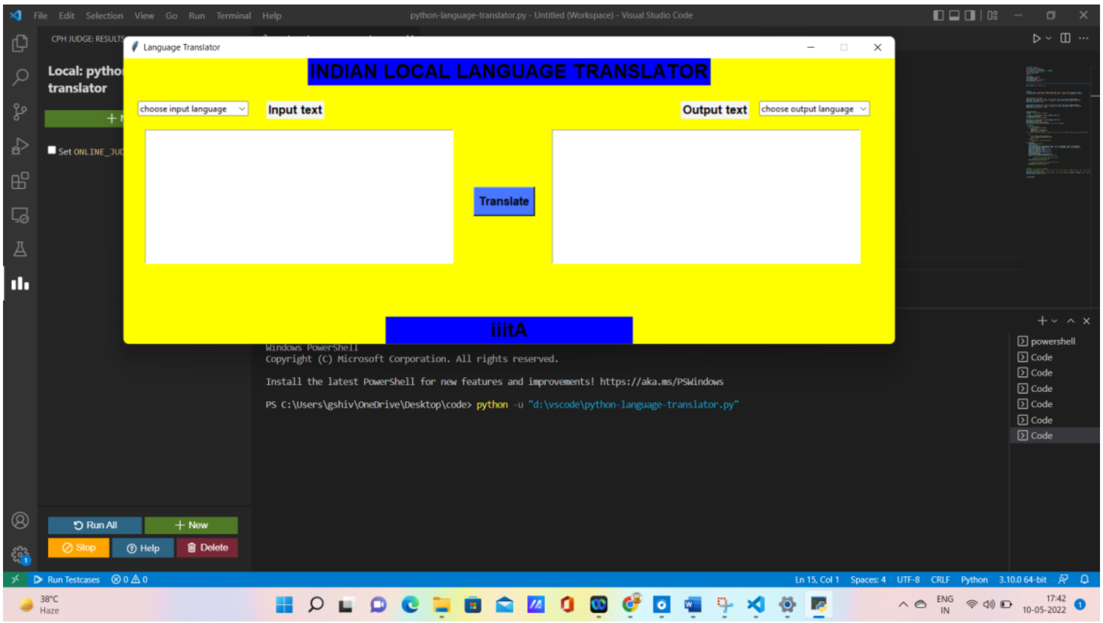
6) Code Implementation

```
from tkinter import *
from tkinter import ttk
from winsound import PlaySound
from googletrans import Translator , LANGUAGES
import speech_recognition as sr
from gtts import gTTS
import os

root = Tk()
root.geometry('1080x400')
root.resizable(0,0)
root.title("Language Translator")
root.config(bg = 'yellow')



dic=('kannada','kn','telugu','te')




#heading
Label(root, text = "INDIAN LOCAL LANGUAGE TRANSLATOR", font = "arial 20 bold",
bg='blue').pack()
Label(root,text ="iiitA", font = 'arial 20 bold', bg ='blue' , width =
'20').pack(side = 'bottom')



#INPUT AND OUTPUT TEXT WIDGET
Label(root,text ="Input text", font = 'arial 13 bold', bg ='white
smoke').place(x=200,y=60)
Input_text = Text(root,font = 'arial 10', height = 11, wrap = WORD, padx=5,
pady=5, width = 60)
Input_text.place(x=30,y = 100)
```

```
Label(root,text ="Output text", font = 'arial 13 bold', bg ='white
smoke').place(x=780,y=60)
Output_text = Text(root,font = 'arial 10', height = 11, wrap = WORD, padx=5,
pady= 5, width =60)
Output_text.place(x = 600 , y = 100)



##################
language = list({'kannada','telugu'})

src_lang = ttk.Combobox(root, values= language, width =22)
src_lang.place(x=20,y=60)
src_lang.set('choose input language')

dest_lang = ttk.Combobox(root, values= language, width =22)
dest_lang.place(x=890,y=60)
dest_lang.set('choose output language')
########################################  Define function #######

def Record():
    r = sr.Recognizer()
    with sr.Microphone() as source:
        print("Talk")
        audio_text = r.listen(source)
        print("Time over, thanks")
# recoginize_() method will throw a request error if the API is unreachable,
hence using exception handling

    try:
        # using google speech recognition
        text = r.recognize_google(audio_text)
        return text
    except:
         print("Sorry, I did not get that")

def Translate():
    translator = Translator()
    stext = Record()
    stranslate=translator.translate(text= stext , src = src_lang.get(), dest =
src_lang.get())
    translated=translator.translate(text= stext , src = src_lang.get(), dest =
dest_lang.get())
    Input_text.delete(1.0, END)
    Input_text.insert(END, stranslate.text)
    Output_text.delete(1.0, END)
    Output_text.insert(END, translated.text)
    text=translated.text
    to_lang=dic[dic.index(dest_lang.get())+1]
```

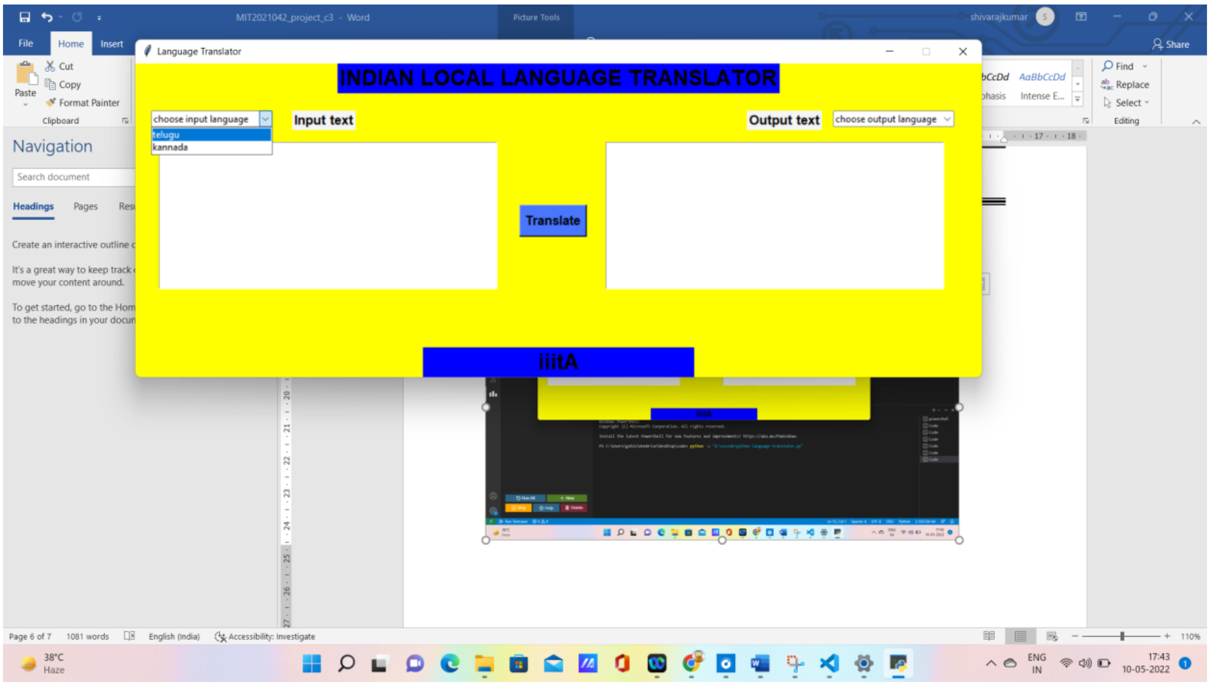
```
  speak = gTTS(text=text, lang=to_lang, slow= False)

      # Using save() method to save the translated
      # speech in capture_voice.mp3
  speak.save("captured_voice.mp3")

      # Using OS module to run the translated voice.
  os.system("start captured_voice.mp3")
 #PlaySound('captured_voice.mp3')



##########  Translate Button ########
#record_btn = Button(root, text = 'record',font = 'arial 12 bold',pady =
5,command = Record , bg = 'royal blue1', activebackground = 'sky blue')
#record_btn.place(x=250, y=100)
trans_btn = Button(root, text = 'Translate',font = 'arial 12 bold',pady =
5,command = Translate , bg = 'royal blue1', activebackground = 'sky blue')
trans_btn.place(x = 490, y = 180)


root.mainloop()
```

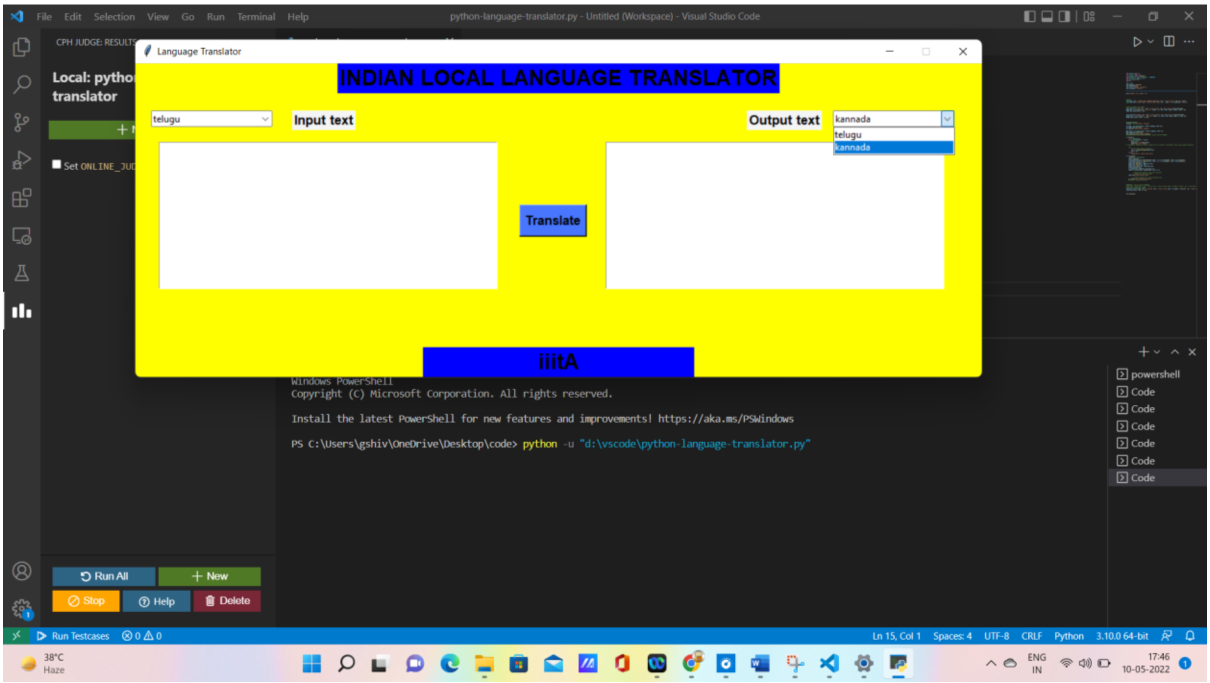
7) Output Screenshots: -



```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\gshiv\OneDrive\Desktop\code> python -u "d:\vscode\python-language-translator.py"
```

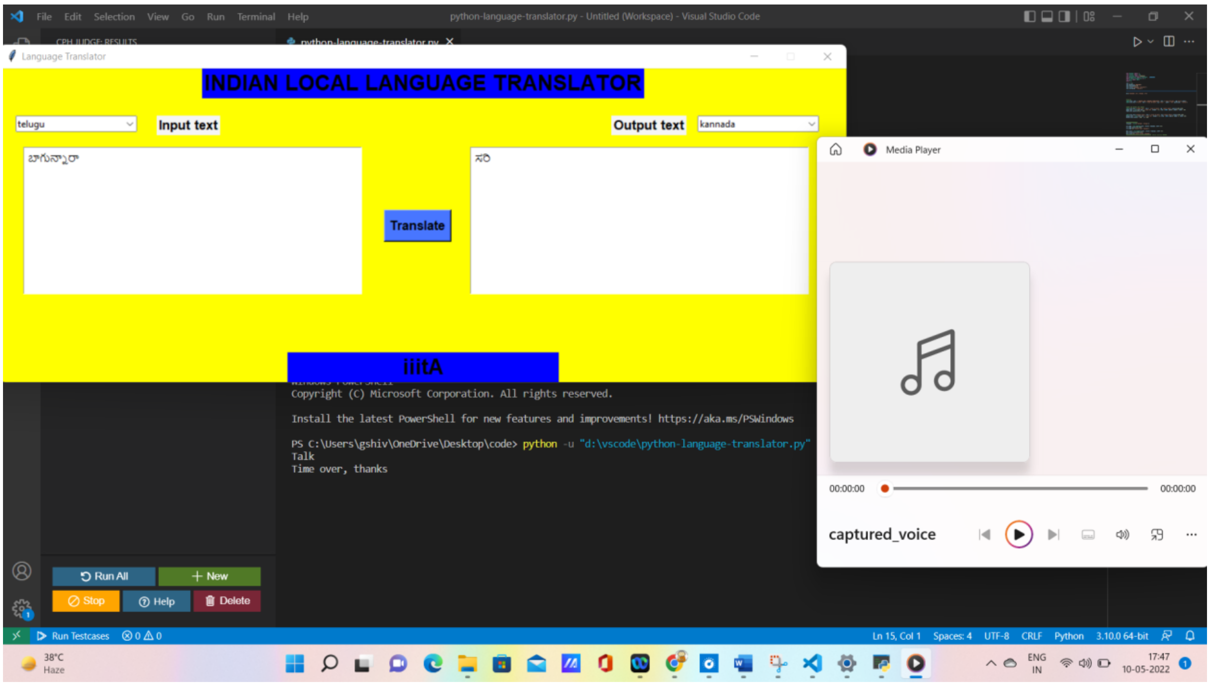
Fig: - main interface





```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\gshiv\OneDrive\Desktop\code> python -u "d:\vscode\python-language-translator.py"
```

Fig: - selection of languages



python-language-translator.py - Untitled (Workspace) - Visual Studio Code
Language Translator
INDIAN LOCAL LANGUAGE TRANSLATOR
Translate
iiitA
Media Player
Run All
New
Stop
Help
Delete

telugu Input text

బాగున్నారా

Output text kannada

ಸರಿ

```
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\gshiv\OneDrive\Desktop\code> python -u "d:\vscode\python-language-translator.py"
Talk
Time over, thanks
```

captured_voice 00:00:00 00:00:00

Fig: - output

8) Drawbacks of current work: -

In above algorithm I am using intermediate language as English so it is leading to a less accuracy. and it creating dependency of language.

9) Future work: -

To increase the accuracy of model proposed below algorithm.

Algorithm

Step 1: - maintaining dictionaries of words for source and destination language.

Step 2: - taking a dataset of both the languages which contains large number of sentences of the languages.

Step-3: - breaking the words of source language and find appropriate words of destination language.

Step-4: - by performing some pattern matching identify proper sentence from the dataset as output.

10) Conclusion

In present project translation will going smoothly but due to usage of English it has less accuracy. to improve the model in future the proposed new algorithm will be implemented.