INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Text Interpreter and Converter

Arpitha Vasudev[1], Karthik M R[2], K. Pranav[3], Monsih[4], Sai Davya[5]
[1]Professor, [2, 3, 4, 5]B.Tech CSE Department, DSATM Bengaluru

Abstract: With the exponential growth in data, it is essential to analyse it. The data is a valuable source of information and knowledge that should be effectively interpreted to be helpful. The data should be interpreted in different forms based on the user requirements to get precise information and knowledge. Nowadays smartphones are the most commonly used electronic device. A smartphone is not only a communication device, it is also a powerful computing device. So it is possible to apply translation, text extraction, summary, and much more techniques, which require much computational work. This paper presents an application to analyse the text from documents on smartphones. However, it is challenging to interpret the documents on smartphones. The proposed application converts the documents or images to searchable and editable digital text, and further, it can be used to analyse them into different forms. The objective of this application is four-fold 1) To recognize text from documents or images by using optical Character Recognition 2) Summarization of the text 3) Translate the extracted text to different languages 4) Generation of speech from the text by using a text-to-speech algorithm.
Keywords: Optical Character Recognition, Summarization, Translate, text-to-speech

I. INTRODUCTION

The application accepts documents as input and generates a text file by using tesseract OCR [1], that extracts the text from the document. This text file is uploaded to cloud storage. The text file can be accessed to summarize, translate, and generate speech. The text recognition module uses Tesseract OCR to extract text. Initially, it converts the color image to a binary image, separates the characters, extracts the information from the images, and finally, it does post-processing.This application uses extractive summarization to summarize the text. Extractive summarization extracts the most important subset of information of the sentences from the text to generate a summary. The machine learning model uses the Text-Rank algorithm [2]. Text translation analyses the structure of sentences, syntax and grammar in the source language and, based on the grammar and rules of the target language, it translates the text. This application uses the Firebase ML kit's API [3] to translate the text into different languages. A Text-to-speech contains two parts. First, it does normalization, converts the text into tokens, and then assigns phonetic transcriptions and prosodic units, then they are converted to the waveform, and by using a synthesizer, the speech is generated. The text file can be accessed from cloud storage and used to generate speech with the help of Text-to-speech modules which use android text-to-speech API [4].

II. RELATED WORK

OCR, Text translation, Text-to-speech, and Text summarization are the key technologies used to interpret the text in different forms for better understanding. Some applications use these technologies for the interpretation of thetext.[5] This paper presents the web app that extracts text from the images and uses the real-time OCR for extraction. This application mainly extracts text from an image. The extracted text can be editable. It is not capable of extracting text from a document. It is not able to store the extracted text. [6] proposed an android application that extracts text from the images, and the extracted text can be stored in the local storage. It cannot extract the text from the document, and the text cannot be interpreted into other forms. The text is stored in local storage, and there may be a chance of data loss when the app is uninstalled or if the device is crashed. [7], [8] These papers have researched an android application capable of live translation of the text by accessing each frame from the device's camera. This application is bounded to a small set of users willing to read the text in their local language. It directly performs translation of a given frame to the local language. [9] this paper has researched automatic text summarization. The summarizer used in this reference paper uses a sentence ranking algorithm that accepts text from the document, which is used to generate a summary. The summarized text can be converted into audio format.[10] this paper tellsus about an android application that converts the text from the image to speech by translating it into another language. The text is extracted using OCR, and the text is translated, and the translated text is converted to speech. This application does not have any storage facility. Moreover, it works only on the image, not on any document. [11] this paper researched OCR in Android OS. This paper mainly focuses on OCR, which works using Tesseractfor Android. [12] this paper presents the desktop application that can extract the text using Tesseract OCR.

[13]It also supports operations on the text like translation, text to speech. It is also capable of speech recognition. This application is bounded to only desktops. All the above-mentioned papers use the technologies to extract the text, and some use them to interpret the text, but all together is not embedded in a single application.

We proposed an android application "Text Interpreter & Converter", accepts the document as input, extracts the text from documents, and converts it into digital editable and searchable text. This text is stored in the cloud. Then, the text file can be summarized, translated, converted to speech. Here we embedded all the above existing technologies in an android application.

III. PROPOSED WORK

In this paper, we proposed to create an android application. This application consists of mainly five modules. Each module has its functionality. The five modules are Text recognition, Text to speech, Text translator, Textsummarizer, and Database & Authentication.

A. Text Recognition

Many open-source OCR engines are ready to use, but there are fewer OCR engines for mobile because mobile has less power and less capacity to process. Among the fewer OCR engines, Tesseract is one of them. Coming to android applications mobile, Tesseract OCR [1] provides an open-source library for android, which is helpful for text extraction from scanned images. In this Application, the tesseract engine is provided English trained data.

Method: Firstly, the Application accepts images or documents (the documents should be in pdf format). After getting the document or image then, the text extraction takes place. If the input is a document, the record is rendered first through a PDF Renderer while rendering each page of a given document is copied into a bitmap, and the bitmap is passed for text extraction. All set of pages in the given document follows the same procedure. If the input is an image, the image is converted into a bitmap, which is passed for text extraction.After collecting the bitmap from the preceding process for text extraction, the bitmap is passed to the Tesseract engine, which has some trained English data. The tesseract engine accepts bitmap, and it does all pre-processing operations like grayscale conversion, binarization, and text segmentation. Moreover, after performing all the pre-processing operations, it conducts a text recognition operation through which it extracts the characters present in a bitmap. Finally, the output given by the Tesseract is appended to the text editor through which the user can edit and perform other operations.

B. Text-To-Speech

Text to speech is one of the unique modules through which users can listen to text converted to speech or audio. Text-to-speech is mainly helpful for people who are visually impaired. Using Text-to-speech, the user can listen to the content of their document, which is readen out by the Text-to-speech engine. For Text-to-speech, android's text- to-speech API is used.

Method: The Text-to-speech module is initialized by the android Text-to-speech (TTS) API. While initializing, the TTS (Text-to-speech module) is assigned with certain voices and other parameters. After initializing, the text is accessed, the text is given to the TTS API as input, and the TTS API generates output as speech. Here user can select a male/female voice and stop the speech in between.

C. Text Translator

ML Kit's on-device translation API [3] is used for text translation. ML Kit is a mobile SDK that helps access Google's machine learning expertise in mobile applications. By using on-device translation, the translations are performed quickly. Because of the complexity of Natural Language Processing (NLP), the translations may not be appropriate for all users. The primary source language is English. The text can be translated into different languages provided. Text translator is mainly helpful for users who are comfortable reading in their local language instead of English. For every other language translation, there are different machine learning models to translate. If the user chooses the language for the first time, the model is downloaded. If the user has already translated to that language, the downloaded model is used for translation. Every model has a size of 30 MB.

Method: In Text translator module, the translator is initialized. After initializing, the text is provided as input to the translator. When the user selects a target language, the translator translates the text into the target language and displays it to the user.

D. Text Summarizer

There are mainly two types of summarizers: abstractive summarizers and extractive summarizers. An abstractive summarizer helps understand the core content of the original text.

An extractive summarizer helps in extracting the most critical subset of sentences. An extractive summarizer is quicker to implement using unsupervised learning as it requires no training data. The similarities among the sentences are calculated, and the summary is generated based on the maximum similarity score. The summarizer uses the Text-Rank algorithm [2], to generate a summary in this Application. We deployed the summarization model on the Heroku cloud by using flask framework.

Steps in Text-Rank:

1) Extract sentences from the original text.
2) Creating vectors for all text units (sentences) based on the words (tokens) present in the original text.
3) Calculating cosine similarity between each pair of the sentence.
4) Creating an n*n matrix where n represents the number of sentences.
5) Creating a graph using the similarity matrix where each vertex is a sentence, and the edge is similarity.
6) Ranking the text units based on similarity score and returning the top n sentences that are to be included in the summarized text.
7) Cosine distance between two vectors is calculated using the Cosine of the angle between them.

Cosine Distance

$$(va, vb) = 1 - Cosi(Angle\ between\ va, vb)$$

We can say that for the same vectors, the Cosine distance will be 0, and the Cosine similarity will be 1 for perpendicular vectors.

Method: In the Text summarization module, the text is accessed from cloud. The accessed text and the number of lines to be summarized are the two parameters taken as input for the post request. These post requests are handled and forwarded to the text summarizer. This summarizer generates the summary. The summary is converted to JSON format. The JSON object is dealt with in the Application. Finally, the summary of the original text is displayed to the user. The Application works with all the modules mentioned above as its main features. In addition, database and Authentication services are also integrated with the Application. When tesseract OCR extracts the text, the extracted text is stored in a database (Cloud Storage) in the text file. This text file is accessed when Text Translator, Text Summarizer, Text-to-speech modules are invoked. So to keep track of users and display user's files Authentication Service is also required. For Authentication, users can log in using their phone number, Email Id, or Google Account. Above mentioned are the three authentication service providers provided in this Application.

For Database & Authentication, we have used Firebase. For the database, we have used Firebase Firestore. For the storage of text files, we have used Firebase Cloud Storage. Finally, for Authentication, we have used Firebase Authentication. The user must register using an authentication service to use this Application.

E. Working of Application

When a user opens the Application if the user is a registered user, all the files of that particular user are displayed. If the user is not registered, the user needs to be registered using any particular service provider (phone number, Email, or Google) provided by the Application. Then, users can upload documents or images by using the text recognition module, which helps extract characters from the documents or images. After extraction of text, the user can edit the data, and by clicking on the save button, the text file is uploaded to the cloud storage and database. Finally, the uploaded file uploaded is displayed in the main activity.

The user can click on any of the files which he/she are desired to translate or summarize or listen to it. In all the other modules like Text translation, Text Summarization, and Text-to-speech, the text is accessed from the text file stored in cloud storage.
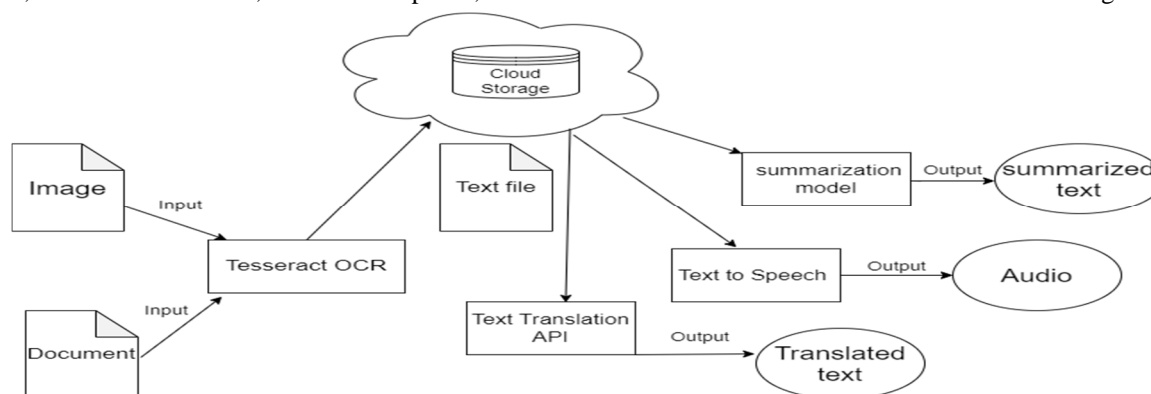

FIGURE.1. System Architecture

IV. RESULTS

The Text recognition module in this application uses the Tesseract OCR library. The existing text recognition models have an accuracy of 89.03%, whereas the proposed tesseract OCR has 90%. To achieve maximum accuracy for an image, the image should be ideal. The ideal conditions depend on three factors, they are

1) Background of an image
2) Font size of an image
3) Font style of an image

The summarizer model uses an algorithm called the TextRank algorithm. In the TextRank algorithm, the frequency of the sentences is compared with the whole given text. The accuracy of this model is around 67%. For further information, refer to [9].

The Text Translation module of this application uses ML kit's on-device text translation API, which uses Google vision The accuracy of the Google vision translate API is 85%.

Text to Speech API accepts only 4000 characters to convert into speech so to handle that for large text the text is divided   into chunks, each chunk consists of 4000 characters and each chunk is passed to the Text to speech engine. The optimized approach can be developed for this solution.

V. CONCLUSION

This paper introduced "Text Interpreter & Converter," an android application for analyzing lengthy text. The volume of text is a huge source of information that should be analyzed to extract useful information. This paper proposed an android application to recognize, summarize, translate, and convert text to speech. The proposed application extracted the text from documents, and the text can be summarized, translated, or converted to speech. The UI is so friendly that users can easily interact with it. By using the above-mentioned technologies, we are ableto interpret the text in different ways.

In this application, the source language is English. In the future, we can add different languages, where we can extract characters from local languages and translate, summarize and generate speech for that particular language

REFERENCES

[1]   "Android OCR Application Based on Tesseract", Codeproject.com, 2022. [Online]. Available: https://www.codeproject.com/Articles/1275580/Android-OCR-Application-Based-on-Tesseract. [Accessed:14- Feb- 2022].

[2]   "Automatic Text Summarization Using TextRank Algorithm", Analytics Vidhya, 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/. [Accessed:14- Feb- 2022].

[3]   "Translation | ML Kit | Google Developers", Google Developers, 2022. [Online]. Available: https://developers.google.com/ml-kit/language/translation. [Accessed: 14- Feb- 2022].

[4]   "TextToSpeech | Android Developers", Android Developers, 2022. [Online]. Available: https://developer.android .com/reference/android/ speech/tts/TextToSpeech. [Accessed: 14- Feb- 2022].

[5]   S. Dome and A. P. Sathe, "Optical Charater Recognition using Tesseract and Classification," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 153-158, doi: 10.1109/ESCI50559.2021.9397008.

[6]   S. Pattnaik, S. R. Laha, B. K. Pattanayak and B. C. Pattanaik, "A Framework to Detect Digital Text Using Android Based Smartphone," 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), 2021, pp. 1-6, doi: 10.1109/ODICON50556.2021.9428993.

[7]   S. Revathy and S. Nath, "Android Live Text Recognition and Translation Application using Tesseract," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 1259-1263, doi: 10.1109/ICICCS48265.2020.9120973.

[8]   Ravindra Bandal, Adesh Jadhav, Vitthal Kale., 2014, Mobile Camera Based Text Detection and Translation, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 03, Issue 01 (January 2014).

[9]   J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 International Conference on Data Science and Communication (IconDSC), 2019, pp. 1-3, doi: 10.1109/IconDSC.2019.8817040.

[10]   Ahuja, D., Amesar, J., Gurav, A., Sachdev, S., &Zope, V. (2018). Text Extraction and Translation from Image using ML in Android. International Journal of Innovative Research in Science, Engineering and Technology, 7(1), 176–179. https://doi.org/10.15680/IJIRSET.2018.0701028

[11]   Zaki, Muhammad &Zai, Sammer& Ansari, Muhammad &Zaki, Urooba. (2019). Development Of An Android App For Text Detection. Journal of Theoretical and Applied Information Technology. 97. 2485 - 2496.

[12]   Journal, I. J. E. R. T. (2016). IJERT-Multilingual Speech and Text Recognition and Translation using Image. International Journal of Engineering Research and Technology (IJERT)

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)