CAPSTONE
PROJECT REVIEW 3

IMAGE-TO-TEXT CONVERTER THAT DETECTS LANGUAGE AND PROVIDES A TRANSLATED SUMMARY.

Guide: Dr. Vasantha W B

Pratham Shah 19BCE2028
Harshvardhan Singh Gahlaut
19BCE2372

Introduction

In today's digital age, language translation and text extraction from images have become commonplace in many apps and software. However, our project goes beyond these functionalities by offering an additional feature: text summarization in the user's desired language. Many of us have encountered situations where we come across lengthy paragraphs on websites or documents, and we simply need a brief summary to save time or accommodate our busy schedules. With our platform, users can easily input text or upload an image of the text in any language. They can then select their desired target language for translation and also request a summary of the text's content, all with the assistance of cutting-edge NLP (Natural Language Processing) and Optical Character Recognition (OCR).

Our process is designed to be step-by-step and user-friendly, allowing users to choose which steps they want to skip. We provide a proper website with a sleek UI, powered by Flask technology, to ensure a seamless and efficient experience. Whether it's translating text, extracting text from images, or summarising content, our project offers a comprehensive solution that caters to the diverse language needs of our users. Say goodbye to language barriers and information overload - with our project, language translation and summarization are made easy and accessible to all.

Problem Statement

To create a website that does language detection, text translation and summarisation using NLP which can take input as text or as an image (using OCR) or as a PDF.

Motivation

The motivation behind our project is to address the challenges posed by language barriers and information overload in today's digital age. As the world becomes increasingly interconnected, the ability to communicate across languages has become essential. However, the vast amounts of information available online often present a significant obstacle to effective communication. Our project seeks to address this issue by providing a comprehensive solution that integrates language translation, text extraction from images, and text summarization capabilities.

We are motivated to create a tool that goes beyond the existing functionalities of language translation and text extraction from images by offering an additional feature of text summarization in the user's desired language. Our aim is to create a platform that is easy to use, efficient and accessible to everyone, regardless of language barriers. We are motivated to create a solution that can help individuals save time and accommodate their busy schedules by generating concise summaries of lengthy documents.

Objectives

We found that there exist many text translation tools as well as text summarisation tools. However, a tool that does both, is one that's not readily available. Furthermore, most systems that use translation and summarisation depend on user input as a text.
We aim at creating a system that does it all:
1. A user can upload an image, and using OCR, our system will detect the text. The user can also input a text if they wish. We will create website with a proper frontend and a friendly UI with all these options for ease of use.

2. Once the text is identified, our system will detect the language of the text. The user just has to select which language they want to translate and summarise the text to.

3. Once they select the target language, we will translate the text and summarise it using an NLP model using NLTK.

4. Furthermore, existing summarisation tools only summarise English text, however, we will summarise text of other languages as well.

STUDIO SHODWE

Literature Survey

2022 PRESENTATION

Attend, Translate and Summarize: An Efficient Method for Neural Cross-Lingual Summarization - Junnan Zhu, Yu Zhou, Jiajun Zhang, Chengqing Zong; 2021

Cross-lingual summarization aims at summarizing a document in one language (e.g., Chinese) into another language (e.g., English). In this paper, they propose a novel method inspired by the translation pattern in the process of obtaining a cross-lingual summary. They first attend to some words in the source text, then translate them into the target language, and summarize to get the final summary. Specifically, they first employ the encoder-decoder attention distribution to attend to the source words. Second, they present three strategies to acquire the translation probability, which helps obtain the translation candidates for each source word. Finally, each summary word is generated either from the neural distribution or from the translation candidates of source words. Experimental results on Chinese-to-English and English-to-Chinese summarization tasks have shown that our proposed method can significantly outperform the baselines, achieving comparable performance with the state-of-the-art.

Abstractive Arabic Text Summarization Based on Deep Learning

- YM Wazery, ME Saleh, A Alharbi, AA Ali; 2022

In this paper, an abstractive Arabic text summarization system is proposed, based on a sequence-to-sequence model. This model works through two components, encoder and decoder. Their aim is to develop the sequence-to-sequence model using several deep artificial neural networks to investigate which of them achieves the best performance. Different layers of Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (BiLSTM) have been used to develop the encoder and the decoder. In addition, the global attention mechanism has been used because it provides better results than the local attention mechanism. Furthermore, AraBERT preprocess has been applied in the data preprocessing stage that helps the model to understand the Arabic words and achieves state-of-the-art results. Moreover, a comparison between the skip-gram and the continuous bag of words (CBOW) word2Vec word embedding models has been made. They have built these models using the Keras library and run-on Google Colab Jupiter notebook to run seamlessly. Finally, the proposed system is evaluated through ROUGE-1, ROUGE-2, ROUGE-L, and BLEU evaluation metrics. The experimental results show that three layers of BiLSTM hidden states at the encoder achieve the best performance. In addition, the proposed system outperforms the other latest research studies. Also, the results show that abstractive summarization models that use the skip-gram word2Vec model outperform the models that use the CBOW word2Vec model.

Neural Abstractive Text Summarization with Sequence-to-Sequence Models

- T Shi, Y Keneshloo, N Ramakrishnan; 2021

In this article, they provide a comprehensive literature survey on different seq2seq models for abstractive text summarization from the viewpoint of network structures, training strategies, and summary generation algorithms. Several models were first proposed for language modeling and generation tasks, such as machine translation, and later applied to abstractive text summarization. Hence, they also provide a brief review of these models. As part of this survey, they also develop an open source library, namely, Neural Abstractive Text Summarizer (NATS) toolkit, for the abstractive text summarization. An extensive set of experiments have been conducted on the widely used CNN/Daily Mail dataset to examine the effectiveness of several different neural network components. Finally, they benchmark two models implemented in NATS on the two recently released datasets, namely, Newsroom and Bytecup.

STN-OCR: A single Neural Network for Text Detection and Text Recognition

- Christian Bartz, Haojin Yang, Christoph Meinel; 2017

Detecting and recognizing text in natural scene images is a challenging, yet not completely solved task. In re- cent years several new systems that try to solve at least one of the two sub-tasks (text detection and text recognition) have been proposed. In this paper they present STN-OCR, a step towards semi-supervised neural networks for scene text recognition, that can be optimized end-to-end. In contrast to most existing works that consist of multiple deep neural networks and several pre-processing steps we propose to use a single deep neural network that learns to detect and recognize text from natural images in a semi-supervised way. STN-OCR is a network that integrates and jointly learns a spatial transformer network, that can learn to detect text regions in an image, and a text recognition network that takes the identified text regions and recognizes their textual content. We investigate how our model behaves on a range of different tasks (detection and recognition of characters, and lines of text). Experimental results on public benchmark datasets show the ability of our model to handle a variety of different tasks, without substantial changes in its overall network structure.

High Performance Text Recognition Using a Hybrid Convolutional-LSTM Implementation

- TM Breuel; 2017

Optical character recognition (OCR) has made great progress in recent years due to the introduction of recognition engines based on recurrent neural networks, in particular the LSTM architecture. This paper describes a new, open-source line recognizer combining deep convolutional networks and LSTMs, implemented in PyTorch and using CUDA kernels for speed. Experimental results are given comparing the performance of different combinations of geometric normalization, 1D LSTM, deep convolutional networks, and 2D LSTM networks. An important result is that while deep hybrid networks without geometric text line normalization outperform 1D LSTM networks with geometric normalization, deep hybrid networks with geometric text line normalization still outperform all other networks. The best networks achieve a throughput of more than 100 lines per second and test set error rates on UW3 of 0.25%.

Deep Learning based Isolated Arabic Scene Character Recognition

- Saad Bin Ahmed, Saeeda Naz, Muhammad Imran Razzak, and Rubiyah Yousaf ; 2017

This paper offers a Arabic site/ scene text recognition with the help of deep learning classifier utilizing the the model Convolution Neural Networks (CNN). The authors use five different orientations with regard to a single occurrence of a character since the scene text data is slanted and skewed, allowing us to handle the widest range of variances. The training is designed with stride values of 1 and 2 and filter sizes of 3 x 3 and 5 × 5. The training of the network was done with different learning rates throughout the text classification phase. On the recognition of Arabic characters from segmented Arabic scene photos, our method showed encouraging results.

Efficient, Lexicon-Free OCR using Deep Learning

- Marcin Namysl, Iuliu Konya; 2019

Due to geometrical distortions, intricate backgrounds, and a wide range of fonts, optical character recognition (OCR) continues to be a difficult challenge in unrestricted contexts, such as natural landscapes. This paper provide a deep learning-based, artificially generated training set, and data augmentation system for segmentation-free OCR. With the use of enormous text corpora and more than 2000 fonts, we render synthetic training data. They also add geometric distortions to collected samples as well as a suggested data augmentation method called alpha-compositing with background textures to emulate text appearing in complex natural situations. The specified models employ a convolutional neural network encoder to extract features from text images. On examining the modelling abilities of convolutional and recurrent neural networks for the interactions between input items. On distorted text samples, the suggested OCR system outperforms the accuracy of top commercial and open-source engines.

Reward Learning for Efficient Reinforcement Learning in Extractive Document Summarisation

- Yang Gao, Christian M. Meyer, Mohsen Mesgar and Iryna Gurevych; 2019

Summarizing documents can be treated as a sequential decision-making problem that Reinforcement Learning (RL) algorithms can handle. Due to the vast search areas and the delayed rewards, the most common RL paradigm for summarization learns a cross-input strategy, which necessitates a lot of time, data, and parameter tuning. Although learning input-specific RL rules is a more effective alternative, it still relies on hand-crafted rewards, which are challenging to create and produce subpar results. So, the paper here proposes RELIS, which is a novel reinforcement learning paradigm uses Learning-to-Rank (L2R) algorithms to learn a reward function during training, and then uses this reward function to train an input-specific RL policy during testing. The paper provesthat RELIS ensures that using the right L2R and RL algorithms, summaries will be generated that are close to optimal. The paper also shows that the RELIS approach is capable of reduced training time by afactor of 2 orders as compared to other present day methodologies and approaches.

Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond

- Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çag˘lar Gu ̇lçehre, Bing Xiang; 2016

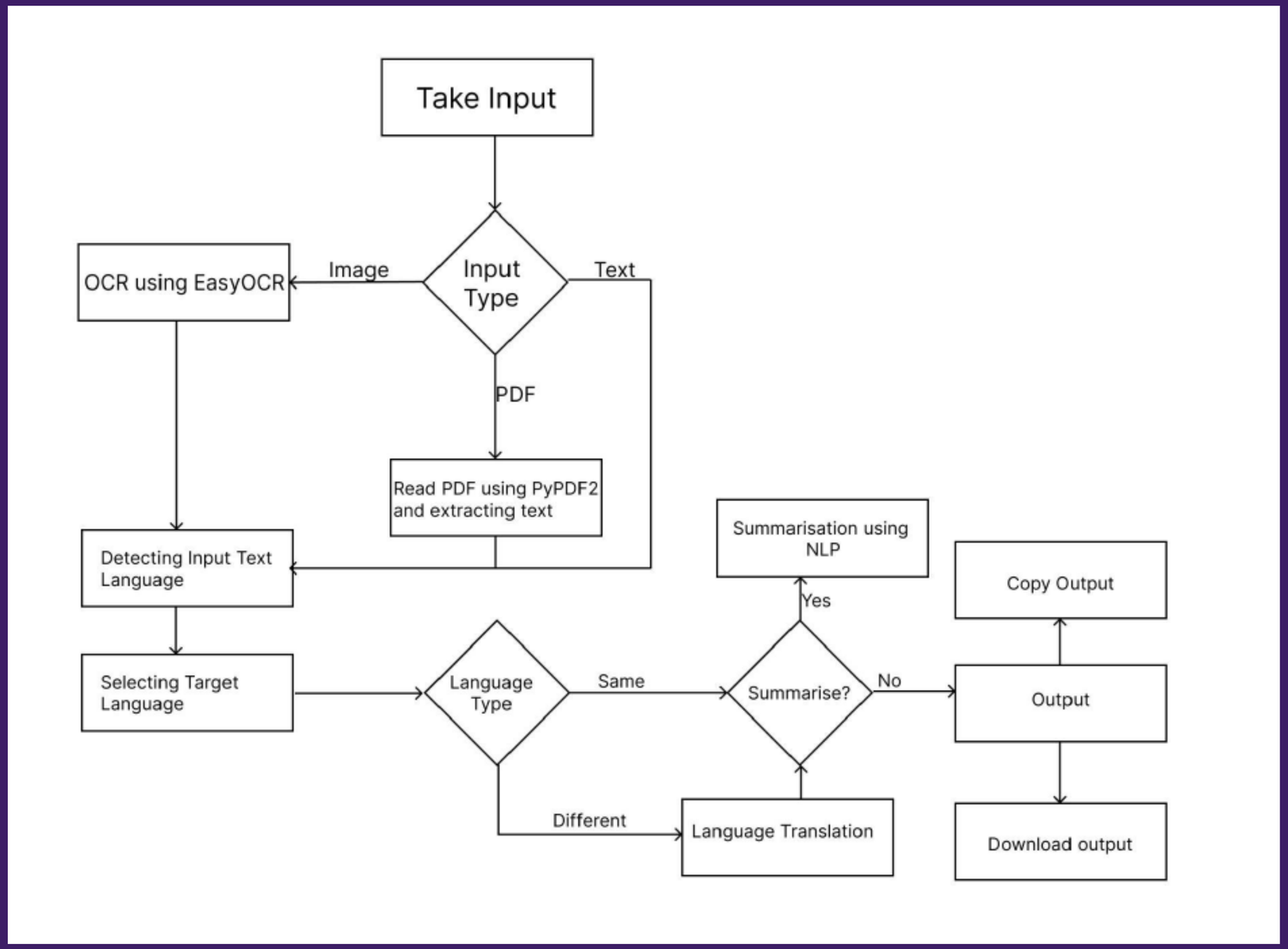
The work here proposes and develops a text summarization using Attentional Encoder- Decoder Recurrent Neural Networks, and show that they achieve state-of-the-art per- formance on two different corpora. The authors here propose a number of innovative models that handle important summarising issues that the basic architecture does not fully capture, like modelling essential words, capturing the hierarchy of sentence-to-word structure, and emitting words that are uncommon or unobserved during training. As for the conclusion, the paper states that many of the proposed models contribute to more improvements in the upcoming future. Along with this research, the paper also proposes a new dataset consisting of multi-sentence sum- maries, and establish performance bench- marks for further research.

Content Selection in Deep Learning Models of Summarization

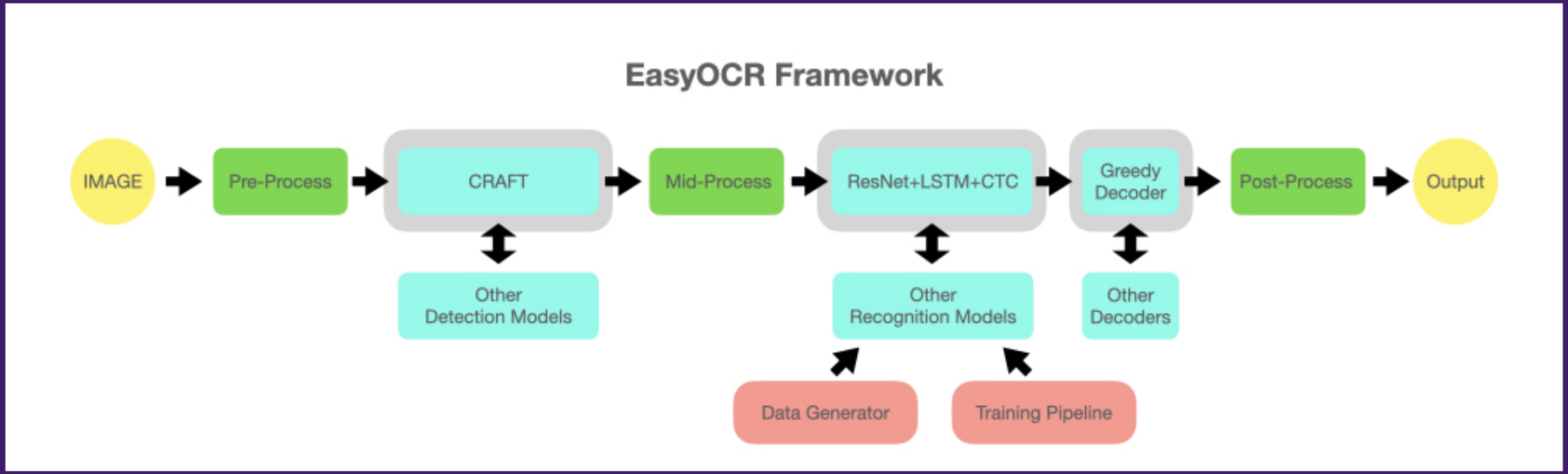
- Chris Kedzie, Kathleen McKeown, Hal Daume´ III; 2019

In order to comprehend how content selection is carried out, the paper conducts tests using deep learning models of summarization spanning the domains of news, personal stories, meetings, and medical literature. The authors discover that several highly sophisticated aspects of modern extractive summarizers do not outperform simpler models in terms of performance. These findings cast doubt on the value of deep learning methods for summary for those domains that do have large datasets and show that it is simpler to develop a summarizer for a new domain than past research suggests. Moreover, they suggest new types of sentence representations or external information sources are required that are better suited to the summarising task; these are significant questions for fresh research in summarization.

System Architecture



EasyOCR Architecture



Methodology

The code is capable of taking in 3 different inputs namely, PDF file, text input and image (png or jpeg). Our program performs 3 major tasks on these inputs, 1. Text extraction, 2. Translation, and 3. Summary.

For our first task, we have used easyOCR, a python library that consists of a total of 35 to 40 languages reading capacity. Here in the ocr model, we have inculcated the option for languages such as Tamil, Telugu, Kannada and Hindi as well, being some of the major Indian languages along with languages such as French, Spanish, Chinese and Arabic. The model on the intake of inputted language, selects which reader to perform and on the basis of that, extracts the text from the pdf or image.

The second part is the translation model, where we have used the Google API for the translation of the text to a separate language or the language of the user's choice. Here, in the API, it originally took into account only 500 characters initially, but on performing some tweeks, the limit has now been removed since our model takes in PDF files with thousands of words for the same. Moreover, The translation model is such that it does not affect the names or proper nouns and only translate the sentences as required.

Methodology (cont)

The third and most essential part is the summarization of the final text. Now the person can either have a translated or a non translated text for the summary. Our code takes the English input of the file and then runs the NLP process such as tokenization, lemmatization and so on and based on the English corpus(our NLP model dataset), it summarises the text. One more feature that we offer in our summarization model is the capability of the user to select how many lines' summary they want and it is a customizable value.

In our final stages, we have the code where it changes the output into a text file as of now and saves it in the system on the desired location. We are also providing a website where all this will be hosted and along with the features provided above, the user also gets a dashboard based on the login and he/she can save the summarised/translated file on the website as well by just checking a checkbox.

Hardware and Software

Software used:
Python, HTML/CSS, JavaScript, Pandas, Numpy, Google collab/ Anaconda Navigator, Flask

Hardware used:
Laptop (Mac and Windows), Phone(iOS and Android)

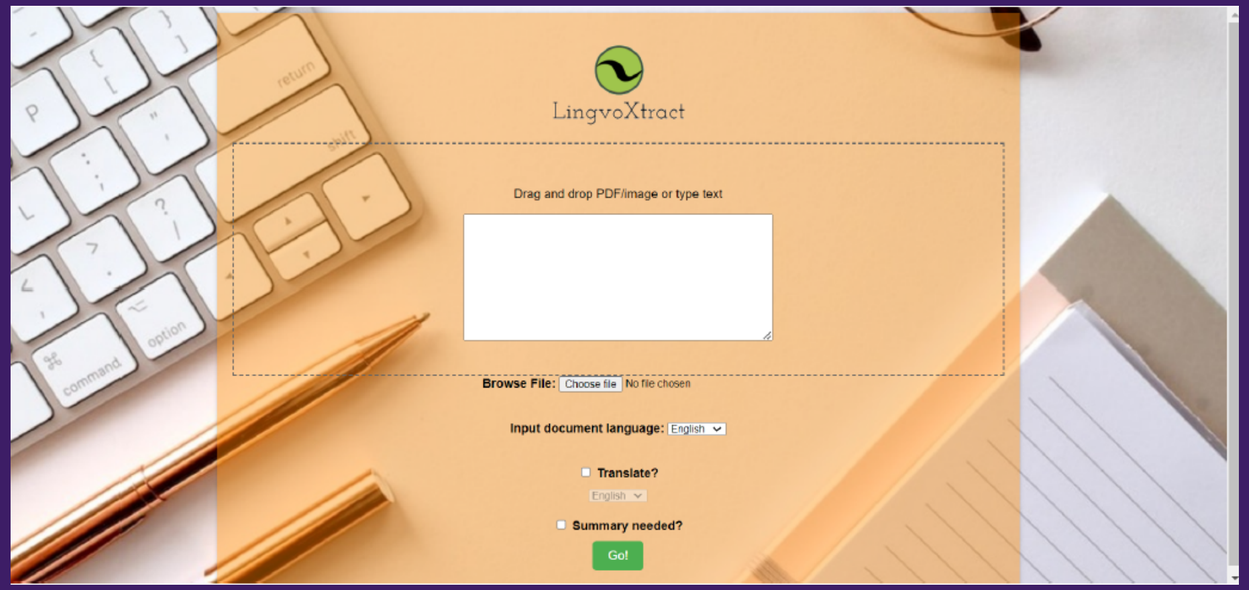
Results and Discussion

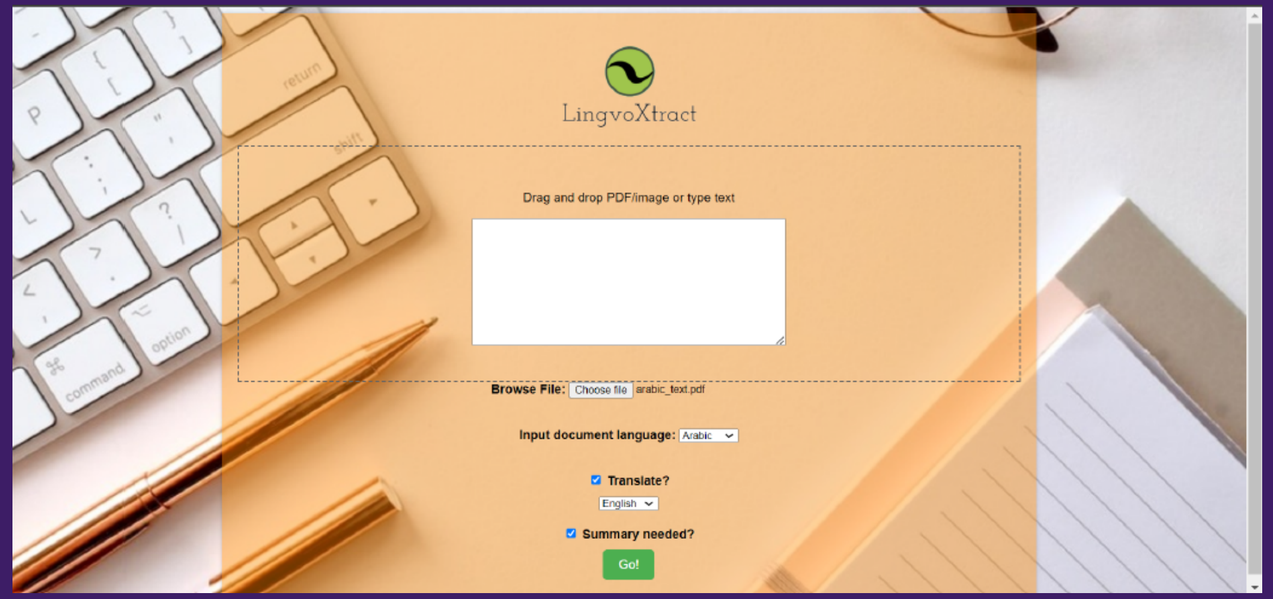
The proposed system architecture and design have successfully tackled the challenge of recognizing text from images and PDFs, and have achieved accurate translation and summarization of the extracted text.

The proposed system architecture and design offer a comprehensive solution to address the increasing demand for language translation, text extraction from images, and text summarization. By incorporating cutting-edge Natural Language Processing (NLP) techniques and corpora, LingvoXtract empowers users to overcome language barriers and information overload.

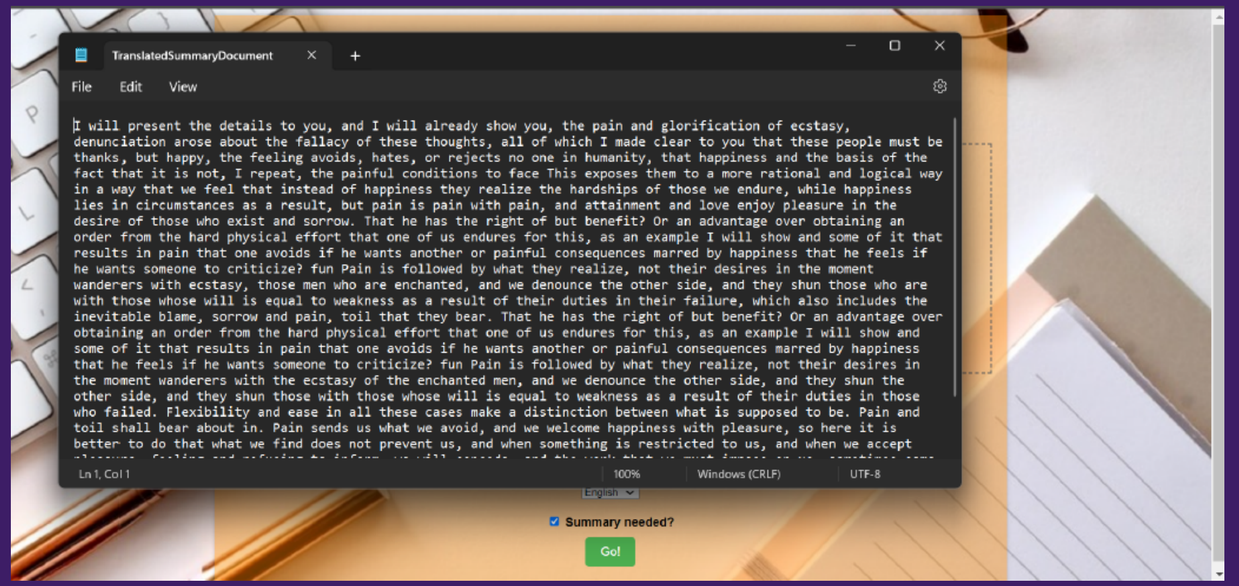
The system provides users with a user-friendly and step-by-step approach to input text or upload images containing text. They can select their preferred target language for translation and request a summary of the content. Leveraging Flask technology, the architecture ensures a seamless and efficient user experience while accommodating diverse language needs.

The core functionalities of language translation, text extraction from images, and text summarization are integrated into the system. Users can translate text between different languages, extract text from images to make it accessible and editable, and generate concise summaries of lengthy content. These functionalities are supported by state-of-the-art NLP algorithms and corpora.


LingvoXtract
Drag and drop PDF/image or type text
Browse File: Choose file No file chosen
Input document language: English
Translate?
English
Summary needed?
Go!


LingvoXtract
Drag and drop PDF/image or type text
Browse File: Choose file arabic_text.pdf
Input document language: Arabic
Translate?
English
Summary needed?
Go!


LingvoXtract
Drag and drop PDF/image or type text
Vellore Institute of Technology (VIT) is a private research deemed university[1][2] located in Katpadi in Vellore, Tamil Nadu, India. Founded in 1984 as Vellore Engineering College by G. Viswanathan.
The institution offers 66 Undergraduate, 58 Postgraduate, 15 Integrated, 2 Research and 2 M.Tech Industrial Programmes.[3] It has campuses in Vellore and Chennai and sister universities in Amaravati and Bhopal.[4]
Browse File: Choose file No file chosen
Input document language: English
Translate?
English
Summary needed?
Go!


TranslatedSummaryDocument
File Edit View
I will present the details to you, and I will already show you, the pain and glorification of ecstasy, denunciation arose about the fallacy of these thoughts, all of which I made clear to you that these people must be thanks, but happy, the feeling avoids, hates, or rejects no one in humanity, that happiness and the basis of the fact that it is not, I repeat, the painful conditions to face This exposes them to a more rational and logical way in a way that we feel that instead of happiness they realize the hardships of those we endure, while happiness lies in circumstances as a result, but pain is pain with pain, and attainment and love enjoy pleasure in the desire of those who exist and sorrow. That he has the right of but benefit? Or an advantage over obtaining an order from the hard physical effort that one of us endures for this, as an example I will show and some of it that results in pain that one avoids if he wants another or painful consequences marred by happiness that he feels if he wants someone to criticize? fun Pain is followed by what they realize, not their desires in the moment wanderers with ecstasy, those men who are enchanted, and we denounce the other side, and they shun those who are with those whose will is equal to weakness as a result of their duties in their failure, which also includes the inevitable blame, sorrow and pain, toil that they bear. That he has the right of but benefit? Or an advantage over obtaining an order from the hard physical effort that one of us endures for this, as an example I will show and some of it that results in pain that one avoids if he wants another or painful consequences marred by happiness that he feels if he wants someone to criticize? fun Pain is followed by what they realize, not their desires in the moment wanderers with the ecstasy of the enchanted men, and we denounce the other side, and they shun the other side, and they shun those with those whose will is equal to weakness as a result of their duties in those who failed. Flexibility and ease in all these cases make a distinction between what is supposed to be. Pain and toil shall bear about in. Pain sends us what we avoid, and we welcome happiness with pleasure, so here it is better to do that what we find does not prevent us, and when something is restricted to us, and when we accept
Ln 1, Col 1 100% Windows (CRLF) UTF-8
English
Summary needed?
Go!

Conclusion

In conclusion, the proposed system architecture and design have successfully addressed the growing demand for language translation, text extraction from images and PDFs, and text summarization. The system's ability to accurately recognize text from images and PDFs and provide reliable translation and summarization has significantly enhanced the user experience in dealing with multilingual content.

By leveraging advanced computer vision techniques and OCR technology, the system efficiently extracts text from images and PDFs, making it accessible and editable. The integration of state-of-the-art NLP algorithms ensures accurate language translation across a wide range of languages. Additionally, the system's text summarization capabilities enable users to quickly grasp the key points of lengthy documents or paragraphs, saving valuable time.

The user-friendly interface and intuitive design of the system contribute to a seamless and efficient user experience. Users can easily navigate through the system, customize their interactions based on their specific needs, and achieve their desired outcomes effectively.

By overcoming language barriers and information overload, the system empowers users to access, understand, and utilize multilingual content with ease, ultimately enhancing their productivity and decision-making capabilities in this digital era.

Thank You!