

REAL-TIME TEXT TRANSLATION

Rutij Rajesh Shenvi Navelkar (201105047)

Project Report submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF ENGINEERING

Branch: COMPUTER ENGINEERING

of

GOA UNIVERSITY



2025

Under the Guidance of **Prof. Nitesh Naik**

COMPUTER ENGINEERING DEPARTMENT

GOA COLLEGE OF ENGINEERING

(GOVERNMENT OF GOA)

FARMAGUDI, PONDA, GOA – 403401

INDEX

1. INTRODUCTION.....	2
1.1 Motivation	2
1.2 Description	3
1.3 Objectives	4
2. LITERATURE REVIEW	5
2.1 OCR	5
2.2 Translation Using Recurrent Neural Networks	7
2.3 Speech Recognition (OpenAI's Whisper)	14
2.4 Hardware and Software Requirements	16
2.5 Summary of Existing Work	17
3. DESIGN	19
3.1 Authentication	20
3.2 Text Translation	21
3.3 Document Translation	22
3.4 Image Text Extraction and Translation	23
3.5 Video Captioning	24
3.6 Database Integration	25
4. TESTING AND RESULTS	26
4.1 Text Translation	26
4.2 Document Translation	27
4.3 Image Translation	29
4.4 Video Captioning	31
4.5 Displaying Previous Translations	33
5. PROJECT TIMELINE	35
CONCLUSION	35
REFERENCES	36

INTRODUCTION

In an increasingly interconnected world, linguistic diversity remains a vital component of our cultural tapestry. The Konkani Language, a vibrant vernacular with a rich history, has a unique place in this mosaic. To bridge the gap between languages and cultures, we introduce the "Konkani Language Translation Tool," a powerful and versatile platform. With simplicity at its core, this system ensures a seamless transition from Konkani to English Language, promoting effective communication.

Users can effortlessly input Konkani text, receiving accurate translations in real-time and vice versa. Beyond simple text translation users can also use this tool for translating PDF files, text from images. It also provides a video captioning service to create Konkani closed captions for English videos.

1.1 MOTIVATION

In a rapidly globalizing world, Konkani, a vernacular language, confronts the risk of diminishing significance and accessibility. Traditional translation methods and tools, such as Google Translate, have inadequately addressed the nuanced complexities of Konkani, often resulting in inaccurate or awkward translations.

Additionally, auto-generated captions on Youtube videos do not offer an option for Konkani. This research project is motivated by the dire need to develop a specialized text-based language translation system for Konkani that respects its unique nuances and cultural context.

By creating a platform tailored to Konkani's linguistic intricacies, we aim to facilitate communication, bridge cultural divides, and ensure the endurance of this beautiful language in an increasingly interconnected world.

1.2 DESCRIPTION

This translation service aims to facilitate seamless and real-time translation for Konkani. We aim to provide precise translation from Konkani, breaking down communication barriers and promoting the preservation of Konkani as a language.

It aims to be able to translate text in the form of images or documents in Konkani, to shed light on older texts written in the language, to make them accessible/understandable to a larger userbase. It also aims to provide Konkani captions for videos with intelligible English audio, given the video data is also available.

A main part of this project is the translation methods used, in order to provide understandable and sensible translations. Another part is being able to preserve the richness of the Konkani language, where some traditional models have fallen short of doing so, giving awkward or inaccurate translations.

Some features aimed for implementation include-

- Real-time Translation: Enable users to have real-time conversations with people who speak different languages. The translator should provide instant translation of spoken words.
- Let user translate from raw text, an image (jpeg) or document form (pdf) by extracting the text from the media mentioned, and then running it through the translation model.
- Provide the output of the image/document translation in a usable format (pdf/word file).
- Let users view their previously translated texts, PDF and images and download them in .docx format.

1.3 OBJECTIVES

This project's core objectives are as follows:

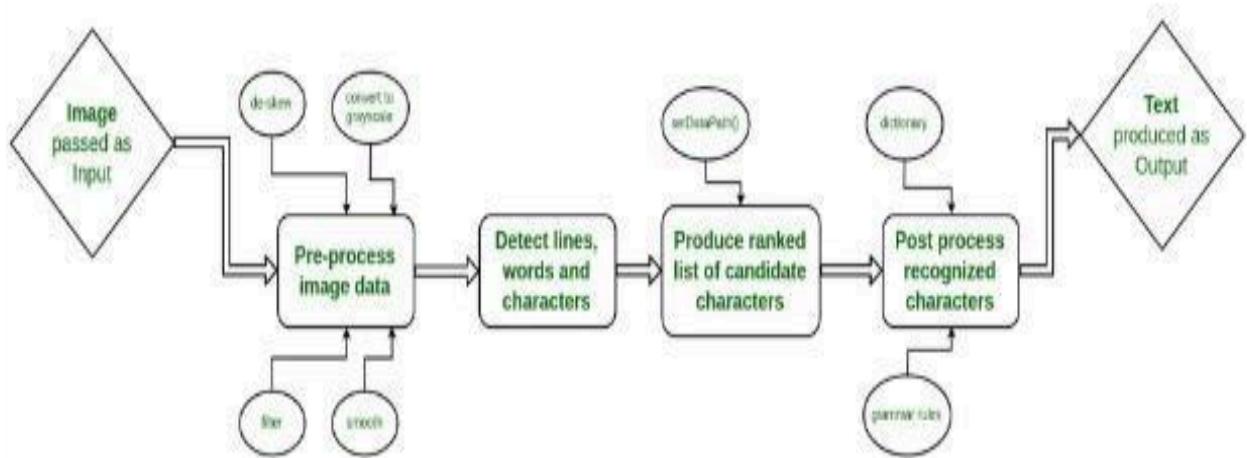
- **Efficient Translation System:** Develop an efficient and real-time text-based language translation system for the Konkani language, ensuring precision and fluency in translations.
- **Breaking Language Barriers:** Enable accurate translation between Konkani and English, bridging communication gaps and promoting cross-cultural understanding.
- **Preservation of Konkani Culture:** Contribute to the preservation of Konkani culture and heritage by empowering the language's speakers to communicate effectively in a globalized world.
- **Respect for Linguistic Nuances:** Design a platform that respects the unique linguistic intricacies and cultural context of Konkani, ensuring that translations are both accurate and culturally sensitive.
- **User-Friendly Interface:** Create a user-friendly interface that allows individuals, regardless of their technological proficiency, to easily use the translation tool, thereby encouraging its adoption.
- **Long-Term Sustainability:** Ensure the long-term sustainability of the translation system, keeping it up-to-date with evolving language usage and technological advancements.

By addressing these objectives, this project aims to make the Konkani Language Translation Tool a valuable resource for Konkani speakers, learners, and anyone interested in exploring this beautiful language. It's a step towards preserving linguistic diversity, fostering global communication, and celebrating the unique cultural heritage of Konkani.

LITERATURE REVIEW SUMMARY

2.1 OCR (OPTICAL CHARACTER RECOGNITION)

General working of OCR



OCR technology solves the problem by converting text images into text data that can be analyzed by other business software. You can then use the data to conduct analytics, streamline operations, automate processes, and improve productivity. Tesseract OCR is an optical character reading engine developed by HP laboratories in 1985 and open sourced in 2005. Since 2006 it is developed by Google. Tesseract has Unicode (UTF-8) support and can recognize more than 100 languages “out of the box” and thus can be used for building different language scanning software also. Latest Tesseract version is Tesseract 4. It adds a new neural net (LSTM) based OCR engine which is focused on line recognition but also still supports the legacy Tesseract OCR engine which works by recognizing character patterns.

2.1.1 HOW OCR WORKS

The OCR engine or OCR software works by using the following steps:

- a. **Image acquisition:** A scanner reads documents and converts them to binary data. The OCR software analyzes the scanned image and classifies the light areas as background and the dark areas as text.
- b. **Preprocessing:** The OCR software first cleans the image and removes errors to prepare it for reading. These are some of its cleaning techniques:
 - De-skewing or tilting the scanned document slightly to fix alignment issues during the scan.
 - De-speckling or removing any digital image spots or smoothing the edges of text images.
 - Cleaning up boxes and lines in the image.
- c. **Text recognition:** The two main types of OCR algorithms or software processes that an OCR software uses for text recognition are called pattern matching and feature extraction.
 - **Pattern matching:** Pattern matching works by isolating a character image, called a glyph, and comparing it with a similarly stored glyph. Pattern recognition works only if the stored glyph has a similar font and scale to the input glyph. This method works well with scanned images of documents that have been typed in a known font.
 - **Feature extraction:** Feature extraction breaks down or decomposes the glyphs into features such as lines, closed loops, line direction, and line intersections. It then uses these features to find the best match or the nearest neighbor among its various stored glyphs
- d. **Post-processing:** After analysis, the system converts the extracted text data into a computerized file. Some OCR systems can create annotated PDF files that include both the before and after versions of the scanned document.

Features

The advantages of OCR are numerous, but namely:

- It increases the efficiency and effectiveness of office work
- The ability to instantly search through content is immensely useful, especially in an office setting that must deal with high volume scanning or high document inflow.
- OCR is quick ensuring the document's content remains intact while saving time as well.

2.2 TRANSLATIONS USING RECURRENT NEURAL NETWORKS

Recurrent Neural Networks (RNN) Encoder-Decoder model consists of two recurrent neural networks. One RNN encodes a sequence of words into the fixed-length vector by first passing it through an embedding layer which is used to form an embedding matrix and then it is fed into a recurrent neural network. In the encoder section, there is no output given through the softmax layer and is fed to the next RNN cell.

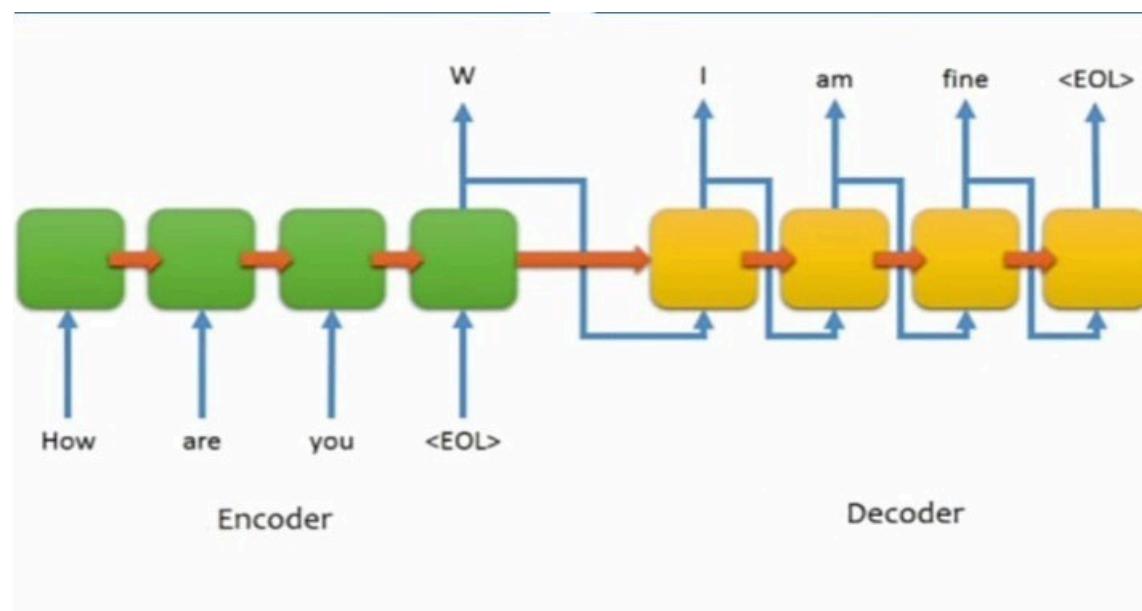


Fig. 2.3.1 – Basic Diagram of an RNN encoder-decoder model

Encoder and decoder both use the same neural network model but play a somewhat different role. The encoder is used to encode all the word embeddings and extract context and long-term dependencies which are then passed over to decoder to generate output sentence.

The problem of long-term dependencies is that sometimes we just need to look at recent information to perform the present task. For example, consider a language model trying to predict the last word in “the clouds are in the sky”. Here it’s easy to predict the next word as sky based on the previous words. But consider the sentence “I grew up in France. I speak fluent French.” Here it is not easy to predict that the language is French directly. It depends on previous input also. In such sentences it is entirely possible for the gap between the relevant information and the point where it is needed to become very large. In theory, RNN’s are capable of handling such “long-term dependencies.”

Sadly, in practice, recurrent neural networks don’t seem to be able to learn them. This problem is called Vanishing gradient problem. The neural network updates the weight using the gradient descent algorithm. The gradients grow smaller when the network progress down to lower layers. The gradients will stay constant meaning there is no space for improvement. The model learns from a change in the gradient. This change affects the network’s output. However, if the difference in the gradients is very small network will not learn anything and so no difference in the output.

Therefore, a network facing a vanishing gradient problem cannot converge towards a good solution.

LSTM – SOLUTION TO LONG-TERM DEPENDENCIES

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) that is well-suited for learning long-term dependencies in sequential data. RNNs are a type of neural network that can learn from sequential data, such as text or time series data. However, traditional RNNs have difficulty learning long-term dependencies in

the data. LSTM networks overcome this problem by using a special type of memory cell that can store information for long periods of time. This allows LSTM networks to learn how past events can influence future events.

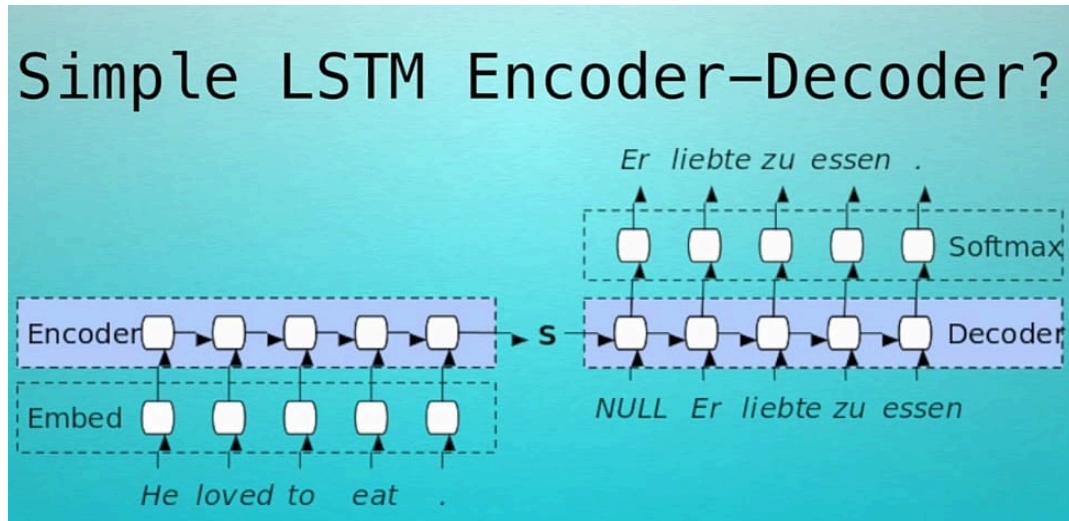


Fig. 2.4.1- Simple LSTM Encoder-Decoder, using memory cells to store information long-term.

However, there is a drawback to this architecture, it has limited memory. The hidden state 's' of the LSTM is where we are trying to fit the whole sentence we want to translate. 's' is usually only a few hundred floating point numbers long. The more we try to force our sentence into this fixed dimensionality vector the more lossy our neural network is forced to be.

We could increase the hidden size of the LSTM, after all they are supposed to remember long term dependencies but what happens is as we increase the hidden size 'h' of the LSTM the training time increases exponentially.

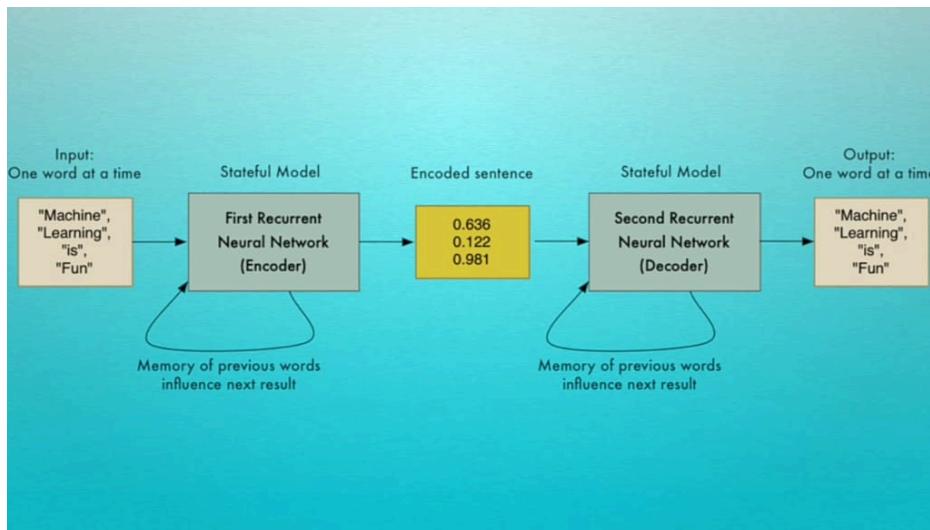


Fig. 2.4.2 – LSTM model using hidden states to ‘memorize’ previous words

So, to solve this we are going to bring '**Attention**' into the mix. If we were to translate a long sentence, we would probably glance back at the source sentence a couple times to make sure we are capturing all the details, iteratively paying attention to the relevant parts of the source sentence. We can let neural nets do the same by letting them store and refer to previous outputs of the LSTM.

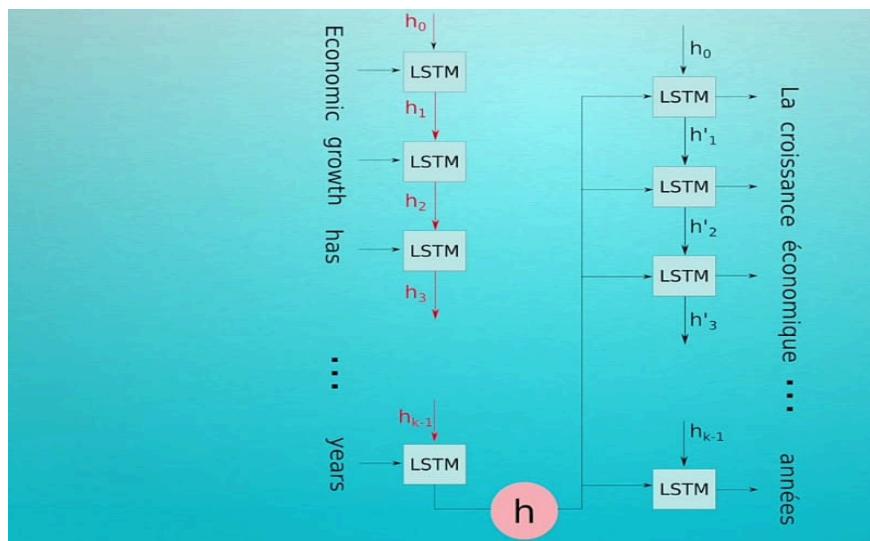


Fig. 2.4.3 – LSTM model without ‘Attention’ Mechanism

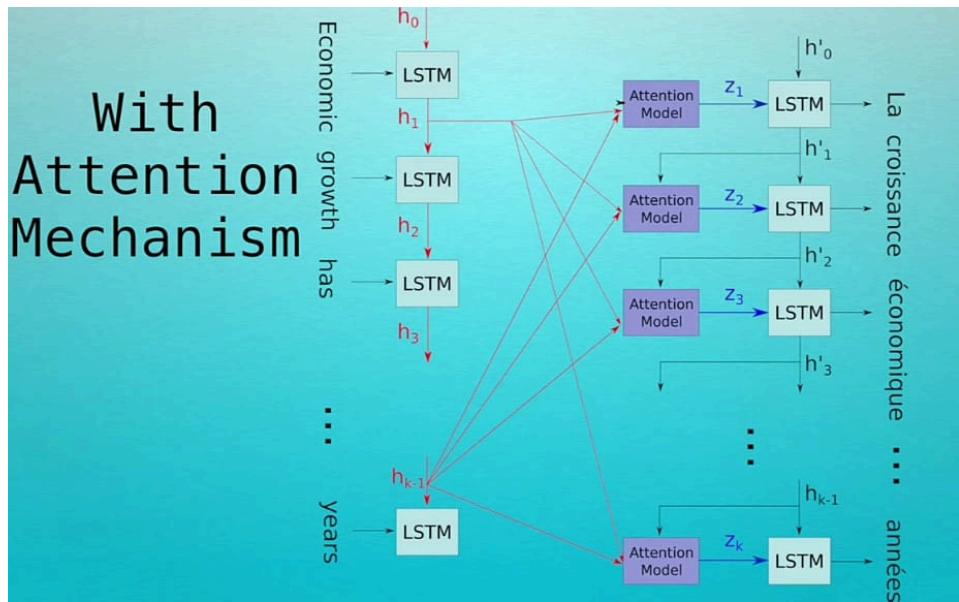


Fig. 2.4.4 – LSTM model using ‘Attention’ Mechanism, using previous relevant states to use in later stages.

Using Attention Mechanism increases the storage of our model without changing the functionality of the LSTM. So, the idea is once we have LSTM outputs from the encoder stored, we can query each output asking how relevant they are to the computation happening in the encoder. Each encoder output gets a relevancy score which we can convert to a probability score by applying a softmax activation to it.

Key Features of LSTM:

- Sequential Data Handling: LSTMs can capture and model sequential data, essential for analyzing price movements over time.
- Long-Term Dependencies: They can identify and model long-term dependencies, crucial in financial predictions.
- Memory Cells: LSTMs store and retrieve relevant information while discarding noise.

LSTMs are used for learning long-term dependencies and used for processing and predicting time-series data. General RNNs have a single neural network layer. LSTMs, on the other hand, have four interacting layers communicating extraordinarily. An LSTM network is made up of a series of LSTM cells. Each cell has three gates: an input gate, a forget gate, and an output gate.

- **Forget gate:** The forget gate controls how much information is removed from the cell state. It does this by taking the previous hidden state as input and

outputting a value between 0 and 1. This value is then used to multiply the cell state, and the product is subtracted from the cell state.

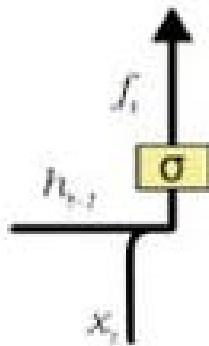


Fig. 2.4.5 – Forget Gate controls amount of information removed from a cell state.

- **Input gate:** The input gate controls how much new information is added to the cell state. It does this by taking the input token at the current time step and the previous hidden state as input and outputting a value between 0 and 1. This value is then used to multiply the new information, and the product is added to the cell state.

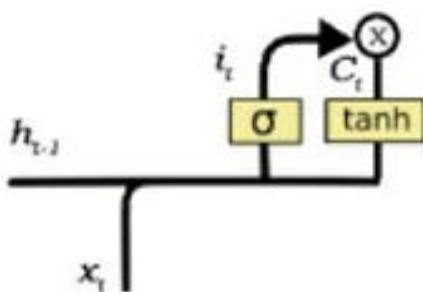


Fig. 2.4.6 – Input Gate regulates amount of new information added to cell state.

- **Output gate:** The output gate controls how much information is output from the cell state. It does this by taking the cell state and the previous hidden state as input and outputting a value between 0 and 1. This value is then used to multiply the cell state, and the product is the output of the LSTM cell.

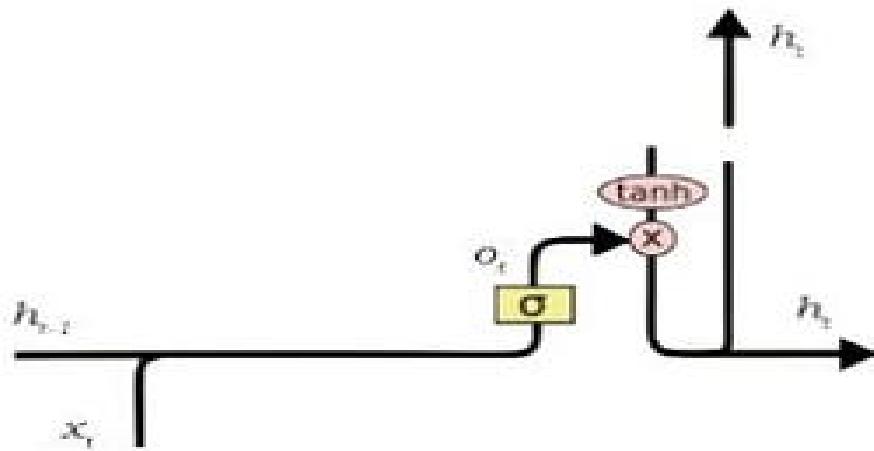


Fig. 2.4.7 – Output Gate regulating the amount of information outputted from the cell state.

An LSTM network works by processing the input sequence one token at a time. At each time step, the LSTM network updates the cell state and the hidden state. The hidden state is the output of the LSTM network and it contains information about the entire input sequence up to that point.

2.3 SPEECH RECOGNITION (OpenAI's Whisper model)

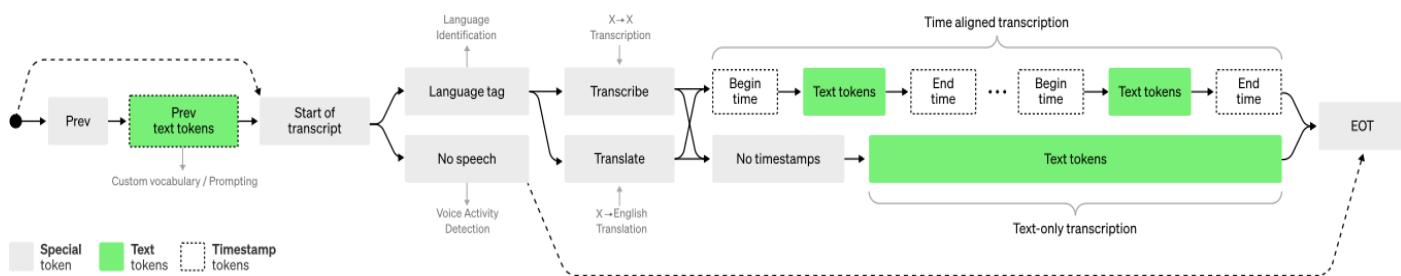
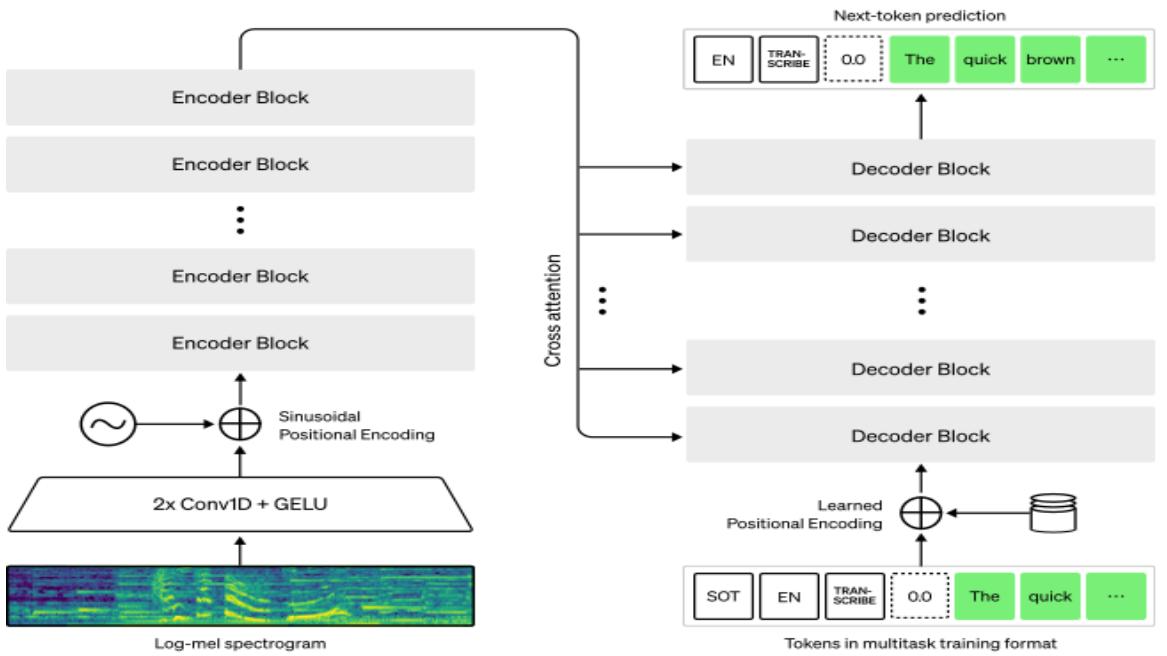
Whisper is an **automatic speech recognition (ASR) system** trained on **680,000 hours** of multilingual and multitask supervised data collected from the web.

A Transformer sequence-to-sequence model is trained on various speech processing tasks, including multilingual speech recognition, speech translation, spoken language identification, and voice activity detection. These tasks are jointly represented as a sequence of tokens to be predicted by the decoder, allowing a single model to replace many stages of a traditional speech-processing pipeline. The multitask training format uses a set of special tokens that serve as task specifiers or classification targets.

There are six model sizes, four with English-only versions, offering speed and accuracy tradeoffs. Below are the names of the available models and their approximate memory requirements and inference speed relative to the large model.

Size	Parameters	Required VRAM	Relative speed
tiny	39 M	~1 GB	~10x
base	74 M	~1 GB	~7x
small	244 M	~2 GB	~4x
medium	769 M	~5 GB	~2x
large	1550 M	~10 GB	1x
turbo	809 M	~6 GB	~8x

The speed may vary significantly depending on many factors including the language, the speaking speed, and the available hardware.



The Whisper architecture is a simple end-to-end approach, implemented as an encoder-decoder Transformer. Input audio is split into 30-second chunks, converted into a log-Mel spectrogram, and then passed into an encoder. A decoder is trained to predict the corresponding text caption, intermixed with special tokens that direct the single model to perform tasks such as language identification, phrase-level timestamps, multilingual speech transcription, and to-English speech translation.

2.4 HARDWARE AND SOFTWARE REQUIREMENTS

- Stable and consistent internet connection.
- Desktop computer with reasonably powerful CPU for reasonable translation efficiency.
- Dedicated GPU, preferably Nvidia (optional). This can accelerate training and interface of large language models.
- Enough RAM to handle language models and processing larger chunks of data.
- Basic I/O devices (keyboard, mouse and monitor).
- Browsers supported: Google Chrome, Microsoft Edge, Mozilla Firefox, Opera(Javascript enabled)

Technologies Used-

- React (Frontend)
- NodeJS (Backend)
- Python (Scripts for OCR and Captioning)
- MySQL (Storing the user credentials and translations)
- Imagemagick Library (To display Konkani captions on videos)
- PytesseractOCR (Python-tesseract is a wrapper for Google's Tesseract OCR engine for python. That is, it will recognize and “read” the text embedded in images)

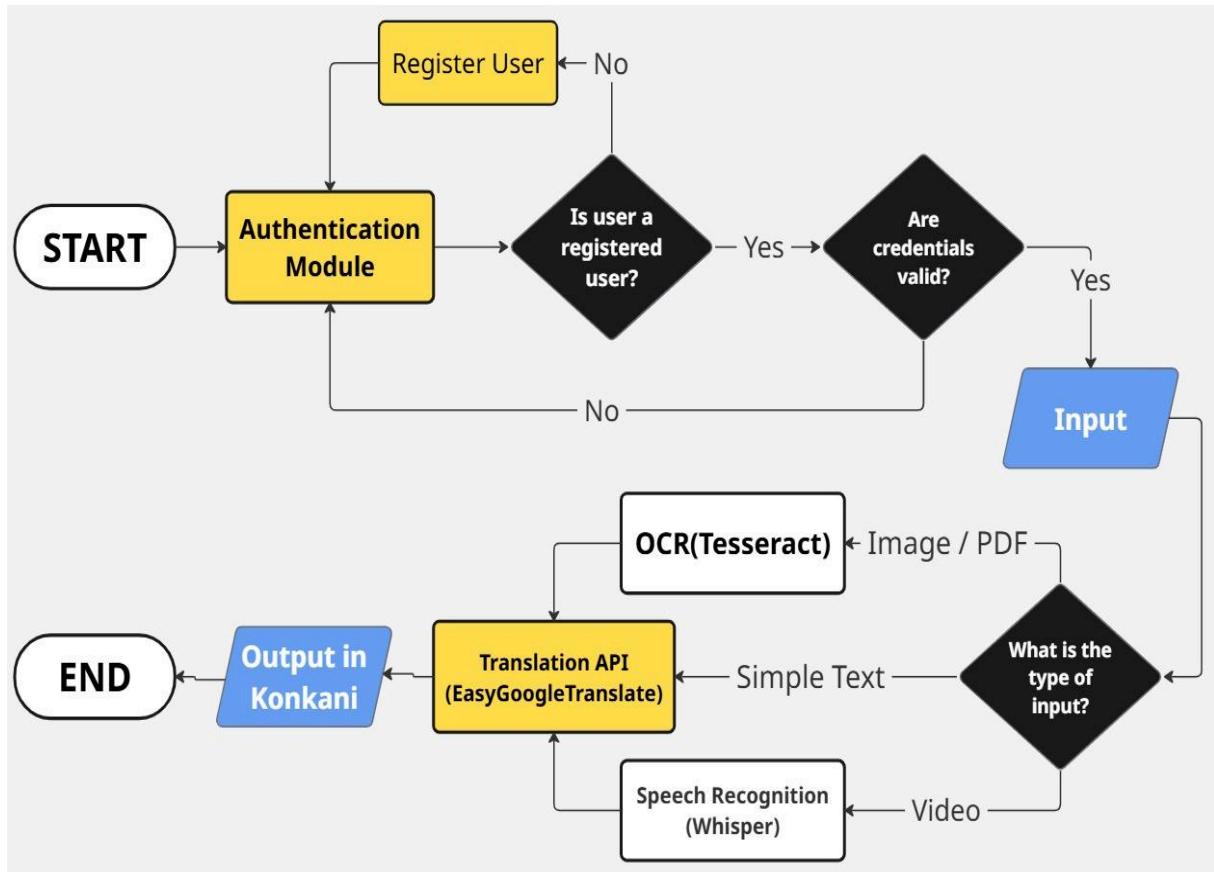
2.5 SUMMARY OF EXISTING WORK

Name of Literature	Inference Derived
1] KOP: An Opinion Mining System in Konkani	The research paper introduces the proposed work, detailing the architecture of the Konkani Opinion Mining System (KOP), including tokenization, stop word elimination, the development of Konkani-SentiWordNet (K-SWN), polarity determination, and the calculation of polarity at word, sentence, paragraph, and document levels. The document concludes with a method for calculating the overall polarity of the document based on the summation of individual paragraph polarities.
2] A Survey of Machine Translation Approaches for Konkani English	The research paper discusses the importance of Konkani, the native language of Goa, India, and its limited translation into other languages, hindering the dissemination of Konkani literature and Goa's history. It introduces the field of language translation, mentioning three methods: dictionary-based, corpus-based, and machine translation (MT). The paragraph then emphasizes the proposed use of Statistical Machine Translation (SMT) for Konkani, using a parallel corpus. The aim is to compare various language translation systems, focusing on automated processes with minimal human intervention.
3] Handwritten Devanagari Numeral and Vowel Recognition using Invariant Moments	The research paper describes the recognition of handwritten Devanagari numerals and vowels, covering sections on Devanagari script, data preparation, invariant moment, affine moment invariant, recognition system, experimental results, and conclusion. The Devanagari script, is discussed in terms of its structure and features. The lack of a standardized database for Devanagari handwritten characters is highlighted, and the paper presents a created database consisting of 2000 samples of handwritten Devanagari numerals and 1250 samples of vowels, collected from various writers. The proposed system focuses on the Konkani language, written in the Devanagari script. It explains the segmentation process involving line, word, and character segmentation. The OCR system is designed to recognize and extract text from scanned images of handwritten Konkani documents
4] Konkani Script to Speech Conversion by Concatenation of recognized Hand written Konkani Text Using Neural Network	The Research proposed model for POS tagging in Konkani involves three main steps: data collection, training phase, and testing phase. The model utilizes the Hidden Markov Model (HMM) and the Viterbi algorithm. The training phase involves collecting pre-tagged Konkani data, training the model using the HMM, and calculating transition and emission probabilities. The testing phase applies the Viterbi algorithm to find the best tag sequence for unseen Konkani text.

5] Parts of Speech Tagging for Konkani Language	The Research proposed POS tagging model for Konkani aims to annotate words in a text with their corresponding parts of speech using the HMM and Viterbi algorithm. The text discusses the challenges posed by the morphologically rich nature of Konkani and suggests potential future optimizations, such as using different training methods and incorporating diverse domains of data. The model's accuracy and efficiency are highlighted, with potential applications in information retrieval, speech recognition, and other NLP tasks.
6] Robust Speech Recognition via Large-Scale Weak Supervision	Whisper is a general-purpose speech recognition model trained on 680,000 hours of multilingual and multitask supervised data collected from the web. The model is designed to perform automatic speech recognition (ASR) and can also handle speech translation, language identification, and voice activity detection.

DESIGN

This chapter aims to present the methodology of various aspects of the project – the text translation, document translation, image text extraction and translation and lastly the video captioning service.



3.1 AUTHENTICATION

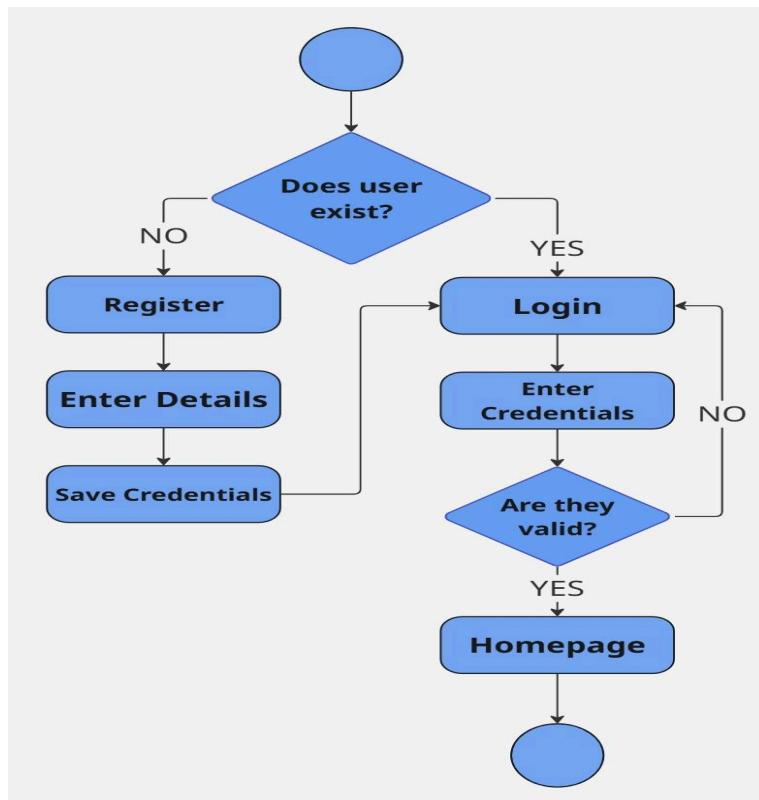


Fig 3.1: Authentication Flowchart

1. The user is presented with the login page.
2. If the user is not registered, they will be prompted to register.
3. The user is then redirected to the registration page, where they will have to enter details (desired username and password).
4. After submission (if username meets required criteria), the credentials are saved and the user is to then login using said credentials.
5. If the details are valid, the user is now logged in, else they are prompted to try again.

3.2 TEXT TRANSLATION

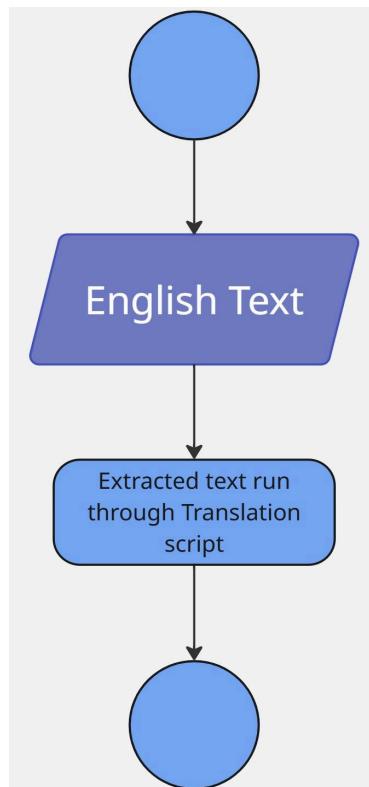


Fig. 3.2.1 - Text Translation Flowchart

1. After the user is logged in, they can click the ‘Translate Now’ button to begin text translation.
2. The user must then enter the text to be translated into the text-box provided.
3. The text is then sent to the translation scripts.
4. The output is rendered in the text field to the right.

3.3 DOCUMENT TRANSLATION

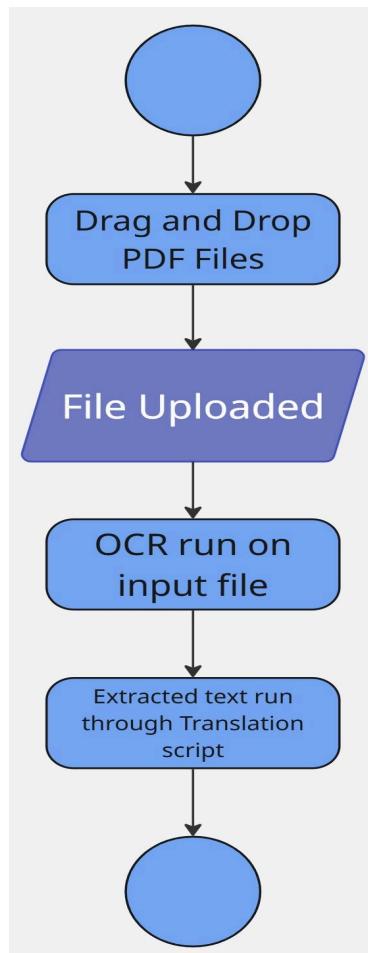


Fig. 3.3.1 - Document Translation Flowchart

1. The user must select the ‘Translate Files’ option to translate PDF files.
2. The user can drag-and-drop a file or browse on the local machine for a pdf file to translate the text from.
3. The PDF is then uploaded and temporarily stored.
4. The file is then fed to the OCR algorithm to extract text from it.
5. The text extracted from the previous step is then translated.
6. The translated text is then outputted in the form of a .doc file

3.4 IMAGE TEXT EXTRACTION AND TRANSLATION

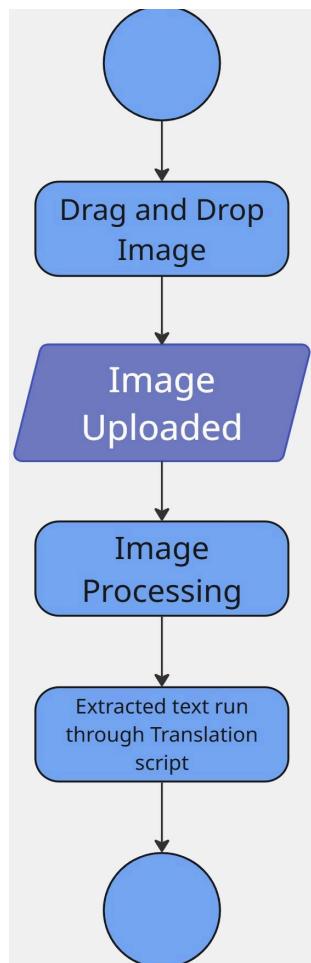


Fig. 3.4.1 - Image Text Translation Flowchart

1. The user must select the ‘Translate Image’ option to translate Images.
2. The user can drag-and-drop an image or browse on the local machine for a pdf file to translate the text from.
3. The image is then uploaded and temporarily stored.
4. The image is then subject to image processing algorithms and converted to greyscale for text extraction.
5. The text extracted from the previous step is then translated.
6. The translated text is then outputted.

3.5 VIDEO CAPTIONING

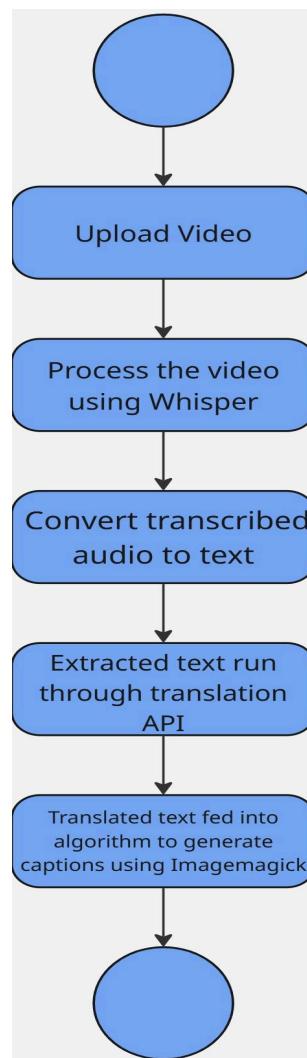


Fig. 3.5.1 - Video Captioning Flowchart

1. The user can drag-and-drop a video or browse the machine.
2. The video is loaded using the **MoviePy** library.
3. It is processed using **OpenAI's Whisper model** to generate accurate time-aligned transcriptions of the spoken content.
4. The transcribed audio is converted to text which is then translated into the Konkani language using the **EasyGoogleTranslate API**.
5. Translated subtitles are rendered into images (frames) using **ImageMagick (Wand)**. This approach ensures that complex scripts (like Devanagari) are accurately displayed with proper font support and styling.

3.6 DATABASE INTEGRATION

MySQL relational database is used to store the user's details such as Username, Email and Password, and also the text, PDF and images that they have translated.

3.6.1 Database Schema Design

Below is the MySQL schema showing two tables and their purposes:

- **users**

Stores registered user details:

- id (Primary Key)
- username
- email
- password

- **translations**

Stores user-translated text and translated PDF/Images in docx format:

- id (Primary Key)
- user_id (Foreign Key → users)
- type (e.g., "text", "pdf", "image", "video")
- input_text
- translated_text
- file_data (Stores the data from the .docx generated after translation)
- filename
- created_at

TESTING AND RESULTS

4.1 TEXT TRANSLATION

To test the translation of text using our scripts, the following English text was entered into the text-box:

“Exploring the Mysteries of the Deep Ocean

The ocean, with its vastness and depth, has long intrigued mankind. From the shimmering surface to the dark abyss, it holds countless secrets waiting to be discovered. Explorers and scientists alike have embarked on daring expeditions to unravel the mysteries of the deep ocean.

One of the most fascinating aspects of the ocean is its diverse marine life. From tiny plankton to massive whales, every corner of the ocean teems with biodiversity. Researchers study these creatures to understand their behaviors, adaptations, and roles in the marine ecosystem.””

The output can be seen in attached image. The obtained translation is satisfactory and preserves the meaning and context of the inputted text.

खोल म्हासागराच्या गूढतायेचो सोद घेवप .

दर्यात व्हडपण आनी खोलाय आशिल्ल्यान मनीसजातीक खूब काळ गूढ आसा. झिलमिलपी पृथग्सावन काळोख अथांग मेरेन तो सोदून काडपाची वाट पळोवपी असंख्य गुपीत धरता. खोल दर्यातल्या गूढ गजालींचो सोद घेवपाच्या धैर्यपूर्ण मोहिमेक संशोधक आनी शास्त्रज्ञांनी सारकेच सुरवात केल्या.

दर्यातिले एक आकर्षक आंग म्हळ्यार ताचो दर्याचो विविधताय आशिल्लो. ल्हान ल्हान प्लवंग ते व्हड क्हेल मेरेन, दर्याच्या दर एका कोनशाक जैवविविधताय आशिल्ली टीम आसता. संशोधकांनी ह्या प्राण्यांचो अभ्यास करून दर्याच्या पर्यावरण यंत्रणेत तांची वागणूक, अनुकूलन आनी भुमिका समजून घेवपाखातीर अभ्यास करतात.

Fig 4.1.1: Translated Konkani Text

4.2 DOCUMENT TRANSLATION

To test the document translation, the following pdf file was uploaded in the drag-and-drop field.

OPEN SOURCE TOOLS FOR IAAS

IaaS is a provision model in which an organization outsources the equipment used to support operations, including storage, hardware, servers, and networking components.

The service provider owns the equipment and is responsible for housing, running, and maintaining it.

The client typically pays on a per-use basis.

The following subsections explain some of the open source tools available for providing IaaS in cloud computing.

OPENNEBULA

OpenNebula is a flexible tool that orchestrates storage, network, and virtualization technologies to enable the dynamic placement of services on distributed infrastructures.

OpenNebula provides a modular architecture intended to be flexible.

One of these modules is the *scheduler*, an interesting algorithm that places virtual machines (VMs) depending on their requirements.

The client communication is also managed by modules that offer interfaces based on web services.

It is perhaps the only open management platform that has invested into a tailorabile VM placement algorithm.

As such, it may provide a nice environment for those researchers seeking to compare and develop different resource allocation strategies.

Fig 4.2.1: Sample PDF file

The following is obtained as an output in the form of a .doc file.

आयए खातीर मुक्त सोत साधनं

आयएएस हें एक तरतूद मॉडेल आसून तातूत एक संघटना उपकरणांक आऊटसोर्स करता.

साठवण, हार्डवेअर, सर्वर, आनी नेटवर्किंग सयत कार्यांक आदार दिवपाक वापरतात.

घटक.

सेवा पुरवणदाराक उपकरणांची मालकी आसा आनी घरां खातीर जापसालदार आसा,

धांवप, आनी ती सांबाळप.

ग्राहक सादारणपणान दर वापराच्या आदाराचेर फारीक करता.

सकयल दिल्ल्या उपविभागांनी उपलब्ध आशिल्ल्या कांय मुक्त सोत साधनांचो स्पृश्टीकरण मेळटा.

मेघ संगणनांत आयए दिवप.

ओपननेबुला 10000000000000000000000.

OpenNebula हें एक लवचीक साधन आसा जें साठवण, नेटवर्क, आनी आभासीकरण वाद्यवृद्धाचें वाद्यवृद्ध करता.

वितरीत मुळाट्या साधनांचेर सेवांचें गतिशील प्लेसमेंट सक्षम करपाक तंत्रज्ञान.

ओपननेबुला लवचीक आसपाचो हेतू आशिल्ली मॉड्यूलर वास्तुकला पुरवण करता.

तातूतलो एक मॉड्यूल म्हळ्यार वेळापत्रक, एक मनोरंजक अल्गोरिदम जो आभासी मशीनां (VMS) तांच्या गरजांचेर आदारून आसतात.

ग्राहक संचारणाचें वेवस्थापन मॉड्यूलां वरवीय वेवस्थापन करतात जे संवादाचेर आदारीत संवाद दितात.

वेब सेवा.

घडये तो एकूच उक्तो वेवस्थापनमंच आसतलो जो एक भनुरूप व्हीएमांत गुंतवणूक केला.

प्लेसमेंट अल्गोरिदम.

अशेतरेन, तुळा करपाक सोदपी संशोधकांक तें एक छान वातावरण दिवंक शक्ता.

आनी वेगवेगळीं संसाधन वाटपाचीं रणनितीं विकसीत करप.

Fig 4.2.2: Translated Text from PDF (.doc file)

Like the text translation, the obtained Konkani translation retains the meaning and tone of the original English text from the PDF file. Although the OCR may struggle with complex documents containing **multiple fonts, sizes, and layouts**.

4.3 IMAGE TRANSLATION

To test the image translation, the following image was fed to the site:

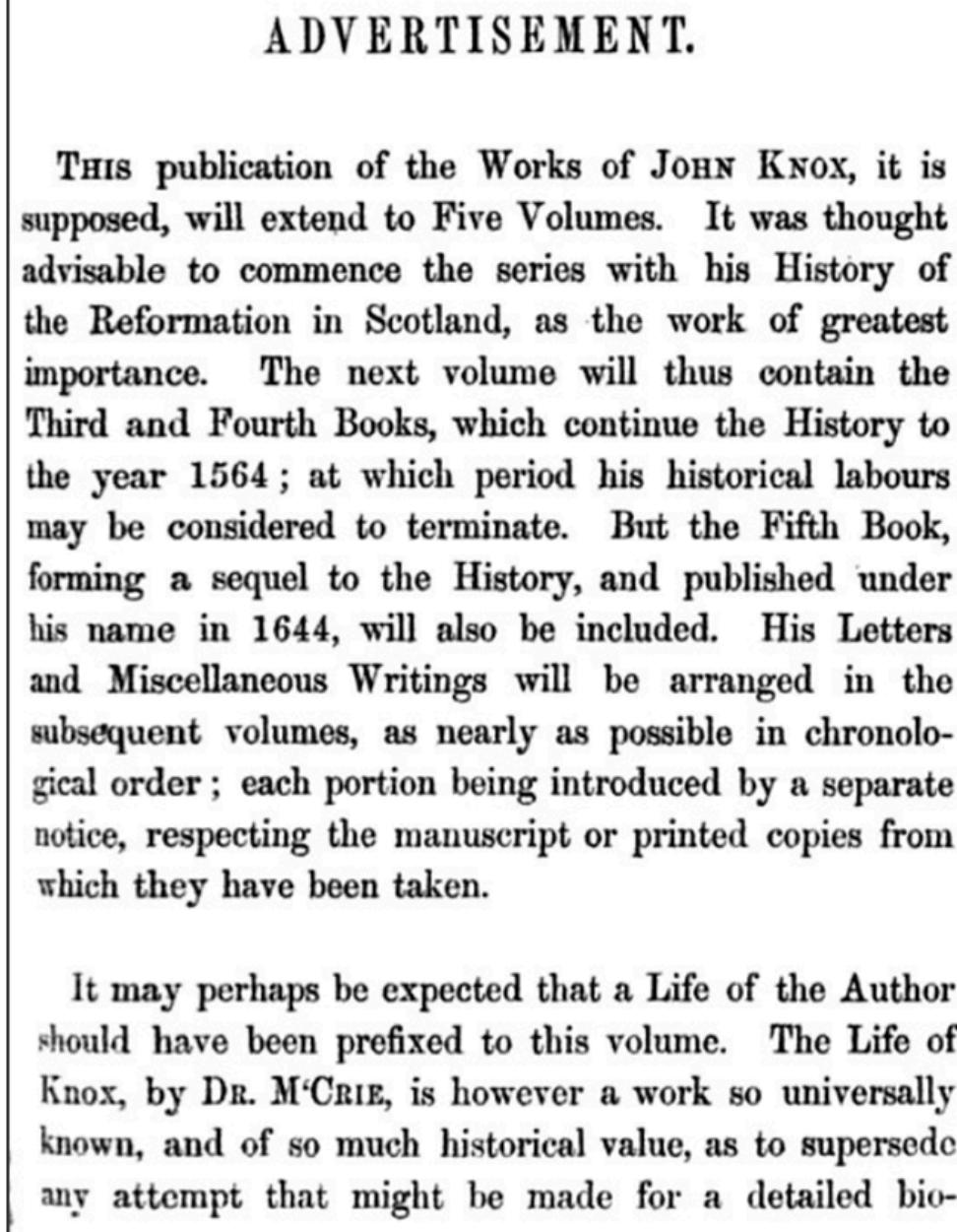


Fig 4.3.1: Sample Image (Screenshot)

The following is obtained as an output—

जायरात.

जोन क्योक्साच्या वक्साचिं टस प्रकाशन, तें आसा
मानतात, पांच खंडांमेरेन पातळटले. विचार केल्लो
ताच्या इतिहासांतल्यान माळेची सुरवात करपाची सल्ला दितात
स्कॉटलंडांतली सुदारणा, सगळ्यांत व्हडलें काम म्हणून
म्हत्व आसा. अशे तरेन फुडल्या खंडांत...
तिसरें आनी चवथें पुस्तक, जें इतिहास ते...
१५६४ हें वर्स; ज्या काळांत ताचे इतिहासीक श्रम
समाप्त करपाचो मानू येता. पूण पांचवें पुस्तक,
इतिहासाचो सिकेल तयार करून, आनी ताचेखाला उजवाडाक हाडलो
१६४४ वर्सा ताचें नांव, ताचोय आस्पाव जातलो. ताचीं पत्रां
आनी विविध लेखनाची वेवस्था आसतली
उपरांतच्या खंडांनी, कालक्रमाप्रमाण शक्य तितले लागीं-
गिकल क्रम ; दरेक भाग वेगळ्या कडल्यान वळख करून दिवप
नोटीस, हातबरपाचो वा छापून आयिल्या प्रतींचो आदर करून
जीं तांकां घेतल्यांत.

लेखकाचें एक जिवीत अशी घडये अपेक्षा आसूं येता
ह्या खंडाक उपसर्ग आसूंक जाय आशिल्लो. द लायफ ऑफ
डॉ.एम'क्रिग हाचें नॉक्स हें पुस्तक मात इतलें सावंत्रिकपणान
वळखतात, आनी इतलें इतिहासीक मोलाचें, की ताचे सुवातेर उरता

सविस्तर जैव-

Fig 4.3.2: Translated Text from Image

4.4 VIDEO CAPTIONING

To test the video captioning capabilities, a test video clip was uploaded from the movie **Star Wars : Revenge of the Sith**.



Fig 4.4.1: Video Captions Generated

The system was able to caption the video, save for some parts where there was background music.

There is some time needed in audio transcription and segmentation. This video was generated using the “**Small**” model size of Whisper to transcribe the audio. For a **3 minute** video it takes around **4~5 minutes** to obtain the captioned video.

The obtained captions are satisfactory. A few instances may occur where some spoken audio is wrongly identified by whisper model. This is due to usage of smaller model sizes. For model sizes “**Medium**” and “**Large**”, the speech is accurately identified but enough memory is needed to run these models.

Also transcripts in both English and Konkani language are generated which are offered for download in a Zip file once the video is translated.

You will not take her from me!
Your anger and your lust for power have already done that.
You have allowed this dark lord to twist your mind until now.
Until now you have become the very thing you swore to destroy.
Don't lecture me Obi-Wan.
I see through the lies of the Jedi.
I do not fear the dark side as you do.
I have brought peace, freedom, justice and security to my new empire.
Your new empire?
Don't make me kill you.
Anakin, my allegiance is to the Republic, to democracy!
I want to say, if you're not with me, then you're my enemy.
Only a Sith deals in absolute so.
I will do what I must.
You will try.

Fig 4.4.2: English transcript

तुमी तिका म्हजे कडल्यान व्हरचे नात!
तुमचो राग आनी तुमच्या सत्ते खातीर कामवासना पयलींच तशें केल्या.
ह्या काळखांतल्या प्रभूचेर आतां मेरेन मन घुंवडावपाक मेळ्णां.
आतां मेरेन तुमी नश्ट करपाक सोपूत घेतल्या तीच गजाल जाल्या.
म्हाका ओबी-वान व्याख्यान दिवचे न्हय.
जेडीच्या फटांतल्यान म्हाका दिसता.
तुमी करतात तशें काळखांतल्या बाजूक हांव भियेना.
हांवे म्हज्या नव्या साम्राज्याक शांतताय, स्वातंत्र्य, न्याय आनी सुरक्षा हाडल्या.
तुमचे नवे साम्राज्य?
तुका मारूळक दिवचे न्हय.
अनाकिन, म्हजी निश्ठा प्रजासत्ताकाक, लोकशायेक!
म्हाका सांगपाचे आसा, तूं म्हजे वांगडा ना जाल्यार मागीर तूं म्हजो दुस्मान.
फक्त एक सिथ निरपेक्ष सो हातूंत वेव्हार करता.
हांव जाय तें करतलों.
तुमी प्रयत्न करतले.

Fig 4.4.3: Konkani transcript

4.5 DISPLAYING PREVIOUS TRANSLATIONS

One of the key features of this project is the ability for users to **view their previously translated content**. This functionality was implemented to enhance usability by maintaining a translation history for each logged-in user.

4.5.1 On Website

Following figures show the implementation for displaying previous translations.

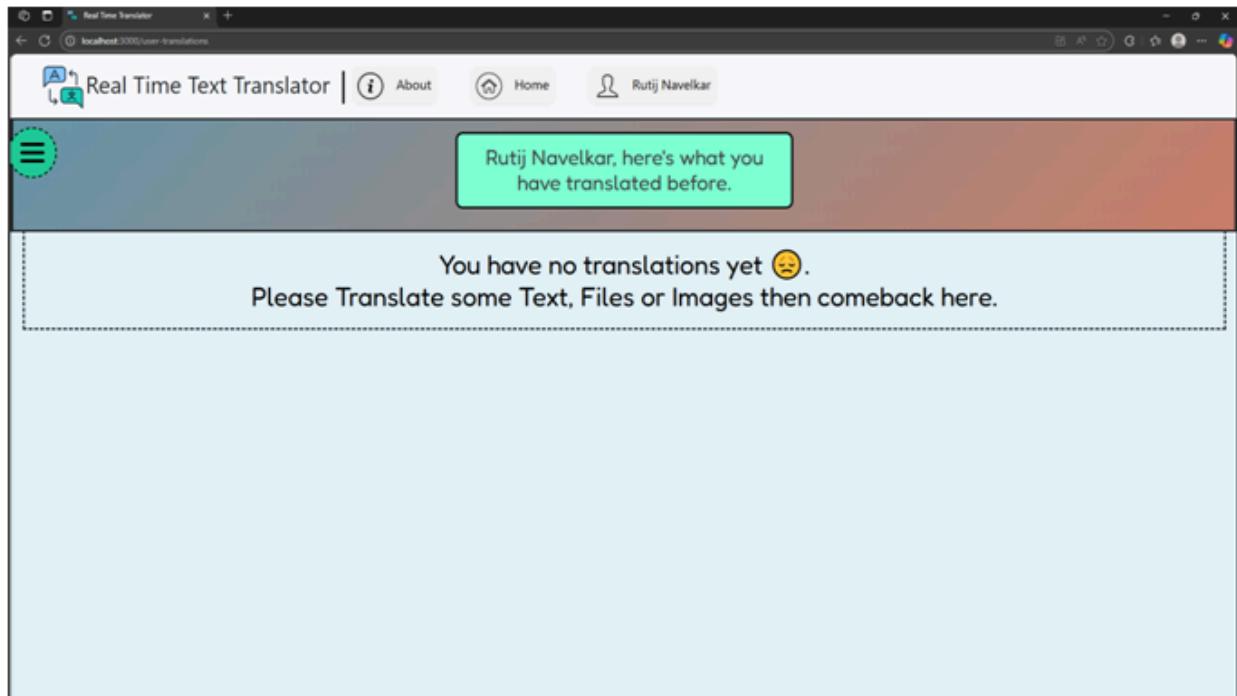


Fig 4.5.1.1: User has not performed any translation

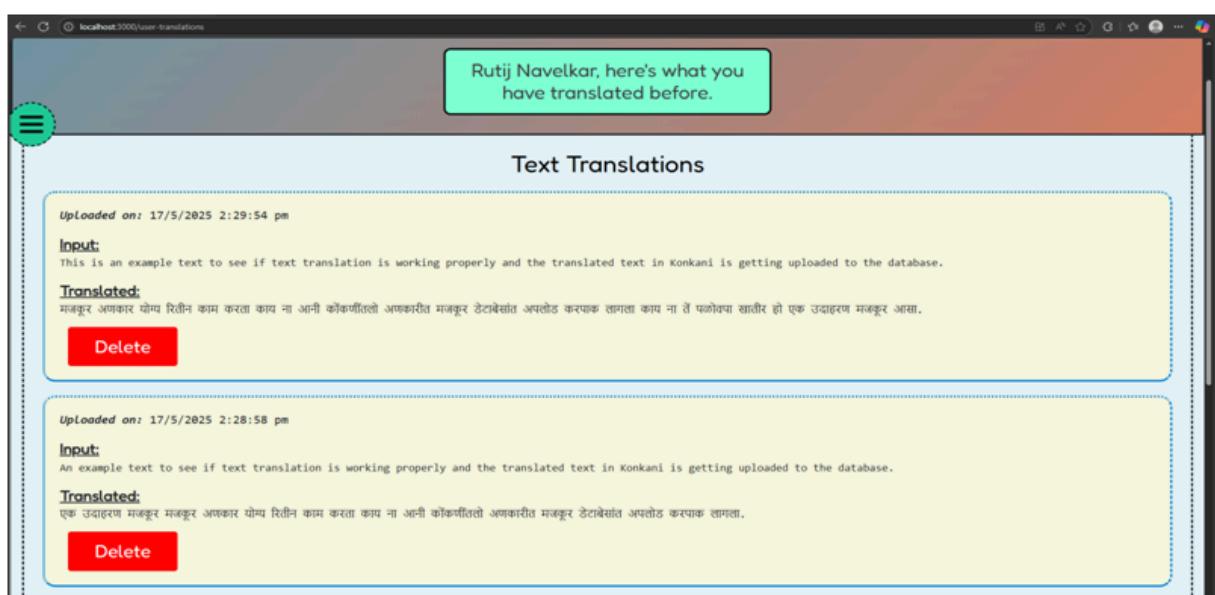


Fig 4.5.1.2: User has performed 2 Text Translations.

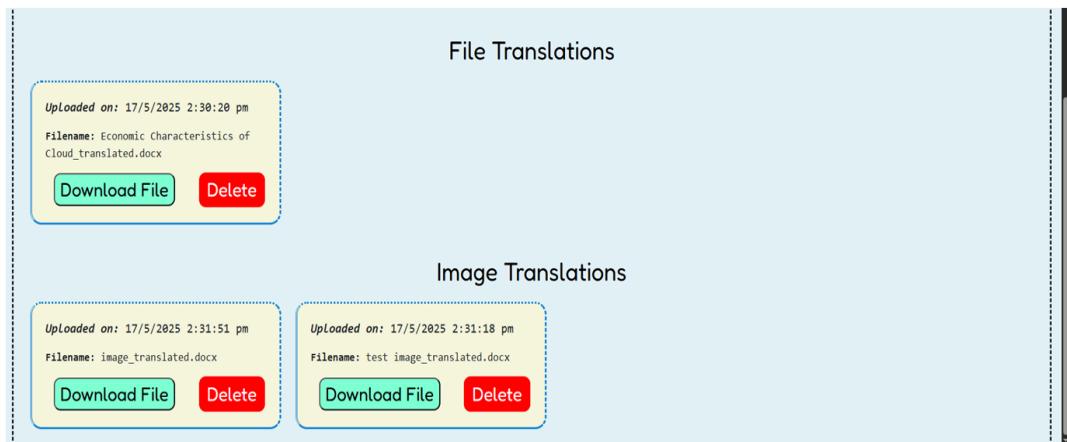


Fig 4.5.1.3: User has performed a File Translation and 2 Image Translations.

A side navigation button (**Green Circle with 3 Lines**) is also provided for users to navigate easily when there are multiple translations.

4.5.2 In the Database

Following figure displays a user and the translations generated by them stored on the Database.

	← T →	id	username	email	password
	<input type="checkbox"/> Edit Copy Delete	3	Rutij Navelkar	rnn@gmail.com	password

Fig 4.5.2.1: A registered user's credentials stored on database.

	← T →	id	user_id	type	input_text	translated_text	file_data	filename	created
	<input type="checkbox"/> Edit Copy Delete	15	3	image	NULL	NULL	[BLOB - 36.2 KiB]	image_translated.do	2025-05-17 14:31:51
	<input type="checkbox"/> Edit Copy Delete	19	3	file	NULL	NULL	[BLOB - 53.9 KiB]	Exploring the Mysteries of the Deep Ocean _transla...	2025-05-17 15:15:19
	<input type="checkbox"/> Edit Copy Delete	20	3	image	NULL	NULL	[BLOB - 53.9 KiB]	Advertisement_transla...	2025-05-17 15:16:11
	<input type="checkbox"/> Edit Copy Delete	22	3	file	NULL	NULL	[BLOB - 38.0 KiB]	Economic Characteristics of Cloud_translated.doc	2025-05-17 15:18:15
	<input type="checkbox"/> Edit Copy Delete	23	3	file	NULL	NULL	[BLOB - 53.9 KiB]	Open Source Tools_translated.doc	2025-05-17 19:30:51
	<input type="checkbox"/> Edit Copy Delete	25	3	text	Exploring the Mysteries of the Deep Ocean	ખોલ મહાસાગરચાય ગુરૂત્વાયેચો સોદ ઘેવય	NULL	NULL	2025-05-19 16:22:08
					The oce...	દર્શાત ...			
	<input type="checkbox"/> Edit Copy Delete	26	3	image	NULL	NULL	[BLOB - 36.2 KiB]	Atomic Habits_translated.doc	2025-05-19 16:44:05

Fig 4.5.2.1: Some of the translations that the user has performed.

TIMELINE

TIMELINE The following timeline outlines the major milestones and deadlines for the our project:

<u>TASKS</u>	<u>MONTH</u>
Video Captioning Improvements and Implementation of Previous Translation Viewing for Users.	April 2025
Testing and Fixing issues and other improvements.	April - May 2025

CONCLUSION

In conclusion, our project journey has followed a methodical path to meet its objectives. Beginning with defining our goals, we conducted a practical market study to gather valuable insights. The literature review provided a solid foundation by summarizing previous work in the field. Breaking down the project into manageable modules facilitated a well-organized development process, and prototyping allowed for continuous refinement. System requirements were carefully pinpointed and documented, setting the stage for a robust framework.

Furthermore, effective project management included the division of tasks among team members, ensuring seamless collaboration. The development timeline, a crucial outcome of this coordination, served as a practical guide aligning our efforts with key project milestones. In essence, each project phase has been executed with attention to detail, from initial planning to designing, showcasing a systematic approach that positions our Real-time Text Translation Tool for successful development and implementation, and achievement of its intended goals.

REFERENCES

	PAPERS	LINKS
1.	KOP: An opinion mining system in Konkani.	https://ieeexplore.ieee.org/document/7807914
2.	Konkani Script to Speech Conversion by Concatenation of recognized Hand written Konkani Text Using Neural Network.	https://ieeexplore.ieee.org/document/8821812
3.	Handwritten Devanagari numeral and vowel recognition using invariant moments.	https://ieeexplore.ieee.org/document/7955307/
4.	A Survey of Machine Translation Approaches for Konkani to English.	https://ieeexplore.ieee.org/document/9077842
5.	PARTS OF SPEECH TAGGING FOR KONKANI LANGUAGE.	https://ieeexplore.ieee.org/document/8487620
6.	Robust Speech Recognition via Large-Scale Weak Supervision	https://cdn.openai.com/papers/whisper.pdf