



Basic Details of the Team and Problem Statement

Ministry/Organization Name/Student Innovation:

Ministry of Commerce and Industries

PS Code: SIH1385

Problem Statement Title: Developing a software that can translate resource material and other texts from English to other Indian regional languages

Team Name: Team 404

Team Leader Name: Mustafa Masuldar

Institute Code (AISHE): C-15716

Institute Name: N B Navale Sinhgad College of Engineering , Solapur.

Theme Name: Smart Education

Idea/Approach Details

Vision:

Our vision is to create a unified platform where language is no longer a barrier to understanding Intellectual Property Rights. We aim to empower people across India by providing seamless translation of IPR resources. Through our solution, we envision a society where knowledge flows freely, fostering innovation and collaboration, irrespective of linguistic differences

1. User Input:

Users can type in English or upload documents directly on our website. This provides flexibility for users to input their content in a way that's convenient for them.

2. Clear Images (OpenCV):

We use OpenCV, a computer vision library, to enhance image clarity. OpenCV sharpens images, making the text within them clearer and improving the accuracy of text extraction.

3. Text Extraction (Tesseract OCR):

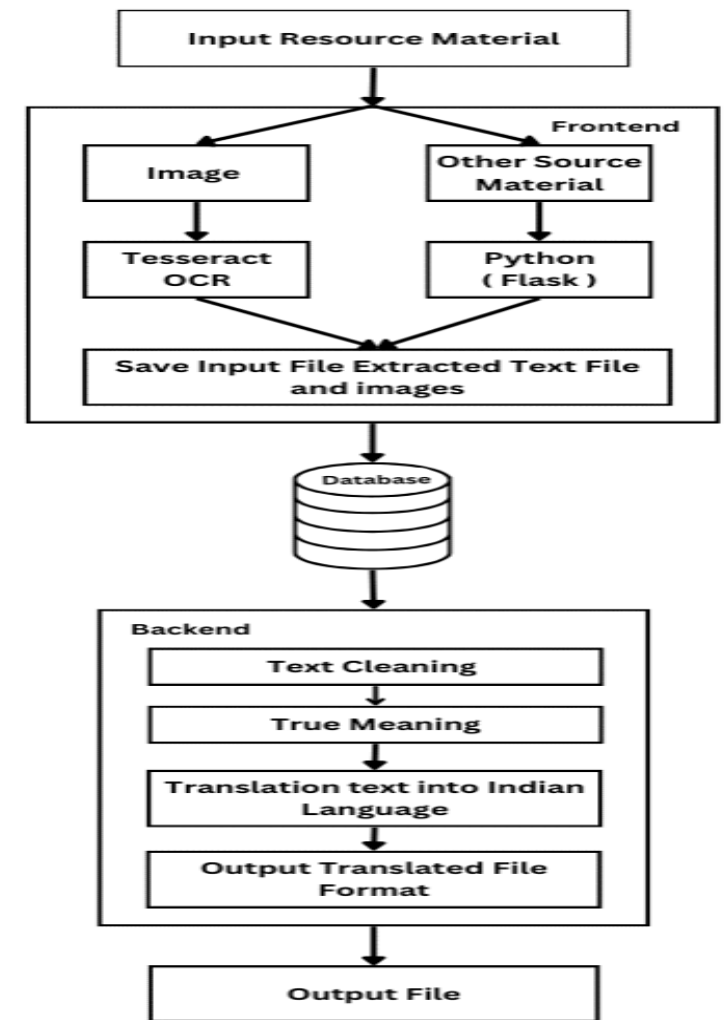
For scanned images or documents, we employ Tesseract OCR (Optical Character Recognition) technology. Tesseract converts images into editable text, enabling seamless processing of scanned content.

4. Save Your Work (Database Integration):

Our system integrates with a secure database where users can store both their original inputs and translated outputs. This feature allows users to retrieve their work later for reference.

5. Correct Spelling:

We implement spelling correction algorithms that automatically identify and rectify spelling errors in the extracted text. These algorithms ensure the accuracy of the content being translated.



Technology stack:

- Translation API+Flask(Framework)
- Tesseract OCR
- NLP + Python libraries (PyPDF2, Gramformer, Translator ,cv2)
- HTML, CSS, JS (Frontend) , PHP

Idea/Approach Details

6. Fix Grammer (Gramformer Library):

Gramformer a powerful library, is used to check and correct grammar errors in the text .This step ensures that the translated content is not only grammatically correct but also reads naturally and fluently

7. Translate Correctly:

We employ specialized translation algorithms that take the corrected text and translate it into the user's preferred language. These algorithms consider context, idiomatic expressions, and cultural nuances, delivering accurate and meaningful translations.

8. Quality Check:

A combination of automated checks and human review ensures the translation's quality. Algorithms verify linguistic accuracy, ensuring that the translated content is faithful to the original meaning and Intent.

9. Choose Your Format:

Users have the flexibility to receive the translated text in various formats such as Word or PDF. This choice accommodates different user preferences and requirements.

❑ Features:

Multi-Modal Input:

Supports text, document uploads, and speech recognition for diverse user inputs..

Custom Terminology Support:

Allows users to add domain-specific terms for accurate translations in specialized fields.

Use Cases:

- **Students' Education:** Enable translation of educational materials from English to Indian languages, enhancing accessibility for students.
- **Government Services:** Translate official announcements and documents into regional languages, ensuring vital information for citizens.
- **CIPAM's Resources:** Translate educational tools, including images and PDFs, into Indian regional languages for broader reach.
- **Business Communication:** Provide real-time translation for effective communication between individuals and businesses speaking different languages.

❑ Dependencies:

- Source Material
- Development of a machine translation library
- Access to a diverse and large dataset for training the translation models
- Development of an intuitive and user-friendly interface for users to input text and receive translations .

Team Member Details

Team Leader Name : Mustafa Masuldar

Branch : Btech

Stream : CSE

Year : IV

Team Member 1 Name : Sakshi Deshmukh

Branch : Btech

Stream : CSE

Year : IV

Team Member 2 Name : Pooja Pawar

Branch : Btech

Stream : CSE

Year : IV

Team Member 3 Name : Nandini Madre

Branch : Btech

Stream : CSE

Year : IV

Team Member 4 Name : Iram Muchale

Branch : Btech

Stream : CSE

Year : IV

Team Member 5 Name: Nikita Jagdale

Branch : Btech

Stream : CSE

Year : III