TEAM HALAQAT
SIH 2022

NLP Based Educational Translation Website

TEAM MEMBERS

1. Hari Prasad
2. Saravanan
3. Revanth Ramesh
4. Kabitha
5. Priyanka
6. Sharumathi

Contents

National Education Policy 2020 :

Early childhood care and education: Foundation of learning Curtailing dropout rates and ensuring universal access to education at all levels Optimal learning environments and support for students
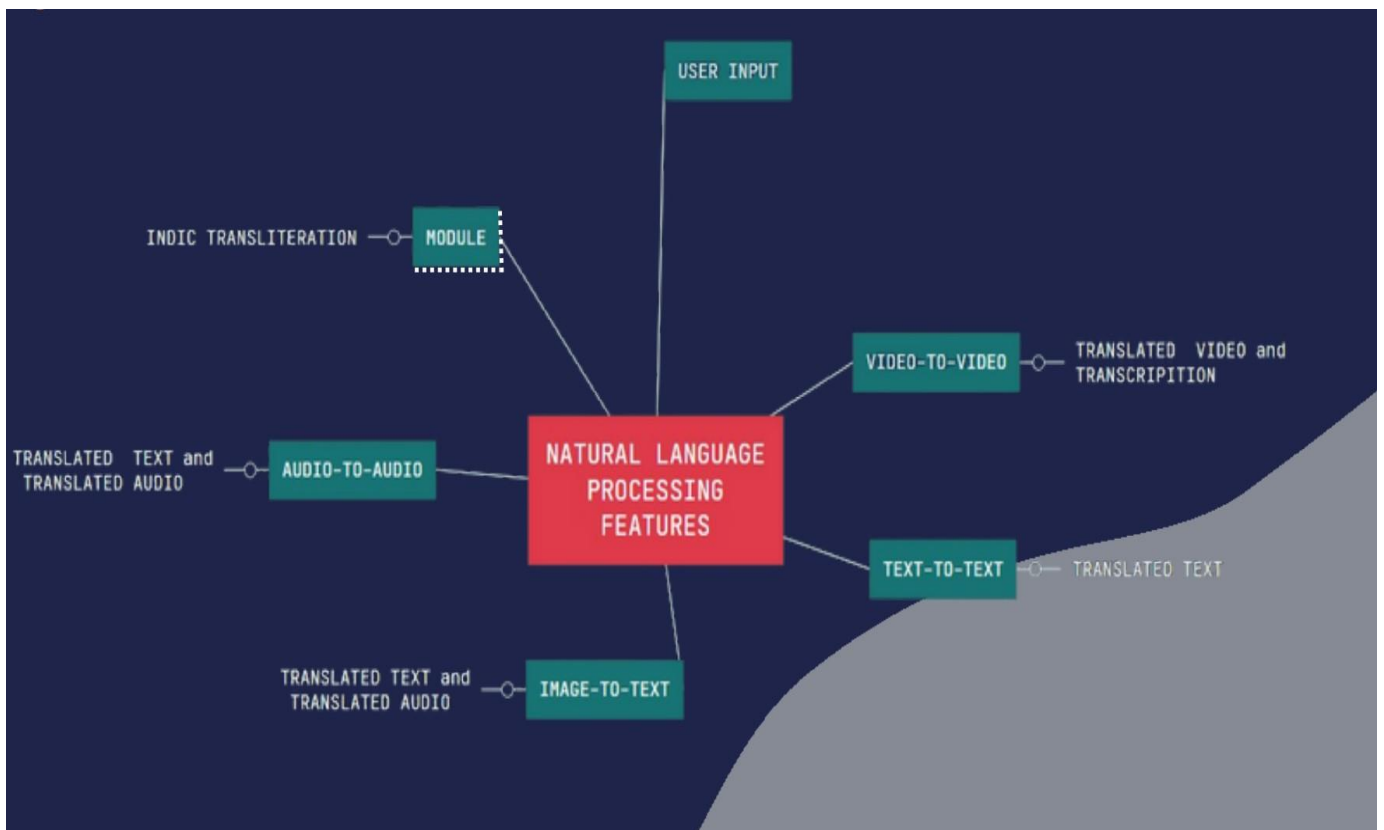
Problem Statement :

Traditionally, lot of challenges are faced by vernacular-medium students as compared to English-medium students, in terms of learning, opportunities and overall personality development. Hence, inline with the new National Education Policy 2020 promulgated by GoI, it is need of the hour to bridge the above gaps faced by Indian students using disruptive technology solutions.

Proposed Solution :

We have developed a web application using NLP (Natural Language Processing) technology for mass effective conceptual learning and education in vernacular languages (Mother Tongue), which consists of some modules such as,

- Text to Text ,
- Audio to Audio,
- Video to Video ,
- Image to Text translation in vernacular languages.

System Flow:

Software Specification :

- Front End - HTML/CSS
- Framework - Flask, Bootstrap
- Programming Language – Python

Technologies

NLLB - is a open source AI Model which can directly translate between 200 languages including lowresource languages like Asturian, Luganda, Urdu and more

Speech Recognition - it incorporates computer science and linguistics to identify spoken words and converts them into text. Technologies gTTS - a Python library to interface with Google Translate's text-to-speech API. It features flexible preprocessing and tokenizing

gTTS - a Python library to interface with Google Translate's text-to-speech API. It features flexible preprocessing and tokenizing

PyDub - it is a Python library to work with " .wav" files. By using this library we can play, split, merge, edit.

MoviePy - it is a module for video editing, which can be used for video compositing , video processing. It can read and write the most common video formats.

Pytesseract - is an OCR tool. It can read and recognize text in images and is commonly used in python ocr image to text use cases.