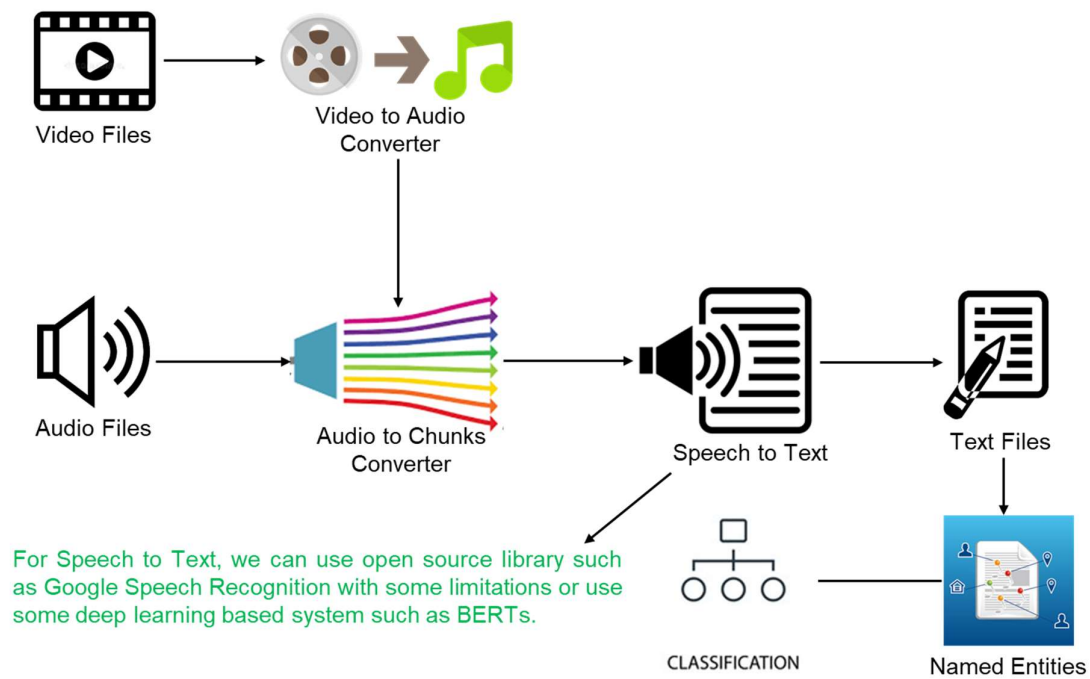# Text Extraction from Video and Audio Files and Classification

Aim of the Project

Aim of this project is to extract text from video and audio files and classify text into different categories using Natural Language Processing (NLP) and Machine Learning (ML) techniques. This project consist of two parts; one the text extraction and other is classification.

Proposed System

For Speech to Text, we can use open source library such as Google Speech Recognition with some limitations or use some deep learning based system such as BERTs.

Note: Most of the components will be open source

Main Components of the System

- Video Converter

In this project we are interested in text/ spoken words in videos, hence we need to extract audio from video first and then text from audio. To convert video into audio we need some converter/ engine. Here we will write module using MOVIEPY open source library.

MoviePy is a Python module for video editing, which can be used for basic operations (like cuts, concatenations, title insertions), video compositing (a.k.a. non-linear editing), video processing, or to create advanced effects. It can read and write the most common video formats, including GIF.

pip install moviepy

- Audio Converter (to WAV file)

Most of the libraries work only with WAV file format of audio. We need some engine to convert any file which is audio (in other formats such as .auc, mp3 etc.) to WAV format.

Waveform Audio File Format is an audio file format standard, developed by IBM and Microsoft, for storing an audio bit stream on PCs. It is the main format used on Microsoft Windows systems for uncompressed audio. The usual bit stream encoding is the linear pulse-code modulation format.

For this we need FFMPEG library.

FFmpeg is a free and open-source software project consisting of a suite of libraries and programs for handling video, audio, and other multimedia files and streams. At its core is the command-line ffmpeg tool itself, designed for processing of video and audio files.

- Audio Splitter

For large size audio files, most of the open source didn't allow us to use large files. Files must be of small size. In this system, we need audio splitter module, which will split audio into smaller chunks if of large size. Smaller chunks will be then used to extract text from.

- Speech to Text Converter

Speech to text is a speech recognition software that enables the recognition and translation of spoken language into text through computational linguistics.
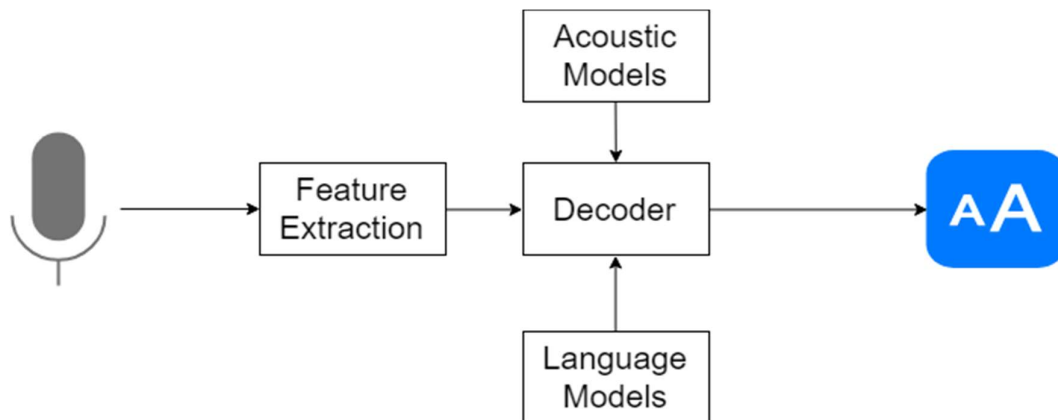
pip install SpeechRecognition

- Named Entities Recognition

Named-entity recognition is a subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.
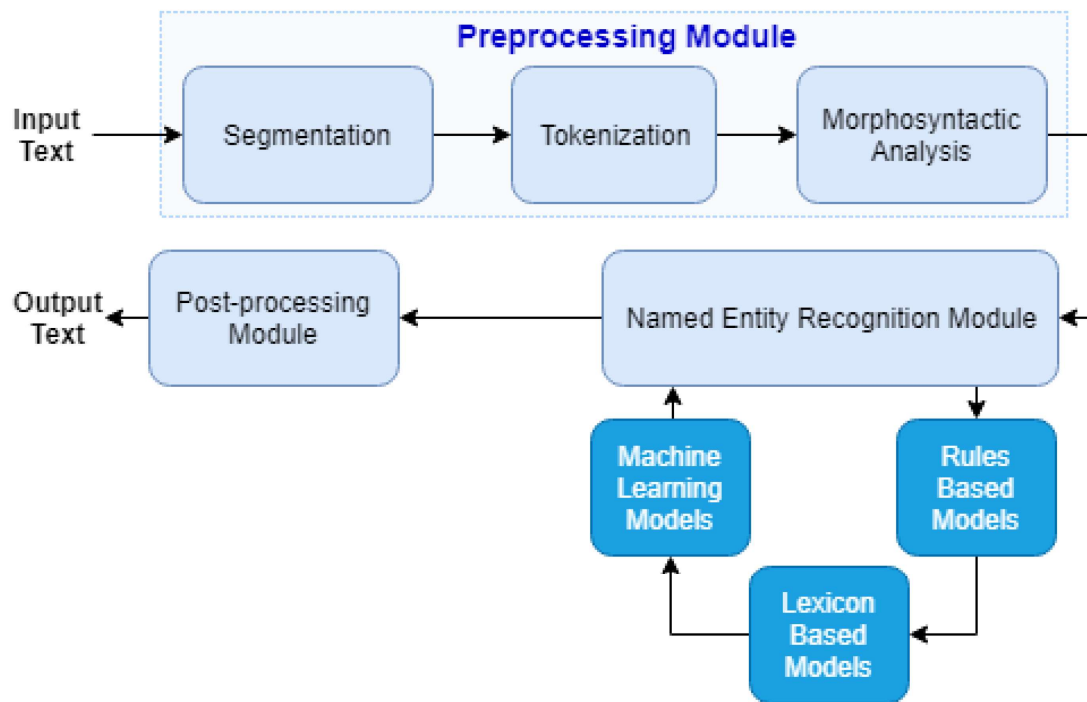
- Text Classifier

In the last step we need some classifier depends on the type of problem. Let say we can do sentiment analysis on the text, we can do classification of text into categories.
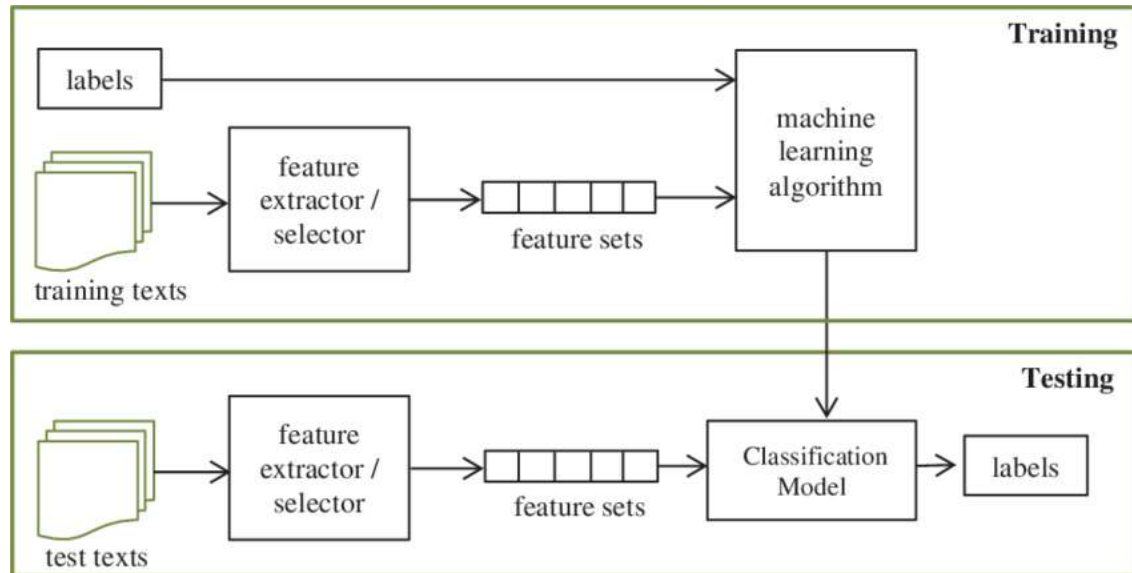
Technical Requirements

- Computer System
  - Usually fast computer system because system will process large files
- Internet Connection
  - This system will be need internet connection as speech to text will be online process in case of open source
- Machine Learning Algorithms
  - Probabilistic models for text categorization
  - Naïve Bayes or any other machine learning model
- Python Libraries
  - speech_recognition or any other available library
  - moviepy
  - NLTK
  - Spacy
  - Scikit-learn
  - ffmpeg for audio and video
  - VADER for sentiment analysis