

Automated Notes Maker from Audio Recordings

Dr. Saurabh Goel (Professor), Pratham, Gaurav, Gautam
(Students – CSE, Panipat Institute of Engineering and Technology, Samalkha, Haryana)

Date of Submission: 10-12-2022

Date of Acceptance: 20-12-2022

ABSTRACT:

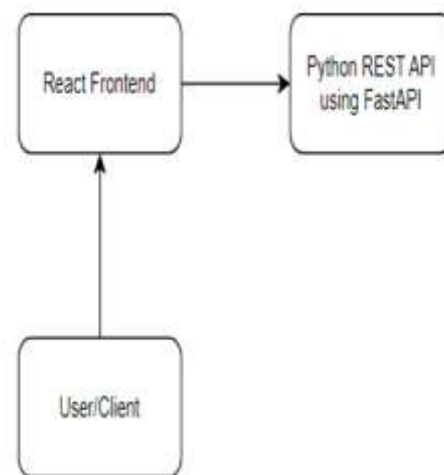
Today millions of students learn online from video tutorials which cover the theory as well as the practical part of their subjects. Similarly, most of the important news is delivered online in the form of videos and audios. Not only that, but many types of media content and online content is also delivered in the form of audio. But for various purposes, there is a need of a system which is used to convert these audio as text in various document formats. The system will also be able to summarize the audio clips and generate keywords. Also text translation and statistics of text should be provided by the system. Our system is an all in one website which includes all of these functionalities to be used.

Keywords: Speech-to-text, Machine Learning, Artificial Intelligence, Google Translation API, REST, ReactJS.

I. INTRODUCTION:

Today millions of people has access to the audio and video recordings of their work. Primarily students, politicians, journalists, government agents etc has their data in the form of audio. This data is of a great importance to the users but in the cases where these audios need to be searched or analysed they must be in a suitable text format. We have addressed this problem and developed a unique system which not only converts the audio to text notes, but is also helpful in summarizing and analysing the text. The system is a web application which at frontend used Javascript (React – taken reference from the book by Robin Weiruch (Ref. II)), but at the backend uses Python (taken reference from the book Head First Python (Ref. I)) and some popular speech to text services.

Working:



The project's design is based upon presentational and data access architecture. In other words it is based upon frontend and backend architecture. For frontend ReactJS Library (Ref. II) is used to write modular and reusable code. The working from the user's side as well as the behind the scene details is described stepwise below:

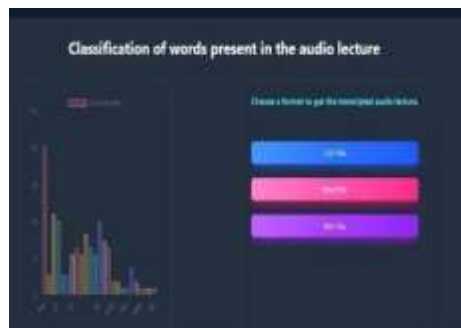
Step 1: The user first enters to the portal through the portal's URL.

Step 2: Then the user has to log-in. In case the user is not registered, the user has to register himself first. In this step the Webpage displays a form which the user needs to fill-in in order to login or sign-up. The user's details are captured by the form and then an API is called to the backend which returns positive or negative result based on the information entered by the user. If the details entered were correct, then the user is redirected to the dashboard.

Step 3: The user has options for different functions on the left navigation panel. These options are:

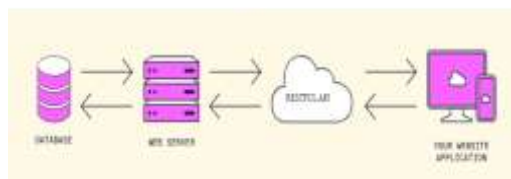
- (i) Dashboard: The main task of conversion of audio to notes is performed on this page. The user has to upload the audio file in the file upload section. After uploading, the user is

shown a popup, which shows the statistics of the text and displays the options to download the text file in 3 different formats: .docx, .txt, .pdf. When the user uploads the file the file is uploaded through javascript and then sent to the backend through the api. The backend then has the controller function which uses the Google's text-to-speech API to perform the conversion. To analyse the text, the nltk library is used. The results are sent back by the api as a json response.



(ii) Translation: This page has a text input field, in which the user enters the text to be translated. The text is then sent to the translation API, the API uses an external service to perform the translation. For our purpose, we have implemented English to Hindi translation and vice versa.

The easy implementation is made possible by using the REST Architecture (Ref. III –Clean Code).



Benefits: This system has a lot of benefits as listed below:

- They save time and money. The user get do not have to jump from website to website or form page to page.
- They result are instant.
- The system runs on its own and produces desired and good results.
- It becomes easier to maintain record of relevant data.
- The user can do multiple tasks like translation, conversion and get statistics on a single place.
- The system is safe and fast as it uses trusted services and APIs (Ref. I and Ref. III).

Need of this system:

In today's world audio and video media is really important to communicate whether it is in the field of education or it is in the field of journalism or it is politics or any other field. The quantity of these audio and video media increases day by day and it becomes more difficult and difficult to track changes, search and analysis in the media which is in the form of video and audio. So a system was surely needed which could convert these type of media into a suitable format which can be then used to compare, analyse and search within the media files. So this system which converts these audio files into notes in the form of text data in various formats is really important. However, the project can also serve an another purpose of generating notes which can be useful for a user to have a quick revision or recap of the audio.

Limitation:

- User can only get statistics of the text data. He is not able to compare or analyse the results of different conversions.
- We cannot rely completely on the translated or converted text because the machine learning

models used behind the scenes are not 100% accurate.

- Two factor authorization is not available and user verification upon registration is not implemented.
- The user's personal data might be used to train the models used behind the scenes which can be a risk.
- Few of the libraries which are used are not of the latest versions, which mean that some of the functionalities might deprecate in the future.
- Tech giant companies sell this data in the market, which is another invasion of privacy.

Future scope: The automated notes maker system will be more personalized and the conversion and translation will be more accurate as the used machine learning models get more and more trained.

II. CONCLUSION:

Audio to text conversion systems are great which can help the users in analysis and comparison of audio and video media. These systems are also capable of generating the text in proper formats as notes. The accuracy and personalization of these systems can be improved in future versions.

REFERENCES:

- [1]. Paul Berry: Head First Python (Second Edition)
- [2]. Robin Weiruch: The Road to Learn React: Your Journey to Master Plain Yet Pragmatic React.js.
- [3]. Steve McConnell: Code Complete (Second Edition)
- [4]. Speech to text conversion and summarization for effective understanding and documentation (https://www.researchgate.net/publication/342147736_Speech_to_text_conversion_and_summarization_for_effective_understanding_and_documentation)
- [5]. Speech to text and text to speech recognition systems (<https://www.igntu.ac.in/eContent/IGNTU-eContent-815947141046-MA-Linguistics-4-HarjitSingh-ComputationalLinguistics-5.pdf>)