Proceedings of the 16th INDIACom; INDIACom-2022; IEEE Conference ID: 54597
2022 9th International Conference on "Computing for Sustainable Global Development", 23rd - 25th March, 2022
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Summary Generation using NLP Techniques

Sweta Gupta
Dept. of Information Technology
Vidyavardhini's College of
Engineering & Technology
Mumbai, India
shwetagupta22031607@vcet.edu.in

Yash Jobalia
Dept. of Information Technology
Vidyavardhini's College of
Engineering & Technology
Mumbai, India
jobaliayash@gmail.com

Isheet Shetty
Dept. of Information Technology
Vidyavardhini's College of
Engineering & Technology
Mumbai, India
isheetshetty@gmail.com

Anagha Patil
Dept. of Information Technology
Vidyavardhini's College of
Engineering & Technology
Mumbai, India
anagha.patil@vcet.edu.in

***Abstract*—Due of the large amount of information and electronic documents available on the internet, it is impossible for a human to study, examine, and analyze this material. The main concept because the core idea of summarization allows humans to swiftly read the summary of a large amount of content and makes a decision whether or not to go deeper into the intricacies. Detailed summaries create a summary by extracting a certain collection of sentences from a document in which NLP approaches like Genism, Text rank, Rule based and Pegasus can be used on a single document or several documents. Thereupon, synopsis generation is the charge of generating incisive summaries unescorted by human assistance while conserving the genuine sense of the overlong document.**

***Keywords*—Speech to Text, Text Processing, Encoding, Summarization, Pegasus, Genism, Rule Based, Text rank, Decoding**

I. Introduction

The human attention span is less than 8 seconds so, if you need to capture someone's attention or highlight an important topic, you need to have a strong headline or summary. A prime example of this phenomenon is that whenever you open a newspaper you glance through the headlines and read through only the ones which have captivating headlines or an interest which aligns with yours. Hence, in this ever growing and vastly expanding world with abundance of data, you need to make sure that the important topics gain your priority attention and you get precise information and knowledge from the vast abyss of data. Information is knowledge and knowledge is power. With the huge increase in the amount of data available the fine distinction between the knowledge information and data is thinning. The motive is to use the huge amount of data provided, abstract the required information and try to highlight emphasize on the necessary knowledge that it contains. Moreover no one has the luxury of spending their precious time on reading reports which are more than a few pages long just of the regular day to day meeting or a general conversation. Hence in order to save your valuable time the meeting summarizer can summarize the entire meeting into a few paragraphs and also highlight the entire gist of the content in a few lines. If the user receives an excellent summary, he or she will be able to grasp the text at a glance without having to read it completely, saving time and effort. Analysis, transformation, and synthesis [1][2] are the three processes in the text summarization process. The project's purpose is to learn about natural language processing concepts and construct a text summary machine learning tool that only incorporates the most important information from the material. The process takes place as follows. We first record the conversation or the meeting from a voice recorder on a device like phone or a mic then we use highly sophisticated STT (speech To Text) algorithms which recognise speech and convert them to text as our first input. These algorithms have the capacity to convert huge audio files with large number of recorded minutes into smaller chunks of audio files slashed at regular intervals of time in order to reduce discrepancy and for better time complexity and management with better results. This converted text file will be then encoded and decoded by algorithms by either extractive summarisation or abstractive summarisation based on the user's requirements in order to convert to convert the lengthy conversation or meetings into small summaries of the entire topic containing of the most important topics and covering the gist of the text.

II. Motivation

The amount of data these days is exponentially increasing through the internet and various other sources. To avoid browsing through these overutilized and long-drawn-out documents and converting them to succinct summary, we stand in need of a tool that can help withdraw summary by clipping of the data in these documents and giving the foremost sentences with pivotal meaning from the prolix document or from a cluster of documents. It is difficult for the human mind to reminisce all this data, so a synopsis generator plays a pivotal role to save the human effort and time. Our work aims to build a synopsis generator that provides the user the liberty to make a selection from n number of summarization methods in accordance with their needs. Furthermore humans are highly intellectual species who can communicate through sophisticated measures and in

HEAD
Dept. of Information Technology