IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10528

An Automated Text Summary Generator of a Video

Anmol Shah[1], Amit Rane[2], Rimpa Sarak[3]

Student, B.E. Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, India[1-3]

Abstract: In today's world where there is tremendous amount of information available on the internet, it is really important to perceive right information quickly and efficiently. The answer to this problem is an "Automated text Summarization". As it can be done manually as well but, there is a chance that humans can be biased during the process and another disadvantage of doing it manually is that it can be very time consuming. It will also help us to extract the most important information from a large amount of data which is going to achieved using Natural Language processing algorithm. We have made our project using Python as a programming language along with React for the front end.

Keywords: Summarization, Natural Language processing, Python, React.

I. INTRODUCTION

Before, actually knowing how text summarization is done, we actually need to understand what Text summarization or a summary is. A summary is a basically a reduced text information extracted from large data. The important advantage of creating a summary is that it reduces reading time, so it's an efficient way of perceiving information. There are two ways of generating a summary, first one is Extractive summarization where the important sentences, words or paragraphs are taken into account from the original document. Second one is called the Abstractive method where you understand the important concept of the related topic and then only use those concepts in the summary. We have used the first approach in our project. There are also two different groups of text summary firstly Inductive where you get 10% of the original data and second one is Informative which gives concise information of about 20% to 30% of the main text.

II. LITERATURE SURVEY

Chin-Yew Lin. In this paper author introduced Recall Oriented Understudy for Gisting Evaluation ROUGE. That is an automatic evaluation package for text summarization. The paper also introduced four different measures of ROUGE: - ROUGE-N, ROUGE-L, ROUGE-W and ROUGE-S. It measures the quality of summary by comparing the generated summary with other ideal summaries that are created by humans. Akshil Kumar et al. In this paper author has analyze and compared the performance of three different algorithms. Firstly, the different text summarization techniques explained. Extraction based techniques are used to extract important keywords to be included in the summary. For comparison three comparison three keyword extraction algorithms namely TextRank, LexRank, Latent Semantic Analysis (LSA) were used. Three algorithms are explained and implemented in python language. The ROUGE 1 is used to evaluate the effectiveness of the extracted keywords. In the end, the TextRank Algorithm gives a better result than other two algorithms.

III. PROPOSED SYSTEM AND IMPLEMENTATION

The proposed Real-Time Scorecard system is meant to ease the process of score-keeping and broadcasting these scores to people all around the globe. A wireframe prototype of the system is shown in Fig 3.1. Text summarization of video lectures means to convey the video information as a short summary in text form. The main objective still remains to convey accurate information of video content in the form of summary. There are many strategies to generate summary but in the considered model an Abstractive approach to get summary is used. Before moving on to using NLP techniques to get summary we must do the following steps to get the entire video content in form of text.

1. Get audio from video

2. Split audio into smaller audio segments for easy computation 3. Convert audio to text

For step1 and step2 we use Ffmpeg (Fast Forward mpeg). Ffmpeg is a freely available opensource project add-on which mainly focuses on working with multimedia that includes audio and video files. We use a command line query that extracts the audio from the video file that is passed in the query. This audio file of the video
lecture is passed inside an ffmpeg query to split it into smaller audio segments each being serially ordered and safely saved inside the audio folder on the server system. The output of step2 is now considered as the input for step3. The acquired audio files are converted into text format using the python speech recognizer module. This module uses Google API client library which allows fast and accurate transcription of audio to text. The sequentially ordered audio files are one by one passed through the recognizer and finally all text output are appended together to get the final text file for the contents of video.

The Next phase focuses on generating a text summary as an output of the input text that is obtained from phase 2. This phase mentions the major steps taken in the algorithm of summarization. SpaCy library is used.
SpaCy is a free, open-source advanced natural language processing library, written in the programming language Python.

1. Word Tokenization, leaning and word Frequency Generation This is the first step of summarization. Here first tokenize all the words (make a list of all the words). If looked clearly in the list then we will find that list has two unnecessary elements which won’t
add much value to the overall meaning of content i.e. 1) Punctuation
2) Stop words (most commonly used words like a, an, the etc). So, iterate through the list and remove punctuations & stop words and at the same time we count the frequency of each word is stored in

2. Normalize Frequency
Here normalized frequency, for this find maximum value from all the values and divide the maximum value throughout

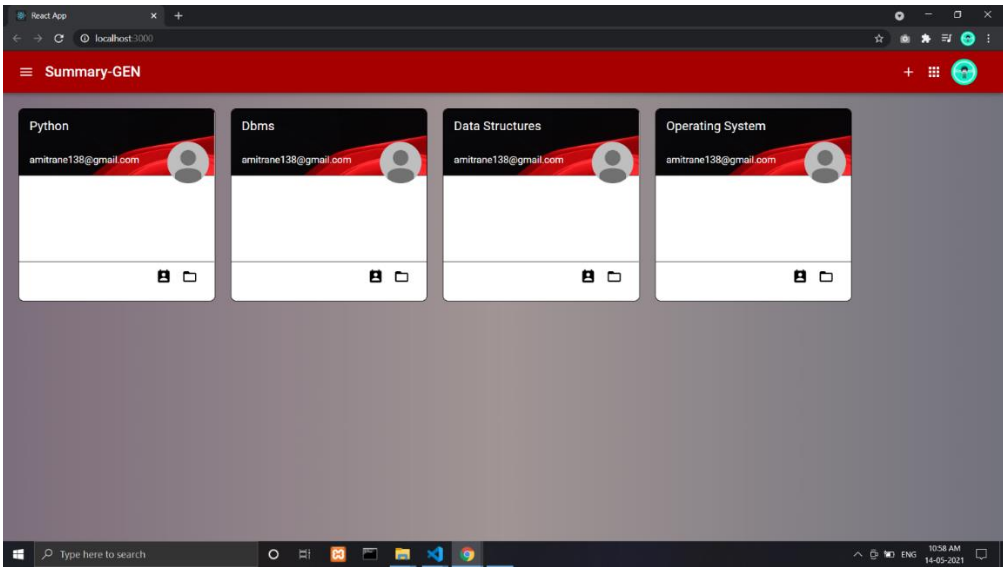
3. Sentence Tokenization and Sentence Scoring
So now we need to calculate scores of each sentence in the data to generate an optimal summary of the given data. For that first tokenize (make list of all sentence) Now, having generated a list of sentences and also a list of words. Generate a score for each sentence by adding the normalized frequencies of the word that occur in a particular sentence and store it in dictionary format.

4. Sentence Selection and Generating Summary Now, select the top N sentences by passing our sentence scores values to the n-largest function. Finally, we will obtain the desired summary

IV. RESULT

By using Extractive method and NLP algorithm, we had generated about 30% summary of the whole content and it has about 80% efficiency. It just takes few minutes to generate the summary from the time of uploading of the file which is very fast even though it involves so many steps before the actual summarization starts.

HOME PAGE:



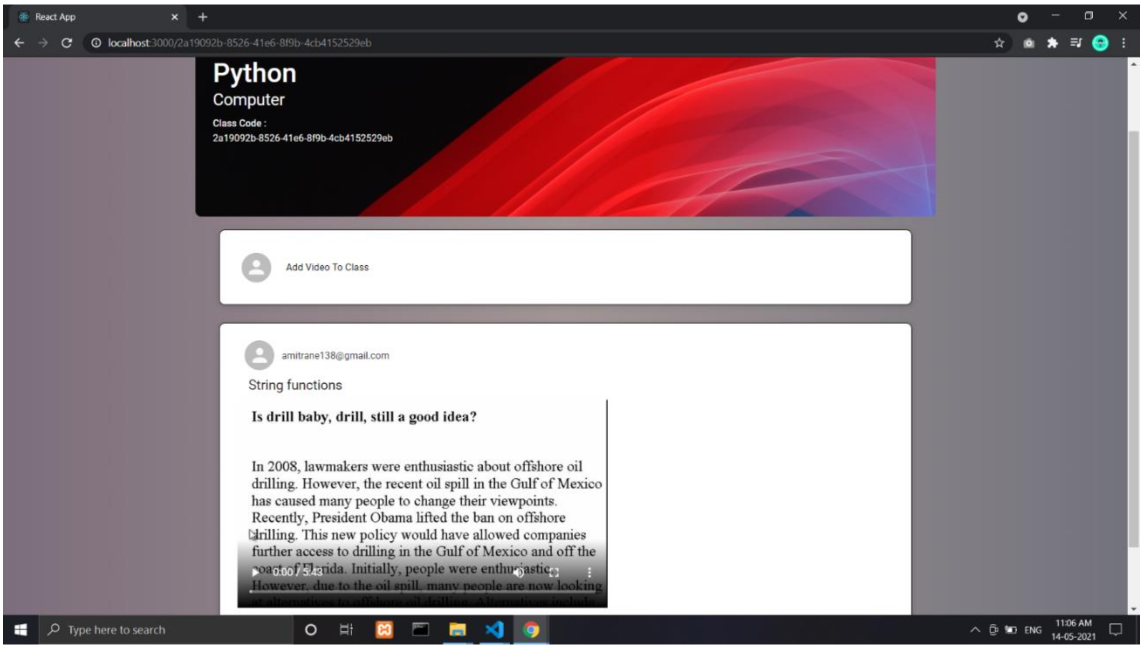
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10528

DIFFEFRENT SUMMARY ROOM:



Python

Computer

Class Code :
2a19092b-8526-41e6-8f9b-4cb4152529eb

Add Video To Class

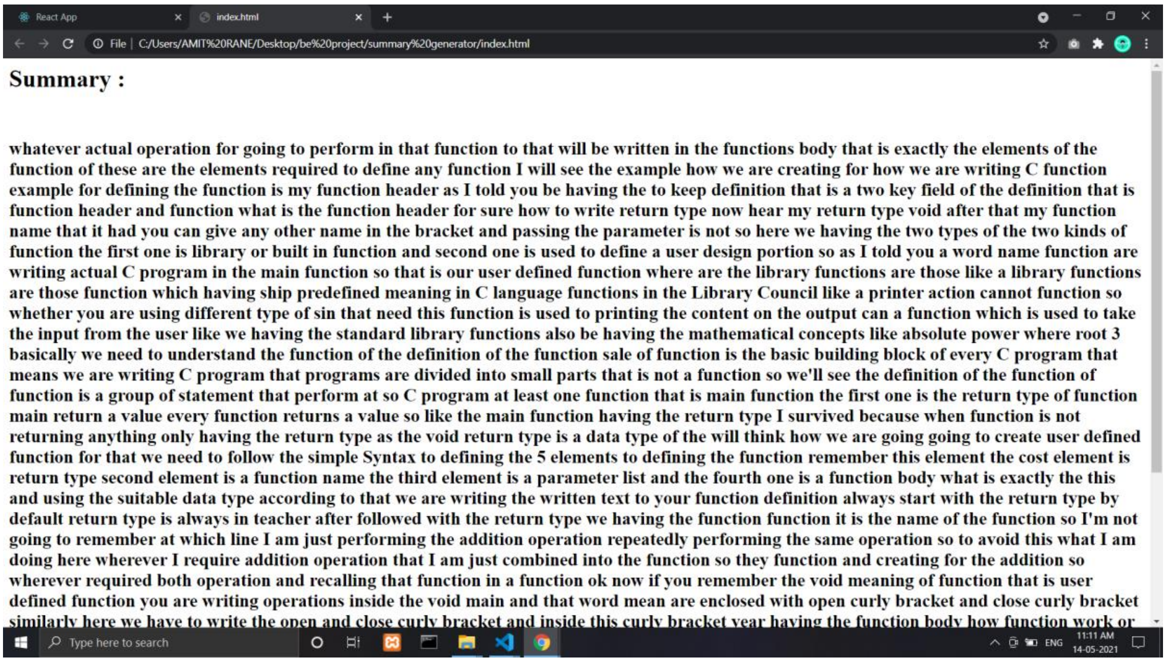
amitrane138@gmail.com

String functions

Is drill baby, drill, still a good idea?

In 2008, lawmakers were enthusiastic about offshore oil drilling. However, the recent oil spill in the Gulf of Mexico has caused many people to change their viewpoints. Recently, President Obama lifted the ban on offshore drilling. This new policy would have allowed companies further access to drilling in the Gulf of Mexico and off the coast of Florida. Initially, people were enthusiastic. However, due to the oil spill, many people are now looking

SUMMARY GENERATED:



React App
index.html
File | C:/Users/AMIT%20RANE/Desktop/be%20project/summary%20generator/index.html
Type here to search
ENG
11:11 AM
14-05-2021

Summary :

whatever actual operation for going to perform in that function to that will be written in the functions body that is exactly the elements of the function of these are the elements required to define any function I will see the example how we are creating for how we are writing C function example for defining the function is my function header as I told you be having the to keep definition that is a two key field of the definition that is function header and function what is the function header for sure how to write return type now hear my return type void after that my function name that it had you can give any other name in the bracket and passing the parameter is not so here we having the two types of the two kinds of function the first one is library or built in function and second one is used to define a user design portion so as I told you a word name function are writing actual C program in the main function so that is our user defined function where are the library functions are those like a library functions are those function which having ship predefined meaning in C language functions in the Library Council like a printer action cannot function so whether you are using different type of sin that need this function is used to printing the content on the output can a function which is used to take the input from the user like we having the standard library functions also be having the mathematical concepts like absolute power where root 3 basically we need to understand the function of the definition of the function sale of function is the basic building block of every C program that means we are writing C program that programs are divided into small parts that is not a function so we'll see the definition of the function of function is a group of statement that perform at so C program at least one function that is main function the first one is the return type of function main return a value every function returns a value so like the main function having the return type I survived because when function is not returning anything only having the return type as the void return type is a data type of the will think how we are going going to create user defined function for that we need to follow the simple Syntax to defining the 5 elements to defining the function remember this element the cost element is return type second element is a function name the third element is a parameter list and the fourth one is a function body what is exactly the this and using the suitable data type according to that we are writing the written text to your function definition always start with the return type by default return type is always in teacher after followed with the return type we having the function function it is the name of the function so I'm not going to remember at which line I am just performing the addition operation repeatedly performing the same operation so to avoid this what I am doing here wherever I require addition operation that I am just combined into the function so they function and creating for the addition so wherever required both operation and recalling that function in a function ok now if you remember the void meaning of function that is user defined function you are writing operations inside the void main and that word mean are enclosed with open curly bracket and close curly bracket similarly here we have to write the open and close curly bracket and inside this curly bracket year having the function body how function work or

V. CONCLUSION

We have made a basic automatic text summarizer using nltk library using python and it is working on small documents. We have used extractive approach to do text summarization. The limitation is that if the video is of larger size, it will take more time for text summarization. Also, if the audio is not proper in the video, then the summarization may miss some of the words that might be related to the topic. It is still pretty efficient than most of the automated system available.

VI. REFERENCES

(I)https://www.researchgate.net/publication/278661954_Auto matic_Text_Summarization_Past_Present_and_Future
(II)https://towardsdatascience.com/text-summarization through-use-of-spacy-library-b8a5902747a4
(III)https://www.ijraset.com/fileserve.php?FID=1213 (IV)https://hal.archives-ouvertes.fr/hal-00782442/document (V) https://kgptalkie.com/text-summarization-using-nlp/