NMAM Institute of Technology

(An Autonomous Institute Affiliated to VTU, Belagavi)

(A unit of NITTE Education Trust)

NITTE – 574110, UDUPI DIST., KARNATAKA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MINI PROJECT ON

WEB SUMMARIZER

(https://web-summarizer.herokuapp.com/)

Team Members:

1) Mr. Manjunath Patkar (4NM17CS100)

2) Mr. Maukashyap (4NM17CS101)

Under the Guidance

DR DK Sreekantha

Professor

Department of Computer Science and Engineering

NMAMIT Nitte.

ABSTRACT

Our main aim through is to showcase the possibilities of cloud computing through our project i.e web summarizer. Cloud computing has come a long way since its materialization that it is accessible to even a person who does not have any experience in the field of programming. There are technologies available that will help a common man to create a website and host it in the cloud with no programming skill required in less than a day.

Web summarizer in short aims to summarise the contents of any webpage in a fraction of time it will take a human to read, comprehend and summarise it. We use advanced web scraping technology to scrape the desired webpage for textual content and pass it through our word weightage based natural language processing model to analyze the contribution of each word to the article and rank them accordingly. Thus we create the summary which we hope will be meaningful in accordance to the desired article.

Index

CERTIFICATE

Certified that the project work carried out by Manukashyap U.V. (4NM17CS101) and Manjunatha Patkar (4NM17CS100) bonafide students of NMAM Institute of Technology, Nitte in fulfilment for the Cloud Project in Computer Science and Engineering during the academic year 2019-2020.

Signature of the examiner:

Signature of the Guide:

ACKNOWLEDGEMENT

The satisfactions that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible. So we acknowledge all those whose guidance and encouragement served as a beacon of light and crowned our efforts with success.

We are thankful to our project guide, DR D K Sreekanta, Dept. of CSE for his valuable guidance and advice. His willingness to motivate us contributed tremendously to our project.

We would like to place on record our deep sense of gratitude to Dr K. R. Uday Kumar Reddy, HOD-Dept. of Computer Science and Engineering, NMAMIT, Nitte for his generous guidance, help and useful suggestions. We also acknowledge and express our sincere thanks to our beloved Dr Niranjan N. Chiplunkar, Principal, NMAMIT, Nitte who is a source of inspiration to us.

We thank all the Teaching and Non-Teaching staff members of the department of CSE for providing resources for the completion of the project. A special thanks go to our parents, friends and relatives for supporting and encouraging us in all ways thus making our project successful. Finally, we thank all those who have contributed directly or indirectly in making this project a grand success.

Manukashyap U.V. (4NM17CS101)

Manjunatha Patkar (4NM17CS100)

Introduction

What is summarizing?

Summarising is an important skill in academic writing. It enables you to extract the most important points from a text and rewrite them in your own words, in a shortened form. Such skills are invaluable when you are note taking and researching for an essay. (source:http://aeo.sllf.qmul.ac.uk/Files/Summarizing/Summarising.html)

Why do we need text summarizer?

Automatic text summarization is the data science problem of creating a short, accurate, and fluent summary from a longer document. Summarization methods are greatly needed to consume the ever-growing amount of text data available online. In essence, summarization is meant to help us consume relevant information faster.
(Source:https://blog.frase.io/what-is-automatic-text-summarization/)

Our tool utilises the power of cloud processing and Natural language processing to create an automatic text summarise that will summarise a desired webpage for the user.

TECHNOLOGY STACK

Languages

1. Python
2. HTML
3. CSS

Libraries

1. NLTK
2. Beautiful Soup
3. LXML
4. Python Regex
5. Heapq

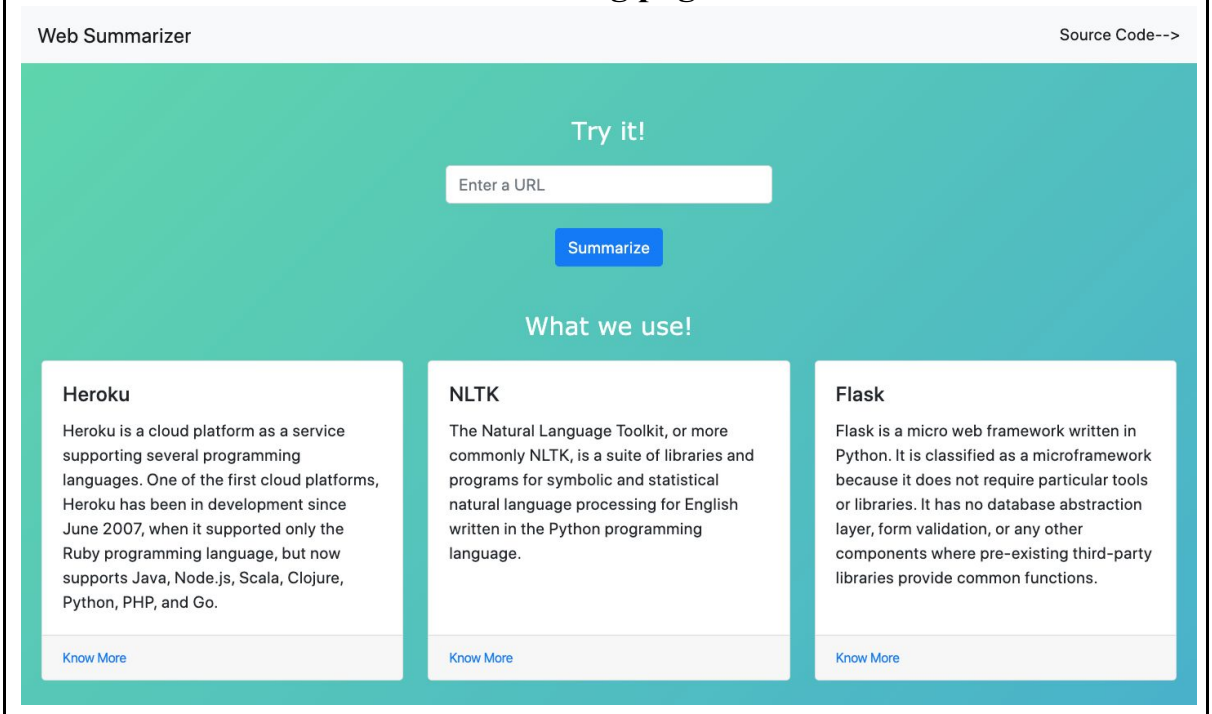
Framework

1. Flask
2. Bootstrap
3. Materialize

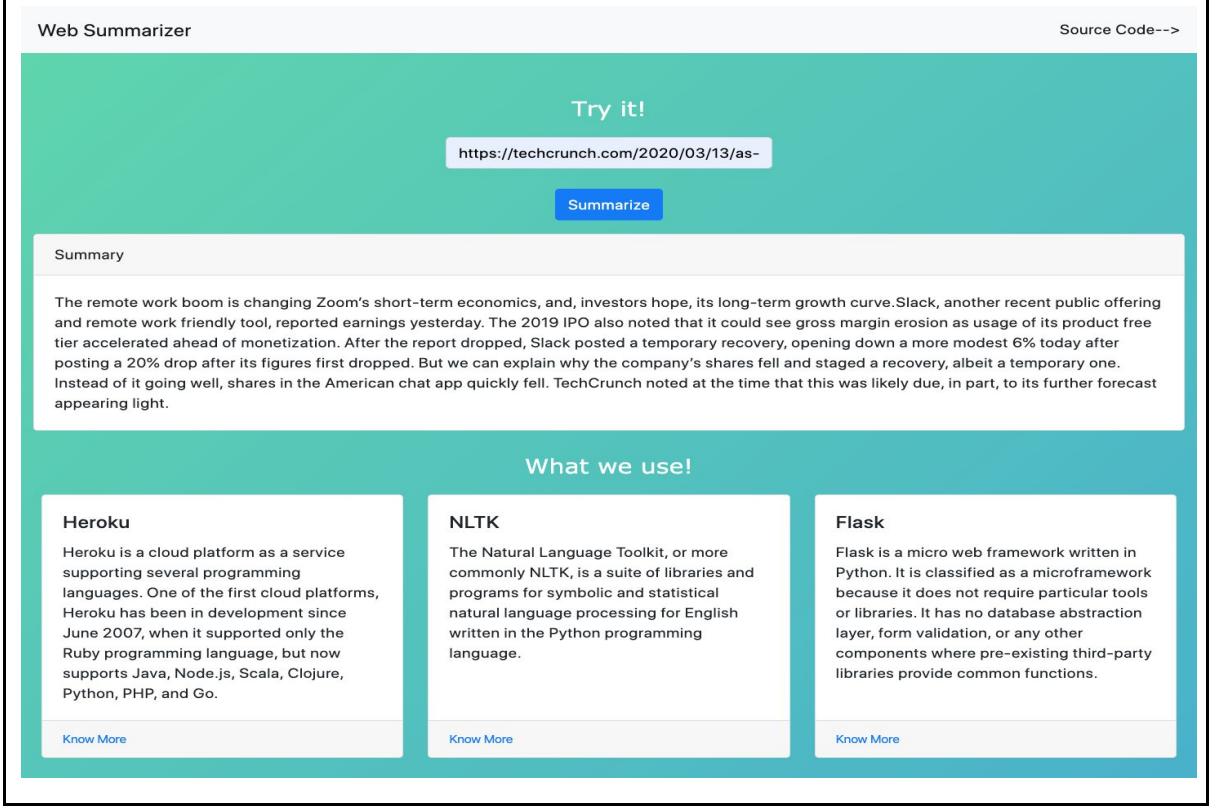
Platform

1. Heroku cloud application platform

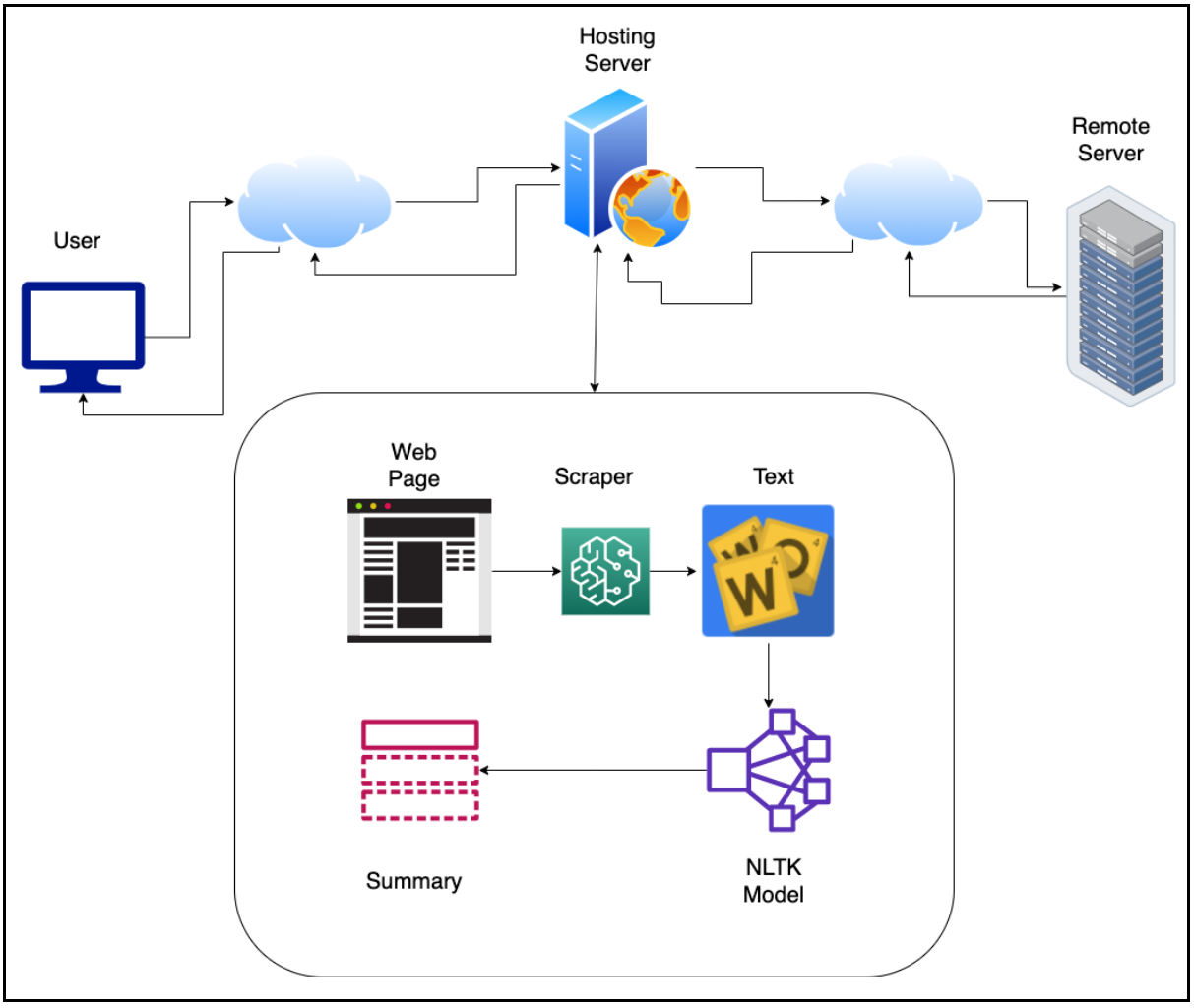
Landing page



Result



Working



The user provides a link to the web page, which he wishes to be translated. Our server sends a request to the desired webpage server and the returned webpage is scared and its textual data is stored in our server. We then extract the relevant data from the textual data and break them into words. Then the NLTK model ranks the words based on their occurrence and priority and subsequently every sentence is assigned a priority. Then at the end desired number of sentences having the highest priority is selected to be outputted as a summary to the user. We consider the user as a thin client thus all the computing is done on the server.

CONCLUSION

The web summarizer is designed using the most basic text summarising model since the main intention of this project was to demonstrate the prowess of cloud computing. That said this particular idea can be further developed using more efficient algorithms and can be given more user-friendly functions if so desired.

REFERENCE

- www.stackoverflow.com
- www.w3schools.com
- www.freecodecamp.org