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.

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

CERTIFICATE

"Web Summarizer"

is a bonafide work carried out by

Manjunath Patkar - 4NM17CS100

Manukashayp U V - 4NM17CS101

in partial fulfilment of the requirements for the award of Bachelor of Engineering degree in computer science and engineering prescribed by the Vishvesvaraya Technological University, Belagavi during the year 2019 - 2020

It is certified that all the corrections/suggestions indicated for internal assessment have been incorporated in the report.

The mini-project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the Bachelor of Engineering Degree.

Signature of Guide

Signature of HOD

ACKNOWLEDGEMENT

We believe that our project will be complete only after we thank the people who have contributed to making this project successful.

First and foremost, we express our deep sense of gratitude and indebtedness to our guide DR DK Sreekantha, Professor, Department of Computer Science and Engineering, for his inspiring guidance, constant encouragement, support and suggestions for improvement during the course of our project.

We sincerely thank Dr K.R. Udaya Kumar Reddy, Head of Department of Computer Science and Engineering, Nitte Mahalinga Adyantaya Memorial Institute of Technology, Nitte.

Our sincere thanks to our beloved principal, Dr Niranjan N. Chiplunkar for giving us an opportunity to carry out our project work at our college and providing us with all the needed facilities.

We also thank all those who have supported us throughout the entire duration of our project.

Finally, we thank the staff members of the Department of Computer Science and Engineering and all our friends for their honest opinions and suggestions throughout the course of our project.

Manukashyap U V - 4NM17CS101

Manjunath Patkar - 4NM17CS100

WEB SUMMARIZER **1**

CERTIFICATE **2**

ACKNOWLEDGEMENT **3**

Chapter 1: Introduction **5**

1.1 Scope 5

1.2 Importance 5

1.3 Objective 5

Chapter 2: Literature Survey **6**

2.1 Technical Background 6

2.2 Existing System 6

2.3 Proposed System 6

Chapter 3: System requirement and specification **7**

3.1 Functional Requirements 7

3.2 User Requirements 7

3.3 Software Requirements (For developers) 7

3.4 Hardware Requirement 7

Chapter 4: Technical Stack **8**

4.1 Languages 8

4.2 Libraries 8

4.3 Framework 8

4.4 Platform 8

Chapter 5: Implementation **9**

Chapter 6: ScreenShots **10**

Chapter 7: Conclusion and Future Work **11**

7.1 Conclusion/Result 11

7.2 Future Work 11

Chapter 8: Reference **12**

Chapter 1: Introduction

1.1 Scope

The rise of cloud storage and computing along with virtualization has become a boon for the technological development of humankind. With the conception of Web 2.0 the development of technical devices relying on the internet went into hyperspeed. With our project Web Summarizer, we intend to demonstrate the use of cloud processing.

1.2 Importance

In our everyday life much of the technological devices we come into contact with harness the power of the cloud. Our project follows this trend by harnessing the power of the cloud to run our natural language model to provide our users with the summary of the textual content present on the website the user wishes to summarize.

1.3 Objective

This project encompasses a series of objectives to visualize our idea of building a web summarizer. They include building a natural language model which can fetch the contents of any website just with its URL and then extract and analyze its textual content to create a summary with a negligible amount of time.

The second objective is to create a simple yet functional interface for the commoners to interact with our model. This is made possible using the micro framework flash which enables our model to run on the web.

The third objective is the use of the power of cloud storage and processing to enable people from all over the world to make use of our service. We host our web application in a reliable and cost-effective hosting platform and thus making it an accessible worldwide.

Chapter 2: Literature Survey

2.1 Technical Background

The initial approach is to create a web application and host it in the Heroku hosting platform which is a subsidiary of Salesforce. We use python to create our natural language model and use the Flask framework to make it work on the web. We use bootstrap to build an interactive and attractive user interface for our model.

2.2 Existing System

There are two kinds of existing solutions to the problem we intend to counter. First one will be to manually read the website and create a summary. Even though it will be more meaningful it will require time and human effort both of which is rare these days. The other summarizers available online ask the user to input the text to be summarized where we differ is we let the user input the URL of the webpage they wish to be summarised and we take care of the rest.

2.3 Proposed System

Our natural language model is built using python and we make use of the Flask framework to make it work on the web.

We use bootstrap to create an attractive and interactive UI to the user to interact with. It will be simple to use and will not require much resources to run since we will be treating our client's devices as thin clients.

We will host our app in Heroku hosting platform which is an efficient and cost-effective hosting platform to make our application available to users worldwide

Chapter 3: System requirement and specification

3.1 Functional Requirements

- User should be able to enter an URL.
- The website should return the summary for the entered URL.
- The website should be lite and responsive.
- The website should work on both computers and smartphones.
- The website should not be ambiguous.

3.2 User Requirements

- Mobile phone or computer.
- Active internet conneciton.
- Web browser.

3.3 Software Requirements (For developers)

- IDE capable of handling flask (Recommended: PyCharm by Jetbrains).
- Heroku CLI installed.
- Python 3.

3.4 Hardware Requirement

- Computer running Windows, macOS or Linux.
- Intel i series or equivalent Ryzen processor.
- 4GB or more RAM.
- 100GB of free disk space.

Chapter 4: Technical Stack

4.1 Languages

1. Python
2. HTML
3. CSS

4.2 Libraries

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

4.3 Framework

1. Flask
2. Bootstrap
3. Materialize

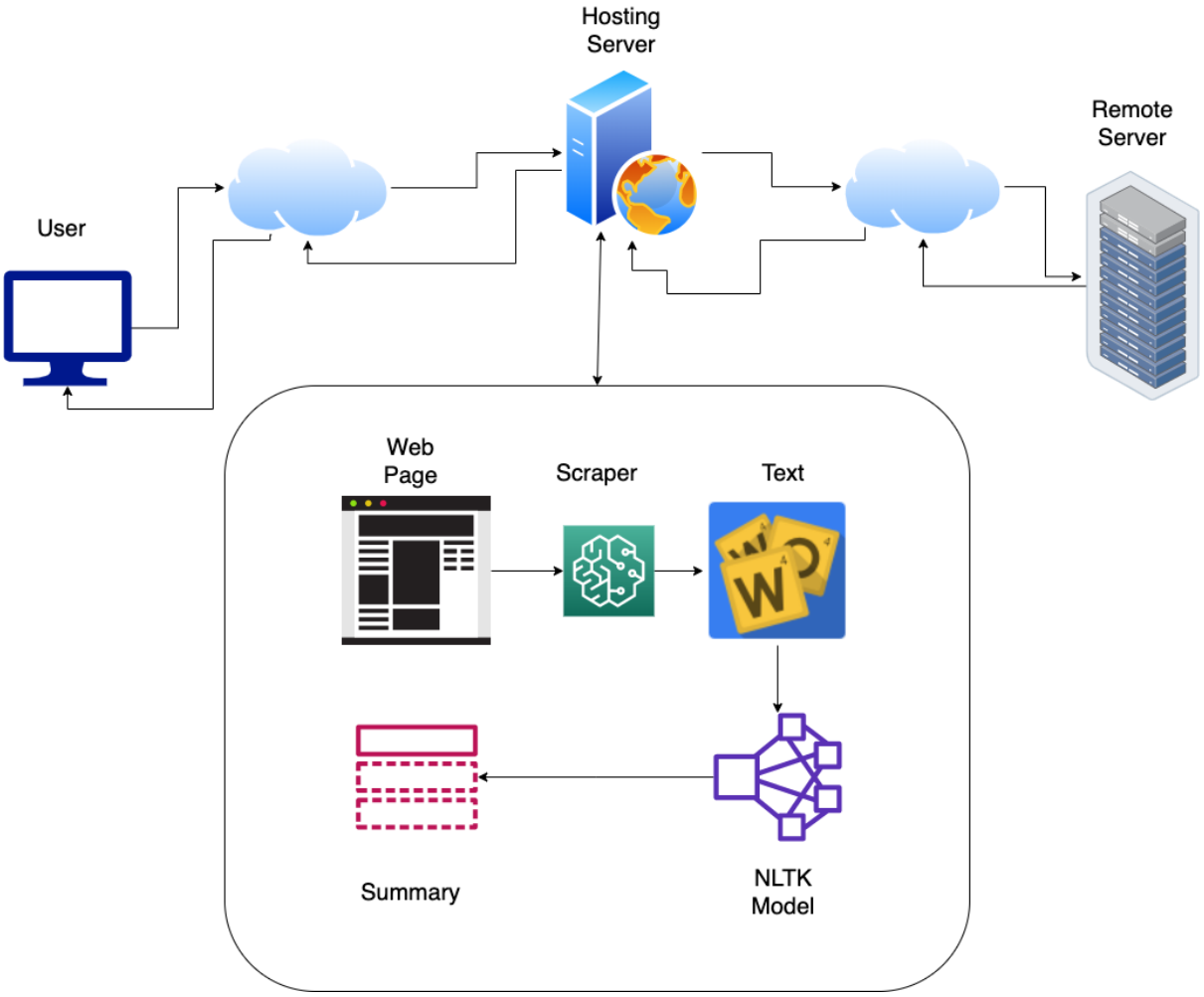
4.4 Platform

1. Heroku cloud application platform

Chapter 5: Implementation

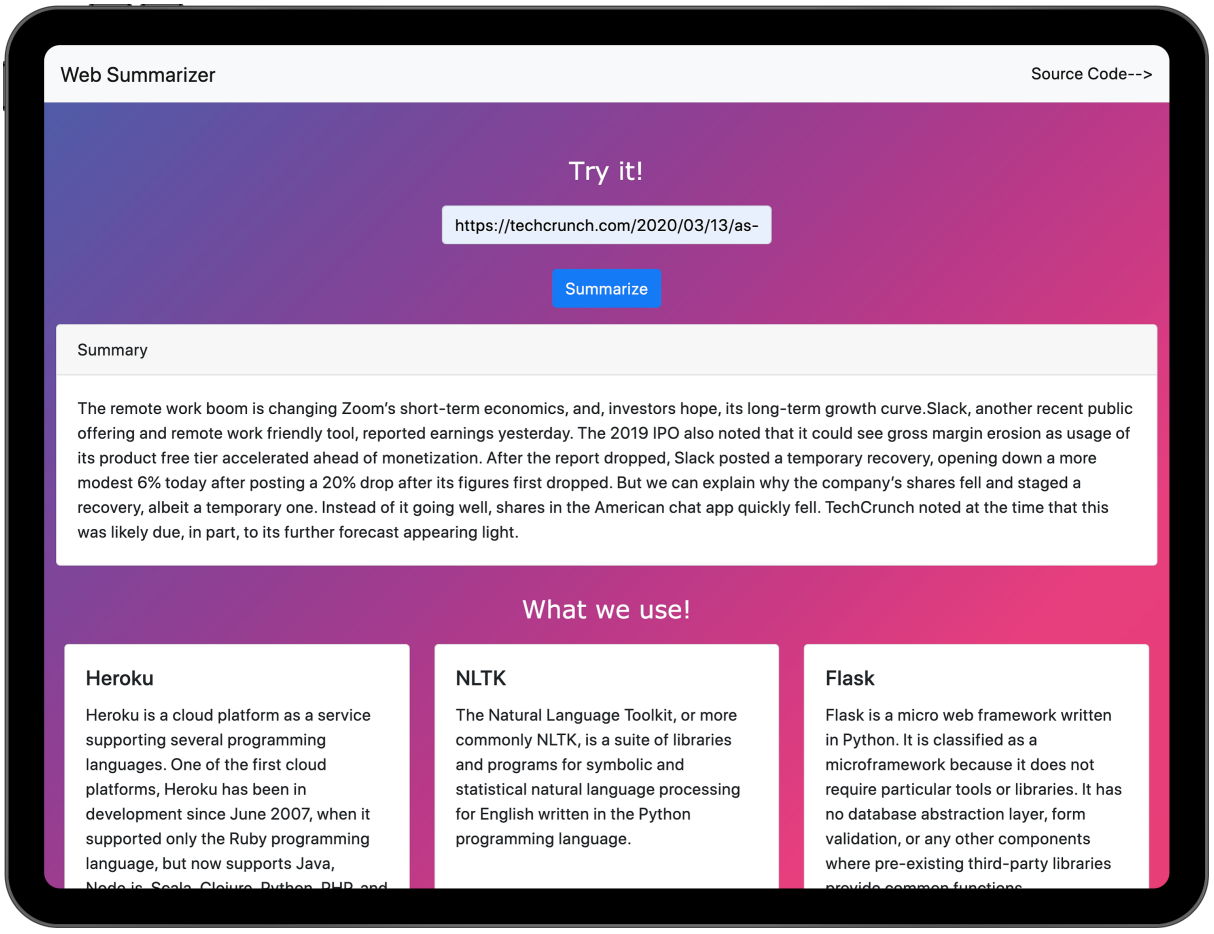
The full source code of this project can be found on

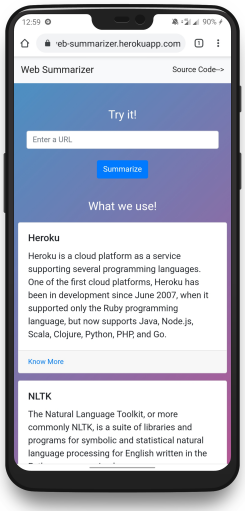
https://github.com/kashyapmanu/Web-summarizer

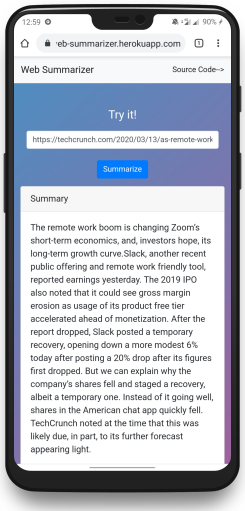


The user input is transferred to our server and our server fetches the webpage at the end of the user input and we scrape the entire page for textual data. We then partition the textual data into sentences and our model goes through all of these sentences removing stop words and other unnecessary punctuation marks and ranks all the words by the usage and assigns specific weights them. Based on the word weight we run the sentences through a heap queue which will sort the sentences based on priority and it will select those few of highest priority to create a Summary.

Chapter 6: ScreenShots







WEB SUMMARIZER

Chapter 7: Conclusion and Future Work

7.1 Conclusion/Result

- We have successfully created and deployed a web app that can summarize a web page.
- With this project we demonstrate the efficiency, scalability and flexibility of web processing and storage to automatically adapt to surge in users and protect against attacks without the need for the developer to work on this sector.
- Thus we can utilize cloud infrastructure to scale our application and reach millions of people.

7.2 Future Work

- Add ability to manually add text to be summarized.
- Add ability to specify the number of lines in summary.
- Create a better and more efficient model to handle the surge in user requests.

Chapter 8: Reference

- Stack Overflow: https://stackoverflow.com/
- FreeCodeCamp: https://www.freecodecamp.org/
- Heroku documentation: https://devcenter.heroku.com/categories/reference
- Flask documentation: https://flask.palletsprojects.com/en/1.1.x/
- Bootstrap documentation: https://getbootstrap.com/