TWITTER BOT FOR SUMMARIZING ARTICLES BASED ON TWITTER TRENDS

A Project Report

Submitted by

MS. SAMREEN REYAZ

MR. DISHANT BHATT

MR. GABRIEL FRANCIS

Under the guidance of

PROF. PANKTI DOSHI

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN COMPUTER SCIENCE

At

MUKESH PATEL SCHOOL OF TECHNOLOGY, MANAGEMENT AND ENGINEERING, NMIMS, MUMBAI

October 2019

DECLARATION

We, Samreen Reyaz, Dishant Bhatt and Gabriel Francis, Roll Nos. B005, B013 and B027 B.Tech (Computer Engineering), VIII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source: IEEE, The institute, Dec. 2004)

4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Student:

Name:	Samreen Reyaz	Dishant Bhatt	Gabriel Francis
Roll No.:	B005	B013	B027
Place:		Mumbai	
Date:		12th October 2019	

CERTIFICATE

This is to certify that the project entitled “Twitter Bot For Summarizing Articles Based On Twitter Trends” is the bonafide work carried out by Samreen Reyaz, Dishant Bhatt and Gabriel Francis of B.Tech (Computer Engineering), MPSTME (NMIMS), Mumbai, during the VIII semester of the academic year 2019-2020, in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

Prof. Pankti Doshi

Internal Mentor

Examiner 1

Examiner 2

Dean

Dr. N. T. Rao

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	LIST OF FIGURES	i
	LIST OF TABLES	ii
	LIST OF EQUATIONS	iii
	ABBREVIATIONS	iv
	ABSTRACT	v
1	INTRODUCTION	1
	1.1 Background	1
	1.2 Project Overview	2
	1.3 What is Text Summarization?	2
	1.4 Motivation of Problem	3
	1.5 Problem Statement	3
	1.6 Scope of the Problem	4
	1.7 Assumptions and Limitations	4
	1.8 Organization of Report	4
2	REVIEW OF LITERATURE	5
	2.1 Approaches to Text Summarization	5
	2.2 Extractive Text Summarization	6
	2.3.1 Word Frequency	7
	2.3.2 Term Frequency – Inverse Document Frequency	7
	2.3.3 Machine Learning	8
	2.3.4 Neural Networks	9
	2.3 Evaluation of Summarization	10
	2.3.1 Bilingual Evaluation Understudy (BLEU)	11
	2.3.1 Recall-Oriented Understudy for Gisting Evaluation (ROUGE)	13
	2.4 Twitter Trends	14
	2.4.1 Fetching Twitter Trends	14

2.4.2 Twitter API 14
2.4.3 Twitter Web App 14
2.4.4 Posting New Tweets 15
2.4.5 WOEID 15
2.4.6 Trend Manipulation 15
2.4.7 Drawbacks 15
2.5 Web Scraping 16
2.5.1 BeautifulSoup 16
2.5.2 Selenium 16

3 PROPOSED SOLUTION
3.1 Selection of Summarization and Evaluation Algorithms 17
3.2 Components 18
3.1.1 Twitter Interface 18
3.1.2 Web Scraper 19
3.1.3 Text Modifier 20
3.3 General Flow of Execution 21

4 IMPLEMENTATION 23
4.1 Summarization – Word Frequency 23
4.2 Summarization – Term Frequency – Inverse Document Frequency 24
4.3 Twitter – Fetching and Manipulating Trends 25
4.4 Twitter – Tweeting Summary 25
4.5 Web Scraping – BeautifulSoup and Selenium 26
4.5.1 BBC 26
4.5.2 Al-Jazeera 26

RESULTS AND DISCUSSION 28
CONCLUSION AND FUTURE SCOPE 33
REFERENCES 34
APPENDIX 35