AUTOMATIC TEXT SUMMARISATION

by

Richa Singh (1302710125)
Riya Jain (1302710128)
Ruchita Chandel (1302710130)
Shalu Sengar (1302710147)

Submitted to the Department of Computer Science & Engineering

in partial fulfillment of the requirements

for the degree of

Bachelor of Technology

in

Computer Science & Engineering

Ajay Kumar Garg Engineering College, Ghaziabad

Dr. APJ Abdul Kalam Technical University, Lucknow

May, 2017

TABLE OF CONTENTS

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature :

Name :Richa Singh

Roll No : 1302710125

Date :

Signature :

Name : Riya Jain

Roll No : 1302710128

Date :

Signature :

Name :Ruchita Chandel

Roll No : 1302710130

Date :

Signature :

Name :Shalu Sengar

Roll No : 1302710147

Date :

CERTIFICATE

This is to certify that Project Report entitled “Automatic Text Summarization” which is submitted by Richa Singh(1302710125) ,Riya Jain(1302710128) ,Ruchita Chandel(1302710130) , Shalu Sengar(1302710147) in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science and Engineering of Dr. APJ Abdul Kalam Technical University,is a record of the candidate own work carried out by him under our supervision. The matter embodied in this thesis is original and has not been submitted for the award of any other degree.

Supervisor

Ms Shiva Tyagi

(Assistant Professor)

CSE Department

Date:

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Ms. Shiva Tyagi and Mr. Akhilesh Verma, Department of Computer Science & Engineering, Ajay Kumar Garg Engineering College, Ghaziabad for their constant support and guidance throughout the course of our work. Their sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only their cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of Professor Mamta Bhusry, Head, Department of Computer Science & Engineering, Ajay Kumar Garg Engineering College, Ghaziabad for his full support and assistance during the development of the project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty and staff members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Signature :

Name :Richa Singh

Roll No : 1302710125

Date :

Signature :

Name : Riya Jain

Roll No : 1302710128

Date :

Signature :

Name :Ruchita Chandel

Roll No : 1302710130

Date :

Signature :

Name : Shalu Sengar

Roll No : 1302710147

Date :

ABSTRACT

The project titled “Automatic Text Summarisation” aims to the process of reducing a text Document with a computer program in order to create a summary that retains the most important points of the original document. As The problem of information overload has grown, and as the quantity of data has increased, so has interest in automatic summarization. It is very difficult for human beings to manually summarize large documents of text. Text Summarization methods can be classified into extractive and abstractive summarization. An extractive summarization method consists of selecting important sentences, paragraphs etc. from the original document and concatenating them into shorter form. The importance of sentences is decided based on statistical and linguistic features of sentences. Extractive methods work by selecting a subset of existing words, phrases, or sentences in the original text to form the summary. The extractive summarization systems are typically based on techniques for sentence extraction and aim to cover the set of sentences that are most important for the overall understanding of a given document.

.

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

=	equal to
≈	almost equal to
/	division slash
∞	infinity
(	ornate left parenthesis
)	ornate right parenthesis
[	left square bracket
]	right square bracket
<	less than
>	greater than
≤	less-than or equal to
≥	greater-than or equal to
,	single low-9 quotation
-	en dash
+	plus sign
.	decimal point
*	multiplication sign
{	left curly bracket
}	right curly bracket

LIST OF ABBREVIATIONS

IR	Information Retrieval
TF	Term Frequency
IDF	Inverse Document Frequency
NLP	Natural Language Processing
DBMS	Data Base Management System
KBS	Knowledge Base System
HWR	Handwriting recognition
UML	Unified Modeling Language
WORA	write once, run anywhere
JVM	Java Virtual Machine
JRE	Java Runtime Environment
CSIS	Cross-Sentence Information Subsumption