

## CONTENTS

TOPIC	PAGE NO
ABSTRACT	i
1. INTRODUCTION	
1.1. Introduction	1
1.2. Text mining and NLP	1
1.3. Information Extraction	2
1.4. Text Summarization	4
1.4.1. Stages of Text Summarization	4
1.4.2. Methods of Text Summarization	5
1.5. Steps for preprocessing	5
1.5.1 Text Normalization	5
1.5.2 Eliminating stop words	5
1.5.3 Stemming	6
1.6. Text document classification	6
1.6.1. Single Document Summarization	7
1.6.2. Multiple Document Summarization	7
1.7. Evaluation measures	7
1.8. Motivation for work	9
1.9. Problem statement	9
1.10. Organization of project report	9
2. LITERATURE SURVEY (EXISTING SYSTEM)	
2.1. Introduction	11
2.2. Methods of generating document summary	11
2.2.1) Studied papers title	11
2.3. Different methods in Text Summarization	26
2.4. Feature extraction methods	30
3. METHODOLOGY	
3.1. Overall system architecture of text summarizatio	32
3.1.1. Text processing	33

3.1.2. Preprocessing	33
3.1.3. Proposed system for Text summarization	34
3.1.3.1. Feature Extraction Algorithm	34
3.1.3.2. Algorithm Illustration	37
4. EXPERIMENTAL ANALYSIS AND RESULTS	
4.1. Software requirements	66
4.2. Hardware requirements	69
4.3. Sample code	69
4.4. Screenshots	74
5. CONCLUSION AND SCOPE FOR FURTHER STUDIES	
5.1 conclusion	83
5.2 Scope for further studies	83
REFERENCES	84
APPENDIX A: LIST OF STOP WORDS	86
APPENDIX B: LIST OF ABBREVIATIONS	88

## LIST OF TABLES

S.No	Table No	TableName	Pg. No
1	1.1	Measuring the correctness of the system	8
2	3.1	Sample input Document 1	37
3	3.2	Sample input Document 2	39
4	3.3	List of words for Document 1 after Stop Word Removal	43
5	3.4	List of words for Document 2 after Stop Word Removal	45
6	3.5	Calculation of probability for the list of words for Document 1	49
7	3.6	Calculation of probability for the list of words for Document 2	52
8	3.7	Calculation of Frequency score for Document 1	53
9	3.8	Calculation of Frequency score for Document 2	54
10	3.9	Calculation of sentence score for Document 1 based on header method	57
11	3.10	Calculation of sentence score for Document 2 based on header method	58
12	3.11	Calculation of score of each sentence in Document 1 using length method	59
13	3.12	Calculation of score of each sentence in Document 2 using length method	60
14	3.13	Calculation of score of each sentence in Document 1 using position method	61
15	3.14	Calculation of score of each sentence in Document 2 using position method	62
16	3.15	Calculation of final score of each sentence in Document 1	63
17	3.16	Calculation of final score of each sentence in Document 2	64

## List of Figures

S.No	Figure No.	Figure Name	Pg. No.
1	3.1	Architecture of the Proposed System	32
2	4.1	Input Document 1	75
3	4.2	Input Document 2	76
4	4.3	Output when compression ratio is 10%	77
5	4.4	Output when compression ratio is 20%	78
6	4.5	Output when compression ratio is 40%	79
7	4.6	Output when compression ratio is 40%	80