

Contents

ACKNOWLEDGEMENTS	V
ABSTRACT	VII
CONTENTS	IX
LIST OF FIGURES.....	XII
LIST OF TABLES	XIV
CHAPTER 1: INTRODUCTION TO TEXT SUMMARIZATION.....	1
1.1 INTRODUCTION	1
1.2 DATA MINING	2
1.3 TEXT MINING.....	5
1.4 MOTIVATION FOR TEXT SUMMARIZATION.....	8
1.5 PROBLEM STATEMENT AND OBJECTIVES.....	9
1.5.1 <i>Problem Statement</i>	9
1.5.2 <i>Objectives</i>	9
1.6 THE RESEARCH CONTRIBUTION	9
1.7 LAYOUT OF THE THESIS.....	10
CHAPTER 2: LITERATURE SURVEY	12
2.1 HISTORY OF SUMMARIZATION	12
2.2 A DEEPER LOOK INTO SUMMARIZATION.....	13
2.3 THE INPUT SOURCE	14
2.3.1 <i>Single Document Text Summarization</i>	14
2.3.2 <i>Multi-Document Text Summarization</i>	15
2.4 SUMMARIZATION APPROACHES	15
2.4.1 <i>Text Summarization with Supervised Approaches</i>	15
2.4.2 <i>Classification in Text Summarization</i>	17
2.4.3 <i>Regression in Text Summarization</i>	18
2.4.4 <i>Text Summarization with Unsupervised Approaches</i>	20
2.4.5 <i>Cluster-Based Methods</i>	20
2.4.6 <i>Term Frequency Based Methods</i>	21
2.4.7 <i>Latent Semantic Analysis Methods</i>	22
2.4.8 <i>Graph-Based Methods</i>	23
2.4.9 <i>Text Summarization with Natural Language Processing</i>	25
2.4.10 <i>Deep Learning in Text Summarization</i>	26
CHAPTER 3: TEXT PRE-PROCESSING.....	29
3.1 INTRODUCTION	29
3.2 MORPHOLOGICAL ANALYSIS	29
3.3 TOKENIZATION.....	30

3.4 FILTERING	30
3.5 STEMMING	31
3.5.1 <i>Introduction to Stemming</i>	31
3.5.2 <i>Stemming and Lemmatizing</i>	33
3.6 SYNTACTICAL AND SEMANTICAL ANALYSIS	35
3.6.1 <i>Syntactical Analysis</i>	35
3.6.2 <i>Semantical Analysis</i>	35
3.7.2 <i>Term Frequency</i>	37
3.7.4 <i>Length Normalization</i>	40
3.8 ATTRIBUTE SELECTION	41
3.8.1 <i>Introduction</i>	41
3.8.2 <i>Comparison of Attribute Selection Methods</i>	41
CHAPTER 4: LATENT SEMANTIC ANALYSIS	46
4.1. INTRODUCTION	46
4.2. INPUT MATRIX GENERATION	48
4.2.1. <i>Tokenization</i>	48
4.2.2. <i>Stop Words Removal</i>	48
4.2.3. <i>Stemming</i>	49
4.2.4. <i>Term Frequency</i>	49
4.2.5. <i>Weighting Cell Values</i>	49
4.3. SINGULAR VALUE DECOMPOSITION (SVD)	50
4.4. THE APPROACH TO SENTENCE SELECTION	52
4.4.1. <i>Gong and Liu Approach to Sentence selection (2011)</i>	52
4.4.2. <i>Steinberger and Jezek (2004)</i>	53
4.4.3. <i>Murray, Renals, and Carletta (2005)</i>	54
4.4.4. <i>Cross Method</i>	55
4.4.5. <i>Topic Method</i>	56
4.5. LIMITATIONS OF LSA	58
CHAPTER 5: TEXT SUMMARIZATION WITH NAÏVE BAYES	60
5.1 INTRODUCTION	60
5.2 CLASSIFICATION IN MACHINE LEARNING	61
5.2.1 <i>Classification - Text Handling</i>	63
5.2.2 <i>Classifying Summary Sentences</i>	63
5.3 THE NAÏVE BAYES CLASSIFIER	64
5.3.1 <i>The Bayes Theorem</i>	64
5.3.2 <i>The Multinomial Model</i>	66
5.3.3 <i>The Bernoulli Model</i>	68
5.3.4 <i>Comparison of the Multinomial and Bernoulli Models</i>	69
5.4 THE PROPOSED MODEL	70
5.4.1 <i>The Dataset description</i>	70
5.4.2 <i>Pre-processing the Data</i>	71
5.4.3 <i>Feature Extraction</i>	72

5.4.4 Latent Semantic Analysis.....	73
5.4.5 Threshold Generation	73
5.5 RECURSIVE FEATURE ELIMINATION	73
5.6 CLASSIFICATION USING NAÏVE BAYES	75
5.7 MODEL EVALUATION	75
5.7.1 The Evaluation Measures.....	76
5.8 EXPERIMENTAL RESULTS	81
5.8.1 Visualization of Training and Testing Data	84
CHAPTER 6: A HYBRID APPROACH TO TEXT SUMMARIZATION	88
6.1 INTRODUCTION	88
6.2 AN INSIGHT INTO DEEP LEARNING	89
6.3 THE PROPOSED MODEL	92
6.3.1 Pre-processing the Data	95
6.3.2 Feature Extraction	95
6.3.3 Latent Semantic Analysis.....	96
6.3.4 Threshold Generation	96
6.4 SELF-ORGANIZING MAPS (SOM)	96
6.4.1 SOM Based Text Summarization.....	99
6.4.2 SOM Visualization	100
6.5 MAPPING OF INPUT AND OUTPUT	100
6.6 INTRODUCTION TO DEEP NEURAL NETWORK.....	101
6.6.1 The Architecture of Deep Neural Network	103
6.6.2 Neural Network Learning	104
6.6.3 Gradient Descent.....	106
6.6.4 Activation Function	110
6.7 EXPERIMENTAL SETUP AND RESULT ANALYSIS.....	115
6.7.1 Dataset Description	115
6.7.2 Result Analysis	117
CHAPTER – 7 CONCLUSION AND FUTURE WORK	129
APPENDIX A.....	132
APPENDIX B.....	138
PUBLICATIONS	141
BIBLIOGRAPHY.....	143