Table of Contents

Abstract **V**
Table of Contents **VII**
List of Tables **IX**
List of Figures **XI**

Chapter 1 Introduction to Lemmatization **1**
1.1 Introduction 1
1.2 Morphology 1
1.3 History of Morphology 2
1.4 Morpheme 3
1.5 Types of Words 6
1.6 Approaches of Morphological Analyzer 9
1.7 Application of Morphological Analyzer 10
1.8 History of Application of Morphological Analysis 11
1.9 Stemming and Lemmatization 12
1.10 Applications of Stemming and Lemmatization 14
1.11 Research Gap and Research Formulation 16
1.12 Problem Statement 17
1.13 Objectives of the Study 17
1.14 Research Contribution 18
1.15 Organization of the thesis 18
1.16 Conclusion and Summary 19

Chapter 2 Literature Survey **20**
2.1 Introduction 20
2.2 Approaches to Stemming 20
2.3 Analysis of Affix Removal Stemming Algorithms 22
2.4 Lemmatization 37
2.5 Analysis of Lemmatizers 37
2.6 String Distance Similarity Measures 42
2.7 Conclusion and Summary 46

Chapter 3 Comparative Study of Stemmers and Lemmatizers **47**
3.1 Introduction 47
3.2 Algorithm for Comparative Study of Stemmers 47
3.3 Comparative Output of Stemmers 48
3.4 Popular Lemmatization Tools 53
3.5 Comparative Output of Existing Lemmatizers 54
3.6 Conclusion and Summary 56

Chapter 4 LemmaChase: A Lemmatizer **57**
4.1 Introduction 57
4.2 Understanding of LemmaChase 57
4.3 Input DataSet 60
4.4 Algorithm of LemmaChase 61
4.5 Parsing of Surface Word 69
4.6 Tools used in LemmaChase 74
4.7 Result and Discussion 76
4.8 Conclusion and Summary 88

Chapter 5 LemmaQuest: A Lemmatizer **89**
5.1 Introduction 89
5.2 Understanding LemmaQuest 89
5.3 Input DataSet 91
5.4 LemmaQuest Model 91
5.5 The LemmaQuest Algorithm 92
5.6 Details of LemmaQuest: Lemmatizer 92
5.7 Result and Discussion of LemmaQuest 98
5.8 Comparative Study 110
5.9 LemmaQuest vs. LemmaChase 114
5.10 Conclusion and Summary 117

Chapter 6 Conclusion and Future Enhancement **119**
6.1 Summary 119
6.2 Outcome of the Research Work 120
6.3 Research Output 121
6.4 Future Enhancement 122

Publications **123**

References **124**