TABLE OF CONTENTS

APPROVAL ii

SELF DECLARATION AGAINST PLAGIARISM iii

ABSTRACT iv

ABSTRAKSI v

DEDICATION vi

ACKNOWLEDGEMENT vii

TABLE OF CONTENTS viii

LIST OF TABLES xi

LIST OF FIGURES xii

CURRICULUM VITAE xiv

CHAPTER 1 THE PROBLEM 1

1.1 Rationale / Motivation 1

1.1 Theoretical Framework 2

1.2 Conceptual Framework/Paradigm 3

1.3 Problem Statement 3

1.4 Objectives 4

1.5 Hypothesis 4

1.6 Assumption 4

1.7 Scope and Delimitation 5

1.8 Importance of the Study 5

CHAPTER 2 REVIEW OF LITERATURE AND STUDIES 7

2.1. *Bahasa Indonesia* Morphology 7

2.2. Related Works 9

2.2.1 Nazief – Adriani Stemmer Algorithm (S_NA) 9

2.2.2 CS Stemmer Algorithm 12
2.2.3 Enhanced Confix Stripping Stemmer 15
CHAPTER 3 RESEARCH METHODOLOGY 17
3.1 Morphological Observation 17
3.1.1 Weakness of S_NA Stemmer 17
3.1.2 Weakness of CS Stemmer 19
3.1.3 Weakness of Enhanced Confix Stripping 20
3.1.4 New Stemmer Algorithm (S_HA) 21
3.2 Experiment with Real Data 27
3.2.1 Experiment Design 27
3.2.2 Experiment Scenario 30
3.2.3 Instrument and Data Collection 31
3.2.3.1 “Laskar Pelangi” Novel Book 31
3.2.3.2 “Riyadhus Shalihin” Hadits Book 32
3.2.3.3 “Republika” Online News 33
3.2.3.4 “Kamus Besar Bahasa Indonesia Tahun 2008” (KBBI 2008) 34
3.2.4 Tools for Data Analysis 36
3.2.4.1 Python 2.7 36
3.2.4.2 Apache Open Office 4.1.1 36
3.2.4.3 Stemming Process Output 36
3.2.4.4 Categorized Failed Stemmed Words 38
CHAPTER 4 PRESENTATION, ANALYSIS AND INTERPRETATION OF DATA 39
4.1. Presentation of Data 39
4.1.1 Performance Comparison S_NA, CS, and New Stemmer Algorithm 39
4.1.2 Cause Failures Category for *Laskar Pelangi* Data 43
4.1.3 Cause Failures Category for *Riyadhus Shalihin* 44
4.1.4 Cause Failures Category for Republika 45

4.1.5 Result Comparison S_NA vs CS Stemmer 45
4.1.6 Result Result Comparison CS Stemmer vs New Stemmer (S_HA) 47
4.2 Analisis of Data 48
4.2.1 Analisys of Data Impact to Stemmer Performance 48
4.2.2 Analysis of Failed Cause 49
4.2.3 Analysis of Comparison of S_NA vs CS Stemmer 50
4.2.4 Analysis of Comparison of CS Stemmer vs New Stemmer 53
CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS 55
5. 1 Conclusions 55
5.2 Recomendations 55
REFFERENCES 56