
CONTENTS

APPROVAL	ii
SELF DECLARATION AGAINST PLAGIARISM	iii
ABSTRACT	iv
ABSTRAK	v
CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
1 INTRODUCTION	1
1.1 Rationale	1
1.2 Problem Formulation	2
1.3 Objective	3
1.4 Hypothesis	3
1.5 Problem Limitation	3
1.6 Research Methodology	4
1.7 Systematic Writing	5
2 LITERATURE REVIEW	6
2.1 Previous Research	6
2.2 Literature Study	9
2.2.1 Multilabel Text Classification	9
2.2.2 Semantic Features	10
2.2.3 Lesk Algorithm	11
2.2.4 Word Embedding	12
2.2.5 Semantic Naive Bayes	13
2.2.6 Tensor Space Model	14
2.2.7 Performance Evaluation	15
3 RESEARCH METHODOLOGY AND SYSTEM DESIGN	16
3.1 Research Methodology	16
3.1.1 Problem Identification	16
3.1.2 Requirement Identification	16
3.1.3 System Design	17
3.1.4 Data Collection	17

3.1.5	System Implementation	17
3.1.6	Testing System System Evaluation	17
3.1.7	Analysis of Test Result	17
3.2	System Design	17
3.2.1	Dataset	17
3.2.2	Preprocessing	19
3.2.3	Collecting Wikipedia Pages	20
3.2.4	Splitting Dataset	21
3.2.5	Word Embedding	21
3.2.6	Classifier Model	21
3.2.7	Classification	22
3.2.8	System Evaluation	23
3.3	Proposed Method	23
3.4	Scientific Contribution	25
4	TESTING AND RESULT ANALYSIS	27
4.1	System Testing	27
4.1.1	Testing Objectives	27
4.1.2	Testing Scenario	27
4.2	Test Result and Analysis	28
4.2.1	Result and Analysis Use of Different Concept Windows in Building Lesk Model	28
4.2.2	Result and Analysis Use of Feature of Lesk and Word Embedding	29
4.2.3	Result and Analysis Use of Combination Feature of Lesk with Word Embedding	29
5	CONCLUSIONS AND FUTURE OF WORK	31
5.1	Conclusions	31
5.2	Future of Work	31
	BIBLIOGRAPHY	32