



TABLE OF CONTENTS

APPROVAL PAGE	I
STATEMENT	II
FOREWORD	III
TABLE OF CONTENTS	V
TABLE OF TABLES	VII
TABLE OF FIGURES	IX
ABSTRACT	XI
ABSTRAKSI	XII
CHAPTER I	1
1.1 RESEARCH BACKGROUND	1
1.2 RESEARCH PROBLEM	3
1.3 RESEARCH SCOPE	3
1.4 RESEARCH OBJECTIVE	4
1.5 RESEARCH ADVANTAGE	4
1.6 RESEARCH METHODOLOGY	4
1.7 THESIS ORGANIZATION	5
CHAPTER II	7
CHAPTER III	12
3.1 TEXT MINING	12
3.2 TOPIC MODELING	12
3.3 LATENT DIRICHLET ALLOCATION	13
3.4 GIBBS SAMPLING	15
3.5 JENSEN-SHANNON DIVERGENCE	17
3.6 WORD INTRUSION TASK	18
CHAPTER IV	19
4.1 GENERAL ANALYSIS	19
4.2 DATA ACQUISITION ANALYSIS	20
4.3 DATA PRE-PROCESSING DESIGN	22
4.3.1 Case Folding	23
4.3.2 Stopwords	23
4.3.3 Lemmatization	24
4.3.4 Tokenization	24
4.4 EXPERIMENT	25
4.4.1 Create document matrix and vocab matrix	25



4.4.2	Topic modeling	27
4.4.3	Topic chain.....	33
4.5	EVALUATION.....	35
CHAPTER V.....		37
5.1	SPECIFICATION.....	37
5.2	DATA ACQUISITION IMPLEMENTATION	37
5.3	DATA PRE-PROCESSING IMPLEMENTATION	39
5.3.1	Stopwords.....	40
5.3.2	Tokenization.....	40
5.3.3	Lemmatization.....	40
5.4	EXPERIMENT IMPLEMENTATION.....	42
5.4.1	Create document matrix and vocab matrix	42
5.4.2	Topic modeling	44
5.4.3	Topic chain.....	50
5.5	EVALUATION IMPLEMENTATION	51
CHAPTER VI		54
6.1	DATA ACQUISITION RESULT	54
6.2	DATA PRE-PROCESSING RESULT	56
6.3	EXPERIMENT RESULT.....	56
6.3.1	Document matrix and vocab matrix result.....	56
6.3.2	Topic modeling result	58
6.3.3	Topic chain result.....	73
6.3.4	Topic chain result.....	87
6.4	EVALUATION RESULT	90
CHAPTER VII.....		91
7.1	CONCLUSIONS	91
7.2	FUTURE WORKS.....	92
REFERENCES.....		93
ATTACHMENT		96