

TABLE OF CONTENTS

UNDERGRADUATE THESIS	i
UNDERGRADUATE THESIS	ii
APPROVAL PAGE	iii
STATEMENT	iv
DEDICATION PAGE.....	v
FOREWORD	vi
ABBREVIATIONS AND ACRONYMS	viii
TABLE OF CONTENTS	ix
LIST OF TABLES	xii
LIST OF FIGURES	xiii
ABSTRACT	xvi
CHAPTER I INTRODUCTION	1
1.1 Background	1
1.2 Problem Formulation.....	4
1.3 Research Scope.....	4
1.4 Research Objectives	5
1.5 Advantages	5
1.6 Thesis Organization.....	5
CHAPTER II LITERATURE REVIEW	7
CHAPTER III THEORITICAL BASIS	14
3.1 Electronic Health Records	14
3.1.1 Discharge Summary	15
3.2 Natural Language Processing	15
3.2.1 Part of Speech Tagging	16
3.2.2 Corpus	23
3.3 N-Gram Tagger	23
3.3.1 Unigram Tagger	24
3.3.2 Bigram Tagger	25
3.3.3 Trigram Tagger	26

3.3.4	Combination of Taggers.....	27
3.4	Natural Language Toolkit (NLTK) Python Library.....	27
3.6	Evaluation Performance	29
3.6.1	Gold Standard	30
3.6.2	K-Fold Cross Validation	31
CHAPTER IV ANALYSIS AND SYSTEM DESIGN		32
4.1	System Analysis	32
4.1.1	General Description	32
4.1.2	Data Analysis	34
4.2	Research Design	35
4.2.1	Intra-institution Scenario.....	37
4.2.2	Inter-institution Scenario.....	38
4.2.3	Mix-Institution Scenario	39
4.3	System Design	41
4.3.1	Pre-Processing.....	41
4.3.2	Assigning Tag on Data Training	41
4.3.3	POS Tagger	43
4.3.4	Evaluation Measurement.....	44
CHAPTER V IMPLEMENTATION		45
5.1	Implementation Description	45
5.2	Pre-processing	45
5.3	Assigning Tag on Data Training	50
5.4	POS Tagger	53
5.4.1	Unigram Tagger Step	53
5.4.2	Bigram Tagger Step	56
5.4.3	Trigram Tagger Step	57
5.4.4	Combined Tagger Step.....	58
5.5	Scenario Implementation.....	59
5.6	Evaluation Performance	62
CHAPTER VI EXPERIMENTAL RESULTS AND DISCUSSION		64
6.1	Results	64
6.1.1	Intra-Institution Scenario Results.....	64

6.1.2	Inter-Institution Scenario Results.....	65
6.1.3	Mix-Institution Scenario Results	66
6.1.4	Scenario Comparison	66
6.2	Discussion	69
CHAPTER VII CONCLUSION AND FUTURE WORK.....		72
7.1	Conclusion.....	72
7.2	Future Work	73
ACKNOWLEDGMENT.....		xvii
BIBLIOGRAPHY		xviii
APPENDIX		xxii