

TABLE OF CONTENTS

APPROVALiii

SELF DECLARATION AGAINST PLAGIARISM.....iv

ABSTRACTv

ABSTRAKSIvi

DEDICATIONvii

ACKNOWLEDGMENTS.....viii

PREFACEix

TABLE OF CONTENTS 1

LIST OF FIGURE 3

LIST OF CODE 4

LIST OF TABLE 4

1. CHAPTER 1: INTRODUCTION 5

 1.1 Rationale..... 5

 1.2 The Problem..... 6

 1.3 Theoretical Framework..... 6

 1.4 Conceptual Framework/Paradigm 6

 1.5 Hypothesis 7

 1.6 Assumption 7

 1.7 Scope and Delimitation..... 7

 1.8 Importance of the Study 8

2. CHAPTER 2: REVIEW OF LITERATURE AND STUDIES..... 9

 2.1 Related Works..... 9

 2.2 Chart Image Processing Overview..... 12

 2.3 Natural Language Processing Overview 12

 2.4 Natural Language Processing Method 13

 2.4.1 Lexicons 13

 2.4.2 Corpus..... 13

 2.4.3 *N*-Grams..... 13

 2.4.4 Training *N*-Gram Models..... 14

 2.4.5 Grammar Rule..... 15

3. CHAPTER 3: RESEARCH METHODOLOGY 18

 3.1 Research Design..... 18

 3.2 System/product/method Implementation 19

 3.2.1 Image Processing 20

 3.2.2 Loading the image..... 20

 3.2.3 Pixel Projection 22

 3.2.4 OCR 25

 3.2.5 OCR to integer (scale) 26

 3.2.6 OCR to String (title,legend,axis title) 27

 3.2.7 Following chart line to produce pair of (x, y) values of the chart line 27

 3.2.8 Pixel to Cartesian conversion 27

 3.3 Bridge..... 28

3.3.1	Seed Words Segmentation.....	29
3.3.2	Pattern Recognition	29
3.3.3	Knowledge Base	34
3.3.4	Pattern Abstraction.....	34
3.4	Natural Language Processing.....	34
3.4.1	Seed Words Selection	34
3.4.2	Corpus.....	35
3.4.3	Calculating the probability of Bigram Grammar Rules	35
3.4.4	Calculating the Probability For Sentence Generation	35
3.4.5	Generating Sentences	35
3.5.	Experiment Scenario	36
3.6.	Population/Samples.....	37
3.7.	Tools for Data Analysis.....	37
4	CHAPTER 4: PRESENTATION, ANALYSIS AND INTERPRETATION OF DATA.....	39
4.1	Running Program	39
4.1.1	Main GUI	39
4.1.2	Knowledge Base	39
4.1.3	Probability Count For Sentence Generation.....	40
4.1.4	Generating Sentences	40
4.2	Presentation and Analysis of the Data	40
4.2.1	Error Calculation for the Chart Value Extraction process.....	40
4.2.2	Accuracy calculation for the Optical Character Recognition (OCR) process	44
4.2.3	Acceptance rate of the automatic report generation	45
4.2.4	Keyword Sufficiency.....	45
4.2.5	Chart Happening Possibility	46
4.3	Summary of Findings	47
4.3.1	Suggestion From The Expert	47
5.	CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS	49
5.1	Conclusions.....	49
5.2	Recommendations	49
	REFERENCES	50
	APPENDIX A : SUBMIT PAPER TO ICOICT 2013.....	53
	(Under IEEE Conference)	53
	APPENDIX B : Questionnaire With Dummy Data	54
	APPENDIX C : Questionnaire Result From Experts	56
	APPENDIX D : Extracted Value VS Original Value to calculate RMSE	61
	APPENDIX E : Source Of The Corpus	65
	APPENDIX F : Realistic Scanned Chart.....	69
	CURRICULUM VITAE	72