TABLE OF CONTENT

COVER PAGE

ORIGINALITY STATEMENT OF THE THESIS

THESIS SUPERVISOR'S APPROVAL

THESIS EXAMINATION COMMITTEE

ABSTRACT v

FOREWORD vii

TABLE OF CONTENT viii

LIST OF FIGURES x

LIST OF TABLES xi

LIST OF ATTACHMENTS xii

1. CHAPTER I 13

1.1 Background 13

1.2 Problem Identification 18

1.3 Problem Limitation 19

1.4 Problem Definition 20

1.5 Research Purposes 21

1.6 Outline of the Thesis 21

2. CHAPTER II 24

2.1 Named Entity Recognition 24

2.2 Image Pre-processing 27

2.3 Optical Character Recognition 34

2.4 Rule-Based Text Matching 44

2.5 Evaluation Method 446

3. CHAPTER III 46

3.1 Overview 47

3.2 Text Data Extraction from Letter of Credit 48

3.3 Data Extraction from Invoice 52

3.4 Rule-based Matching 58

3. CHAPTER IV 61

4.1 Paper Summary 1: OPTICAL CHARACTER RECOGNITION (OCR) FOR TEXT RECOGNITION AND ITS POST-PROCESSING METHOD: A LITERATURE REVIEW 61

4.2 Paper Summary 2: AN AUTOMATION OF INVOICE EXAMINATION UNDER LETTER OF CREDIT BASED ON UCP 600 USING OCR AND RULE-BASED TEXT MATCHING 62

4.3 Research Connectivity 62

3. CHAPTER V 65

5.1 Summary 65

5.2 Future Research 65

BIOGRAPHY ..... **Error! Bookmark not defined.**

REFERENCES 67

ATTACHMENT 72

LIST OF FIGURES

Figure 2.1. Illustration of Image filtering 32
Figure 2.2. Illustration of Padding 32
Figure 2.3. Original Image (Left) and Sharpened Image (Right) 33
Figure 2.4. Illustration of blurring on noisy image with their OCR result 34
Figure 2.5. Architecture of Tesseract OCR 37
Figure 2.6. Sample Image 38
Figure 2.7. Output from Tesseract v 3.0 (a), 4.0 (b) and 5.0 (c) 38

Figure 2.8. Neural Network Architecture 41
Figure 2.9. RNN Node Architecture 42
Figure 2.10. LSTM Node Architecture 44
Figure 3.1. Block Diagram of the whole process 47
Figure 3.2. L/C Example 49
Figure 3.3. Variation of Name and Address 49
Figure 3.4. Coding example to extract Beneficiary/Applicant Name 50
Figure 3.5. Coding to extract Applicant Country 51
Figure 3.6. Original Image 54
Figure 3.7. De-skewed Image 54
Figure 3.8. Cleaned border image 54
Figure 3.9. Image after box lines deleted 55
Figure 3.10. Blurred Image 55
Figure 3.11. OCR result of image without blurring (left) and with blurring (right) 55

Figure 3.12. (a) OCR output in string type, (b) Invoice illustration 56
Figure 3.13. Original image (left) and signature location extracted (right) 57

Figure 3.14. Beneficiary Name checking output 59
Figure 3.15. Result of City/Country search on Invoice 60
Figure 3.16. Currency Checking result 60

LIST OF TABLES

Table 4.1. Experiment Results 62

LIST OF ATTACHMENTS

Paper 1 (OPTICAL CHARACTER RECOGNITION (OCR) FOR TEXT RECOGNITION AND ITS POST-PROCESSING METHOD: A LITERATURE REVIEW) A-1

Paper 2 (AN AUTOMATION OF INVOICE EXAMINATION UNDER LETTER OF CREDIT BASED ON UCP 600 USING OCR AND RULE-BASED TEXT MATCHING) A-2