TABLE OF CONTENTS

DECLARATION ii

ACKNOWLEDGEMENTS iii

ABSTRAK iv

ABSTRACT v

TABLE OF CONTENTS vi

LIST OF TABLES x

LIST OF FIGURES xi

1. INTRODUCTION 1

1.1. Background 1

1.2. Motivation 2

1.2.1. Classifying Expert 2

1.2.2. Automated Classification 2

1.3. Objectives of the Research 3

1.4. Research Problem 3

1.5. Assumptions and Constraints 4

1.6. Scope of Research 5

1.7. Outline of Thesis 6

2. BACKGROUND & RELATED WORK 7

2.1. Introduction 7

2.2. Expert Classification 7

2.3. Classification Model 8

2.3.1. Manual Classification Model 8

2.3.2. Automated Classification Model 8

2.3.3. Supervised Classification Model 9

2.3.4. Summary 10

3. EXPERT CLASSIFICATION (ExpClass) FRAMEWORK 11

3.1. Introduction 11

3.2. An Overview of the Expert Classification (ExpClass) Framework 11

3.3. Classification System Training 12

3.3.1. Category and Training Texts 12

3.3.2. Feature Pre-processing 13

3.3.3. Feature Selection 14

3.3.3. (a) Term Weighting 14

3.3.3. (b) Term Frequency (TF) 14

3.3.3. (c) Term Frequency – Inverse Document Frequency (TF-IDF) 15

3.4. Expert Classification 16

3.4.1. Expert and Bibliography (Bib) 16

3.4.2. Feature Pre-processing 16

3.4.3. Bibliography – Category Similarity Matching 17

3.4.4. Expertise Ranking 18

3.5. Application of Expert Classification Framework 19

3.5.1. Classification Expert 19

3.5.2. Expert Filtering using Advance Search 20

3.6. Summary 21

4. EVALUATION 22

4.1. Introduction 22

4.2. Experiment Setup 22

4.3. Evaluation of Classification using Broad Classification Model 24

4.3.1. Motivation 24

4.3.2. Test Collections 25

4.3.2. (a) Data Collection 25

4.3.2. (b) Topic Set 27

4.3.2. (c) Relevance Assessment 28

4.3.2. (d) Performance Metrics 29

4.3.3. Results and Findings 30

4.3.3. (a) Evaluation on Performance of Broad Classification System 30

4.3.3. (b) Evaluation on the Impact of the Number of Training Text 32

4.3.3. (c) Evaluation on the Impact of the Number of Category per Bibliography 35

4.4. Evaluation of Classification using Specific Classification Model 37

4.4.1. Motivation 37

4.4.2. Test Collections 37

4.4.2. (a) Data Collection 37

4.4.2. (b) Topic Set 38

4.4.2. (c) Relevance Assessment 39

4.4.2. (d) Performance Metrics 39

4.4.3. Results and Findings 39

4.5. Finding and Discussion 40

4.5.1. Accuracy of Classification 40

4.5.2. Number of Training Texts 41

4.5.3. Verification on Different Models 41

4.5.4. Limitations 42

4.6. Summary 42

5. CONCLUSION & FUTURE WORK 43

5.1. Summary of Results 43

5.2. Research Contribution 44

5.3. Limitation of the Research 45

5.4. Future Work 46

5.4.1. Optimizing categories 46

5.4.2. Increasing the number of test case 46

5.4.3. Improvement of preparation of golden standard 47

6. REFERENCES 48