# Contents

1. Introduction

1.1 Aim ..... 5

1.2 Motivation ..... 5

1.3 Organization of our report ..... 6

2. Background Study

2.1. Text Normalization ..... 7

2.2. Text Summarization ..... 7

2.2.1. Definition .....7

2.2.2. Process of automatic text summarization ..... 8

2.2.3 Approaches to text summarization ..... 8

2.2.3.1 Surface level ..... 9

2.2.3.2 Entity level ..... 9

2.2.3.3 Discourse level ..... 10

2.2.4 Measures of evaluation for text summarization ..... 10

2.2.4.1 Intrinsic Evaluation ..... 11

2.2.4.1.1 Summary Coherence ..... 11

2.2.4.1.2 Summary Informativenes ..... 11

2.2.4.1.3 Sentence Precision and Recall ..... 11

2.2.4.1.4 Sentence Rank ..... 11

2.2.4.1.5 The Utility Method ..... 11

2.2.4.1.6 Content Similarity ..... 12

2.2.4.1.7 BLEU Scores ..... 12

2.2.4.2 Extrinsic Evaluation ..... 12

2.2.5 Evaluation Tools ..... 12

3 Related Approaches

3.1 CECS ..... 13

3.2 An extractive approach by Chandan et.al ..... 15

3.3 Phrase Reinforcement Algorithm ..... 15

4.Work Done

4.1 Corpus collection ..... 17

4.2 Detection of Non-Standard Words ..... 17

4.3 Analysis and categorization of Non-Standard Words..... 17

4.4 Normalization Flow Chart ..... 23

4.5 Text Summarization ..... 25

5. Observation and Results ..... 27

6. Conclusion and Future Work ..... 29

# List of figures

Fig 3.1 CECS ..... 14

Fig 4.1 Normalization Flow Chart ..... 23

# List of Tables

Table 4.1 NSWs Category table ( i ) ..... 19

Table 4.2 NSWs Category table ( ii ) ..... 22

Table 5.1 Results table ..... 27