Novel Predictive Text Entry Method for Sinhalese Mobile Users

H. A. N. Ishara

(Reg. No.:MS20911980)

M.Sc. in IT

Specialized in EAD

Supervisor: Dr. Shyam Reyal

July 2021

Department of Information Technology

Faculty of Graduate Studies and Research

Sri Lanka Institute of Information Technology

Table of Contents

1. Chapter 1 Introduction: 4
   1.1. Introduction to Chapter 4
   1.2. Context and Background 4
      1.2.1. Emerging of Typing Keyboards 4
      1.2.2. Human Computer Interaction 4
      1.2.3. User Interface 5
      1.2.4. Design Thinking 5
   1.3. Scope 5
   1.4. Problem Definition 5
   1.5. Research Objectives 6
      1.5.1. Research Questions 6
      1.5.2. Primary Objectives 6
      1.5.3. Secondary Objectives 6
   1.6. Significance of the study 6
2. Literature Review 7
   2.1. Introduction to Chapter 7
   2.2. Error Handling in Text entry 7
   2.3. Future Impact of Text Entry Methods 7
   2.4. Challenges in Designing 8
   2.5. Word Prediction 8
   2.6. Native languages for Mobile Text Entry 9
   2.7. History and Evaluation 9
   2.8. Evaluation of Sinhalese Text Entry 10
   2.9. Evolution Of Keyboards 11
   2.10. Mobile keyboards 11
   2.11. Architecture of Mobile Keyboards 12
      2.11.1. Best practices 12
   2.12. Factors to consider prior to implement a mobile Keyboard 14
      2.12.1. Existing keyboards 14
3. Methodology 15
   3.1. Introduction to the chapter 15
   3.2. Conceptual Framework 15
   3.3. Data Collection 16
   3.4. Procedure 17

3.4.1. Phase 1 – Data Collection and Preprocessing 17
3.4.2. Phase 2 – Language Model Building 17
3.4.2.1. Tokens generation 18
3.4.2.2. Language model generation 18
3.4.2.3. Language model generation with SRILM Toolkit 18
3.4.2.4. Cyqwin installation 18
3.4.2.5. Perplexity calculation 19
3.4.2.6. Smoothing Techniques 19
3.4.2.6.1. Knesner-Ney Smoothing 19
3.4.3. Phase 3 - Building REST API for next Word prediction 20
4. Results 21
4.1. Introduction to Chapter 21
4.2. Results of Phase 1 21
4.3. Results of Phase 2 21
..... 21
4.4. Results of phase 3 23
5. Discussion 24
5.1. Introduction to Chapter 24
5.2. Discussion for Phase 1 24
5.3. Discussion for the Phase 2 24
5.4. Discussion for the Phase 3 25
6. Conclusion 26
6.1. Introduction to Chapter 26
6.2. Conclusion 26
6.3. Limitation of present study 26
6.4. Recommendation for future Study 26
Appendix 1: Token Generation and Language model Python Script 27
Appendix 2: Python Script for the Twitter API data Scrape 30
Appendix 3: Extracted sample data from Twitter (for Given Sinhala Token) 32
Appendix 4: Python script for combining all the text files 33
Appendix 5: REST API python script for Next word Prediction 34
Appendix 6: Tokens Generated (sample) 35
7. References 36
8. Table of Figures 38