INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Real time Spelling & Grammar checker for Sinhala Language: A Multi-Agent Approach

A dissertation by

Ms. Samudi Hansani

W1800728 - 20200007

Supervised by

Mr. Mohamed Cassim Farook

Submitted in partial fulfillment of the requirements for the MSc in Advanced Software Engineering degree at the University of Westminster.

August 2022

ABSTRACT

Sinhala language is an agglutinative and morphologically rich language which is spoken by the majority of the Sri Lankans. It is apparent that the complexity of the language and the lack of knowledge on the spelling and grammar rules result in the spelling and grammar errors. The incorrect use of Sinhala language spelling and grammar in the electronic media leads to the circulation of invalid and meaningless information among a community. Thus, a spelling and grammar checker for Sinhala language is an important piece of work. However, owing to the limited availability of natural language processing resources, there exists an absence of a potential spelling and grammar checker for Sinhala language.

The research aims to design, develop and evaluate a multi agent system for real time spelling and grammar checking in Sinhala language. The ultimate prototype is an extension to Google Chrome browser which is capable of detecting spelling and grammar errors in real time and generating suggestions for grammar errors. The multi agent system is composed of two major agents for spell checking and grammar checking. A dictionary lookup approach is implemented for spell checking while a rule-based approach is applied in the grammar checker agent.

Module integration testing, functional testing and non-functional testing are conducted on the implemented prototype. Accuracy testing is conducted as a part of the non-functional testing under three major criteria as spelling error detection, grammar error detection and grammar suggestions generation. The spelling error detection functionality has a 90.90% lexical recall, 94.12% of error recall and 80% of precision. The accuracy of grammar error detection is 95% while recall, precision and f-measure are 91.91%, 100% and 95.78% respectively. The grammar suggestions generation accuracy is tested by the language experts with 99 grammatically incorrect sentences. According to the evaluation of the language experts, 75 of the generated suggestions are perfect, 3 of them are basically OK and 21 of the suggestions are incorrect.

Keywords - Multi Agent Systems, Natural Language Processing, Dictionary Look Up, Rule Based Approach, Spell Checking, Grammar Checking, Suggestions Generation