

SINHALA GRAMMAR CHECKER USING PARTS OF SPEECH TAGGING

P.A.S. Fernando* and T. Arudchelvam

Department of Computing and Information Systems, Wayamba University of Sri Lanka

*anneshanaf@gmail.com

Sinhalese was formerly called Sinhalese and is the language of Sinhalese. Sinhala has two forms such as spoken language and written language. Sinhala is very gentle in grammatical form and rigidity. Sinhala has its own writing system; it is descendants of Brahmi's book in India. There are many applications for Sinhala typing, Sinhala translation and also although there are many more tutorials for learning Sinhala. It is really difficult task to check the grammatical errors of a sentence. Detecting grammatical error is really important while doing documentations in Sinhala. The aim of this work is to detect grammatical mistakes, if any, of a given text. Parts of speech tagging is not a new concept. For Sinhala Language also many researches have been carried out to detect words and tag. In those works, statistical approaches such as Hidden Markov Model and stochastic models were used. But, in this study, tagging is used with the use of database built within the system. That type of approach has been used for Tamil language also. Sinhala Grammar Checker was developed with the influence of that work. In this work, the grammatical errors are found based on the agreement of "Subject" and the "Finite Verb" with the help of sub tags. As the language consists of many sub tags, we have identified most important ten sub-tags with the guidance of a Sinhala Language Expert. These sub-tags play the main role in detecting grammatical mistakes which are due to the disagreement between subjects and finite verbs with respect to person, number and/or gender, and tense disagreement between finite verbs and words that indicate tense aspects in the sentences. In this work, grammar checking for simple sentences is successfully done. But for complex sentences, some additional tags and methods are to be used. Further, a web based real time system is intended to be developed to check the grammatical correctness of a context written in Sinhala language.

Keywords: Sinhala Grammar, parts of speech, speech tagging, Hidden Markov model