Informatics Institute of Technology

In Collaboration With

University of Westminster, UK

Thamizhi – A Tamil Grammar Checker using NLP

A Dissertation

by

Miss. Rajkumar Shrikrishmethaa w1761222/ 2019217

Supervised by

Mr. Saman Hettiarachchi

April 2023

This document is submitted in partial fulfilment of the requirements for the BEng (Hons) Software engineering degree at the University of Westminster

ABSTRACT

Tamil Language is popularly known for its critic and rich in literature which has history for more than millenniums. It is complex and has evolved throughout the years and has been changed a lot in its writing methods. Lately, researchers and domain experts are making many attempts to automate the process of checking the grammatical errors in Tamil literature. After the pandemic, the world started working more virtually than working physically. So, there arose a need for digital copies of all critics and literature for easily accessible for students, teachers, writers, poets and etc. When trying to digitalize them it is hard for only man power to type and error check them. There is a sudden urge for people for an automated grammatical error checker.

Thamizhi is a web application which is capable of detecting the subject – verb disagreement errors in simple structured Tamil sentence (Subject – Object - Verb) using rule-based approach in Natural Language Processing. The application uses a set of predefined grammatical rules referenced from grammar books to ensure the accuracy in detecting the agreement errors without any defects.

The application has undergone many phases of testing to ensure the performance, accuracy, scalability and security and has shown excellent results. This application can be considered as an initial step to create an automated error checker for Tamil and help in preserving the valuable resources of the language.

Subject Description:

Tamil Error Detector, Natural Language Processing, Baseline Application, Stanza

Keywords:

Thamizhi, Tamil Grammatical Error Checker, Error Detector, Tamil, Natural Language Processing, Rule based learning, Stanza, POS tagging

Rajkumar Shrikrishmethaa | W1761222