

NLP based Autonomous grading system for Sinhala Language Essays of grade 5 students

P.W.Maddumage
Faculty of Computing, Sri Lanka
Institute of Information Technology
Malabe, Sri Lanka
it21007538@my.sliit.lk

W.M.C.I.Wijesinghe
Faculty of Computing, Sri Lanka
Institute of Information Technology
Malabe, Sri Lanka
it21006098@my.sliit.lk

J.H.T.Amarasena
Faculty of Computing, Sri Lanka
Institute of Information Technology
Malabe, Sri Lanka
it20604394@my.sliit.lk

Abstract— This study looks into the development and use of an autonomous grading system based on natural language processing (NLP) for essays written in Sinhala by students in the fifth grade. The system provides accurate and unbiased assessments of student writings by utilizing advanced natural language processing (NLP) techniques tailored to the grammatical and syntactic peculiarities of Sinhala. The primary goals are to ensure evaluation consistency, expedite the grading process, and provide students with helpful feedback. This article includes a detailed description of the system design, data collection and preprocessing methods, NLP algorithms employed, and first testing results.

Keywords— Automated Essay Grading, NLP, Text Processing, Essay Evaluation, Sinhala

I. INTRODUCTION

Essay writing is an essential part of language instruction since it helps pupils become more creative and coherent thinkers. But marking essays by hand requires a lot of work, subjectivity, and consistency issues. Due to differences in interpretation and exhaustion, teachers frequently struggle to maintain consistent grading standards. These issues can be resolved by an autonomous grading system that makes use of NLP by offering a standardized and effective evaluation procedure. Because of the distinctive linguistic features of the Sinhala language, which is mainly spoken in Sri Lanka, creating an NLP-based grading system is both difficult and fruitful. Computational models must handle the language's rich morphology, intricate script, and

variety of syntactic structures with particular care. The goal of this research is to develop an autonomous natural language processing (NLP) grading system designed for essays written by fifth-grade pupils and specifically customized for the Sinhala language. Essays are evaluated based on several factors, including as originality, coherence, language, and topic relevance. The system's mission is to increase student learning while decreasing instructor workload by automating grading and offering unbiased, consistent feedback.

II. LITERATURE REVIEW

As technology continues to advance rapidly, the realm of education is not immune to its transformative effects. One area that has seen significant progress is exam evaluation. Traditionally, grading exams has been a time-consuming and subjective process, prone to human error and bias. However, with the emergence of automated grading systems powered by Artificial Intelligence (AI), the landscape of exam evaluation is undergoing a revolutionary change.[1] Setting up an automatic grading system is one of the advantages of using an online examination system. This way of grading is much more convenient than the traditional approach. The expansion of technology in e-learning and the rising number of participants need the development of an error-free scoring process that lightens the load of the instructors, saves time, and guarantees impartiality.[2] The research problem at hand is the development of an autonomous grading system tailored for Sinhala language tests, leveraging the capabilities of Natural Language Processing (NLP). With the increasing integration of

technology in education, there is a growing need for automated assessment tools that can handle the nuances of languages beyond English, particularly in regions where diverse languages are predominant.

The motivation for this research stems from the unique linguistic characteristics of the Sinhala language, spoken by the majority of the population in Sri Lanka. Sinhala possesses a rich morphology and syntax, presenting distinct challenges for automated language processing. Existing autonomous grading systems primarily designed for English may not effectively capture the intricacies of Sinhala, limiting their applicability in educational contexts where Sinhala is the medium of instruction.[3],[4] The significance of this problem is underscored by the potential impact on education in Sri Lanka. By developing an autonomous grading system that understands and analyzes Sinhala language nuances, users can benefit from more efficient and accurate assessments, ultimately improving the learning experience for students.

Difficulties with Sinhala NLP

Given its rarity as a language studied in NLP, Sinhala poses particular difficulties. Specialized preprocessing models and procedures are needed because of its rich morphology, syntactic structures, and complex script. Prior research in Sinhala natural language processing (NLP) has concentrated on text classification, sentiment analysis, and machine translation; this work laid the foundation for more advanced applications, like essay grading.

Orthographic Variations: Standardization is crucial because there are several ways to express specific sounds in Sinhala script.

Morphological Richness: Strong morphological analysis methods are required since Sinhala words frequently have several morphemes.

Syntactic Complexity: Sentence structure changes in Sinhala require precise parsing and interpretation due to the language's flexibility.

III. FUNCTIONAL DESCRIPTION

1. Identifying native words and replace them with generic words

This function processes Sinhala text to identify and replace native words with generic equivalents. The input text is tokenized using a Sinhala-specific tokenizer. Each token is checked against a pre-defined dictionary of native and generic words. Identified native words are replaced with their corresponding generic words. Finally, the tokens are rejoined to form the transformed text, which is then output.

2. Misspelled Words & Complex Words Detection and Correction

This function identifies and corrects misspelled and complex words in a given text. The input text is first tokenized into individual words. Each word is checked against a standard dictionary to detect misspellings, and complex words are identified based on predefined criteria (e.g., length or rarity). Misspelled words are corrected using a spell-check algorithm, and complex words are replaced with simpler synonyms from a thesaurus. The corrected and simplified text is then reconstructed and returned.

3. Check the Grammar rules in the sentence

This function checks the grammar of a given sentence by analyzing its structure according to predefined grammatical rules. The sentence is first tokenized into words and phrases. Each token is then parsed to identify its part of speech (POS) and syntactic role. The function evaluates the sentence structure, agreement (subject-verb, noun-pronoun), and proper use of tenses against standard grammar rules. Any grammatical errors identified are flagged, and suggestions for correction are provided. Finally, the function returns a report detailing the grammatical accuracy and any necessary corrections.