

AUTONOMOUS GRADING SYSTEM FOR SINHALA LANGUAGE ESSAYS OF GRADE 5 STUDENT

Wijesinghe W.M.C.I-IT21006098

B.Sc. (Hons) Degree in Information Technology specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology Sri Lanka

February 2024

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Wijesinghe W.M.C.I	IT21006098	Sethania

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Supervisor Date

Jerry 29/02/2024

Ms. Jenny Krishara

Co-Supervisor

Date

Ms. Wishalya Thisara

ABSTRACT

This research focuses on enhancing the writing skills of Grade 5 scholarship students in Sinhala language through an innovative Automated Educational Assessment System. The system employs a set of 10 carefully crafted questions to evaluate children's knowledge, critical thinking, and expression skills. The specific component addressed in this research is the Missed Spelling Words Detection and Correction module.

The Missed Spelling Words Detection and Correction component plays a crucial role in ensuring the accuracy and language proficiency of students' responses. The system employs advanced Natural Language Processing (NLP) techniques to identify and rectify misspelled words in the answers provided by the students.

The research methodology involves two main processes: spelling error detection and correction. In the spelling error detection phase, the system identifies words with spelling mistakes, considering the context of the sentence. The correction phase utilizes a comprehensive dataset of correct spellings to replace the identified errors, ensuring the overall coherence and correctness of the response.

This component contributes to the broader system architecture, working in conjunction with other modules such as Grammar Check, Informal to Formal Word Replacement, Answer Extraction, Generalization, Grading System, and Similarity Check. The integrated approach ensures a holistic evaluation of students' responses, covering correctness, grammar, language proficiency, creativity, and critical thinking.

The Missed Spelling Words Detection and Correction module not only aids in improving the language proficiency of students but also contributes to the system's overall goal of providing accurate and culturally sensitive evaluations of Sinhala language proficiency. The research utilizes a brute-force searching approach to identify misspelled words and emphasizes the collection of a comprehensive dataset to enhance the system's accuracy.

By addressing the specific linguistic and cultural challenges posed by Sinhala language assessments, this research aims to advance the field of educational technology, offering a tailored solution that elevates the overall educational experience for Grade 5 scholarship students.

TABLE OF CONTENTS

DECLARATION	i
ABSTRACT	ii
LIST OF FIGURES	iv
LIST OF TABLES	iv
1.0 INTRODUCTION	v
1.1 BACKGROUND AND LITERATURE SURVEY	vi
1.2 RESEARCH GAP	viii
1.3 RESEARCH PROBLEM	X
3.0 OBJECTIVES	xii
3.1 Primary objectives	xii
3.2 Specific objectives	
4.0 METHODOLOGY	xv
4.1 System Architecture	xv
4.1.2 Overall System	
4.2 Data Collection	xvi
4.3 Algorithm Implementation:	xvi
4.4 Tools and technologies	xvi
4.5 Commercialization plan	xvii
4.6 Budget	xviii
5.0 PROJECT REQUIRMENTS	xix
5.1 Nonfunctional requirements	xix
5.2 Functional requirement	
5.3 System requirement	
6.0 WORK BREAKDOWN CHART	
8.0 REFERENCES	xxiii

LIST OF FIGURES

Figure 1	V
Figure 2	
Figure 3	
Figure 4	
Figure 5	
Figure 6	

LIST OF TABLES

Table 1: Comparisons between former research and the systems

Table 2: Budget

1.0 INTRODUCTION

In the multifaceted landscape of educational assessment, the endeavor to enhance the writing skills of Grade 5 scholarship students in Sinhala language is met with an innovative approach – the Automated Educational Assessment System. Within the intricate architecture of this system, my focused contribution revolves around the vital aspect of Missed Spelling Words Detection and Correction. This integral component aims to refine and elevate the language proficiency of students by addressing the nuances of spelling errors in their responses.

In the Sinhala language, where the intricacies of meaning can be context-dependent, the importance of accurate spelling cannot be overstated. The Missed Spelling Words Detection and Correction module employ advanced Natural Language Processing (NLP) techniques to meticulously identify and rectify spelling mistakes in the students' answers. This not only ensures correctness but also aligns with the broader objectives of the system in evaluating grammar, language proficiency, creativity, and critical thinking.

The methodology employed in this research encapsulates a comprehensive approach to spelling error detection and correction. Leveraging a brute-force searching technique, the system diligently examines the context of sentences to identify misspelled words. Furthermore, a curated dataset is utilized for correction, ensuring the replacement of errors maintains coherence and aligns with the linguistic context of Sinhala language usage.

As an indispensable part of the overall system architecture, the Missed Spelling Words Detection and Correction module collaborates synergistically with other components, such as Grammar Check, Informal to Formal Word Replacement, Answer Extraction, Generalization, Grading System, and Similarity Check. This collaborative approach ensures a holistic evaluation that goes beyond mere spelling correctness, contributing to the overarching goal of providing accurate, culturally sensitive assessments of Sinhala language proficiency.

Through this research, the intention is not only to improve the language skills of Grade 5 scholarship students but also to address the specific linguistic and cultural challenges inherent in Sinhala language assessments. By delving into the intricate details of spelling correctness, this research aims to make a meaningful contribution to the broader field of educational technology, creating a tailored solution that enhances the educational experience for students in the Sinhala language context.

1.1 BACKGROUND AND LITERATURE SURVEY

The landscape of educational assessment has witnessed a significant transformation with the advent of innovative technologies and methodologies. In the context of Sinhala language assessments for Grade 5 scholarship students, there exists a unique set of linguistic and cultural challenges. Ensuring accurate evaluation of students' language proficiency, creativity, and critical thinking skills demands a tailored approach. The Automated Educational Assessment System, designed to address these challenges, incorporates various components, each playing a specific role. This literature survey delves into the existing knowledge landscape, exploring relevant studies and technologies that form the foundation for the Missed Spelling Words Detection and Correction module.



Figure 1

Literature Survey:

Automated Assessment Systems: Numerous studies have highlighted the efficacy of automated assessment systems in enhancing the objectivity and efficiency of evaluating students' responses. These systems often employ Natural Language Processing (NLP) techniques to analyze and grade written content (Burstein et al., 2003). In the context of Sinhala language assessments, however, there is a scarcity of comprehensive studies addressing the specific linguistic intricacies.

NLP in Educational Technology: Natural Language Processing has been widely utilized in educational technology to analyze and understand human language. Studies in this domain emphasize the potential of NLP in improving language learning outcomes (Mostafa & Chen, 2019). In the Sinhala language context, leveraging NLP for spelling error detection and correction is a novel and unexplored area, warranting specific attention.

Language Specificity in Automated Assessment: Language-specific challenges play a pivotal role in the success of automated assessment systems. Studies highlight the need for tailored solutions that consider the linguistic nuances of specific languages to ensure accurate evaluations (Jing et al., 2014). Understanding the unique aspects of Sinhala language structure and usage becomes crucial in developing effective assessment tools.

Spelling Error Detection: Existing research on spelling error detection and correction underscores the importance of context-aware approaches. Incorporating contextual information improves the accuracy of identifying and rectifying spelling mistakes in written content (Agirre & Martinez, 2001). In the context of Sinhala language, where word meanings can be context-dependent, a similar approach becomes imperative.

Educational Technology in Sinhala Language Learning: A limited but growing body of literature focuses on the integration of technology in Sinhala language learning. These studies highlight the potential of technology in addressing linguistic challenges and improving language proficiency (Ranasinghe & Boralessa, 2018). However, there is a notable gap in specific research addressing automated spelling error detection in the educational context.

1.2 RESEARCH GAP

While the Automated Educational Assessment System outlined in this research offers a comprehensive approach to evaluating and enhancing the writing skills of Grade 5 scholarship students in Sinhala language proficiency, the focus of this section is on the specific research gap within the Missed Spelling Words Detection and Correction component.

Despite the strides made in natural language processing (NLP) and educational technology, there exists a noticeable research gap concerning the accurate detection and correction of misspelled words, especially within the context of Sinhala language assessments. The following areas represent the identified research gap:

- Limited Dataset for Sinhala Misspelled Words
- Complexities in Sinhala Language Patterns
- Integration of Binary Search for Efficient Correction
- Adaptation to Children's Writing Styles

Comparisons between former research and the systems

	[1] "Language model- based spell-checker for sri lankan names and addresses"	[2] "Erroff: A Tool to Identify and Correct Real-word Errors in Sinhala Documents"	[3]"An automated essay evaluation system using natural language processing and sentiment analysis"	[4]" Automated Essay Grading System using NLP Techniques"	"Dhara" Android Application
Integrate with mobile app	NO	NO	NO	NO	YES
Spell checking	YES	YES	YES	YES	YES
Correct Spellings	NO	YES	YES	YES	YES
Targeting school students scope	NO	NO	NO	YES	YES
Using only Sinhala Language	YES	YES	NO	NO	YES

Figure 2

	[1]"Language model-based	[2]"Erroff: A Tool to	[3]"An automated	[4]"Aut omated	"Dhara
	spell-checker for Sri Lankan names and address"	identify and correct Real-word Errors in Sinhala	essay evaluation system using natural language	essay grading system using NLP Techni	Androi d Applica tion
		Document s"	processing and sentiment analysis"	ques"	
Integrate with mobile app	NO	NO	NO	NO	YES
Spell checking	YES	YES	YES	YES	YES
Correct spellings	NO	YES	YES	YES	YES
Targeting school student scope	NO	NO	NO	YES	YES
Using only Sinhala Language	NO	YES	NO	NO	YES

Table-1

1.3 RESEARCH PROBLEM

In the pursuit of developing an innovative Automated Educational Assessment System for Grade 5 scholarship students, the Missed Spelling Words Detection and Correction component encounters several research problems that warrant focused attention.

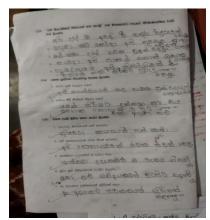


Figure 3

• Inadequate Sinhala Misspelling Datasets:

The lack of a comprehensive and diverse dataset containing a range of misspelled words in Sinhala poses a significant challenge. Without a robust dataset, the system may struggle to accurately identify and correct misspelled words across various contexts and linguistic intricacies.

• Complexity of Sinhala Language Patterns:

The complex and nuanced nature of Sinhala language patterns introduces difficulties in accurately detecting misspelled words. The research problem revolves around understanding how the system can effectively navigate the intricate sentence structures and linguistic nuances inherent in Sinhala language assessments.

• Optimizing Binary Search for Sinhala Language:

While binary search presents a promising approach for efficiently correcting misspelled words, adapting and optimizing this algorithm to suit the specific linguistic characteristics of Sinhala is a challenging problem. The research must explore how to tailor binary search to effectively handle Sinhala language intricacies.

• Dynamic Adaptation to Children's Writing Styles:

Children exhibit diverse writing styles that evolve as their language proficiency develops. The research problem involves devising a mechanism within the Missed Spelling Words Detection and Correction module that dynamically adapts to the varying writing styles of children, ensuring accurate correction across different proficiency levels.

• Balancing Precision and Processing Speed:

Achieving a balance between the precision of misspelled word correction and the processing speed of the system poses a research challenge. It is essential to design the correction mechanism in a way that maintains high accuracy while ensuring swift and efficient processing, especially within the constraints of educational assessment settings.

• Cultural Sensitivity in Correction Suggestions:

The research problem extends to addressing the cultural nuances in suggesting corrections. Ensuring that the system provides culturally sensitive and contextually appropriate corrections for misspelled words in Sinhala is a vital consideration.

Addressing these research problems is crucial for the successful implementation and effectiveness of the Missed Spelling Words Detection and Correction module within the broader context of the Automated Educational Assessment System. By tackling these challenges, the research aims to contribute to the development of a sophisticated and culturally aware tool for enhancing Sinhala language proficiency among Grade 5 scholarship students.

3.0 OBJECTIVES

3.1 Primary objectives

The primary objectives of the Missed Spelling Words Detection and Correction component within the Automated Educational Assessment System are outlined as follows

• Develop a Comprehensive Sinhala Misspelling Dataset:

Create an extensive and diverse dataset containing a wide array of misspelled words in Sinhala. This dataset should cover various linguistic contexts, sentence structures, and proficiency levels to ensure the robust training and testing of the Missed Spelling Words Detection and Correction module.

• Enhance Accuracy in Sinhala Language Pattern Recognition:

Improve the accuracy of the misspelled word detection mechanism by incorporating advanced natural language processing (NLP) techniques. Focus on understanding and addressing the intricate language patterns inherent in Sinhala, ensuring precise identification of misspellings within diverse sentence structures.

• Optimize Binary Search Algorithm for Sinhala Correction:

Fine-tune and optimize the binary search algorithm for efficient correction of misspelled words in Sinhala. Explore adaptations and enhancements to the algorithm that specifically cater to the linguistic characteristics and nuances of Sinhala language patterns, maintaining both precision and speed.

• Dynamic Adaptation to Children's Writing Styles:

Develop a dynamic adaptation mechanism within the Missed Spelling Words Detection and Correction module. This mechanism should be capable of understanding and accommodating the evolving writing styles of children, ensuring accurate corrections across different proficiency levels and linguistic developmental stages.

• Balance Precision and Processing Speed:

Strike a balance between the precision of misspelled word correction and the processing speed of the system. Design the correction mechanism to deliver high accuracy in identifying and rectifying misspellings while optimizing processing speed to align with the time constraints of educational assessments.

• Ensure Cultural Sensitivity in Correction Suggestions:

Incorporate cultural sensitivity into the correction suggestions provided by the system. Tailor the correction recommendations to align with the cultural nuances of Sinhala language usage, ensuring that the suggested corrections are contextually appropriate and respectful of the cultural diversity inherent in Sinhala assessments.

By accomplishing these primary objectives, the Missed Spelling Words Detection and Correction module aims to contribute significantly to the overarching goal of the Automated Educational Assessment System, offering a sophisticated and culturally aware tool for improving Sinhala language proficiency among Grade 5 scholarship students.

3.2 Specific objectives

Building upon the primary objectives, the specific objectives of the Missed Spelling Words Detection and Correction component within the Automated Educational Assessment System are delineated as follows:

• Compile a Sinhala Misspelling Dataset:

Curate a comprehensive dataset consisting of commonly misspelled words in Sinhala. Include variations in spelling errors, considering the linguistic nuances and diverse sentence contexts. This dataset will serve as the foundation for training and evaluating the accuracy of the correction mechanism.

• Implement Part-of-Speech Tagging for Misspelled Words:

Integrate part-of-speech tagging techniques to accurately identify misspelled words within the Sinhala text. This involves analyzing the grammatical structure and context to pinpoint potential errors, enhancing the precision of the misspelling detection process.

• Develop Linguistic Rules for Correction Suggestions:

Formulate linguistic rules specific to Sinhala language patterns to guide the correction suggestions. These rules should encompass variations in spelling based on context, ensuring that the correction mechanism provides accurate and culturally appropriate suggestions for different sentence structures.

• Refine Binary Search for Sinhala Correction:

Refine the binary search algorithm to efficiently correct misspelled words in Sinhala. Optimize the algorithm parameters considering the intricacies of Sinhala linguistic structures, and conduct rigorous testing to validate its effectiveness in providing accurate corrections.

• Incorporate Machine Learning for Adaptive Proficiency Levels:

Integrate machine learning components to enable the system to adapt dynamically to varying proficiency levels exhibited in children's writing styles. Train the model to recognize evolving linguistic patterns and adjust the correction mechanism accordingly to accommodate the diverse skill levels of Grade 5 scholarship students.

• Evaluate Processing Speed and Accuracy Trade-offs:

Conduct a systematic evaluation of the misspelling correction mechanism, considering both processing speed and accuracy. Explore trade-offs and adjustments to strike a balance between

swift processing, essential for real-time assessments, and maintaining a high level of precision in identifying and correcting misspelled words.

• Ensure Ethical Handling of Cultural Sensitivity:

Implement ethical considerations in the development of correction suggestions, ensuring cultural sensitivity and respect for the diverse linguistic and cultural backgrounds of Sinhala language users. Regularly assess and refine the system's output to minimize biases and promote inclusivity. By achieving these specific objectives, the Missed Spelling Words Detection and Correction module aims to contribute a refined and adaptive tool to the Automated Educational Assessment System, fostering improved Sinhala language proficiency among Grade 5 scholarship students through accurate and culturally sensitive spelling error correction

4.0 METHODOLOGY

The methodology for developing and implementing the Missed Spelling Words Detection and Correction component within the Automated Educational Assessment System involves a systematic approach encompassing data preparation, algorithm implementation, and evaluation procedures.

4.1 System Architecture

4.1.2 Overall System

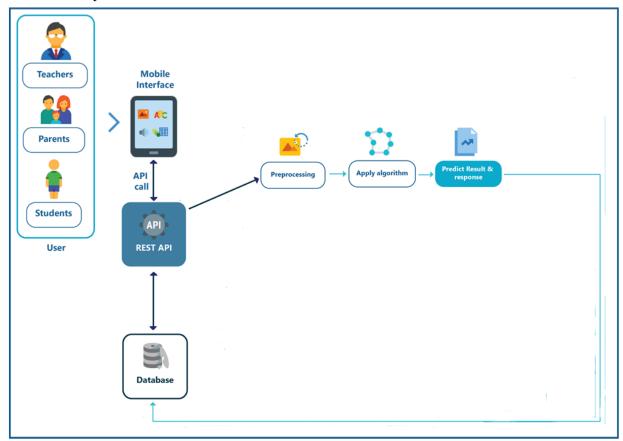


Figure 4

4.2 Data Collection:

• Compilation of Sinhala Misspelling Dataset:

Source diverse Sinhala texts, including educational materials, literature, and common writing samples. Manually identify and annotate misspelled words within the selected texts, considering variations in spelling errors. Organize the compiled dataset into categories based on linguistic contexts and sentence structures.

4.3 Algorithm Implementation:

• Integration of Part-of-Speech Tagging:

Implement part-of-speech tagging techniques using NLP libraries to identify potential misspelled words within the Sinhala text.Utilize linguistic rules to filter false positives and refine the list of identified misspelled words.

• Linguistic Rule Formulation:

Develop linguistic rules specific to Sinhala language patterns, considering variations in spelling based on grammatical context and sentence structures.

Create a rule-based system to guide the correction suggestions for identified misspelled words. Refinement of Binary Search Algorithm:

Adapt the binary search algorithm for efficient correction of misspelled words in Sinhala.

Optimize algorithm parameters, taking into account the linguistic intricacies and nuances of Sinhala language patterns.

• Incorporation of Machine Learning Components:

Integrate machine learning components to enable the system to dynamically adapt to varying proficiency levels in children's writing styles. Train the model on the compiled dataset to recognize evolving linguistic patterns and adjust the correction mechanism accordingly.

4.4 Tools and technologies

Web frameworks & libraries

- React JS JavaScript frontend framework to UI development
- Python Django Python framework to web development
- Node JS JavaScript framework to backend development

Machine learning & deep learning libraries

- Open CV (Open Source Computer Vision Library) A software library which enables users to use machine learning and open source computer vision in various manners
- YOLO (you only look once) An algorithm which detects objects real-time.
- TensorFlow An end to end open source platform which is used to machine learning
- CNN (Convolutional Neural Network) A class of neural network, specializing in processing data and having a topology which is grid.

Database

- AWS Cloud (Cloud computing platforms)
- Database provides API on a metered payment basis.

Tools

- VS Code for developers Visual Studio code code editor.
- PyCharm Python IDE to professional developers

4.5 Commercialization plan

The commercialization plan for the Missed Spelling Words Detection and Correction component involves strategic steps to introduce and monetize the technology within the educational sector

Target Audience:

Primary Focus: Educational institutions, schools, and language learning centers.

Secondary Focus: Educational technology companies, publishers, and language assessment organizations

Marketing and Promotion:

Develop a comprehensive marketing strategy targeting educators, school administrators, and decision-makers in the education sector.

Utilize digital marketing channels, educational conferences, and partnerships with educational associations for promotion.

Create engaging content, including case studies and demonstrations, to showcase the effectiveness of the technology.

By implementing these elements in the commercialization plan, the Missed Spelling Words Detection and Correction component can effectively penetrate the educational market, providing a valuable tool for improving Sinhala language proficiency among Grade 5 scholarship students.

4.6 Budget

Item	(USD) Budget	(LKR) Budget
Desing tool	25.03	9000.00
Printing documents	10.26	4500.00
field visits	16.52	5000.00
Server Cost	6.50	3000.00
Total	52.31	21500.00

5.0 PROJECT REQUIRMENTS

5.1 Nonfunctional requirements

- ✓ Performance- Response Time: The system should provide real-time correction suggestions with a response time not exceeding [specified time] to ensure seamless integration within the assessment process.
- ✓ Scalability- The solution must be scalable to accommodate an increasing number of users and a growing dataset without compromising performance.
- ✓ Accuracy and Precision-The misspelling detection and correction mechanism should achieve an accuracy rate of [specified percentage] or higher to ensure reliable and precise identification and correction of misspelled words in Sinhala.
- ✓ Adaptability-The system should adapt dynamically to different proficiency levels exhibited in children's writing styles, ensuring accurate corrections across a spectrum of linguistic development.
- ✓ Cultural Sensitivity-Correction suggestions should align with cultural norms and sensitivities of Sinhala language usage, avoiding offensive or inappropriate recommendations. Regular reviews and updates should be conducted to address any cultural biases.
- ✓ Security-All collected datasets, especially those containing personally identifiable information, should adhere to strict security protocols to safeguard the privacy of individuals.

.

5.2 Functional requirement

✓ Misspelling Detection

The system should accurately identify misspelled words within Sinhala language texts.

Utilize part-of-speech tagging and linguistic rules to enhance the precision of misspelling detection. Consider variations in spelling errors, including typos, phonetic variations, and contextual misspellings.

✓ Correction Suggestions

Provide correction suggestions for identified misspelled words.

Implement a rule-based correction mechanism guided by linguistic rules specific to Sinhala language patterns. Offer multiple correction options when applicable, prioritized based on linguistic context and common usage.

✓ Adaptive Proficiency Levels:

Dynamically adapt to varying proficiency levels in children's writing styles.

Utilize machine learning components to recognize evolving linguistic patterns and adjust the correction mechanism accordingly .Accommodate diverse skill levels and linguistic developmental stages exhibited by Grade 5 scholarship students.

✓ Integration with System Architecture:

Seamlessly integrate with the overall Automated Educational Assessment System architecture.

Communicate effectively with other modules such as Answer Extraction, Generalization, and Grading System to contribute to a comprehensive evaluation process.

5.3 System requirement

The system requirements for the Missed Spelling Words Detection and Correction component within the Automated Educational Assessment System encompass both hardware and software specifications necessary for optimal performance:

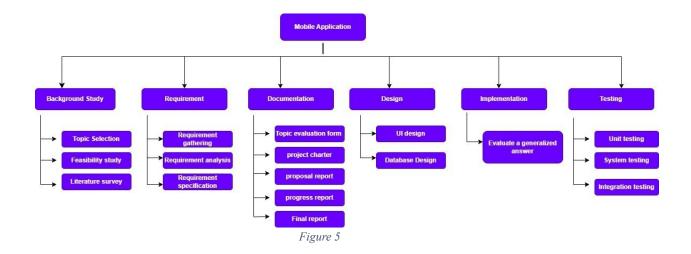
Hardware Requirements:

- Processing Power:Minimum: Dual-core processor
- Recommended: Quad-core processor or higher
- Memory (RAM):Minimum: 4 GB
- Recommended: 8 GB or higher
- Storage: Minimum: 100 GB HDD or equivalent SSD
- Recommended: 256 GB SSD or higher for efficient data retrieval
- Network Connectivity: Stable internet connection for real-time correction suggestions and access to external language resources.

Software Requirements:

- Operating System: Windows 10 or Linux (Ubuntu 18.04 LTS or later)macOS support for client-side interfaces (optional)
- Programming Languages:Python 3.x for algorithm implementation Java for system integration
- Database Management System:MongoDB for efficient storage and retrieval of linguistic datasets
- Web Framework (if applicable): Django or Flask for web-based interfaces
- Natural Language Processing Libraries: NLTK (Natural Language Toolkit) for part-of-speech tagging
- spaCy for linguistic analysis and rule-based corrections
- Machine Learning Framework (if applicable):
- TensorFlow or PyTorch for implementing adaptive proficiency level recognition
- Security Tools:Secure Sockets Layer (SSL) for secure data transmission Encryption algorithms for securing personally identifiable information
- Documentation and Collaboration Tools:Sphinx or similar tools for creating comprehensive documentation ,Collaboration tools (e.g., Git, Jira) for version control and issue tracking

6.0 WORK BREAKDOWN CHART



7.0 GANTT CHARTS

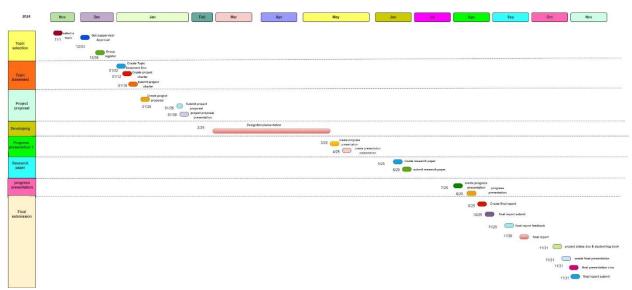


Figure 6

8.0 REFERENCES

- [1] Y. Udagedara, B. Elikewela, and U. Thayasivam, "Language model-based spell-checker for sri lankan names and addresses," in 2022 Moratuwa Engineering Research Conference (MERCon). IEEE, 2022, pp. 1–6
- [2] P. Sudesh, D. Dashintha, R. Lakshan, and G. Dias, "Erroff: A Tool to Identify and Correct Realword Errors in Sinhala Documents," in 2022 Moratuwa Engineering Research Conference (MERCon). IEEE, 2022, pp. 1–6.
- [3] Vijaya Shetty, SKadagathur Raghavendra, Rao GuruvyasPranav, Prashantha PatilShow, Gunakimath Suryakanth, "An automated essay evaluation system using natural language processing and sentiment analysis," December 2022
- [4] Kaushal Yadav, "Automated Essay Grading System using NLP Techniques," 2020, International Journal of Engineering and Advanced Technology
- 41. S. G C and Y. Zhang, "nc-nd/4.0/). Weed and crop species classification using computer vision and deep learning technologies in greenhouse conditions," sciencedirect, Sep. 09, 2022. http://creativecommons.org/licenses/by ncnd/4.0/