

Plagiarism Detection System

Mrs. M. Sathya, S Nithikamalini, S Pradeepa, G Preethi Roja, R Ragavi

Professor, Department of Information Technology

Students, Department of Information Technology

Sri Shakthi Institute of Engineering and Technology, Tamil Nadu, India

Abstract: Plagiarism is a significant issue in academic, research, and professional writing, undermining originality and ethical standards. This Plagiarism Detection System is designed to address this challenge by providing an efficient, accurate, and user-friendly solution for detecting similarities in textual content. Using N-gram analysis and Cosine Similarity, the system identifies exact matches, paraphrased content, and partial similarities, ensuring a comprehensive evaluation of documents. Users can upload text files, manually input content, or use a voice-to-text feature for plagiarism detection, making it accessible to a wide range of users, including students, educators, and content creators. The system generates detailed reports, highlighting the plagiarized sections, their similarity percentage, and possible sources while ensuring data security and user privacy. It is built with a focus on efficiency, accuracy, and usability, minimizing false positives and improving the reliability of results. The platform's intuitive interface allows for easy navigation, while its real-time processing enables quick analysis of large documents. This system not only helps users maintain academic integrity but also fosters ethical writing practices in professional and educational environments. Future enhancements may include multilingual support, and integration with large-scale databases to improve accuracy and expand its applicability across diverse fields

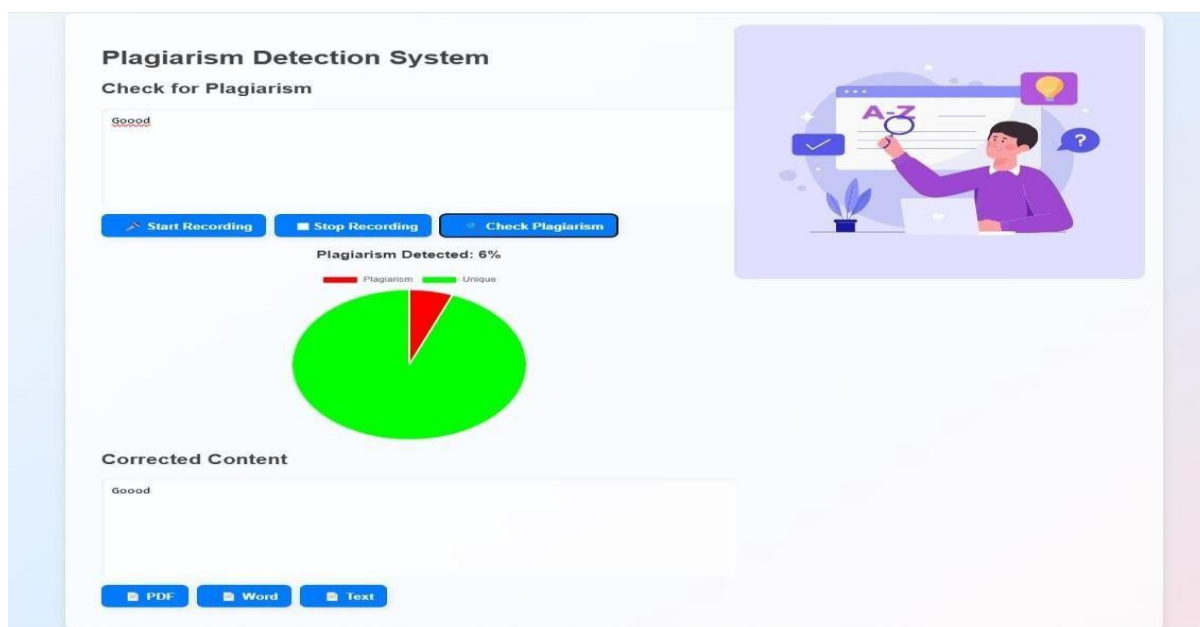
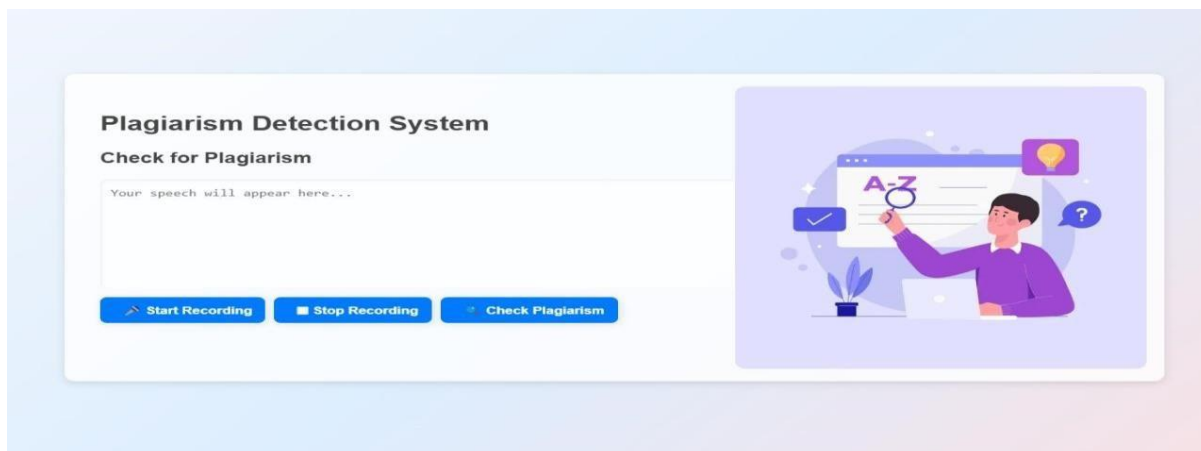
Keywords: Plagiarism, N-gram analysis, Cosine Similarity, Text Processing, Academic Integrity

I. INTRODUCTION

Plagiarism detection is an essential tool in modern academia and content creation. With the vast amount of information available online, distinguishing between original and copied content has become increasingly challenging. This paper presents a plagiarism detection system that utilizes advanced text analysis techniques such as N-gram similarity and Cosine Similarity to ensure the authenticity of written documents. This system is particularly beneficial for students, educators, researchers, and publishers, providing a robust tool for maintaining integrity in academic and professional writing.

Figures





II. CONCLUSION

This study introduced a plagiarism detection system that leverages N-gram analysis and Cosine Similarity to identify textual similarities with high precision. The system successfully detects both direct plagiarism and paraphrased content, aiding users in maintaining academic and professional integrity. Future work aims to integrate deep learning models for enhanced semantic analysis, expand database coverage for more comprehensive detection, and introduce multilingual support. Additionally, improvements in detection methodologies and broader data source integration will further refine the system's ability to detect complex similarities. By continuously refining detection techniques and improving user experience, this system aspires to be a vital tool in promoting ethical writing and safeguarding intellectual property rights.



ACKNOWLEDGMENT

The authors would like to acknowledge the contributions of their respective institutions in supporting this research. Special thanks to the research and development teams for their invaluable insights and technical assistance. The financial support provided by the institutions has played a crucial role in the successful implementation of this study. We also extend our gratitude to colleagues and peer reviewers for their constructive feedback and encouragement, which significantly contributed to refining the research. Their expertise and guidance have been instrumental in improving the accuracy and reliability of the proposed plagiarism detection system.

REFERENCES

- [1]. Baldwin, M., Chang, C., Gravano, L., & Paepcke, A.: Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108-121
- [2]. Bruce, K.B., Cardelli, L., & Pierce, B.C.: Comparing Object Encodings. Springer-Verlag, Berlin Heidelberg New York (1997)

