International Journal of Research Publication and Reviews

Plagiarism and Grammar Checker using Speech Recognition

T S Sowmya[1], Adisree R[2], Aarthi R[3], Vadivelu A[4]

[1,2,3]AIML Students, [4]Guide

SRM Institute of Science and Technology, Ramapuram, Chennai

ABSTRACT:

Nowadays, plagiarism and poor grammar in written content keep to pose full-size challenges to the integrity of instructional, professional, and innovative works. There is a growing want for innovative gear that may help writers and educators in retaining the very best requirements of authenticity and language talent. The mission is to create a grammar and plagiarism checker that makes use of speech reputation to take a look at spoken data for faults in grammar, plagiarism, and stylistic enhancements. The answer focusses on delivering a powerful and precise software that helps writers, content creators, experts, students, and others create unique, excellent, and nicely-structured written material. It ambitions to create algorithms that transform spoken speech with a big library of previously published texts to spot feasible plagiarism and use present day strategies to distinguish between suitable paraphrasing and blatant copying. It primarily involves integrating natural language processing (NLP) algorithms to observe the spoken content's grammar, sentence production, writing style and deliver recommendation on a way to make sentences greater succinct, coherent, and clear. It utilizes modern-day speech recognition generation to faithfully convert spoken input to textual content. This transcription will serve as the input for plagiarism and grammar evaluation.

Keywords: Speech recognition, Grammar Checker, Plagiarism Detection, Natural Language Processing.

Introduction:

In an era characterized by the ubiquity of information and the relentless pursuit of excellence in written communication, plagiarism and grammatical errors continue to be pervasive challenges. These issues compromise the integrity and quality of academic papers, professional documents, and creative works, undermining the credibility of writers and the educational institutions they represent. However, in the pursuit of originality and linguistic precision, there is a growing need for innovative tools that can not only detect instances of plagiarism and grammar issues but also bridge the gap between spoken and written language. The traditional approach to plagiarism detection has relied primarily on the scrutiny of written text, comparing documents against vast databases of pre-existing works to uncover instances of unoriginal content. While these tools have been invaluable in upholding academic and professional integrity, they have limitations, particularly when it comes to detecting plagiarism within spoken language or identifying poorly structured sentences. This is where the integration of speech recognition technology becomes indispensable.

The integration of speech recognition technology into a plagiarism and grammar checker presents several compelling advantages. For writers, it offers a sophisticated means of self-auditing their work, mitigating the risk of unintentional plagiarism and improving the overall quality of their writing. Educators benefit by having a robust tool that facilitates the verification of student submissions for originality and offers valuable feedback on grammar and language proficiency. Additionally, this tool serves as an inclusive resource, aiding those with disabilities who find speech-to-text technology more accessible in their content creation processes.

Problem Statement:

The proliferation of content creation across various platforms, such as academic essays, professional documents, and online content, has raised significant concerns about plagiarism and language quality. Existing plagiarism and grammar checkers primarily focus on written text, often missing instances of plagiarism or grammatical errors in spoken language. Additionally, the current tools lack the integration of voice recognition technology, making it challenging for users to effectively address these issues in both their written and spoken communication. There is a growing need for a comprehensive solution that combines plagiarism detection, grammar correction, and speech recognition capabilities. The current tools fall short in effectively identifying and rectifying plagiarism in spoken language, thereby compromising the integrity of academic and professional discourse.

Furthermore, users often resort to separate tools for grammar checking, resulting in a disjointed and less efficient editing process. Therefore, the problem at hand is to develop a sophisticated plagiarism and grammar checker that seamlessly integrates speech recognition technology, enabling users to detect and correct plagiarism in spoken language while simultaneously providing real-time grammar and style suggestions. This solution should address the existing limitations, improve the accuracy of plagiarism detection, enhance the quality of spoken and written content, and offer a user-friendly interface to cater to the diverse needs of academics, professionals, and content creators.

Proposed Work:

The proposed system of a plagiarism and grammar checker that integrates speech recognition technology aims to comprehensively address the challenges of plagiarism detection, grammar checking, and linguistic analysis across both spoken and written language. The system employs NLP algorithms to extract meaning, context, and semantics from both spoken and written content. This ensures that the system comprehends language structure, grammar rules, and context, facilitating accurate grammar checking and plagiarism detection. The proposed system utilizes plagiarism detection algorithms that compare transcribed text against an extensive database of existing works. It excels in identifying similarities, paraphrasing, and cross-lingual plagiarism. Additionally, advanced grammar-checking software identifies and corrects grammatical errors while offering contextually appropriate suggestions for improvements. It is characterized by its versatility, accuracy, and user-centric design. It empowers educators to maintain academic integrity, assists writers and professionals in producing high-quality content, and supports language learners and non-native speakers in their linguistic journey. By embracing the auditory dimension of communication, the system reflects the evolving needs of users in an interconnected world where spoken language is of paramount importance.

Scope and Motivation:

A speech recognition-based plagiarism and grammar checker may have many applications and provide creative approaches to evaluating spoken content. The program might be used by educators and students to assess the veracity of oral assignments and presentations, assisting in the fight against plagiarism. The grammar, pronunciation, and fluency of the pupils' spoken language might also be assessed by teachers. Podcasters and YouTubers might use the program to check the originality of their audio material and get grammatical recommendations for their scripts. With the use of technology, people with speech or language problems may be able to communicate more effectively. Additionally, it could provide feedback on their spoken material for those with visual impairments through accessibility tools. Language learners may get immediate feedback on their usage of the language, grammar, and pronunciation, which would help them get better at speaking. Based on spotted grammar errors, the application might provide workouts or drills.

Project Domain:

Natural Language Processing (NLP) is the backbone of innovation in the domain of plagiarism and grammar checking, especially when combined with speech recognition technology. NLP plays a pivotal role in enabling this multifaceted tool to transcend the boundaries of traditional written text analysis and address the complexities of spoken language. Here, we delve into the essential ways in which NLP is harnessed in this project. NLP is the driving force behind the transformation of spoken words into written text. Speech recognition technology, a subset of NLP, accurately transcribes oral content, allowing for further analysis.

NLP algorithms are employed to compare the transcribed text against an extensive database of existing works. By analysing the semantics, structure, and context of text, the tool can detect instances of plagiarism more effectively. It can flag similarities in meaning, even if the words are rephrased, thus enhancing the tool's capacity to identify unoriginal content. NLP is instrumental in grammar checking by recognizing the rules and structures of a given language. It identifies grammatical errors, punctuation issues, and style inconsistencies, and offers suggestions for improvements. The tool's ability to provide meaningful and contextually appropriate grammar suggestions relies heavily on NLP's understanding of language rules.

Review of Related Literature:

- Vilmar Santos Nepomuceno Sergio Soares in year 2020, Avoiding Plagiarism in Systematic Literature Reviews: An Update Concern. This paper's approach includes conducting a survey and semi-structured interviews with specialists performing Systematic Literature Reviews (SLRs) to determine their knowledge of SLR plagiarism and how to avoid it. The writers also used a plagiarism analyzer to compare SLR updates to their original papers in order to discover potential plagiarism scenarios. The report outlines a number of best practices for avoiding plagiarism in SLR maintenance.

- Vanja Pupovac in the year 2021, The frequency of plagiarism identified by text-matching software in scientific articles: a systematic review and meta-analysis. A two-step process was used to assess plagiarism, first identifying textual similarity using text-matching software and then inspecting detected similarity in the human verification process. As a result, it's safe to assume that the text-matching software used an algorithm to detect textual similarity between articles.

Module Description:

•NUMPY - In a plagiarism and grammar checker system using speech recognition, the NumPy module can play a significant role in enhancing performance and functionality, particularly in the data processing and analysis aspects. NumPy, a fundamental library for numerical computations in Python, provides support for multidimensional arrays, mathematical functions, and various data manipulation capabilities. Incorporating NumPy into the plagiarism and grammar checker using speech recognition can greatly enhance the system's data handling and processing capabilities. It empowers the system to efficiently manage audio and text data, perform linguistic analysis, and conduct statistical assessments. Additionally, NumPy's support for data

normalization and performance optimization contributes to the overall efficiency and accuracy of the system, ensuring that it can effectively address the challenges of content evaluation and enhancement.

•GRAMFORMER - Gramformer is a powerful natural language processing model designed to improve the quality and correctness of text. While it is not typically used as a speech recognition tool, it can play a valuable role in a plagiarism and grammar checker system that utilizes both written and spoken language. While Gramformer is not a speech recognition tool, its strength lies in its ability to refine and enhance text, which is highly valuable in a plagiarism and grammar checker system that deals with both spoken and written content. By integrating Gramformer, the system can offer users a more comprehensive and sophisticated solution for content evaluation, ensuring that their text is not only free from plagiarism but also grammatically correct and contextually appropriate, whether it originates from written or transcribed sources.

•NLP - Incorporating NLP into a plagiarism and grammar checker using speech recognition significantly enhances the system's accuracy, capability, and sophistication. By leveraging NLP's ability to understand language structure, meaning, and context, the system provides users with more accurate and contextually relevant suggestions for grammar improvement and offers robust plagiarism detection. This comprehensive approach ensures that both spoken and written content is evaluated at a high linguistic standard, meeting the evolving needs of users in an interconnected world where spoken language is increasingly significant.

•LEVENSHTEIN DISTANCE - The Levenshtein distance, also known as the edit distance, is a valuable algorithm in a plagiarism and grammar checker system that employs speech recognition technology. It is used to quantify the difference between two strings in terms of the minimum number of single-character edits required to transform one string into the other. These edits can include insertion, deletion, or substitution of characters. It is more commonly associated with text comparisons and string manipulation. The Levenshtein distance can assist in evaluating the consistency of language and style in a piece of text. By comparing various sections of text, the system can assess whether the style remains consistent throughout the document. Significant deviations may indicate potential issues that need to be addressed. By leveraging this algorithm, the system can identify and correct transcription errors, assess the similarity of text segments, evaluate grammar and style consistency, and provide valuable feedback to users. It contributes to the overall quality and reliability of the system's analysis, ensuring that both spoken and written content meets high linguistic and authenticity standards.

Conclusion

In conclusion, the integration of speech recognition technology into plagiarism and grammar checkers marks a significant leap forward in the realm of language assessment and content evaluation. This innovative approach enhances the functionality and accessibility of these tools, providing users with a comprehensive solution for spoken and written language analysis. There are numerous benefits of integrating speech recognition with plagiarism and grammar checkers.Firstly, it greatly improves accessibility, accommodating individuals with diverse needs and preferences. Users who rely on spoken language or have accessibility requirements can now effectively utilize language assessment tools. This inclusivity is a fundamental aspect of modern software development, and the integration of speech recognition exemplifies this commitment to a wider user base.

Contextual analysis and advanced error detection are made possible through speech recognition. The technology goes beyond individual words and takes into account pronunciation, intonation, and spoken language nuances. This depth of analysis enriches the system's understanding of language, enabling more meaningful and contextually relevant feedback. The integration of speech recognition, however, is not without its challenges. These include ensuring the accuracy of transcriptions, handling diverse accents and languages, and maintaining a seamless user experience. As technology continues to advance, these challenges are being addressed and overcome. In essence, the integration of speech recognition technology into plagiarism and grammar checkers is a step towards a more inclusive, efficient, and powerful language assessment solution. It exemplifies the commitment to innovation and user-centric design in the world of language technology. With ongoing advancements in speech recognition and natural language processing, the future holds great promise for the continued evolution of these essential language tools, further enhancing communication and content quality for users across various domains.

References:

Research Papers:

1. Carroll, J. (2002) A Handbook For Deterring Plagiarism In Higher Education, Ocsld, Oxford Brookes University, Oxford.

2. Dr Jeremy B. Williams, Universitas 21 Global ,Plagiarism: Deterrence, Detection And Prevention, Zayed University.

3. Hsien-Chin Liou (2000) ,Development Of An English Grammar Checker A Progress Report .

4. https://ieeexplore.ieee.org/abstract/document/8285846

5. https://ieeexplore.ieee.org/document/6657120

6. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1362619

7. https://www.researchgate.net/publication/271339485_Effectiveness_of_Coaching_Interventions_using_Grammarly_Software_and_Plagiarism_Detection_Software_in_Reducing_Grammatical_Errors_and_Plagiarism_of_Undergraduate_Researches

8. https://www.researchgate.net/publication/331062965_Overview_of_Plagiarism_Checkers_and_Plagiarism_Detection_Tools_A_Study

9. https://iopscience.iop.org/article/10.1088/1742-6596/1651/1/012135/pdf

10. https://www.researchgate.net/publication/347180164_A_Research_on_Online_Grammar_Checker_System_Based_on_Neural_Network_Model

11. https://www.hindawi.com/journals/sp/2022/4082082/

12. https://www.researchgate.net/publication/331062965_Overview_of_Plagiarism_Checkers_and_Plagiarism_Detection_Tools_A_Study