IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

LinguaGuardian: Your Language Companion

Sayali Jadhav[1], Omkar Karpe[2], Anshul Pathrabe[3], Amruta Kate[4], Prof. Rashmi Kale[5]
Students, Department of Computer Engineering[1,2,3,4]
Professor, Department of Computer Engineering[5]
Smt. Kashibai Navale College of Engineering, Pune, India

Abstract: *Lingua Guardian: Your Language Companion is a comprehensive solution for modern communication needs, offering precise translations, grammar correction, and plagiarism detection. Leveraging cutting-edge language models and machine learning, our Text Translator facilitates seamless communication across linguistic boundaries, empowering users to express themselves with clarity. The Grammar Checker ensures that writing meets the highest standards of accuracy and style, providing real-time feedback for academic, professional, and creative projects. Upholding academic integrity, our Plagiarism Detector scans content against a vast database, aiding users in attributing sources and maintaining credibility. From students to professionals, our program caters to diverse users, prioritizing user-friendliness, security, and performance. LinguaGuardian stands as an indispensable tool for those who value effective communication and original thought in the digital age.*

Keywords: *Lingua Guardian*

I. INTRODUCTION

In today's digitally interconnected world, the demand for effective and efficient language processing tools has never been more critical. The rapid globalization of communication and the proliferation of digital content have underscored the need for sophisticated solutions that enhance language proficiency and ensure originality. This versatile web application addresses these needs by offering translation, grammar correction, and plagiarism detection functionalities. Users can input text and utilize features such as translation to another language, grammar correction, and plagiarism detection, making it a multifaceted solution for diverse linguistic needs.

The importance of such a tool cannot be overstated. Grammar and language proficiency issues can significantly compromise the quality and impact of written materials, while the abundance of online information increases the risk of inadvertently incorporating others' work. This web application leverages advanced translation APIs, language processing techniques, and extensive databases to provide accurate translations, correct grammatical errors, and verify originality. By ensuring clarity, professionalism, and content integrity, it addresses critical challenges faced by individuals and organizations in content creation.

Current research in this domain focuses on refining the accuracy and reliability of language processing tools. Innovations in natural language processing (NLP), machine learning, and AI-driven algorithms are continuously being integrated into these solutions to improve their performance. This web application represents a significant stride in language processing technology, embodying the latest advancements and addressing the practical needs of users. By integrating robust translation, grammar correction, and plagiarism detection capabilities, it provides a holistic solution for content creation and verification, paving the way for a future where linguistic proficiency and originality are universally accessible.

II. METHODOLOGY

Senyeu Hao and Gang Hao discusses research on an online grammar checker system based on a neural network model called Transformer. The system uses the Transformer model along with pre-processing methods such as byte pair encoding algorithm, tokenization, and spellchecker. The pre-processing methods used in the online grammar checker system are tokenization, spellchecker, and byte pair encoding (BPE) algorithm. [1]

Xiuhua Wangl and Weixuan Zhong discusses the use of deep learning technology to propose an ASS grammar detection model for English grammar check and error correction. The results show that the model has improved

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

accuracy and efficiency in detecting grammatical errors. The accuracy rate of the ASS model is the highest among several models, reaching 99.71%. In comparison, the accuracy rate of the ordinary model is 51.74%. This indicates that the ASS model performs significantly better in terms of accuracy compared to the ordinary model. [2]

Yash Thakare and Tejas Sridhar worked on the development and application of grammar checking and correction systems in the field of natural language processing. It discusses the challenges and approaches used in developing grammar checkers for different languages, including Indian languages. [3]

Shashank Shetty and Shreyas Udaya presents a grammatical error correction (GEC) system that provides suggestions to users for correcting incorrect sentences. The system uses sequence tagging, where words or phrases are classified using a predefined label set. The model is pretrained on synthetically generated grammatical errors and trained on various datasets. The system improves the inference power of the model by repeatedly feeding the output for further improvement. [4]

Iwara Arikpo and Iniobing Dickson discusses the development of an automated English-to-local-language translator using Natural Language Processing. The system uses the transfer-based approach and Java technology for implementation. It analyses English texts morphologically, transforms their grammatical structure to fit the local language, and replaces source text with local language synonyms. [5]

Taresh Bokade and Tejas Chede discussed that plagiarism is a big problem in academics, and this project aims to develop a system for plagiarism detection in which student assignments are compared with each other using data mining algorithms and natural language processing. The system will check for plagiarism by comparing assignments syntactically and semantically, and it will generate a plagiarism detection report. The proposed system also includes features such as adding missing citations and rewriting text. The document discusses various algorithms used in plagiarism detection, including the Rabin Karp algorithm, the KMP algorithm, and WordNet expansion. [6]

Zhan Su and Byung-Ryul Ahn discusses a hybrid plagiarism detection method that combines the use of the Levenshtein distance and the Smith-Waterman algorithm. The approach aims to detect plagiarism in texts by considering the insertion, deletion, and substitution of words. The results show that the approach improves efficiency and offers practicality for plagiarism detection. [7]

Shalini Sharma and Shekhar Sharma discusses the development of a plagiarism detection tool called Parikshak, which is designed to detect plagiarism in source codes of students learning programming languages. The tool supports six programming languages and uses tokenization and the Greedy String Tiling algorithm for comparison. The tool has been well-received by teachers and has been effective in detecting plagiarism.[8]

Brinardi Leonardi and Seng Hansun discusses the use of Rabin-Karp and Jaro-Winkler Distance algorithms for detecting plagiarism in text documents. The Rabin-Karp algorithm is effective for detecting multiple string patterns, while the Jaro-Winkler Distance algorithm is faster in terms of processing time. The algorithms were tested on different types of documents and were found to be effective. [9]

III. PROPOSED ALGORITHM

1. Grammar Correction:

- Implement algorithms for grammar correction using natural language processing (NLP) techniques and libraries.
- Develop a web-based tool for grammar correction and rewriting using HTML, CSS, and JavaScript within a web application framework.
- Integrate a Generative AI (GenAI) API for grammar correction and rewriting, likely utilizing machine learning algorithms to analyze and correct grammatical errors within the user-provided tex

2. Plagiarism Detection:

- Integrate third-party APIs or develop custom algorithms for plagiarism detection, considering factors such as text similarity metrics and source comparison methods.
- Use Natural Language Processing (NLP) to analyze the text and break it down into its component parts for plagiarism detection.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

- Employ tools such as Selenium, BeautifulSoup, NLTK, and scikit-learn for browsing the internet, reading web pages, understanding text, and comparing content for potential plagiarism.

3. Translation:

- Utilize a Translation API for translating text from one language to another.
- Leverage machine translation models trained on multilingual corpora for accurate and contextually nuanced translations.
- Provide an interface for users to input text and receive the translated content.

4. Output Text:

Provide the final corrected, rewritten, or translated text to the user based on the chosen functionality.

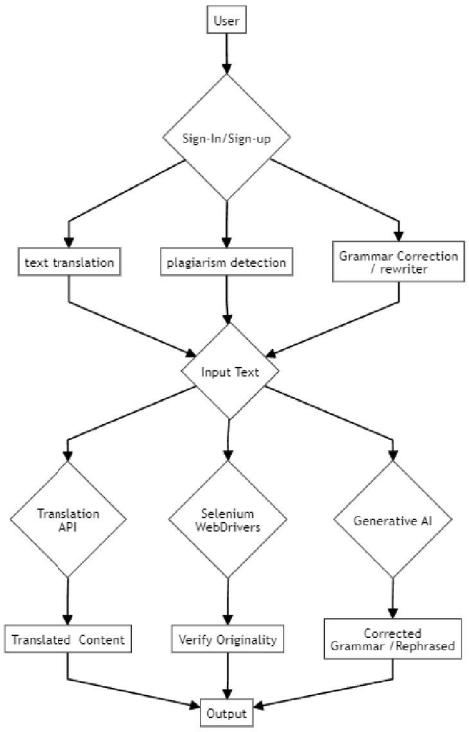


Fig. Data Flow Diagram

IV. SYSTEM ARCHITECTURE

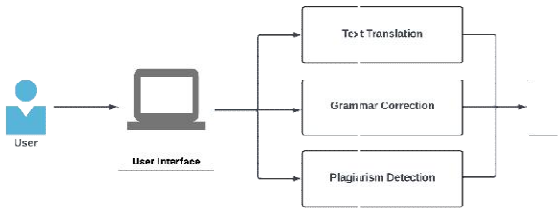


Fig. System Architecture

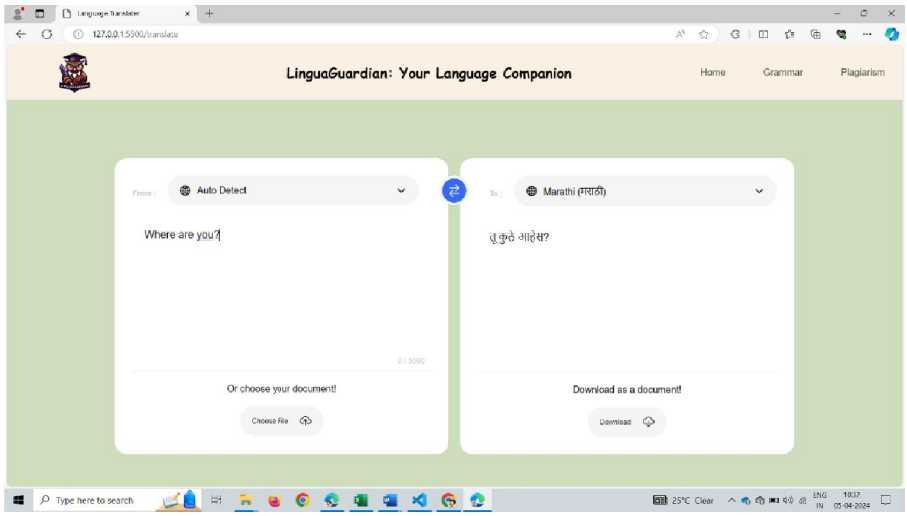
This is a system architecture diagram for a web application that translates text, corrects grammar, and detects plagiarism. Here's a breakdown of the components and their interactions:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

- User Interface (UI): This is the part of the system that users interact with to provide text input and receive the translated, corrected, and plagiarism-checked text.
- API Gateway: This component acts as a single-entry point for all API requests coming from the web application layer. It routes the requests to the appropriate service, in this case, the Google Translate API.
- Google Translate API: This external service translates the text entered by the user into the target language.
- NLTK: This Python library is used for grammar correction. It analyses the user-entered text for grammatical errors and suggests corrections.
- Selenium Web drivers: These web automation tools are used for the plagiarism detection functionality. They simulate a web browser and allow the system to access web pages potentially containing plagiarized content. The system then compares the user-entered text with the content on those webpages to identify plagiarism.
- Database: This component is responsible for storing and retrieving data used throughout the system. This may include user input, translation results, grammar corrections, plagiarism detection reports, and potentially other data.
- Web Application Layer: This layer serves as the intermediary between the UI components and the other functional components. It receives user input, interacts with the API gateway, NLTK, Selenium web drivers, and the database to process the input, and returns the results to the UI for display.

V. RESULTS

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024



VI. FUTURE SCOPE

Advancements in language models are set to revolutionize text interaction, particularly improving translation accuracy and context-awareness. These models will capture nuances better and provide enhanced grammar and style suggestions, improving communication quality. Integrating machine learning and AI will make our program more intuitive and adaptable, offering personalized recommendations based on user interactions. In plagiarism detection, we aim to expand source coverage and provide detailed reports, promoting academic integrity. We prioritize accessibility, ensuring compatibility with assistive technologies and adherence to standards, empowering all users.We're also exploring integration with various platforms and text editors to streamline the writing experience. Data privacy and security are paramount, and we commit to robust protection measures, aligning with regulations and best practices to ensure user trust and confidence.

VII. CONCLUSION

In today's globalized world, effective communication transcends boundaries. LinguaGuardian is your reliable ally in written expression, whether you're a student, researcher, professional, or language enthusiast. LinguaGuardian offers a rich array of functionalities, including accurate text translation, meticulous grammar checking, and robust plagiarism detection, all designed to enhance your writing quality. Our commitment to accuracy, reliability, and user-friendliness ensures high-quality results for every task. Distinguishing itself with an integrated solution, LinguaGuardian eliminates the need for multiple applications, offering all linguistic tools in one intuitive platform. We continuously innovate, updating our services to stay ahead of linguistic and technological advancements. Our ongoing investment in advanced language models, customization options, and expanded language support harnesses the latest in natural language processing and AI. LinguaGuardian is more than a tool; it's a catalyst for connection, understanding, and empowerment, enriching lives and fostering global communication.

REFERENCES

[1]. Senyue Hao, Gang Hao," A Research on Online Grammar Checker System Based on Neural Network Model", Journal of Physics: Conference Series

[2]. Shashank, Shetty Shreyas Udaya, Sumukha N Shilge, Mohammed Yasir, Mrs. Sreevidya B S, "Free Writing Grammatical Error Correction System: Sequence Tagging", International Research Journal of Engineering and Technology (IRJET), Sep 2022

[3]. Xiuhua Wang1 and Weixuan Zhong, "Research and Implementation of English +Grammar Check and Error Correction Based on Deep Learning", Hindawi, January 2022

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

[4]. Taresh Bokade, Tejas Chede, Dhanashri Kuwar, Prof. Rasika Shintre, "Online Assignment Plagiarism Checking Using Data Mining and NLP", International Research Journal of Engineering and Technology (IRJET), January 2021

[5]. Yash Thakare, Tejas Sridhar, Navanit Srisangkar, Pankaj Vanwari, "Application for Grammar Checking and Correction", International Research Journal of Engineering and Technology (IRJET), June 2020

[6]. Shalini Sharma, Shekhar Sharma, Veena Tyagi,"Plagiarism Detection Tool "Parikshak", International Conference on Communication, Information &Computing Technology (ICCICT), Jan. 16-17

[7]. Brinardi Leonardo, Seng Hansun, "Text Documents Plagiarism Detection using Rabin-Karp and Jaro-Winkler Distance Algorithms", Indonesian Journal of Electrical Engineering and Computer Science, February 2017

[8]. Iwara ARIKPO, Iniobong DICKSON, "Development of an automated English-to-locallanguage translator using Natural Language Processing", International Journal of Scientific & Engineering Research, July-2018

[9]. Zhan Su, Byung-Ryul Ahn, Ki-yol Eom, Min-koo Kang, Jin-Pyung Kim, Moon-Kyun Kim, "Plagiarism Detection Using the Levenshtein Distance and Smith-Waterman Algorithm", IEEE, 2008