IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 1, January 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201045 |

LinguaGuardian: Your Language Companion

Sayali Jadhav, Omkar Karpe, Anshul Pathrabe, Amruta Kate, Prof. Rashmi Kale

Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India

Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India

Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India

Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India

Professor, Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, India

ABSTRACT: This study presents a multifunctional language enhancement tool designed to address the growing need for accurate and efficient text translation, plagiarism detection, and grammar correction. Leveraging advanced natural language processing techniques, and a robust algorithm that analyzes and refines text. Our analysis demonstrates its exceptional performance in enhancing written content quality by providing accurate translations, identifying and preventing plagiarism, and rectifying grammatical errors. The integration of these functionalities into a single platform makes it a valuable resource for writers and researchers, ultimately improving the quality and authenticity of their work.

KEYWORDS: Text Translation, Plagiarism Detection, Grammar Correction.

I. INTRODUCTION

In today's digitally interconnected world, effective written communication is at the heart of countless endeavours. Whether you are a student completing an assignment, a professional composing an email, or a researcher publishing groundbreaking work, the quality of your writing profoundly impacts your ability to convey ideas, establish credibility, and foster understanding. The challenges of crafting well-written content are manifold, ranging from language barriers to unintentional plagiarism and grammatical missteps. To address these challenges and empower individuals to communicate more effectively, the development of multifunctional Language Enhancement Tools has become increasingly significant.

In this era of globalization and multiculturalism, the need for text translation tools is more evident than ever. Businesses expanding into new markets, international travellers, and language enthusiasts all rely on the ability to effortlessly convert text from one language to another. These tools bridge linguistic gaps, fostering cross-cultural communication and supporting the exchange of ideas across borders. They have transcended mere convenience to become vital assets for international collaboration and understanding.

Concurrently, the academic and professional realms are acutely aware of the importance of originality and the perils of unintentional plagiarism. Academic institutions and publishing houses have stringent standards for content authenticity, and writers need tools to ensure their work is unique. Plagiarism checkers have emerged as essential tools in this context, helping individuals maintain the integrity of their writing while avoiding the legal and ethical consequences associated with unauthorized content duplication.

Furthermore, the nuances of grammar and writing mechanics play a pivotal role in the clarity and persuasiveness of communication. Effective writing, free from grammatical errors, ensures that ideas are presented with precision and impact. Consequently, grammar correctors have evolved to serve as indispensable aids in improving writing quality, helping writers achieve fluency, coherence, and conciseness in their prose.

To meet the burgeoning demand for comprehensive language enhancement, we have been dedicated to creating multifunctional tools that seamlessly integrate text translation, plagiarism checking, and grammar correction. This tool have undergone significant advancements, employing state-of-the-art algorithms and machine learning techniques to provide more accurate, user-friendly, and efficient solutions.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201045 |

II. LITERATURE REVIEW

[1] This document 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.
[2] The research article 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 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.
[3] The main focus of this paper is 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.
[4] This document 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.
[5] The document 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.
[6] The document 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.
[7] The paper 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.
[8] The document 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.
[9] The document 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.

III. METHODOLOGY

The methodology used in the systems developed for Grammar Correction, Language (Text) Translation, Plagiarism Detection focuses on advancements in the field of computer vision, particularly in scene text detection and recognition, under the influence of deep learning. Here are the key components of the methodology:

Grammar Correction:
Utilizes a Neural Machine Translation (NMT)-based approach for grammatical error correction. Implements a Sequence Tagging Model using pre-trained BERT-like transformers for error identification. Involves preprocessing, token transformation, and training with specific hyperparameters. This System has achieved F0.5 scores of 65.3/66.5 on CoNLL-2014 (test) and 72.4/73.6 on BEA-2019 (test). It is Described as faster, simpler, and more efficient than traditional transformer-based seq2seq systems.
Another System proposed in which Core of the system is the Transformer model, addressing limitations of simple encoder-decoder models. Utilizes Byte Pair Encoding (BPE) Algorithm for data compression and open vocabulary issues. Implements SpaCy Tokenization for effective text preparation. Incorporates a spellchecker using a gated convolutional neural network language model. For this model, experiments and evaluation were as follows, Tested on

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201045 |

CoNLL-2014 shared task test set, evaluating precision, recall, and F0.5 scores for different beam search sizes. Comprehensive approach combining neural network models, pre-processing, and spell-checking techniques.

Text Translation:

A System was developed which focuses on augmented reality (AR) applications in text translation. Conducts literature review, component analysis, and a comparative study of existing AR translation applications. Aims to develop an Android application for real-time Arabic text translation. The components used for this system were Optical Character Recognition (OCR) for text extraction, techniques and algorithms for text detection, translation APIs, and programming languages. It was aimed at developing an Android mobile application for real-time Arabic text translation.

Another System was developed with a goal to create a machine translator for converting English to Efik language in Nigeria. It Utilizes object-oriented principles, Java technology, and a transfer-based approach. The key components were Morphological, syntactic, and semantic analysis, grammar transformation, bilingual dictionary, and corpus use. System features include morpho-syntactic-semantic analysis, grammar transformation, and word replacement using a bilingual dictionary.

Plagiarism Detection:

It Involves data collection, pre-processing, classification, text analysis, processing and analyzing trigrams, similarity measures, clustering, semantic plagiarism detection, and cross-lingual plagiarism detection. Implements specific algorithms like Rabin-Karp, KMP, and SCAM for plagiarism detection. The Tools and Techniques used were Incorporates data mining and NLP techniques for effective plagiarism detection. Emphasizes automation for efficiency and accuracy.

Another System Addresses source code plagiarism detection using the 'Parikshak' tool. Employs tokenization, N-gram representation, and the Greedy String Tiling algorithm for comparing source codes. The System Focuses on developing a plagiarism detection system using string matching algorithms (Rabin-Karp and Jaro-Winkler Distance) and plagiarism analysis using Jaccard's Similarity coefficient. Comparative analysis with other plagiarism detection tools.

IV. MODELING AND ANALYSIS

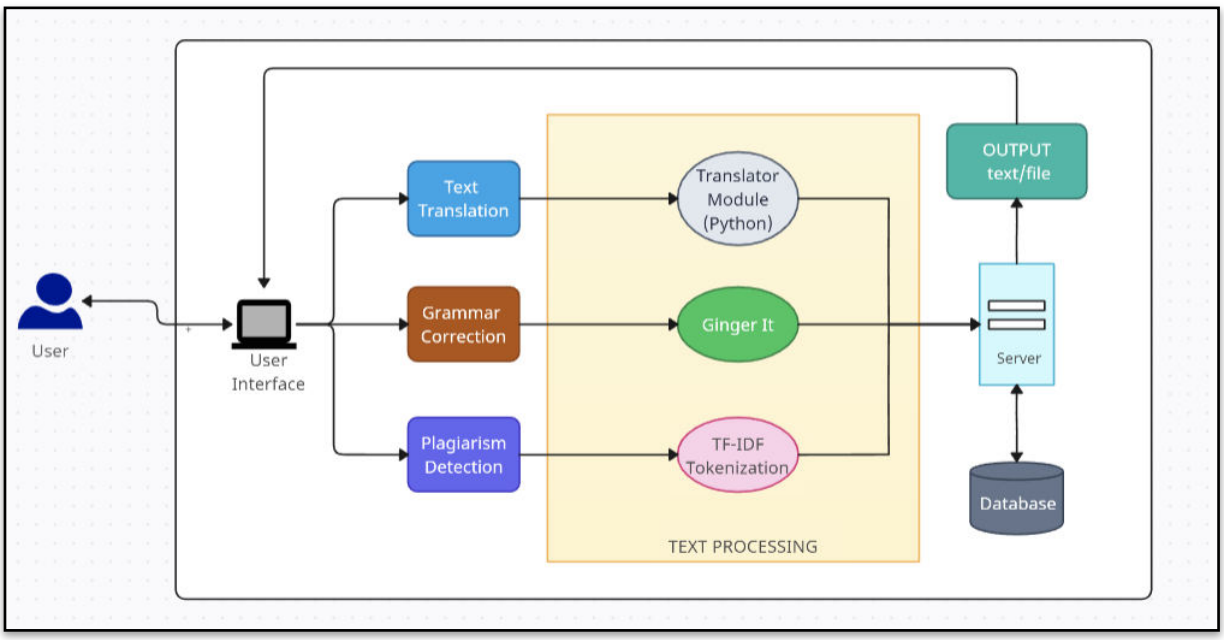

Fig. 1. System Architecture

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201045 |

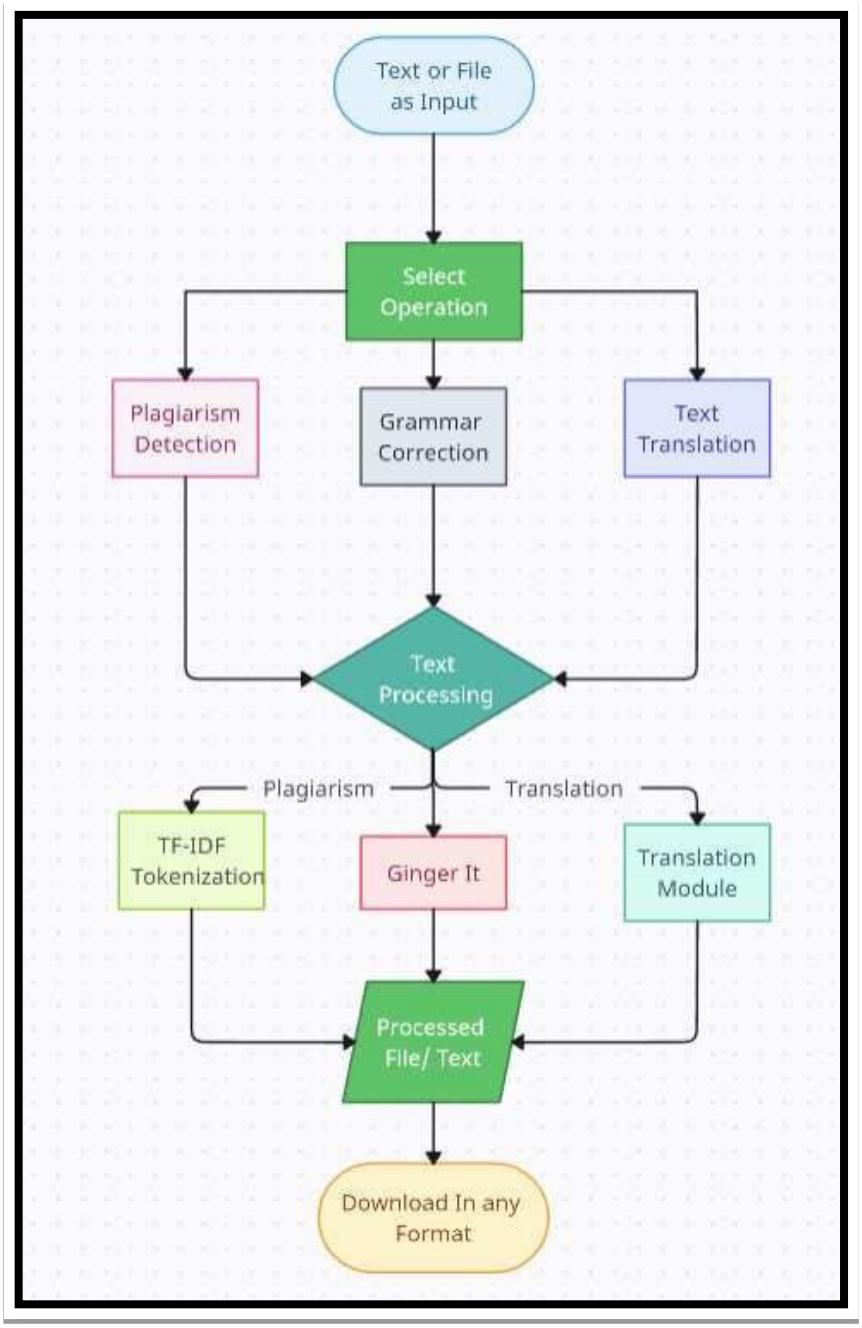

Fig. 2. Data Flow Diagram

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201045 |

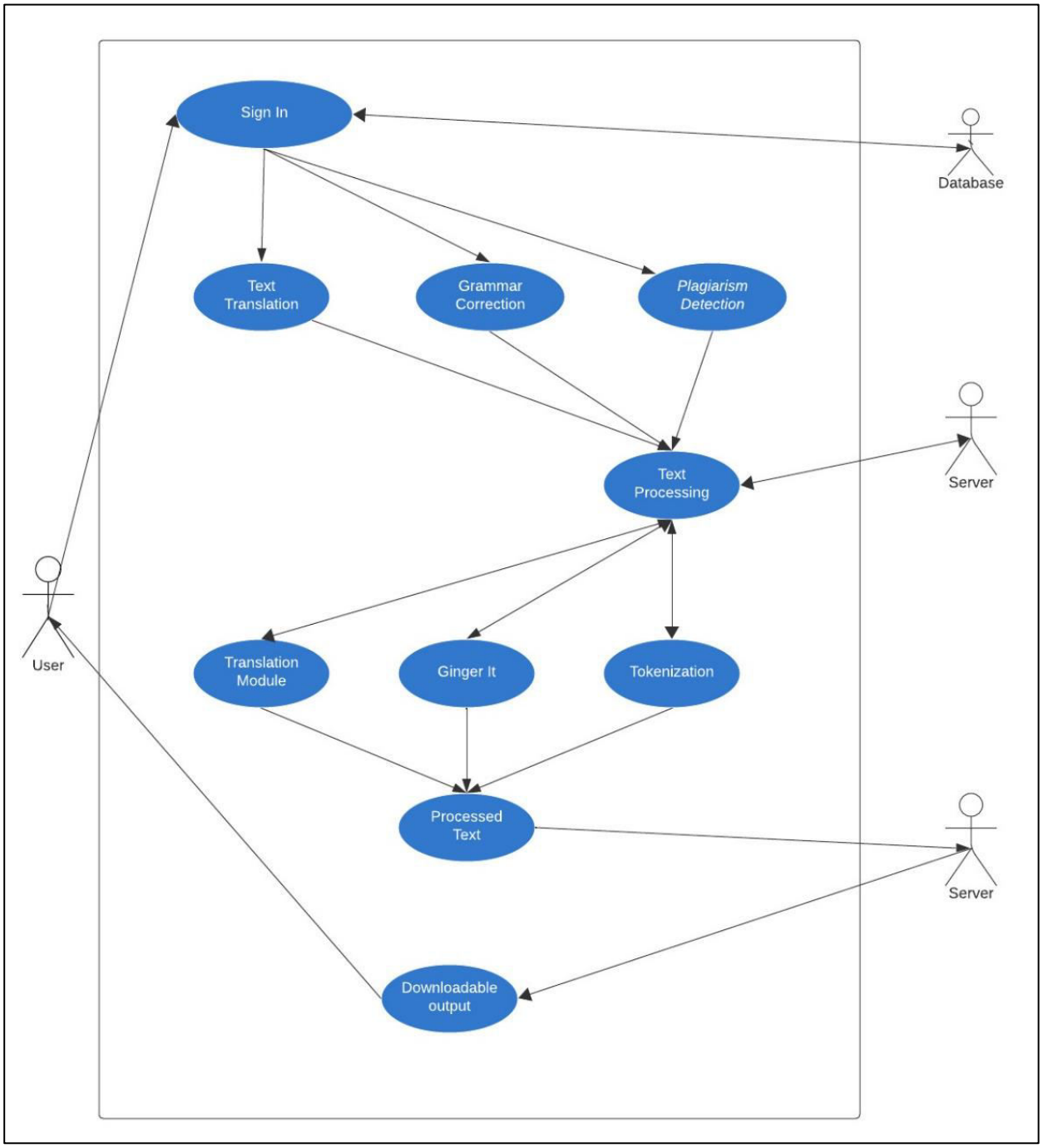

Fig. 3. Use Case Diagram

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201045 |

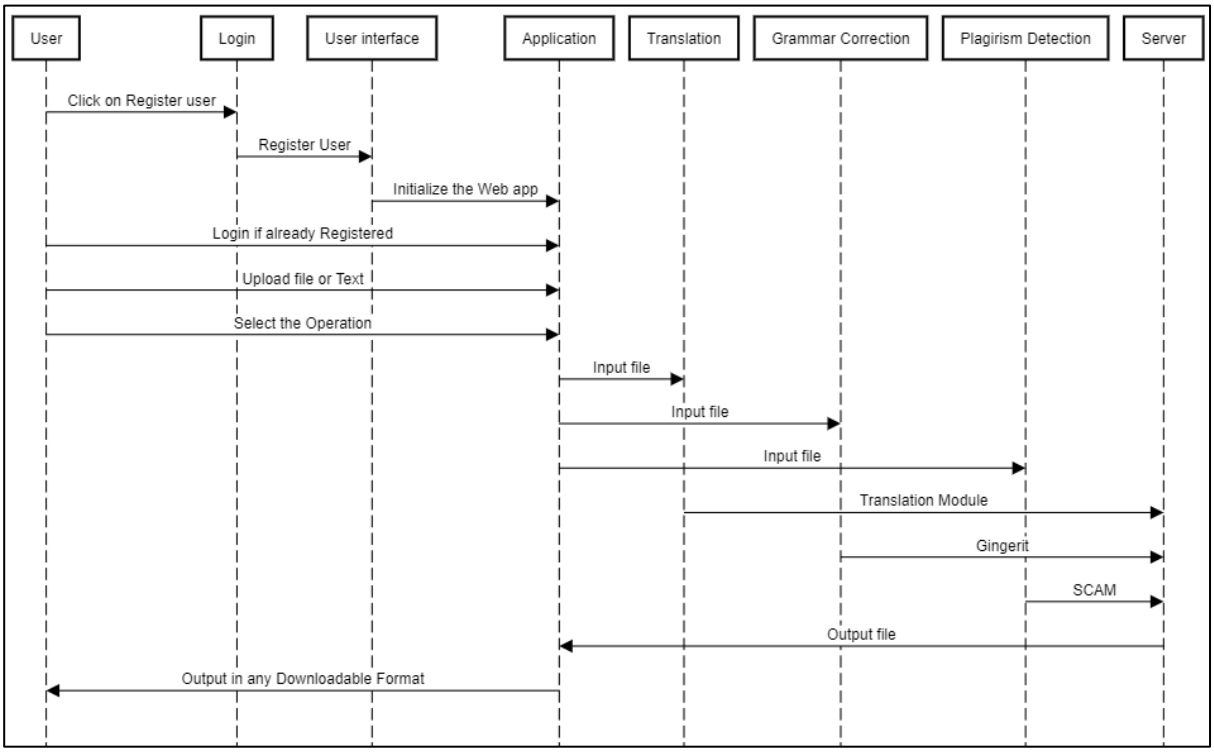

Fig. 4. Sequence Diagram

V. CONCLUSION

In conclusion, the methodologies employed in the development of systems for Grammar Correction, Language (Text) Translation, and Plagiarism Detection demonstrate a strong reliance on advancements in computer vision, particularly in scene text detection and recognition, within the realm of deep learning. Each system addresses specific challenges and objectives within its domain.

For Grammar Correction, the use of a Neural Machine Translation (NMT)-based approach, Sequence Tagging Model with pre-trained transformers, and innovative techniques like SpaCy Tokenization results in a faster, simpler, and more efficient system compared to traditional transformer-based seq2seq models.

In Text Translation, the applications leverage components like Optical Character Recognition (OCR), translation APIs, and programming languages to achieve real-time translation capabilities. The emphasis on addressing unique linguistic challenges, demonstrates a comprehensive and adaptive approach.

For Plagiarism Detection, the systems involve a multi-step process encompassing data collection, pre-processing, classification, and the utilization of specific algorithms like Rabin-Karp, KMP, and SCAM. Additionally, there is an emphasis on the application of NLP techniques and automation to enhance efficiency and accuracy. Specific attention is given to source code plagiarism detection, employing tools like 'Parikshak' and algorithms such as Jaccard's Similarity coefficient.

VI. ACKNOWLEDGEMENTS

The success and final outcome of this research required a lot of guidance and assistance and we are extremely privileged to have got this all. All that we have done is only due to such supervision and assistance and we would not forget to thank them.

We respect and thank Prof. Rashmi Kale, for providing us insight and expertise that greatly assisted the research. We are extremely thankful to her for providing such a nice support and guidance.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1201045 |

REFERENCES

1. Senyue Hao, Gang Hao," A Research on Online Grammar Checker System Based on Neural Network Model", Journal of Physics: Conference Series
2. Xiuhua Wang1 and Weixuan Zhong, "Research and Implementation of English Grammar Check and Error Correction Based on Deep Learning", Hindawi , January 2022
3. Yash Thakare, Tejas Sridhar, Navanit Srisangkar, Pankaj Vanwari, "Application for Grammar Checking and Correction", International Research Journal of Engineering and Technology (IRJET), June 2020
4. 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
5. Iwara ARIKPO, Iniobong DICKSON, "Development of an automated English-to-locallanguage translator using Natural Language Processing", International Journal of Scientific & Engineering Research, July-2018
6. 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), Jan 2021
7. 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
8. Shalini Sharma, Shekhar Sharma, Veena Tyagi, "Plagiarism Detection Tool "Parikshak" ", International Conference on Communication, Information &Computing Technology (ICCICT), Jan. 16-17,
9. 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

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details