



IJIRCCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Phrase Composing Tool using Natural Language Processing

Mayur Ahirrao, Atharva Gandhe, Sumeet Kotgire, Yash Joshi

UG Students, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa, Pune, India

ABSTRACT: In this fast-running world, machine communication plays a vital role. To compete with this world, human-machine interaction is a necessary thing. To enhance this, Natural Language Processing technique is used widely. Using this technique, we can reduce the interaction gap between the machine and human. Till now, many such applications are developed which are using this technique. This tool deals with the various methods which are used for development of grammar error correction. These methods include rule-based method, classifier-based method and machine translation-based method. Also, models regarding the Natural Language Processing (NLP) pipeline are trained and implemented in this project accordingly. Additionally, the tool can also perform speech to text operation.

KEYWORDS: Speech-to-Text, Recurrent Neural Network, Computational Linguistics, Natural Language Processing, Grammatical Error Correction, Transformer Model.

I. INTRODUCTION

In this professional world, email communication as well as legal documentation plays a vital role. In this world of email communication and legal documentation, language is very important to communicate in the corporate world as well as in day-to-day life. In the global region English is considered as the universally accepted language and most of the countries have different mother tongue, so they lack in proposing the exact English grammar. Also, due to lack of communication, they are unable to crack the global clients and at many times the senders are unable to convey the exact message to the receiver. To get rid of this problem, this paper shows the solution to this issue with the help of NLP which helps the computer to understand human language in a better manner. Our main motive is to solve grammatical errors with Grammatical Error Detection/Correction. GEC is a method which helps to correct different types of errors in sentence such as punctuation, spelling and word choice. GEC has a kind of formulation such that error-ful sentences act as source language, and error free sentences correspond to the target language. To follow this formulation GEC follows these approaches 1) Rule based approach, 2) Classifier-based approach, 3) Machine translation approach, 4) Statistical-based Approach, 5) Neural Machine Translation Approach, 6) Sequence Labelling Approach. The extensive approaches used to address GEC are normally based on statistical analysis, neural networks or mostly based on rules. These approaches may or may not be exclusive. The first proposal of transformer network to decode machine learning obstacles is proposed lately.

II. LITERATURE REVIEW

This paper is written by G. Krishna Chaitanya, Supervisor: Prof. Pushpak Bhattacharyya -Department of Computer Science and Engineering Indian Institute of Technology Bombay Mumbai 400076 (India), "GRAMMATICAL ERROR CORRECTION". This paper was published in 2017. It gives an overview of translation task which helps in the conversion of ungrammatical to grammatical correct text. [6]

This paper is written by Kostiantyn Omelianchuk, Vitaliy Atrasevych, Artem Chernodub, Oleksandr Skurzhanyski "GECToR – Grammatical Error Correction: Tag, Not Rewrite". This paper was published in 2020. It gives an overview of a system architecture to act as base of the project. [11]

This paper is written by Lossom Manchanda Assistant Professor Department of CSE GNDEC Ludhiana, Vijay Anant Athavale Director Gulzar Group of Institutions Khanna, Sanjeev kumar Sharma Assistant professor DAV University Jalandhar, "Various Techniques Used for Grammar Checking" This paper was published in 2016. It gives an overview of Statistics and Syntax based technique which helps in Grammar Checking. [9]

This paper is written by Gayathri Parthasarathy, Aspen Olmsted, Paul Anderson Department of Computer Science College of Charleston Charleston, SC parthasarathyg@g.cofc.edu, olmsteda@cofc.edu, andersonpe2@cofc.edu "Natural Language Processing Pipeline for Temporal Information Extraction and Classification

from Free Text Eligibility Criteria" This paper was published in 2016, it gives an overview of NLP pipeline which deals with the cleaning of Corpora. [1]

This paper is written by Shrimai Prabhumoye, Ruslan Salakhutdinov, Alan W Black Students School of Computer Science Carnegie Mellon University Pittsburgh, PA, USA "Topological Sort for Sentence Ordering" This paper was published in 2020, It gives an overview of the Topological sorting which is helpful for the ordering of the grammatically correct statements. [12]

III. PROPOSED ALGORITHM

A. Description of the Proposed Algorithm:

Aim of the proposed algorithm is to minimize the grammatical errors of the input sentence. The proposed algorithm is consisting of following steps.

Step 1: Pre-Training of Synthetic Data: 9M parallel sentence with generated grammar.

Step 2: Fine Tuning: 1. On errorful corpus 2. On errorful and error free corpora.

Step 3: Maps each token from source sentence to subsequent token for Token-level transformations: 1. To increase the coverage of GEC. 2. To check spellings, Noun Number, subject-verb agreement and verb form.

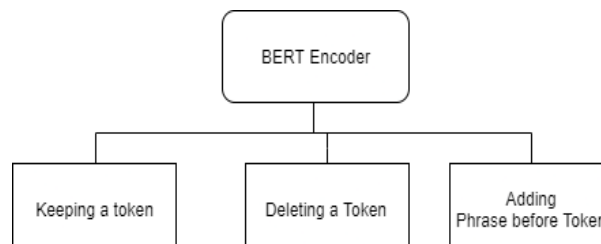


Figure 1: BERT Encoder

Step 4: G-transformations: 1. To change the case of the current token. 2. Merge the current token and next token into one. 3. Split the current one into new token.

Step 5: Sentence Reordering: 1. Set of statements S. 2. Random order of statements O. 3. To find the right order O*.

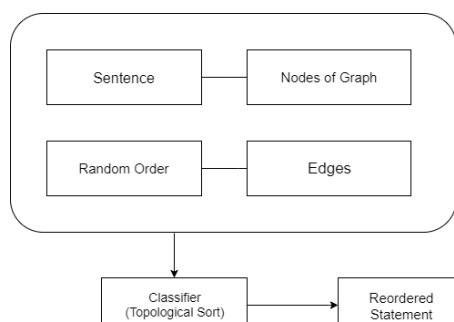


Figure 2: Sentence Reordering

Step 6: Topological Sort: 1. Linear ordering of vertices of directed graph. 2. Uses Depth-first search. 3. To visit each node n and prepends it to the output ordering.

IV. PSEUDO CODE

Step 1: Take text as an input from the user.

Step 2: Check the Grammatical accuracy of the text.

Step 3: Check the above condition for each iteration.

if (Accuracy > 90)

 Continue to Step 5.

else

Pass on the sentence to GECToR Model for Grammatical Correction.

end

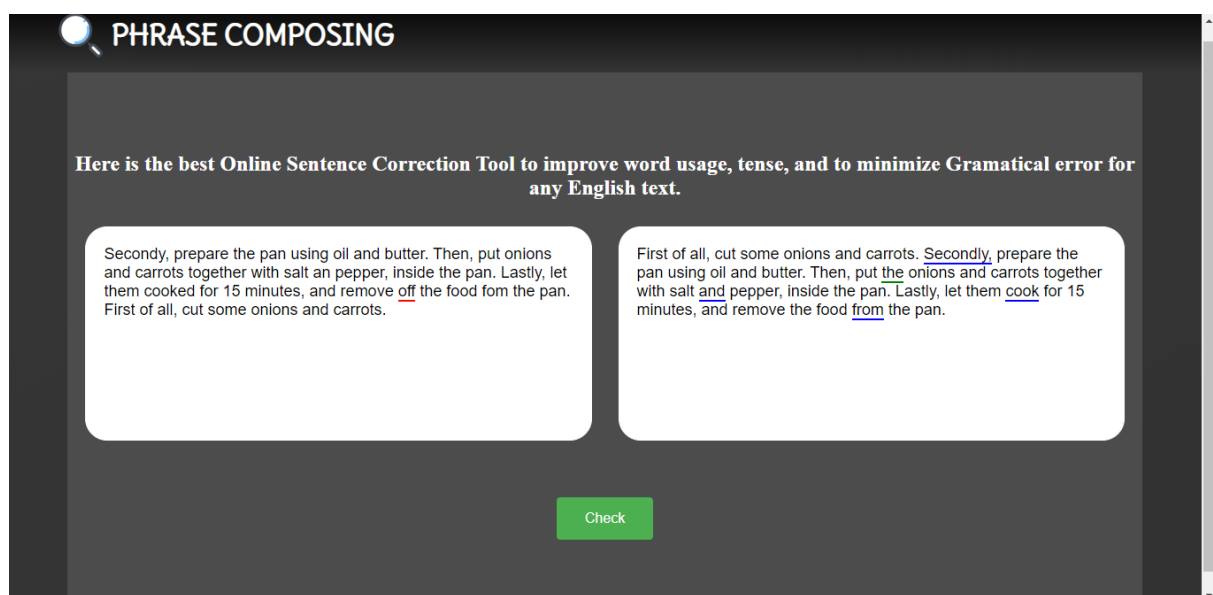
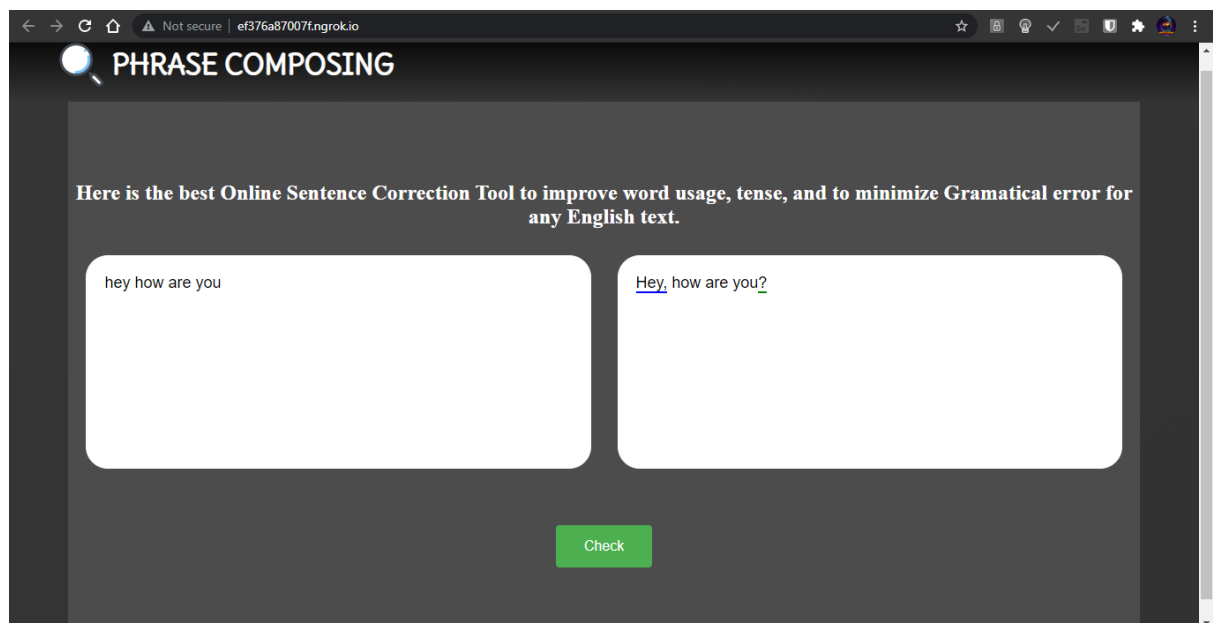
Step 4: Go to step 3 for next iteration.

Step 5: Reorder the sentence using Topological Sort

Step 6: End.

V. SIMULATION RESULTS

The simulation studies show the results as if the tool is provided a paraphrase which has accuracy more than the sufficient barrier, then it will output the same sentence as given as input. In accordance, if a grammatically correct sentence with its order shuffled is given to the tool, then its purpose is to check its accuracy and then further sort it into the most logically correct sentence. The studies showed that if grammatically incorrect as well as unordered sentence is given, the tool provides accurate results with sufficient accuracy.



VI. CONCLUSION AND FUTURE WORK

Natural Language Processing is a very advancing field and has huge scope in real world. With the continuous growth of corporate world, official email writing and legal documentations are in dire need. Basically, Grammar Error Correction and Checking are the most important things, as we have to write most of the professional write-ups. To make it work, multiple approaches such as syntax-based, rule-based, machine learning approaches as well as works that have approached automating Grammatical Error Correction and Detection by using Transformer Model have evolved into existence. As research is continuously in process, the need of development of the most optimal model is required.

REFERENCES

1. Gayathri Parthasarathy, Aspen Olmsted, Paul Anderson Department of Computer Science College of Charleston Charleston, SC parthasarathyg@g.cofc.edu, olmsteda@cofc.edu, andersonpe2@cofc.edu "Natural Language Processing Pipeline for Temporal Information Extraction and Classification from Free Text Eligibility Criteria".
2. H. T. Ng, S. M. Wu, Y. Wu, C. Hadiwinoto, and J. Tetreault, "The CoNLL-2013 Shared Task on Grammatical Error Correction," in Seventeenth Conference on Computational Natural Language Learning: Shared Task, 2013, pp. 1–12.
3. H. Yannakoudakis, M. Rei, Ø. E. Andersen, and Z. Yuan, "Neural Sequence-Labeling Models for Grammatical Error Correction," in 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2795–2806.
4. Chandhana Surabhi. M Velalar College of Engineering and Technology, Erode, S. India – 638 012. Surabhimanoahar.be@gmail.com "Natural Language Processing Future".
5. Nora Madi Information Technology Department King Saud University Riyadh, Saudi Arabia nmadi@ksu.edu.sa, Hend S. Al-Khalifa Information Technology Department King Saud University Riyadh, Saudi Arabia hendk@ksu.edu.sa "Grammatical Error Checking Systems: A Review of Approaches and Emerging Directions".
6. G. Krishna Chaitanya, Supervisor: Prof. Pushpak Bhattacharyya - Department of Computer Science and Engineering Indian Institute of Technology Bombay Mumbai 400076 (India), "GRAMMATICAL ERROR CORRECTION".
7. A. Schmaltz, Y. Kim, A. M. Rush, and S. M. Shieber, "Sentence Level Grammatical Error Identification as Sequence-to-Sequence Correction," 2016.
8. Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik and Supriya Agrawal - Department of Computer Science, NMIMS University, Mumbai, India. Corresponding Author: Navya Pant, "Speech to text and text to speech recognition systems".
9. Lossom Manchanda Assistant Professor Department of CSE GNDEC Ludhiana, Vijay Anant Athavale Director Gulzar Group of Institutions Khanna, Sanjeev Kumar Sharma Assistant professor DAV University Jalandhar, "Various Techniques Used for Grammar Checking".
10. C. Leacock, M. Chodorow, and M. Gamon, Automated Grammatical Error Detection for Language Learners, 2nd ed. Morgan & Claypool, 2014.
11. Kostiantyn Omelianchuk, Vitaliy Atrasevych, Artem Chernodub, Oleksandr Skurzhanyski "GECToR – Grammatical Error Correction: Tag, Not Rewrite", 2020.
12. Shrimai Prabhumoye, Ruslan Salakhutdinov, Alan W Black Students School of Computer Science Carnegie Mellon University Pittsburgh, PA, USA "Topological Sort for Sentence Ordering", 2020.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



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