

Rescript Clamping

Nikita Anand¹, Ishika Rathore², Vidhi Verma³

Students, Department of Computer Science and Information Technology^{1,2,3}
Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India

Abstract: *This project is all about paraphrasing of single or multi-lined texts, using Python's Parrot library and algorithm naming, T5 (Text-to-Text Transfer Transformer) algorithm. The algorithm treats a wide variety of many-to-many and many-to-one NLP (Natural Language Processing) tasks in a unified manner by encoding the different tasks. We are in the era of fast changing technology with a very busy schedule. Sometimes we are loaded with so much stuffs that we get confused in terms of any project, understanding any article, framing sentences, etc. So, we need a technological helper such that we can resolve these problems. Our project is one such solution of many problems regarding summarization, translation, knowledge extraction, etc. Previously only single-line text was paraphrased but now multiple - line text can also be added to it. Moreover, Algorithm which is used for this paraphrasing is T5 which has a large scope in NLP. As per considering future aspects this multi-line paraphrasing can be extended to AI based voice recognition and its summarization such that people briefly understand the things..*

Keywords: Python, Parrot Library, NLP (Natural Language Processing), T5 (Text-to-Text Transfer Transformer) Algorithm, Google Collaboratory

I. INTRODUCTION

“Understand It, Think It, Build It” ----We understand the things, think on it and build it by own. Rescript Clamping is something which explains the same thing. This project is all about converting things in own way by understanding them which can further make it easier and more understandable to read and write whomsoever accesses it. The project is going to enhance the user's way of understanding such that by reading a line or a paragraph he/she could write the same thing in his/her own way. This is going to be helpful to all such community who are willing to have a good understanding of someone's thoughts. When we read something, it might happen that we do not understand it's expression but para-phrasal text converts that into readable one without changing the meaning of text. This will help user to understand things properly and write again into their own language. Thus, our main motive is to provide different ways to express oneself in various forms with the help of this technology. Importance of this research paper shows that the project is having numerous future scope and users can use it in wide varieties like summarization, knowledge extraction, translation, etc. Some advantages to this paper is that it saves cost and time, minimal hardware as well as software requirement, access to requirements is feasible and easieras algorithm used in this project has huge future scope.

II. RELATED WORK OR LITERATURE STUDIES

A previous research paper says that paraphrasing tool is used to rewrite or rephrase a sentence without altering its meaning. This is accomplished by substituting any number of alternate versions for specific words, phrases, sentences, or even whole paragraphs to create a slightly different variant. Rephrasing a sentence follows the same process as paraphrasing, but the most important consideration is to make the sentence clearer. Rephrasing may or may not be coupled with rewording or synonym adjustments. It may only entail rearranging the original sentence as long as clarity is obtained. Another paper describes how to use the paraphrasing concept by just few clicks. So, it states that how to paraphrase text in Python using the Parrotlibrary. Particularly, under the hood Parrot's paraphrasing technology is based on the T5 algorithm (an acronym for Text-To-Text Transfer Transformer) that was originally developed by Google (for more information refer to the T5 resource at Papers with Code). At a high-level, text generation is niche area of the exciting area of natural language processing (NLP), which is generally referred to as artificial intelligence or AI when explained to the general audience.

III. MOTIVATION

Our motivation to make this research paper regarding proposed project is to aware students, teachers including all working professionals to utilize paraphrasing technology such that they can have their own taste of understanding things as well as writing it. We are having solution for those sentences also which is not understandable properly and people face issue in acknowledging them. Algorithm used in this project also helps in summarization of paragraphs, translation from one language to another, etc.

IV. PROBLEM DOMAIN

With reference to the text typed by a writer or user, this project refers to the detection of meaning under paraphrasing with the help of Python using “Parrot” library. The problem statement says that if more than one person wants to convey his/her message then there should be different ways to it.

V. PROBLEM DEFINITION

Text paraphrasing with the help of parrot library of python using T5 algorithm is machine learning based project such that single or multi-lined paraphrases can be converted into different forms which would be easier to read and understand. T5 algorithm is also used for classification and regression problems, conversion of languages from one form to another.

VI. STATEMENT

This project shows use of python and Google Collaboratory notebook to paraphrase texts and change its form without changing the meaning and it is easy to use as well as handle it. One more important thing is that it is free to use.

VII. INNOVATIVE CONTENT

Paraphrasing was done only for single lined text but we are extending it for paragraphs and multi lined texts as well using some establishing proper frontend and backend connections. NLP is the main root for paraphrasing as in T5 algorithm is one of the most famous algorithms. This project is not only good from present aspect but also from future perspectives. AI ChatGPT (Artificial Intelligence Chatbot Generative Pre-Trained Transformer 3) is an exclusive example of language learning models(LLM) that relies on Deep Learning. Moreover, these all are connected with NLP whose one form is Paraphrasing. This project is all about innovation as well as blend of ML (Machine Learning), AI (Artificial Intelligence) and DL (Deep Learning). Regression and Classification problems are also considered in solving problems related to training models used in plagiarism and paraphrasing.

VIII. PROBLEM REPRESENTATION

Rescripting is done on the basis of certain procedure. There is a diagram which shows the problem related to understanding of the text. Here you can see that quote written below is quite tough to understand but paraphrasing makes it simpler and easier to understand. Language used here is English but one who will read this have to go through once again to understand the meaning completely written here. Thus, Paraphrasing is important to decode such things accordingly.

This states the problem domain of project and necessity of tools like paraphrasing.

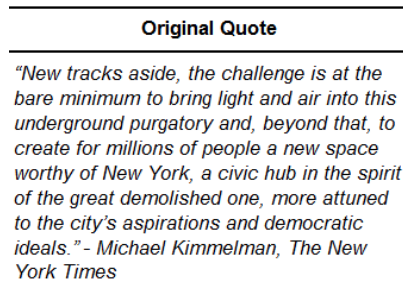


Fig. 1. Problem Domain
DOI: 10.48175/568

IX. SOLUTION METHODOLOGIES OR PROBLEM SOLVING

This is with the reference to paragraphing of texts; we have proposed the methods through which this problem can be resolved. The most important thing is that availability of requires sources for this project is minimal and free too. Google Collaboratory access which is easily available who are having a g-mail account. This project is based on T5 algorithm which has huge scope in solving different types of machine learning based problems e.g., classification, regression, etc.

The block diagram below will show clear path that explain use of this project:

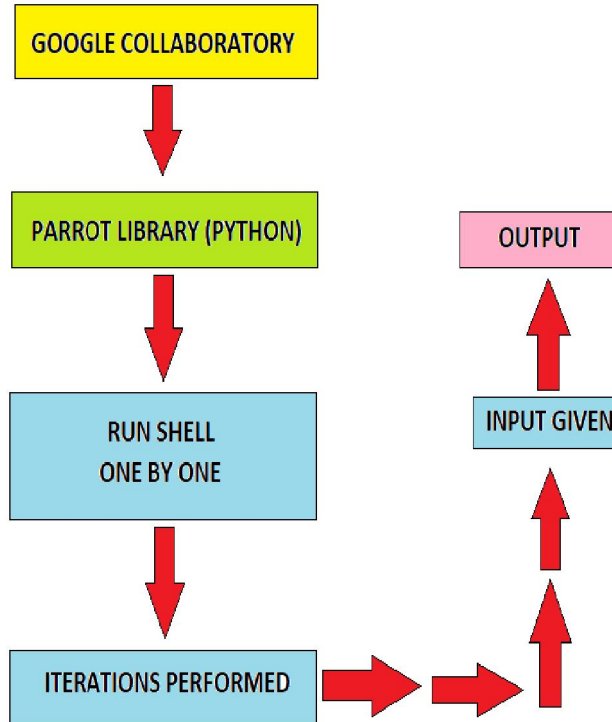


Fig. No. 2. Block Diagram

```

Collecting git+https://github.com/PrithivirajDanodaran/Parrot.git
Cloning https://github.com/PrithivirajDanodaran/Parrot.git to /tmp/pip
Running command git clone -q https://github.com/PrithivirajDanodaran/
Collecting transformers
  Downloading https://files.pythonhosted.org/packages/d5/43/fe4ee79bb
  [Progress bar] 2.3MB 27.8MB/s
Collecting sentencepiece
  Downloading https://files.pythonhosted.org/packages/f5/99/e080ct947b
  [Progress bar] 1.2MB 33.9MB/s
Collecting python-Levenshtein
  Downloading https://files.pythonhosted.org/packages/2a/dc/37f2b62ef0f
  [Progress bar] 51kB 6.9MB/s
Collecting sentence-transformers
  Downloading https://files.pythonhosted.org/packages/cc/25/d144101cd
  [Progress bar] 81kB 9.3MB/s
Collecting fuzzywuzzy
  Downloading https://files.pythonhosted.org/packages/43/ff/74f23998ad2
Collecting tokenizers<0.11,>=0.10.1
  Downloading https://files.pythonhosted.org/packages/d4/e2/d13543e8ff
  [Progress bar] 3.3MB 41.6MB/s
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dis
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dis
Requirement already satisfied: requests in /usr/local/lib/python3.7/dis
Collecting huggingface-hub==0.6.8
  Downloading https://files.pythonhosted.org/packages/a1/88/7b1e45720e
  [Progress bar] 901kB 32.0MB/s
Requirement already satisfied: numpy=>1.17 in /usr/local/lib/python3.7/
Requirement already satisfied: importlib-metadata; python version < "3.4
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/pyth
Collecting sacremoses
  Downloading https://files.pythonhosted.org/packages/75/ee/57241dc87f2
  [Progress bar] 901kB 32.0MB/s
Requirement already satisfied: tqdm=>4.27 in /usr/local/lib/python3.7/d
Requirement already satisfied: setuptools in /usr/local/lib/python.7/d
Requirement already satisfied: torch=>1.6.0 in /usr/local/lib/python3.7/
Requirement already satisfied: torchvision in /usr/local/lib/python3.7/
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.7/
29s completed at 7:49 PM
  
```

Fig No.3(a). Packages and Parrot Library Installation

Now let's see glimpses of procedures involved in the rescript clamping project. Below are the figures that shows packages and libraries of python required for this project is being installed.

X. RESULTS AND SENSITIVITY ANALYSIS

This diagram gives an overview of how the things will work. It states various use of paraphrasing that means we can convey our message in many forms. Moreover, if something is there which we don't get properly then we can use paraphrasing there also. The example below shows inputs and their outputs corresponding to it. Hence, the diagram is as stated below:

```
-----
Input_phrase: What's the most delicious papayas?
-----
("what's the best papaya to eat?", 29)
('which are the best papayas?', 27)
('what is your most delicious papaya?', 19)
('list the most delicious papayas?', 16)
```

Fig. 4.Single - Lined Paraphrasing Of Texts

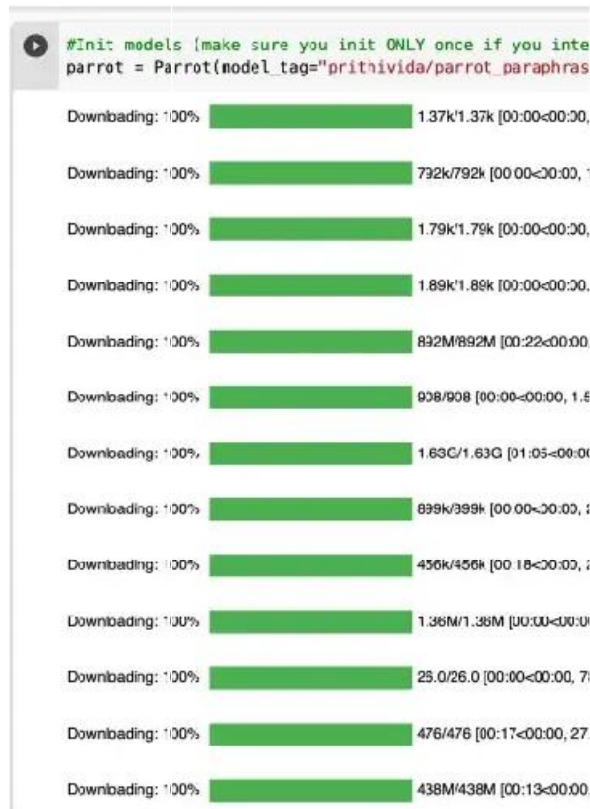


Fig. 3(b). Installation of INIT Models

XI. COMPARISON OF RESULTS

Here is the comparison of results of different research papers and their views regarding various aspects of paraphrasing.

<p><u>2. PARAPHRASE TEXT USING PYTHON</u></p> <ol style="list-style-type: none"> 1. <u>Author</u>- Chanin Nantasenamat 2. <u>Platform</u>- Python 3. <u>Algorithm</u>- NLP - (T5) 4. <u>Description</u>- This project focuses on paraphrasing of texts using Python of single -lined phrase. 	<p><u>3. PLAGIARISM CHECKER AND PARAPHRASING TOOL USING PYTHON</u></p> <ol style="list-style-type: none"> 1. <u>Author</u>- Kavya B, Nandita Sahu, Dr. JP Patra 2. <u>Platform</u>- Python 3. <u>Algorithm</u>- NLP (Natura Language Processing) 4. <u>Description</u>- Explains Plagiarism detection like textual or source code and paraphrasing tool technique.
---	---

Paraphrase	Direct Quote
In his 2012 study, Miller discusses recent innovations in interferon drugs. He writes that increasing levels of tumor hypoxia significantly reduced the uptake of Regorafenib (Miller, 2012).	As Miller (2012) notes , "uptake of 300cc injected Regorafenib was reduced by one-third for each .05% of increased hypoxia in the tumored organ."

Fig 5. Rescripting Paragraph(1)

Source text	Paraphrase
While many governments have taken steps to implement privatisation programmes, progress has been slower than was anticipated in the early 1980's.	According to Bayliss (2003), although the implementation of privatisation programmes has been a priority for many governments, progress in the area has not been as rapid as expected in the early 1980's.

Fig 6. Rescripting Paragraph(2)

<p><u>1. RESCRIPT CLAMPING</u></p> <ol style="list-style-type: none"> 1. <u>Authors</u>- Nikita Anand, Ishika Rathore, Vidhi Verma 2. <u>Platform</u>- Python (Parrot's Library) 3. <u>Algorithm</u>- T5 (Text-to-Text Transfer Transformer) 4. <u>Description</u>- It's about altering texts and its form without changing meaning of text. Applicable for single-lined and multi-lined texts.
--

4. ML MODELS FOR PARAPHRASE IDENTIFICATION AND ITS APPLICATION IN PLAGIARISM DETECTION

1. Author- Ethan Hunt, Binay Dahal, Laxmi Gevali
2. Platform- ML Models
3. Algorithm- SVM, KNN, Maximum Entropy
4. Description- Describes paraphrase identification through ML pre-trained models and its use for plagiarism checking.

5. CORPUSBASED PARAPHRASE DETECTION EXPERIMENTS AND REVIEW

1. Author-Tedo Vrbanec, Ana Mestrobic
2. Platform- Python
3. Algorithm- NLP, DL
4. Description- Overview of various types of corpus-based models, especially deep learning (DL) models, with paraphrase detection.

XII. JUSTIFICATION OF THE RESULTS

Considering the references for this project we can conclude that there are a number of ways by which solution can be found out. There are many ways to perform paraphrasing, say, for example, with the help of DL or ML models; SVM (Support Vector Machine); KNN (K-Nearest Neighbors) and so on. This gives a combined approach for multi-level data classification which ultimately saves from preventing cyber-attacks where traditional machine learning algorithms can usually be used such as decision trees and naive Bayes. In NLP, paraphrasing is generally studied from a machine learning perspective, with notable interest surrounding paraphrase identification for plagiarism detection.

XIII. CONCLUSION

The project explains use of paraphrasing and its future scope. Single lined rescripting including multi lined also. Classification and regression problems with the help of proper training models and performing iterations on the machine will contribute to AI based paraphrasing in a modified way. Other use cases are summarization of, say, an article; translation of one form of language to another and ChatGPT. Hereby, We conclude our part of this project and research paper as well.

XIV. FUTURE WORK

As we have mentioned earlier also that this project has its scope extended in future too. As per considering future aspects this multi-line paraphrasing can be extended to AI based voice recognition and its summarization such that people briefly understand the things. There are many ways to perform paraphrasing, say, for example, with the help of DL or ML models; SVM (Support Vector Machine); KNN (K-Nearest Neighbors) and so on.

XV. ACKNOWLEDGEMENT

This is an opportunity to all of us to show our ideas in the field of technology by preparing relevant project. We thank our faculty members for giving such chance to prove ourself. This project is related to conversion of texts i.e., paraphrasing with the help of Python and its packages and libraries. The things which we do not understand in its general or original form but its converted form is easy to catch, for example, any paragraph which is tough to get it but paraphrasing makes it easier and decodable. The most riveting picture of this paper is that it contains glimpse of ML/DL (Nantasenamat, 30 Sept. 2021) and advanced form of Python. Thus, this project is having concerns in future too and is important with our studies perspective also. I and my team members are obliged and grateful to my project guide Prof. Nidhi Nigam as well as Project In-charge Prof. Garima Joshi for guiding us at each and every step of this project.

REFERENCES

- [1]. Nantasenamat Chanin, How to Paraphrase text using Python, 30 Sept. 2021
- [2]. Laxmi Gewali, Binay Dahal, Ethan Hunt, Machine Learning Models for Paraphrase Identification and its Application on Plagiarism Detection, February 24, 2023
- [3]. Chih Yu Lin, How to paraphrase text in Python using the PARROT library, 24 Feb 2022
- [4]. Asad Shehzad, how to paraphrase text using python AI tools, 19 Dec 2022
- [5]. Abdou Rockikiz, Paraphrase Text Using Transformer in Python, Sept 2022

- [6]. Rajesh Kumar, AI Working Paraphrasing Using Python, 1 March 2022
- [7]. <https://www.sciencedirect.com/science/article/abs/pii/S0889490617300480> Ling Shi, Ismaeil Fazel, Nasrin Kowkabi, Paraphrasing, July 2018
- [8]. https://www.researchgate.net/publication/363255779_Paraphrase_Identification_Lightweight_Effective_Methods_Based_Features_from_Pre-trained_Models Abu Bakar Siddiquir Rahman, Thang Hoang Ta, Alexander Gelbukh; Paraphrase Identification Lightweight Effective Methods Based Features from Pre-Trained Models ; Sept. 2022
- [9]. Simona Ramanauskaitė, Pavel Stefanovič, Combined Approach for Multi-Label Text Data Classification, June 2022
- [10]. Noura Al Moubayed, Paraphrasing for multi-task question answering, Oct 2021
- [11]. Animesh Nigohjkar, Improving Paraphrase Detection with the Adversarial Paraphrasing Task, January 2021
- [12]. Mounir Zrigui, BLSTM-API_Bi-LSTM Recurrent Neural Network-Based Approach for Arabic Paraphrase Identification, Feb 2021
- [13]. Andreas Chandra, Experiments on Paraphrase Identification Using Quora Question Pairs Dataset, June 2020
- [14]. Experiments on Paraphrase Identification Using Quora Question Pairs Dataset, Ruben Stefanus, June 2020
- [15]. <https://www.enago.com/academy/how-to-paraphrase-in-a-research-paper/>, Enago Academy, Nov 26 2021
- [16]. <https://edintegrity.biomedcentral.com/articles/10.1007/s40979-016-0013-y>, Ann M. Rogerson & Grace McCarthy, 26 Jan 2017
- [17]. <https://www.jstor.org/stable/42636364>, Paraphrasing, plagiarism, and misrepresentation in scientific writing, Mark E. Eberle, 2013
- [18]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3732494, Ismail Ruslan, Abdul Haris Sunubi, Ali Halidin, Amzah, Nanning, Andi Kaharuddin, 28 Jan 2021
- [19]. <https://blog.wordvice.com/how-to-paraphrase-in-research/>, Wordvice KH, Oct 5 2022
- [20]. <https://projectstore.com.ng/tips-and-guidelines-for-paraphrasing-research-paper/>, Project store, Tips and Guidelines For Paraphrasing Research Paper, Oct 15 2022
- [21]. <https://www.mindtools.com/axggxkv/paraphrasing-and-summarizing>, MindTools, Feb 25 2023
- [22]. <https://www.iedunote.com/best-paraphrasing-tools>, iEduNote, July 02, 2022
- [23]. <https://owl.excelsior.edu/research/drafting-and-integrating/drafting-and-integrating-paraphrasing-activity/>, Paraphrasing Activity, May 13 2017
- [24]. https://www.unitedinnovator.com/research-article-writing-services.php?gclid=CjwKCAiA3KefBhByEiwAi2LDHED90KXZg_KI_msf43BEh15sj_yAz0Pli8ex7aYG1dKG7nrzyybSRoCQtwQAvD_BwE, Research Article Writing Assistance & Publication Support, Feb 25 2023