

# **Plagiarism Removal**

## Abstract

Maintaining data uniqueness is one of the important features for many areas like in colleges and universities. In order to maintain plagiarism free content, there is a need for effective methods, wherein the existing method used, should rewrite entire content manually if there are many pages of content to be written, then it takes a lot of time and what if you don't like it, all the efforts would be futile.

But, if the same task can be done by our machines using NLP (Natural Language Processing), then why not? Plagiarism Removal is made by using a simple concept of synonyms set that comes with wordnet. Lemmas can be used and then their names will be synonyms. Randomly, any synonym can be assigned without changing the context. Everytime, program is run, the user can get a different flavour of plagiarism removed text each time, making the user have different options and then the user can simply go with the one which he/she finds most appropriate as per his/her need. But after certain iterations, user may start getting same content.

## Existing System

In the existing system, in order to remove plagiarism for the content, a manual process was involved in which the user should understand each meaning of the sentence and rewrite the entire content with its own words, which is a time taking process.

## Proposed System

Plagiarism Removal is made by using a simple concept of synonyms set (synset) that comes with wordnet. Lemmas can be used and then their names will be synonyms. Randomly, any synonym can be assigned without changing the context. Everytime, program is run, the user can get a different flavour of plagiarism removed text each time, making the user have different options and then the user can simply go with the one which he/she finds most appropriate as per his/her need.

<b>Language</b>	Python 3.x
<b>Libraries</b>	For Logic <ul style="list-style-type: none"><li>● nltk</li><li>● Random</li></ul> For GUI Design <ul style="list-style-type: none"><li>● tkinter</li><li>● functools</li></ul>

## Code

```
import nltk
from nltk.corpus import wordnet
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
import random

stop_words = stopwords.words("english")

def plagiarism_remover(i):
    word = i
    synonyms = []
    if word in stop_words:
        return word
    if wordnet.synsets(word)==[]:
        return word
    for syn in wordnet.synsets(word):
        for lemma in syn.lemmas():
            synonyms.append(lemma.name())
    pos_tag_word = nltk.pos_tag([word])
    pos = []
    for i in synonyms:
        pos.append(nltk.pos_tag([i]))
    final_synonyms = []
    for i in pos:
        if pos_tag_word[0][1] == i[0][1]:
            final_synonyms.append(i[0][0])
    final_synonyms = list(set(final_synonyms))
    if final_synonyms == []:
        return word
    if word.istitle():
        return random.choice(final_synonyms).title()
    else:
        return random.choice(final_synonyms)

def plagiarism_removal(para):
    para_split = word_tokenize(para)
    final_text = []
    for i in para_split:
```

```

        final_text.append(plagiarism_remover(i))
    final_text = " ".join(final_text)
    return final_text

import tkinter as tk
from functools import partial
def call_result(label_result, n):
    text = n.get()
    result = plagiarism_removal(text)
    label_result.config(text="Text after plagiarism removal is:\n %s" % result, wraplength=500)
    return
root = tk.Tk()
root.geometry('1000x1000')
root.title('Plagiarism Removal')
number1 = tk.StringVar()
labelNum1 = tk.Label(root, text="Enter text to remove plagiarism")
labelNum1.grid(row=1, column=0)
labelResult = tk.Label(root)
labelResult.grid(row=7, column=2)
entryNum1 = tk.Entry(root, textvariable=number1)
entryNum1.grid(row=1, column=2)
call_result = partial(call_result, labelResult, number1)
buttonCal = tk.Button(root, text="Remove Plagiarism", command=call_result)
buttonCal.grid(row=3, column=0)
root.mainloop()

```

## Sample Text 1

Nature is an important and integral part of mankind. It is one of the greatest blessings for human life; however, nowadays humans fail to recognize it as one. Nature has been an inspiration for numerous poets, writers, artists and more of yesteryears. This remarkable creation inspired them to write poems and stories in the glory of it. They truly valued nature which reflects in their works even today. Essentially, nature is everything we are surrounded by like the water we drink, the air we breathe, the sun we soak in, the birds we hear chirping, the moon we gaze at and more. Above all, it is rich and vibrant and consists of both living and non-living things. Therefore, people of the modern age should also learn something from people of yesteryear and start valuing nature before it gets too late.

## Text after Plagiarism Removal

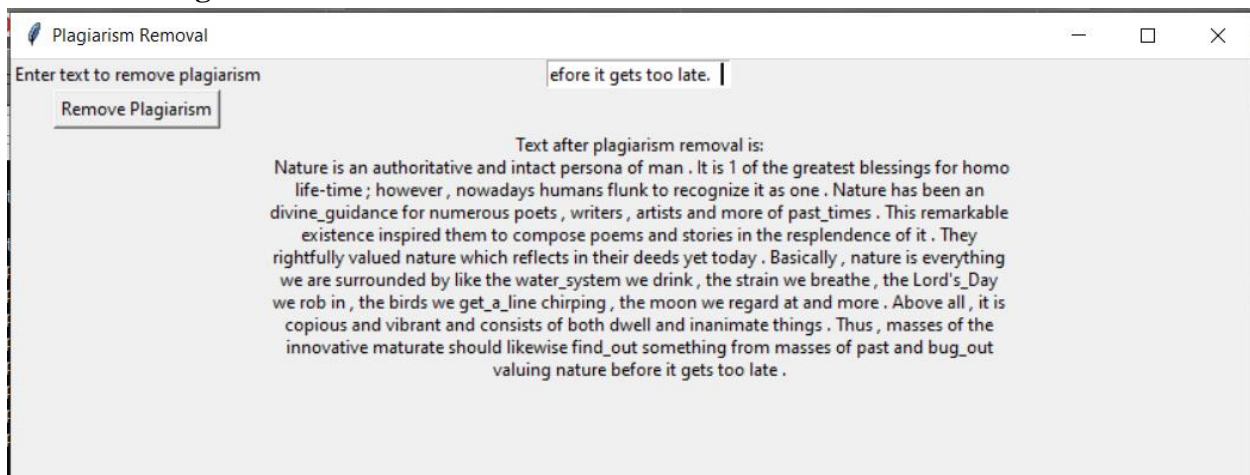


Figure 1: Output for Plagiarism Removal for Sample Text 1

## After rerun (User can get different flavour of Plagiarism Removed Text)

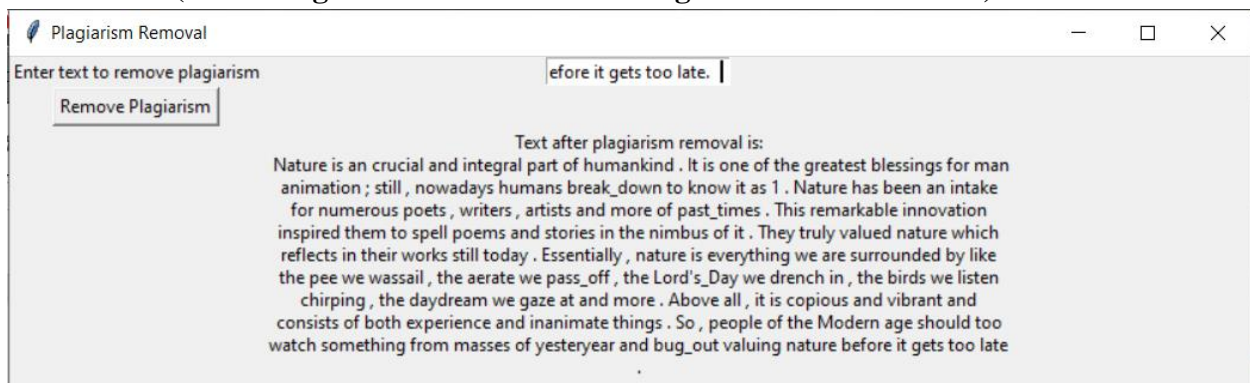


Figure 2: Output for Plagiarism Removal for Sample Text 1 (rerun)

**After rerun (User can get different flavour of Plagiarism Removed Text once again)**  
**(Note: User can rerun any no. of times)**

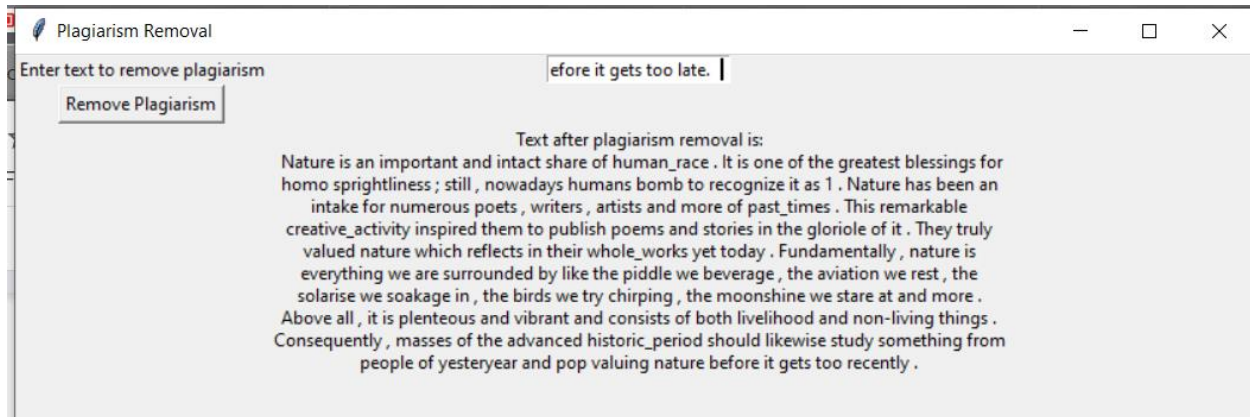


Figure 3: Output for Plagiarism Removal for Sample Text 1 (rerun once again)

## Sample Text 2

Friendship is one of the greatest blessings that not everyone is lucky enough to have. We meet a lot of people in the journey of life but there are only a few who leave a mark on us. My best friend is one such person who has been able to make a positive impact on my life. We have been a part of each other's lives for the longest time and our friendship is still developing. Most importantly, I feel extremely fortunate to have someone as a best friend in my life. In this essay on my best friend, I will tell you about how we became friends and about her best qualities.

## Text after Plagiarism Removal

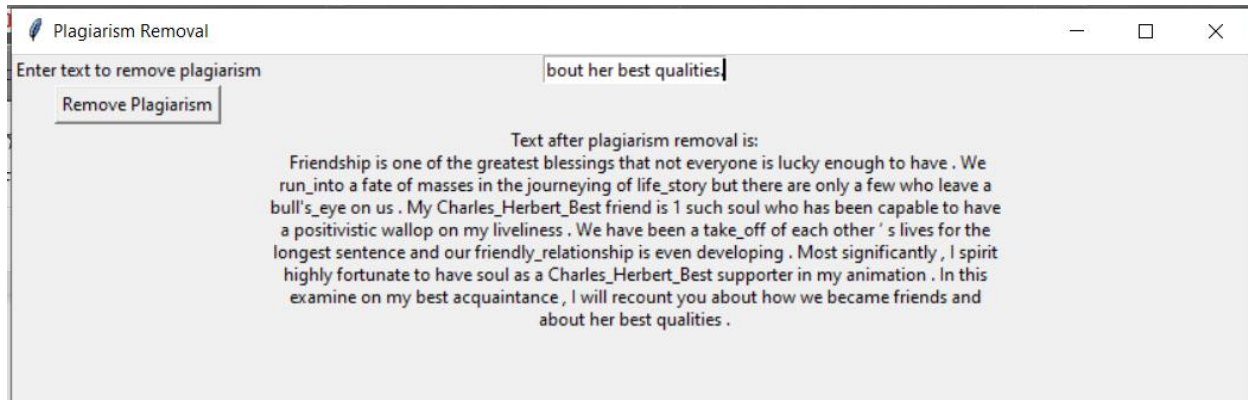


Figure 4: Output for Plagiarism Removal for Sample Text 2

## After rerun (User can get different flavour of Plagiarism Removed Text)

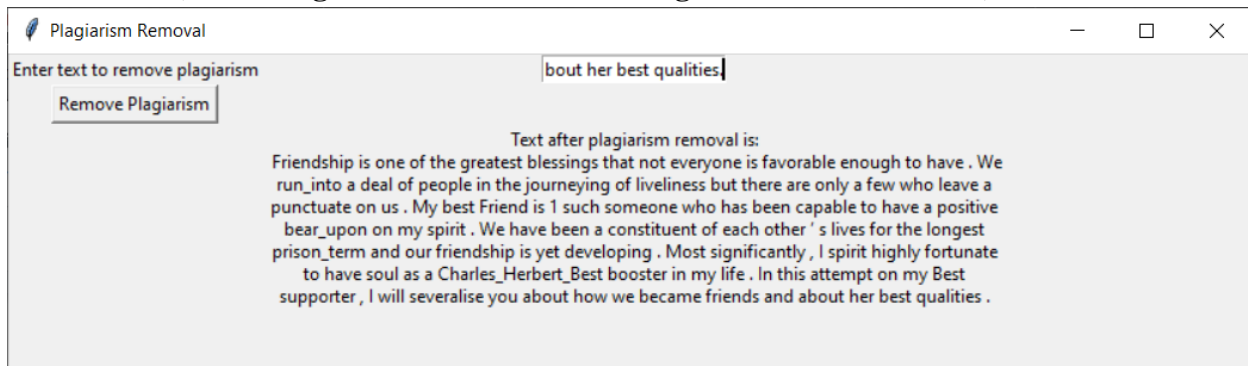


Figure 5: Output for Plagiarism Removal for Sample Text 2 (rerun)