



# Industrial Project Report

*Submitted in partial fulfillment of the degree of*



## **B-tech in Electronics And Communications Engineering**

**By**

***PARTHIB SAHA[11900322003]***

***DEBAJYOTI LOHA [11900322006]***

***ONKAR CHOWDHURY[11900322009]***

***SAMARPAN THAPA [11900322019]***

***PRASANT PRADHAN [11900322012]***

**First-year student of**

**SILIGURI INSTITUTE OF TECHNOLOGY**

*THIS IS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF*  
**AFFILIATED TO**

**Maulana AbulKalam Azad University of Technology**



**Under the supervision of :- Mr. RipamKundu**

**PROJECT ON :- PLAGARISM DETECTOR**

By

**PARTHIB SAHA [11900322003]**

**DEBAJYOTI LOHA [11900322006]**

**ONKAR CHOWDHURY[11900322009]**

**SAMARPAN THAPA [11900322019]**

**PRASANT PRADHAN [11900322012]**

UNDER THE GUIDANCE OF

**Mr. Ripam Kundu**

**Project Guide**

**SikharthyInfotechPvt. Ltd.**



*THIS IS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF*

**B.Tech**

IN

Electronics and Communication Engineering

**SILIGURI INSTITUTE OF TECHNOLOGY**

**AFFILIATED TO**

**MaulanaAbulKalam Azad University of Technology**

# **Department of Electronics And Communications**

## **Engineering**

I hereby forward the documentation prepared under my supervision by **Ripam Kundu Sir** entitled **Siliguri Institute Of Technology** to be accepted as fulfillment of the requirement for the Degree of Bachelor of Technology in Electrical Engineering, **Siliguri Institute Of Technology** affiliated to **Maulana Abul Kalam Azad University of Technology (MAKAUT)**.

---

**Mr. Ripam Kundu**  
**(Software Developer)**  
Project Guide  
Sikharthy Infotech Pvt. Ltd.

---

**HOD**  
**Department Of Electronics and Communication**  
**Engineering, SIT**

---

**Shilpi Ghosal**  
**(Director)**  
Sikharthy Infotech Pvt. Ltd.

---

**TPO**  
**Siliguri Institute of Technology**

## **Certificate of Approval**

The foregoing project is hereby approved as a creditable study for the B.Tech in Electronics and Communications Engineering presented in a manner of satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorsed or approved any statement made, opinion expressed or conclusion therein but approve this project only for the purpose for which it is submitted.

Final Examination for  
Evaluation of the Project

-----

-----

-----

**Signatures of Examiners**

## **ABSTRACT**

*The paper explains the working of a Plagiarism Checker. Plagiarism is presenting work or ideas from another source as your own, with or without consent of the original author, by incorporating it into your work without full acknowledgement. A plagiarism checker uses advanced database software to scan for matches between your text and existing texts. The results are usually presented in percentages, allowing you to compare how much of your work is unique and how much of it is copied. It also provides the source or sources with similar texts used in your work. Plagiarism checkers are important tools that do more than make sure you aren't copying someone's work. They also protect your reputation and keep you out of legal trouble.*

## **ACKNOWLEDGEMENT**

It is a great pleasure for me to acknowledge the assistance and participation of a large number of individuals in this attempt. Our project report has been structured under the valued suggestion, support, and guidance of **Mr.Ripam Kundu**. Under his guidance, we have accomplished the challenging task in a very short time.

Finally, we express our sincere thankfulness to our family members for inspiring me all throughout and always encouraging us.

### **Group Member Signature**

---

---

---

---

---

## **TABLE OF CONTENTS**

- Introduction
- Flask
- Cosine\_similarity
- Database for comparison
- Loops comes in play
- Output and return the end
- What user will see
- Code & Screenshot
- Conclusion

## INTRODUCTION

We will use [Python](#) and its different libraries to complete the plagiarism checker

## WHAT LIBRARIES WE USED

### Importing Libraries

The analysis will be done using the following libraries :

- Flask: Flask is a light-weight web framework for python.
- Re: A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).
- Math: Math is a built-in module in the Python 3 standard library that provides standard mathematical constants and functions. You can use the math module to perform various mathematical calculations, such as numeric, trigonometric, logarithmic, and exponential calculations.

To importing all these libraries, we can use the below code :

```
from flask import Flask, request, render_template
import re
import math
```

### Creating a flask web app with html interface in Post method

```
app = Flask("__name__")

q = ""

@app.route("/")
def loadPage():
    return render_template('index.html', query="")

@app.route("/", methods=['POST'])
```



## Creating a cosine Similarity function:

Cleaning Cosine Similarity is a metric that allows you to measure the similarity of the documents.

```
def cosineSimilarity():
    try:
        universalSetOfUniqueWords = []
        matchPercentage = 0

        #####

        inputQuery = request.form['query']
        lowercaseQuery = inputQuery.lower()

        queryWordList = re.sub("[^\w]", " ", lowercaseQuery).split() #Replace punctuation by space and split
        # queryWordList = map(str, queryWordList) #This was causing divide by zero error

        for word in queryWordList:
            if word not in universalSetOfUniqueWords:
                universalSetOfUniqueWords.append(word)

        #####
```

## DATABASE :-

READING AND APPENDDDING THE DATABASE TO A NEW LIST WITH CONDITIONS REQUIRED FOR THE COMPARISON.

```
fd = open("database1.txt", "r")
database1 = fd.read().lower()

databaseWordList = re.sub("[^\w]", " ", database1).split() #Replace punctuation by space and split
# databaseWordList = map(str, databaseWordList) #And this also leads to divide by zero error

for word in databaseWordList:
    if word not in universalSetOfUniqueWords:
        universalSetOfUniqueWords.append(word)
```

# LOOPS FOR COMPARISONS AND TO GET THE PERCENTAGE OF PLAGIARISM SPOTTED .

```
queryTF = []
databaseTF = []

for word in universalSetOfUniqueWords:
    queryTfCounter = 0
    databaseTfCounter = 0

    for word2 in queryWordList:
        if word == word2:
            queryTfCounter += 1
    queryTF.append(queryTfCounter)

    for word2 in databaseWordList:
        if word == word2:
            databaseTfCounter += 1
    databaseTF.append(databaseTfCounter)

dotProduct = 0
for i in range (len(queryTF)):
    dotProduct += queryTF[i]*databaseTF[i]

queryVectorMagnitude = 0
for i in range (len(queryTF)):
    queryVectorMagnitude += queryTF[i]**2
queryVectorMagnitude = math.sqrt(queryVectorMagnitude)

databaseVectorMagnitude = 0
for i in range (len(databaseTF)):
    databaseVectorMagnitude += databaseTF[i]**2
databaseVectorMagnitude = math.sqrt(databaseVectorMagnitude)

matchPercentage = (float)(dotProduct / (queryVectorMagnitude * databaseVectorMagnitude))*100
```

## RETURNING THE OUTPUT :-

```
    output = "Input query text matches %.02f%% with database."%matchPercentage

    return render_template('index.html', query=inputQuery, output=output)
except Exception as e:
    output = "Please Enter Valid Data"
    return render_template('index.html', query=inputQuery, output=output)

app.run()
```

# THE USER INTERFACE OF PLAGIARISM CHECKER

Input Text :

A large, empty rectangular text area with a thin black border, intended for the user to paste or type text for plagiarism checking.

CHECK PLAGIARISM

## ENTERING INPUT IN THE TEXTAREA

Input Text :

The cat (*Felis catus*) is a domestic species of small carnivorous mammal.[1][2] It is the only domesticated species in the family Felidae and is commonly referred to as the domestic cat or house cat to distinguish it from the wild members of the family.[4] Cats are commonly kept as house pets but can also be farm cats or feral cats; the feral cat ranges freely and avoids human contact.[5] Domestic cats are valued by humans for companionship and their ability to kill rodents. About 60 cat breeds are recognized by various cat registries.[6]

CHECK PLAGIARISM

## **RESULT AFTER CHECKING PLAGIARISM WIH THE DATABASE**

Input Text :

The cat (Felis catus) is a domestic species of small carnivorous mammal.[1][2] It is the only domesticated species in the family Felidae and is commonly referred to as the domestic cat or house cat to distinguish it from the wild members of the family.[4] Cats are commonly kept as house pets but can also be farm cats or feral cats; the feral cat ranges freely and avoids human contact.[5] Domestic cats are valued by humans for companionship and their ability to kill rodents. About 60 cat breeds are recognized by various cat registries.[6]

CHECK PLAGIARISM

**Input query text matches 87.54% with database.**

# **FUNCTIONAL REQUIREMENTS OF THE SYSTEM**

## ***SOFTWARE:***

- Windows OS 11
- VS Code

## ***WEB BROWSER:***

- Brave Browser
- Firefox
- Chrome
- Opera

## ***CODING LANGUAGE :***

- Python

## **Conclusion :**

The Plagiarism checker with the help of the database can check the input and compare the two documents letter by letter and to check for any similarities present in both of the documents . On the basis of which it will show the percentage representing the amount of plagiarism is there between both the documents. With the help of Flask framework and some other modules, the plagiarism checker does it's job . Plagiarism is definitely one of the problem our society is facing .It causes the real creator loses it's value because of some other person just copy pasted the real work and masked it with showing the false claims of owning the thing. It's result in the recruitment of the person without proper skillset and hence it's going to affect the company as well.

# REFERENCE

**MR. RIPAM KUNDU SIR**

**[WWW.GEEKSFORGEEKS.IN](http://WWW.GEEKSFORGEEKS.IN)**

**GITHUB**