

**361 CCS-4 – Artificial Intelligence**

**Term 2 –2022/2023**

**Plagiarism detection software in**

**text files**

* Group #:
* Members of the group:

|  |  |  |  |
| --- | --- | --- | --- |
| # | Name | ID | Role |
| 1 |  |  | Leader |
| 2 |  |  | Member |
| 3 |  |  | Member |
| 4 |  |  | Member |
| 5 |  |  | Member |

# 1.0 Introduction

The project is in place to help reduce plagiarism and cheat assignments.

It is a program in the Python language whose first task is to compare the two files suspected of cheating, which will show the percentage result.

Depending on this result, the administrator can then determine whether it was a fraud or not.

For example, it is possible to specify a general percentage (30%) as a maximum of similarity between the files, but if the percentage exceeds the aforementioned, it will be considered as fraud and a plagiarism process that has been applied against the duty and the monitoring of grades in this way is fair for everyone.

# .2.0 Methodology

Plagiarism checkers work by using advanced codes software to scan for matches between the inner text in files.

The accuracy is determined by two factors:

* the algorithm (which recognizes the plagiarism)
* the size of the file (which text is inside it).

# 2.1 Tools

**Hardware:**

Computer with 64bit or 32bit is good also.

About the RAM it should be 4G and above.

Windows 10 at least for defied any error with software.

**Software:**

Notepad in windows to create text files and save the texts inside it.

Jupyter light to start coding:

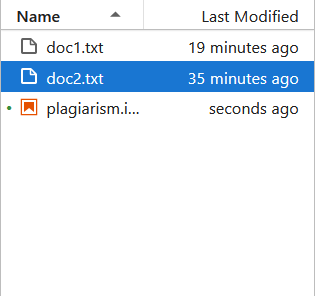
In the jupyter we used **difflib** Python package.

Which offer classes and functions for comparing sequences.

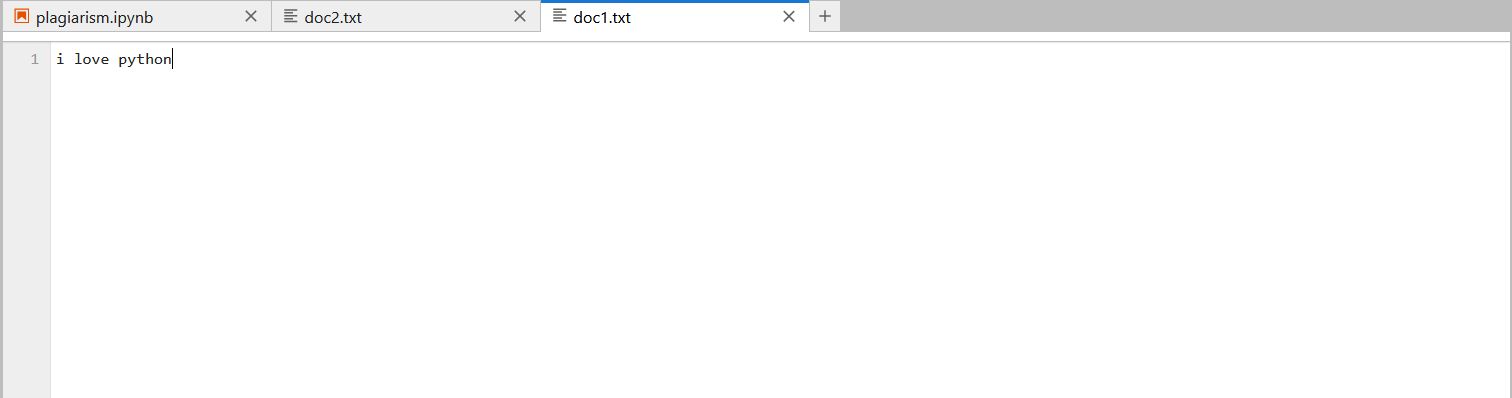
And from this package we used SequenceMatcher

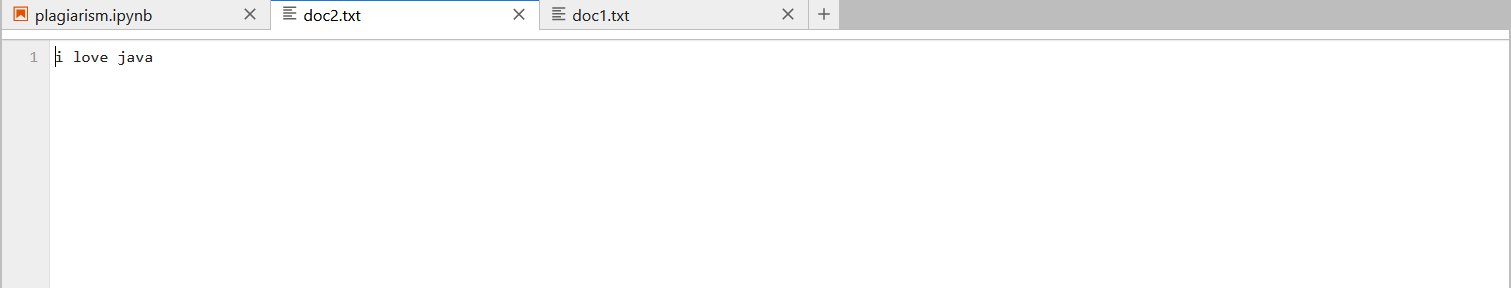
Which offered compare files and can produce information about file differences in various formats.

The files we used:



Inside the files:





## 2.2 Code

**# importing SequenceMatcher of difflib module**

from difflib import SequenceMatcher

with open('doc1.txt') as first\_file,

     open('doc2.txt') as second\_file:

    # Reading Both Text Files

    file1 = first\_file.read()

    file2 = second\_file.read()

    # Comparing Both Text Files

    ab = SequenceMatcher(None, file1,

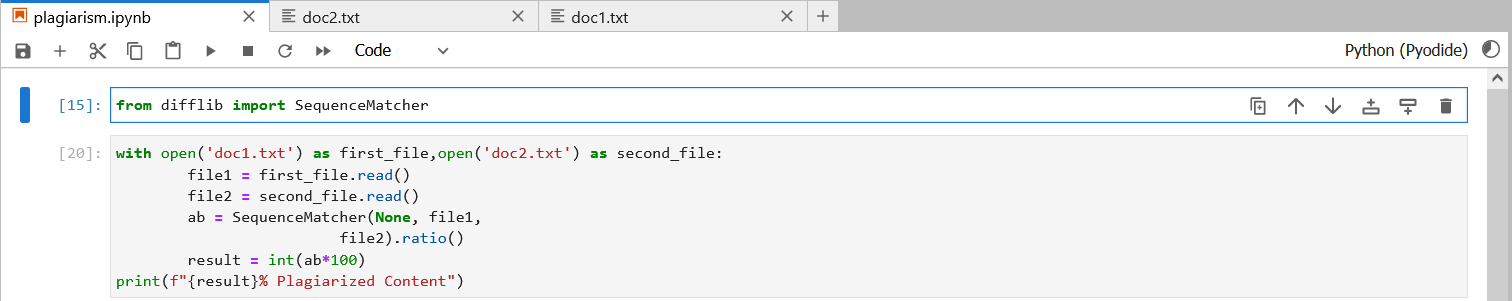
                         file2).ratio()

    # converting decimal output in integer

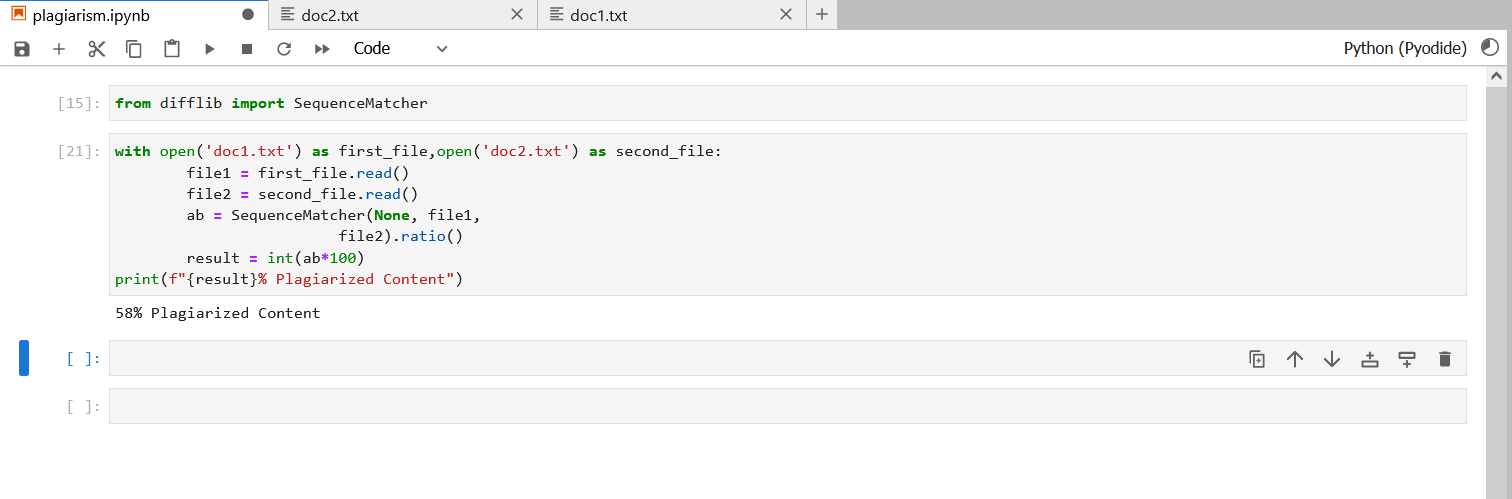
    result = int(ab\*100)

    print(f"{result}% Plagiarized Content")

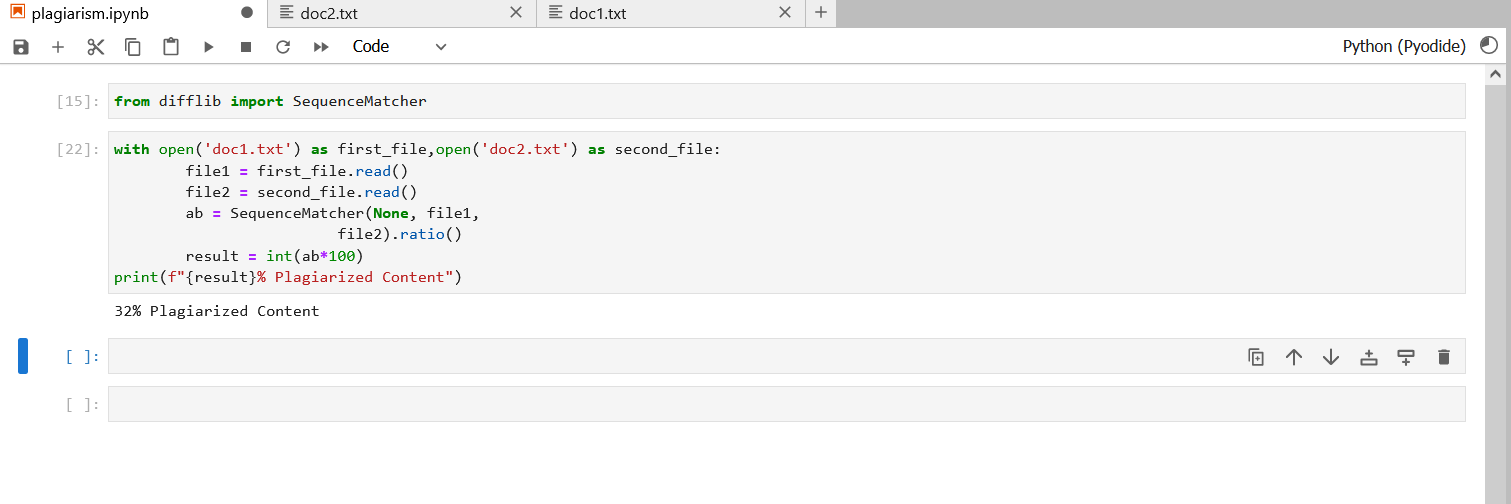
screenshot:



## 3.0 Result



After change the text inside the docs:



# References

1. <https://www.educative.io/> . to learn about the difflib

2.python.org . learn about library

3.jupyter online