

# Plagiarism Checker Documentation

## Overview

This document explains a simple Python based plagiarism checker. The program compares two text files and calculates the percentage of common words to estimate plagiarism.

## How the Code Works

- The function accepts two text file names as input.
- Both files are opened and read in lowercase format to ensure fair comparison.
- The text from each file is split into individual words.
- Words are stored in sets to automatically remove duplicate entries.
- Common words between the two files are identified using set intersection.
- Plagiarism percentage is calculated based on the ratio of common words to total unique words in the first file.
- The final similarity percentage is printed as output.

## Complete Code

```
def plagiarism_check(file1, file2):  
    # Open the first file in read mode and convert all text to lowercase  
    with open(file1, "r", encoding="utf-8") as f1:  
        text1 = f1.read().lower()  
  
    # Open the second file in read mode and convert all text to lowercase  
    with open(file2, "r", encoding="utf-8") as f2:  
        text2 = f2.read().lower()  
  
    # Split text into words and convert them into sets  
    words1 = set(text1.split())  
    words2 = set(text2.split())  
  
    # Find common words present in both files  
    common_words = words1.intersection(words2)  
  
    # Calculate similarity percentage  
    similarity = (len(common_words) / len(words1)) * 100  
  
    # Display the plagiarism percentage  
    print("Plagiarism Percentage:", round(similarity, 2), "%")  
  
plagiarism_check("file1.txt", "file2.txt")
```

## Author

Satyam Gajjar