Combat Online Plagiarism with AI

```
from difflib import SequenceMatcher
```

The SequenceMatcher class from the difflib module in Python is used to compare pairs of sequences, such as strings or lists, and to identify similarities and differences between them. It's a powerful tool for tasks like comparing text files, finding the longest contiguous matching subsequence, and more.

```
with open(r"/content/1.txt") as file1,open(r"/content/3.txt") as
file2:
  file1_data = file1.read()
  file2_data = file2.read()

  similarity_ratio = SequenceMatcher(None, file1_data,
file2_data).ratio()
  print(similarity_ratio)
```

Let's break down the provided code which compares the contents of two text files and calculates their similarity using SequenceMatcher from the difflib module.

Step-by-Step Explanation:

Import SequenceMatcher from difflib:

```
from difflib import SequenceMatcher
```

This imports the SequenceMatcher class which is used for comparing sequences.

Open the Files:

```
with open(r"/content/1.txt") as file1,open(r"/content/3.txt") as
file2:
```

- with open(r"/content/1.txt") as file1 opens the first file in read mode and assigns the file object to file1.
- with open(r"/content/3.txt") as file2 does the same for the second file and assigns the file object to file2.
- The with statement ensures that the files are properly closed after their contents are read, even if an error occurs.

Read File Contents:

```
  file1_data = file1.read()
  file2_data = file2.read()
```

- file1.read() reads the entire content of file1 and stores it in the variable file1_data.
- file2.read() reads the entire content of file2 and stores it in the variable file2_data.

Create a SequenceMatcher Object and Calculate Similarity Ratio:

```
similarity_ratio = SequenceMatcher(None, file1_data,
file2_data).ratio()
```

- SequenceMatcher(None, file1_data, file2_data) creates a SequenceMatcher object with file1_data and file2_data as the sequences to be compared.
- The first argument None is a placeholder for a custom function to filter out unwanted elements from the sequences, which is not used in this case.
- .ratio() computes the similarity ratio of the two sequences. This ratio is a float between 0 and 1, where 1 indicates that the sequences are identical, and 0 indicates that they are completely different.

Print the Similarity Ratio:

```
print(similarity_ratio)
```

This prints the calculated similarity ratio to the console.