Design and Analysis of Algorithms, Fall 2019

Term Project

Simple Plagiarism detection

Plagiarism is a serious problem in academic integrity. In this project, you will implement a simple plagiarism detection algorithm. The following is the assumptions:

1. The input is plain English text (no equation, graph, picture, or computer code. There might be numbers such "45,456".
2. You can pick your programming language of your choice.
3. The length of input document varies. During the test, you will be given a moderate size.
4. The input file will be in plain text with no formatting metadata such as font.
5. During the test, you will be two text files: one as source (document S) and the second is the document being reviewed (document R).
6. You must report:
   a. The similarity score: the degree of plagiarism of the document R from document S (from 0 to 1, where 1 means document S is a full copy of S and zero means completely original).
   b. The matching parts in both documents.
7. I know the similarity and also the matching parts in advanced. Your work will be marked based on the quality of your detection.
8. Deadline: November 29, 2019