Assignment No. B 2

Aim

Plagiarism Detection

Problem Definition

Write a web application using Scala/ Python/ Java /HTML5 to check the plagiarism in the given text paragraph written/ copied in the text box.
Give software Modeling, Design, UML and Test cases for the same using Analysis Modeling (Static Modeling, Object Structuring, Dynamic Modeling).

Learning Objectives

- Understanding concepts of Plagiarism Detection
- Learn how to implement Plagiarism detection algorithms & Web application Development
- Learn about USE-CASE modeling in Modelio Software

Learning Outcome

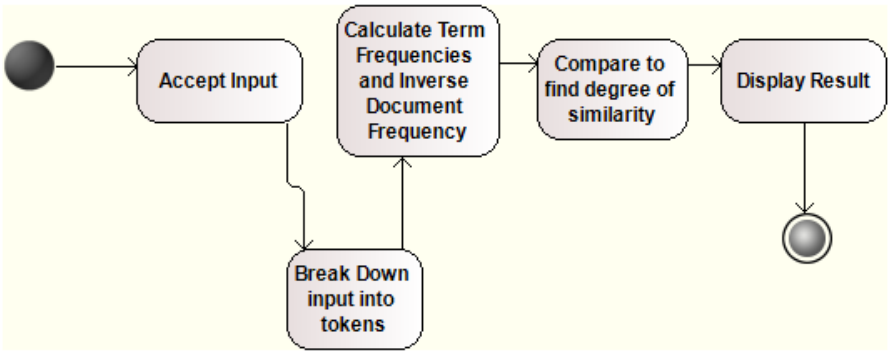
- Learnt concepts of Plagiarism Detection
- Implemented Plagiarism detection algorithm in a Web application
- Learnt USE-CASE modeling in Modelio Software

Software And Hardware Requirements

- Latest 64-BIT Version of Linux Operating System
- Eclipse with Python and Dynamic Web Project Module
- Modelio Software

Mathematical Model

```
Let S be the system of solution set for given problem statement such that,
S={ s,e,X,Y,F,DD,NDD,Su,Fu }
where,
s= Start State, where  Y={}

e= End State, where Y={Y1|Y2}
    where, Y1= Plagiarism Detected , Y2= Plagiarism Not Detected

X= Input Set,
    X={X1,X2}
    where,
    X1= Predefined File with which comparison takes place
    X2= User Input Text

Y= Output Set,
    Y={Y1|Y2}
    where, Y1= Plagiarism Detected , Y2= Plagiarism Not Detected

DD= Deterministic Data Set
    DD={User Input is same or different than the predefined file}

NDD= Non Deterministic Data Set
    NDD={User Input is null || File content is null}

Su= Success Case
    {User Input is not empty & File is not empty}

Fu= Failure Case
    {Either User input of File is Empty}

F= Set of functions
    F= {F1,F2,F3}
    where,
    F1= Accept User Input
    F2=Compare with File
    F3=Display Output
```



State Diagram

Theory

Plagiarism Detection

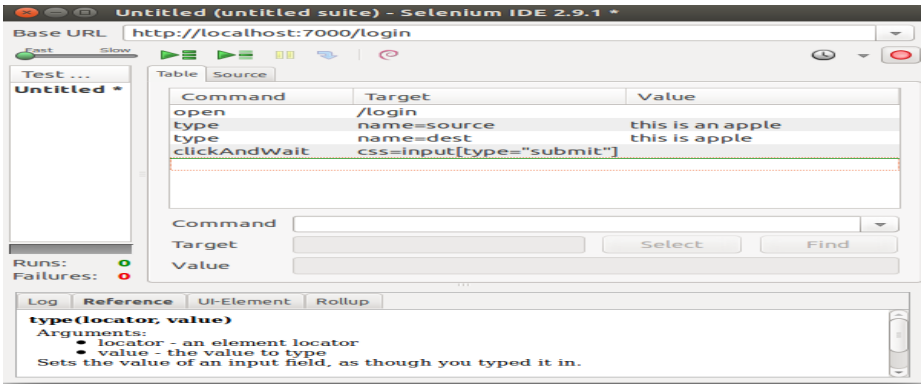
- Plagiarism detection is the process of locating instances of plagiarism within a work or document. The widespread use of computers and the advent of the Internet has made it easier to plagiarize the work of others. Most cases of plagiarism are found in academia, where documents are typically essays or reports. However, plagiarism can be found in virtually any field, including scientific papers, art designs, and source code.
- Detection of plagiarism can be either manual or software-assisted. Manual detection requires substantial effort and excellent memory, and is impractical in cases where too many documents must be compared, or original documents are not available for comparison. Software-assisted detection allows vast collections of documents to be compared to each other, making successful detection much more likely.
- The practice of plagiarizing by use of sufficient word substitutions to elude detection software is known as rogeting.
- Systems for text-plagiarism detection implement one of two generic detection approaches, one being external, the other being intrinsic.External detection systems compare a suspicious document with a reference collection, which is a set of documents assumed to be genuine. Based on a chosen document model and predefined similarity criteria, the detection task is to retrieve all documents that contain text that is similar to a degree above a chosen threshold to text in the suspicious document. Intrinsic PDS solely analyze the text to be evaluated without performing comparisons to external documents. This approach aims to recognize changes in the unique writing style of an author as an indicator for potential plagiarism. PDS are not capable of reliably identifying plagiarism without

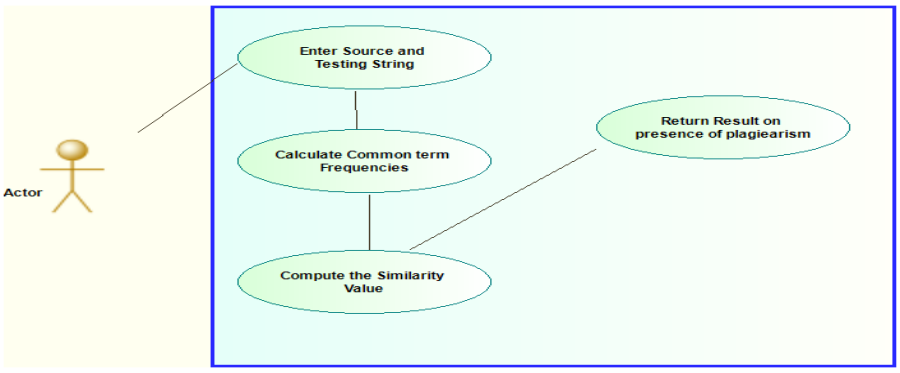
human judgment. Similarities are computed with the help of predefined document models and might represent false positives.

USE-CASE Diagram

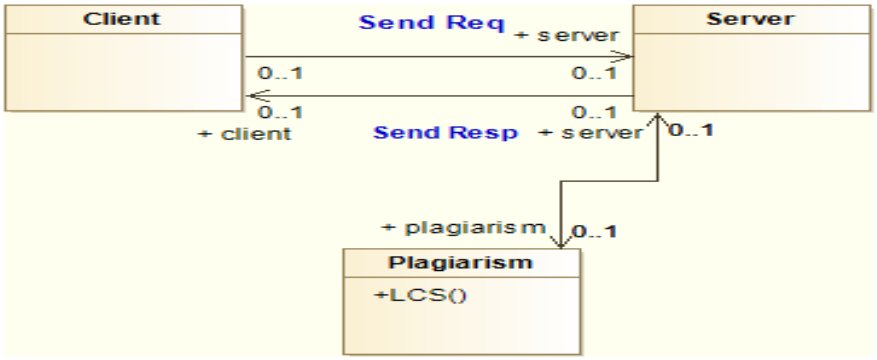
.

The plagarism percent is 11.1111111111 %



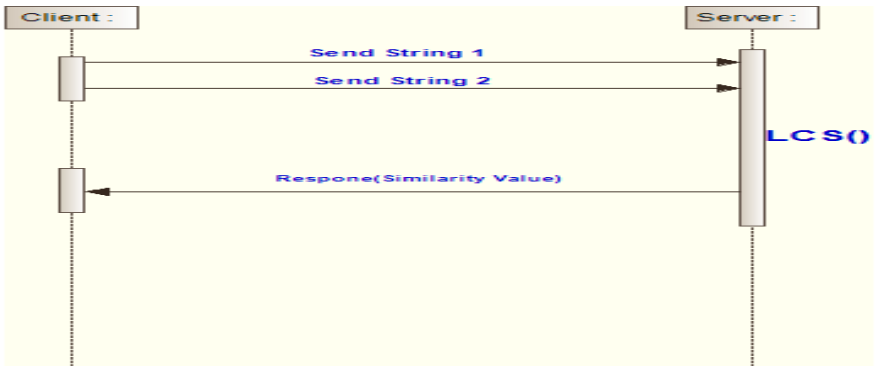


Class Diagram



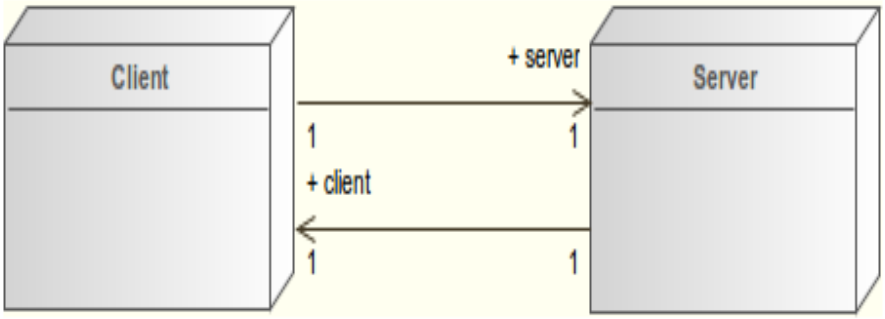
.

Sequence Diagram



.

Deployment Diagram



Conclusion

Thus, We have successfully implemented a plagiarism checker web application in Flask in python. Here we are checking the user input with different files in the database and concluding whether the given text input is plagiarised or not.