Northeastern University

MSD Project Plan

Plagiarism Detection System

Varun Nandu
(nandu.v@husky.neu.edu)

Krrish Mittal
(mittal.kr@husky.neu.edu)

Harsha Rahul Boggaram Krishna Prasad
(boggaramkrishnapra.h@husky.neu.edu)

Mingchao Wu
(wu.ming@husky.neu.edu)

Supervised by
Dr. Jose Annunziato
Dr. Mike Weintraub

January 30, 2018

1 Introduction

Plagiarism is the "wrongful appropriation" and "stealing and publication" of another author's "language, thoughts, ideas, or expressions" and the representation of them as one's own original work. Plagiarism is considered academic dishonesty and a breach of journalistic ethics. This is a growing problem in today's times in educational institutions. Thus in order to protect one's academic integrity plagiarism detection has gained a lot of importance these days. Hence, we are building a Plagiarism Detection System which will provide a similarity score between two projects written in **Python 3**.

2 Technologies

In this project, we are going to use the following technologies to implement our system:

- The Front-End of the system will be built using HTML, CSS, Bootstrap and Angular JS
- The Back-End will have Java used for the comparison logic.

3 The Plan

Our plan to implement the system will consist of three phases.

- In Phase A, the project objectives and the major deliverables are identified. The project is initiated to deliver the solution. A mockup for the user interface is designed and developed.
- In Phase B, Design Documents will be developed wherein the team will identify the steps used in the design of the application. First, the general system characteristics will be defined. The application logic will be designed and documented in the form of UML diagrams mainly class and sequence diagrams.
- In Phase C, the project plan is put into motion and the project team will accomplish the work defined in the plan. People will be carrying out the tasks and progress will be reported through regular team meetings. Appropriate adjustments will be made when the project factors change. The implementation of the system will consist of the following steps:
    - Creating Project Upload Page for the User.
    - After Uploading files creating a front-end controller to pass the files to Java Back-End.
    - Using ANTLR library to create Abstract Syntax Tree (AST) of the files.
    - Finding Similarity between AST of the two projects and assigning a Similarity score.
    - Sending the score data back to the front end controller.
    - Creating a Result Page for the User.