

Plan for Plagiarism Detector System

We plan to develop a plagiarism detection system for python. It is targeted to be used by Instructors and TA's to check the submissions for plagiarism.

We'll be checking plagiarism of the following forms:

1. Two submissions differ in only the names of variables.
2. Two submissions might have similar comments.
3. A method from one of the submissions is broken down into multiple methods in the other.
4. Two submissions have the same piece of code written in different order.
5. Detection through metadata of files.

We plan to develop the system in incremental phases in about 10 weeks.

Week 1-2 (Jan 29- Feb 9)

Our target for week 1 and 2 is to design the architecture for the system and decide on the design patterns. We plan to implement the login page, and the functionality of batch upload of student's source code files.

Week 3-4 (Feb 10- Feb 23)

Our target for week 3, 4 is to implement logic for Abstract Syntax Trees and find out plagiarism by analyzing metadata and by looking for similarity in comments. We also plan to design test suites.

Week 5-6 (Feb 24- Mar 2)

Our target for week 5, 6 is to implement logic for Abstract Syntax Trees and find out plagiarism by analyzing variable name changes, modularity, and moving code around.

Week 7-8 (Mar 3-Mar 16)

Our target for week 7, 8 is to design the front end along with highlighting the differences in files and download plagiarism detection report option(optional).

Week 9-10 (Mar 17-Mar 30)

Our target for week 9, 10 is to do exhaustive testing of application.

We will be doing unit test side by side for each module.

We plan to contribute 20 hrs. per week to this project.