Code Similarity using Program Dependence Graph

Tobin Pereira (tob5232@rit.edu)

ADVISOR: Dr. Carlos Rivero

ROCHESTER INSTITUTE OF TECHNOLOGY

Objective

- Use Graph Alignment Techniques[2] on PDG representations of code to perform approximate subgraph isomorphism to find “similar” code(s).

Program Dependence Graph

- A graph representation of a piece of code.
- The representation tries to capture the data and control dependency of the code[1].

Alignment

- The data used to check the results of the system were 3 assignments picked from codechef[3] each having an average of about 600 submissions.

Results

- Program Dependence Graph is a very good abstract representation of code which frees it from constrains which make it difficult to identify plagiarized code or even “similar” code.
- This project lays the foundation of a system which can acts as a plagiarism checker for code or also as system which provides large scale personalized feedback for programming assignments.

References

