

Thesis Topic

Plagiarism detection in Git/GitLab

Degree level

Bachelor or Master

Company

LNU

Description

Programming education at Linnaeus university are using GitLab to store, submit and handle student code-submissions. Currently the plagiarism detection toolset included in gitlab is nonexistent. This thesis is aimed at filling that hole.

How to create plagiarism tools based on Git-technology that deters students from copying code from others.

Current codeplagiarsm tools are based on comparing end-results. Git offers the opportunity to include the evolution of the code into evidence of code-duplication using git-history. This would also help providing transparency on how the code was built that could be used to encourage better study habits. Building a database with code both from the university and external open resources.

Objectives

This project may include

- Creating and overview of existing tools for plagiarism detection.
- Investigate plagiarismdetection over time using git-history together with evolution of code to create an simple overview for student and teachers.
- Create visualisation of git-history (Time, Tests, Code-size, Code-Duplication)
- Creating solution database. Inclusion of code from other sources for similar projects. Building a database with code both from the university and external open resources.
- Build prototype. Integration with existing tools (For example JPlag) in gitlab. Code is compared on git push, visualisations and reports are built.
- Evaluation of the prototype using end to end test. Creating datasets to evaluate different scenarios

Daniel Toll daniel.toll@lnu.se