

# Detection of Software Plagiarism

Degree programme : BSc in Computer Science | Specialisation : Digital Business Systems  
Thesis advisor : Prof. Dr. Olivier Biberstein  
Expert : Igor Metz (Glue Software Engineering AG)

With the advent of computer science, we talk more and more about piracy. Every year a lot of companies and schools all over the world open legal actions and report incorrect behavior of their employees or their students. Plagiarism is an illegal activity in which a person takes possession of a job carried out by an external person without citing him

## Premise

For schools, computing plagiarism means that a student is assessed incorrectly because his abilities do not reflect his personality. For a company this means a loss of internal knowledge, outsourcing the source code to people who could use it to fulfill secondary purposes.

## Solution

Discovering plagiarism between programs has always been a challenge since the earliest days of computer science. Today there are no official solutions to fulfill this task. What exists is bachelor, master or doctorate thesis carried out by persons of some universities in the world or simply internal research of a specific university. Most of the time these projects are then supplied Open Source to the internet community in the hope of continuous research and development by users.

The results of plagiarism provided by these tools differ a lot due to the strategies adopted during the process analysis.

## The Idea

The idea of my bachelor thesis is to develop a utility that allows you to discover the plagiarism between two software written in Java 8 language and then analyze the results in a concise and orderly manner

## Prototypes

There are three ways considered critical for analyzing plagiarism between software. The first way is to analyze the bytecodes generated by the compilation of two projects. The second way is to analyze the two structures of the abstract syntax tree of the source files, while the third way analyzes the dependencies that occur at the level of packages and classes. Depending on the way you want to refer to, an algorithm is used that analyzes the main phase of plagiarism.

## Optional Tasks

Next to this software lies a graph database in which it's possible to save the structures of the various projects that you wish to analyze in the future. Through a graphical interface, the user can then interact with the plagiarism software whose requests are transmitted through a REST service.



Davide Andrea Vanoni  
davide95.v@gmail.com

