

## Detecting the Code Clone Density and Measuring the Complexity of Source Code Using SonarQube

T. Vijayanandan\*, B.T.G.S. Kumara, and K. Banujan

Department of Computing and Information Systems, Faculty of Applied Sciences,  
Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.

\*[mvthanoshan@std.appsc.sab.ac.lk](mailto:mvthanoshan@std.appsc.sab.ac.lk)

In software development, developers often copy and paste the code, with or without modifications. In the software development sector, maintenance accounts for roughly 90% of the software lifetime cost. The concept of code cloning is regarded as one of the aspects of bad smell which complicates the maintainability of the software. In the past, the adverse effects of code clones were measured in terms of their bug-proneness, fault-proneness, tendency of experiencing constant modifications, including bug fixes, and a lifetime of clones across the software project. However, none of the studies analyzed the source code complexity in relation to the density of code clones. This study aims to analyze the association between code clone density and complexity. The study design is two-folded: analysis of duplicated files at the overall project-level and analysis of duplicated files at the individual file-level. In this study, we investigated twenty-two well-known Apache Software Foundation subject systems written in Java programming language for overall project-level and analyzed 416 source files among ten subject systems considered above for individual file-level. The SonarQube tool was utilized to detect code clones and to measure the complexity. A positive correlation coefficient of 0.927 was identified among code clone density and the complexity of overall project-level duplicated files but with a less statistical significance ( $p=5.201$ ). A minor negative correlation coefficient of 0.099 was identified amongst code clone density and the complexity of its duplicated file with a statistical significance ( $p=0.042$ ). According to the results, it is concluded that there is non-significant correlation between code clone density and overall project-level duplicated files' complexity and a significant negative correlation between code clone density with its individual duplicated file's complexity. Although no conclusive evidence that code clone density increases software difficulty was identified, this study can serve as a foundation for future research on code clones and software complexity.

Keywords: Code Clones, Bug-Proneness, Density, Complexity