

STUBBER: Compiling Source Code into Bytecode without Dependencies for Java Code Clone Detection

André Schäfer

`andre.schaefer@uni-jena.de`

Friedrich Schiller University Jena, Jena, Germany

Wolfram Amme

`wolfram.amme@uni-jena.de`

Friedrich Schiller University Jena, Jena, Germany

Thomas S. Heinze

`thomas.heinze@dlr.de`

German Aerospace Center, Jena, Germany

In recent years, many clone detection tools for Java have been introduced. On the one hand, many of these tools work with Java source code and can therefore be conveniently evaluated with the modern evaluation framework BigCloneEval and the benchmark BigCloneBench. On the other hand, certain clone detectors deliberately address bytecode as input and cannot be applied to Java source code. Simple compilation often does not solve the problem due to missing required dependencies, as in the case of the Java files in BigCloneBench. Therefore, we present the STUBBER tool for compiling Java source code into bytecode without dependencies. We can show that STUBBER can successfully generate bytecode for over 95% of all Java source files and 92.5% of all code clones contained in BigCloneBench. Consequently, the evaluation of bytecode-based clone detectors with BigCloneEval on BigCloneBench becomes possible and such tools can thus also be compared with source code-based clone detectors.