NOVELTY STATEMENT

Journal First Submission of
"Toxic Code Snippets on Stack Overflow"
C. Ragkhitwetsagul, J. Krinke, M. Paixao, G. Bianco, R. Oliveto

The paper reports completely new results that are not reported before in any prior work. This study is among, if not the first, to address the important issues of toxic code snippets, including outdated and license-violating online code clones on Stack Overflow. The paper presents a study incorporating several state-of-the-art tools and techniques to automatically extract and filter online clone pairs between Stack Overflow and 111 open source projects in the curated Qualitas corpus. We analysed 2,302 non-trivial online clone candidates and reported 101 cases of outdated code snippets and 214 cases of potential software license violations. Our online survey of 201 Stack Overflow high reputation users (33% response rate) is valuable since it supports our findings of toxic code snippets, and also reveals the awareness of Stack Overflow users regarding the problems. Currently, there is no literature on finding the origins of code examples copied to Stack Overflow, their potential ramifications, and the awareness of Stack Overflow developers in doing so. Since Stack Overflow is actively used in research and software development, this study gives insights to the problem and lays a foundation to an automatic technique for finding toxic code snippets on Stack Overflow in the future.