

## Abstract

Online code clones are code fragments that are copied from software projects or online sources to Stack Overflow as examples. Due to an absence of a checking mechanism after the code has been copied to Stack Overflow, they can become toxic code snippets, i.e., they suffer from being outdated or violating the original software license. We present a study of online code clones on Stack Overflow and their toxicity by incorporating two developer surveys and a large-scale code clone detection. A survey of 201 high-reputation Stack Overflow answerers (33% response rate) showed that 131 participants (65%) have ever been notified of outdated code and 26 of them (20%) rarely or never fix the code. 138 answerers (69%) never check for licensing conflicts between their copied code snippets and Stack Overflow's CC BY-SA 3.0. A survey of 87 Stack Overflow visitors shows that they experienced several issues from Stack Overflow answers: mismatched solutions, outdated solutions, incorrect solutions, and buggy code. 85% of them are not aware of CC BY-SA 3.0 license enforced by Stack Overflow, and 66% never check for license conflicts when reusing code snippets. Our clone detection found online clone pairs between 72,365 Java code snippets on Stack Overflow and 111 open source projects in the curated Qualitas corpus. We analysed 2,289 non-trivial online clone candidates. Our investigation revealed strong evidence that 153 clones have been copied from a Qualitas project to Stack Overflow. We found 100 of them (66%) to be outdated and potentially harmful for reuse. Furthermore, we found 214 code snippets that could potentially violate the license of their original software and appear 7,112 times in 2,427 GitHub projects.