A Survey of Software Clone Detection Techniques

Volume 137
Number 10

Year of Publication: 2016

Authors:
Abdullah Sheneamer, Jugal Kalita

Abstract

If two fragments of source code are identical or similar to each other, they are called code clones. Code clones introduce difficulties in software maintenance and cause bug propagation. Software clones occur due to several reasons such as code reuse by copying pre-existing fragments, coding style, and repeated computation using duplicated functions with slight changes in variables or data structures used. If a code fragment is edited, it will have to be checked against all related code clones to see if they need to be modified as well. Removal, avoidance or refactoring of cloned code are other important issues in software maintenance. However, several research studies have demonstrated that removal or refactoring of cloned code is sometimes harmful. In this study, code clones, common types of clones, phases of clone detection, the state-of-the-art in code clone detection techniques and tools, and challenges faced by clone detection techniques are discussed.

References


24. X. Yan, J. Han, and R. Afshar. Clospan: Mining closed sequential patterns in large datasets, 2003.


52. Han, Jiawei. Data Mining: Concepts and Techniques. (2006).


Index Terms

Computer Science

Software Engineering
Keywords

Software Clone, Code Clone, Duplicated Code Detection, Clone Detection