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
20. J. Mayrand, C. Leblanc and E. Merlo, Experiment on the automatic detection of function
244-253.
21. E. Merlo, detection of plagiarism in university projects using metrics-based spectral
similarity, Proc. Dagstuhl Seminar 06301: Duplication, Redundancy, and Similarity in Software,
2006.
Software Engineering, 2008. ICSE’08. ACM/IEEE 30th International Conference on. IEEE,
2008.
24. X. Yan, J. Han, and R. Afshar. Clospan: Mining closed sequential patterns in large
Addison-Wesley, 1986.
26. Ducasse, Stéphane, Oscar Nierstrasz, and Matthias Rieger. On the effectiveness of clone
detection by string matching. Journal of Software Maintenance and Evolution: Research and
27. Roy, Chanchal K., James R. Cordy, and Rainer Koschke. Comparison and evaluation of
code clone detection techniques and tools: A qualitative approach. Science of Computer
Programming 74.7 (2009): 470-495.
29. Raheja, Kanika, and Rajkumar Tekchandani. An emerging approach towards code clone
detection: metric based approach on byte code. International Journal of Advanced Research in
THAPAR UNIVERSITY, 2011.
31. Krinke, Jens. Identifying similar code with program dependence graphs. Reverse
32. Liu, Chao, et al. GPLAG: detection of software plagiarism by program dependence
graph analysis. Proceedings of the 12th ACM SIGKDD international conference on Knowledge
33. Prechelt, Lutz, Guido Malpohl, and Michael Philippsen. Finding plagiarisms among a set
35. Patenaude, J-F., et al. Extending software quality assessment techniques to java
Working Notes of the Third Workshop on AI and Software Engineering: Breaking the Toy Mold
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