Abstract

Code cloning is the procedure where the developers reuse the code fragments implementing the paste option. They may or may not make the modification in the source code. The code thus developed after copying is known as clone. It is the synonym of duplicate. In the year 2002, Ira Baxter coined the term clones as the segments of code that are similar according to some definition of similarity. The similarity can be based on text, syntactic or semantic. Studies have revealed that almost 10-15% of the source code in large software are part of single or more clones[1]. Clones have adverse impact on the software maintenance, thus identification of clones is beneficial. In the past decade many tools have been developed to detect the clones but none was able to correctly identify all types of clones. In this paper the literature survey of all the clone detection techniques has been done. Along with this it also propose an approach which will use a combination of tree and token based approach in order to detect the code clones.
token-based code clone detection system for large scale source code." IEEE Transactions on

study of clones in the STL and some general implications." Proceedings of the 27th international

21. Mayrand, Jean, Claude Leblanc, and Ettore M. Merlo. "Experiment on the automatic
detection of function clones in a software system using metrics." Software Maintenance 1996,

on.IEEE, 1999.


24. Komondoor, Raghavan, and Susan Horwitz. "Using slicing to identify duplication in

25. Higo, Yoshiki, and Shinji Kusumoto. "Code clone detection on specialized PDGs with
heuristics." Software Maintenance and Reengineering (CSMR), 2011 15th European
Conference on. IEEE, 2011.


27. Tool SimScanhttp://www.blue-edge.bg/download.html

28. Project Bauhaus http://www.bauhaus-struggart.de


intentional clones using flexible pretty-printing and code normalization." Program

31. Roy, Chanchal K. "Detection and analysis of near-miss software clones." Software

in large scale systems." Program Comprehension (ICPC), 2010 IEEE 18th International

33. Roy, Chanchal K., and James R. Cordy. "Are scripting languages really different?."

34. Martin, Douglas, and James R. Cordy. "Analyzing web service similarity using contextual

35. Gupta, Sonam, and P. C. Gupta. "Literature survey of clone detection techniques."


language-independent detection of near-miss clones." Proceedings of the 2004 conference of
40. Monden, Akito, et al. "Software quality analysis by code clones in industrial legacy
41. Nguyen, Tung Thanh, et al. "Scalable and incremental clone detection for evolving
2009.
43. Perumal, A., Kanmani, S., & Kodhai, E. (2010, September). Extracting the similarity in
detected software clones using metrics. In Computer and Communication Technology (ICCCT),
2010 International Conference on (pp. 575-579). IEEE.
44. Li, Z. O., & Sun, J. (2010, April). A metric space based software clone detection
approach. In Information Management and Engineering (ICIME), 2010 The 2nd IEEE
International Conference on (pp. 393-397). IEEE.
problems with GPU-based algorithms. In Proceedings of the 4th International Workshop on
Software Clones (pp. 25-32). ACM.
accurate tree-based detection of code clones. In Proceedings of the 29th international
conference on Software Engineering (pp. 96-105). IEEE Computer Society.
47. Tairas, R., & Gray, J. (2010, March). Sub-clone refactoring in open source software
2373-2374). ACM.
clone management. In Proceedings of the 30th international conference on Software
engineering (pp. 843-846). ACM.
World Academy of Science, Engineering and Technology, 60, 350-354.

Index Terms

Computer Science
Software Engineering

Keywords
Clones, textual comparison, LWH approach, token based approach, PDG approach, metric comparison, AST approach.