Abstract

Source code plagiarism is becoming a common practice among higher education community. People duplicate and modify the source code of other people and show the program as their own program. In this paper, we want to draw researchers’ attention towards this problem and projected a novel approach which detects plagiarism in C language code by converting it into assembly language which is done with the help of GCC compiler. Assembly language converted by the compiler is not sensitive to all type of different code transformation, for example-swapping variable names, reformation of language, adding extra comment or blanks. Therefore, assembly language gives rise to reduced amount of variations, if there is a modification in the original code. Previous works in plagiarism compares the whole program but in this paper, we proposed a method which split the C program into assembly language code and divide each function of program into blocks and blocks are transformed into token strings. This method compares each function with other program function and provides a statistical output, according to the token string likeness of that function. If the output is above assigned specific plagiarism similarity threshold value then it counts under the case of plagiarism.
Plagiarism Detection of C Program using Assembly Language

References

12. Dick Grune website regarding to similarity measure URL: http://www.dickgrune.com/Programs/similarity_tester/
13. JPlag tool site URL: http://jplag.ipd.kit.edu
16. Shan S., Guo F., Ren J.: similarity detection method based on assembly language and string matching
18. Moussiades L., and vakali A.,: a clustering approach for detecting plagiarism in source code datasets, in proceeding of the computer journal (November 2005)

Index Terms

Computer Science

Information Sciences

Keywords

Plagiarism, assembly language, string similarity, Plagiarism detection method, token string