Code Clone Detection Model: A SWOT Analysis Perspective

Al-Fahim Mubarak-Ali1, Rahiwan Nazar Romli2, Nilam Nur Amir Sjarif3

1Software Engineering Research Group, Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia.
3Advanced Informatics School, Level 5, Menara Razak, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

Corresponding author Email: fahim@ump.edu.my
Received: 22 March 2017   Accepted: 29 August 2017

Code clone is similar source code that occurs in a program. Code clone affects the maintainability of software. Therefore, various code clone detection techniques such as text based comparison technique, token based comparison technique, tree based comparison technique, metric based comparison technique and graph based comparison technique have been proposed to detect code clone. Apart from these techniques, models also have been used to detect code clone. Model consist of step by step process that is utilized to detect code clones. This work looks into analyzing existing code clone detection models through SWOT analysis. The strength and weakness of these code clone detection models is discussed. The output of this work shows that code clone detection is a more procedural way of detecting code clone through process and the opportunity of expanding the research area is vast.

Keywords: Code Clone, Code Clone Detection Techniques, Code Clone Detection Model.

1. INTRODUCTION

Code clone is the usual terminology applied by the researchers. It refers to the source codes replicated many times in an application. There are also a variety number of terminologies used to address code clone. The difference in terminologies happens because of the similarity divergent and variety tolerance level allowed permitted for the source code that had been cloned1. The most usually used phrase for clone is classified into 4 types2. Type-1 refers to the code clone that is exact to each other. The source code does not have any alterations with exclusion to the white space and comments that occur with the source file. Type-2 is code clone that has identical copy that contains modifications to the access modifiers, identifiers of the function, variable and type. Type-3 is a code clone with additional alterations from Type-2. Additional alteration involves source code that were altered, added, or removed. Type-4 refers to code clones altered syntactically.

Intentionally or unintentionally done, code clone is a contributing factor that affect maintenance of software3. As an example, if a code being copied contains a bug, there is possibility that the other code clone contains the same bug. Hence, this causes the increase of maintenance effort not only due to the existing bug propagation in the existing source code but also because of the new bugs that may happen if the structure of copied source code is recycled without making any alterations. This is same case with malware14-16. Apart from introducing new bugs through copying the source code, there are also other residing affects to the software due to code clone. The impacts includes difficulty in improvising the system and increasing the resource of the system7.

1Email Address: fahim@ump.edu.my