



INVESTIGANDO A EFETIVIDADE DE FERRAMENTAS DE GERAÇÃO AUTOMÁTICA DE CASOS DE TESTE.

Beatriz Bezerra de Souza¹, Patrícia Duarte de Lima Machado²

RESUMO

Ferramentas para geração automática de casos de teste tem sido ativamente desenvolvidas. Porém, resultados de estudos mostram que a efetividade dos conjuntos de teste gerados pode não ser satisfatória. Este artigo compara coleções de testes automaticamente geradas com coleções de testes manualmente escritas por desenvolvedores. O estudo empírico deste artigo utiliza 10 programas Java, totalizando 119236 linhas de código, que possuem coleções de testes manuais e aplica duas sofisticadas ferramentas para geração automática de casos de teste: Randoop e EvoSuite. Para avaliar a qualidade das coleções de testes, nós utilizamos cobertura de linha e de mutação. Os resultados mostram que há uma diferença estatística significativa entre as coleções de testes manuais e as automaticamente geradas em relação às coberturas de linha e de mutação. Além do mais, há uma diferença estatística significativa entre as coleções de testes geradas por EvoSuite e Randoop em relação à cobertura de mutação. Os resultados também indicam que a quantidade de casos de teste não está correlacionada a efetividade de coleções de testes. Estes resultados indicam que ferramentas de geração automática de casos de teste ainda precisam evoluir para que possam produzir coleções de testes equivalentes as manualmente escritas por desenvolvedores. Além do mais, quantidade de casos de teste pode não indicar a efetividade de coleções de testes.

Palavras-chave: Coleções de teste , Geração automática, Efetividade.

¹Graduanda em Ciência da Computação, Departamento de Sistemas e Computação, UFCG, Campina Grande, PB, e-mail: beatriz.souza@ccc.ufcg.edu.br

²PhD, Professora, Departamento de Sistemas e Computação, UFCG, Campina Grande, PB, e-mail: patricia@computacao.ufcg.edu.br



INVESTIGATING THE EFFECTIVENESS OF TOOLS THAT AUTOMATICALLY GENERATE TEST CASES.

ABSTRACT

Tools that automatically generate test cases have been actively developed. However, some studies show that the effectiveness of automatically generated test suites may not be satisfactory. This paper compares automatically generated test suites to developer-manually written test suites. The empirical study of this paper uses 10 Java programs, a total of 119236 lines of code, that already have manually-written test suites and applies two sophisticated tools that automatically generate test cases: Randoop and EvoSuite. To evaluate the test suites' quality, we used line and mutation coverage. The results show that there is a statistical significant difference between manually-written test suites and automatically generated test suites in relation to both line and mutation coverage. Moreover, there is a statistical significant difference between test suites generated by EvoSuite and Randoop in relation to mutation coverage. The results also indicate that the amount of test cases is not correlated to test suites' effectiveness. These results indicate that tools that automatically generate test cases have to be improved in order to produce test suites equivalent to manually-written test suites. Moreover, the amount of test cases may not indicate test suites' effectiveness.

Keywords: Test suites, Automatic generation, Effectiveness.