

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES  
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

GUSTAVO DA MOTA RAMOS

**Seleção entre estratégias de geração automática de dados de teste por meio  
de métricas estáticas de softwares orientados a objetos**

São Paulo

2018

GUSTAVO DA MOTA RAMOS

**Seleção entre estratégias de geração automática de dados de teste por meio  
de métricas estáticas de softwares orientados a objetos**

Versão original

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Orientador: Prof. Dr. Marcelo Medeiros Eler

São Paulo

2018

Ficha catalográfica

Dissertação de autoria de Gustavo da Mota Ramos, sob o título **“Seleção entre estratégias de geração automática de dados de teste por meio de métricas estáticas de softwares orientados a objetos”**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_ pela comissão julgadora constituída pelos doutores:

\_\_\_\_\_  
**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Presidente

\_\_\_\_\_  
**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

\_\_\_\_\_  
**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

\_\_\_\_\_  
**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

*Aos meus pais, Nivaldo e Izildinha, que não mediram esforços para que eu chegasse até aqui. Eles são responsáveis pela maior herança da minha vida: meus estudos.*

## Resumo

RAMOS, Gustavo da Mota. **Título do trabalho:** Seleção entre estratégias de geração automática de dados de teste por meio de métricas estáticas de softwares orientados a objetos. 2018. 75p f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2018.

Produtos de software com diferentes complexidade são criados diariamente através da elicitacão de demandas complexas e variadas juntamente a prazos restritos. Enquanto estes surgem, altos níveis de qualidade são esperados para tais, ou seja, enquanto os produtos tornam-se mais complexos, o nível de qualidade pode não ser aceitável enquanto o tempo hábil para testes não acompanha a complexidade. Desta maneira, o teste de software e a geração automática de dados de testes surgem com o intuito de entregar produtos contendo altos níveis de qualidade através de baixos custos e atividades rápidas de teste. Porém, neste contexto, os profissionais de desenvolvimento dependem das estratégias de geração automáticas de testes e principalmente da seleção da técnica mais adequada para conseguir maior cobertura de código possível, este é um fator importante dados que cada técnica de geração de dados de teste possuem particularidades e problemas que fazem seu uso melhor em determinados tipos de software. A partir desde cenário, o presente trabalho propõe a seleção da técnica adequada para cada classe de um software com base em suas características, expressas por meio de métricas de softwares orientados a objetos a partir do algoritmo de classificação naive bayes. Inicialmente foi realizado uma revisão bibliográfica dos dois algoritmos de geração estudos, algoritmo de busca randômico e algoritmo de busca genético, compreendendo assim suas vantagens e desvantagens tanto de implementação como de execução. As métricas CK também foram estudadas com o intuito de compreender como estas podem descrever melhor as características de uma classe. O conhecimento adquirido possibilitou coletar os dados de geração de testes de cada classe como cobertura de código e tempo de geração a partir de cada técnica e também as métricas CK, permitindo assim a análise destes dados em conjunto e por fim execução do algoritmo de classificação. Os resultados desta análise demonstraram que um conjunto reduzido e selecionado de métricas é mais eficiente e descreve melhor as características de uma classe além de demonstrarem que as métricas CK possuem pouca ou nenhuma influência no tempo de geração dos dados de teste e no algoritmo de busca randômico. Entretanto, as métricas CK demonstraram média correlação e influência na seleção do algoritmo genético, participando assim na sua seleção pelo algoritmo naive bayes.

Palavras-chaves: Métricas CK. Teste de software. Geração de testes. Cobertura de testes. Naive bayes. Algoritmo genético.

## Abstract

RAMOS, Gustavo da Mota. **Work title:** Selection between whole test generation strategies by analysing object oriented software static metrics . 2018. 75 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, DefenseYear.

Software products with different complexity are created daily through analysis of complex and varied demands together with tight deadlines. While these arise, high levels of quality are expected for such, as products become more complex, the quality level may not be acceptable while the timing for testing does not keep up with complexity. In this way, software testing and automatic generation of test data arise in order to deliver products containing high levels of quality through low cost and rapid test activities. However, in this context, software developers depend on the strategies of automatic generation of tests and especially on the selection of the most adequate technique to obtain greater code coverage possible, this is an important factor given that each technique of data generation of test have peculiarities and problems that make its use better in certain types of software. From this scenario, the present work proposes the selection of the appropriate technique for each class of software based on its characteristics, expressed through object oriented software metrics from the naive bayes classification algorithm. Initially, a literature review of the two generation algorithms was carried out, random search algorithm and genetic search algorithm, thus understanding its advantages and disadvantages in both implementation and execution. The CK metrics have also been studied in order to understand how they can better describe the characteristics of a class. The acquired knowledge allowed to collect the generation data of tests of each class as code coverage and generation time from each technique and also the CK metrics, thus allowing the analysis of these data together and finally execution of the classification algorithm. The results of this analysis demonstrated that a reduced and selected set of metrics is more efficient and better describes the characteristics of a class besides demonstrating that the CK metrics have little or no influence on the generation time of the test data and on the random search algorithm . However, the CK metrics showed a medium correlation and influence in the selection of the genetic algorithm, thus participating in its selection by the algorithm naive bayes.

Keywords: CK metrics. Software testing. Test data generation. Code coverages. Naive bayes. Genetic algorithm.

## Lista de figuras



## Lista de algoritmos

## Lista de quadros

## Lista de tabelas

## Lista de abreviaturas e siglas

CBO	<i>Coupling between objects</i>
CK	<i>Chidamber &amp; Kemerer</i>
CUT	<i>Class under test</i>
DIT	<i>Depth inheritance tree</i>
LCOM	<i>Lack of cohesion of methods</i>
LOC	<i>Lines of code</i>
NOC	<i>Number of children</i>
NOF	<i>Number of fields</i>
NOM	<i>Number of methods</i>
NOPF	<i>Number of public fields</i>
NOPM	<i>Number of public methods</i>
RFC	<i>Response for a Class</i>
SUT	<i>Software under test</i>
WMC	<i>Weight method class</i>

## Sumário

<b>1</b>	<b>Introdução . . . . .</b>	<b>13</b>
<b>2</b>	<b>Conclusão . . . . .</b>	<b>14</b>
2.1	<i>Uma seção secundária . . . . .</i>	14
2.1.1	Uma seção terciária . . . . .	14
2.1.2	Outra seção terciária . . . . .	15
2.1.3	Mais uma seção terciária . . . . .	15
2.2	<i>Outra seção secundária . . . . .</i>	15
2.3	<i>Mais uma seção secundária . . . . .</i>	16
	<b>Anexo A – Resumo das normas . . . . .</b>	<b>17</b>

## 1 Introdução

Este *template* apresenta as regras básicas para a elaboração do trabalho segundo as normas ABNT. Estas regras devem ser seguidas rigorosamente a fim de que o mesmo possa receber sua ficha catalográfica e ser posteriormente aprovado para publicação na Biblioteca Digital de Teses e Dissertações (BDTD) da USP. Qualquer desvio realizado nas configurações e recomendações deste *template* poderá causar atrasos nesses processos, uma vez que o texto precisará ser corrigido antes.







