



VALID 2023



AN EXPLORATION OF MAVEN-BASED JAVA REPOSITORIES AND THEIR TESTING PRACTICES

Canol Simsek, Tugkan Tuglular

Presenter: Tugkan Tuglular
tugkantuglar@iyte.edu.tr
Izmir Institute of Technology

ASSOC. PROF. TUGKAN TUGLULAR, PH.D.

IZMIR INSTITUTE OF TECHNOLOGY, TURKEY

Tugkan Tuglular received the B.S., M.S., and Ph.D. degrees in Computer Engineering from Ege University, Turkey, in 1993, 1995, and 1999.

He worked as a research associate at Purdue University from 1996 to 1998.

He has been with Izmir Institute of Technology since 2000.

After becoming an Assistant Professor at Izmir Institute of Technology, he worked as Chief Information Officer in the university from 2003-2007.

In addition to his academic duties, he acted as IT advisor to the Rector between 2010-2014.

In 2018, he became an Associate Professor in the Department of Computer Engineering of the same university.

He has more than 75 publications and an active record of duties with international and national conferences.

His current research interests include model-based testing and software quality with machine learning support.

OBJECTIVES



To collect and analyze *Maven*-based Java repositories on GitHub



To compare test class & method counts with total class & method counts

THIS RESEARCH AIMED TO FIND ANSWERS TO THESE TWO QUESTIONS:

How does the test usage change with the size of the repositories in terms of test classes and overall classes?

How does the test usage change with the size of the repositories in terms of the test methods and overall methods?



INTRODUCTION

- The research aims to reach test usage data on Java language of two different topics in GitHub.
- Two topics selected as Gaming and E-Commerce.
- After cloning, projects are scanned with SonarQube and a Python Script for internal data.
- We used this data to compare the test usage rate between projects.

COLLECTION OF JAVA REPOSITORIES VIA GITHUB API

Our aim was to explore, aggregate, and analyze the top gaming and e-commerce repositories that are written in Java on GitHub.



Our aim also included to find repositories that used Maven. So, we tried to find and clone that contained specifically pom.xml file.

INCORPORATING SONARQUBE FOR MORE INSIGHTS



SonarQube, a tool known for its code quality analysis, was used to gain deeper insights.



It provides a systematic approach to quantify project size, allowing us to understand the scale of these projects better and gather critical data related to the complexity and structure of the codebases.

EXTRACTION OF CLASS AND METHOD DATA

- We have used two methods to gather overall data from cloned repositories.
- First method is SonarQube for gathering overall code analysis like size, code smells and coverage metrics.
- The second method is our python script for gathering test-usage data.

ANALYSIS PARAMETERS

The gathered parameters from SonarQube and our script are as follows:

Class count : Number of classes written in the project (without test classes).

Test Class count : Number of test classes written in the project.

Method count : Number of methods written in the project (without test methods).

Test Method count : Number of test methods written in the project.

GitHub Projects	Class count	Test class count	Method count	Test method count
ezyfox-server	663	337	2553	943
base	785	314	3742	4256
Open Realm of Stars	302	190	3471	1245
jeveassets	1024	76	8987	743
mmo	184	14	1633	44

ANALYSIS & FINDINGS - GAMING

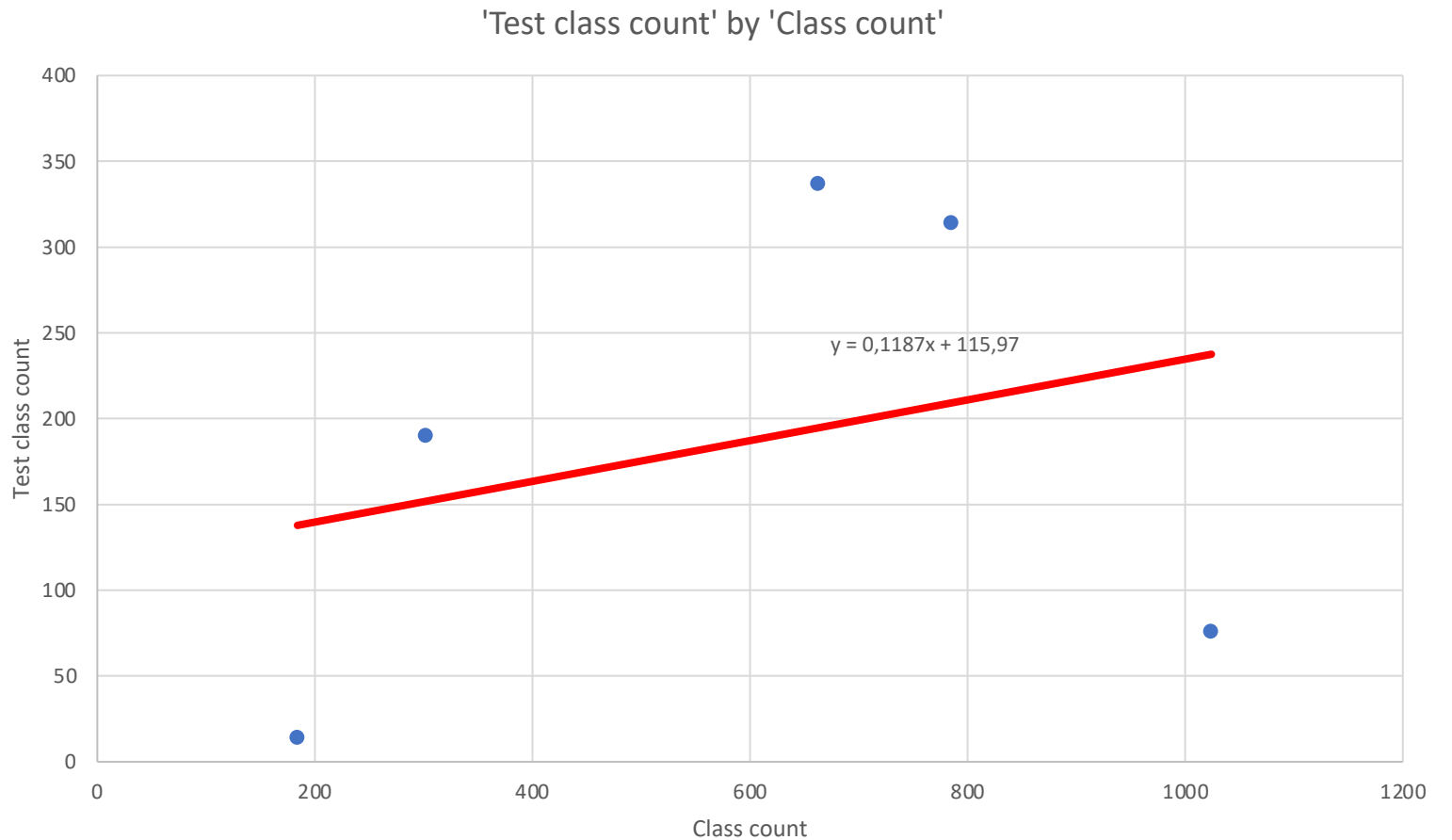
CLASS AND TEST CLASS COUNTS COMPARED TO METHOD AND TEST METHOD COUNTS FOR GAMING TOPICS

GitHub Projects	Class count	Test class count	Method count	Test method count
io.spot-next:spot-framework	430	11	2025	25
microservices-event-sourcing/parent	203	9	479	16
IOM Test Framework	337	6	1667	20
productsv	27	6	73	12
[Tool] IOM Project Bootstrap Archetype	16	4	56	6

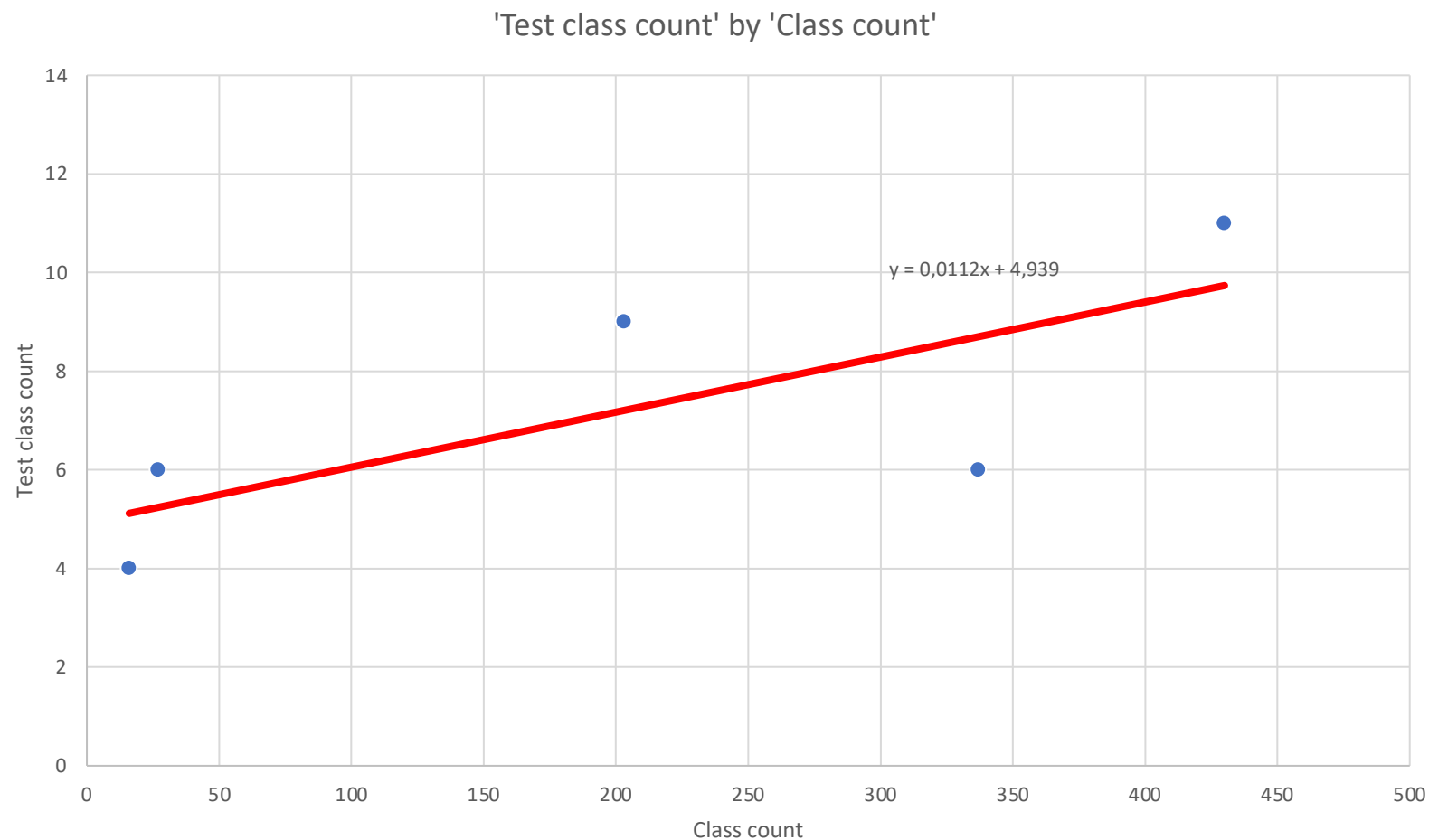
ANALYSIS & FINDINGS – E-COMMERCE

CLASS AND TEST CLASS COUNTS COMPARED TO METHOD AND TEST METHOD COUNTS FOR E-COMMERCE TOPICS

Scatter plots of total class count vs test class count for gaming topic



Scatter plots of total class count vs test class count for e-commerce topic



ANSWER TO QUESTION 1:

How does the test usage change with the size of the repositories in terms of test classes and overall classes?



Every Java repository has different characteristics.

With the Gaming and E-Commerce topics, test usage changes from topic to topic and even between projects.

While gaming repositories are much larger than e-commerce repositories, their test usage is unsatisfactory.

On the other hand, in e-commerce topic, results are slightly better than gaming.

Their repository sizes are small.

However, they have a better class ratio than gaming topics.

So, test usage is expected to have a positive relationship with the class counts in the source code.

ANSWER TO QUESTION 2:

How does the test usage change with the size of the repositories in terms of the test methods and overall methods?



Method counts and test method usage depend on the software projects' characteristics.

For two topics, methods are like class counts in the gaming topic, and test usage is the same as expected.

In the e-commerce topic, classes have fewer methods than gaming topic, and test classes have fewer test methods than gaming topic.



THREATS TO VALIDITY

The results of this research cannot be generalized neither to the domains under consideration nor to other domains due to the following reasons:

- the study uses a small sample size, and the results are not statistically significant enough to represent the larger population.
- the study sample might not be representative of the whole population.
- the study uses a non-random sampling method, which can introduce bias.
- the study uses a tool, developed by the authors, that might have some inaccuracies or limitations.

CONCLUSION

- Recap of the objectives & methods
 - We searched useable repositories and filtered them for our needs to examine the relationship of test classes and methods density with overall code sizes in two topics.
- Brief on key findings
 - Despite there might be a level of relationship, java projects may differ quite a lot between each other. Even in the same topic, test usage changes from project to project.

CONCLUSION

For the existing research; in the Gaming topic, there are significantly more usage of test when we compare it to e-commerce topic.

Note that our data is quite limited. With more data examined and expanding code language options with version flexibility, results might differ.