

Project Nominal Information

This dataset can be used to built better machine learning classifiers for God Class Design Smell detection. The dataset is formed by the 12,588 classes of 24 open source systems written in Java obtained from SourceForge source code repository. The projects in the dataset can be classified into three main categories regarding the project size, domain, and status. A source code analyzer tool was used to compute a set of important metrics which include several software package-level and class-level metrics regarding complexity, coupling, cohesion, etc.. Five different design smell detection tools were used to detect God Class in each system. The systems were selected as randomly as possible from the repository.

DATASET DESCRIPTION

Number of Projects: 24.

Number of Classes: 12,588.

Number of God CLasses: 1,958.

Source Code analyzer tool: RefactorIt v2.7.0.

Number of Metrics: 16.

God Class Design Smell detection tools:

* PMD v5.3.2 * iPlasma v6. * Décor v1.0 * JDeodorant v5.0.13 * Borland Together v12.6.

DATASET FORMAT

1. Project name.
2. Project version.
3. Project domain.
4. Project Status.
5. Project size.
6. Package name.
7. Class name.
8. Full name.
9. Class stereotype.
10. LOC. Total Lines of Code
11. NCLOC. Non-Comment Lines of Code
12. CLOC. Comment Lines of Code
13. EXEC. Executable Statements
14. DC. Density of Comments
15. NOT. Number of Types
16. NOTa. Number of Abstract Types
17. NOTc. Number of Concrete Types
18. NOTe. Number of Exported Types
19. RFC. Response for Class
20. WMC. Weighted Methods per Class
21. DIT. Depth in Tree
22. NOC. Number of Children in Tree
23. DIP. Dependency Inversion Principle
24. LCOM. Lack of Cohesion of Methods
25. NOA. Number of Attributes
26. Output of Borland Together tool (Binary value).

27. Output of iPlasma tool (Binary value).
28. Output of JDeodorant tool (Binary value).
29. Output of PMD tool (Binary value).
30. Output of DECOR tool (Binary value).
31. Total output of all tools (Binary value).

DATASET CATEGORIES

1. The size:
 - Large.
 - Medium-Large.
 - Medium.
 - Small-Medium.
 - Small.
2. The domain:
 - Application Software.
 - Software Development.
 - Client Server.
 - Diagram Generator/ Data visualization.
3. The status:
 - Production/Stable.
 - Beta.
 - Mature.
 - Multi.

REFERENCE

This Dataset is related to the paper "An Exploratory Study of the Impact of Project Domain, Status and Size on the Detection of the God Class Design Smell" submitted to the Journal of Software evolution and process (**This paper is under review**).

```
@conference{KhalidKsem16,  
author = {Khalid Alkharabsheh, Yania Crespo, José R. Viqueira, José A. Taboada},  
title = {An Exploratory Study of the Impact of Project Domain, Status and Size on the Detection of the God Class  
Design Smell},  
booktitle = {Under review in Journal of Software evolution and process},  
year = 2017,  
editor = {XXX},  
volume = 000,  
series = 000,  
pages = 000,  
address = {},  
month = 3,  
organization = {XXX},  
publisher = {Springer},  
}
```



LICENSE

This information is under the license [Creative Commons Reconocimiento-Compartir Igual 4.0 Internacional](#). You can use this dataset on your publication as long as you include a citation to the reference on this [page](#). When including a link to this dataset, please use this page instead of linking the file directly.

INFORMACIÓN

Investigadores
Khalid Alkharabsheh
José Ángel Taboada González
Yania Crespo

DESCARGAR

-  Repositorio Gitlab
-  Descargar de Gitlab

PUBLICACIONES

Assessing the Influence of Size Category of the Project in God Class Detection, an Experimental Approach based on...
International Conference on Software Engineering & Knowledge Engineering, 2019