

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 two sets: Set A is formed by the 12,588 classes of 24 open source systems written in Java obtained from SourceForge source code repository. Set B includes 18,441 classes of the original available in <https://figshare.com/s/5da162e21b8d54fbfce81377> by selecting a version of each of its 13 projects. A total of 16 metrics are calculated from the A and B datasets using the RefactorIt tool and manually classified into two categories according to the size and domain of the project. Both sets are analyzed to identify their god classes. The dataset A is classified semi-automatically by 5 tools and the second one, dataset B, has its original manual evaluation.

DATASET DESCRIPTION

Dataset A

Number of Projects: 24.

Number of Classes: 12,588.

Number of God CLasses: 1,958.

God Class Design Smell detection tools:

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

Dataset B

Number of Projects: 13.

Number of Classes: 18,441.

Number of God CLasses: 95.

God Class Design Smell detection: Manual

DATASET FORMAT

1. Project domain.
2. Project size.
3. RFC. Response for Class
4. WMC. Weighted Methods per Class
5. DIT. Depth in Tree
6. NOC. Number of Children in Tree
7. DIP. Dependency Inversion Principle
8. LCOM. Lack of Cohesion of Methods
9. NOA. Number of Attributes
10. NOT. Number of Types
11. NOTa. Number of Abstract Types
12. NOTc. Number of Concrete Types
13. NOTe. Number of Exported Types
14. LOC. Total Lines of Code
15. NCLOC. Non-Comment Lines of Code
16. CLOC. Comment Lines of Code
17. EXEC. Executable Statements
18. DC. Density of Comments

Dataset A

- 19. Output of Borland Together tool (Binary value).
- 20. Output of iPlasma tool (Binary value).
- 21. Output of JDeodorant tool (Binary value).
- 22. Output of PMD tool (Binary value).
- 23. Output of DECOR tool (Binary value).
- 24. Total output of all tools (Binary value).

Dataset B

- 19. Manual God Class Identification (Binary value)

DATASET CATEGORIES

1. The size:

- 6: Very Large
- 5: Large.
- 4: Medium-Large.
- 3: Medium.
- 2: Small-Medium.
- 1: Small.

2. The domain:

- 1: Application Software.
- 2: Software Development.
- 3: Client Server.
- 4: Diagram Generator/ Data visualization.

REFERENCE

This Dataset is related to the paper "Exploratory Study of the Impact of Project Domain and Size Category on the Detection of the God Class Design Smell" submitted to Software Quality Journal (This paper is under review).

```
@article{Khalid20,  
  author = {Khalid Alkharabsheh, Yania Crespo, Manuel Fernández-Delgado, José R. Viqueira, José A. Taboada},  
  title = {Exploratory Study of the Impact of Project Domain and Size Category on the Detection of the God Class  
  Design Smell},  
  journal = {Under review in Software Quality Journal},  
  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
José Ángel Taboada González
Khalid Alkharabsheh
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