

# CaM: Classes and Metrics

YEGOR256/CAM

This report was generated automatically by the script, which is located in the GitHub repository [yegor256/cam](https://github.com/yegor256/cam).

In total, 1000 were found and retrieved from GitHub. The following search criteria have been used: at least 400, at most 10000 stars, at least 100Kb size of Git repo. The exact query string for GitHub API was the following: “stars:400..10000 size:>=100 language:java is:public mirror:false archived:false NOT android”. The full list of them is in the `repositories.csv` file. The `hashes.csv` file has a list of Git hashes of their latest commits. The filtering process was the following:

1. There were 3589803 files total, 1310217 of them were `.java` files; all other files, which were not `.java`, have been deleted: 2279586 total;
2. There were 14713 files named as `package-info.java`, all of them were deleted;
3. There were 1295504 files total; 405052 of them were test files with `Test` or `ITCase` suffixes and that's why they were deleted;
4. There were 1042584 files total; 14314 of them were Java files with broken syntax and that's why were deleted;
5. There were 1028270 files total; 2769 of them had at least one line longer than 1024 characters, which most probably is a symptom of an auto-generated code; that's why they all were deleted;
6. There were 1025501 files total. 150563 of them were not Java classes but interfaces or enums, that's why they were deleted;
7. There were 874938 files total; 12433 of them had more than 1 Java class, that's why were deleted;
8. There were 6883 symlinks total; all of them were deleted;
9. There were 579781 empty directories total; all of them were deleted;

The structure of the archive is the following:

- `data/` is the collection of `.csv` files with the data (34363 total);
- `github/` is the entire collection of 862517 source files “as is” (8.7G total).

There are 33 metrics:

- `blanks`: Blank Lines
- `comments`: Commenting Lines
- `loc`: Lines Of Code
- `attrs`: Number of Non-Static Attributes
- `sattrs`: Number of Static Attributes

- `ctors`: Number of Constructors
- `mtds`: Number of Non-Static Methods
- `smtds`: Number of Static Methods
- `ncss`: Non-Commenting Source Statements (NCSS)
- `impls`: Number of Implemented Interfaces
- `extnds`: Number of Extended Classes
- `gnrcs`: Number of Type Parameters (generics)
- `final`: Class is Final
- `annts`: Number of Annotations
- `coco`: Total Cognitive Complexity of All Methods
- `acoco`: Average Cognitive Complexity of a Method
- `mxco`: Max Cognitive Complexity of a Method
- `mncoco`: Min Cognitive Complexity of a Method
- `cc`: Cyclomatic Complexity
- `MMAC`: Method-Method through Attributes Cohesion.
- `CAMC`: Cohesion Among Methods in Class
- `NHD`: Normalized Hamming Distance
- `LCOM5`: Revision of the initial LCOM metric
- `SCOM`: Sensitive Class Cohesion Metric
- `MMAC-cvc`: Same as MMAC, but constructors are excluded
- `CAMC-cvc`: Same as CAMC, but constructors are excluded
- `NHD-cvc`: Same as NHD, but constructors are excluded
- `LCOM5-cvc`: Same as LCOM5, but constructors are excluded
- `SCOM-cvc`: Same as SCOM, but constructors are excluded
- `hsdif`: Halstead Difficulty
- `hsef`: Halstead Effort
- `hsvol`: Halstead Volume
- `midx`: Maintainability Index

The dataset was built by 12 CPUs, in 233h16m.

If any questions or suggestions, please [email](#).