

CaM: Classes and Metrics

YEGOR256/CAM

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

We found and retrieved 959 repositories from GitHub.

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 3483589 files total, 1284687 of them were `.java` files; all other files, which were not `.java`, have been deleted: 2198902 total;
2. There were 13160 files named as `package-info.java`, all of them were deleted;
3. There were 1271527 files total; 400154 of them were test files with `Test` or `ITCase` suffixes and that's why they were deleted;
4. There were 1022012 files total; 13406 of them were Java files with broken syntax and that's why were deleted;
5. There were 1008606 files total; 2710 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 1005896 files total. 149472 of them were not Java classes but interfaces or enums, that's why they were deleted;
7. There were 856424 files total; 12419 of them had more than 1 Java class, that's why were deleted;
8. There were 6694 symlinks total; all of them were deleted;
9. There were 564803 empty directories total; all of them were deleted;

The structure of the archive:

- `data/` is the collection of `.csv` files with the data you are looking for (56218 total);
- `github/` is the entire collection of 844083 source files "as is" (9.7G total);
- `measurements/` is where we keep individual metrics calculated per file, which you may use to verify our algorithms;

There are 29 metrics that we use:

- `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
- `coc`: Total Cognitive Complexity of All Methods
- `acoc`: Average Cognitive Complexity of a Method
- `mxcoc`: Max Cognitive Complexity of a Method
- `mncoc`: 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

It took -471269619232098.44 hours to generate this report.

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