

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 repositories were found and retrieved 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:

- 1,513,253 files out of 2,335,845 without the `.java` extension were deleted;
- 8,832 files named as `package-info.java` were deleted;
- 1,107 files named as `module-info.java` were deleted;
- 247,554 files out of 812,653 with `Test` or `ITCase` suffixes were deleted;
- 14,368 files out of 651,496 with an unparseable Java syntax were deleted;
- 348 files out of 637,128 with at least one line longer than 1,024 characters, which most probably is a symptom of an auto-generated code, were deleted;
- 97,891 files out of 636,780 with interfaces or enums (instead of classes) inside were deleted;
- 6,539 files out of 538,889 with more than one Java class inside were deleted;
- 4,417 symlinks were deleted;
- 379,432 empty directories were deleted;

The structure of the archive is the following:

- `data/` is the collection of `.csv` files with the data (43854 total);
- `github/` is the entire collection of 532394 source files "as is" (77G total).

There are 48 metrics:

- `acoco`: Average Cognitive Complexity of a Method
- `ahf`: Method Attribute Factor (AHF)
- `CAMC`: Cohesion Among Methods in Class
- `CAMC-cvc`: Same as CAMC, but constructors are excluded
- `cc`: Total Cyclomatic Complexity of all methods
- `coco`: Total Cognitive Complexity of All Methods
- `final`: Class is Final
- `fout`: Fan-Out
- `hsd`: Halstead Difficulty
- `hse`: Halstead Effort

- `hsv`: Halstead Volume
- `LCOM5-cvc`: Same as LCOM5, but constructors are excluded
- `LCOM5`: Revision of the initial LCOM metric
- `loc`: Lines Of Code
- `mhf`: Method Hiding Factor (MHF)
- `midx`: Maintainability Index
- `MMAC-cvc`: Same as MMAC, but constructors are excluded
- `MMAC`: Method-Method through Attributes Cohesion.
- `mncoco`: Min Cognitive Complexity of a Method
- `mx coco`: Max Cognitive Complexity of a Method
- `mxnomp`: Maximum of Method Parameters (MxNOMP)
- `mxnosmp`: Maximum of Static Method Parameters (MxNOSMP)
- `napc`: Number of Ancestor (Parent) Classes
- `ncss`: Non-Commenting Source Statements (NCSS)
- `NHD-cvc`: Same as NHD, but constructors are excluded
- `NHD`: Normalized Hamming Distance
- `nobl`: Number of Blank Lines
- `noca`: Number of Class Annotations
- `nocc`: Number of Class Constructors
- `nocl`: Number of Commenting Lines
- `nocm`: Number of Static (Class) Methods
- `noga`: Number of Git Authors (Different Committers)
- `noii`: Number of Implemented Interfaces
- `nom`: Number of Overriding Methods (NOM)
- `nomp`: Number of Method Parameters (NOMP)
- `nooa`: Number of Non-Static (Object) Attributes
- `noom`: Number of Non-Static (Object) Methods
- `nop`: Number of Polymorphic Methods (NOP)
- `nosa`: Number of Static Attributes
- `nosmp`: Number of Static Method Parameters (NOSMP)
- `notp`: Number of Type Parameters (Generics)
- `nulls`: Number of NULL References
- `raf`: Relative Age of File (To repository existence)
- `sahf`: Static Attribute Hiding Factor (SAHF)
- `SCOM-cvc`: Same as SCOM, but constructors are excluded
- `SCOM`: Sensitive Class Cohesion Metric
- `smhf`: Static Method Hiding Factor (MHF)
- `varcomp`: Average number of parts in variable names

The dataset was built by 8 CPUs, in 24h5m.

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