Qualitas.class Corpus: A Compiled Version of the Qualitas Corpus

Ricardo Terra, Luis Fernando Miranda, Marco Tulio Valente, Roberto S. Bigonha

Department of Computer Science, UFMG, Brazil

{terra,luisfmiranda,mtov,bigonha}@dcc.ufmg.br

Abstract

This paper documents a compiled version of the Qualitas Corpus named *Qualitas.class Corpus*. We provide compiled Java projects for the 111 systems included in the corpus. We also gathered a large amount of metrics data (such as measurements from complexity, coupling, and CK metrics) about the systems. By making *Qualitas.class Corpus* public, our goal is to assist researchers by removing the compilation effort when conducting empirical studies.

1 Introduction

Qualitas.class Corpus is a compiled version of the Qualitas Corpus proposed by Tempero et al. [6]. In short, it provides compiled Eclipse Java projects for the 111 systems included in the last release of the corpus.[1] Table 1 provides an overview on our compiled corpus. It contains more than 18 million LOC, 200,000 compiled classes, and 1.5 million compiled methods. As another contribution, we have gathered a large amount of metrics data (such as measurements from size, coupling, and CK metrics) about the systems (see Section 3).

Table 1: Overall Total

Systems	111
Lines of Code (LOC)	18,548,026
Internal Projects (NOIP)	802
Packages (NOP)	16,509
Classes (NOCL)	202,052
Interfaces (NOI)	22,115
Methods (NOM)	1,464,893

The original Qualitas Corpus provides a huge contribution for experimentation in software engineering. However, there are several scenarios—e.g., experiments that rely on Abstract Syntax Trees (AST) or bytecode—in which researchers need to import and compile the source code. Since this task is not trivial in the case of systems with many external dependencies, this work addresses such effort by providing a compiled variant of the Qualitas Corpus.

2 Compilation Process

The corpus contains a collection of **systems**, each of which contains *one or more* **projects**. For instance, the AspectJ system is divided in four internal projects: `matcher`, `rt`, `tools`, and `weaver`. Thus, considering the 111 systems, the *Qualitas.class Corpus* has a total of 802 internal projects (including 461 projects from NetBeans).

Since the Qualitas Corpus provides us with the source code of the systems, our main work consisted in organizing the code in well-defined Java projects according to the following guidelines:

- In the case of outdated systems, we compiled their most recent version, instead of thoses in the Qualitas Corpus.
- Source code distributions usually do not include third-party libraries, which are required by the compilation process. We hence searched for these libraries using Maven repositories[2] or version control systems.
- Some projects rely on very particular libraries. For instance, JTOpen 7.8 relies on the IBM AS/400 library, which is not publicly available. In these cases, we created *stub* JAR files to simulate the internal structure of the libraries (e.g., `ibm_as400_stub.jar`).
- Some projects presented compilation errors. In these cases, we fixed the error and explained our fixing procedure in a comment.[3] For instance, package `etc.testcases` on Ant 1.8.2 has some classes whose file name differs from the public class name. Therefore, we renamed the class names and inserted comments into the code to explain the changes.

3 Measurements using Metrics

Qualitas.class Corpus also includes the values of the following 23 source code metrics measured at the level of classes [1, 4, 3, 5]:

- *Basic Metrics*: Number of lines of code (LOC)[4], Number of packages (NOP), Number of classes (NOCL), Number of interfaces (NOI), Number of methods (NOM), Number of attributes (NOA), Number of overridden methods (NORM), Number of parameters (PAR), Number of static methods (NSM), and Number of static attributes (NSA).
- *Complexity Metrics*: Method lines of code (MLOC), Specialization index (SIX), McCabe cyclomatic complexity (VG), Nested block depth (NBD), and Normalized distance (RMD).

[1] The current release is 20120401

[2] Maven repositories: search.maven.org and mvnrepository.com

[3] Comments like `//Qualitas.class:...` to be easily identified.

[4] Our metric counts non-blank and non-comment lines of codes.

- *CK Metrics*: Weighted methods per class (WMC), Depth of inheritance tree (DIT), Number of children (NOC), and Lack of cohesion in methods (LCOM).[5]

- *Coupling Metrics*: Afferent coupling (CA), Efferent coupling (CE), Instability (I), and Abstractness (A).

Table 2 presents a subset of the metrics gathered for the systems in the corpus. As can be noticed, the corpus is very heterogeneous. For example, systems' size ranges from 3.5 KLOC (fitjava) to 2,500 KLOC (Eclipse). There are lowly- (e.g., jasml, LCOM 0.08) and highly-cohesive systems (e.g., freecs, LCOM 0.57). Analogously, there are lowly- (e.g., xmojo, CE 0.6) and highly-coupled systems (e.g., megamek, CE 38).

We relied on Google CodePro Analytix[6] and Metrics[7] to compute the metrics. For each project P, the *Qualitas.class Corpus* provides a `XML` file with a `Metric` element for each metric M (identified by the attribute `id`). For example, the element <`Metric id` = ″`NOM`″ `avg` = ″`4.04`″ ... > contains the average value of the metric *Number of Methods*.

In a summarized perspective, Figure 1 illustrates the distribution of the average value for a subset of metrics. Basically, each circle represents a system and the values in the vertical indicate the overall average for each metric. For example, the MLOC metric ranges from 3.35 (fitlibraryforfitnesse) to 23.4 (jparse), but the overall average is 7.88 ± 2.7.

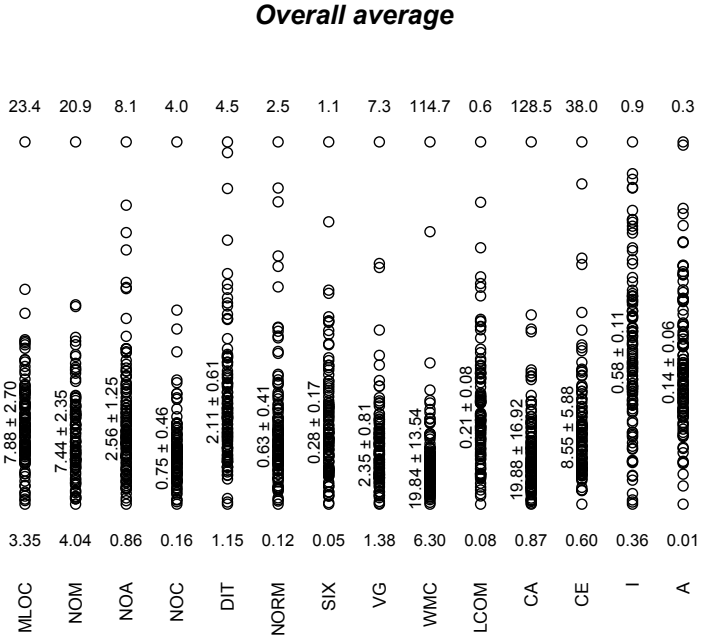

Figure 1: Distribution of the systems (average)

[5]Our metric relies on the LCOM HS (Henderson-Sellers) method [3].
[6]CodePro Analytix 7.1.0, http://developers.google.com/java-dev-tools
[7]Metrics 1.3.8, http://metrics2.sourceforge.net

4 Final Remarks

The original Qualitas Corpus provides the source code of a large curated collection of open source Java systems. However, there is a broad spectrum of scenarios—e.g., experiments that rely on Abstract Syntax Tree (AST) or bytecode—in which researchers need to import and compile the source code. In practice, this compilation process is not trivial and may impact other researchers' ability to replicate the study. This work therefore addresses such hard-working and time-consuming tasks by providing a compiled version of the original Qualitas Corpus. For instance, we used the *Qualitas.class Corpus* to evaluate a refactoring recommendation system we are working on [8, 7]. As future work, we intend to integrate the *Qualitas.class Corpus* with our previous data set with historical source code metrics values [2].

The *Qualitas.class Corpus* is publicly available at:

http://java.labsoft.dcc.ufmg.br/qualitas.class

Acknowledgments: This research was supported by grants from CNPq, CAPES, and FAPEMIG.

References

[1] S.R. Chidamber and C.F. Kemerer. A metrics suite for object oriented design. *IEEE Transactions on Software Engineering*, 20(6):476–493, 1994.

[2] Cesar Couto, Cristiano Maffort, Rogel Garcia, and Marco Tulio Valente. Comets: A dataset for empirical research on software evolution using source code metrics and time series analysis. *Software Engineering Notes*, pages 1–3, 2013.

[3] Brian Henderson-Sellers. *Object-oriented metrics: measures of complexity*. Prentice-Hall, 1996.

[4] Michele Lanza, Radu Marinescu, and Stéphane Ducasse. *Object-Oriented Metrics in Practice*. Springer-Verlag, 2005.

[5] Robert Martin. OO design quality metrics – an analysis of dependencies. In *Workshop Pragmatic and Theoretical Directions in Object-Oriented Software Metrics, OOPSLA'94*, pages 1–8, 1994.

[6] Ewan Tempero, Craig Anslow, Jens Dietrich, Ted Han, Jing Li, Markus Lumpe, Hayden Melton, and James Noble. The Qualitas Corpus: A curated collection of Java code for empirical studies. In *17th Asia Pacific Software Engineering Conference (APSEC)*, pages 336–345, 2010.

[7] Ricardo Terra, João Brunet, Luis Miranda, Marco Tulio Valente, Dalton Serey, Douglas Castilho, and Roberto Bigonha. Measuring the structural similarity between source code entities. In *25th International Conference on Software Engineering and Knowledge Engineering (SEKE)*, pages 753–758, 2013.

[8] Ricardo Terra, Marco Tulio Valente, Krzysztof Czarnecki, and Roberto Bigonha. Recommending refactorings to reverse software architecture erosion. In *16th European Conference on Software Maintenance and Reengineering (CSMR), Early Research Achievements Track*, pages 335–340, 2012.

Table 2: *Qualitas.class Corpus* (1 of 2)

System	Version*	KLOC	NOP	NOCL	NOI	MLOC	NOM	NOA	NOC	DIT	NORM	SIX	VG	WMC	LCOM	CA	CE	I	A
ant	1.8.2	127.6	127	1627	97	6.30±12.4	7.77±10.3	2.46±4.6	0.83±6.2	2.45±1.4	0.68±1.3	0.33±0.6	2.12±3.0	17.19±28.0	0.25±0.4	14.43±69.7	7.45±14.6	0.81±0.3	0.09±0.1
antlr	3.4	47.4	20	381	20	8.80±13.3	10.14±16.2	2.73±5.5	1.45±10.8	1.90±1.2	0.82±2.3	0.35±0.8	2.05±2.9	21.66±38.4	0.20±0.3	15.60±22.4	8.60±10.8	0.58±0.4	0.07±0.1
aoi	2.8.1	110	23	604	32	12.38±29.0	10.33±11.5	6.27±7.6	0.60±3.0	2.16±1.8	1.67±3.3	0.32±0.6	3.39±6.7	37.34±62.8	0.38±0.3	62.56±93.3	18.56±19.0	0.55±0.4	0.09±0.1
argouml	0.34	105.8	77	1408	107	7.50±15.5	6.01±8.7	1.29±3.1	0.79±4.4	3.03±2.1	0.91±2.2	0.57±1.1	2.32±3.8	14.92±25.6	0.13±0.3	31.78±63.4	13.83±23.8	0.48±0.3	0.14±0.2
aspectj	1.6.9	501.8	144	3600	564	8.40±23.3	11.01±20.8	3.18±12.1	0.85±2.5	1.85±1.5	1.53±5.4	0.42±0.8	3.03±8.6	36.58±100.4	0.25±0.3	37.19±63.9	13.91±20.8	0.41±0.3	0.25±0.3
axion	1.0-M2	24.2	13	257	38	5.03±11.6	11.81±22.1	2.08±6.5	1.02±2.5	1.60±1.0	0.96±2.1	0.21±0.4	2.22±4.4	26.45±89.7	0.20±0.3	24.31±41.6	14.54±11.2	0.53±0.2	0.21±0.2
azureus	4.7.0.2	495.5	473	4038	1077	8.68±26.0	8.15±14.4	2.60±5.7	1.27±10.8	1.20±1.1	0.21±0.7	0.11±0.4	2.36±4.4	20.97±45.3	0.22±0.3	20.56±64.0	4.61±6.5	0.43±0.3	0.28±0.3
batik	1.7	194.7	117	2624	288	7.90±22.1	5.96±8.0	2.77±19.0	0.86±3.8	2.19±2.0	0.58±1.6	0.33±0.8	2.37±5.5	14.90±34.5	0.17±0.3	16.76±38.8	9.65±19.2	0.61±0.4	0.20±0.3
c_jdbc	2.0.2	95.5	145	777	16	10.60±24.1	7.53±15.5	2.50±5.0	0.65±2.9	2.88±1.9	0.43±1.2	0.25±0.7	2.47±4.1	19.45±44.7	0.22±0.3	14.88±28.9	4.26±4.6	0.51±0.4	0.11±0.2
castor	1.3.3*	219.7	381	2803	156	7.01±20.0	7.35±11.4	2.08±4.0	0.54±3.8	2.04±1.3	0.95±2.9	0.37±0.8	2.06±4.5	15.77±46.6	0.27±0.4	7.03±20.4	5.34±6.3	0.76±0.3	0.08±0.2
cayenne	3.0.1	198.2	235	3193	160	6.43±12.4	5.73±8.3	1.32±2.6	0.88±5.7	2.60±1.6	0.48±1.7	0.30±0.7	1.82±2.9	10.81±21.5	0.15±0.3	24.67±81.7	10.02±15.4	0.61±0.3	0.12±0.2
checkstyle	5.6*	36.6	42	553	18	7.40±9.5	5.00±5.0	1.20±2.2	0.76±7.4	2.52±1.5	0.64±1.1	0.44±0.8	1.92±2.4	10.01±13.0	0.14±0.3	13.90±46.6	9.57±9.3	0.88±0.3	0.08±0.2
cobertura	1.9.4.1	54.6	34	160	11	12.18±25.1	20.91±90.1	6.82±19.8	0.32±1.4	1.66±1.1	0.29±0.7	0.14±0.4	5.24±11.3	114.72±533.7	0.25±0.4	2.18±4.6	2.06±2.1	0.64±0.4	0.11±0.2
collections	3.2.1	55.4	23	676	27	6.12±10.2	8.42±9.1	1.07±1.7	0.91±3.1	3.16±2.4	1.74±2.8	0.85±1.3	1.70±1.7	16.31±26.6	0.12±0.2	22.09±48.0	17.04±12.0	0.61±0.3	0.13±0.1
colt	1.2.0	35.9	24	381	67	7.11±14.7	8.19±10.9	1.53±4.0	2.02±7.8	1.95±1.7	1.20±2.9	0.33±0.7	2.54±3.8	25.76±42.4	0.12±0.2	22.92±26.9	9.21±8.9	0.45±0.3	0.26±0.3
columba	1.0	71.9	212	1183	110	6.52±11.4	5.26±6.2	1.80±3.3	0.51±4.3	2.54±1.9	0.31±1.5	0.17±0.5	1.80±2.0	9.97±13.9	0.19±0.3	10.91±25.2	4.25±5.4	0.53±0.3	0.17±0.3
compiere	330	400.5	78	2622	57	6.50±19.5	13.24±19.1	2.83±6.7	0.78±14.0	2.53±1.4	0.57±1.4	0.23±0.6	2.26±4.1	32.00±50.4	0.27±0.5	128.51±382.9	25.35±94.6	0.51±0.4	0.09±0.2
derby	10.9.1.0*	651.1	217	3010	388	12.00±54.0	11.47±26.7	3.57±25.2	0.78±6.6	2.28±1.8	0.93±2.9	0.31±0.7	2.62±5.9	34.64±134.3	0.22±0.3	45.79±135.1	10.37±23.5	0.53±0.4	0.22±0.4
displaytag	1.2	20.5	32	320	16	6.70±13.7	5.19±8.6	1.44±4.4	0.69±6.7	2.88±1.6	0.59±2.2	0.25±0.8	1.86±2.5	9.93±19.3	0.12±0.3	13.91±27.5	8.28±20.1	0.65±0.4	0.12±0.2
drawswf	1.2.9	27.7	34	311	25	6.21±13.7	8.23±14.8	2.67±3.8	0.54±1.8	2.40±1.8	0.49±1.7	0.23±0.7	2.02±3.5	17.73±30.3	0.26±0.3	8.18±10.4	4.15±5.0	0.38±0.3	0.14±0.3
drjava	20100913-r5387	89.5	30	1210	95	7.43±22.8	6.94±15.8	1.93±8.3	1.01±6.1	2.34±1.7	0.40±1.7	0.18±0.6	1.86±2.5	13.53±37.0	0.14±0.3	34.23±50.4	14.23±16.5	0.44±0.3	0.20±0.1
eclipse_SDK	3.7.1	2484.3	1425	24871	3410	8.06±17.4	7.75±12.9	2.42±6.9	1.17±12.8	2.02±1.7	0.75±3.3	0.33±0.7	2.73±5.3	23.30±62.9	0.22±0.3	61.61±287.5	11.42±15.2	0.46±0.3	0.25±0.3
emma	2.0.5312	23.1	33	321	57	7.47±20.0	4.93±5.3	2.90±12.3	0.57±1.5	1.42±1.2	0.38±0.9	0.14±0.4	2.53±4.3	14.28±22.8	0.16±0.3	9.24±12.5	3.97±4.5	0.40±0.3	0.34±0.3
exoportal	1.0.2	96	413	2162	257	4.34±8.4	5.53±6.9	1.44±2.5	0.64±6.0	2.20±1.8	0.44±1.1	0.37±0.9	1.58±1.7	9.06±12.2	0.18±0.3	8.50±24.6	3.04±3.8	0.59±0.4	0.15±0.3
findbugs	1.3.9	110.8	67	1432	127	7.15±21.7	6.14±11.3	2.49±5.4	0.64±3.2	1.98±1.8	0.61±2.8	0.34±0.8	2.85±7.8	19.33±40.8	0.23±0.3	26.66±66.6	12.25±26.6	0.59±0.4	0.18±0.3
fitjava	1.1	3.5	5	95	0	4.38±5.4	4.37±4.3	1.82±2.7	0.70±2.8	2.36±0.9	0.60±0.7	0.66±1.0	1.88±2.1	9.14±10.8	0.19±0.3	7.00±14.0	8.00±5.7	0.82±0.3	0.00±0.0
fitlibraryforfitnesse	20110301*	46.6	157	1337	102	3.34±6.0	4.64±7.1	1.19±1.9	0.71±4.2	1.96±1.5	0.23±0.7	0.13±0.4	1.53±1.6	7.65±15.5	0.12±0.3	13.15±29.7	5.01±8.3	0.58±0.4	0.13±0.2
freecol	0.10.3	106.4	51	952	32	9.63±20.8	7.71±15.2	2.54±4.3	0.83±5.4	2.88±1.9	0.60±1.2	0.32±0.8	2.76±5.1	22.95±59.3	0.20±0.3	29.39±53.7	10.69±19.7	0.45±0.3	0.07±0.1
freecs	1.3.20100406	22.6	12	146	16	12.30±34.2	8.58±14.8	4.56±12.3	0.59±4.4	1.42±0.7	0.90±0.9	0.23±0.2	4.22±10.2	40.46±70.1	0.57±0.4	30.75±34.2	10.08±14.2	0.42±0.3	0.14±0.3
freemind	0.9.0	52.8	45	751	79	5.80±13.7	7.21±12.8	2.10±5.6	0.65±2.1	2.15±1.6	0.41±1.9	0.17±0.5	1.90±2.6	14.19±28.1	0.19±0.3	22.62±46.9	7.87±10.1	0.51±0.3	0.13±0.2
galleon	2.3.0	61.1	35	533	30	11.23±26.8	6.60±10.6	3.90±5.6	0.58±3.9	3.30±2.5	0.56±0.9	0.44±0.7	2.85±6.3	21.38±32.7	0.33±0.4	12.49±28.4	6.57±7.0	0.77±0.3	0.04±0.1
ganttproject	2.1.1*	48.6	55	785	115	5.64±13.7	6.42±9.6	2.67±6.0	0.72±2.3	1.83±1.6	0.39±0.9	0.19±0.5	1.72±2.9	11.28±22.4	0.26±0.4	20.91±38.1	7.02±7.8	0.50±0.3	0.15±0.2
gt2	9.2*	876.5	894	9991	1560	6.55±24.6	7.82±15.9	2.19±14.6	0.50±2.9	2.19±1.7	0.69±2.6	0.31±0.7	1.96±3.7	16.60±39.4	0.15±0.3	29.93±95.2	8.78±12.5	0.63±0.4	0.13±0.3
hadoop	1.1.2*	319.9	238	3968	189	8.81±17.0	5.42±9.3	2.26±4.9	0.64±3.9	1.79±1.0	0.44±1.4	0.24±0.6	2.19±3.0	13.52±29.4	0.18±0.3	27.93±93.6	7.84±16.4	0.56±0.4	0.10±0.2
heritrix	1.14.4	64.9	48	656	51	7.75±13.3	7.72±12.0	1.85±3.9	0.76±2.9	3.08±2.4	0.58±1.2	0.50±1.1	2.22±3.0	18.94±33.6	0.18±0.3	17.98±36.3	8.31±8.3	0.53±0.3	0.16±0.2
hibernate	4.2.0*	431.7	856	7119	661	5.18±12.2	6.47±11.8	1.88±3.6	0.70±6.1	1.69±1.3	0.60±2.0	0.25±0.6	1.52±1.9	10.14±23.4	0.26±0.3	13.74±72.5	4.46±7.0	0.75±0.3	0.11±0.2
hsqldb	2.0.0	149.7	34	660	58	11.36±27.7	13.33±20.1	5.93±22.0	0.48±1.8	1.74±1.2	1.27±3.7	0.23±0.6	3.30±7.0	48.57±94.9	0.32±0.4	29.26±49.7	10.65±18.4	0.53±0.3	0.19±0.2
htmlunit	2.8	100.8	43	903	32	8.61±14.3	9.00±21.6	1.16±7.9	0.91±8.3	3.00±1.9	0.42±1.3	0.28±0.8	1.67±2.5	15.59±37.5	0.10±0.3	26.05±70.6	16.44±29.0	0.62±0.3	0.04±0.1
informa	0.7.0-alpha2	218	120	1771	45	8.75±26.5	8.87±12.4	3.28±7.6	0.47±4.0	4.43±2.3	2.18±6.8	1.08±1.2	2.06±3.3	18.87±37.5	0.28±0.4	25.89±92.2	12.70±25.0	0.64±0.3	0.04±0.1
ireport	3.7.5	13.9	26	223	46	5.05±14.5	6.70±10.1	1.80±3.7	0.67±2.1	1.55±1.4	0.23±0.6	0.08±0.4	1.74±2.9	12.62±17.1	0.20±0.3	9.31±23.6	5.77±5.6	0.73±0.3	0.11±0.2
itext	5.0.3	78.3	34	583	42	8.89±21.4	9.08±16.0	3.73±7.0	0.50±2.4	1.76±1.3	0.46±1.6	0.14±0.4	2.98±5.8	30.55±62.2	0.26±0.4	14.09±33.3	8.59±21.0	0.44±0.3	0.10±0.2
ivatagroupware	0.11.3	24.7	94	228	24	8.19±19.0	8.63±9.2	2.78±4.0	0.20±0.5	2.05±1.3	0.60±1.0	0.32±0.6	2.20±3.8	19.53±32.6	0.33±0.4	5.31±11.9	2.05±2.1	0.55±0.4	0.13±0.3
jFin_DateMath	1.0.1	8.9	26	121	2	5.57±12.8	6.42±7.8	5.28±29.7	0.18±1.0	2.08±0.9	0.14±0.4	0.05±0.2	1.49±1.7	10.30±11.5	0.16±0.3	1.81±3.6	3.65±2.7	0.81±0.4	0.06±0.2
jag	6.1	15.7	18	136	7	8.11±20.9	9.10±10.9	4.21±7.0	0.33±1.2	1.95±1.6	0.29±0.5	0.14±0.6	2.36±4.1	23.28±32.2	0.38±0.4	6.89±7.3	4.28±5.4	0.36±0.3	0.05±0.1
james	2.2.0	42.8	53	531	78	8.02±19.6	6.38±10.0	2.25±4.5	0.67±2.2	1.95±1.4	0.57±3.1	0.29±0.7	2.32±4.0	15.11±30.4	0.18±0.3	8.25±18.8	6.02±9.1	0.65±0.4	0.17±0.2
jasml	0.1	5.7	5	50	1	15.26±39.8	4.04±7.2	3.42±4.6	0.48±2.0	1.50±0.7	0.24±0.4	0.22±0.4	5.30±13.8	26.82±67.7	0.08±0.2	3.00±3.0	2.00±1.4	0.57±0.4	0.02±0.0
jasperreports	3.7.4*	169.8	61	1709	285	6.62±16.0	8.98±18.9	2.60±6.3	0.97±4.9	1.55±1.1	0.16±0.5	0.08±0.4	1.85±2.8	17.77±40.6	0.22±0.4	47.13±138.1	20.39±24.6	0.58±0.3	0.20±0.3
javacc	5.0	18.3	12	135	6	12.47±34.9	5.03±6.9	2.30±3.7	0.78±3.9	1.90±1.0	0.29±1.7	0.10±0.3	3.65±8.0	30.95±71.9	0.20±0.4	2.33±4.9	2.33±3.0	0.76±0.4	0.10±0.1
jboss	5.1.0	524.9	1170	7362	1546	6.52±14.3	6.07±8.7	1.44±3.3	0.41±2.0	2.00±2.1	0.34±1.0	0.20±0.7	1.77±2.4	11.37±20.4	0.15±0.3	2.45±8.0	3.53±5.6	0.71±0.4	0.23±0.3
jchempaint	3.0.1	212.8	201	2134	122	7.76±20.5	8.06±14.8	1.32±3.4	0.89±7.1	2.26±1.7	0.48±1.5	0.31±0.9	2.10±6.7	18.06±41.0	0.12±0.3	31.73±116.3	8.66±12.5	0.58±0.3	0.09±0.2
jedit	4.3.2	109.5	38	1017	69	10.98±20.8	6.01±16.4	3.10±8.6	0.42±2.1	2.14±1.9	0.59±1.2	0.38±0.8	3.11±5.6	21.83±82.5	0.20±0.3	19.08±43.9	8.53±11.5	0.67±0.3	0.12±0.1
jena	2.6.3	65.8	48	1279	192	3.94±11.8	6.98±14.3	1.42±4.6	1.33±6.1	1.86±1.6	0.65±1.5	0.40±0.7	1.81±4.8	13.78±32.4	0.12±0.3	35.90±67.2	14.40±13.6	0.50±0.3	0.23±0.2
jext	5.0	60.2	59	761	35	10.77±36.9	4.57±8.4	2.23±4.9	0.60±6.1	2.28±1.6	0.76±1.7	0.57±1.0	3.02±8.6	15.55±36.0	0.18±0.3	10.83±34.6	6.70±6.6	0.65±0.4	0.06±0.1
jfreechart	1.0.13	143.1	69	1037	103	9.54±16.3	10.60±17.3	2.10±5.2	0.64±2.2	2.32±1.3	0.68±1.3	0.15±0.4	2.02±3.0	22.97±43.9	0.17±0.3	28.96±56.2	12.25±11.2	0.63±0.4	0.10±0.1
jgraph	5.13.0.0	31.8	34	298	23	7.77±15.4	8.70±17.2	3.07±6.3	0.70±1.7	2.52±2.0	0.80±1.7	0.57±1.1	2.56±3.8	24.38±54.1	0.24±0.4	7.85±17.5	4.09±4.1	0.60±0.3	0.09±0.2
jgraphpad	5.10.0.2	24.2	22	375	16	8.94±21.3	4.40±9.6	1.57±5.5	0.72±6.1	3.62±1.8	0.46±1.3	0.46±1.1	2.46±3.5	11.83±27.2	0.12±0.3	15.54±31.2	12.00±25.1	0.66±0.4	0.14±0.2

* a more recent version than the one in the original Qualitas Corpus [6]

Table 2: *Qualitas.class Corpus* (2 of 2)

System	Version*	KLOC	NOP	NOCL	NOI	MLOC	NOM	NOA	NOC	DIT	NORM	SIX	VG	WMC	LCOM	CA	CE	I	A
jgrapht	0.8.1	17.2	30	324	33	6.79±10.5	4.29±4.8	1.47±2.4	0.72±2.0	1.96±1.4	0.26±1.0	0.15±0.6	1.79±1.7	8.08±10.1	0.12±0.2	19.47±40.3	6.53±10.1	0.50±0.4	0.15±0.3
jgroups	2.10.0	96.3	28	1036	80	7.77±12.3	7.07±9.9	3.51±5.3	0.55±3.7	1.50±1.0	0.83±2.0	0.27±0.5	2.45±3.2	20.40±39.1	0.29±0.3	46.07±87.9	15.25±17.5	0.52±0.3	0.13±0.2
jhotdraw	7.5.1	79.7	66	765	60	7.41±17.5	8.58±9.9	2.18±3.7	0.74±2.9	2.55±1.7	1.27±2.2	0.53±0.9	2.20±3.5	20.57±32.5	0.22±0.3	22.15±39.0	6.65±8.2	0.37±0.3	0.12±0.2
jmeter	2.5.1	94.8	175	1038	80	6.67±12.4	8.04±9.8	2.21±4.5	0.76±3.5	2.70±2.1	0.54±1.7	0.26±0.7	1.86±2.5	16.18±22.6	0.18±0.3	15.38±44.4	4.33±5.9	0.60±0.4	0.12±0.3
jmoney	0.4.4	8.2	4	83	3	9.77±29.6	6.93±8.3	5.18±8.0	0.45±1.7	3.20±2.0	0.49±1.5	0.46±1.1	1.88±2.3	13.39±19.4	0.32±0.4	13.00±11.4	8.50±7.4	0.50±0.3	0.08±0.1
joggplayer	1.1.4s	29.9	17	300	17	9.18±26.5	6.15±7.4	3.96±8.8	0.28±1.4	2.37±2.0	0.54±3.0	0.23±0.6	2.46±4.2	18.22±37.4	0.30±0.4	8.00±11.0	7.94±15.5	0.40±0.4	0.11±0.2
jparse	0.96	24.8	4	75	6	23.44±66.6	10.92±13.5	8.08±28.7	1.33±5.3	4.53±2.2	0.77±1.1	0.50±1.0	7.28±19.1	87.84±311.8	0.26±0.3	21.50±20.5	16.75±7.5	0.62±0.3	0.16±0.1
jpf	1.5.1	13.3	10	140	31	7.26±16.5	8.23±7.2	3.15±3.6	0.73±1.5	1.59±1.4	0.35±0.8	0.20±0.6	2.33±3.5	20.95±27.4	0.37±0.4	12.10±18.3	7.80±5.5	0.61±0.4	0.28±0.3
jrat	1.0-beta1*	14	77	382	37	3.63±5.5	4.29±4.9	1.58±2.2	0.30±1.1	1.56±1.3	0.49±1.8	0.18±0.5	1.42±1.1	6.85±7.8	0.19±0.3	5.82±17.0	3.26±2.6	0.66±0.3	0.12±0.2
jre	1.6.0	923.9	425	9923	1700	7.33±20.3	8.22±19.6	2.06±5.3	4.01±173.7	1.81±1.5	0.91±3.6	0.32±0.8	2.37±4.6	21.46±52.4	0.16±0.3	44.76±155.3	8.31±16.6	0.37±0.3	0.34±0.3
jrefactory	2.9.19	123.4	113	1668	53	6.82±14.2	6.09±18.8	1.67±6.1	0.57±3.2	2.92±2.1	0.51±1.2	0.38±0.8	2.50±5.1	17.40±165.5	0.13±0.3	19.04±44.1	8.88±14.0	0.56±0.3	0.07±0.1
jruby	1.7.3*	244.3	139	3689	176	5.81±20.5	5.69±20.0	1.23±5.6	2.22±46.2	1.64±1.2	0.66±2.2	0.25±0.6	2.11±4.5	14.63±59.3	0.10±0.2	36.94±119.4	8.94±18.9	0.54±0.4	0.12±0.2
jsXe	04_beta	18.5	14	251	11	8.93±19.1	5.04±8.4	2.04±3.5	0.21±1.2	2.05±1.8	0.39±0.9	0.22±0.6	2.71±5.8	15.29±31.5	0.19±0.3	15.07±17.2	6.50±4.3	0.42±0.2	0.09±0.1
jspwiki	2.8.4	60.2	70	582	36	9.69±16.4	6.51±11.9	1.82±3.5	0.47±2.3	2.03±1.1	0.53±1.2	0.25±0.6	2.24±3.2	16.02±27.7	0.19±0.3	14.24±40.2	6.16±7.5	0.54±0.3	0.11±0.2
jtopen	7.8*	342	15	1915	100	8.89±27.5	11.63±20.6	3.24±7.0	0.59±2.8	2.09±1.3	1.24±2.7	0.39±0.7	2.76±7.0	33.67±70.6	0.25±0.4	30.00±86.9	33.67±45.7	0.78±0.3	0.09±0.1
jung	2.0.1	37.9	44	588	63	6.20±14.4	5.87±8.7	2.34±3.8	0.67±2.1	1.79±1.6	0.40±1.2	0.19±0.6	1.86±2.3	11.44±15.8	0.22±0.3	20.36±34.6	7.14±6.9	0.51±0.3	0.17±0.2
junit	4.1	6.6	28	171	15	4.02±4.9	4.68±5.1	0.93±1.2	0.70±1.8	1.57±1.2	0.29±0.9	0.16±0.5	1.52±1.1	8.63±11.5	0.11±0.2	8.32±12.5	4.21±3.2	0.56±0.3	0.22±0.2
log4j	2.0-beta*	33.3	64	616	55	5.98±10.2	4.40±8.6	1.54±2.8	0.43±1.6	1.45±1.0	0.31±0.9	0.13±0.3	1.98±2.5	10.48±17.2	0.15±0.3	16.11±38.2	5.75±5.3	0.63±0.4	0.12±0.2
lucene	4.2.0*	413	387	4629	142	9.99±27.7	5.14±6.3	2.18±3.6	0.98±9.7	2.28±1.2	0.43±1.1	0.21±0.6	2.50±9.3	13.79±33.0	0.18±0.3	23.16±100.3	6.66±11.8	0.69±0.3	0.07±0.2
marauroa	3.8.1	17.7	41	247	14	6.22±9.1	6.57±7.7	1.98±3.4	0.41±2.2	1.55±0.9	0.37±0.7	0.10±0.2	1.81±2.0	12.72±19.9	0.18±0.3	10.98±19.7	4.39±6.0	0.60±0.4	0.10±0.2
maven	3.0.5*	70.9	181	916	175	6.25±13.8	6.72±12.9	1.80±3.4	0.41±1.0	1.68±1.5	0.38±2.0	0.10±0.4	1.81±2.3	12.70±30.1	0.18±0.3	6.61±16.9	3.76±5.4	0.66±0.4	0.20±0.3
megamek	0.35.18	242.8	37	1859	64	13.93±43.4	6.53±23.8	3.52±12.5	0.81±5.8	4.10±2.3	0.61±2.5	0.40±0.9	3.80±12.5	27.30±145.9	0.16±0.3	67.57±211.8	38.00±122.5	0.47±0.4	0.14±0.2
mvnforum	1.2.2-ga	105.3	75	784	144	8.20±23.0	9.88±15.6	2.85±7.0	0.51±2.7	1.15±0.9	0.20±1.0	0.07±0.3	2.39±6.2	27.44±48.2	0.22±0.4	20.12±41.8	7.11±8.8	0.53±0.4	0.15±0.3
myfaces_core	2.1.10*	343.3	277	3426	265	7.71±19.4	6.92±10.6	2.12±8.0	0.90±3.4	1.97±1.4	0.74±2.8	0.24±0.6	2.17±4.0	17.02±32.4	0.17±0.3	16.14±47.9	4.84±7.3	0.53±0.4	0.20±0.3
nakedobjects	4.0.0	133.9	496	2975	470	3.54±6.5	5.77±8.6	1.16±2.4	1.01±4.0	1.82±1.3	0.41±1.2	0.19±0.5	1.40±1.3	8.48±13.7	0.13±0.3	17.27±54.8	4.72±6.3	0.57±0.4	0.22±0.3
nekohtml	1.9.14	7.6	7	64	5	9.56±18.6	7.56±9.9	3.42±7.4	0.39±1.3	1.67±1.1	0.73±2.1	0.22±0.5	3.30±5.7	27.36±57.3	0.22±0.3	4.86±5.0	4.86±2.6	0.63±0.3	0.09±0.2
netbeans	7.3*	2153.2	1959	25054	2374	7.75±30.8	6.44±10.1	2.43±5.8	0.66±5.2	1.97±1.6	0.56±1.6	0.27±0.7	2.45±5.8	18.03±47.6	0.22±0.3	6.39±12.9	6.11±7.3	0.64±0.3	0.17±0.3
openjms	0.7.7-beta-1	39.4	66	616	79	5.85±9.6	6.12±7.5	1.62±2.8	0.66±1.5	1.86±1.5	0.34±0.8	0.14±0.4	1.86±2.0	12.44±18.3	0.18±0.3	11.21±17.8	5.44±6.2	0.52±0.4	0.17±0.2
oscache	2.3*	7.6	22	115	10	6.82±12.2	6.09±8.3	2.21±3.2	0.36±0.9	2.03±1.2	0.50±1.2	0.23±0.5	2.36±3.8	15.42±32.9	0.22±0.3	4.18±7.9	3.23±1.8	0.63±0.3	0.13±0.2
picocontainer	2.10.2	9.3	15	206	29	3.79±6.1	5.85±7.4	0.98±1.6	0.97±2.2	1.75±1.4	0.45±1.0	0.25±0.6	1.58±1.7	9.71±16.4	0.09±0.2	12.40±26.5	8.87±9.7	0.55±0.3	0.23±0.3
pmd	4.2.5	60.7	88	872	52	6.37±17.0	6.30±21.9	1.70±8.0	0.71±4.6	2.27±1.4	0.53±1.2	0.44±0.8	2.61±6.8	17.59±128.1	0.11±0.3	10.41±38.8	5.36±6.7	0.75±0.4	0.09±0.2
poi	3.6	203.1	212	2414	131	6.63±16.0	7.34±18.3	2.05±7.8	0.55±3.8	2.12±1.2	0.41±1.1	0.15±0.4	1.81±3.3	14.68±30.5	0.18±0.3	16.63±48.8	7.71±13.7	0.54±0.3	0.09±0.2
pooka	3.0-080505	44.5	28	491	36	7.67±13.2	8.31±12.2	2.31±4.0	0.48±1.6	2.48±1.9	0.85±3.1	0.27±0.8	2.53±3.2	21.50±38.8	0.26±0.4	16.93±27.4	7.50±12.6	0.46±0.3	0.13±0.2
proguard	4.9*	62.6	35	648	42	5.62±16.3	8.35±13.7	2.31±5.1	1.48±10.2	1.71±1.2	2.48±6.8	0.54±0.8	1.97±4.5	17.33±29.6	0.17±0.3	48.17±80.3	12.89±12.8	0.46±0.3	0.15±0.2
quartz	1.8.3	28.6	51	269	36	7.08±18.0	10.28±19.0	2.57±8.2	0.94±4.0	1.47±1.1	0.40±1.2	0.13±0.4	2.00±3.8	21.40±47.1	0.20±0.3	5.41±19.1	3.82±4.2	0.77±0.3	0.10±0.2
quickserver	1.4.7	18.3	28	196	23	8.33±24.0	7.50±15.2	3.65±8.0	0.40±1.4	1.50±1.3	0.47±2.3	0.15±0.4	2.44±5.7	20.03±44.6	0.26±0.4	8.64±17.3	4.71±3.5	0.61±0.4	0.13±0.3
quilt	0.6-a-5	8	20	113	9	6.63±13.3	6.99±7.9	3.40±4.6	0.49±1.7	1.55±1.0	0.46±0.9	0.18±0.4	1.91±2.5	13.81±14.7	0.31±0.4	6.05±8.1	4.05±4.7	0.57±0.4	0.14±0.2
roller	5.0.1*	65.7	95	738	76	6.08±13.2	8.06±9.9	2.34±3.6	0.40±2.1	1.78±1.3	0.67±1.2	0.30±0.6	1.96±2.8	17.46±22.1	0.28±0.4	8.21±16.8	6.22±7.0	0.65±0.3	0.13±0.2
rssowl	2.0.5	100.6	54	771	110	11.34±25.1	7.55±9.7	2.91±7.1	1.09±5.3	1.77±1.4	0.37±1.0	0.13±0.4	2.60±4.0	21.29±36.8	0.24±0.4	39.18±68.1	11.50±9.5	0.48±0.3	0.16±0.3
sablecc	3.2	28.4	5	242	5	9.45±41.9	8.56±23.2	1.81±3.6	0.96±5.9	2.15±1.0	2.09±5.4	0.58±0.7	2.08±6.8	18.34±53.1	0.15±0.3	33.40±47.1	26.00±33.3	0.56±0.3	0.10±0.1
sandmark	3.4	93.2	127	1045	42	9.47±24.6	6.15±8.5	2.16±3.2	0.62±2.4	2.10±1.4	0.52±1.3	0.20±0.4	2.77±5.9	19.42±40.3	0.19±0.3	10.83±39.0	5.07±8.4	0.72±0.3	0.08±0.2
springframework	3.0.5	329.4	598	5999	692	5.73±10.0	5.76±9.4	1.35±3.1	0.74±4.0	1.81±1.7	0.28±0.8	0.17±0.6	1.58±1.9	9.46±17.2	0.15±0.3	6.57±16.0	5.39±6.1	0.71±0.3	0.20±0.3
squirrel_sql	3.1.2	6.9	3	73	16	7.61±14.7	7.71±9.3	3.96±5.6	0.26±0.6	1.62±1.8	0.41±1.0	0.15±0.5	1.74±1.6	13.51±22.0	0.42±0.4	6.67±6.6	12.67±14.4	0.59±0.3	0.09±0.1
struts	2.2.1	143.4	261	2239	169	6.07±12.8	6.29±8.8	2.06±4.4	0.74±4.5	2.24±1.7	0.50±1.9	0.23±0.6	1.89±3.4	12.28±28.8	0.22±0.3	13.35±50.3	5.37±7.8	0.68±0.4	0.11±0.2
sunflow	0.07.2	22	22	209	19	10.28±25.7	6.13±7.5	3.95±6.2	0.64±2.5	1.17±0.9	0.12±0.4	0.07±0.3	2.80±5.5	19.68±29.6	0.20±0.3	18.54±29.4	6.18±6.3	0.52±0.3	0.03±0.1
tapestry	5.1.0.5	97.8	144	2133	406	4.88±7.8	4.31±9.5	1.41±2.7	0.74±4.2	1.39±1.5	0.15±1.6	0.05±0.3	1.38±1.2	6.30±13.8	0.14±0.3	29.66±84.3	11.12±25.4	0.60±0.4	0.17±0.3
tomcat	7.0.2	181.2	157	1876	187	7.38±17.4	8.15±15.0	2.69±5.7	0.61±2.9	1.68±1.2	0.71±2.3	0.30±0.7	2.53±4.5	22.24±49.0	0.23±0.3	14.62±31.8	4.69±7.4	0.54±0.4	0.16±0.3
trove	2.1.0	5.8	4	72	9	8.35±12.2	6.75±10.3	0.86±1.3	1.78±6.0	1.61±1.2	0.51±1.5	0.14±0.4	1.81±1.6	13.15±20.2	0.09±0.2	1.25±1.3	4.25±3.6	0.83±0.2	0.11±0.1
velocity	1.6.4	37	42	445	47	7.68±23.8	6.16±10.4	1.76±5.6	0.65±3.5	2.00±1.4	0.63±2.3	0.26±0.5	2.57±9.8	17.37±95.5	0.13±0.3	14.07±22.9	7.29±13.2	0.66±0.4	0.14±0.2
wct	1.5.2	52.3	130	677	82	4.62±11.9	8.99±11.2	3.24±5.2	0.39±1.7	1.93±1.9	0.23±0.5	0.14±0.5	1.66±2.4	15.36±23.7	0.36±0.4	10.29±22.1	3.61±4.9	0.51±0.4	0.13±0.2
webmail	0.7.10	10.1	19	115	11	6.42±15.4	8.66±10.3	1.98±2.8	0.77±2.7	1.57±1.1	0.34±0.7	0.15±0.4	2.11±3.4	19.25±34.4	0.23±0.3	12.05±15.3	3.58±4.8	0.43±0.4	0.12±0.3
weka	3.6.9*	272.6	113	1766	122	9.92±23.2	10.41±12.5	3.88±6.4	1.29±10.3	2.48±1.5	1.10±1.8	0.33±0.6	2.68±4.5	30.68±44.3	0.32±0.4	20.50±85.9	9.97±11.3	0.66±0.3	0.13±0.2
xalan	2.7.1	183.9	89	1409	238	9.88±36.9	8.21±15.3	1.85±4.1	1.01±3.5	2.05±1.6	1.11±2.6	0.65±1.2	2.41±5.7	21.47±42.9	0.17±0.3	27.86±56.2	8.70±16.3	0.56±0.4	0.22±0.3
xerces	2.10.0	126	53	947	157	8.98±23.1	9.76±13.6	2.98±7.0	0.76±2.8	2.00±1.8	1.21±3.4	0.43±1.0	3.06±6.3	31.79±67.4	0.22±0.3	21.11±38.2	10.15±12.8	0.56±0.3	0.19±0.3
xmojo	5.0.0	4	15	32	4	9.31±16.3	8.31±6.8	5.16±5.3	0.16±0.4	1.50±1.5	0.22±0.5	0.07±0.2	2.25±3.4	20.56±24.5	0.48±0.4	0.87±0.8	0.60±0.5	0.47±0.4	0.10±0.2

* a more recent version than the one in the original Qualitas Corpus [6]