# Project 67: Evaluating Identifier Quality

Students: Jack Chu & Jafar Maash

Supervisor: Ewan Tempero

UNIVERSITY OF AUCKLAND Waipapa Taumata Rau NEW ZEALAND

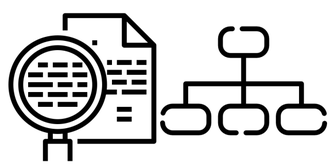
ENGINEERING
DEPARTMENT OF ELECTRICAL, COMPUTER, AND SOFTWARE ENGINEERING

## Introduction + Motivation

- 70% of code is made up of identifiers
- The literature has guidelines for how to name these identifiers
- Following these guidelines is associated with higher quality and more comprehensible code
- But, how closely are they really followed in practice?

## Our Approach

Extract identifiers from code



Parse identifiers

Apply rules to identifiers

Analyse data

## Identifier Naming Guidelines (Semantics)

J. Gosling et al, The Java® Language Specification, 2022

| | Rule | Example |
|---|---|---|
| 1 | Fields should have names that are nouns, noun phrases, or abbreviations for nouns. | `shuffled` `shuffledCards` |
| 2 | Method names should be verbs or verb phrases | `randomRoom` `randomiseRoom` |
| 3 | A method that tests a boolean condition V about an object should be named isV. | `landscape` `isLandscape` |

D. Binkley et al, “Improving identifier informativeness using part of speech information”, 2011

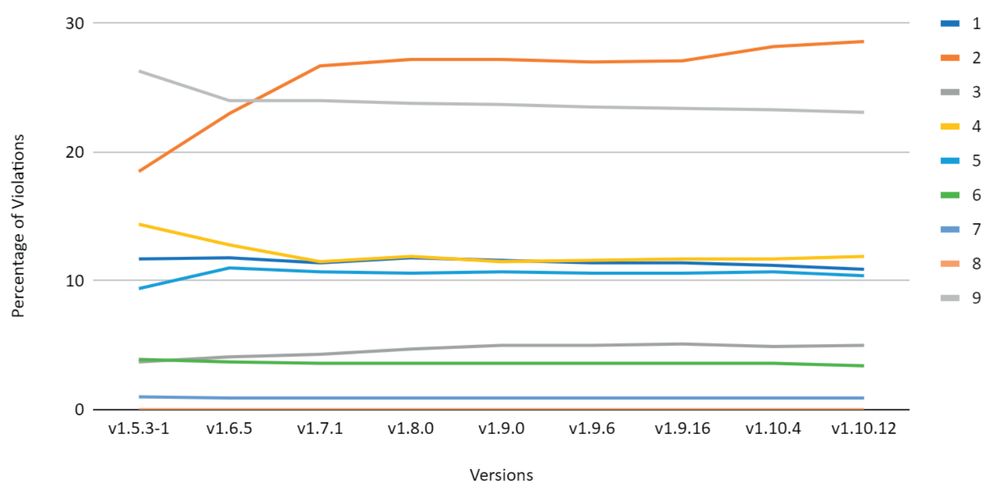
| | Rule | Example |
|---|---|---|
| 4 | Non-boolean field names should not contain a present tense verb | `create_mp4` `created_mp4_container_type` |
| 5 | Field names should never be only a verb | `int recycle` `int recycledCount` |
| 6 | Field names should never be only an adjective | `interesting` `interestingItems` |
| 7 | Boolean field names should contain third-person forms of the verb “to be” or the auxiliary verb “should” | `deleted` `was_deleted` |

P.A. Relf, “Achieving Software Quality through Source Code Readability”, 2004

| | Rule | Example |
|---|---|---|
| 8 | Identifier names should not be composed only from numeric words or values | `OneHundred` |
| 9 | Identifiers should consist of 2,3, or 4 words | `recycle` `recycledCount` |

Note: further rules were implemented but have been omitted due to rarity of violation or simplicity

## Analysis




| Violations | v1.5.3-1 | v1.6.5 | v1.7.1 | v1.8.0 | v1.9.0 | v1.9.6 | v1.9.16 | v1.10.12 | v1.10.4 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 11.7 | 11.8 | 11.4 | 11.8 | 11.6 | 11.4 | 11.4 | 10.9 | 11.2 |
| 2 | 18.5 | 23 | 26.7 | 27.2 | 27.2 | 27 | 27.1 | 28.6 | 28.2 |
| 3 | 3.7 | 4.1 | 4.3 | 4.7 | 5 | 5 | 5.1 | 5 | 4.9 |
| 4 | **14.4** | **12.8** | **11.5** | **11.9** | **11.5** | **11.6** | **11.7** | **11.9** | **11.7** |
| 5 | **9.4** | **11** | **10.7** | **10.6** | **10.7** | **10.6** | **10.6** | **10.4** | **10.7** |
| 6 | **3.9** | **3.7** | **3.6** | **3.6** | **3.6** | **3.6** | **3.6** | **3.4** | **3.6** |
| 7 | **1** | **0.9** | **0.9** | **0.9** | **0.9** | **0.9** | **0.9** | **0.9** | **0.9** |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 26.3 | 24 | 24 | 23.8 | 23.7 | 23.5 | 23.4 | 23.1 | 23.3 |

### Libraries / Tools

| Violations | Java Design Patterns | JUnit | Arduino | Mindustry | Signal Android | sndcpy | Termux | Tsunami | Algorithms | Baritone | Dropwizard Jobs | Guava | Exercism | Mockito | OpenRefine | Realm | Spring Boot | Zerocode |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11.2 | 15.1 | 23.6 | 17.8 | 14.4 | 7.9 | 5.9 | 5.6 | 16.3 | 17.1 | 14.3 | 14.1 | 13.6 | 14.8 | 9.5 | 6.4 | 8.2 | 5.5 |
| 2 | 34.1 | 30.5 | 39.4 | 28.8 | 24.7 | 27.8 | 19.9 | 15.8 | 58.2 | 28.6 | 21.8 | 45.4 | 44 | 19.2 | 35.1 | 26.1 | 23.9 | 33.5 |
| 3 | 2.2 | 1.8 | 13.5 | 12 | 8.4 | 0 | 7.3 | 2 | 4.1 | 11.8 | 1.8 | 2.3 | 1.5 | 2.3 | 9.5 | 2.9 | 1.3 | 1.3 |
| 4 | **13.5** | **11.1** | **30** | **17.3** | **23.3** | **21.1** | **23.8** | **12.7** | **8.4** | **21.4** | **15.2** | **15** | **9.1** | **17.3** | **12.2** | **11.1** | **12.8** | **13.8** |
| 5 | **1.5** | **4.2** | **9.5** | **13** | **10.8** | **0** | **6.6** | **1.9** | **3.7** | **6.5** | **1.9** | **3.9** | **1.6** | **3.2** | **4.6** | **2.7** | **4.8** | **1.9** |
| 6 | **3.3** | **5.4** | **8.8** | **8.2** | **6.1** | **0** | **1.7** | **1.9** | **4.8** | **5.6** | **5.7** | **4.4** | **4.4** | **4.6** | **3.7** | **1.7** | **3.8** | **2.7** |
| 7 | **1.4** | **1.7** | **2.8** | **2** | **1.7** | **0** | **0.6** | **1.2** | **1.1** | **2** | **0** | **1.5** | **2.5** | **1.5** | **0.7** | **0.6** | **1** | **0.1** |
| 8 | 0 | 0.8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.2 | 0 | 0.2 | 0 | 0.2 | 0.1 | 0 |
| 9 | 33.7 | 34.6 | 63.5 | 34.6 | 32.4 | 39.5 | 36.5 | 20.6 | 39.1 | 34.8 | 48.6 | 24.9 | 41.5 | 33.5 | 26.6 | 28.8 | 21.1 | 22.5 |

### Applications

| Violations | AntennaPod | appsmith | book project | glide | LanguageTool | News Android | Open LaTeX Studio | Teammates | SeriesGuide |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 10.5 | 4.2 | 7.5 | 9.8 | 8.8 | 9.3 | 8.8 | 4.5 | 8 |
| 2 | 39.6 | 21.5 | 11.3 | 17.2 | 24.3 | 23 | 12 | 12.7 | 22.9 |
| 3 | 6.5 | 1 | 2 | 2.1 | 5.9 | 4.3 | 1.4 | 1.7 | 8.2 |
| 4 | **20.5** | **15.7** | **13.8** | **12.3** | **14.2** | **21.5** | **13.3** | **17.3** | **23.2** |
| 5 | **5.9** | **0.9** | **1.2** | **2.5** | **7.1** | **3.9** | **3.1** | **1.5** | **4.9** |
| 6 | **3.6** | **1.3** | **1.4** | **3.2** | **2.5** | **3.5** | **3.1** | **1.2** | **2.7** |
| 7 | **0.6** | **0.3** | **1.2** | **1.3** | **0.7** | **0.6** | **1** | **0.2** | **0.6** |
| 8 | 0 | 0 | 0 | 0.1 | 0 | 0 | 0 | 0 | 0 |
| 9 | 29.7 | 17.7 | 28.3 | 25 | 19 | 30.1 | 35 | 21.8 | 25.5 |

### Games

| Violations | FreeCol | Kroniax | Stendhal | Terasology | Cengball | Tic tac toe | EbbeFlut | Tempest | PuzzleGame | Pixel dungeon |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 18.3 | 11.9 | 11.3 | 8.2 | 10.2 | 31.1 | 25.4 | 14.7 | 14.9 | 13.2 |
| 2 | 15.2 | 11 | 23.3 | 21 | 10.3 | 28 | 22.7 | 20.3 | 35.2 | 39.6 |
| 3 | 24.1 | 6.8 | 6.2 | 3.6 | 2.9 | 24 | 9.2 | 3.8 | 4.7 | 9.9 |
| 4 | **13** | **50** | **14.8** | **13.8** | **15.4** | **11.1** | **15.9** | **10.4** | **15.4** | **7.4** |
| 5 | **5.1** | **4.8** | **4.9** | **5.3** | **0.3** | **4.4** | **11.1** | **5.7** | **6.7** | **6** |
| 6 | **7.3** | **1.8** | **4.2** | **3.1** | **2.6** | **8.9** | **4.8** | **4** | **4.6** | **6.1** |
| 7 | **3** | **0.6** | **1.1** | **0.8** | **1.2** | **0** | **3.2** | **1.3** | **1.4** | **1.5** |
| 8 | 0 | 0 | 0.1 | 0 | 0 | 0 | 0 | 0 | 0.2 | 0 |
| 9 | 42.5 | 27.4 | 28.7 | 21.4 | 25.6 | 77.8 | 43.9 | 34.4 | 35.5 | 46 |

## Next Steps

- Perform analysis on humans to investigate the relationship between code readability and number of guideline violations.
- Implement more complex rules (e.g. rules dependent on code context, or more abstract semantics)
- Find suitable metrics for code quality and investigate correlations between it and guideline adherence

## Conclusions

- Certain guidelines are rarely ever followed while others are consistent globally.
- Identifier names can often be quite cryptic or vague
- Generally speaking, percentage violations do not get worse with time or as the same repository's size increases
- Some rules, in particular older ones, had many examples of violations that were subjectively valid given context