Preliminary exploratory analysis: Incremental PMD Alerts + Commits

Bruno Crotman

02/11/2019

Contents

1 Incremental alerts and commits

This report shows how incremental PMD alerts were identified on a Java project. The incremental alerts are those included and excluded in each commit of the Twitter4j project. These are the alerts that begin in lines that were included or excluded from the project in a commit. Lines that are modified are considered excluded from the previous commit code and included in the posterior commit code.

2 Mapping lines of code using diff

A diff was run on each pair of sequential commits.

Using diff information, a map was built for each pair of sequential commits. The map links the line on previous commit with the equivalent line on the posterior commit. The lines that were modified are not mapped to any lines in the other commit of the pair.

Figure 1 shows an example. New lines are included in the posterior commit:

- Lines ...-114 in the previous commit are linked to ...-114 in the posterior commit.
- Lines 115-124 in the posterior commit are not linked with any lines in the previous commit
- Lines 115-... in the previous commit are linked with 125-... in the posterior commit

Figure 2 shows an example. A line is modifies and new lines are included in the posterior commit. The line that is modified is treated as if a line was removed and another one was added:

- Lines ...-114 in the previous commit are linked to ...-231 in the posterior commit.
- Line 219 in the previous commit is not linked with any lines in the posterior commit
- Line 250 in the posterior commit is not linked with any lines in the previous commit
- Lines 250-253 in the posterior commit not linked with any lines in the previous commit
- Lines 220-... in the previous commit are linked with 254-... in the posterior commit

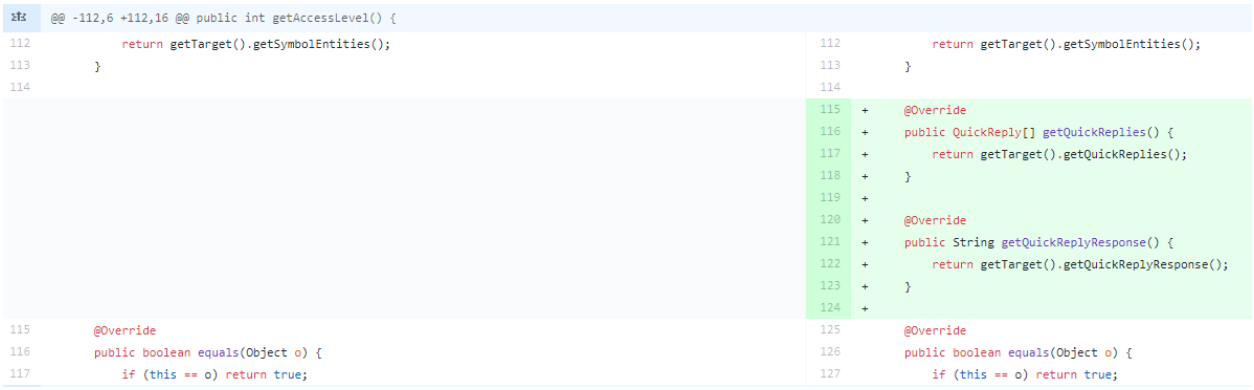

Figure 1: Diff example 1: linking lines from previous and posterior commmits

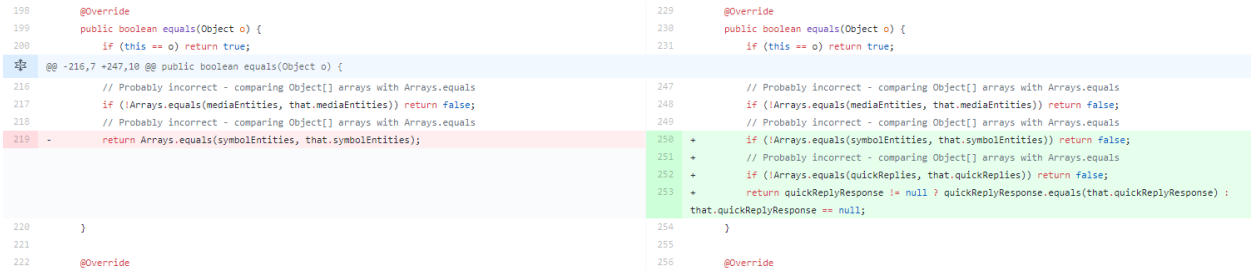

Figure 2: Diff example 2: linking lines from previous and posterior commmits

3 Identifying incremental alerts

PMD alerts are linked to the line of code where the problem begins. So it's possible to identify those alerts that are added or removed in the event of the commit.

In the example on Table 1, we can identify an alert that occurs in a line of code that exists in the posterior version and is not linked to any line in the previous version. So we categorize this alert as a new alert added at the moment of the commit of the posterior version.

4 Some results

We present the some statistics about the data.

It´s only a preliminar analysis to gather some properties of the data.

Figure 3 shows the histogram of the number of lines changed at each commit.

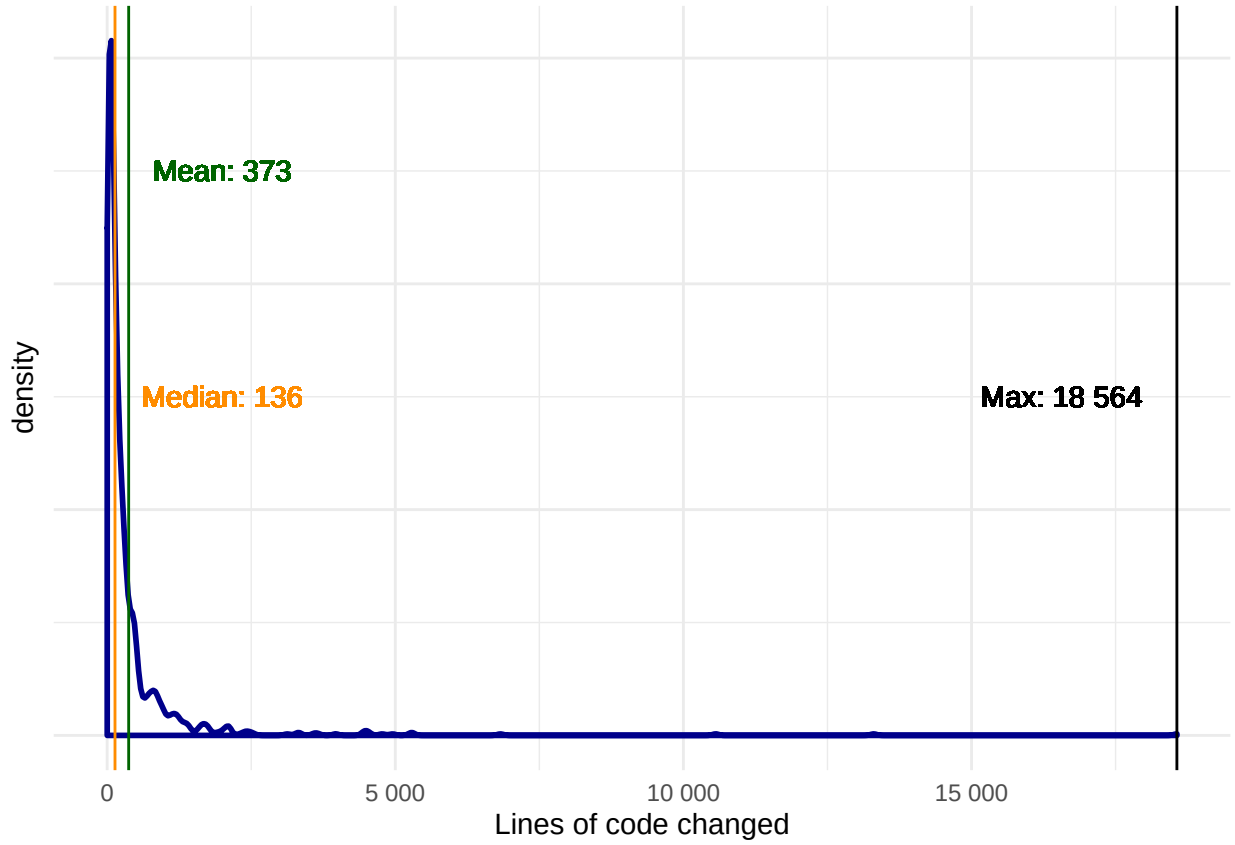

Figure 3: Histogram of changed lines of code

Table 1: Alert rules linked to previous and posterior lines of code

file	map_prev	map_post	prev_rule	post_rule
...ssage.java	32	32	UnnecessaryFullyQualifiedName	UnnecessaryFullyQualifiedName
...ssage.java	117	127	ControlStatementBraces	ControlStatementBraces
...ssage.java	118	128	ControlStatementBraces	ControlStatementBraces
...NImpl.java	76	80	EmptyStatementNotInLoop	EmptyStatementNotInLoop
...NImpl.java	NA	115	NA	OptimizableToArrayCall
...NImpl.java	200	231	ControlStatementBraces	ControlStatementBraces
...NImpl.java	201	232	ControlStatementBraces	ControlStatementBraces
...NImpl.java	205	236	ControlStatementBraces	ControlStatementBraces
...NImpl.java	206	237	ControlStatementBraces	ControlStatementBraces
...NImpl.java	207	238	ControlStatementBraces	ControlStatementBraces