Do Comments follow commenting conventions? A case study in Java and Python

Pooja Rani, Suada Abukar, Nataliia Stulova,

Alexandre Bergel, Oscar Nierstrasz

University of Bern, Switzerland

```java
/** Event fired when a spacer element is hidden or shown in Escalator
 *
 * @author Vaadin Ltd
 * @since 7.7.13
 */

public class SpacerVisibili            {
    System.out.println(..);
}
```

```java
// This class fires an Event when a spacer element is hidden or shown in
the Escalator. I am the author of this class.

// It is present in the project since the version 7.7.13 and might
delete in the next versions.

public class SpacerVisibili                              {
    System.out.println(..);
}
```

No strict convention for their style & syntax

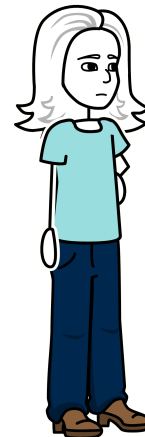# Coding Style guidelines

E.g., Oracle comment guidelines

- First sentence should summarise the class

- Use phrases instead of complete sentences

- Use third person instead of second person

Do developers follow these guidelines?

```
/** Event fired when a spacer element is hidden or shown in Escalator
 *
 * @author Vaadin Ltd
 * @since 7.7.13
 */
```

- First sentence should summarise the class
- Use phrases instead of complete sentences
- Use third person instead of second person

Does this comment follow the guidelines?

Does it not?

Previous works by Rani et.al.,

Empirical Software Engineering
https://doi.org/10.1007/s10664-021-09981-5

What do class comments tell us? An investigation of comment evolution and practices in Pharo Smalltalk

Pooja Rani[1] · Sebastiano Panichella[2] · Manuel Leuenberger[1] · Mohammad Ghafari[3] · Oscar Nierstrasz[1]

The Journal of Systems & Software 181 (2021) 111047

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

How to identify class comment types? A multi-language approach for class comment classification

Pooja Rani [a,*], Sebastiano Panichella [b], Manuel Leuenberger [a], Andrea Di Sorbo [c], Oscar Nierstrasz [a]

[a] Software Composition Group, University of Bern, Switzerland
[b] Zurich University of Applied Science, Switzerland
[c] Department of Engineering, University of Sannio, Italy

Analyzed Smalltalk class comments against the style guidelines but what about other languages? Java, Python

Analyzed Java and Python class comments but not against the style guidelines.

Methodology

Dataset

Do developers follow commenting guidelines for Java and Python comments?

Dataset

Java

Apache Hadoop

Apache Spark

Eclipse.cdt

Google Guava

Google Guice

Python

Django

iPython

Mailpile

Pandas

Pipenv

PyTorch

Requests

Methodology

Projects

Extracting
guidelines

1

Identifying
rules

2

Create
taxonomy

3

Class
comments

4

Validate comments

5

9

1 Extracting guidelines

Java

Apache Hadoop ─────┐
                   │
Apache Spark ──────┼──── Oracle
                   │
Eclipse.cdt ───────┘

Google Guava ──────┐
                   ├──── Google
Google Guice ──────┘

Python

Django ────────┐
               │
Pipenv ────────┤
               │
Mailpile ──────┼──── PEP8/257
               │
Request ───────┤
               │
iPython ───────┘

Pandas ──────── Numpy

PyTorch ──────── Google
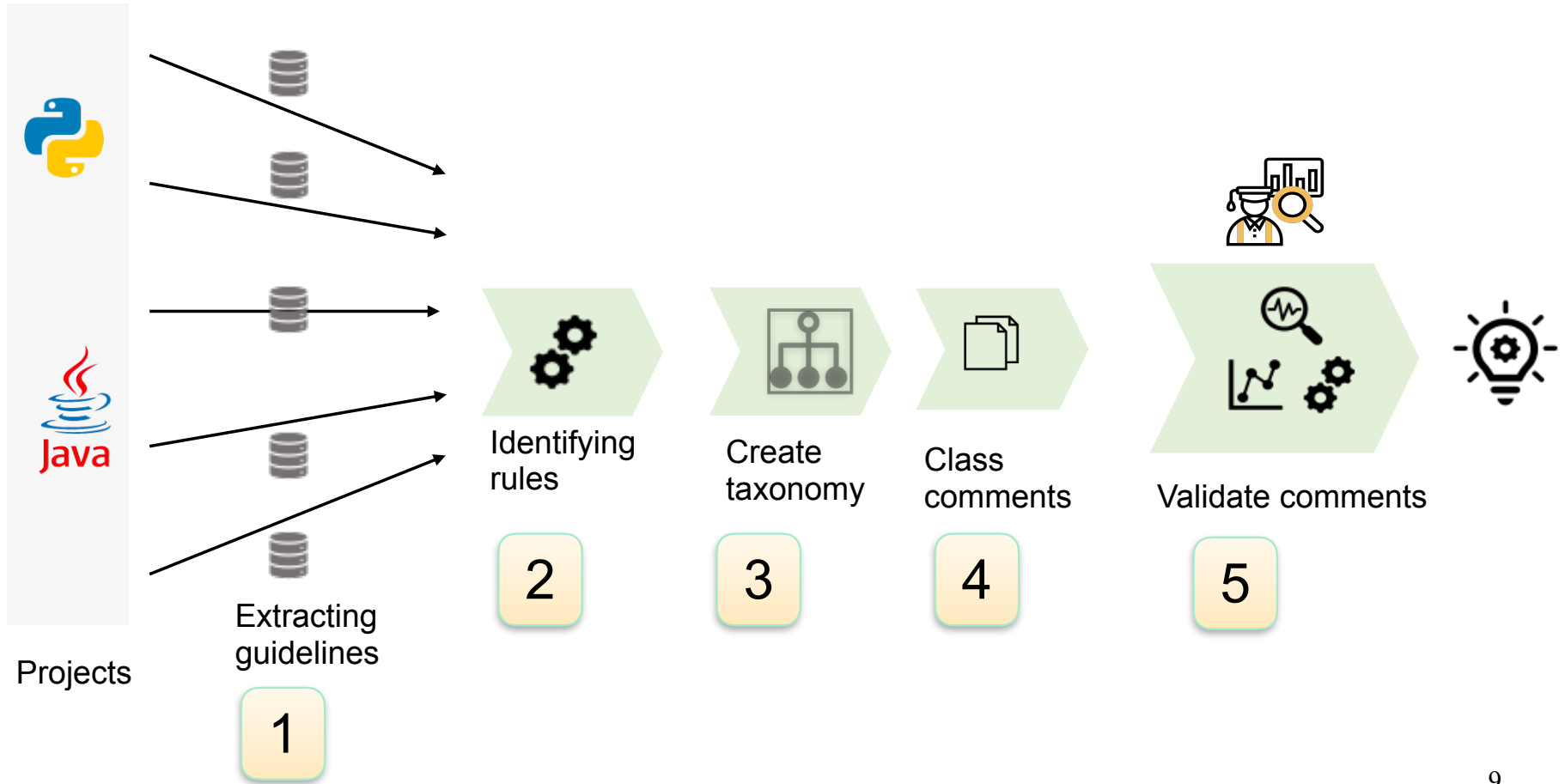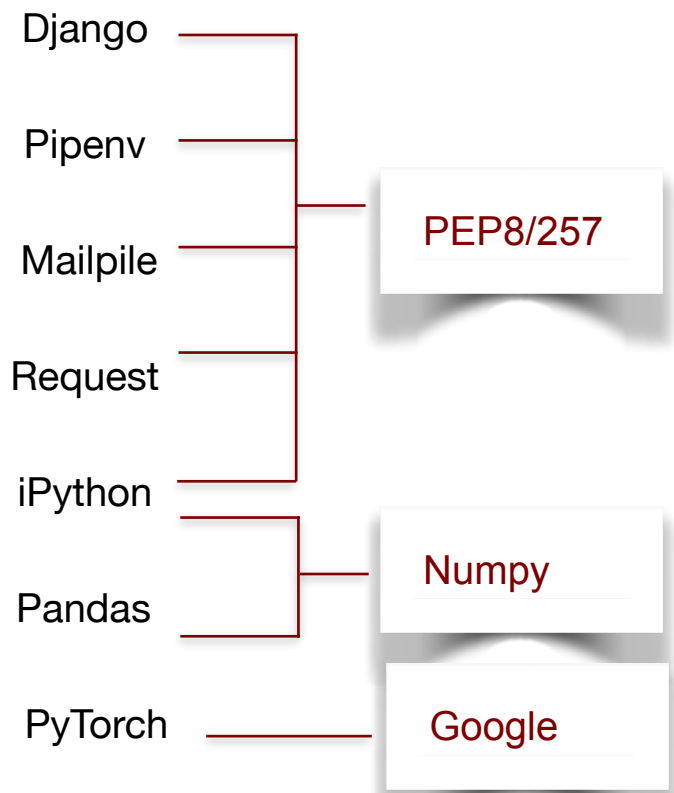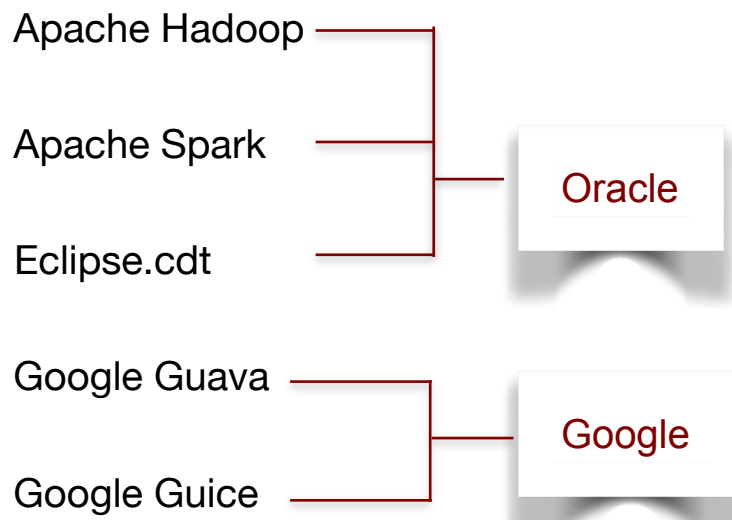
2 Identifying rules

Multi-line Docstrings ①

Multi-line docstrings consist of a summary line just like a one-line docstring, ② followed by a blank line, followed by a more ③ elaborate description. The summary line may be used by automatic indexing tools; it is important that it fits on one line and is ④ separated from the rest of the docstring by a blank line. The summary line may be on the same line as the opening quotes or on the next line. ⑤ The entire docstring is indented the same as the quotes at its first line ⑥ (see example below).

●PEP 257 -- Docstring Conventions | Python.org

Create taxonomy

Content   Formatting   Content

Multi-line Docstrings ①

Multi-line docstrings consist of a summary line just like a one-line docstring, followed by a blank line, followed by a more ② ③

elaborate description. The summary line may be used by automatic indexing tools; it is important that it fits on one line and is ④

separated from the rest of the docstring by a blank line. The summary line may be on the same line as the opening quotes or

on the next line. The entire docstring is indented the same as the quotes at its first line (see example below). Formatting

⑤   ⑥

Formatting   Formatting

●PEP 257 -- Docstring Conventions | Python.org

RQ1: Types of comment conventions?

Percentage of comment conventions in Java

Percentage of comment conventions in Python

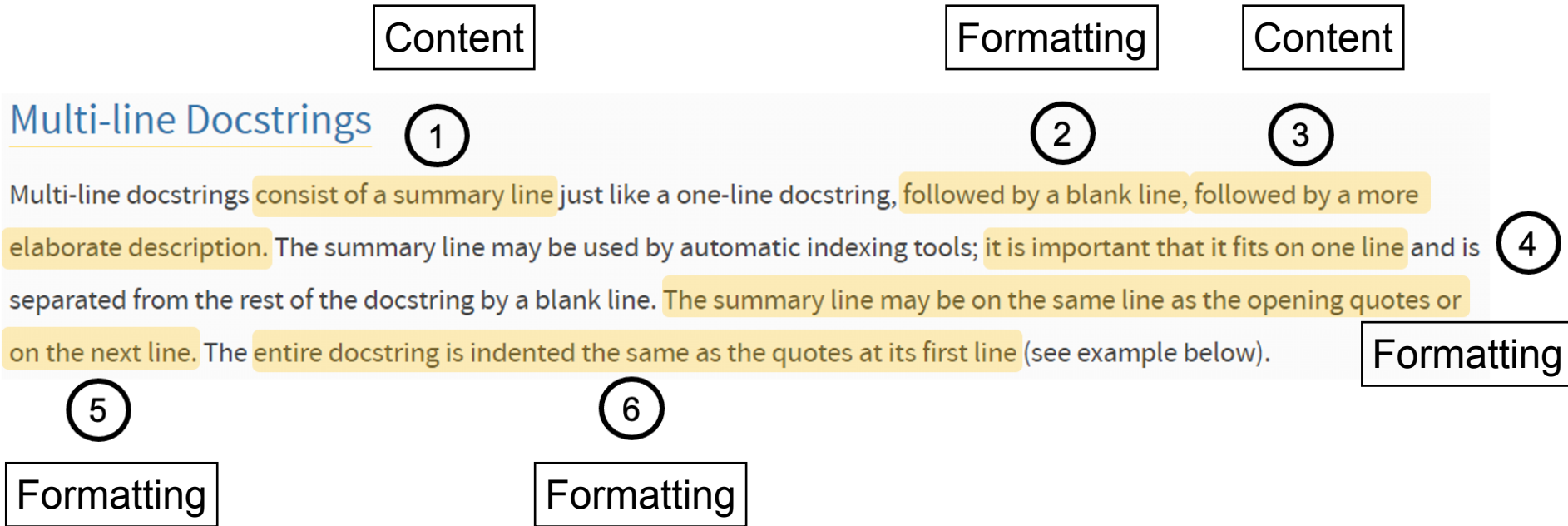Legend: Formatting, Content, Syntax, Structure, Writing style, Other

	Formatting	Content	Syntax	Structure	Writing style	Other
Oracle (149)	14%	38%	21%	14%	11%	1%
Google (19)	26%	21%	11%	26%	16%	
Spark* (190)	13%	42%	22%	12%	11%	1%
Eclipse* (153)	14%	39%	22%	14%	11%	1%
Hadoop* (152)	14%	39%	21%	14%	11%	1%
Vaadin* (150)	15%	38%	21%	14%	11%	1%
Guava (19)	26%	21%	11%	26%	16%	
Guice (19)	26%	21%	11%	26%	16%	
Numpy (111)	10%	48%	29%	11%		1%
Google (59)	15%	42%	12%	14%	17%	
PEP8/257 (55)	27%	38%	9%	15%	11%	
Pandas* (213)	9%	42%	30%	7%	10%	3%
iPython* (170)	18%	44%	22%	12%		5%
Pytorch* (102)	21%	40%	11%	18%	11%	
Django* (66)	30%	36%	9%	12%	12%	
Requests* (66)	30%	33%	11%	12%	14%	
Pipenv (55)	27%	38%	9%	15%	11%	
Mailpile (55)	27%	38%	9%	15%	11%	

Content rules are more prevalent in style guidelines but hard to locate

13

RQ2: Do developers follow conventions?

Legend: Followed rules — Not followed rules — Not applicable rules

Java

Project	Followed rules	Not followed rules	Not applicable rules
Guice	53%	2%	45%
Guava	63%	3%	35%
Hadoop	24%	2%	74%
Spark	25%	2%	73%
Vaadin	32%	3%	65%
Eclipse	27%	3%	70%

Python

Project	Followed rules	Not followed rules	Not applicable rules
Requests	38%	12%	50%
Pytorch	49%	10%	42%
Pipenv	54%	6%	40%
Pandas	11%	3%	86%
Mailpile	47%	10%	43%
iPython	16%	2%	83%
Django	49%	7%	44%
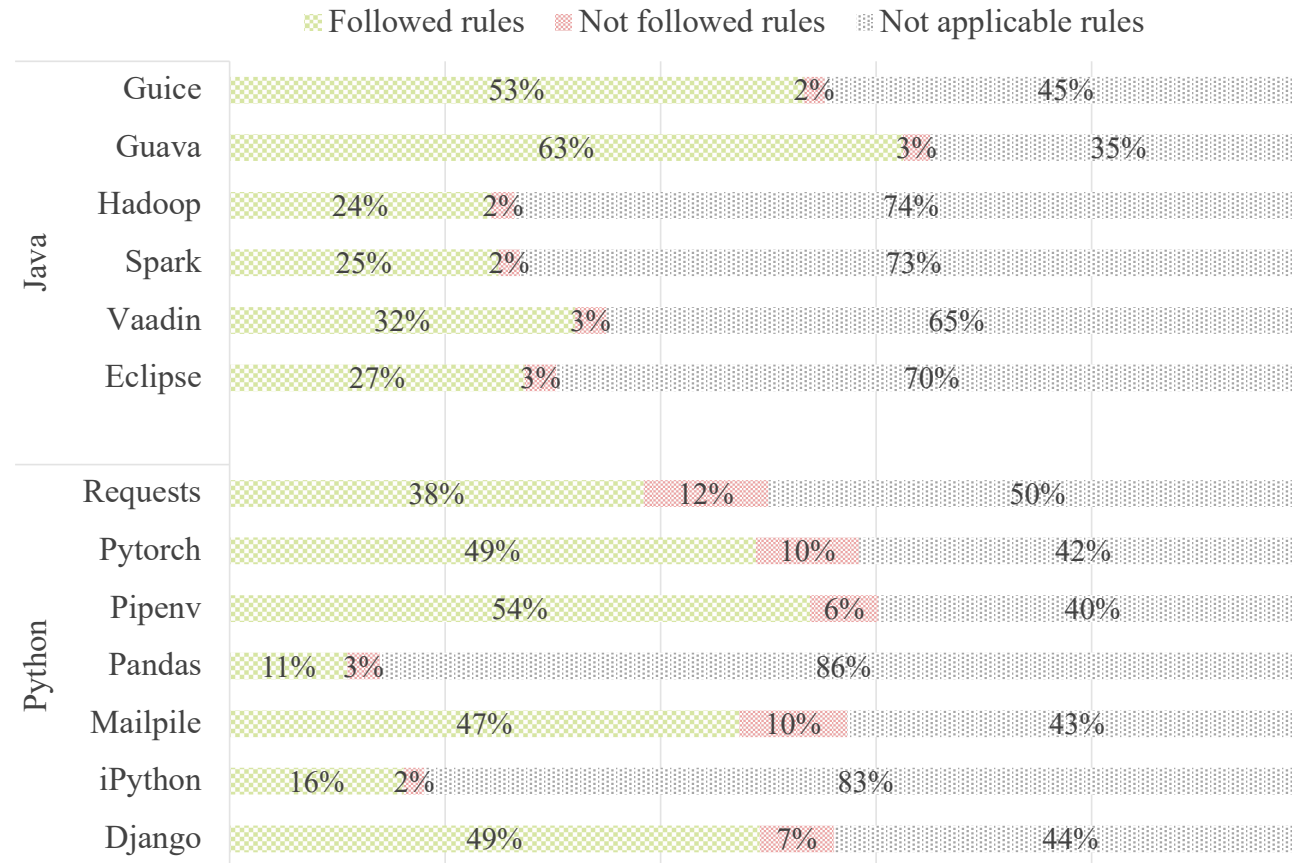
Developers do not always adopt comment conventions
(**Not applicable rules**)

RQ2: Do developers follow conventions?

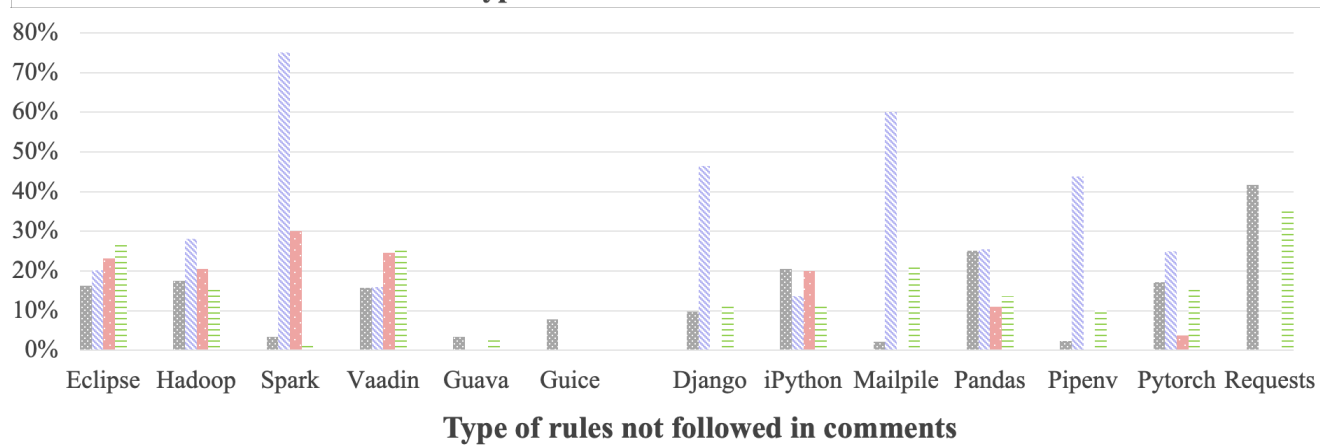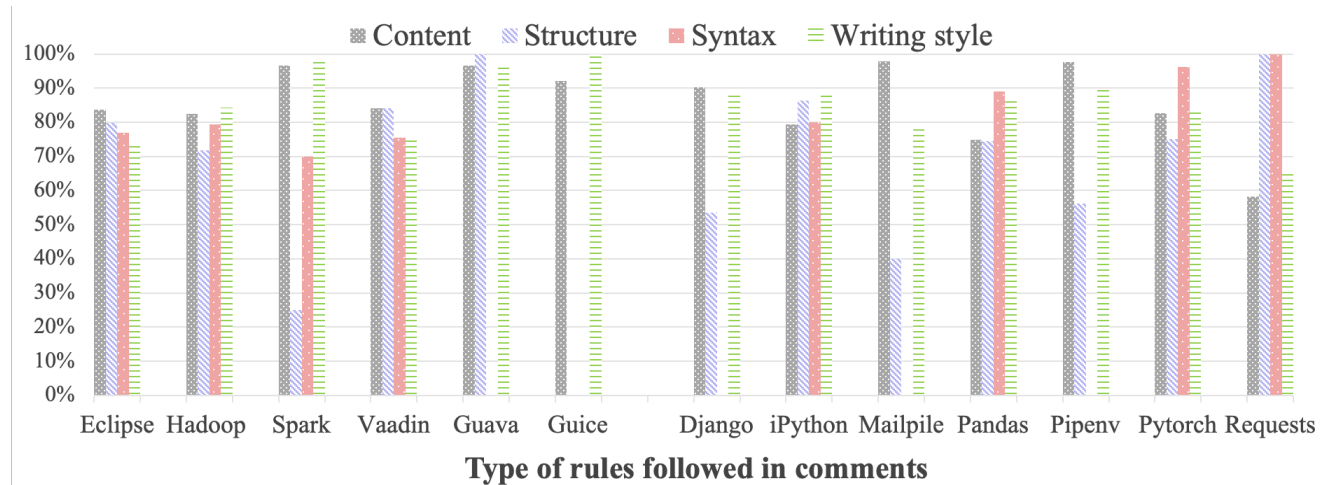```
/** @author Vaadin Ltd
  *
  * @since 7.7.13
  */
```

Not followed rules

Not applicable rules

🔴 First sentence should summarise the class

⚪ Use phrases instead of complete sentences

⚪ Use third person instead of second person

| | Followed rules | Not followed rules | Not applicable rules |
|---|---|---|---|
| **Java** | | | |
| Guice | 53% | 2% | 45% |
| Guava | 63% | 3% | 35% |
| Hadoop | 24% | 2% | 74% |
| Spark | 25% | 2% | 73% |
| Vaadin | 32% | 3% | 65% |
| Eclipse | 27% | 3% | 70% |
| **Python** | | | |
| Requests | 38% | 12% | 50% |
| Pytorch | 49% | 10% | 42% |
| Pipenv | 54% | 6% | 40% |
| Pandas | 11% | 3% | 86% |
| Mailpile | 47% | 10% | 43% |
| iPython | 16% | 2% | 83% |
| Django | 49% | 7% | 44% |

# RQ2: Do developers follow conventions?



Type of rules followed in comments

Type of rules not followed in comments

Comments often follow **writing style** and **content rules** than other rule types

# Future work

Verify other types of comments (Method, inline comments)

Verify comments of other languages (C++, JavaScript)

Develop tools to validate comments against the guidelines

Improve comment quality assessment

# Do Comments follow Commenting Conventions? A Case Study in Java and Python

Preprint on Arxiv

https://arxiv.org/pdf/2108.10766v2

 Video on YouTube

https://youtu.be/mX_9XxQTSxQ

Replication Package on Zenodo

https://doi.org/10.5281/zenodo.5296443

Contact us

$u^b$