Seminar in Software Engineering

T16: Documentation practices of developers in Polyglot environments

Student: Irakli Kelbakiani

Supervisor: Pooja Rani

3rd June 2022

Documentation practices of developers in Polyglot environments

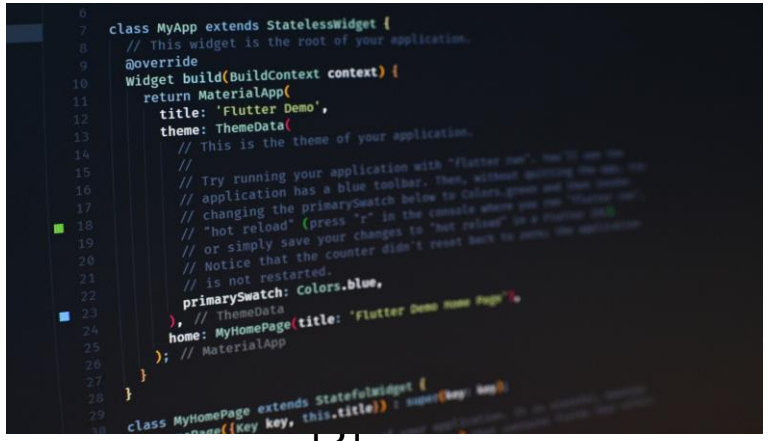
- Introduction
- Project of interest
- Javadoc Documentation
- PEP8 Documentation
- Challenges
- What is WoC and how can it help us?
- Vaadin developers' analysis.
- Future Goals

Introduction

Human Readable Descriptions inside of computer programs

Good for Internal documentation.

Comments complement the coding style, not replace it



[3]

[1]

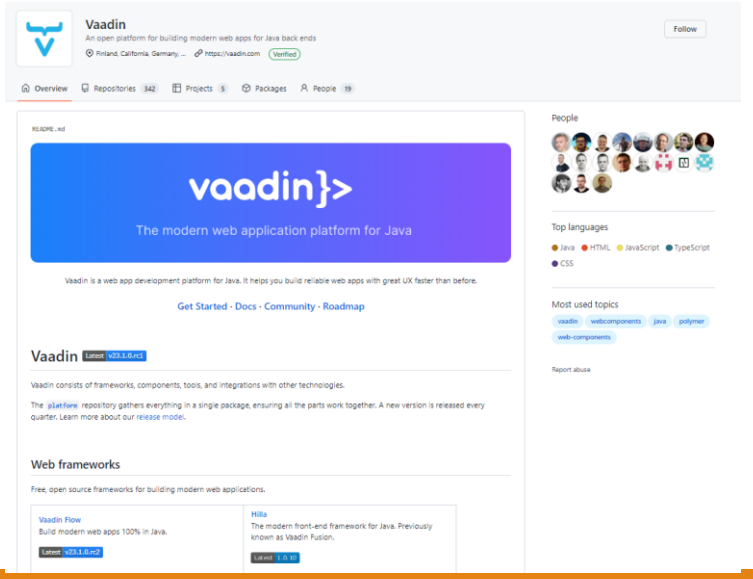
Comments to be investigated

Class comments

Method comments

Content related comments

Java and Python conventions respected?



Javadoc Documentation

```
/**
 * Short one line description.                        (1)
 * <p>
 * Longer description. If there were any, it would be    (2)
 * here.
 * <p>
 * And even more explanations to follow in consecutive
 * paragraphs separated by HTML paragraph breaks.
 *
 * @param  variable Description text text text.          (3)
 * @return Description text text text.
 */
public int methodName (...) {
    // method body with a return statement
}
```

```
// import statements

/**
  * @author         Firstname Lastnam
  * @version        1.6
  * @since          1.2              (the
  */
public class Test {
    // class body
}
```

Tags	Usage
{@author}	Adds the author of a class.
{@docRoot}	Represents the relative path to the generated document's root directory from any generated page.
@since	Adds a comment indicating that this API should no longer be used.
@exception	Adds a **Throws** with the classname and description text.
{@inheritDoc}	Takes description from overridden method
{@link}	Link to other symbol
@see	See Also
@param	Adds a parameter with the specified parameter-name followed by the specified description to the "Parameters" section.
@return	Adds a "Returns" section with the description text.
….	….

PEP8 Documentation

```
class OneHotCategorical:
    """
    Creates a one-hot categorical distribution parameterized by :attr:`probs` or
    :attr:`logits`.

    Samples are one-hot coded vectors of size ``probs.size(-1)``.

    .. note:: :attr:`probs` must be non-negative, finite and have a non-zero sum,
              and it will be normalized to sum to 1.

    See also: :func:`torch.distributions.Categorical` for specifications of
    :attr:`probs` and :attr:`logits`.

    Example::

        >>> m = OneHotCategorical(torch.tensor([ 0.25, 0.25, 0.25, 0.25 ]))
        >>> m.sample()  # equal probability of 0, 1, 2, 3
        tensor([ 0.,  0.,  0.,  1.])

    Args:
        probs (Tensor): event probabilitie
        logits (Tensor): event log probabilities
    """
```

```
def function(self, arg1, arg2, *args, **kwargs):
    """

    :param self:
    :param arg1:
    :param arg2:
    :param args:
    :param kwargs:
    :return:
    """
```

Comment identification Challenges

Developers use multiple email addresses

Multiple names and the different combinations

How to search all the repositories user committed to?

Should all those repositories be cloned?

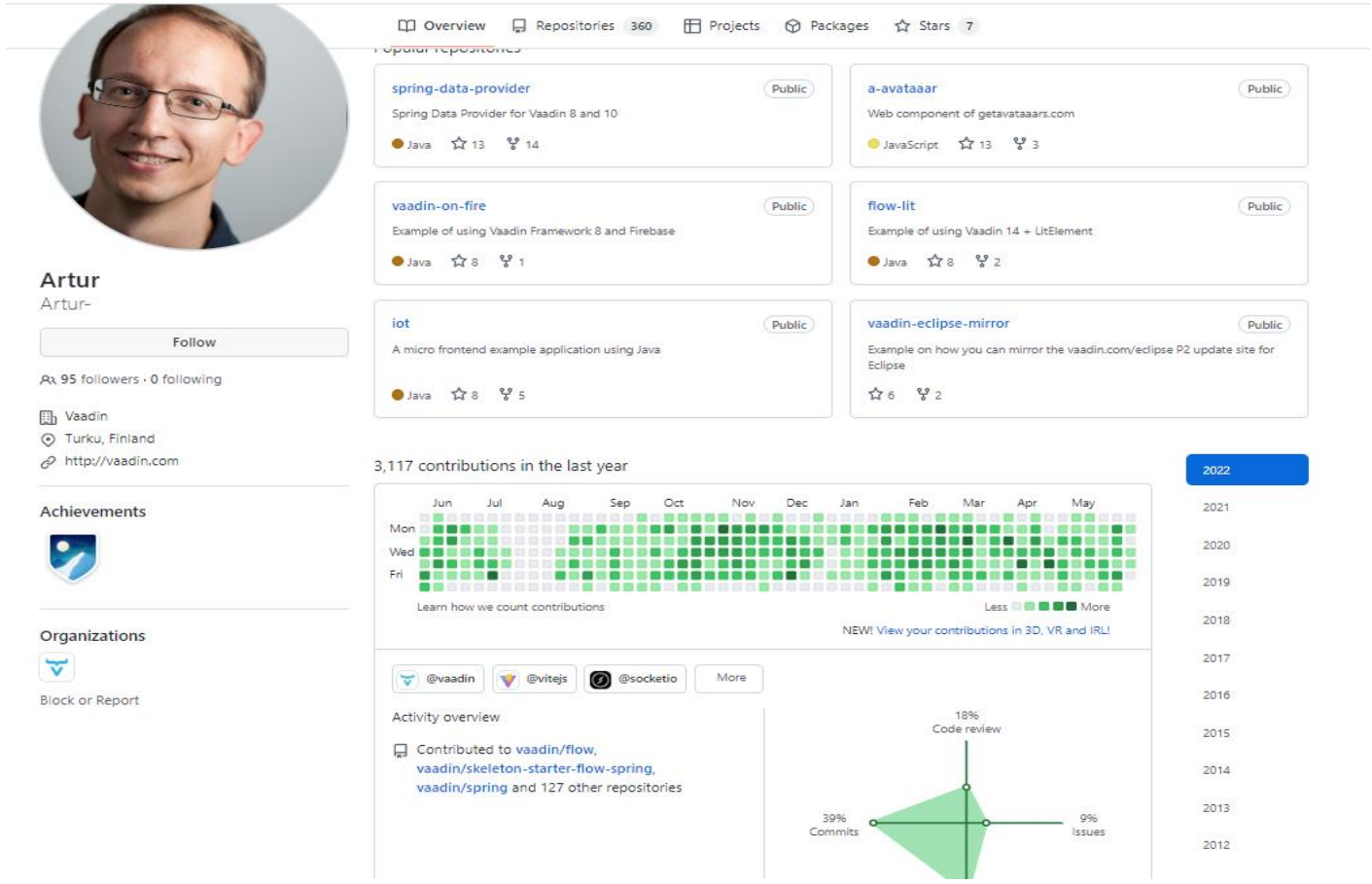
GHttorrent 1 month database dump around 101 GB

How to mine data?

GHTorrent

GHArchive

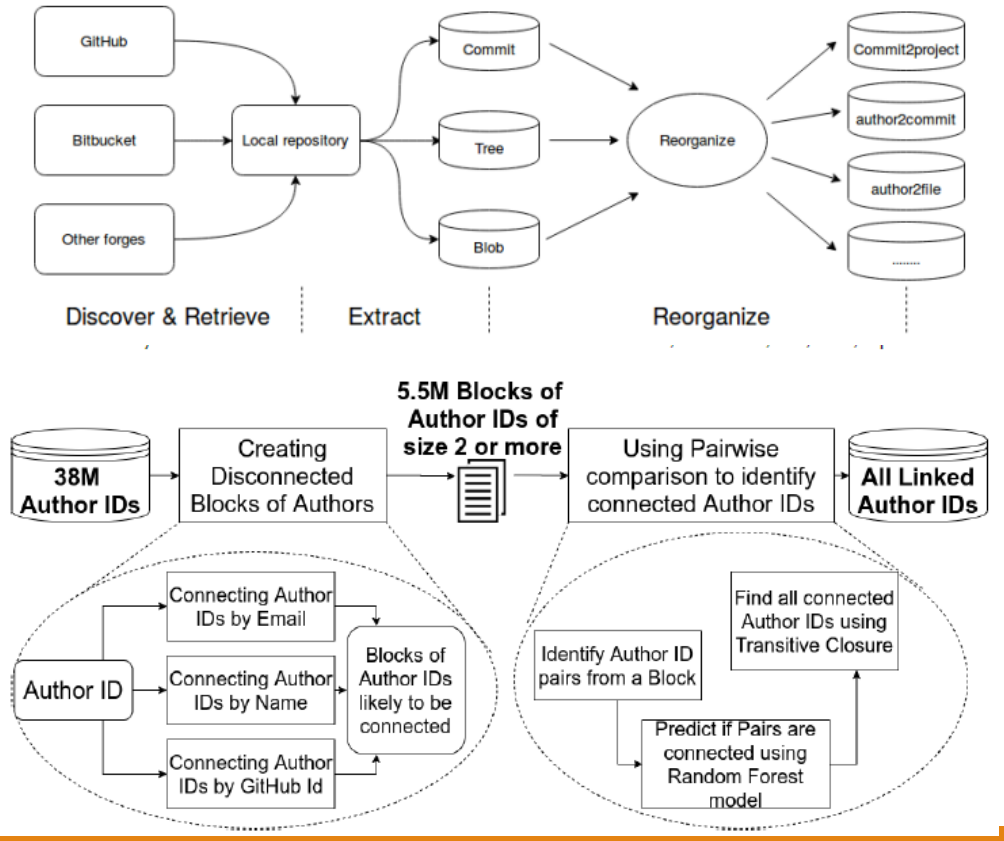
Selenium Program



What is WoC and how can it help us?

World of Code: An Infrastructure for Mining the Universe of Open Source VCS Data

- FLOSS ecosystem developed at University of Tennessee
- Keeps large datasets up-to-date
- Access to 40 TB data
- 12B git objects



Vaadin Contributors

Vaadin Top Contributors (185 unique contributors out of 232):

- 2814 Artur Signell <artur@vaadin.com>
- 2121Artur Signell artur.signell@itmill.com (same person as above)
- 2082 Matti Tahvonen <matti.tahvonen@itmill.com>
- 1964 Leif Åstrand <leif@vaadin.com>
- 883 Teemu Suo-Anttila <teemusa@vaadin.com>

Contributor	Vaadin Commits	Git Commits	Modified Java files	Modified Python files
Artur Signell	4935	21382	94709	13
Matti Tahvonen	2144	12308	9500	3
Leif Astrand	2071	7633	42525	67
Henri Sara	1572	5415	59418	23

Git top contributors

Name	Commits	Python Commits Processed	Java Commits Processed
Kewin Sawicki	44226	1088	10325

Future Goals

Restructure extracted 16GB data

Separate class, method and other comments

Create type models

Make a comparison

Evaluate

References

[1] https://sites.google.com/site/unclebobconsultingllc/

[2] https://medium.com/swlh/10-things-every-programmer-should-know-26ba37cfcaf4

[3]https://www.earthdatascience.org/courses/intro-to-earth-data-science/write-efficient-python-code/intro-to-clean-code/python-pep-8-style-guide/

[4] https://velog.io/@jongbeen_song/Version-Control-with-Git

[5]https://www.researchgate.net/figure/A-class-comment-in-Python_fig2_353364930

[6] https://github.com/woc-hack/tutorial

[7]https://sphinxcontrib-napoleon.readthedocs.io/en/latest/example_numpy.html

[8] A Dataset and an Approach for Identity Resolution of 38 Million Author IDs extracted from 2B Git Commits. Tanner Fry, Tapajit Dey, Andrey Karnauch, Audris Mockus

[9] https://github.com/KelbakianiIrakli/SW-seminar

[10] World of Code: An Infrastructure for Mining the Universe of Open Source VCS Data Yuxing Ma, Chris Bogart, Sadika Amreen, Russell Zaretzki, Audris Mockus

Questions?