Recommending Log Placement at Method-Level Based on Code Vocabulary

1. Problem

Logging is a common programming practice that enables developers to collect runtime information from a system.

Logging is **important** in order to:

- Debug system failures.
- Monitor a system's performance.

Deciding which parts of the code to log is a hard and manual process for which there are no standard practices in the industry.

2. Approach

In this research, we trained machine learning classifiers to predict log placement at method level, using only the **Code Vocabulary** as training data.

We treated Log Placement as a **Binary Document Classification** problem.

Each method is considered to be a document created from the Code Vocabulary.

3. Research Questions

RQ1: What is the performance of a Machine Learning model based on bag-of-words for log placement at method level?

RQ2: What value do different words add to the classifier?

RQ3: How does the size of a method affect the performance of the trained classifier?

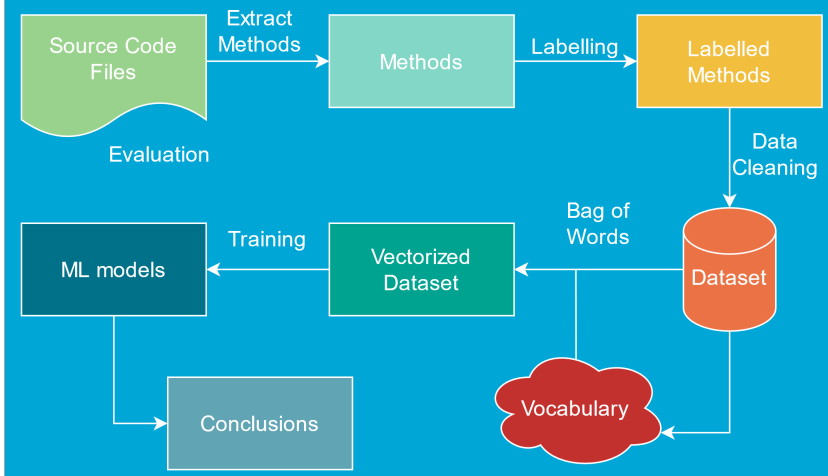
Contact

Konstantinos Lyrakis: k.lyrakis@student.tudelft.nl

Supervisor: Jeanderson Cândido

Responsible Professor: Dr. Maurício Aniche

4. Methodology



6. Conclusions

Code Vocabulary is a great data source that can be solely used to predict log placement at method level.

Classifiers trained solely on the Code Vocabulary are hard to interpret, as most words provide the same amount of information.

The performance of a classifier increases when it has to classify bigger methods, but they perform well even with small methods.

TU Delft

5. Results