

## Automatic Generation of Clone References

### **Background**

Code clones are separate fragments of code that are similar in their structures or semantics. Code clones can occur frequently in an application under development with different developers or various versions. Since the code clones can be propagated in various parts of software products, attending new changes such as security updates or a defect repair should be considered in all of the source code. For solving the problem of clone detection, there are various clone detection tools implemented by academic and commercial organizations.

In order to evaluate the performance and accuracy of existing and new clone detection tools, various clone benchmarking tools are developed. The benchmarking tools contain a repository of clone references to test the functionality and accuracy of clone detection tools. The main limitations of existing clone benchmarks are their dependency on the creators as well as the limited types of their clone references.

### **Task**

The goal of this thesis is to propose and implement a novel approach to create code clone references without any human interference. As a part of this thesis, a literature review of the existing benchmarking tool has to be performed. Machine learning techniques or statistical approaches among various open source software repositories are some of the that the students need to investigate.

The evaluation of the proposed approach consists of testing the generated clone references using some well know clone detection tools and reporting the comparison of the results towards existing clone detection benchmarking frameworks.

The development of the solution idea can be supported by the supervisor.

A report of about 70-90 pages in English or German which contains the concepts, strategies, and results is mandatory.

### **References**

[1] Yuki, Yusuke, et al. "Generating clone references with less human subjectivity." Program Comprehension (ICPC), 2016 IEEE 24th International Conference on. IEEE, 2016.

[2] Svajlenko, Jeffrey, et al. "Towards a Big Data Curated Benchmark of Inter-Project Code Clones." Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on. IEEE, 2014

### **Organisatorisches**

**Supervisor:** Dr. Javad Ghofrani, javad.ghofrani@inf.uni-hannover.de, Raum G308

**Reviewer:** Prof. Dr. Schneider

**Start:** from now