OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG | INF | FAKULTÄT FÜR INFORMATIK

Master Thesis

Supervisor: Dr.-Ing. Sandro Schulze

Highly-Configurable Software Systems enable the development of a set of related variants by using variability implementation mechanisms such as plug-ins, frameworks, module systems, or annotations. A prominent example for annotations is the C preprocessor CPP, frequently used in the C programming language. While it allows for a flexible customization of the final software system (by excluding/including code fragments for compilation), the CPP is often criticized for obfuscating the actual source code, thus, making evolution and maintenance a tedious and error-prone task. In this thesis, the task is to employ machine learning techniques to verify, whether the notion of variability-aware code metrics can establish a relation between badly used preprocessor annotations and code anomalies such as bugs or vulnerabilities.

A Feasibility Study on Bug Prediction Using Variability-Aware Code Metrics and Machine Learning

In recent work, variability-aware code smells have been proposed as a metric-based approach to identify misuse of CPP annotations. This MSc thesis should now employ different machine learning techniques to analyze whether and which of these the metrics, used for defining a smell, correlate with other anomalies (e.g, bugs, vulnerabilities), and thus, can serve as a predictor for future anomalies.

Main Tasks:

- Use and extend an existing set variability-aware code metrics
- Review machine learning techniques and their application in defect prediction
- Choose two techniques and implement a protoype predictor on top of the metrics
- Evaluate your implementation (using open-source systems)

Requirements:

- good programming skills
- quick grasp of subject matter, strong work ethic, work on you on initiative (with guidance by the supervisor)
- background in machine learning is a plus, but not required (can be obtained during MSc thesis)