Master Thesis Description, status: finished, language: EN
Keywords: code review, data mining, python, quantitative data analysis

Exploratory Data Analysis of Code Review Data

Summary

Code reviews are an essential practice to increase code quality. We collected records of all reviews performed at a large software companies and successful open source projects. The records contain information on millions of reviews performed by thousands of developers. In this thesis, the student will perform an exploratory data analysis of a given sample of review data. The student will use the remaining data to cross-validate selected hypotheses / theories resulting from the exploratory data analysis of the initial sample. Knowledge in Python, data mining, statistics, and handling large data sets are a plus.

Details

- Literature review
  - Exploratory data analysis (e.g. Tukey et al.)
  - Code review processes, practices
  - Qualitative and quantitative studies on code review
- Method
  - Exploratory data analysis following an established method
- Results
  - Analysis scripts written in Python, using Pandas and numpy
  - Report on exploratory analysis
    - Descriptive overview of the available data set
    - Discussion of the findings; potentially resulting hypotheses / theories

Supervisor

Michael Dorner, michael.dorner@fau.de
Prof. Dr. Dirk Riehle, dirk.riehle@fau.de

Open Source Research Group
Computer Science Department
Friedrich-Alexander University

More information: http://osr.cs.fau.de/theses/resources/