

Measuring Patch-Flow in the Real World

Summary

With our partners, we want to kick-off inner source (utilization of open source development practices within an organization) within a software product line. In preparation, we want to measure how much collaboration on software code already happens within the product line by measuring the patch-flow (flow of patches across intra-organizational boundaries as project or organizational unit boundaries).

In this thesis, the student analyzes whether existing systems (org. or code repositories, ...) deliver required data for measuring patch-flow, extends our patch-flow crawler to read their data, and measures the patch-flow. If data from org. repositories is insufficient, the student utilizes (and validates) an already theorized clustering algorithm to augment the data from the organizational repositories.

Work Results

- Literature review
 - Metrics and measurements to quantify collaboration based on source code
- Contributions / Research questions
 - Do existing systems deliver sufficient data for measuring the patch-flow? Which systems? What attributes are missing?
 - How much code collaboration is already happening within the selected product line?
 - Exploration of patch-flow graphs (structure, prevalence of patching)
- Work results
 - Evaluation of possible data sources, their APIs, missing attributes
 - Adapted our pre-existing Java patch-flow crawler for use
 - Delivered patch-flow graph and evaluated hypothesis: “Patch-flow exists”
 - (If necessary: adapted and validated our theoretical clustering algorithm to augment org. data)

Supervisor

Maximilian Capraro, maximilian.capraro@fau.de

Prof. Dr. Dirk Riehle, dirk.riehle@fau.de

Open Source Research Group, Computer Science Department, Friedrich-Alexander University

Changes to the topic

Due to medical absence of the student, the project moved forward without him. Some work results have already been realized. We modify the thesis topic as follows so that the student can build on top of the reading and work that he has already performed.

Instead of measuring (and preliminarily analyzing) the patch-flow data, the student will evaluate the fitness for use of the now already measured patch-flow data.

- Modified contributions / research questions
 - What's the quality of the measured data? How fit for patch-flow analysis is the measured data?
- Work results
 - Strategy to evaluate patch-flow data's quality / fitness for use (including metrics and visualizations)
 - Prototypical implementation of the metrics on top of the collected data (e.g. using R, Libre Office Calc, or other generic tooling)
 - Evaluation of problems with the crawled data