

Black box investigation of the Ohloh data source for OSS research

This Master Thesis investigates the reliability of the data provided by Ohloh for Open Source Research. Ohloh provides a directory of Open Source projects and analytics of the project code bases. The thesis employs various methods to validate the quality of the provided data for large-scale analysis work.

Work results

- Literature review - what methods are already known
- Short summary of the research by
 - Wolfgang Mauerer
 - Gottfried Hofmann
- Research methods employed
- Which dimensions and timespans of the data are likely to be valid
- In case of problematic areas
 - Identify them
 - Possible source of the problems
- Feedback - Talk with Ohloh
 - In case of errors report to Ohloh
 - Propose solutions
- Create a pipes-and-filters architecture to gather clean data

Thesis Advisor

Prof. Dr. Dirk Riehle

Friedrich-Alexander-Universität Erlangen-Nürnberg

dirk.riehle@fau.de, <http://osr.cs.fau.de>