RUHR UNIVERSITÄT BOCHUM RUB

Bachelor/Master Thesis Topic

Automated Extraction of Test Oracles with Large Language Models

Motivation and Background:
Automated program repair (APR) [1] is driven by the availability of test cases to validate the auto-generated patches. Evolutionary testing for APR [2] tries to mitigate the issue of overfitting existing test cases by generating new ones. Part of its proposed workflow is an evolutionary process that co-evolves patches and tests. The test generation is based on EvoSuite [3] and uses manually extracted test oracles based on reports from issue trackers. With the power of Large Language Models (LLM), this project explores the automated extraction of test oracles, i.e., test assertions, which can help to fully automate the process of program repair.

Student Task and Responsibilities:

- Make yourself familiar with EvoRepair [2] and the Defects4J benchmark (https://github.com/rjust/defects4j).
- Systematically explore the usage of various LLMs (commercial and open) for the purpose of test oracle extraction based on the available bug issues in the benchmark.
- Combine your approach with EvoRepair for the purpose of automated program repair.
- Curate a dataset with new programs to validate your workflow outside Defects4J. Ideally, this step helps to mitigate the issue of data leakage.
- Design/select evaluation metrics and conduct a thorough evaluation of your approach.
- Analyze the results and document your findings.

Deliverables:

- Extension of EvoRepair that uses LLMs for test oracle extraction
- Evaluation artifacts (dataset, tools, etc.)
- Documented findings of the conducted experiments

Pre-Requisites: (Programming Languages, OS, Skills, Papers, etc)
Knowledge in Java and LLM prompt engineering is helpful for this project.

[1] C. Le Goues, M. Pradel, A. Roychoudhury and S. Chandra, "Automatic Program Repair," in *IEEE Software*, vol. 38, no. 4, pp. 22-27, July-Aug. 2021. https://doi.org/10.1109/MS.2021.3072577

[2] H. Ruan, H. L. Nguyen, R. Shariffdeen, Y. Noller and A. Roychoudhury, "Evolutionary Testing for Program Repair," *2024 IEEE Conference on Software Testing, Verification and Validation (ICST)*, Toronto, ON, Canada, 2024, pp. 105-116. https://doi.org/10.1109/ICST60714.2024.00058

[3] G. Fraser and A. Arcuri. 2011. EvoSuite: automatic test suite generation for object-oriented software. In Proceedings of the ESEC/FSE 2011. ACM, NY, USA, 416–419. https://doi.org/10.1145/2025113.2025179

Contacts
Prof. Dr. Yannic Noller (sq-office@rub.de)
Software Quality group, Faculty of Computer Science, Ruhr University of Bochum