JOHANNES KEPLER
UNIVERSITÄT LINZ

o.Univ.-Prof. Dr.
Hanspeter Mössenböck
Institute for System Software

T +43 732 2468 4340
F +43 732 2468 4345
hanspeter.moessenboeck@jku.at

Secretary:
Karin Gusenbauer
Ext 4342
karin.gusenbauer@jku.at

Master Thesis

Automated Searching and Analyzing Open Source Projects

Student:

Advisor: DI Lukas Makor

Start date: xx.xx.xxxx

Research projects need to be validated based on independent benchmarks. Research in the area of compiler construction and runtimes often tries to improve the speed of programs written by developers. Typically, certain standard benchmarks are used to evaluate program optimizations. However, these benchmarks are often not suitable for the evaluation of optimizations that target specific use-cases. Therefore, utilizing real-world projects for evaluation might be more promising.

Source code hosting platforms such as GitHub, GitLab or BitBucket have fostered open source development over the years. Hence, they might be a vital resource to evaluate program optimizations on practically relevant projects. However, due to the vast number of repositories, manually searching for suitable projects in these repositories is infeasible.

This thesis focuses on implementing an application that allows automatic searching of open source projects in software repositories. Via filters, popular and frequently used projects written in specific programming languages should be prioritized, while inactive or outdated projects should be excluded. In a second step, commonalities between projects should be identified to further narrow down the investigated scope. This could include common project structures (e.g., Node.js projects with a package.json file) or filtering for project dependencies (e.g., projects that use Lodash). For the evaluation, a number of relevant projects should be identified via such criteria that can be used as potential benchmark workloads. Hence, it is required to provide sufficiently large input data sets as well as scripts to execute each application.

The goals of this project are:

- Create a list of potentially applicable projects by searching source code hosting platforms (GitHub search should be prioritized)
- Create a document describing common patterns in these projects
- For a subset of the projects, provide a script to start the program, initialize it with data and trigger the relevant operations

Explicit non-goals are:

- Fully automatic detection or verification whether projects are indeed executable

Modalities

The progress of the project should be discussed at least every two weeks with the advisor. A time schedule and a milestone plan must be set up within the first 3 weeks. It should be continuously refined and monitored to make sure that the thesis will be completed in time. The final version of the thesis must be submitted not later than xx.xx.xxxx.

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenberger Straße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696