



Source code analysis of big package repositories

Bachelor thesis for Fabian Ferihumer

Many programming environments are supported by huge repositories of packages available for download. New implementations of programming languages should routinely check that they can correctly load and execute code in these repositories. The goal of this project is to develop a tool that will download (new) packages from the CRAN repository and perform some basic analysis on it using the FastR [1] runtime. The output of the tool should be a webpage that provides a summary of the analysis results.

The scope of this thesis is as follows:

- Develop a tool that downloads CRAN packages, and keeps a local database with a history of package updates so that the set of packages at a specific date can be reused
- Code that unpacks and analyzes all or specific packages:
 - Types of source files
 - Presence of specific patterns in the source code (configured as regex)
 - Success of parsing R source with FastR, successful loading, etc.
- The result of this analysis should be output as a static webpage that provides a summary of all packages (+ maybe detail pages for each package)
- If time permits: visualize deltas to specific previous version results

The work's progress should be discussed with the supervisor at least every 2 weeks. Please note the guidelines of the Institute for System Software when preparing the written thesis.

The project will start 2015/11/01. Please note the regulations of the Institute for System Software concerning maximum duration of bachelor theses.

Supervisor: Dipl.Ing. Dr. Lukas Stadler, Dipl.Ing. Matthias Grimmer

[1] <https://bitbucket.org/allr/fastr>