A Large Scale Empirical Study of the Relationship Between Build Technology and Build Maintenance

Summary

The authors proposed a large scale empirical study on build tools used on GitHub to examine build tools. The authors collected the majority of the data set from GitHub projects. The authors filtered out projects that's too small by number of source files, has low development activity and contains too much unknown project files. This resulted in 261k projects. By analyzing the data set, the authors found out that language-specific technologies capture more market share; there exists a high correlation between programming language and build tool; large projects are more likely to use newer technologies earlier than smaller projects; framework driven build tools have more features but potentially needs more maintenance, their changes also tightly couple with source code changes.

Things I would like to see discussed

- The study ignores the quality of the build system as a factor when doing statistical analysis. Does these factor play a huge role in development?

- How are projects with multiple build tool considered when doing analysis? RQ1 mentions it but I don't quite understand when doing other analysis in other RQs.

- I remember the reason that `make` is not used now for java is because the structure of java program does not suits `make` very well. Based on the graph in the paper it shows that no build tools correlates with two different class of languages. Can we safely assume that build tools are most likely to be designed for a specific type of languages?