

cHash: Erkennung redundanter Übersetzeraufrufe durch AST Hashing

Softwareprojekte, die mittels eines Übersetzers ins Maschinenformat gebracht werden, erleben im Laufe ihrer Lebensspanne mehrere tausend Iterationen des Übersetzungsprozesses. Der Übersetzer wird immer wieder auf eine Quelltextbasis angewendet, die sich jedoch häufig nur in Details geändert hat. Viele dieser Aufrufe sind jedoch nicht nötig, da die Änderung am Quelltext sich nicht unbedingt in einem sich geänderten Binärartefakt niederschlagen.

Um die Auswirkungen solcher redundanten Aufrufe zu vermindern, haben wir cHash entwickelt. cHash arbeitet innerhalb des Übersetzers, nach der Parse-Phase und berechnet eine Prüfsumme über den abstrakten Syntaxbaum. Diese Prüfsumme ändert sich mindestens immer dann, wenn das Übersetzungsergebnis sich auch ändern wird. Bleibt die Prüfsumme unverändert zum vorherigen Aufruf, kann der Übersetzer einfach an dieser Stelle, vor den zeitintensiven Optimierungen, beendet werden. Die Prüfsumme über den abstrakten Syntaxbaum erlaubt nicht nur die Erkennung von rein syntaktischen Änderungen, sondern auch Sprachelemente, die die aktuell vorliegende nicht beeinflussen, können ignoriert werden, wie beispielsweise unbenutzte Typedeklarationen.