

Bachelorarbeit in Informatik

Erkennung von generiertem Code in Open-Source Software

Der Quelltext von Software lässt sich bezüglich seiner Rolle in der Wartung in Kategorien unterteilen. Neben Quelltext, der von Hand geschrieben und gewartet wird, sind die größten Kategorien generierter Code und Test-Code. In vielen Systemen fallen bis zu 50% des Quelltexts in diese Kategorien.

Statische Analysen erkennen Qualitätsprobleme in Quelltext. Dabei ist die Code-Kategorie maßgeblich für die Relevanz der untersuchten Qualitätskriterien. So können Sicherheitslücken oder Performanceprobleme, die in Produktionscode entscheidend sind, in generiertem Code irrelevant sein, da er während der Wartung nicht direkt editiert wird. Um die Relevanz und Aussagekraft der Ergebnisse statischer Analysen zu erhöhen, muss daher die Kategorie des untersuchten Quelltexts berücksichtigt werden.

Das gilt insbesondere bei der Erstellung von Benchmarks, die Häufigkeiten des Auftretens von Qualitätsdefekten und Verteilungen von Metrik-Werten über eine Vielzahl von Projekten ermitteln. Da eine manuelle Klassifikation von Quelltext bei einer Vielzahl von Projekten in der Praxis nicht durchführbar ist, ist hier ein automatischer Ansatz notwendig.

Ziel der Arbeit ist die Konzeption und prototypische Umsetzung der automatischen Erkennung von generiertem Code. Sie umfasst folgende Schritte:

- Prototypische Implementierung von Heuristiken zur Erkennung von generiertem Code. Für die Erkennung von generiertem Code sollen Heuristiken eingesetzt werden, die u.a. Clone-Detection Techniken auf aus dem Code extrahierte Kommentare anwenden.
- Ableitung einer Liste von Heuristiken mit der generierter Code aus verschiedenen Generatoren anhand der Kommentare erkannt werden kann („Generator-Pattern Repository“).
- Evaluierung der Vollständigkeit und Genauigkeit der entwickelten Heuristik gegen einen Referenzdatensatz
- Automatische Klassifikation von Quelltext in einer großen Sammlung von Open Source Systemen und Evaluierung des Anteils von generiertem Code

Voraussetzungen:

- Kenntnisse in Java

Anzahl der Bearbeiter: 1

Aufgabensteller:

Prof. Dr. Dr. h.c. Manfred Broy
Technische Universität München
Institut für Informatik
Software & Systems Engineering

Betreuer/weitere Information:

Elmar Jürgens
TU München, Lehrstuhl Broy