

# Clustering von Aufrufgraphen für die hierarchische Fehlerlokalisierung

Verschiedene Arbeiten am IPD beschäftigen sich mit der Lokalisierung von Fehlern in Software unter Verwendung von Data-Mining-Techniken. Dazu werden Aufrufgraphen des untersuchten Programmes generiert und untersucht. Unter anderem kommen Frequent-Subgraph-Mining-Algorithmen zum Einsatz. Diese sind NP-vollständig und daher nur für Graphen beschränkter Größe einsetzbar. Um dennoch die Fehlerlokalisierung in großen Softwareprojekten zu ermöglichen, haben wir am IPD ein hierarchisches Analyseverfahren entwickelt. Dabei werden die Aufrufgraphen auf Methodenebene aufgezeichnet, für die Analyse werden die Graphen jedoch auf verschiedenen Hierarchieebenen zusammengefasst, und diese Zusammenfassungen werden analysiert.

Bisher stützt sich die Aggregation der Graphen auf eine in Programmiersprachen vorhandene Hierarchie, nämlich Pakete, Klassen und Methoden. Oft ist es aber so, dass beispielsweise die Zuordnung einer Klasse zu einem Paket willkürlich ist. Um robuste und flexible Zusammenfassungen zu finden, sollen in dieser Studienarbeit verschiedene Clustering/Partitionierungs-Algorithmen hierfür untersucht und verglichen werden.

Das Software-Projekt *Rhino* kommt in einer früheren Arbeit erfolgreich zum Einsatz, um ein Verfahren zur Fehlerlokalisierung zu evaluieren. Die dabei generierten Daten sowie das genutzte Framework stehen zur Verfügung, um eine auf Graph-Clustering oder Graph-Partitionierung basierte Zusammenfassung zu evaluieren. Dazu sind im Einzelnen folgende Schritte zu leisten:

1. Es sind geeignete Algorithmen auszuwählen. Durch Literaturrecherche sollen Algorithmen identifiziert werden, die geeignet erscheinen.
2. Verschiedene Algorithmen sollen implementiert werden und ggf. angepasst werden, oder eine existierende Implementierung für die vorliegenden Daten nutzbar gemacht werden.
3. Die in 1.) ausgewählten Algorithmen sollen auf den vorhandenen Daten angewendet werden. Auch das Zusammenfassen in mehreren Schritten, durch wiederholte Anwendung eines Algorithmus, soll in Betracht gezogen werden.
4. Die aggregierten Graphen sollen als Eingabe für ein vorhandenes Framework zur Fehlerlokalisierung dienen. Das Framework ist dazu anzupassen. Die erzielten Ergebnisse die sollen mit bisherigen Ergebnissen verglichen und interpretiert werden.

## Ansprechpartner

Dipl. Inform. Christopher Oßner    ossner@kit.edu    +49 721 608 44065    Raum: 340

Am Fasanengarten 5    76131 Karlsruhe    Gebäude: 50.34