



Bachelor-, Master- und Doktorandenseminar  
des Instituts für Informatik

## Einsatz von Data- und Text-Mining- Techniken zur Analyse von Quellcode

Oliver Fox, TU Clausthal

Die Größe und Komplexität von Software-Systemen nimmt seit Jahren kontinuierlich zu. Zum einen werden immer höhere Anforderungen an Software-Systeme gestellt, was oftmals komplexere Systeme erfordert, zum anderen nimmt die Komplexität des eigentlichen Entwicklungsprozesses und Lebenszyklus von Software immer weiter zu, da Software meist verteilt entwickelt und über Jahre gewartet und weiterentwickelt wird.

Häufig stehen daher an der Entwicklung solcher Systeme beteiligte Architekten oder Programmierer vor der Aufgabe, einen existierenden Systemteil analysieren und verstehen zu müssen. Fehlt eine adäquate Dokumentation des betrachteten Systemteils oder ist diese nicht aktuell, so bleibt oftmals nur die manuelle Analyse des zugrundeliegenden Source-Codes, dies ist jedoch meistens eine zeitaufwendige Aufgabe.

In dieser Arbeit soll untersucht werden, inwiefern Data- und Text-Mining-Techniken eingesetzt werden können, um Software-Architekten und Programmierern Orientierungshilfen bei der Analyse von Quellcode zu geben. Text-Mining-Techniken werden im Allgemeinen dazu eingesetzt, die für einen Leser eines Textes relevanten und essenziellen Informationen aus einem Text zu extrahieren. Hierbei sollen ausdrücklich nicht nur unterschiedliche Analyseverfahren für ihre Eignung im o.g. Szenario untersucht werden, sondern auch geeignete Repräsentationstechniken.

Zudem sollen geeignete Analyse- und Repräsentationstechnik als Prototypen realisiert werden, der auf dem Quellcode eines Softwaresystems die für den Architekten oder Programmierer relevanten Informationen aufbereitete und präsentiert. Ansatz und Prototyp sollen durch die Analyse des Quellcodes eines Beispielsystems evaluiert werden.

Montag, den 22.09.2014

13:30 Uhr in Raum 106, Ifl, Julius-Albert-Straße 4