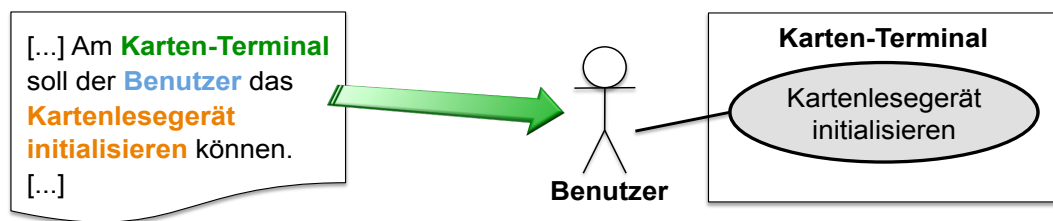


## Masterarbeit

# Effiziente Information-Extraction-Verfahren zur Erstellung von Use-Case-Diagrammen aus Texten

### Motivation

Im Entwicklungsprozess aufwändiger Software-Projekte entstehen enorme Mengen textueller Daten, die manuell kaum zu bewältigen sind. Die jüngst abgeschlossene Projektgruppe „Information-driven Software Engineering“ hat daher unter Anderem Verfahren der Information Extraction entwickelt, mittels derer sich benannte Entitäten wie Akteure, Komponenten und Anforderungsbezeichnungen automatisch in textuellen Software-Artefakten identifizieren lassen. Eine Erweiterung um die Erkennung von Relationen zwischen diesen Entitäten (so genannte „Relation Extraction“) ermöglicht die automatische Erstellung von Use-Case-Diagrammen. Für solche Zwecke kommen Pipelines teilweise berechnungsintensiver Algorithmen der natürlichen Sprachverarbeitung zum Einsatz.



### Aufgabenstellung

Ziel der Arbeit ist es, aufbauend auf den Ergebnissen der Projektgruppe maschinelle Lernverfahren zur Erkennung von Relationen zwischen den erwähnten Entitäten zu entwickeln, in Java zu implementieren und in die von der Projektgruppe erstellte Umgebung einzubinden. Eine Visualisierung der gefundenen Relationen als Use-Case-Diagramm ist bereits vorhanden. Die Güte der Verfahren soll geeignet evaluiert werden. Um den Berechnungsaufwand minimal zu halten, soll die Pipeline aller verwendeten Vorverarbeitungs- und Extraction-Algorithmen optimiert werden, wofür auf Methoden aktueller Forschungsergebnisse der Fachgruppe s-lab – Software Quality Lab der Universität Paderborn zurückgegriffen werden kann.

### Voraussetzungen

Vorwissen im Bereich des Maschinellen Lernens, zum Beispiel durch die gleichnamige Vorlesung

### Bearbeiter

Nikita Korshunov

### Wissenschaftliche Leitung

Prof. Dr. G. Engels

### Ansprechpartner

s-lab – Universität Paderborn  
Henning Wachsmuth  
Raum E4.107  
Tel: +49 (0) 5251 / 60-3359  
Email: hwachsmuth@s-lab.upb.de