

# Spezifikationsvortrag

Fabian Düker, Uli Steinbach

Universität Heidelberg, Institut für Computerlinguistik  
Softwareprojekt, SoSe 2018

Prof Dr. Katja Markert

12.06.2018

# Übersicht I

Übersicht  
Übersicht

inhaltliche Spezifikation  
inh. Spezifikation

Modularisierung und Aufgabenverteilung  
Modularisierung und Aufgabenverteilung

Programmarchitektur, Datenstrukturen  
Programmarchitektur und Datenstrukturen

# Übersicht

Autom. Erstellung eines Lexikons für die Erkennung von  
Abusive Words

Anwendung auf Germeval Task I 2018

Binäre Klassifikation von 5000 Tweets

# Problemstellung

- ▶ Problem: Hatespeech ist in ständiger Veränderung begriffen (Neologismen, Ambiguität, Kontext)
- ▶ Wiegand et al. 2016: Erstellung eines englischen Lexikons mit guten Ergebnissen auf cross-domain Evaluation
- ▶ SentiWS: Lexikon mit negativen Wörtern für das Deutsche

# Lösungsansatz

- ▶ Erstellung Baselexikon aus SentiWS neg. Sentiment-Lexikon
- ▶ halbautomatische Erweiterung des Baselexikons mit deutschen Schimpfwörtern
- ▶ autom. Erweiterung mittels graphbasiertem Label-Propagation-Algorithmus
- ▶ Anwendung auf Germeval 2018 Datenset und Evaluation

# halbautom. Erweiterung mit deutschen Schimpfwörtern

- ▶ Genius API: Erstellung eines Deutschrappkorpus
- ▶ Deutschrapp: zeitgemäße Verwendung von Schimpfwörtern (genrespezifisch, aber auch politisch + rassistisch)
- ▶ autom. Extraktion von Kandidaten mittels syntaktischer Pattern
- ▶ Beispielpattern: Du [NN] , Du [ADJ]\* [NN]
- ▶ manuelle Bereinigung der extrahierten Daten und Auswahl von Schimpfwörtern

# Auszug aus Korpus

- ▶ Liste mit Songs/Artists, Textauszug, Übersicht Top-Schimpfwörter

# Evaluation

- ▶ Baseline 1: Unigram und Bigram SVM:
- ▶ Baseline 2: Feature Selection (Mutual Information) SVM



## Evaluation

Table: Baseline: tf-idf unigram SVM

[illegible]

## Evaluation

Table: Baseline: tf-idf bigram SVM

[illegible]

## Evaluation

Table: Baseline: Feature-Selection m. Mutual Information

[illegible]

# Aufgabenverteilung

# Zeitplan

# Programmarchitektur

# Datenstrukturen

# Literatur