

INSTITUT FÜR MASCHINELLE SPRACHVERARBEITUNG

Universität Stuttgart
Pfaffenwaldring 5B
D-70569 Stuttgart

BACHELORARBEIT IN MASCHINELLER SPRACHVERARBEITUNG

Eine korpusbasierte Analyse zur Unterscheidung von Ironie und Sarkasmus

Autor:

Jennifer LING

Betreuer und 1. Prüfer:

Dr. Roman KLINGER

2. Prüfer:

Prof. Dr. Sebastian PADÓ

Beginn der Arbeit:

14.07.2015

Ende der Arbeit:

14.01.2016

Inhaltsverzeichnis

1. Einleitung	1
1.1. Hinführung	1
1.2. Motivation	3
1.3. Zielsetzung	5
2. Grundlagen	7
2.1. Begrifflichkeiten	7
2.1.1. Token, Typen, N-Gramme und Part-of-Speech-Tagging	7
2.1.2. Maschinelles Lernen	8
2.1.3. Sentimentanalyse	8
2.1.4. Twitter	9
2.2. Unterscheidung von Ironie und Sarkasmus anhand verwandter Arbeiten	10
2.2.1. Charakterisierung von Ironie und Sarkasmus	11
2.2.2. Erkennungssysteme für Ironie und Sarkasmus	15
2.2.3. Differenzierung von Ironie und Sarkasmus	18
3. Methoden zur Unterscheidung von Ironie und Sarkasmus	23
3.1. Korpus	23
3.1.1. Zugriff auf Twitter API	24
3.1.2. Vorverarbeitung der Tweets	27
3.1.3. Fakten zum Korpus	28
3.2. Anwendung von maschinellern Lernen	30
3.2.1. Merkmalsschablonen	30
3.2.2. Wahl des Klassifikators	34
3.2.3. Hashtags als Gold-Standard-Annotation	36
3.3. Ergebnisse und Diskussionen	37
3.3.1. Analyse der Merkmalsverteilung	37
3.3.2. Aufbau der Experimente	45
3.3.3. Ergebnisse der Experimente	49
3.3.4. Diskussion der Experimente	56
3.3.5. Analyse der falsch vorhergesagten Tweets	67
4. Schlussfolgerung und Ausblick	75
Literaturverzeichnis	79
A. Danksagung	85
B. Anhang	87