

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>15</b>
<b>Tabellenverzeichnis</b>	<b>21</b>
<b>Abkürzungsverzeichnis</b>	<b>27</b>
<b>1 Einleitung</b>	<b>29</b>
<b>2 Aufgaben und Problemstellungen des Opinion Mining und der Sentimentanalyse</b>	<b>35</b>
2.1 Begriffliche Abgrenzungen . . . . .	35
2.2 Dimensionen und Facetten des Opinion Mining . . . . .	37
2.2.1 Zielsetzungen und Aufgaben im Opinion Mining	37
2.2.1.1 Erkennen von Subjektivität . . . . .	37
2.2.1.2 Identifizieren von Meinungsäußerungen	38
2.2.1.3 Ermittlung der Polarität und Intensität von Meinungsäußerungen . . . . .	39
2.2.1.4 Identifizieren des Urhebers einer Meinungsäußerung . . . . .	40
2.2.1.5 Identifizieren des Ziels einer Meinungsäußerung . . . . .	42
2.2.1.6 Anforderungen an ein Opinion Mining (OM)-System . . . . .	43
2.2.2 Textsorten und Anwendungskontexte . . . . .	43
2.2.3 Granularität der Untersuchungsebene . . . . .	45
2.2.4 Zusammenfassung . . . . .	45
2.3 Evaluierung . . . . .	46
2.3.1 Evaluierung der Identifikation von Meinungsäußerungen, Urhebern und Zielen einer Meinungsäußerung . . . . .	49
2.3.2 Evaluierung der Ermittlung von Polarität . . . . .	50
2.3.3 Einbeziehen des Nutzers in den Evaluierungsprozess . . . . .	51

<b>3</b>	<b>Stand der Forschung</b>	<b>55</b>
3.1	Aspektbasiertes Opinion Mining . . . . .	56
3.1.1	Identifikation von Meinungsäußerung . . . . .	57
3.1.1.1	Monolinguale Ansätze und Methoden . . . . .	58
3.1.1.2	Multi- und cross-linguale Ansätze und Methoden . . . . .	67
3.1.1.3	Zusammenfassung . . . . .	74
3.1.2	Identifikation der Ziele von Meinungsäußerungen	79
3.1.2.1	Identifikation von Eigennamen und To- pics als Meinungsziele . . . . .	80
3.1.2.2	Identifikation von Produkteigenschaften in Rezensionen . . . . .	86
3.1.2.3	Zusammenfassung . . . . .	104
3.1.3	Identifikation von Relationen zwischen Meinungs- äußerungen und zugehörigen Zielen . . . . .	110
3.1.3.1	Regelbasierte Ansätze . . . . .	110
3.1.3.2	Maschinelle Lernverfahren . . . . .	112
3.1.3.3	Zusammenfassung . . . . .	122
3.1.4	Abschließende Betrachtung zum aspektbasierten Opinion Mining . . . . .	126
3.2	Wahrgenommene Nützlichkeit von Meinungsäußerungen	128
3.2.1	Einflussfaktoren auf die wahrgenommene Nütz- lichkeit von Kundenrezensionen . . . . .	128
3.2.1.1	Modellierung und Vorhersage der wahr- genommenen Nützlichkeit . . . . .	128
3.2.1.2	Ermittlung von Einflussfaktoren auf die wahrgenommene Nützlichkeit . . . . .	133
3.2.2	Einflussfaktoren auf die Relevanzbewertungen von Dokumenten und Textpassagen . . . . .	143
3.2.3	Zusammenfassung . . . . .	149
<b>4</b>	<b>Forschungsfragen und Methoden</b>	<b>151</b>
4.1	Forschungsfragen und Hypothesen . . . . .	151
4.2	Methoden . . . . .	161
<b>5</b>	<b>Entwicklung mehrsprachiger Ressourcen für die Evaluierung feingranularer OM-Systeme</b>	<b>167</b>
5.1	Existierende Korpora . . . . .	167
5.2	Erstellen eines multilingualen Korpus . . . . .	177
5.2.1	Annotationsmethodik . . . . .	179
5.2.2	Werkzeug für den Annotationsprozess . . . . .	183

5.2.3	Inter-Annotator-Agreement . . . . .	184
5.3	Erstellen eines Goldstandards für die Evaluierung von Relationen zwischen Meinungsäußerungen und den zu- gehörigen Zielen . . . . .	191
5.4	Zusammenfassung . . . . .	195
<b>6</b>	<b>Extraktion von Produkteigenschaften und Meinungen</b>	<b>197</b>
6.1	Aufbau des Extraktionsprozesses . . . . .	201
6.2	Linguistische Vorverarbeitung . . . . .	203
6.2.1	Normalisierung nutzergenerierter Inhalte . . . . .	203
6.2.2	OpenNLP-Komponenten . . . . .	204
6.2.3	Konvertieren der morphosyntaktischen Wortklas- sen . . . . .	205
6.2.4	Segmentieren von Teilsätzen . . . . .	206
6.3	Identifizieren von Meinungsäußerungen und zugehö- rigen Zielen . . . . .	208
6.3.1	Extraktion von Head-Modifier-Phrasen unter Ver- wendung von Begrenzerpaaren . . . . .	209
6.3.2	Regeln für Head-Modifier-Relationen . . . . .	210
6.3.3	Priorisieren der extrahierten Kandidaten . . . . .	214
6.3.4	Ermitteln der Meinung sowie des zugehörigen Ziels in extrahierten Phrasen . . . . .	216
6.3.5	Iterativer Filterprozess für Meinungsäußerungen und Ziele von Meinungsäußerungen . . . . .	218
6.3.5.1	Erstellen einer initialen Liste meinungs- tragender Terme für das Englische . . . . .	219
6.3.5.2	Erstellen einer initialen Liste meinungs- tragender Terme für das Deutsche . . . . .	219
6.3.5.3	Erstellen einer initialen Liste meinungs- tragender Terme für das Spanische . . . . .	221
6.3.6	Überprüfen von Phrasen niedrigerer Priorität . . . . .	221
6.3.7	Ermitteln diskontinuierlicher Phrasen . . . . .	222
6.4	Zusammenfassung . . . . .	222
<b>7</b>	<b>Systembezogene Evaluierung</b>	<b>225</b>
7.1	Evaluierung meinungstragender Adjektiv-Nomen- Relationen . . . . .	228
7.2	Evaluierung aller meinungstragenden Head-Modifier- Relationen . . . . .	238
7.3	Evaluierung der Übertragbarkeit des DOHM-Ansatzes auf weitere Sprachen . . . . .	241

7.4	Evaluierung der Identifikation impliziter Meinungsäußerungen und -ziele . . . . .	248
7.5	Zusammenfassung . . . . .	254
<b>8</b>	<b>Untersuchung der wahrgenommenen Nützlichkeit von Meinungsäußerungen</b>	<b>259</b>
8.1	Definition des Untersuchungsgegenstands . . . . .	260
8.2	Untersuchungsdesign . . . . .	260
8.2.1	Variablen . . . . .	262
8.2.2	Skala zur Messung der wahrgenommenen Nützlichkeit . . . . .	265
8.2.3	Aufbau des Fragebogens . . . . .	266
8.2.4	Simulierte Arbeitsaufgabensituationen . . . . .	268
8.2.5	Annotation der Merkmale der Meinungsäußerungen . . . . .	272
8.2.6	Kriterien für die Auswahl der im Fragebogen verwendeten Meinungen . . . . .	274
8.3	Pretest . . . . .	276
<b>9</b>	<b>Crowdsourcing als Form der Datenerhebung</b>	<b>277</b>
9.1	Crowdsourcing . . . . .	277
9.2	Qualität von mittels Crowdsourcing erhobenen Daten .	278
9.3	Nachteile durch die Verwendung von Crowdsourcing-Plattformen . . . . .	282
9.4	Rekrutierung der Teilnehmer . . . . .	284
<b>10</b>	<b>Auswertung</b>	<b>289</b>
10.1	Beschreibung der Daten . . . . .	289
10.2	Methode . . . . .	294
10.2.1	Logistische Regression . . . . .	295
10.2.2	Einbeziehen der subjektspezifischen Varianz in der logistischen Regression . . . . .	298
10.2.3	Kreuzvalidierung . . . . .	299
10.2.4	Anpassung des logistischen Regressionsmodells .	299
10.3	Ergebnisse . . . . .	300
10.3.1	Ergebnisse der Untersuchung des Hauptmodells .	300
10.3.2	Ergebnisse der Untersuchung des Hauptmodells unter Einbezug der Kontrollvariablen . . . . .	302
10.3.3	Überprüfen der Annahme auf proportionale Odds	306
10.3.4	Überprüfung von Multikollinearität . . . . .	311

10.3.5	Haupteffekt für die Nennung der Produkteigen- schaft in einer Meinungsäußerung . . . . .	312
10.3.6	Interaktionseffekte zwischen der Art der Mei- nungsäußerung, dem kulturellen Hintergrund und den Kontrollvariablen . . . . .	314
10.3.6.1	Interaktionseffekt zwischen der Art der Meinungsäußerung, dem kulturellen Hin- tergrund und dem Geschlecht . . . . .	314
10.3.6.2	Interaktionseffekt zwischen Art der Mei- nungsäußerung, kulturellem Hintergrund und Polarität . . . . .	317
10.3.7	Interaktionseffekte zwischen der Nennung von Zusatzinformation, dem kulturellen Hintergrund und den Kontrollvariablen . . . . .	324
10.3.7.1	Interaktionseffekt zwischen der Nennung von Zusatzinformation, dem kulturel- lem Hintergrund deutscher und spani- scher Teilnehmer und dem Geschlecht der Teilnehmer . . . . .	326
10.3.7.2	Interaktionseffekt zwischen der Nennung von Zusatzinformation, dem kulturel- len Hintergrund deutscher und spani- scher Teilnehmer und der Polarität ei- ner Meinungsäußerung . . . . .	331
10.3.7.3	Interaktionseffekt zwischen der Nennung von Zusatzinformation, dem kulturel- len Hintergrund amerikanischer und eu- ropäischer Teilnehmer und dem Geschlecht der Teilnehmer . . . . .	332
10.3.7.4	Interaktionseffekt zwischen der Nennung von Zusatzinformation, dem kulturel- len Hintergrund amerikanischer und eu- ropäischer Teilnehmer und der Polari- tät einer Meinungsäußerung . . . . .	336
10.3.8	Erklärungsgüte des Modells . . . . .	339
10.3.9	Subjektspezifische Varianz . . . . .	340
10.3.10	Zusammenfassung . . . . .	340

<b>11 Fazit</b>	<b>345</b>
-----------------	------------

<b>Literatur</b>	<b>353</b>
------------------	------------

<b>Anhang</b>	<b>379</b>
<b>A Annotationsrichtlinien</b>	<b>379</b>
<b>B Liste der Penn Treebank PoS-Tags</b>	<b>382</b>
<b>C Detaillierte Übersicht der Evaluierungsergebnisse</b>	<b>383</b>
<b>D Befragung</b>	<b>391</b>
D.1 Deutscher Fragebogen (Smartphones) . . . . .	391
D.2 Korrekturen der Meinungsäußerungen für die Befragung	401
<b>E Auswertung der Befragung</b>	<b>402</b>
E.1 Histogramme . . . . .	402
E.2 Bearbeitungsdauer . . . . .	403
E.3 Überprüfen der Annahme auf proportionale Odds . . . .	404