

Universität Konstanz
FB Informatik und Informationswissenschaft
Master-Studiengang Information Engineering

Masterarbeit

**Meinungsanalyse in Textdaten: Chancen der
interaktiven Steuerung des Analyseprozesses**

*zur Erlangung des akademischen Grades eines
Master of Science (M.Sc.)*

Studienfach: Information Engineering
Schwerpunkt: Visual Analytics
Themengebiet: Angewandte Informatik

von

Fabian Zintgraf

(01/622978)

Erstgutachter: Prof. Dr. Daniel A. Keim
Zweitgutachter: Prof. Dr. Marc H. Scholl
Betreuer: Christian Rohrdantz
Einreichung: Juli 2012

Abstract

Opinions affect us in our day-to-day decisions. Especially in purchasing products such as electronic equipment and books or also securities such as bonds and stocks. In the past, these purchase decisions were discussed with family members, whereas today we can also utilize the internet. Here, users can reveal their experiences within reviews to help other buyers with their purchases. Based on these reviews, buyers are able to identify strengths and weaknesses of the respective product features. For a small amount of data this is no noteworthy inconvenience for a human. But caused by the large amount of data, techniques of computational linguistics and computer science shall help - so far, however, with limited success. There are spelling errors, colloquial expressions, complex semantic relationships and ambiguities that can affect these automatic detections. Often, information can only be resolved with its context or a users's world knowledge. This can be integrated into the analysis process by using interactive steering, which can help distinguishing between different domains or to correct inevitable errors. This can be seen as process optimization, which can consequently improve the identification of feature-based sentiments. This thesis describes the chances of an interactive steering process which embodies a connection between man and machine in order to improve the analysis process. For this purpose an application has been developed, which incorporates the interactive steering and assists the analyst in its process optimization. A discussion of challenges and problem solutions within a scientific environment will take place and will be evaluated in the end.

Kurzfassung

Meinungen beeinflussen uns in unseren alltäglichen Entscheidungen. Sie begegnen uns zum Beispiel beim Kauf von Produkten, wie elektronischen Artikeln und Büchern oder auch Wertpapieren wie Anleihen und Aktien. Früher wurden diese Kaufentscheidungen mit Familienangehörigen diskutiert, während man heutzutage auch auf das Internet zurückgreifen kann. Hier können Benutzer mittels Rezensionen ihre Erfahrungen preisgeben, um andere Käufer bei ihrem Kaufabschluss zu beraten. Hieraus kann der Käufer Stärken und Schwächen einzelner Produktmerkmale identifizieren. Für eine geringe Datenmenge stellt dies für den Menschen keine nennenswerte Schwierigkeit dar. Aber bedingt durch die großen Datenmengen sollen hier nun Techniken der Computerlinguistik und der Informatik anknüpfen - bis jetzt aber nur mit mäßigem Erfolg. Dabei können Rechtschreibfehler oder umgangssprachliche Wortgebräuche eine automatische Erkennung beeinflussen, welche noch zusätzlich durch komplexe semantische Beziehungen und Mehrdeutigkeiten erschwert wird. Oftmals können Informationen nur aus dem Kontext heraus aufgelöst werden, wobei dies zum Beispiel auch über das Weltwissen des Menschen bewerkstelligt werden kann. Das kann mithilfe der Interaktiven Steuerung in den Analyseprozess integriert werden, um zum Beispiel verschiedene Domänen untereinander abzugrenzen oder um unvermeidliche Fehler korrigieren zu können. Hierdurch erfolgt eine Prozessoptimierung, welche folglich die Identifikationen von merkmalsbezogenen Meinungen verbessern kann. Diese Arbeit erläutert hierbei die Chancen der Interaktiven Steuerung, wobei untersucht wird, welcher genaue Mehrwert die Verbindung von Mensch und Maschine für die Analyse von Meinungen hervorbringen kann. Hierzu wurde eine Anwendung entwickelt, welche die Interaktiven Steuerung einbindet und den Analysten in seiner Prozessoptimierung unterstützt. Dabei werden Herausforderungen und Problemlösungen innerhalb dieses wissenschaftlichen Umfeldes diskutiert und abschließend evaluiert.

Inhaltsverzeichnis

Abstract / Kurzfassung	2
1 Einleitung	6
1.1 Einleitung	6
1.2 Strukturierung der Arbeit	12
2 Grundlagen	13
2.1 Meinungsanalyse	13
2.2 Terminologie	15
2.3 Textdaten	16
2.3.1 Verkaufsplattformen	16
2.3.2 Internetforen	18
2.3.3 Blogs	18
2.3.4 Nachrichten	19
2.4 Informationen aus unstrukturiertem Text	19
2.4.1 Statistische Ebene	20
2.4.2 Morphologische Ebene	21
2.4.3 Syntaktische Ebene	22
2.4.4 Semantische Ebene	25
2.4.5 Wissensbasen	26
2.5 Identifikatoren	27
2.5.1 Meinungen	27
2.5.2 Merkmale	30
2.5.3 Referenzen	32
2.5.4 Negationen	34
2.6 Interaktive Analyse	35
2.6.1 Visuelle Analyse	36

2.6.2 Analytische Steuerung 37

3 Problemstellung 40

3.1 Herausforderungen 40

3.1.1 Meinungsanalyse 40

3.1.2 Meinungsbegriff 42

3.1.3 Interaktive Steuerung 43

3.1.4 Textdaten 44

3.1.5 Evaluation 45

3.2 Lösungsansätze 45

3.2.1 Meinungsanalyse mit Sicherheiten 45

3.2.2 Meinungsbegriff 47

3.2.3 Interaktive Steuerung 48

3.2.4 Informationen aus unstrukturiertem Text 50

3.2.5 Identifikatoren 51

3.3 Hypothesen 57

4 Entwicklung 58

4.1 Analyseprozess 59

4.1.1 Analyseprozessstruktur 59

4.1.2 Analyseprozesskomponenten 60

4.2 Datenmodell 68

4.2.1 Erweiterbarkeit 70

4.3 Oberfläche 72

4.3.1 Komponenten 72

4.3.2 Schematischer Aufbau 75

4.3.3 Interaktive Steuerung 76

4.3.4 Interaktionsszenario 80

5 Erkenntnisse 83

5.1 Testumgebung 83

5.1.1 Datengrundlage 83

5.1.2 Ausgabe der Anwendung 87

5.2 Evaluationspezifikation 89

5.2.1 Meinungskandidaten 89

5.2.2 Produktmerkmalskandidaten 91

<i>INHALTSVERZEICHNIS</i>	5
5.2.3 Referenzen	93
5.3 Ergebnisse	95
5.3.1 Analyseprozess	95
5.3.2 Interaktive Steuerung	104
5.3.3 Übertragung der Prozessoptimierung	118
5.3.4 Schlussfolgerungen	123
5.4 Ausblick	123
6 Zusammenfassung	126
Literaturverzeichnis	134

Kapitel 1

Einleitung

1.1 Einleitung

Meinungen beeinflussen uns in unseren alltäglichen Entscheidungen. So begegnen sie uns zum Beispiel beim Kauf von Produkten wie elektronischen Artikeln, Büchern oder Aktien. Aber auch bei der Auswahl von passenden Urlaubsorten, Hotels, Ausbildungsplätzen oder Filmen müssen wir uns auf eine von vielen Alternativen festlegen. Früher wurden Fragen wie „Denkst du, ich sollte diese Kamera kaufen?“, „Wen wählst du?“ oder „Kennst du einen guten Zahnarzt?“ mit Familienangehörigen oder Freunden diskutiert, während man heutzutage auf verschiedenste Medien wie Zeitungen, Rundfunk und Fernsehen zurückgreifen kann. Bedingt durch die rasante Entwicklung des Internets und dank Web 2.0 findet man diese Diskussionen auch häufiger im Netz wie in sozialen Gemeinschaften, Foren, Blogs oder anderen Datenbankdienstleistungen. Hier können Benutzer mittels Rezensionen, Kundenfeedback oder Kommentaren ihre Erfahrungen mit einem spezifischen Produkt preisgeben, um andere Käufer bei ihrem Kaufabschluss zu beraten.

Nutzermeinungen werden vor Kaufentscheidungen aktiv genutzt
»Ich lese oft Nutzermeinungen vor einem geplanten Kauf.«

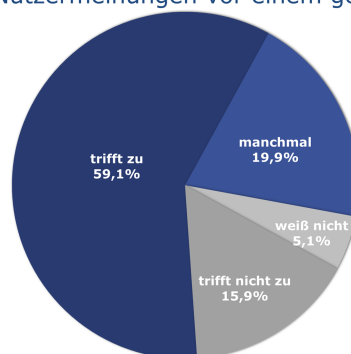


Abbildung 1.1: Bei einer Kaufentscheidung, wie sie oft im Internet von Benutzern getätigt wird, geben 79 % der Befragten an, sich zumindest manchmal von Nutzermeinungen inspirieren zu lassen. Basis: Internetgesamtnutzerschaft. Quelle: 29. WWW-Benutzeranalyse W3B, Herbst 2009.

Im Gegensatz zu früher besteht hierbei keine nahe Bindung mehr zwischen Produktkritiker und Produktkäufer. Dennoch lassen sich nach einer amerikanischen Studie zufolge 73 % bis 87 % der Onlinekäufer von Beurteilungen fremder Personen beeinflussen¹, in Deutschland bisher nur jeder Dritte². Unter diesen Voraussetzungen ist es aber nicht immer einfach, die richtigen Meinungen für die eigene Entscheidung zu finden und zu analysieren. Dies liegt hauptsächlich an der großen Datenmenge und an der Vielfalt an verschiedenen Plattformen, welche diese Informationen anbieten. Um den Käufer bei seiner Suche zu unterstützen, bieten viele dieser Plattformen Hilfestellungen an wie zum Beispiel 5-Sterne-Bewertungen oder automatische Zusammenfassungen der Rezensionen Dritter. Der potentielle Käufer hat somit die Möglichkeit auf einen Blick zwischen verschiedenen Produkten evaluieren zu können. Abbildung 1.2 zeigt einen Auszug aus einer fiktiven Verkaufsplattform.

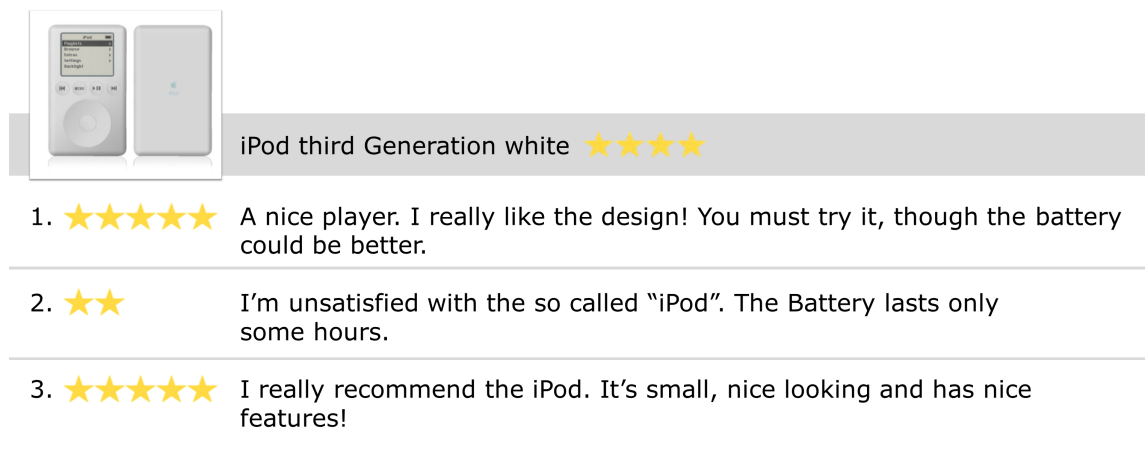


Abbildung 1.2: Das Beispiel zeigt Benutzerbewertungen über den iPod der dritten Generation. Hierbei bewertet Nutzer 1 das Produkt als sehr positiv (5 Sterne), Nutzer 2 bewertet das Produkt als eher negativ (2 Sterne) und Nutzer 3 vergibt ebenfalls 5 Sterne. Hieraus resultiert eine Gesamtbewertung des Produkts von 4 Sternen.

Der amerikanischen Studie zufolge sind 20 % bis 99 % der Konsumenten bereit für ein, mit fünf Sternen bewertetes Produkt mehr zu zahlen als für eines, das nur mit vier Sternen bewertet wurde² - doch diese müssen nicht immer glaubwürdig sein. Unter anderem besteht hier die Gefahr, dass Rezensionen gefälscht werden, um hierdurch kommerziell profitieren zu können³.

Gänzlich ungeeignet sind automatische Bewertungssysteme, wenn es dem potentiellen Käufer um die Stärken und Schwächen einzelner Produktmerkmale geht. Hier können die Geschmäcker der Konsumenten durchaus unterschiedlich ausfallen. Vertraut man nur auf Bewertungen, welche auf einer Gesamtpunktzahl basieren, kann dies zu einer irreführenden Empfehlung durch andere Be-

¹„Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior“, comScore, 29.11.2007

²Bitkom.org-Studie, März 2010

³„'Fake' reviews on TripAdvisor a real problem for hoteliers“, The Irish Times, 23.09.2011

nutzer führen. Betrachtet man das Bewertungskriterium „*Batterielaufzeit*“ in der vorhergehenden Abbildung, so ist dies für *Nutzer 1* niedriger gewichtet als für *Nutzer 2*, der gerade der Batterielaufzeit wegen eine schlechte Kritik geäußert hat. Das Gesamtergebnis von 4 Sternen fällt somit für *Nutzer 1* zu niedrig und für *Nutzer 2* zu hoch aus. Diese Bewertungssysteme sind folglich für einen individuellen Käufer mit seiner spezifischen Bewertungsgewichtung ungeeignet. Doch nicht nur für den potentiellen Käufer ist das Aufdecken von Produktstärken und Produktschwächen von Interesse. Ein Produkthanbieter könnte hierdurch zum Beispiel nicht nur evaluieren, ob sein Produkt beliebt ist, sondern auch welche Produkteigenschaften bei den Konsumenten gut, beziehungsweise schlecht ankommen. Bezogen auf das aktuelle Beispiel wurde die Akkuleistung des iPods in der nächsten Generation deutlich verbessert.

Bei einer geringen Datenmenge stellt die Erfassung dieser Produktstärken und Produktschwächen für den Menschen keine nennenswerte Schwierigkeit dar. Doch beinhalten manche, im Internet angebotenen, Produkte nicht selten mehrere Tausend Benutzerrezensionen - eine Datenflut, die der potentielle Käufer nicht mehr selbst bewältigen kann. Hier wollen nun Techniken der Computerlinguistik und der Informatik anknüpfen, um Konsumenten in ihren Entscheidungen und Produkthanbieter in ihren Produktevaluationen zu unterstützen. Hierzu müssen zunächst die Kommentare nach unterschiedlichen Satzbestandteilen, welche die Meinung über das Produktmerkmal ausdrücken, durchsucht werden. Unter anderem bedarf es der Identifikation von meinungstragenden Wörtern⁴, von Wörtern die für das Produkt⁵ oder das Produktmerkmal⁵ stehen und zuletzt der Identifikation der Beziehungen zwischen ebendiesen⁶. Ein Beispiel hierfür ist in Abbildung 1.3 zu sehen.

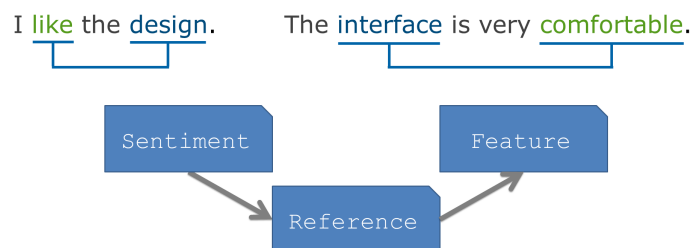


Abbildung 1.3: In diesen beispielhaften Nutzerkommentaren beschreiben „*like*“ und „*very comfortable*“ meinungstragende Worte (Sentiment) mit positiver Ausrichtung und „*design*“ und „*interface*“ stehen hierbei für Produktmerkmale (Feature). Die Verknüpfung (Referenz) dieser zwei Satzbestandteile ergibt also, dass das Design und das Interface für den Nutzer als positiv empfunden wird.

⁴Ein Wort, das eine Nutzermeinung repräsentiert, nennt sich Meinungswort. In dieser Arbeit wird das englische Wort „*Sentiment*“ als gleichwertiger Begriff angesehen

⁵Eine Ansammlung von Wörtern die für ein Produkt, ein Produktmerkmal oder eine Produkteigenschaft stehen, wird in dieser Arbeit auch mit dem englischen Begriff „*Feature*“ umschrieben

⁶Eine Beziehung zwischen einem Meinungswort und einem Produkt oder Produktmerkmal wird in dieser Arbeit als Referenz bezeichnet (in Englisch „*Reference*“)

Diese Identifikationsschritte können durch verschiedene Techniken der Computerlinguistik und der Informatik realisiert werden - bis jetzt aber nur mit mäßigem Erfolg. Dies liegt zunächst einmal an den Daten selbst, da Rechtschreibfehler, umgangssprachliche Wortgebräuche oder technische Ausdrücke eine automatische Erkennung beeinflussen können. Sowohl Ironie, Homographen⁷, wie auch komplexe semantische Beziehungen lassen sich für den Computer nur bedingt aufschlüsseln. Abbildung 1.4 zeigt Beispiele von Mehrdeutigkeiten, welche selbst für den Menschen eine Herausforderung darstellen.

Shankar sah den Affen mit dem Fernrohr.	2 Lesarten
Shankar sah den Affen mit dem Fernrohr auf der Mauer.	4 Lesarten
Fruit flies like a banana.	
	2 Bedeutungen
Visiting relatives can be dangerous.	
	2 Bedeutungen

Abbildung 1.4: Eine Mehrdeutigkeit oder Ambiguität liegt vor, wenn ein Satz oder ein Wort mehrere Bedeutung annehmen kann. Die strukturelle Ambiguität beschreibt hierbei syntaktische Mehrdeutigkeiten eines Satzes. Es kann sowohl der Affe durch ein Fernrohr beobachtet werden, als auch einen Affen mit einem Fernrohr in der Hand. Eine lexikalische Ambiguität liegt vor, wenn Wörter unterschiedliche Bedeutungen besitzen oder wenn sie verschiedenen Wortarten zugeordnet werden können. (Beispiele aus „*Skripte von Wilhelm Geuder (Einf. in die Ling. 2001), Ingrid Kaufmann und Barbara Stiebels*“).

Mehrdeutigkeiten können somit erst durch weitere Informationen des Kontextes aufgelöst werden, welcher aber nur bedingt von einer Maschine berechnet werden kann. Im Umfeld der Meinungsanalyse müssen diese Mehrdeutigkeiten auch für Meinungen betrachtet werden. Dabei ist zusätzlich zum Kontext ein Weltwissen gefordert. Hier müssen zum Beispiel Meinungswörter, welche in anderen Umgebungen⁸ eine unterschiedliche Meinung repräsentieren, identifiziert, unterschieden und voneinander abgegrenzt werden [1], da deren Meinungsrichtung vom jeweiligen Kontext abhängig ist. Dieses Phänomen kann jeweils innerhalb einer Produktparte, aber auch spartenübergreifend auftreten. Abbildung 1.5 erläutert dies anhand zweier Beispiele.

The printer prints fast.
The ink cartridge runs out fast.

The behavior of the software is unpredictable.
The plot of the movie is unpredictable.

Abbildung 1.5: Innerhalb der Druckersparte kann das gleiche meinungstragende Wort „fast“ eine positive Ausrichtung in Bezug auf die Druckgeschwindigkeit besitzen, sowie eine negative Ausrichtung in Bezug auf den Tintenverbrauch. Äquivalent kann dieses Phänomen auch in verschiedenen Sparten, wie hier zwischen einem Softwareprodukt und einem Film, auftauchen. Einerseits ist ein unberechenbares Softwareverhalten schlecht und zum anderen kann die unvorhersehbare Handlung eines Films positiv sein. Ein Weltwissen ist hier unerlässlich.

⁷Homographen sind gleich geschriebene Wörter mit unterschiedlichen Bedeutungen

⁸In dieser Arbeit beschreibt eine Umgebung eine spezifische Sparte von Produkten, wie Kameras, Bücher, oder andere. Eine „Domäne“ wird in dieser Arbeit als gleichwertiges Synonym verwendet

Dieses Kontextverständnis kann mithilfe der Interaktiven Steuerung in den Analyseprozess integriert werden. Somit kann auf das Weltwissen des Benutzers zurückgegriffen werden, um zum Beispiel die verschiedenen Umgebungen untereinander abzugrenzen oder um unvermeidliche Fehler durch Mehrdeutigkeiten korrigieren zu können. Spezielle Interaktionstechniken und Analysewerkzeuge unterstützen die Arbeit des Benutzers und stellen somit den Menschen in den Vordergrund der Meinungsanalyse. Abbildung 1.6 zeigt hierbei bildhaft, wie sich der Mensch und die Maschine in ihrer Leistung ergänzen können.

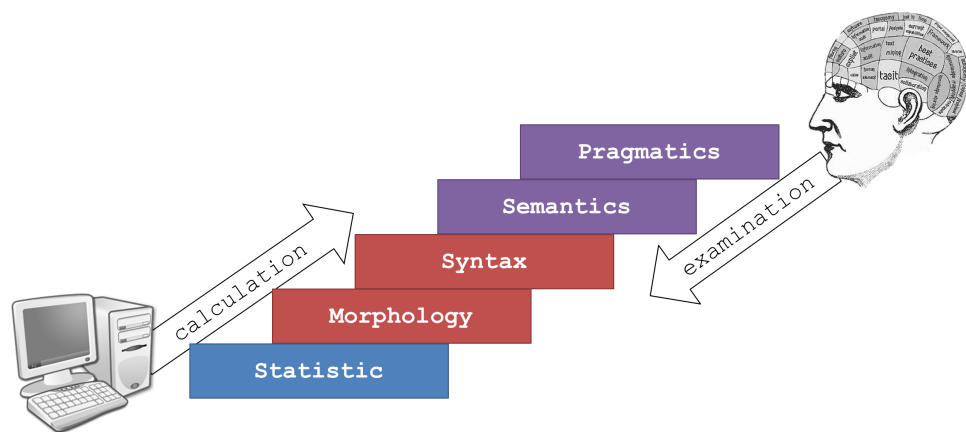


Abbildung 1.6: Die Wissenschaft der Computerlinguistik schreitet immer weiter voran, wodurch neuere Techniken entwickelt werden, um komplexere Strukturen in Texten aufschlüsseln können und somit automatische Algorithmen der Meinungsanalyse zu verbessern. Betrachtet man aber das Feld der Semantik und der Pragmatik, so kann hier bisher nur der Mensch mit seinem Kontextverständnis aushelfen. Arbeiten beide Teile zusammen, so können sie sich gegenseitig ergänzen.

Betrachtet man nur den zugrundeliegenden Meinungsanalyseprozess, so existieren in der wissenschaftlichen Literatur viele unterschiedliche Lösungsansätze, welche sich in ihrer Arbeitsweise und Anwendungsdomäne unterscheiden. Häufig sind diese Techniken nur für bestimmte Anwendungsfälle konzipiert oder decken nur eine spezifische Produktparte ab. Wird zum Beispiel ein Produktmerkmalskatalog aus der Automobildomäne in einer anderen Sparte eingesetzt, können nur noch mittelmäßige Erfolge erzielt werden. Dadurch ergeben sich für verschiedene Anwendungsdomänen spezifische Vorteile und Nachteile der jeweiligen Methoden, welche aber auf wissenschaftlicher Seite selten evaluiert werden. Hier wird meist ein besseres Endergebnis und eine höhere Genauigkeit der Methode angestrebt. Autoren, welche sich mit merkmalsbezogenen Meinungen beschäftigen, verwenden häufig auch mehrere, voneinander abhängige, Methoden, um das Endergebnis verbessern zu können. Hierdurch leidet jedoch die Transparenz, denn der Grund für die resultierende Verbesserung kann nicht mehr exakt zurückverfolgt werden. Dabei kann die Interaktive Steuerung helfen, da hierdurch ein Verständnis für die zugrundeliegenden Arbeitsweisen der Methoden generiert werden kann.

Die Herausforderungen bestehen darin, zunächst einen Analyseprozess zu entwerfen, worin Meinungswörter und Produktmerkmale identifiziert und zueinander in Beziehung gestellt werden, um merkmalsbezogene Meinungen modellieren zu können. Um die jeweiligen Vorteile und Nachteile verschiedener Identifikationsmethoden für unterschiedliche Anwendungsdomänen nutzen zu können, werden im Analyseprozess gleichzeitig mehrere Lösungsansätze modelliert. Hieraus können Unstimmigkeiten zwischen den einzelnen Methoden abgeleitet werden. Der Analyseprozess wird durch das Einbinden eines Benutzers als Fehlerkorrekteur und Wissenslieferant gesteuert, wobei unter anderem visuelle und interaktive Techniken eingesetzt werden.

Das **Ziel dieser Arbeit** ist folglich, die Chancen einer Interaktiven Steuerung des Analyseprozesses zu beleuchten. Dabei soll untersucht werden, welcher genaue Mehrwert die Verbindung von Mensch und Maschine für die Analyse von Meinungen hervorbringen kann. Wie eine solche Anwendung konzipiert und entwickelt wird, welche Herausforderungen hieraus entstehen und welche Ergebnisse erzielt werden können, soll nun innerhalb dieser Masterarbeit beschrieben und erörtert werden.

1.2 Strukturierung der Arbeit

Nachdem nun die Themenstellung dieser Arbeit motiviert wurde, folgt in *Kapitel 2* die Vorstellung der Grundlagen aus dem Bereich der Meinungsanalyse, wofür zunächst ein Überblick über die verschiedenen Sichtweisen auf dieses Themengebiet geschaffen wird. Nach Erläuterung der Begrifflichkeiten, welche in dieser Arbeit verwendet werden, wird anschließend der Meinungsbegriff in unterschiedlichen Textdaten untersucht. Daraufhin werden unterschiedliche Verfahren vorgestellt, welche aus unstrukturierten Textdaten Informationen für eine Meinungsanalyse bereitstellen können. Aufbauend auf diesem Vorwissen werden dann prominente Algorithmen zur Identifikation von Meinungen, Produktmerkmalen, Referenzen und Negationen aufgezeigt und in Hinblick auf die Thematik dieser Arbeit erörtert. Zuletzt gibt *Kapitel 2* Aufschluss über den Ansatz der Interaktiven Steuerung und beleuchtet deren Standpunkt in der Meinungsanalyse.

Das darauf folgende *Kapitel 3* geht auf die Problemstellung dieser Arbeit ein. Hierzu werden zunächst die, aus den Grundlagen abgeleiteten, Herausforderungen analysiert, aus welchen sich dann Anforderungen und Lösungswege für die Umsetzung eines Meinungsanalysewerkzeuges ergeben. Hieraus lassen sich Hypothesen formulieren und Folgerungen für eine spätere Evaluation definieren.

Nach Abhandlung des theoretischen Teils wird nun in *Kapitel 4* näher auf die Entwicklung der Anwendung eingegangen. Diesbezüglich werden die implementierten Komponenten vorgestellt und aufgezeigt, wie diese zu einem Analyseprozess verbunden werden. Auf welche Art und Weise dieser Prozess durch den Benutzer interaktiv gesteuert werden kann, zeigt die Beschreibung der grafischen Benutzeroberfläche und ein Benutzerszenario am Ende dieses Kapitels.

Danach folgt in *Kapitel 5* die Vorstellung der Erkenntnisse, welche sich aus der Anwendung heraus ergeben. Hierzu werden zunächst die Evaluationsdaten und die Testspezifikation vorgestellt. Anhand dieser kann dann die Beantwortung der, in dieser Arbeit vorgestellten, Hypothesen folgen, wobei interessante Funde anhand von Beispielen erläutert werden. Anschließend wird vorausschauend gezeigt, wie noch bestehende Konflikte gelöst werden könnten und welche Erweiterungen die Anwendung ergänzen würden.

Zuletzt werden in *Kapitel 6* die wichtigsten Hauptaugenmerke zusammengefasst, wodurch diese Arbeit abgerundet werden kann.

Kapitel 2

Grundlagen

Dieses Kapitel beleuchtet die Grundlagen aus dem Bereich der Meinungsanalyse. Hierzu wird zunächst näher auf diesen Begriff eingegangen und aufgezeigt, welche verschiedenen Sichtweisen eine Analyse von Meinungen beinhalten kann. Nach der Definition der grundlegenden Begrifflichkeiten, welche innerhalb dieser Arbeit Verwendung finden, wird der Meinungsbegriff in unterschiedlichen Textdaten untersucht. Danach wird dargelegt, welche Informationen aus unstrukturierten Daten die verschiedenen Verfahren der Computerlinguistik extrahieren können. Diese Informationen bilden dann das Grundgerüst, woran sich Algorithmen der Meinungsanalyse bedienen können. Nachdem die prominentesten Verfahren zur Identifikation von Meinungen, Merkmalen, Referenzen und Negationen vorgestellt wurden, soll abschließend der Ansatz einer interaktiven Steuerung diskutiert werden.

2.1 Meinungsanalyse

Der Begriff der *Meinungsanalyse* (auf Englisch *Sentiment Analysis*) kann in der Literatur auch durch deren Synonyme wie *Opinion Mining*, *Review Mining* oder *Subjectivity Analysis* umschrieben sein⁹. Ein Text, wie zum Beispiel eine Benutzerrezension, beinhaltet allgemein zwei unterschiedliche Kategorien von Informationen. Faktische Informationen beschreiben hierbei Ausdrücke über Objekte und Ereignisse sowie deren Eigenschaften. Meinungstragende Informationen beschreiben im Gegensatz dazu eine subjektive Äußerung, welche die Einstellungen, Gefühle und Beurteilungen einer Person gegenüber Fakten ausdrücken [2]. Eine Aufgabe der *Meinungsanalyse* kann zum Beispiel darin bestehen, die enthaltenen Informationen in Fakten und Meinungen zu unterteilen, um somit subjektive Äußerungen zu Analysezielen herauszufiltern.

⁹In dieser Arbeit wird der deutsche Begriff Meinungsanalyse für alle Synonyme von Sentiment Analysis verwendet

Der Bereich der *Meinungsanalyse* umfasst ein recht weitreichendes Forschungsgebiet. Erste Arbeiten lagen schon in den 70er Jahren vor, wobei der Hauptteil der Wissenschaftler erst seit dem Beginn dieses Jahrhunderts ein Forschungsinteresse entwickelten. Mittlerweile existieren bereits eine Vielzahl von Veröffentlichungen in diesem Themengebiet. Zurückzuführen ist dies zum einen auf den stetig wachsenden Bestand von Techniken im Bereich *Maschinelles Lernen*, *Computerlinguistik* und *Information Retrieval*, welche die Grundlagen für Anwendungen bilden, und zum anderen auf die Verfügbarkeit und die weite Verbreitungsmöglichkeiten von Datensätzen mithilfe des Internets. Aber auch die dahinter liegenden, gedanklichen Herausforderungen und die breite Verwendung für kommerzielle und informative Systeme bewirkten den Anstieg des Forschungsvolumens [3].

Somit entstanden eine Fülle von unterschiedlichen Herangehensweisen, die entweder neue Anwendungsbereiche erfassten oder die bereits existierenden Methoden verbesserten. Die in der Literatur zu findenden Arbeiten lassen sich in verschiedene Teilaspekte einordnen, je nachdem, welche Herausforderungen und Ziele die jeweiligen Autoren verfolgen wollen. Die einzelnen Aspekte sollen nun anhand von Fragestellungen aufgezeigt und näher erläutert werden.

Wer ist Verfasser des Textes und worin befindet sich die Meinung? Sind professionelle Journalisten die Urheber oder ist ein Benutzer eines Produkts der Verfasser der Meinung und findet der Meinungs Austausch in Twitterbotschaften oder zum Beispiel in Artikeln eines Computermagazines statt?

Auf welche Informationen stützt sich die Technik? Verwendet sie einfache statistische Einheiten oder verwendet sie komplexere syntaktische oder gar semantische Strukturen?

Wie sehen die Autoren den Meinungsbegriff? Unterscheiden sie lediglich zwischen Objektivität und Subjektivität, beschreiben sie die Subjektivität durch deren positiven und negativen Ausrichtung oder differenzieren sie den Meinungsbegriff noch feiner? Wird die Meinung alleinstehend betrachtet oder stellen sie diese zusätzlich zum Beispiel in Bezug zu einem Produktmerkmal?

Auf welchen Ebenen befindet sich die Meinung? Soll das Dokument gesamtheitlich oder nur einzelne Teilaspekte, wie einzelne Sätze oder Satzbestandteile, als Meinungsträger betrachtet werden?

Worin findet die Methode ihre Verwendung? Ist sie zur Verbesserung eines bestehenden Systems vorgesehen, dürfen Benutzer zum Beispiel am visuellen Mehrwert der Anwendung teilhaben oder wird hier eine Basis für zukünftige Techniken geschaffen?

Inwiefern ist der Benutzer in den Prozess involviert? Arbeitet das Verfahren komplett autonom oder muss der Benutzer einzelne oder viele Entscheidungen treffen?

Diese Fragen spiegeln die Weitläufigkeit der Anwendungen innerhalb der Meinungsanalyse wider, wodurch ein allumfassender Überblick über bisherige Erkenntnisse nahezu unmöglich erscheint. Im Weiteren werden nun Beispiele innerhalb dieser Teilaspekte herausgegriffen und anhand der prominentesten Anwendungen in der wissenschaftlichen Literatur näher erläutert.

2.2 Terminologie

Bevor nun genauer auf spezifische Grundlagen der Meinungsanalyse eingegangen werden kann, müssen die dort erwähnten Begrifflichkeiten beschrieben und erläutert werden.

Eine **Produktparte** beschreibt hierbei den Zusammenschluss von Produkten mit ähnlichen Merkmalen und Eigenschaften. So können zum Beispiel Digitalkameras in die *Kamera-Sparte* und Tintenstrahldrucker in die *Drucker-Sparte* einsortiert werden. Spricht man von einer **Sparte**, so können auch Zusammenschlüsse katalogisiert werden, die keine Produkte enthalten. Hierzu zählen zum Beispiel die *Nachrichten-Sparte* oder die *Film-Sparte*. Anstelle von *Sparte* kann auch der gleichbedeutende Begriff „*Domäne*“ verwendet werden.

Die **Meinung**, oder auf Englisch „*Sentiment*“, stellt den Untersuchungsgegenstand innerhalb der Meinungsanalyse dar. Sie beschreibt einen subjektiven Ausdruck eines Individuums über einen spezifischen Umstand, wodurch dieser zum Beispiel anhand seiner Eigenschaften bewertet wird. Betrachtet man dies in Textdaten, so wird ein solches Wort als *meinungstragendes Wort* bezeichnet. Ist der bewertende Umstand des Wortes noch nicht eindeutig identifiziert, so spricht man in diesem Fall von einem *Meinungskandidaten*. Dieser gibt lediglich an, dass es sich um ein meinungstragendes Wort handelt, aber noch kein Bezug zu einem Umstand hergestellt worden ist. Die Subjektivität der Meinung wird als *Meinungsausrichtung* oder *Meinungsorientierung* bezeichnet, wobei diese Arbeit lediglich die Ausrichtung in die negative oder die positive Richtung betrachtet. Andere Abstufungen sind vom jeweiligen Anwendungsziel abhängig, zum Beispiel könnten emotionale Kategorien erfasst werden [4].

Ein **Merkmal**, oder auf Englisch „*Feature*“, entspricht innerhalb der Meinungsanalyse einem Umstand, worüber eine Meinung geäußert werden kann. Innerhalb der Produktparte kann dieser zum Beispiel für das Produkt selbst, für eine Produkteigenschaft oder für ein Produktbestandteil stehen. Für andere Sparten wird der gleiche Ausdruck verwendet, obwohl es sich hierbei nicht

zwingend um Produktmerkmale handelt. Dies gewährleistet somit eine einheitliche Begrifflichkeit innerhalb dieser Arbeit. Steht der Bezug zu einer Meinung nicht fest, so spricht man von einem *Produktmerkmalskandidaten*.

Eine **merkmalsbezogene Meinung**, Referenz, oder auf Englisch „*Reference*“, wird definiert durch die Beziehung zwischen einem Meinungskandidaten und einem Produktmerkmalskandidaten. Hierdurch werden die Kandidaten in den Stand von tatsächlichen Meinungen und Produktmerkmalen befördert.

Spricht man von **Weltwissen**, so ist innerhalb der Meinungsanalyse der Aspekt des Kontextverständnisses gemeint, welches nur bedingt vom Computer errechnet werden kann. Hierzu gehören unter anderem die Abgrenzung der Begriffe aus verschiedenen Sparten und das Einbeziehen des Kontextes speziell bei merkmalsbezogenen Meinungen. Zusätzlich können über den Kontext auch Mehrdeutigkeiten abgegrenzt werden.

2.3 Textdaten

Im Internet finden sich diverse Plattformen, worin Benutzermeinungen repräsentiert werden. Verkaufsplattformen bieten ihren Benutzern an, Bewertungen über ein Produkt einzusehen oder abzugeben, wohingegen in Foren Diskussionen über vorab definierte Themengebiete geführt werden. In Blogs finden sich meist längere Texte des Autors, welche von Benutzern weiter kommentiert werden können, während das Mikroblogging-Portal *Twitter*¹⁰ ihren Benutzern hierfür nur eine beschränkte Anzahl an Zeichen zur Verfügung stellt. In Nachrichtentexten, wie Aktienneuigkeiten oder Zeitungsartikeln, findet keine direkte Diskussion mit Benutzern statt und es herrscht ein geringes Maß an Subjektivität. Im Weiteren soll nun vereinzelt auf die spezifischen Gruppen eingegangen werden, um zu zeigen, inwiefern sich die dort enthaltenen Meinungen äußern.

2.3.1 Verkaufsplattformen

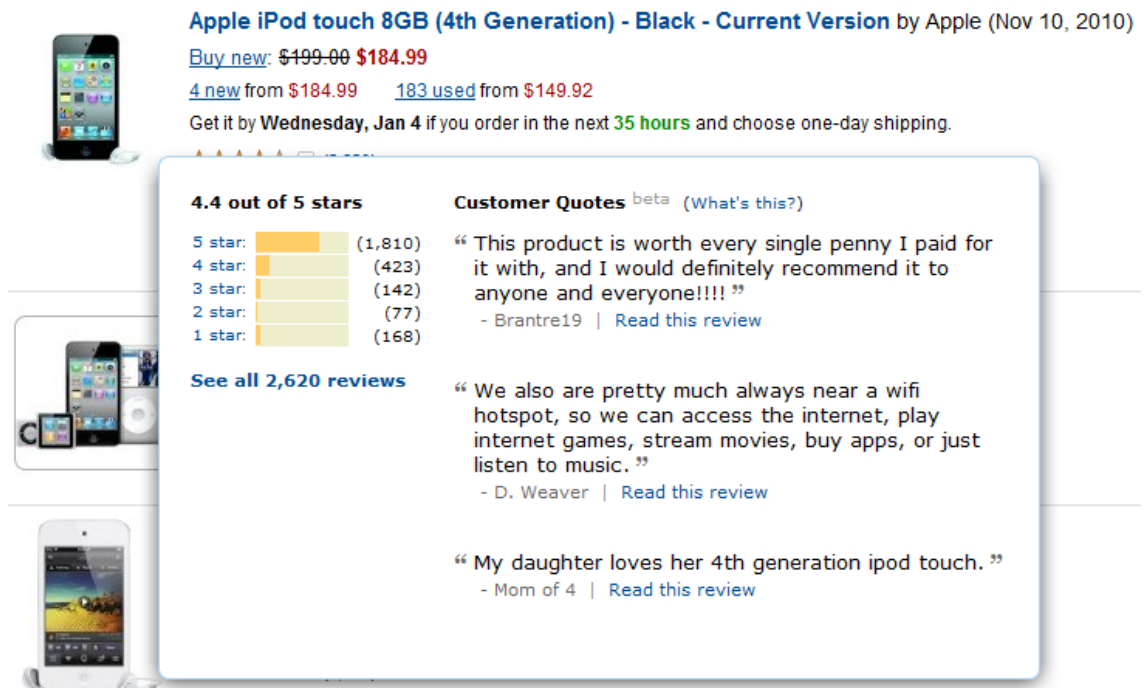
Verkaufsplattformen sind im Internet recht weit verbreitet. Bei *Amazon*¹¹ hat ein Käufer zum Beispiel die Auswahl aus über 85 Millionen verschiedenen Artikeln¹². Hier finden sich meist technische Geräte wie MP3-Player, Mobiltelefone, Drucker oder Digitalkameras, aber auch Bücher, Haushaltswaren, Kleidung oder Filme. Für jedes Produkt bietet Amazon, wie in Abbildung 2.1 zu sehen, eine Zusammenfassung der 5-Sterne-Bewertungen Dritter sowie einen Auszug der verfassten

¹⁰<http://www.twitter.com>

¹¹<http://www.amazon.com>

¹²Amazon Deutschland 85.177.964, Amazon UK 96.312.681, Amazon USA 121.271.296 (am 02.01.2012)

Rezensionen an. Für das Produkt „iPod touch 8GB“ sind zum Beispiel 2.620 Benutzerbewertungen verfügbar. Ein Käufer hat hier nun die Möglichkeiten, sich für seine Kaufentscheidung auf die Gesamtbewertung des Produktes zu verlassen, sich als hilfreich markierte Rezensionen ausgeben zu lassen, oder selbst Kommentare durchzulesen.



Apple iPod touch 8GB (4th Generation) - Black - Current Version by Apple (Nov 10, 2010)

Buy new: ~~\$199.00~~ **\$184.99**

4 new from **\$184.99** 183 used from **\$149.92**

Get it by **Wednesday, Jan 4** if you order in the next **35 hours** and choose one-day shipping.

4.4 out of 5 stars

Star Rating	Count
5 star	(1,810)
4 star	(423)
3 star	(142)
2 star	(77)
1 star	(168)

[See all 2,620 reviews](#)

Customer Quotes beta ([What's this?](#))

“ This product is worth every single penny I paid for it with, and I would definitely recommend it to anyone and everyone!!!! ”
- Brantre19 | [Read this review](#)

“ We also are pretty much always near a wifi hotspot, so we can access the internet, play internet games, stream movies, buy apps, or just listen to music. ”
- D. Weaver | [Read this review](#)

“ My daughter loves her 4th generation ipod touch. ”
- Mom of 4 | [Read this review](#)

Abbildung 2.1: Überblick über Rezensionen Dritter über das Produkt „iPod touch 8GB“ innerhalb der Verkaufsplattform *Amazon*. Hierbei sind 3 beispielhafte Benutzerkommentare angegeben. Zusätzlich werden die einzelnen 5-Sterne-Bewertungen gruppiert und ihre Häufigkeit in einem Diagramm dargestellt.

In Verkaufsplattformen beschreiben die Rezensenten meist innerhalb von Volltexten diverse Vorteile und Nachteile, welche sie im Laufe der Zeit mit ihrem Produkt gesammelt haben. Oftmals schildern sie diese anhand von realen Lebenssituationen und erläutern ihre Erfahrungen bildhaft. Manche Rezensenten bevorzugen es aber, ihre Kommentare stichpunktartig aufzuführen, wodurch die meinungstragenden Aussagen mittels computerlinguistischen Verfahren einfacher identifiziert werden können [3]. Die Texte sind meist in natürlicher Sprache geschrieben und können Rechtschreibfehler und ironische Bemerkungen enthalten. Abkürzungen sowie Emoticons tauchen eher selten auf. Anwendungen innerhalb der Meinungsanalyse nutzen häufig diese, von Benutzern angelegten, Textdaten, da die darin enthaltenen Sprachstile sich sehr gut als Datengrundlage eignen. Auch findet man Lernalgorithmen, die an den 5-Sterne-Wertungen anknüpfen. Dadurch, dass die Rezensenten ihre Erfahrungen mit einem Produkt preisgeben wollen, enthalten diese Daten viele merkmalsbezogenen Meinungen. Zusätzlich verfügen Verkaufsplattformen über eine sehr große Menge an Textdaten aus unterschiedlichen Sparten.

2.3.2 Internetforen

Ein Internetforum, Marktplatz oder auch Diskussionsforum, ist ein virtueller Platz zum Austausch und zur Archivierung von Gedanken, Meinungen und Erfahrungen im Internet¹³. Die Diskussionsbeiträge sind hier meist nach Themengebieten aufgeteilt, wobei verschiedene Foren unterschiedliche Themengebiete abdecken. Klassische Webforen bieten ihren Benutzern eine hierarchische Darstellungsordnung über die Diskussionsverläufe, welche unter anderem von Moderatoren verwaltet werden. Ein Forum unterscheidet sich von einer Verkaufsplattform dahingegen, dass hier die Gespräche nicht nur kaufberaterisch motiviert sind. Internetforen enthalten eher Diskussionen über verschiedene Probleme oder bieten den Benutzern Hilfestellungen an.

2.3.3 Blogs

Ein Blog ist ein, im Internet, geführtes Tagebuch, in dem ein oder mehrere Autoren Aufzeichnungen führen, Sachverhalte protokollieren oder Gedanken niederschreiben¹⁴. Diese Texte sind meist in der Ich-Perspektive verfasst und enthalten nicht zwangsläufig Meinungen über Produkte, Personen oder Unternehmen. Die Schreibstile der Autoren können durchaus unterschiedlich ausfallen. Oftmals sind die Texte ironischer oder humoristischer Art und berichten über alltägliche Erfahrungen oder persönliche Erlebnisse.

Im Unterschied zu Internetforen wird zuvor kein spezielles Themengebiet abgesteckt. Die Meinungen sind in Erzählungen des Autors eingebettet und daher versteckter als in Verkaufsplattformen. Auch die Einhaltung eines einzigen Themas kann nicht gewährleistet werden, da Blogs oftmals eine Mischung an Unterthemen enthalten, wodurch eine ganzheitliche Meinungsbetrachtung auf Themenebene erschwert wird [5].

Für die Betrachtung der Meinungen ist aber auch die Art des Blogs entscheidend. In Foto-Blogs zum Beispiel finden sich Äußerungen zum Bildmotiv sowie zur Qualität der Aufnahme, welche getrennt untersucht werden sollten [6]. Film-Blogs bestehen meist aus der Wiedergabe des Filminhaltes, worin keine spezifische Meinung enthalten ist, und der eigentlichen Meinung des Benutzers zum Film.

Das Microblogging Portal *Twitter* stellt nach Pak et al. eine wertvolle Quelle für Gefühls- und Meinungsäußerungen der Benutzer dar [7]. Durch entsprechende Funktionen können sich Nutzer zu Netzwerken zusammenschließen, wodurch eine schnelle Meinungsverbreitung und eine hohe Diskussionsdichte erreicht wird. Hierbei wird der Autor auf 140 Zeichen beschränkt. Dies führt zu relativ kurzen Sätzen, in denen viele Abkürzungen und Emoticons enthalten sind, welche somit

¹³<http://de.wikipedia.org/wiki/Internetforum>

¹⁴<http://de.wikipedia.org/wiki/Blog>

oftmals zur Meinungsaussage herangezogen werden. Die grammatikalische Struktur kann häufig nicht aufrechterhalten werden und ist daher für Verfahren der Computerlinguistik schwerer zu analysieren.

Bedingt durch die große Menge an Daten, die riesige Themenvielfalt und die große Abdeckung von Benutzern, rückt diese Plattform trotz dieser Herausforderungen immer mehr in das Interesse von Unternehmen [7]. Für die Wissenschaft innerhalb der Meinungsanalyse ist hierbei nicht nur die Identifikation der Meinungen aufschlussreich, sondern auch das Erkennen von sich entwickelnden Trends über die Zeit oder das Untersuchen von spezifischen Themengebieten.

2.3.4 Nachrichten

Nachrichtentexte unterscheiden sich deutlich von den bereits erwähnten Textformen. Bisher fanden primär Diskussionen über spezifische Themengebiete statt, welche Meinungen über Produkte, Firmen oder Personen beinhalten. Ein solcher Erfahrungsaustausch zwischen Benutzern findet bei Nachrichten nicht statt.

Im Allgemeinen beschreiben Nachrichten die objektive Sicht auf ein Ereignis, welches je nach Kontext positiv, negativ oder neutral sein kann. Eine starke und subjektive Betrachtungsweise, wie sie in Verkaufsplattformen vorkommt, kann nicht angenommen werden. Auch verfolgt hier die Meinungsanalyse einen anderen Zweck. Im Finanzsektor besteht zum Beispiel die primäre Aufgabe, den Kontext der Nachrichten auf die Reaktionen am Markt zu abstrahieren [8]. Oftmals werden hier auch die Begriffe von Optimismus und Pessimismus betrachtet, welcher sich besser auf eine Marktreaktion ableiten lassen [9].

2.4 Informationen aus unstrukturiertem Text

Nachdem nun verschiedene Textformen betrachtet wurden, soll nun erörtert werden, welche Informationen der Computer aus diesen unstrukturierten Daten gewinnen kann. Hierauf können dann Techniken der Informatik aufbauen, um merkmalsbezogene Meinungen identifizieren zu können. Die verschiedenartigen Informationen lassen sich anhand ihrer Charakteristiken in Ebenen gruppieren. Die unterste Ebene bildet hierbei die Statistik, während die Ebenen der Computerlinguistik bereits sprachliche Informationen extrahieren. Zuletzt kann sich der Computer auch an externen Quellen bedienen. Diese Informationen werden nun im Folgenden näher erläutert.

2.4.1 Statistische Ebene

Die statistische Ebene umfasst primär mathematische Berechnungen auf den Wortdaten. Hierzu werden die Texte in *Terme* unterteilt, welche für ein Vorkommen eines Wortes stehen. Diese bilden somit die Untersuchungsgegenstände, worüber die verschiedenen statistischen Berechnungen angestellt werden können.

Zunächst können Terme anhand ihres **Vorkommen** bewertet werden. Dabei kann entweder die Anwesenheit oder die Häufigkeit des Terms betrachtet werden. Im traditionellen *Information Retrieval* werden häufig, zur Beschreibung von Dokumenten, sogenannte Term-Vektoren berechnet. Hierzu verwendet man entweder *binäre Vektoren* [10], die die Term-Präsenz modellieren, oder man verwendet Vektoren, die die Häufigkeiten der Terme enthalten [3]. Um die Relevanz der einzelnen Terme der Dokumente zu beschreiben, gewichtet man die Vektorenwerte mit der inversen Dokumentenhäufigkeit. Diese misst die allgemeine Bedeutung des Terms für die Gesamtmenge der betrachteten Dokumente¹⁵. Für die Analyse von Meinungen spielen aber auch die Betrachtung von seltenen Termen (*hapax legomena*) eine wichtige Rolle [11], gerade auch im Hinblick auf unbekannte Wortschöpfungen oder Rechtschreibfehler.

Auch die **Position** des Terms innerhalb des Satzes oder des Dokuments kann für eine Meinungsbestimmung wichtig sein. So befinden sich zum Beispiel gehaltvolle Aussagen am Anfang und am Ende von Absätzen oder innerhalb des Dokumententitels. Die Position kann somit ein Indikator für die Wichtigkeit einzelner Terme sein. Betrachtet man die Position, so können auch **Distanzen** zwischen verschiedenen Termen berechnet werden. Ein prominentes Beispiel hierfür beschreiben die Autoren Ding, Liu und Yu in ihrem Verfahren [12]. Indem sie Meinungswörter höher gewichteten, weil sie in naher Umgebung zu einem Produktmerkmal auftauchen, erzielten sie deutlich bessere Ergebnisse.

Für eine Meinungsanalyse kann das Betrachten der **Nachbarschaften** der Terme zu einem genaueren Ergebnis führen [10]. Betrachtet man zum Beispiel nicht nur einzelne Terme (sogenannte *Unigramme*), sondern auch aufeinanderfolgende Terme (*Bigramme*, *Trigramme*), so können Klassifikationsalgorithmen genauer werden. Denn meist sind bestimmte Termfolgen ein Indiz für eine meinungsgebende Funktion. Die Autoren Dave et al. erzielten zum Beispiel durch die Trigramme „easy to use“, „is the best“ und „it doesn't work“ ein deutlich besseres Klassifikationsergebnis [13]. Für die Meinungsanalyse sind Nachbarschaften dahingegen interessant, da sich Meinungswörter meist in der Nähe von Merkmalen aufhalten [6].

¹⁵<http://de.wikipedia.org/wiki/Tf-idf-Maß>

2.4.2 Morphologische Ebene

Die Morphologie beschreibt eines von zwei Teilgebieten der Grammatik aus der Sprachwissenschaft. Sie beschäftigt sich mit der inneren Struktur der Wörter - den *Morphemen*, welche die kleinsten bedeutungs- und funktionstragenden Elemente einer Sprache beschreiben. Diese werden zur Wortbildung herangezogen, indem sie durch bestimmte Regeln an Wortstämme, welche die Wortfamilie bestimmen, angehängt werden. Für Algorithmen der Computerlinguistik kann es von Vorteil sein, nicht jede Wortveränderung der Terme zu untersuchen, sondern jeweils nur deren Grundform zu betrachten. Zum Beispiel werden Verben mittels *Flexion* konjugiert und mittels *Derivation* verändert. Macht man diese wieder rückgängig, so können alle Verben in deren Grundform untersucht werden. Speziell auch bei der Identifikation von meinungstragenden Wörtern mithilfe von Wörterbüchern kann eine Grundformreduktion von Vorteil sein. Adverbien besitzen zum Beispiel eine ähnliche Meinungsrichtungen wie deren verwandte Adjektive, sind aber meist durch *Derivation* verändert und müssten somit zusätzlich im Meinungsörterbuch auftauchen. Die Grundformreduktion entfernt folglich *Flexions-Morpheme* und *Derivations-Morpheme*, um ähnliche Terme vereinheitlichen zu können und um hierdurch die Menge der Terme einzugrenzen. Abbildung 2.2 veranschaulicht diese zwei Arten der Wortveränderungen anhand von Beispielen.

Inflection:	run-s	batter-ies	arriv-ed
	do-es	compan-ies	person-s
Derivation:	publish-ing	computer-ization	personal-ly
	build-ing	comput-ational	drink-able

Abbildung 2.2: Die Auswirkung, wenn sich Morpheme an Wortstämme anhängen, unterscheiden sich anhand der Art der Wortveränderung. Flexion bildet Pluralbildung und Konjugation ab, während die Derivation hierbei die Wortart verändert.

Einer der verständlichsten *Grundformreduktionsalgorithmen* für englische Texte stellt der sogenannte *Porter-Stemmer* [14] dar und wird somit sehr häufig in Anwendungen verwendet. Der Algorithmus besteht aus fünf Reduktionsphasen, welche der Reihe nach abgearbeitet werden. Jede Phase besteht aus bestimmten Ersetzungsregeln, die auf vorab definierten Suffixe der Terme angewendet werden. Zusätzlich beeinflussen Einschränkungen zu Silbenanzahl und Wortlänge diese Regeln.

Einen ähnlichen Ansatz verfolgt der *Grundformreduktionsalgorithmus* von Kuhlen [15]. Hierbei werden ebenfalls Ersetzungsregeln auf Suffixe angewendet, wobei aber deren Bedingungen an Vorkommen von Vokalen und Konsonanten geknüpft sind. Dadurch lassen sich Terme, im Vergleich zum Ansatz von Porter, durch formschönere lexikalische Grundformen abstrahieren.

Die Reduktion auf die Grundform kann auch durch eine sprachliche Analyse erfolgen. Diese Techniken benötigen hierbei ein vollständiges Vokabular und eine umfassende morphologische Analyse der jeweiligen Sprache. Die Anwendung *morphy*¹⁶ entspricht einem solchen Ansatz und bildet morphologische Regeln anhand der Einträge innerhalb des domänenunabhängigen Wörterbuches *WordNet*¹⁷. Anhand der Wortart und der vorberechneten morphologischen Regeln kann somit die Grundform des Terms bestimmt werden. Unterschiedliche Domänen können hierbei verschiedenartige Regeln verlangen, welche unter Umständen neu erlernt werden müssen.

2.4.3 Syntaktische Ebene

Die syntaktische Ebene erklärt die Zusammenhänge der Wörter innerhalb eines Satzes. Hierzu gehören die Identifikation von Wortarten und vereinzelt Satzbestandteilen sowie das Erkennen von syntaktischen Abhängigkeiten innerhalb der Sätze. Welche Verfahren die Computerlinguistik hierbei anbietet und welche Informationen für eine Meinungsanalyse bereitgestellt werden können, soll nun im weiteren dargestellt werden.

Part-of-speech Tagging beschreibt hierbei den Prozess, den Termen innerhalb der Sätze sogenannte *Tags* zuzuordnen. Diese stehen für spezifische, syntaktische Funktionen, die die Terme in den Sätzen annehmen können. Hierdurch können Terme mit ähnlichen syntaktischen Eigenschaften zusammengefasst und in Wortartgruppen eingeordnet werden. Zum Beispiel wird den verschiedenen Personalpronomen jeweils der gleiche *Tag* zugewiesen, wodurch folglich Wortartinformationen gewonnen werden können. Die Autoren Pak und Paroubek zeigen in einer Vorabuntersuchung, dass man bereits allein bei Betrachtung der Wortarten auf meinungstragende Inhalte schließen kann [7]. Die Verfahren zur Wortartidentifikation greifen hierbei auf unterschiedliche *Tag Sets* zurück. Hierbei sind Wortarten nicht nur in Nomen, Adjektive oder ähnliche eingeteilt, sondern werden viel feiner unterschieden. Zum Beispiel werden Nomen zusätzlich nach Singular- oder Plural-Form kategorisiert. Abbildung 2.3 gibt die verschiedenen Wortartkategorien aus dem *Penn Treebank Tag Set*¹⁸ an.

¹⁶<http://wordnet.princeton.edu/man/morphy.7WN.html>

¹⁷<http://wordnet.princeton.edu/>

¹⁸<http://www.cis.upenn.edu/~treebank/>

CC	Coordinating conjunction	PRP	Personal pronoun
CD	Cardinal number	PRP\$	Possessive pronoun
DT	Determiner	RB	Adverb
EX	Existential there	RBR	Adverb comparative
FW	Foreign word	RBS	Adverb superlative
IN	Preposition or subordinating conjunction	RP	Particle
JJ	Adjective	SYM	Symbol
JJR	Adjective comparative	TO	to
JJS	Adjective superlative	UH	Interjection
LS	List item marker	VB	Verb base form
MD	Modal	VBD	Verb past tense
NN	Noun singular or mass	VBG	Verb gerund or present participle
NNS	Noun plural	VBN	Verb past participle
NNP	Proper noun singular	VBP	Verb non-3rd person singular present
NNPS	Proper noun plural	VBZ	Verb 3rd person singular present
PDT	Predeterminer	WDT	Wh-determiner
POS	Possessive ending	WP	Wh-pronoun
		WP\$	Possessive wh-pronoun
		WRB	Wh-adverb

Abbildung 2.3: Das Penn Treebank Tag Set besteht aus 36 verschiedenen Wortartkategorien. Nomen werden zum Beispiel anhand ihres Numerus unterschieden. Für Verben werden verschiedene Konjugationsformen betrachtet und Adverbien und Adjektive werden nach Komparativ und Superlativ getrennt betrachtet.

Um nun diese Zuordnung für unbekannte Texte vornehmen zu können, müssen Algorithmen zuvor mittels Trainingsdaten geschult werden. Hierzu stellt die Computerlinguistik eine Vielzahl unterschiedlicher Techniken bereit, wobei diese meistens auf stochastischen Berechnungen basieren. Anhand von manuell annotierten Trainingsdaten erlernen die Algorithmen Regeln und Zuordnungswahrscheinlichkeiten. Etwaige Ambiguitäten können anhand von Auftretungshäufigkeiten oder ähnlichen statistischen Berechnungen aufgelöst werden.

Für Anwendungen innerhalb der Meinungsanalyse kann auf unterschiedliche *Tagging-Algorithmen* zurückgegriffen werden, die als Textanalysebibliotheken frei im Netz verfügbar sind. Die *OpenNLP Bibliothek*¹⁹ stellt hierfür einen wahrscheinlichkeitsbasierten Algorithmus bereit. Berger et al. präsentieren in [16] einen Algorithmus, der auf Entropie-Berechnungen basiert und als *Maxent Tagger*²⁰ im Netz veröffentlicht ist. Und die *Stanford Core NLP Library*²¹ beinhaltet eine Implementation des Algorithmuses aus [17]. Die unterschiedlichen Ansätze basieren jeweils auf dem gleichen *Penn Treebank Tag Set* und erreichen eine außerordentlich gute Genauigkeit. Bedingt durch unterschiedliche Trainingsdaten und die daraus erlernten Zuordnungswahrscheinlichkeiten können sich aber dennoch voneinander abweichende Wortartentscheidungen ergeben.

Die syntaktische Computerlinguistik stellt nicht nur Methoden zur Identifikation von Wortarten bereit, sondern es besteht auch die Möglichkeit, auf **Parser** zurückzugreifen. Diese können anhand

¹⁹<http://opennlp.sourceforge.net/README.html>

²⁰<http://maxent.sourceforge.net/about.html>

²¹<http://nlp.stanford.edu/software/corenlp.shtml>

der Wortartinformationen die Syntaxstrukturen der Sätze ableiten und modellieren hierdurch grammatikalische Zusammenhänge. Hierzu wird meist ein Automat verwendet, der mittels Regeln die natürliche Sprache abstrahiert. Abhängig von der *Parsing-Strategie* werden entweder ausgehend von einem Startsymbol (*top-down*) oder ausgehend von den einzelnen Wörtern (*bottom-up*) Ersetzungsregeln des Automaten angewendet, die die syntaktische Satzstruktur modellieren. Anhand der angewendeten Ersetzungsregeln lässt sich ein *Ableitungsbaum*, wie in Abbildung 2.4 dargestellt, ableiten. Aus der so gewonnenen Syntaxstruktur lassen sich zum Beispiel Satzbestandteile wie Nomenketten oder Nebensätze leichter identifizieren. Um strukturelle Ambiguitäten auflösen zu können, werden die unterschiedlichen Strukturen bewertet und die bestmögliche Struktur ausgewählt. Der *Stanford-Parser* [18] verwendet hierfür zum Beispiel Markov-Wahrscheinlichkeiten, die aus den Trainingsdaten berechnet wurden.

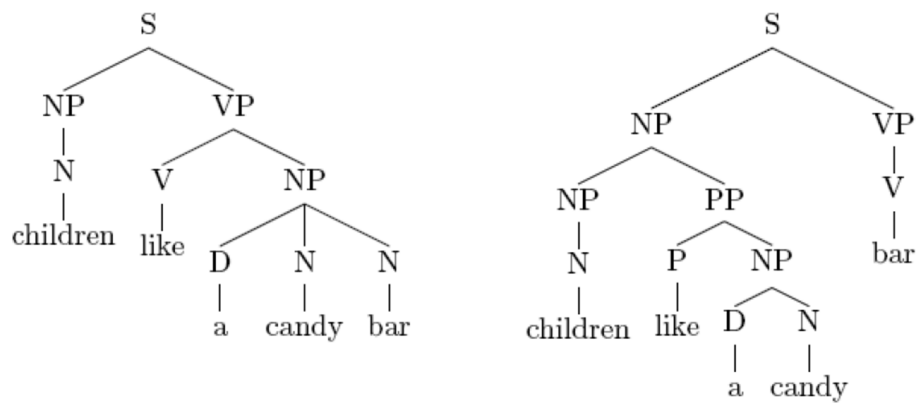


Abbildung 2.4: Die abgebildeten Ableitungsstrukturen resultieren aus dem Beispielsatz „children like a candy bar“. Dieser enthält strukturelle Ambiguitäten, welche zu zwei unterschiedlichen Ergebnissen führen. Hierbei wurde die Nomenkette „a candy bar“ im rechten Ableitungsbaum aufgebrochen.

Eine weitere Quelle für strukturelle Informationen stellen **Dependenzgrammatiken** dar. Hier werden Zusammenhänge zwischen Wortpaaren identifiziert, welche sich aus deren grammatikalischen Abhängigkeiten ergeben [19]. So deklarieren beispielsweise Substantive inhaltliche Aussagen, Verben beschreiben Ereignisse, Adjektive spezifizieren Eigenschaften von Substantiven und Adverbien definieren Eigenschaften von Ereignissen [20]. Die verschiedenen grammatikalischen Auswirkungen der Wörter untereinander werden durch Abhängigkeiten ausgedrückt. Eine solche Abhängigkeit besteht zwischen einem bestimmenden Wort (*governor*) und einem von ihm abhängigen Wort (*dependent*) [20] und definiert eine spezifische grammatikalische Funktion. Im Unterschied zum Syntaxbaum steht bei *Dependenzgrammatiken* das Verb im Zentrum des Satzes. Abbildung 2.5 veranschaulicht die Ausgabe des *Parsers* aus [21].

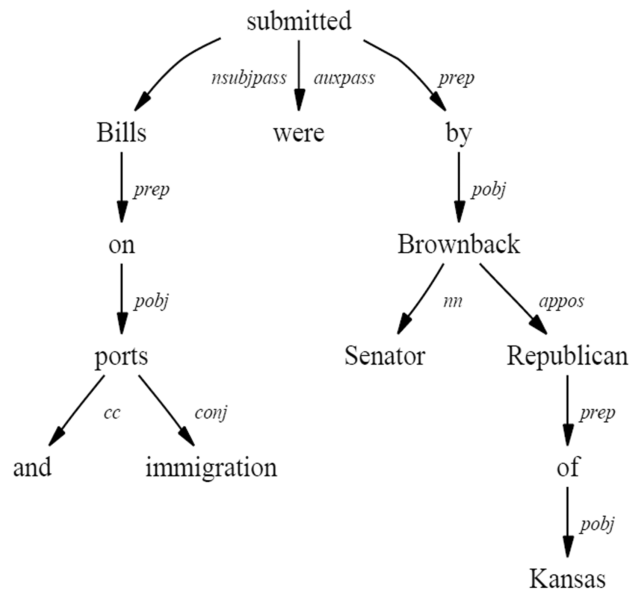


Abbildung 2.5: Ein Beispiel eines *Dependency-Parsers* für den englischen Satz „Bills on ports and immigration were submitted by Senator Brownback, Republican of Kansas“.

Die Ausgabe eines solchen Parsers beschreibt einen gerichteten Graphen, wobei die Wörter des Satzes, die Knoten und die grammatikalischen Abhängigkeiten durch die Kanten beschrieben werden. Transitivität und Reflexivität sind hierbei nicht erlaubt und jeder Knoten besitzt jeweils nur eine ihn bestimmende Relation [19]. Die Autoren Marneffe et al. stellen in [21] einen Ansatz vor, der Abhängigkeitsrelationen, ausgehend von den Syntaxstrukturen der Sätze, extrahiert. Hierbei können 52 verschiedene grammatikalische Relationen berechnet werden, welche unter anderem Negationen, Konjunktionen oder Modifikatoren umfassen. Für die Meinungsanalyse können diese grammatikalischen Zusammenhänge herangezogen werden, um zum Beispiel Beziehungen zwischen Merkmalen (Substantive) und Meinungswörtern (Adjektive) identifizieren zu können [22].

2.4.4 Semantische Ebene

Die semantische Ebene beschreibt die Zuordnung von Bedeutungen für einzelne Sätze oder deren Bestandteile. Hierbei existieren viele verschiedene sprachwissenschaftliche Teilbereiche, wobei für die Computerlinguistik primär das Gebiet der Anaphorik von Bedeutung ist.

Die **Anaphorik** behandelt hierbei Verweise eines Satzteils auf einen anderen. Diese können, wie in Abbildung 2.6 ersichtlich, durch Pronomen, Nominalphrasen oder Zahlwörter ausgedrückt werden.

John found the love of **his** life.
The relationship did not last long.
He started a new **one**.

Abbildung 2.6: Die Pronomen „his“ und „he“ referenzieren jeweils auf die Person „John“. „the relationship“ bezieht sich auf „the love“ des vorherigen Satzes und „one“ bezieht sich auf „the relationship“. Diese entsprechen Relationen zwischen Nominalphrasen.

Die sogenannte *Anapherauflösung* kann diese Verweise mittels computerlinguistischen Techniken zurückverfolgen und auflösen. Die Autoren Lee et al. entwickelten hierzu ein Verfahren, das innerhalb der *Stanford Core NLP Library* als Implementation vorliegt [23]. Speziell für die Identifikation von merkmalsbezogenen Meinungen kann diese Technik zu einem verbesserten Ergebnis führen, da satzübergreifende Relationen zurückverfolgt werden können [24].

2.4.5 Wissensbasen

Die semantischen Informationen innerhalb der Texte sind für eine Maschine nur schwer begreifbar. Dabei behilft sich die Computerlinguistik, indem sie diese mit Informationen aus externen Quellen anreichert. Hier besteht ein breites Angebot an externen Informationsquellen wie unterschiedliche Lexika, Thesauri oder Wissensdatenbanken. Anhand der dort verfügbaren Daten können die Algorithmen der Meinungsanalyse mit semantischen Informationen angereichert werden.

Das Wissen aus einem **Lexikon** wird häufig zur Informationsextraktion von Eigennamen verwendet. Hierzu soll zum Beispiel aus unstrukturiertem Text ein strukturierter Datensatz generiert werden, welcher in einer Datenbank abgespeichert werden kann. Somit müssen Textinhalte in einem vorab bekannten Lexikon nachgeschlagen werden, um diese dann mit einem spezifischen Thema assoziieren zu können. So kann zum Beispiel eine Bewerbung im Informatikumfeld nach Namen von Programmiersprachen oder Plattformen durchsucht werden, um hieraus Fähigkeiten des Bewerbers abzuleiten. Mittels spezieller Lexika ist es folglich möglich, Namen von Personen, Firmen, Städten oder ähnlichem aus Texten zu kategorisieren.

Um unbekannte Kategorien identifizieren zu können, kann auf das Lexikon „Wikipedia“²² zurückgegriffen werden. Dieses wird im Verfahren von Kazama et al. verwendet, um anhand der Beschreibungen innerhalb eines Artikels eine Kategorie für einen unbekanntem Eigennamen abzuleiten. Zum Beispiel kann aus dem Satz „Rare Jimi Hendrix song draft sells for almost \$17,000.“ für den Eigennamen „Jimi Hendrix“ die Kategorie *Musiker* identifiziert werden [25].

²²<http://www.wikipedia.org>

Für Algorithmen der Meinungsanalyse können auch Relationen zwischen Wörtern von Interesse sein. So enthält ein **Thesaurus-Lexikon**, wie zum Beispiel *WordNet*²³, semantische und lexikalische Beziehungen zwischen Wörtern. Hierzu werden diese zunächst anhand ihrer Bedeutung in sogenannte *Synsets* gruppiert und über Relationen miteinander verknüpft. So existieren für Nomen zum Beispiel Beziehungen zu Überbegriffen, Unterbegriffen oder Bestandteilen, während Adjektive mit Synonymen und Antonymen in Verbindung gestellt werden. Diese Beziehungen können dann in der Meinungsanalyse verwendet werden, um beispielsweise ähnliche Produktmerkmale zu gruppieren oder meinungstragende Begriffe zu identifizieren.

2.5 Identifikatoren

Bevor es zu einer Identifikation von merkmalsbezogenen Meinungen kommen kann, müssen zuvor Techniken eingesetzt werden, die Meinungskandidaten und Produktmerkmalskandidaten aus Textdaten extrahieren. Diese können dann anschließend dazu verwendet werden, um merkmalsbezogene Meinungen identifizieren zu können. Dieses Unterkapitel betrachtet nun diese drei Identifikationsschritte als jeweils separate Probleme und stellt die prominentesten Lösungen aus der wissenschaftlichen Literatur vor.

2.5.1 Meinungen

Die Identifikation von meinungstragenden Ausdrücken stellt den Kernpunkt der Meinungsanalyse dar. Ziel ist es, ein Meinungslexikon zu erstellen, welches für Wörter angibt, ob diese eine *positive*, *negative* oder *neutrale* Ausrichtung besitzen. Da nach [26] die Subjektivität sehr stark vom Auftauchen von Adjektiven abhängt, werden häufig nur Adjektive betrachtet. Pang und Lee finden aber auch in Nomen und Verben meinungstragende Bedeutungen [10]. Für eine Identifikation kann auf drei unterschiedliche Herangehensweisen zurückgegriffen werden, welche nun im weiteren erläutert werden.

Eine der trivialsten Methoden beschreibt das **manuelle Erstellen** eines Meinungslexikons, welches aber sehr zeitintensiv ist. Daher wird dieser Ansatz oftmals nur als Kontrollmethode oder in Kombination mit anderen Methoden verwendet.

Der **Wörterbuch-basierte Ansatz** erweitert mithilfe von Thesauri eine geringe, manuell erstellte Meinungswörtermenge, indem nach Synonymen und Antonymen gesucht wird [27]. Die Autoren Hu et al. suchen mittels ihres Verfahrens für noch unbekannte Wörter nach Synonym- oder

²³<http://wordnet.princeton.edu/>

Antonymverknüpfungen zu einem bereits bekannten meinungstragenden Wort. Zum Beispiel folgt für das Adjektiv „*horrible*“ eine negative Ausrichtung, da es ein Antonym zum bereits bekannten positiven Wort „*beautiful*“ darstellt. Somit können bekannte Ausrichtungen dazu verwendet werden, noch unbekannte Meinungswörter zu identifizieren. Die Anwendung *SentiWordNet* [28] verwendet ebenfalls dieses Verfahren und stellt das generierte Meinungswörterbuch der Wissenschaft zur freien Verfügung.

Turney et al. berechnen für die Bestimmung von Meinungen die *Mutual Information* von Textphrasen zu den Wörtern „*excellent*“ und „*poor*“, indem sie Statistiken aus Websuchmaschinen verwenden [29].

Wörterbuch-basierte Methoden eignen sich sehr gut zum Auffinden von Wörtern, welche sich in Wissensbasen nachschlagen lassen. Da diese einen breiten Horizont abdecken, entsteht oftmals ein sehr großes Wörterbuch, das durch einen Benutzer validiert werden muss. Der Standpunkt, dass sich Meinungsbedeutungen besonders in unterschiedlichen Domänen unterscheiden, wird aber von Wörterbuch-basierten Techniken nicht betrachtet.

Um Domänen-spezifische Wörterbücher generieren zu können [4], wird auf **Korpus-basierte Ansätze** zurückgegriffen. Hier wird die Meinungsbedeutung der Wörter meistens mithilfe von Regeln abgeleitet, die sich auf Informationen aus den Textdaten stützen. Dies können Vorkommen bestimmter Wörter sein [30], Informationen über Wortarten [12] oder grammatikalische Abhängigkeitseigenschaften [22] [31]. Ähnlich wie beim Wörterbuch-basierten Ansatz wird eine vorab definierte Liste von Meinungswörtern benötigt, welche dann als Grundlage für das Finden weiterer meinungstragenden Wörter dient. Somit unterscheiden sich Korpus-basierte Techniken anhand ihrer vorab definierten Regeln und den computerlinguistischen oder statistischen Informationen auf welchen die Methoden aufbauen.

Die Autoren Hazivassilogou und McKeown definieren hierzu gewisse linguistische Bedingungen, um die Menge von bekannten Meinungswörtern erweitern zu können [30]. Abbildung 2.7 veranschaulicht hierbei eine Auswahl dieser Regeln. So spricht der Algorithmus zum Beispiel den Wörtern auf beiden Seiten einer *und*-Konjunktion dieselbe Orientierung zu. So kann dem Wort „*fast*“ demnach eine positive Ausrichtung zugewiesen werden. Eine zweite Regel besagt, dass das Vorkommen des Wortes „*but*“ die bisherige Meinungsorientierung für das nachfolgende Adjektiv invertiert. Somit folgt für „*slowly*“ eine negative Ausrichtung.

and-Rule: the camera's zoom is nice and fast.
 but-Rule: the interface is comfortable but runs slowly.

Abbildung 2.7: In zwei Beispielsätzen finden sich zwei unterschiedliche Regeln der Autoren Hazivassilogou und McKeown, um Meinungswörter identifizieren zu können.

Der Ansatz von Qui et al. verwendet Abhängigkeits- und Wortart-Eigenschaften, um Meinungswörter zu identifizieren [22]. Die Autoren beschränken sich hierzu auf spezielle Arten von Abhängigkeiten wie Konjunktionen und Modifikatoren. Damit ist es möglich, nicht nur Abhängigkeiten zwischen Meinungswörtern zu betrachten, sondern auch zwischen Produktmerkmalen und Adjektiven. Abbildung 2.8 stellt diesen Mehrwert beispielhaft dar. Zusätzlich modellieren die Autoren auch indirekte Abhängigkeiten über ein weiteres Wort. Somit lassen sich anhand von vier verschiedenen Regeln meinungstragende Wörter identifizieren. Da nicht alle Regeln eine Orientierung bestimmen können, verwenden die Autoren hierzu den Ansatz von Hazivassilogou und McKeown [30]. Diese Herangehensweise ist besonders dadurch interessant, dass nicht nur Meinungswörter alleinstehend betrachtet werden, sondern sie auch oftmals in Verbindung mit Nomen stehen.

Dependency Rule 1: the camera's zoom is fast and quiet.
↑
CONJ

Dependency Rule 2: the camera's zoom is fast and quiet.
↑
MOD

Abbildung 2.8: In zwei Beispielsätzen finden sich zwei unterschiedliche Regeln von Quiu et al., um Meinungswörter zu identifizieren. *Regel 1* beinhaltet eine konjunktionale Abhängigkeit zwischen einem bekannten Meinungswort und einem Adjektiv. *Regel 2* identifiziert Meinungswörter, indem festgelegt wird, dass diese als Adjektiv im Text vorkommen und Nomen modifizieren. Die Meinungsorientierung kann mithilfe von *Regel 1* identifiziert werden, wohingegen diese durch *Regel 2* nicht berechnet werden kann.

Die Autoren Popescu und Etzioni berücksichtigen zur Identifikation von potentiellen Meinungswörtern ebenfalls Nachbarschaftseigenschaften von Produktmerkmalen [31]. Um die Methode von Hu et al. aus [27] zu verbessern, betrachten sie Nachbarschaften nicht nur innerhalb eines bestimmten Fensters, sondern verwenden Abhängigkeitseigenschaften, welche in Extraktionsregeln zusammengefasst werden. Hierbei verwenden sie Konjunktionen, Diskonjunktionen und manuell erstellte Abhängigkeitsregeln. Abbildung 2.9 gibt Beispiele dieser Regeln an.

Anhand der Regeln können nun potentielle Meinungskandidaten identifiziert werden. Um die Orientierung der Meinungswörter berechnen zu können, verwenden die Autoren Zusatzinformationen aus einem Thesauruslexikon und morphologische Wortzusammenhänge. Die Zuordnung wird über eine sogenannte *Relaxation Labeling*-Technik erreicht. Diese berechnet über eine *Support Function* Wahrscheinlichkeiten, die den Wörtern anhand ihrer Nachbarschaftseigenschaften eine spezifische Meinungsausrichtung zusprechen. Ist zum Beispiel ein noch unbekanntes Adjektiv Teil einer *und*-

Extraction Rules	Examples
if $\exists(M, NP = f) \rightarrow po = M$	(expensive) scanner
if $\exists(S = f, P, O) \rightarrow po = O$	lamp has (problems)
if $\exists(S, P, O = f) \rightarrow po = P$	I (hate) this scanner
if $\exists(S = f, P, O) \rightarrow po = P$	program (crashed)

Abbildung 2.9: Die Tabelle erläutert 4 von 10 Regeln von Popescu et al. zur Identifikation potentieller Meinungsphrasen. Hierfür definiert die erste Spalte die Regel und die zweite Spalte gibt hierzu ein reales Beispiel. po steht für die identifizierte Meinungsphrase, M steht für eine modifizierende Dependency-Eigenschaft, NP beschreibt eine Nomenphrase, S ein Subjekt, P ein Prädikat und O ein Objekt.

Konjunktion mit einem Wort mit bekannter Meinungsausrichtung, so kann die Orientierung mit hoher Wahrscheinlichkeit auf dieses übertragen werden. Dieser iterative Prozess passt pro Schritt die Zuordnungswahrscheinlichkeiten an.

Ding und Liu erweiterten die Idee, die Meinungsausrichtung in der Umgebung von Produktmerkmalen zu untersuchen, indem sie Meinungswörter nicht nur domänenspezifisch betrachten. Denn das gleiche Wort kann selbst innerhalb einer Domäne unterschiedliche Bedeutungen besitzen. Es ist somit abhängig vom Kontext, in welchem es verwendet wird [12]. Zum Beispiel hat in der Kameradomäne das Wort „lang“ einen positiven Bezug auf die Akkulaufzeit, wohingegen eine lange Auslösezeit des Blitzes negativ behaftet ist. Deswegen schlagen die Autoren vor, meinungstragende Wörter nicht nur alleinstehend zu betrachten, sondern zusätzlich in Bezug mit einem Produktmerkmal zu stellen.

Aus den verschiedenen Methoden lässt sich ableiten, dass sich durch unterschiedliche Herangehensweisen ähnliche Meinungslexika erzeugen lassen. Domänen-spezifische Ansätze erzeugen geringere Datenmengen und sind weniger übertragbar. Im Gegensatz dazu beinhalten Domänenübergreifende Lösungen viel mehr meinungstragende Wörter, welche aber nicht Kontext-sensitiv sind. Um den Kontext in eine Anwendung einfließen zu lassen, muss entweder ein Benutzer eingreifen oder der Bezug zu Produktmerkmalen muss mitbetrachtet werden.

2.5.2 Merkmale

Die Identifikation von Produktmerkmalen oder Wörtern, worüber eine Meinung geäußert wird (im weiteren der Einheitlichkeit wegen *Feature* genannt), kann durch unterschiedliche Herangehensweisen modelliert werden. Diese verfolgen das Ziel bestimmte Wörter innerhalb der Texte als Feature zu markieren. Häufig beschränken sich die Techniken hierbei auf Nominalphrasen.

Triviale Lösungen vergleichen einzelne Wörter oder Wortketten mit bereits existierenden Wörterbüchern, um zum Beispiel Namen von Organisationen, Personen oder Länder identifizieren zu können. Innerhalb der Texte kann aber auch nach vorab definierten Mustern wie Adressen oder

Telefonnummern gesucht werden. Diese Methoden eignen sich aber lediglich für spezifische Fälle und nur unter der Bedingung, dass der Suchgegenstand oder zumindest deren Form bekannt ist.

Korpus-basierte Methoden untersuchen die eigentlichen Textdaten und identifizieren Features ähnlich zu den Identifikationsverfahren von Meinungswörtern. Nicht selten basieren diese Methoden auf computerlinguistischen Informationen und finden Features über Strukturregeln. Dabei müssen die identifizierten Features bewertet werden, um eine Menge von tatsächlichen Features zu erhalten.

Die Autoren Hu und Liu verwenden hierzu Assoziationsregeln, welche auf den Wortartinformationen aufbauen [27]. Die resultierenden Nomenketten werden nach ihrer Auftretungshäufigkeit und Länge beurteilt. Seltene Features und Features mit mehr als drei Wörtern werden vorerst nicht weiter betrachtet. Um nun die bestehende Menge weiter zu reduzieren, wenden die Autoren *Compactness Pruning* auf Features mit zwei Wörtern an, siehe Abbildung 2.10, und für Features mit einem Wort wird *Redundancy Pruning* eingesetzt, siehe Abbildung 2.11 [32].

I had searched for a digital camera for 3 months.

This is the best digital camera on the market.

The camera does not have a digital zoom.

Abbildung 2.10: Für kompakte Features gilt, dass deren Wortbestandteile innerhalb der Sätze mindestens zwei Mal gemeinsam vorkommen, wobei deren Abstand nicht größer als 3 sein darf. Somit ist „*digital camera*“ in den ersten beiden Sätzen kompakt, im letzten Beispielsatz aber nicht. Insgesamt ist das kompakte Vorkommen über dem Schwellenwert und reicht aus, um nicht aus der Featuremenge ausgeschlossen zu werden.

Feature	Sentence Occurrence
manual	10
manual mode	4
manual setting	3

Abbildung 2.11: Um redundante Features identifizieren zu können, betrachtet man deren Vorkommen im Satz. Hierbei entscheidet die Anzahl, wie oft dieses Feature nicht als Teilmenge von anderen Features vorkommt. Ist diese unterhalb einem Schwellenwert (kleiner als 3), so wird dieses aus der Featuremenge ausgeschlossen. Das Feature „*manual*“ kommt hierbei in 7 Fällen als Teilmenge von anderen Features vor und somit 3 Mal als eigenständiges Feature. Es muss also nicht ausgeschlossen werden.

Um seltene Features zu identifizieren, betrachten die Autoren diejenigen Sätze, welche Meinungswörter in der Nachbarschaft von Nomen enthalten. Die nächstgelegene Nomenkette stellt somit ebenfalls ein Feature dar.

Informationen zu Produktmerkmalen können aber auch aus **externen Quellen** gewonnen werden. Popescu et al. erweitern die vorhergehenden Ideen von Hu und Liu, indem sie nicht nur die Auf-

tretungshäufigkeiten betrachten, sondern die identifizierten Nomenketten mittels eines Domänen-unabhängigen Informationsextraktionssystems validieren [31]. Hierzu wird jeweils die *Point-wise Mutual Information* anhand von Websuchmaschinenstatistiken zwischen identifizierten Nomenketten und vorab definierten Phrasen berechnet. Diese modellieren in natürlicher Sprache die Domäne und die Art des Produktbezuges. Für die Scanner-Domäne zum Beispiel werden hierzu Sätze wie „of scanner“, „scanner has“ oder „scanner comes with“ angewendet. Dieser Ansatz reichert also mithilfe zusätzlicher Informationsquellen die identifizierten Features mit Vorwissen an, wodurch inkorrekte Features ausgeschlossen werden können.

Eine weitere, im Web frei verfügbare, Informationsquelle bildet das Informationsportal „Wikipedia“, welches von den Autoren Kazama und Torisawa in [25] eingesetzt wird. Die dort enthaltenen Artikel und Beschreibungen werden für eine recht einfache und wirkungsvolle Kategorisierung von Eigennamen verwendet.

Zur Identifikation von Produktmerkmalskandidaten können aber auch interne Informationen wie bereits berechnete Meinungswörter zurate gezogen werden. Der bereits im vorherigen Kapitel vorgestellte Ansatz der Autoren Qiu et al. [22] verwendet nun die Abhängigkeitseigenschaften, um auch unbekannte Features zu identifizieren. Hierzu erweitern die Autoren ihre Regelsammlung um weitere vier Regeln, wobei ein Teil dieser in Abbildung 2.12 dargestellt ist.

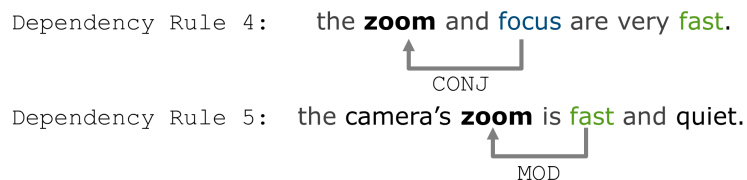


Abbildung 2.12: In zwei Beispielsätzen finden sich zwei unterschiedliche Regeln von Qiu et al., um Features zu identifizieren. *Regel 4* beinhaltet eine konjunktionale Abhängigkeit zwischen einem bekannten Feature und einem Nomen. *Regel 5* identifiziert Features, indem festgelegt wird, dass diese als Nomen im Text vorkommen und von einem bereits bekannten Meinungswort modifiziert werden.

Die verschiedenen Ansätze lassen darauf schließen, dass Features durch deren Wortart und Auftretungshäufigkeit beurteilt werden können. Doch auch seltene Features können einen hohen Bedeutungsgrad innerhalb der Meinungsanalyse haben, wodurch diese ebenfalls mitbetrachtet werden müssen. Die Hinzunahme von externen Informationsquellen kann das Ergebnis verfeinern und bietet sich für Domänen-übergreifende Lösungen an.

2.5.3 Referenzen

Nach getrennter Identifikation von Meinungskandidaten und Produktmerkmalskandidaten können nun diese beiden in Beziehung gesetzt werden, um merkmalsbezogene Meinungen extrahieren zu

können. Frühe Methoden verwenden hierbei den Grundsatz, dass sich eine Meinung in der Nähe des zu beschreibenden Objektes befindet. Meist verwenden sie ein Fenster mit festgelegter Größe und suchen in der Nachbarschaft des Meinungswortes nach dem nächstgelegenen Produktmerkmalskandidaten [27].

Die Einbeziehung der Abstände zwischen Meinungskandidat und Produktmerkmalskandidat findet in [12] als Gewicht für die Meinungsbestimmung des Satzes statt. Für jeden Produktmerkmalskandidaten innerhalb eines Satzes werden die Meinungsrichtungen der einzelnen identifizierten Meinungen aufsummiert und mit deren Abstand zueinander gewichtet. Somit bekommen Meinungswörter, die nahe beim Feature stehen, ein höheres Gewicht zugesprochen.

Betrachtet man die Methode von Qiu et al. in [22], welche bereits zur Identifikation von Meinungskandidaten oder Produktmerkmalskandidaten eingesetzt wird, so kann man diese auch zur Referenzidentifikation gewinnen. Hierzu werden nun nicht mehr unbekannte Informationen gesammelt, sondern die identifizierten Kandidaten werden über die schon bekannten Abhängigkeitsregeln verifiziert. Besteht zum Beispiel zwischen einem Meinungskandidaten und einem Produktmerkmalskandidaten eine bestimmunggebende Abhängigkeit, so können diese in Beziehung zueinander gesetzt werden.

Ein differenzierter Ansatz beschreibt die Arbeit von Kisilevich et al. [6], welche mittels 12 vorab definierten Mustern nach Beziehungen innerhalb der Texte suchen. Diese Muster setzen entweder vor oder nach einer Nomenkette an und beinhalten Beschränkungen zu einer Wortart oder zu einem spezifischen Wort. Kisilevich et al. verwenden diese Muster, um in der Foto-Domäne Referenzen, welche die Fotoqualität betreffen, zu identifizieren. Abbildung 2.13 zeigt eine Übersicht über diese Regeln.

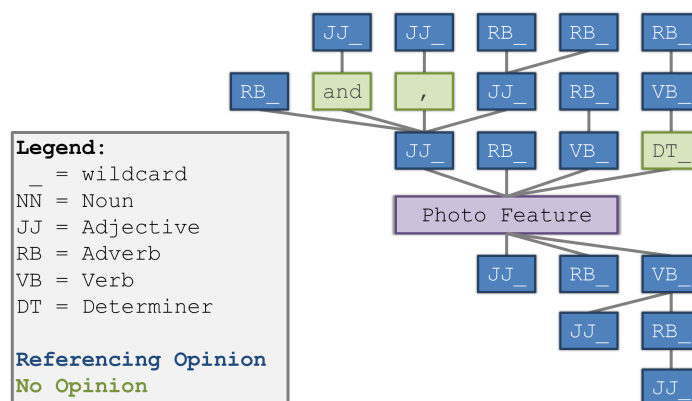


Abbildung 2.13: Eine Übersicht derjenigen Muster, welche zur Identifikation von Referenzen in [6] verwendet werden. Hierbei beschreibt jedes Rechteck ein Wort, sortiert nach der Reihenfolge oben (vor einem Feature) bis unten (nach dem Feature). Hierbei unterscheiden die Autoren nach meinungstragenden Wörtern und Wörtern mit keiner Meinungsbedeutung.

Referenzen bieten den Mehrwert, dass sie Meinungen mit Produktmerkmalen verknüpfen. Hierdurch können erstmals Kontextabhängigkeiten modelliert werden, welche besonders zur Bestimmung der Meinungsausrichtung für unbekannte Wörter Verwendung finden. Abbildung 2.12 beschreibt diese Problematik in zwei beispielhaften Sätzen.

The **printer** prints **fast**.
The **ink cartridge** runs out **fast**.

Abbildung 2.14: Für diese zwei Sätze soll nun bestimmt werden, welche Ausrichtung das Meinungswort „fast“ besitzt. Hier kann nur der Kontext entscheiden, also der Bezug zum Produktmerkmal. Es müsste folglich eine positive Ausrichtung in Bezug zur Druckgeschwindigkeit und eine negative Ausrichtung in Bezug zum Tintenverbrauch resultieren.

Die Autoren Ding, Liu und Yu beschreiben in ihrer Veröffentlichung ein Verfahren, das diese Kontextabhängigkeit erkennen und aufschlüsseln kann [12]. Ihr ganzheitlicher Ansatz betrachtet Meinungen immer im Bezug zu einem Produktmerkmal und bestimmt unbekannte Meinungsausrichtungen anhand von Verbindungsregeln der Sätze. Hierzu setzen die Autoren darauf, dass sich die Meinungsäußerungen der Rezensenten nach einem bestimmten Schema verhalten. Meistens behalten sie ihre Meinungsausrichtung bei, bis ein Signalwort eine Änderung der Orientierungen anzeigt. Dieses Schema soll nun anhand mehrerer Beispiele in Abbildung 2.15 verdeutlicht werden.

This camera takes **great** pictures and has a **long battery life**.
The camera has a **long battery life**, which is **great**.
But the **zoom** is **dreadful**.

Abbildung 2.15: In den Beispielsätzen ist jeweils das Meinungswort mit unbekannter Ausrichtung fett dargestellt. Nach Ding, Liu und Yu folgen die Meinungsäußerungen einem bestimmten Satzverbindungsschema. In *Satz 1* kann über die *und*-Konjunktion die Meinungsausrichtung von „great“ übernommen werden. Dies kann ebenso für Nebensätze (in *Satz 2*) und darauffolgende Sätze (in *Satz 3*) angewendet werden. Da in *Satz 3* das Signalwort „but“ auftritt, resultiert hierfür eine negative Ausrichtung.

Mittels dieser Technik können unbekannte Meinungsausrichtungen aus dem Kontext heraus identifiziert werden. Dieser Ansatz, merkmalsbezogene Meinungen zu extrahieren, ist in der Applikation „Opinion observer“ [12] implementiert. Die Evaluation dieses Systems zeigt, dass sich die Betrachtung von Kontextabhängigkeiten durchweg positiv auf das Ergebnis widerspiegelt.

2.5.4 Negationen

Um die Techniken der Meinungsanalyse abzurunden, sollen auch noch Negationen, welche die Meinungsausrichtung invertieren, berücksichtigt werden. Die Identifikation wird meistens über das Aufsuchen bestimmter Wörter wie „no“, „not“ und „never“ realisiert [12] und dann mit Meinungswörtern in nächster Nachbarschaft in Beziehung gesetzt. Die Autoren Hu und Liu beschränken sich hierbei auf einen maximalen Abstand von fünf Wörtern [27]. Aber auch Abhängigkeitseigenschaften

können zur Identifikation von Negationen beitragen.

Eine genauere Heuristik präsentieren die Autoren Ding, Liu und Yu in [12], indem sie zusätzlich Negationsregeln auf den Wortartinformationen verwenden. Zusätzlich betrachten sie das Vorkommen des Wortes „*but*“ oder dessen Synonyme als Meinungsinvertierer. Abbildung 2.16 gibt hierzu Beispiele für das Greifen dieser unterschiedlicher Negationsregeln.

- (1) NEG neg -> pos This camera has no **problem**.
- (2) NEG obj -> neg The zoom sometimes does **not work**.
- (3) stop VB`ing obj -> neg The camera **stopped working** after 3 days.
- (4) not just -> no I like this camera **not just** because it is beautiful.

Abbildung 2.16: Für *Satz 1* gilt die einfache Regel, dass das Vorkommen eines Negationswortes die Meinungsorientierung invertiert. In *Satz 2* wird ein objektives Wort durch Negation mit einem negativen Kontext behaftet. *Satz 3* beschreibt eine der auf POS-Tags aufgebauten Regeln. Hier liegt das Signalwort als Verb vor, auf das ein anderes Verb in einer *ing*-Form folgt. Da Zweites objektiv ist, wird *Regel 2* angewendet. Doch nicht jedes Vorkommen eines Negationswortes muss zwingend eine invertierende Bedeutung besitzen. *Satz 4* beschreibt hier beispielhaft eine Ausnahme, die nicht als Negation betrachtet werden darf.

Anhand von einfachen Negationsregeln kann also die Ausrichtung von Meinungswörtern in der Nachbarschaft invertiert werden, was zu einem genaueren Endergebnis führen kann.

2.6 Interaktive Analyse

Für die Analyseanwendung ist es von Vorteil, wenn auf das intrinsische Domänenwissen des Menschen zurückgegriffen werden kann. Dieses kann die Effektivität einer Anwendung im Gegensatz zu automatischen Methoden deutlich steigern [33]. Hier werden Techniken zur Verfügung gestellt, womit die Analysefragen des Benutzers innerhalb eines Informationsfindungsprozesses beantwortet werden können. Ein Basiswerkzeug für die Analyse von Daten beschreibt zum Beispiel die Suchanfrage. Diese ist sehr mächtig und kann eine breite Palette von Analyseprozessen abbilden. Ist eine Suchanfrage abgeschickt, so werden die zurückgelieferten Informationen vom Benutzer interpretiert und die bisherige Anfrage verfeinert. Dies geschieht so lange bis schließlich die Analysefrage beantwortet werden kann. Ein Nachteil dieses Werkzeug ist aber, dass der jeweilige Analyst bereits zu Anfang wissen muss, wonach er suchen soll [33]. Eine Analyseanwendung muss folglich einem Benutzer die Freiheit geben, seine Analysefrage sowie Analysestrategie selbst zu entwickeln. Dabei muss gegeben sein, dass während des Informationsfindungsprozesses keine neuen oder unerwarteten Inhalte, welche für den Kontext der Analyse wichtig sein könnten, übergangen werden [33]. In der wissenschaftlichen Literatur existieren hierbei unterschiedliche Herangehensweisen, wobei sich diese Arbeit auf die Konzepte der etablierten *Visuellen Analyse* und der Idee der *Analytischen Steuerung* stützt, welche nun im weiteren erörtert werden sollen.

2.6.1 Visuelle Analyse

Das Konzept der Visuellen Analyse beschreibt den Prozess, analytische Schlussfolgerungen anhand von visuellen und interaktiven Darstellungen treffen zu können [34]. Sie wird meist dort eingesetzt, wo der Datenraum durch seine Größe und Komplexität nicht mehr durchschaubar ist. Die Visuelle Analyse unterstützt beliebige heterogene Datenquellen, welche für die Analyse zunächst vorverarbeitet werden. Hierzu zählen zum Beispiel die Bereinigung von Störquellen oder die Normalisierung von Datenwerten. Diese werden schließlich durch Techniken des *Data Minings* in Modelle transformiert. Der Datenraum wird durch Visualisierungskomponenten dargestellt, worin ein Benutzer Muster oder Trends wahrnehmen kann. Mittels verschiedener Interaktionstechniken kann dieser durch den Datenraum navigieren, um nähere Erkenntnisse über die Daten gewinnen zu können. Hierbei orientiert sich die Visuelle Analyse an folgenden Richtlinien: [34]

- **Analyse First:** Der Benutzer hat zunächst die Möglichkeit den Datenraum zu überblicken. Anhand von Mustern innerhalb der Visualisierung können interessante Gebiete im Datenraum entdeckt werden.
- **Show the Important:** Dem Benutzer werden jeweils nur wichtige Informationen vorgelegt, welche sich im Kontext seiner Analyse befinden.
- **Zoom, Filter and Analyse Further:** Der Benutzer kann über Filtertechniken die Datenmenge eingrenzen oder über Zoomen in interessante Datenbereiche navigieren.
- **Details on Demand:** Der Benutzer kann jederzeit detailliertere Informationen anfordern.

Hierbei existiert ein gegenseitiges Wechselspiel zwischen den visuellen Komponenten und den automatischen Prozessen. Durch Parameterverfeinerungsschritte wird der Benutzer auf dem Weg seiner Analyse unterstützt, sodass er sich der Lösung seiner Frage annähern kann.

Die Anwendung *IN-SPIRE*²⁴ entspricht einem visuellen Analysesystem für Textdaten. Dabei können beliebig große Kollektionen von Textdokumenten im System verarbeitet werden. Diese werden zunächst vorverarbeitet und dann in hochdimensionale Worthistogrammvektoren umgerechnet. Danach werden diese Dokumente anhand ihrer Termähnlichkeiten gruppiert und über multidimensionale Skalierung auf einen zweidimensionalen Raum transformiert, wodurch die Dokumente folglich in einem Streudiagramm dargestellt werden können. Dabei entspricht ein Bildpunkt einem Dokument, wobei Dokumente mit ähnlichen Themeninhalten nahe beieinander platziert werden. Aus Mangel an zur Verfügung stehenden Bildpunkten werden Konzentrationen

²⁴IN-SPIRE, <http://in-spire.pnnl.gov>

von Dokumenten durch eine Hitzekarte visualisiert. Für die einzelnen Gruppen werden die Termrepräsentationen dargestellt. Abbildung 2.17 zeigt unterschiedliche Visualisierungen von *IN-SPIRE*.

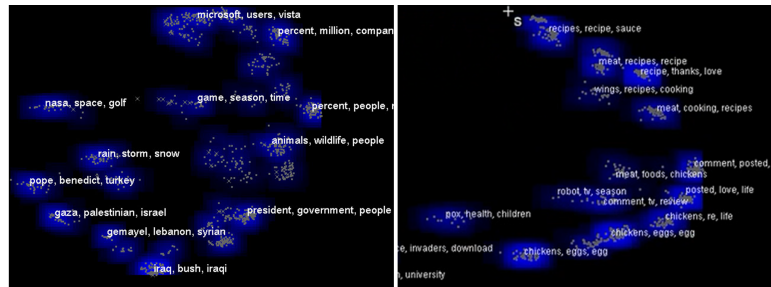


Abbildung 2.17: Das System *IN-SPIRE* beschreibt ein Visuelles Analyse System für Textdaten. Hierbei werden die Dokumente in einem Streudiagramm dargestellt, wobei ähnliche Dokumente mit ähnlichen Termen in Wolken gruppiert werden. Hier sind die Ergebnisse zweier unterschiedlicher Korpora abgebildet.

Der Analyst kann nun in diesen Daten über verschiedene Interaktionen navigieren. Über Suchanfragen können Dokumente und ihr Themenkontext hervorgehoben werden. Bedingt durch die Visualisierung, kann der Analyst die verschiedenen Gruppen von Dokumenten erkennen. Mittels Suchanfragen und Filtertechniken kann der Datenraum untersucht werden, wobei der Analyst die Möglichkeit hat, auch einzelne Dokumente zu betrachten. Für seine Analyse kann er zusätzlich Dokumente oder Terme als Ausreißer markieren, um somit den Datenraum einzugrenzen und seine Sicht auf die Daten zu verfeinern.

Der Vorteil der Visuellen Analyse besteht in den frei definierbaren Fragestellungen des Analysten. Dabei kann er im Verlauf seiner Analyse auf interessante Funde stoßen, welche wiederum neue Hypothesen generieren können. Im Gegensatz zu einer Suchanfrage, beschreibt das Ziel der Visuellen Analyse die Generierung eines Wissens über die Daten, welches über Interaktionen in Erfahrung gebracht wurde.

2.6.2 Analytische Steuerung

Die Analytische Steuerung beschäftigt sich mit der Herausforderung, den Anwendungskontext einer Analyseaufgabe in den Informationsfindungsprozess einzubeziehen. Dieser wird durch die Thematik der Analyse definiert, wie zum Beispiel eine Untersuchung über verschiedene menschliche Krankheiten im Gesundheitsbereich [33]. Die Analytische Steuerung stellt nicht nur das Sammeln von Informationen aus den Daten in den Vordergrund, sondern versucht auch innerhalb dieser, Beziehungen, Themen oder Muster passend zur Anwendungsaufgabe darzustellen. Der Analyst kann hierbei die Sichten innerhalb des unüberwachten Raumes kontrollieren und steuern. Dadurch, dass sich die Daten an den vorab definierten Anwendungskontext anpassen, fällt es dem Analysten einfacher, seine Analysefragen zu beantworten. Die Autoren Bohn, Calapristi et al. betrachten hierbei

folgende Herausforderungen für einen Analyseprozess [33]:

- Der Analyst muss zunächst seine Analysefrage und Analysestrategie definieren. Diese müssen bereits im Vorfeld bekannt sein, damit gemessen werden kann, welche Informationen aus den Daten für den Analyseprozess am relevantesten sind.
- Betrachtet man das Themenspektrum innerhalb der Daten, so können sich dort durchaus Unterschiede ergeben. Für eine Analyse ist es wichtig, schwer wahrnehmbare sowie untergeordnete Themen aus den Daten zu isolieren. Denn diese können häufig eine entscheidende Aussagekraft für den Anwendungskontext des Analysten besitzen und dürfen folglich nicht von dominanten Themenbereichen überdeckt werden.
- Ebenso ist es wichtig Beziehungen zwischen Themen entlang der Hauptinteressenachse zu identifizieren. Denn der Analyst ist nicht nur am Aufdecken von existierenden Themen interessiert, sondern möchte auch untersuchen, welche Beziehungen in den Daten existieren und welche Relevanz diese Themen im Bezug zum Analysekontext besitzen.
- Während des Informationsfindungsprozesses dürfen trotz der Manipulation der Datenansicht keine neuen oder unerwarteten Inhalte, welche sich im Kontext der Analyse befinden, übergangen werden.

Diese theoretischen Aspekte finden sich im Ansatz der Analytischen Steuerung wieder. Dabei entwickelten die Autoren eine Anwendung, welche auf das visuelle Analysesystem *IN-SPIRE*²⁵ aufgebaut ist. Die Autoren übernahmen hierfür lediglich die Darstellungstechnik, um sie in ihrem eigenen Analyseprozess einzubetten [33]. Um den Informationsaustausch zwischen Mensch und Maschine zu ermöglichen, legten sich die Autoren auf die folgende drei Prozesse fest [33]:

- Der Analyst spezifiziert eine Menge an Termen, welche zusammen einen Kontext formulieren. Dieser stellt eine spezielle Bedeutung für die Analyserichtlinie dar.
- Innerhalb der Systemebene werden diese Terme in eine Charakterisierung des Korpus über-
setzt und mit automatisch abgeleiteten thematischen Schlüsselwörtern der Dokumente ver-
mischt. Dadurch werden schließlich die Wahrnehmungs- und Navigations-Interessen des Ana-
lysten unterstützt.
- Es werden Steuerelemente bereitgestellt, welche dem Benutzer nach eigenem Ermessen erlau-
ben, den Einfluss seiner Interaktionen zu kontrollieren.

²⁵IN-SPIRE, <http://in-spire.pnnl.gov>

Der interaktive Steuerungsprozess wird durch zwei unterschiedliche Betrachtungsweisen modelliert. Hierbei kann der Analyst den Kontext durch eine negative oder positive Steuerung kontrollieren. Eine negative Steuerung wird durch Entfernen von benutzerspezifischen Dokumenten und Termen erreicht. Es können zum Beispiel Dokumenten-Ausreißer entfernt werden, da sie visuell sehr gut ersichtlich sind. Dieses Steuerungsprinzip entspricht einer Korrektur von Fehlern oder Defekten, wodurch sich eine Annäherung an eine fokussiertere Sicht auf die Daten ergibt. Term-Ausreißer sind hierbei durch ihre Einbettung innerhalb eines Themenkontextes weniger visuell wahrnehmbar. Eine Entfernung dieser entspricht einer Korrektur von „analytischen Defekten“, um zum Beispiel dominante Termcharakteristiken abwerten zu können [33].

Eine positive Steuerung beschreibt das Einfügen von Kontextwissen in den Analyseprozess. Dieses wird durch kleine fokussierte Termgruppen ausgedrückt, welche zur Manipulation der Datenrepräsentation verwendet werden. Der Analyst beschreibt mit diesen Termgruppen somit seine Sichtweise auf die Daten [33]. Diesem wird hierdurch die Möglichkeit gegeben, die Darstellung der Daten an seinen Analysekontext anzupassen.

Der Analyseprozess wird folglich durch die positiven und negativen Korrekturen von Termen gesteuert, wobei hier der Datenraum manipuliert wird. Hierdurch nähert sich der Analyst an die optimale Darstellung, passend auf seinen jeweiligen Anwendungskontext, an.

Kapitel 3

Problemstellung

Bevor nun auf die Umsetzung eingegangen werden kann, muss zunächst die Problemstellung erläutert werden. Diese beginnt mit der Analyse von vorab bekannten Schwierigkeiten im Meinungsanalyseumfeld, aus welchen sich dann die Anforderungen und Lösungswege für die Umsetzung ergeben. Zusätzlich werden die einzelnen Hypothesen dieser Arbeit beleuchtet und hieraus abschließend Folgerungen für die Umsetzung und eine spätere Evaluation aufgezeigt.

3.1 Herausforderungen

Dieses Unterkapitel beleuchtet die Herausforderungen, die sich aus den Inhalten des vorhergehenden Grundlagenkapitels ergeben. Ein besonderes Augenmerk ist hierbei auf die Themenstellung dieser Arbeit gerichtet. Hierfür sollen zunächst die Begriffe *Meinungsanalyse* und *Meinung* beleuchtet werden. Auch der Schwerpunkt der Interaktiven Steuerung sowie die Herausforderungen für den Analyseprozess und eine spätere Evaluation werden näher betrachtet.

3.1.1 Meinungsanalyse

Diese Arbeit beschäftigt sich primär mit der **Betrachtung von merkmalsbezogenen Meinungen**. Es reicht also nicht aus, Meinungswörter und Produktmerkmale getrennt zu betrachten, sondern deren Beziehungen stehen im Vordergrund. Hierzu werden drei unterschiedliche Identifikationsverfahren benötigt, wobei jeder Arbeitsschritt alleinstehend betrachtet bereits sehr herausfordernd sein kann [12]. Sind diese Informationen noch zusätzlich implizit im Text enthalten, müssen sie aus dem Kontext des Textes abgeleitet werden, wobei viele der automatischen Identifikationsmethoden bereits an ihre Grenzen stoßen.

Eine weitere Herausforderung ist die **Vorabauswahl der computerlinguistischen Informationen**, auf welchen Identifikationen stattfinden können. Kapitel 2.5 gab bereits Aufschluss über verschiedene Ansätze, welche sich an unterschiedlichen Informationsquellen bedienen. Grenzt man diese Auswahl ein, so beschränkt man ebenfalls die Menge von verwendbaren Identifikationsmethoden, beziehungsweise es verringern sich die Möglichkeiten, auf verschiedene Weisen, Meinungen und Produktmerkmale sowie deren Beziehungen zu identifizieren. Die Analyseanwendung soll folglich möglichst viele Informationen bereitstellen, um eine möglichst breite Palette an Algorithmen bedienen zu können. Häufig werden morphologische Einheiten, wie Tokenisierung, Stammformreduktion oder Wortarterkennung eingesetzt sowie grammatikalische Beziehungen vorab berechnet. Aber auch externe Informationsquellen können als Wissenslieferant dienen. Es muss demnach ein harmonisches Maß gefunden werden, welches eine ausreichend hohe Anzahl an unterschiedlichen Identifikationsmethoden zulässt, jedoch den Aufwand für eine Umsetzung einer solchen Analyseanwendung nicht sprengt.

Die Verwendung von **mehreren Methoden für die gleiche Identifikationsaufgabe** kann hierbei die Chance eines domänenabdeckenden Methodenumfanges erhöhen. So können sich die unterschiedlichen Vorteile der Techniken in verschiedenen Textformen und Textsparten gegenseitig ergänzen, wodurch sich unterschiedliche Betrachtungsweisen modellieren lassen. Unter dieser Voraussetzung besteht nun für eine Identifikationsaufgabe, wie zum Beispiel dem Erkennen von Meinungswörtern, eine Auswahlmöglichkeit zwischen unterschiedlichen Methoden. Da die Analyseanwendung nicht domänenspezifisch arbeiten soll, sondern gleichzeitig und ohne Vorwissen verschiedene Sparten abdecken soll, müssen diese Identifikationsmethoden parallel arbeiten und deren einzelnen Entscheidungen auf eine Gesamtentscheidung abstrahiert werden. Werden zum Beispiel zur Identifikation von Meinungen verschiedene Wörterbuchmethoden eingesetzt, setzt sich die daraus resultierende Gesamtentscheidung über das tatsächliche Bestehen eines Meinungswortes somit aus den Einzelmethodenentscheidungen der Wörterbücher zusammen. Die Herausforderung ergibt sich folglich aus der Auswahl der Identifikationsmethoden und der Errechnung einer Gesamtentscheidung.

Betrachtet man hier den **Begriff der Unsicherheit**, so kann diese über die Entscheidungen der Einzelmethode modelliert werden. Treffen mehrere Methoden die gleiche Entscheidung, zum Beispiel über das tatsächliche Bestehen eines positiven Meinungswortes, so kann von einer höheren Genauigkeit der Gesamtentscheidung ausgegangen werden, als wenn beispielsweise alle Entscheidungen unterschiedlich ausfallen würden. Ein Algorithmus kann somit einer sicheren Gesamtentscheidung mehr vertrauen als einer unsicheren. Anhand dieser Vertrauensbeziehungen kann die

resultierende Gesamtentscheidung für die Analyse und für einen Benutzer transparent dargestellt werden.

Transparenz ist aber auch gefordert, sobald sich **Abhängigkeiten zwischen verschiedenen Methoden** ergeben. In Kapitel 2.5.1 wurde hierzu eine Meinungswörterbuchmethode vorgestellt, welche eine bestimmte Wortart einfordert. Hier entsteht folglich ein methodenübergreifender Zusammenhang. Betrachtet man abermals den Begriff der Unsicherheit so wird klar, dass eine bestehende Unsicherheit auch auf die von ihm abhängige Methode übertragen werden sollte. So kann sich eine Wörterbuchmethode nicht mehr so sicher sein, wenn bereits bekannt ist, dass das Meinungswort nicht eindeutig ein Adjektiv repräsentiert. Für die Meinungsanalyse ist es demnach wichtig, diese methodenübergreifenden Beziehungen zu modellieren und durch Aufdecken von Unsicherheiten der Analyse und dem Benutzer eine nötige Transparenz über sämtliche zu treffenden Entscheidungen zu geben.

3.1.2 Meinungsbegriff

Der Meinungsbegriff beinhaltet drei unterschiedliche Facetten. Der Meinungsinhaber äußert eine Meinung über ein Objekt, wobei diese eine spezifische Meinungsausrichtung besitzt [3].

Betrachtet man diesen Begriff für unterschiedliche Textdaten, so kann sich dieser durchaus unterscheiden. Im Produktumfeld werden Meinungen über Produkte, Produktmerkmale oder Produktbestandteile geäußert, wohingegen im Nachrichtenumfeld Bewertungen der Autoren viel schwerer zu identifizieren sind. Zum Beispiel enthält ein Bericht über ein negatives Ereignis zwar negative Wörter, diese müssen aber nicht zwingend bewertende Funktionen annehmen. Ebenso variieren die Meinungsäußerungen bei verschiedenen Autoren, alleine durch deren unterschiedlichen Schreibstile.

Das Bestehen eines Meinungswortes wird oftmals auf spezifische Wortartgruppen wie Adjektive und Adverbien festgelegt [26]. Dies muss aber nicht immer der Fall sein. Denn ebenso kann eine subjektive Meinung durch Verben oder Nomen ausgedrückt werden [3]. Auch können ganze Sätze als meinungstragend gelten oder die Meinung wird durch bestimmte Umstände impliziert, wodurch sie nicht direkt im Text auftaucht [4]. Betrachtet man die Beziehungen von Meinungswörtern, so kann eine Meinung mehrere im Satz befindlichen Produktmerkmale modifizieren oder der Bezug reicht über die Satzgrenzen hinaus. Ebenso kann ein einzelnes Produktmerkmal durch verschiedene Meinungswörter bewertet werden. Innerhalb eines Satzes können sich die Meinungsrichtungen unterscheiden, wobei sich die Orientierungswechsel an bestimmte Regeln halten [12]. Ein weiterer wichtiger Punkt ist das Mitbetrachten von Negationen, welche die Orientierungen der Meinungswörter invertieren. Auch das Vorkommen von spartenspezifischen Produktmerkmalen beeinflusst

die Meinungsausrichtung, welche zusätzlich durch spezifische Wörter verstärkt oder abgeschwächt werden kann.

Um die Subjektivität eines Textes zu bestimmen, reicht es nicht aus, sich zum Beispiel nur auf das vom Benutzer angegebene Rating zu verlassen. So kann zum Beispiel eine positive Rezension nahezu keine Meinungen enthalten, wohingegen eine neutrale Bewertung sehr viele unterschiedliche Meinungen beinhalten kann [3].

Die Herausforderungen bestehen hier im Konzeptentwurf des Datenmodells, um verschiedene Meinungsbedeutungen aufnehmen zu können. Gesucht wird folglich eine möglichst einheitliche und spartenübergreifende Begriffsinterpretation, welche resistent gegenüber den Meinungs- und Produkt-spezifischen Begriffen innerhalb der Texte ist.

3.1.3 Interaktive Steuerung

Um das in Kapitel 2.6 angesprochene Weltwissen des Benutzers mit in die Anwendung fließen zu lassen, muss auf visuelle und interaktive Techniken zurückgegriffen werden, um ihn bei der Meinungsanalyse mitwirken lassen zu können. Hierzu existieren bereits viele wissenschaftliche Arbeiten, welche sich mit der Analyse von Meinungen aus Textdaten beschäftigen. Dabei wird aber der für diese Arbeit wichtige Aspekt, über die Interaktive Steuerung den zugrundeliegenden Analyseprozesses verbessern zu können, nicht betrachtet.

Eine Analyseanwendung muss zunächst einmal den Analyst bei seinem Informationsfindungsprozess unterstützen. So sollte er zum Beispiel in den Daten navigieren können oder über Suchanfragen Themenkontexte hervorheben können. Durch geeignete Visualisierungen kann der Analyst die verschiedenen Informationen in den Daten erkennen. Filtertechniken helfen ihm dabei, den Datenraum einzugrenzen und näher zu untersuchen. Dabei soll der Analyst die Möglichkeit haben, auch einzelne Datenpunkte betrachten zu können. Für eine Analyseanwendung besteht die Herausforderung hierbei bei der Auswahl der passenden Interaktionen, welche den Analysten in effektiver und effizienter Weise bei der Beantwortung seiner Analysefragen unterstützen. Dabei müssen auch passende Visualisierungen eingesetzt werden, worin Muster oder Trends innerhalb der Daten ersichtlich sind.

Für die Meinungsanalyse ist zusätzlich der Ansatz einer interaktiven Manipulation des zugrundeliegenden Analyseprozesses interessant. Hierbei besteht die Herausforderung darin, eine analytische Steuerung zu konzipieren, welche das Weltwissen des Analysten einbindet und den Analyseprozess durch positive oder negative Korrekturen zu verbessern. Diese ermöglichen es, die Datenansicht

an den jeweiligen Anwendungskontext wie zum Beispiel eine spezifische Produktparte anzupassen. Werden die dafür nötigen Interaktionen des Analysten gespeichert, entsteht demnach eine auf einen spezifischen Anwendungsfall konzipierte Prozessoptimierung.

3.1.4 Textdaten

Die Herausforderungen für Textdaten sind zum einen die Beschaffung dieser und zum anderen der spartenspezifische Umgang mit ihnen.

Im **Internet** besteht ein sehr großes Angebot an Portalen, die Benutzerrezensionen zulassen. Diese bieten aber keinen direkten Zugang zu ihren Datenbanken, was die Beschaffung erschwert. Sie müssen zunächst heruntergeladen und von Strukturdaten wie HTML²⁶-Tags oder ähnlichem bereinigt werden. Es existieren aber auch einige Korpora, die von Autoren innerhalb des Meinungsanalyseumfeldes erstellt und der Allgemeinheit zur Verfügung gestellt werden.

Die Abstimmung der zur Verfügung stehenden **Identifikationsmethoden** auf verschiedene Textdaten muss ebenfalls bedacht werden, da dort die Schreibstile durchaus unterschiedlich ausfallen können. Algorithmen zur Wortartbestimmung sind zum Beispiel auf bestimmte Bereiche trainiert oder sind auf kurzen Sätzen angewendet weniger robust. Ebenso enthalten Wörterbücher meist nur eine Auswahl von Begriffen, welche aus spezifischen Kollektionen stammen und somit den Begriffraum nicht komplett abdecken. Die Identifikation von produktarttypischen Begriffen wie Herstellernamen, Produktnamen oder speziellen Produktmerkmalen muss ebenso über spartenübergreifende Techniken erfolgen. Insgesamt besteht hierbei die Herausforderung, dass die Identifikationsmethoden unterschiedliche Textarten und Textformen unterstützen sollen, um nicht nur spartenspezifisch gute Ergebnisse liefern zu können.

Das Datenmodell muss also resistent gegenüber **verschiedener Eingabedaten** sein und diese ohne Probleme verarbeiten können. Hierzu zählen zum Beispiel spezielle Verwendungen wie Emoticons, Abkürzungen oder kurze Sätze. Ebenso stehen verschiedene Sprachen zur Auswahl, welche innerhalb der Daten verwendet werden. Eine weitere Herausforderung besteht in der Datengröße. So sollen große Datenmengen genauso robust und effektiv bearbeitet werden können wie kleine Datenmengen.

²⁶Hypertext Markup Language

3.1.5 Evaluation

Für Anwendungen des Meinungsanalyseumfeldes ist besonders die Evaluation der berechneten Informationen interessant. Eine Herausforderung ist hier, den fehlenden Goldstandard innerhalb dieses wissenschaftlichen Gebietes zu umgehen. In der Literatur sind zwar einige Anreizpunkte zu finden, jedoch müssen diese auch zum gegebenen Anwendungsfall passen. Oftmals beschreiben die annotierten Dokumentensammlungen aber die Ausgabe von automatischen Klassifikatoren oder beinhalten zu viel oder zu wenig Annotationsmaterial. Der Weg, selbst Daten zu annotieren, ist durch den Umfang dieser Arbeit eingeschränkt. Eine besondere Herausforderung ist es aber auch, die Benutzerinteraktionen zu erfassen und auf das Ergebnis zu abstrahieren. Die Evaluation muss folglich nicht nur die unterschiedlichen Identifikationsmethoden untersuchen, sondern muss auch beantworten können, warum eine Benutzerintervention nötig ist und wie sich diese auf das Ergebnis auswirkt.

3.2 Lösungsansätze

Nachdem nun die Herausforderungen für Anwendungen innerhalb der Meinungsanalyse erörtert wurden, kann nun auf die Bewältigung dieser eingegangen werden.

Zunächst wird der Lösungsansatz zur Modellierung von Sicherheiten vorgestellt. Dieser vereint Methoden zur Lösung desselben Identifikationsproblems in einem Modul, wodurch Sicherheiten für einen Analyseprozess abgeleitet werden können. Zusätzlich wird auf die Folgerungen dieses Ansatzes eingegangen, insbesondere welche Chancen sich hieraus ergeben. Danach wird dieser Lösungsansatz auf die Interaktive Steuerung angewendet und erörtert wie das Weltwissen des Benutzers in einen interaktiven Analyseprozess integriert werden kann. Daraufhin wird aufgezeigt, welche computerlinguistischen Informationen aus unstrukturierten Texten extrahiert werden sollen. Auf diesen setzen die Identifikationen von Meinungen, Produktmerkmalen und Referenzen auf, welche danach vorgestellt werden. Hier wird genauer auf die Unterschiede der Verfahren eingegangen, um den domänenübergreifenden Standpunkt dieser Arbeit unterstreichen zu können.

3.2.1 Meinungsanalyse mit Sicherheiten

Im Bereich der Meinungsanalyse haben sich sehr viele unterschiedliche Herangehensweisen zur Lösung spezifischer Identifikationsprobleme etabliert. Die Forschungsziele innerhalb dieses Themengebietes beruhen meistens auf Verbesserungen von bestehenden Techniken und evaluieren diese unter Berechnung eines Genauigkeitswertes. Ein genauer Einblick in die unterschiedlichen Arbeitsweisen der Verfahren wird nicht betrachtet. Die Identifikation derselben Information wie zum Beispiel

einer Wortart können durch verschiedene Verfahren auf unterschiedliche Wege erreicht werden. So besitzen diese Verfahren spezifische Charakteristiken, welche in verschiedenen Fällen im Hinblick auf ein korrektes Ergebnis von Vorteil, aber auch von Nachteil sein können. Diese Stärken und Schwächen sind zum Beispiel auf das erlernte Wissen aus unterschiedlichen Trainingsdaten zurückzuführen. Diese Arbeit versucht nun diese unterschiedlichen Charakteristiken miteinander zu kombinieren, um hieraus Vorteile für die Meinungsanalyse zu gewinnen. Im weiteren wird nun aufgezeigt, wie dieser Lösungsansatz definiert ist und welche Folgerungen hieraus ableitbar sind.

Lösungsansatz Methoden, welche zur Lösung des gleichen Identifikationsproblems beitragen, sollen zu einem Modul zusammengeführt werden. Der schematische Grundsatz ist hierbei in Abbildung 3.1 dargestellt.

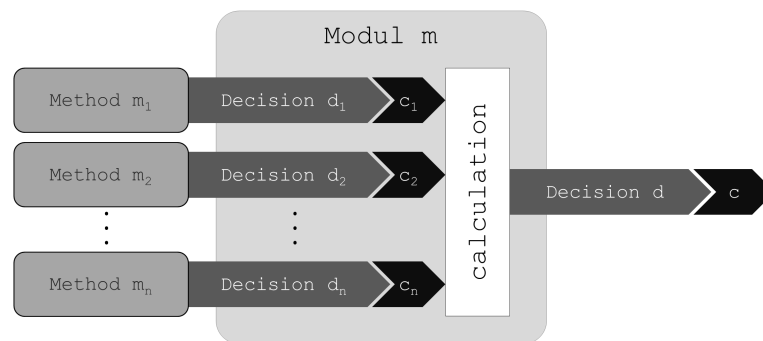


Abbildung 3.1: Ein Modul beinhaltet mehrere Methoden, welche unterschiedliche Herangehensweisen desselben Identifikationsproblems modellieren. Diese erzeugen jeweils spezifische Entscheidungen, welche innerhalb des Moduls zu einer Gesamtentcheidung verrechnet werden müssen. Jede Entscheidung ist zusätzlich mit einem Faktor behaftet, welcher deren Sicherheit ausdrückt.

Eine Identifikationsmethode beschreibt demnach ein Modul (*Modul m*), welches mehrere Einzelmethoden enthält. Jede Einzelmethode ($m_1 \dots m_n$) identifiziert nach ihrem Verfahren ein bestimmtes Ergebnis ($d_1 \dots d_n$) und teilt diese einer Berechnungsfunktion (*calculation*) innerhalb des Moduls mit. Zusätzlich wird an jede Entscheidung ein spezifischer Faktor ($c_1 \dots c_n$) angehängt. Aus den Einzelentscheidungen wird dann eine Modulentscheidung (d) und ein daraus resultierender Sicherheitsfaktor (c) errechnet.

Der Sicherheitsfaktor beschreibt die Sicherheit der Entscheidung und wird Vertrauenswert oder auf Englisch „*confidence*“ genannt. Er reicht von 0 wie überhaupt keine Sicherheit bis 1, wodurch eine vollkommene Sicherheit ausgedrückt wird. Dieser Faktor kann zum Beispiel anhand der Übereinstimmungen der Einzelentscheidungen abgeleitet werden und somit die prozentuale Übereinstimmung einer Mehrheitsentscheidung beschreiben. Diese Berechnung kann aber auch anhand anderer Merkmale erfolgen.

Anhand des Vertrauenswertes lassen sich nun verschiedene Folgerungen für den Meinungsanalyseprozess ableiten. Die Werte geben zum Beispiel Aufschluss über die **Vertrauenswürdigkeit der Entscheidungen des Moduls**. Hierdurch können Entscheidungen priorisiert werden. Ein Analyseprozess könnte zum Beispiel gänzlich unsichere Entscheidungen ignorieren, um ein Mindestmaß an Sicherheit zu gewährleisten. Auch hat ein Benutzer nun die Möglichkeit, auf Unsicherheiten des Systems hingewiesen zu werden, um diese zum Beispiel korrigieren zu können. Wird eine **Modulentscheidung in anderen Methoden wiederverwendet**, so kann der Sicherheitsfaktor miteinbezogen werden, um die neue Modulentscheidung transparenter zu gestalten. Diese kann somit nur noch mindestens so vertrauenswürdig sein, wie deren abhängigen Modulentscheidungen. Der Analyseprozess kann folglich nachvollziehen, warum eine Unsicherheit entsteht, woher diese stammt und ein Benutzer kann diese dann zu einem späteren Zeitpunkt korrigieren. Dadurch, dass die Einzelmethoden ebenfalls einen Sicherheitsfaktor besitzen, können einzelne **Entscheidungen manipuliert werden**. So entspricht der Wert einem spezifischen Gewicht für die Einzelentscheidung. Der Analyseprozess könnte zum Beispiel unsichere Methoden durch Senkung dieser Gewichte bestrafen. Auch ein Benutzer kann diese Gewichte zu analytischen Zwecken beeinflussen, um zum Beispiel Fehler zwischen Methoden aufzudecken oder um das Endergebnis zu verfeinern.

3.2.2 Meinungsbegriff

Der Meinungsbegriff innerhalb dieser Arbeit beinhaltet zwei unterschiedliche Definitionen, welche sich während des Analyseprozesses ändern. Zunächst werden meinungstragende Wörter als Meinungskandidaten identifiziert. Erst durch die Beziehung zu einem Produktmerkmal wandeln sich Meinungskandidaten in tatsächliche Meinungen, welche die in Kapitel 3.1 vorgestellten Facetten wiedergeben. Diese werden durch die Referenz erfasst. Diese Trennung ermöglicht somit eine transparente Einsicht in den Analyseprozess. Für die Themenstellung dieser Arbeit beschreibt das Feststellen des Meinungsinhabers keine Bedeutung, wodurch dieser aus dem Konzeptentwurf des Datenmodells herausgenommen werden kann. Ebenso wird nur die englische Sprache berücksichtigt, um die Komplexität der Anwendung zu verringern.

Die Identifikation von Meinungskandidaten behandelt hierbei noch keine kontextabhängigen Informationen. Dieser Identifikationsschritt beschreibt lediglich das Erkennen von eventuell meinungstragenden Begriffen innerhalb der Texte. Folglich muss der Kontext innerhalb der Referenzen miteinbezogen werden. Dort können spartenspezifische Informationen mittels der Abhängigkeit zu Produktmerkmalen gespeichert werden. Beispielsweise trägt der Meinungskandidat „*rasch*“ eine positive Ausrichtung. Steht er aber in Beziehung mit dem negativen Produktmerkmal wie „*Tintenverbrauch*“, so invertiert diese Begebenheit die Meinungsausrichtung der Referenz, wobei der

Meinungskandidat unberührt bleiben kann. Zusätzlich können in der Referenz Negationen mitbetrachtet werden. So wird im Beispiel „*kein rascher Tintenverbrauch*“ die Meinungsrichtung der Referenz abermals invertiert ohne die Begebenheit des negativen Merkmals anzutasten. Dieser Konzeptentwurf lässt folglich kontextsensitive Informationen zu, während die Meinungskandidatidentifikation als alleinstehender Prozess betrachtet werden kann. Eine Meinung ist demnach auf nur ein Wort reduziert. Das Modell wird durch den Einbezug von Negationen komplettiert.

3.2.3 Interaktive Steuerung

Bisher wurde der Aspekt einer Interaktiven Steuerung anhand der Analyse von Dokumenten und den darin enthaltenen Termen veranschaulicht. Dieser Ansatz soll nun auf die Themenstellung dieser Arbeit angewendet werden, wobei hier Meinungen aus Textdaten analysiert werden.

3.2.3.1 Integration eines Weltwissens

Der Anwendungskontext der Analyse wird durch eine spezifische Domäne definiert. Zunächst stehen dem Analyseprozess keine Informationen über Domänenabhängigkeiten wie zum Beispiel Kontextsensitive Produktmerkmale oder Mehrdeutigkeiten zur Verfügung. Die benötigten Informationen sind im Weltwissen des Analysten verankert. Hier muss folglich ein Transfer des Anwendungswissens des Analysten in den Analyseprozess stattfinden. Dieser wird durch eine *negative Steuerung* erreicht [33]. Dabei kann der Analyst die berechneten Informationen des Analyseprozesses korrigieren, wobei ihm die wichtigen Informationen zuerst vorgelegt werden [34]. Bezieht man hier den Begriff der Unsicherheit mit ein, so können die Vertrauenswerte des Analyseprozesses ein Indiz für diese geforderte Wichtigkeit sein. Abbildung 3.2 zeigt die schematische Integration des Weltwissen des Benutzers in den bisherigen Lösungsansatz.

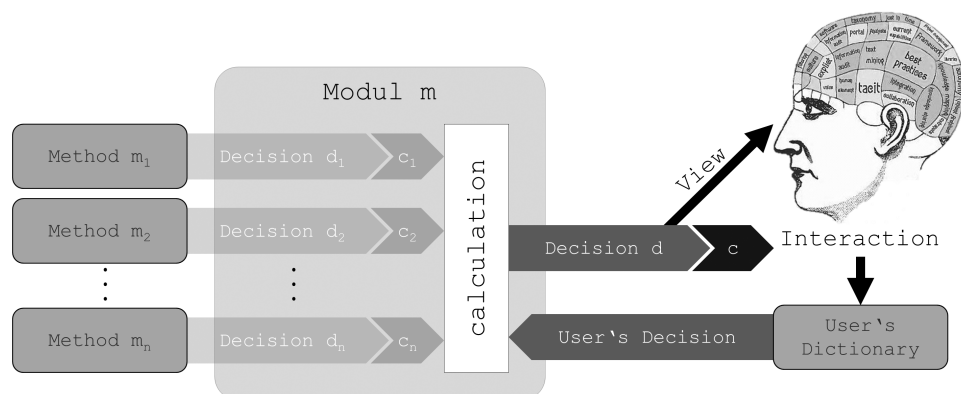


Abbildung 3.2: Eine Handlung des Benutzers wird durch die Vertrauenswürdigkeit der Modulentcheidung angeregt. Dieser interagiert mit dem System, wobei die getätigte Benutzerentscheidung innerhalb eines Benutzerwörterbuchs abgelegt wird. Diese Entscheidung genießt die höchste Vertrauenswürdigkeit und übergeht somit die bisherigen Methodenentscheidungen.

Der bisherige Lösungsansatz der Modulentscheidungen wird durch den Einbezug des Benutzers erweitert. Diesem werden mittels visueller Techniken Informationen zu den Vertrauenswerten der Modulentscheidungen vorgelegt. Bedingt durch sein Kontextverständnis und seiner Erfahrung kann er hierbei Fehler des Systems erkennen. Durch interaktive Techniken kann der Benutzer nun eigene Entscheidungen treffen, welche innerhalb eines Benutzerwörterbuches abgelegt werden. Hierbei können fehlerhafte Entscheidungen korrigiert oder unsichere Entscheidungen bestärkt werden. Dadurch, dass dem Benutzer die höchste Vertrauenswürdigkeit zugesprochen werden kann, können die bisherigen Methodenentscheidungen für diesen spezifischen Änderungsfall übergangen werden. Um den Ansatz der Interaktiven Steuerung zu komplettieren, können auch *positive Steuerungsinformationen* des Analysten erfolgen. Betrachtet man zum Beispiel einen Korpus, worin vorab keine Identifikationen berechnet wurden, so kann hier ein positiver Transfer des Anwendungswissens eines Analysten erfolgen. Dies entspricht dem eigenständigen Entwickeln eigener Identifikationsmodule.

3.2.3.2 Korrekturen

Diese Arbeit befasst sich hierbei mit Korrekturen von vier unterschiedlichen Informationen, welche den Ansatz der Interaktiven Steuerung modellieren. Dabei können die Identifikationsmodule der Wortarten, Meinungskandidaten, Produktmerkmalskandidaten und Referenzen vom Analysten beeinflusst werden. Um die Interaktionen effizienter zu gestalten, werden dem Analysten zusätzlich automatisch berechnete Vorschläge zum Beispiel für ähnliche Fehlerfälle vorgelegt. Dabei kann er relevante Fälle auswählen, wodurch die Änderungen auch auf diese übertragen werden.

Wortarten Eine Änderung der Wortart wird erreicht, indem ein Term einen neuen *Tag* zugewiesen bekommt, welcher über das zugrundeliegende *TagSet* definiert ist. Der Analyst möchte Wortarten korrigieren, da diese Informationen von unterschiedlichsten Modulen wiederverwendet werden. Dadurch, dass die Identifikationsmethoden auf Trainingsdaten basieren, können die Wortarten zu unbekanntem Wörtern wie Eigennamen oder technischen Begriffen nicht eindeutig bestimmt werden. Diese müssen folglich vom Analysten korrigiert werden. Ähnliche Fehlerfälle werden durch die Gleichartigkeit der Terme definiert. So können zum Beispiel alle Vorkommen des technischen Terms „*nokia*“ in Nomen umgewandelt werden.

Meinungen Die Änderung eines meinungstragenden Wortes beinhaltet die Modifikation dessen Meinungsorientierung. Dabei kann der Analyst sich auf eine positive, negative oder neutrale Ausrichtung festlegen. Ähnliche Fehlerfälle werden hier durch die Gestaltung der Terme definiert. Es macht zum Beispiel Sinn, eine Änderung auf identische Terme mit gleicher Wortart zu übertragen. Dies kann auch für deren Basisformen gelten, da Adverbien oder Superlative die gleiche Meinungs-

orientierung wie deren verwandten Adjektive besitzen können. Hier kann der Analyst zum Beispiel mit einer Korrektur das Adjektiv „*slow*“ als negativ markieren, während seine Verwandten „*slowest*“ und „*slowly*“ ebenfalls die gleiche Ausrichtung zugesprochen bekommen.

Produktmerkmale Ein Produktmerkmalskandidat kann sich aus mehreren Wörter zusammensetzen. Eine Korrektur wird dadurch erreicht, dass der Analyst einzelne Wörter markieren kann, welche entweder zum Merkmal gehören oder gelöscht werden können. Ähnliche Fehlerfälle beschreiben diejenigen Produktmerkmalskandidaten, welche diese markierten Wörter enthalten. Der Analyst kann sich nun entscheiden Merkmale anhand seiner Markierungen zu ändern oder diese aus dem Analyseprozess auszuschließen. Zum Beispiel erhält der Analyst durch die Markierung des Wortes „*day*“ innerhalb des Produktmerkmalskandidaten „*business day*“ Vorschläge wie „*today*“, „*saturday*“ oder „*birthday*“, welche folglich gelöscht werden können.

Referenzen Für die Änderungen der Referenzen sind viele verschiedenen Anwendungsfälle möglich. Zunächst einmal kann der Analyst Referenzen löschen oder in Produktmerkmale umwandeln. Hierdurch können zum Beispiel Produktmerkmalskandidaten, welche kein tatsächliches Produktmerkmal darstellen, aus dem Analyseprozess entfernt werden. Da die Referenz einen spezifischen Kontext ausdrückt, kann der Analyst diesen korrigieren. Korrekturen von Referenzen können auf ähnliche Produktmerkmale oder ähnliche Meinungswörter übertragen werden, da hierdurch eine Kombination von Meinung und Merkmal modelliert wird.

3.2.3.3 Visualisierung

Da der Aspekt der Visualisierung für diese Arbeit nicht die höchste Priorität besitzt soll zunächst eine Tabelle als Datenvisualisierung dienen. Diese gibt folglich die berechneten Daten in tabellarischer Form aus, wobei der Benutzer die Sortierreihenfolge beliebig beeinflussen kann. Wird diese Ansicht nach Vertrauenswert sortiert, so sieht der Analyst zum Beispiel entweder die unsichersten oder die sichersten Entscheidungen. Innerhalb des Themengebietes dieser Arbeit sind bereits etliche wissenschaftliche Verfahren zur Informationsvisualisierung bekannt, wodurch die einfache aber dennoch zweckmäßige Tabelle ausgetauscht werden könnte.

3.2.4 Informationen aus unstrukturiertem Text

Eine Extraktion von computerlinguistischen Informationen aus Textdaten kann mittels unterschiedlichen im Internet frei verfügbaren Bibliotheken erzielt werden. Eine der wohl größten und umfangreichsten Bibliotheken beschreibt die Algorithmensammlung der *Stanford Core NLP Library*²⁷.

²⁷<http://nlp.stanford.edu/software/corenlp.shtml>

Diese enthält eine Vielzahl von Techniken, wobei alle in Kapitel 2.4 angesprochenen sprachlichen Ebenen abgedeckt werden. Dort sind zum Beispiel die Segmentierung der Sätze in einzelne Terme, die Berechnungen deren Grundformen, das Parsen der grammatikalischen Syntaxstrukturen oder das Berechnen von Abhängigkeitseigenschaften enthalten. Zusätzlich können semantische Satz-zusammenhänge aufgelöst werden. Die Bibliothek ist für wissenschaftliche Zwecke innerhalb des Meinungsanalyseumfeldes konzipiert worden und ist leicht in eine Anwendung zu integrieren.

Zur Bestimmung von Wortarten kann auf mehrere Standardmethoden zurückgegriffen werden, welche ebenfalls innerhalb von Bibliotheken frei im Netz verfügbar sind. Hierbei handelt es sich um die bereits in Kapitel 2.4 vorgestellten Bibliotheken *Stanford Core NLP Library*²⁸, *Maxent Tagger Library*²⁹ und *Open NLP Library*³⁰. Diese Bibliotheken implementieren unterschiedliche Algorithmen zur Wortartbestimmung, wobei sie allesamt auf dem *Penn Treebank Tag Set* [35] aufbauen. Die berechneten Informationen können somit untereinander verglichen werden.

Zur Identifikation von Stopwörtern kann auf ein Standardwörterbuch oder ein auf den Finanzsektor ausgerichtetes Wörterbuch mit Währungs- und Landes-typischen Zusatzinformationen zurückgegriffen werden.

3.2.5 Identifikatoren

Da nun computerlinguistische Informationen mittels verschiedener Bibliotheken extrahiert werden können, sollen nun die darauf aufbauenden Techniken der Informatik vorgestellt werden. Dabei werden unterschiedliche Herangehensweisen zur Identifikation von Meinungskandidaten, Produktmerkmalskandidaten und Referenzen betrachtet, welche bereits in Kapitel 2.5 erläutert wurden. Das Hauptaugenmerk ist hierbei auf die jeweiligen Unterschiede der Verfahren gerichtet, wobei in dieser Arbeit eine möglichst domänenübergreifende Lösung erreicht werden soll.

3.2.5.1 Meinungen

Um Meinungen in Texten zu identifizieren, soll auf unterschiedliche Wörterbücher zurückgegriffen werden. Diese stammen aus verschiedenen wissenschaftlichen Arbeiten und sind im Internet frei verfügbar. Die in dieser Arbeit verwendeten Wörterbücher bilden unterschiedliche Sparten ab und unterscheiden sich zusätzlich in ihren zur Verfügung gestellten Informationen. Im Folgenden sollen nun die verschiedenen Wörterbücher näher vorgestellt werden.

General Inquirer Dictionary Dieses Wörterbuch wurde am *Massachusetts Institute of Technology* als Software zur Inhaltsanalyse entwickelt und beinhaltet mittlerweile 182 Kategorien, welche

²⁸<http://nlp.stanford.edu/software/corenlp.shtml>

²⁹<http://maxent.sourceforge.net/about.html>

³⁰<http://opennlp.sourceforge.net/README.html>

syntaktische, semantische und pragmatische Informationen zu englischsprachigen Wörtern bereitstellen [36]. Zur Bestimmung von Meinungsrichtungen sind aber nur die Listen der positiven und negativen Wörter interessant. In dieser Arbeit wird nicht die Originalversion, sondern eine bereits verfeinerte Auswahl verwendet. Die resultierende Wortliste ist unabhängig von Wortarten und umfasst 1559 positive und 1928 negative Wörter.

SentiWordNet Dictionary Dieses Wörterbuch stellt die Meinungsrichtung der Wörter in Abhängigkeit mit deren Wortart dar und ist als Ausgabe der bereits in Kapitel 2.5.1 vorgestellten Anwendung *SentiWordNet* [28] frei im Netz verfügbar. Die Meinungsrichtungen wurden anhand eines Thesauruslexikons berechnet. Dieses Wörterbuch umfasst alle Wörter innerhalb des Wörterbuches *WordNet* und teilt die Meinungsrichtung anhand der Wortarten auf. Es enthält folglich eine Vielzahl von gering subjektiven Ausdrücken, wobei diese domänenunabhängig gewonnen wurden. Abbildung 3.3 gibt hierbei einen Überblick über die Aufteilung der Meinungswörter innerhalb des Wörterbuches.

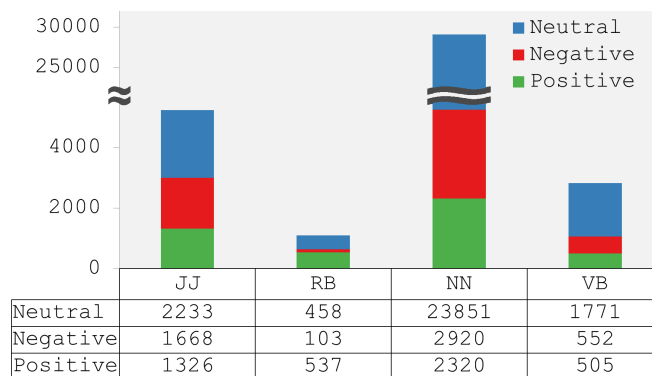


Abbildung 3.3: SentiWordNet unterteilt die berechneten Meinungswörter nach positiv, negativ und neutral und ist abhängig von der Wortart. Hierbei beschreiben die einzelnen Spalten jeweils eine Wortartklasse (*JJ*=Adjektiv, *RB*=Adverb, *NN*=Nomen, *VB*=Verb).

Bing Liu Dictionary Dieses Wörterbuch stammt aus Berechnungen der Autoren Hu et al. [27] [37]. Hierbei extrahierten die Autoren anhand ihrer Methode 2006 positive und 4783 negative Meinungswörter aus Textdaten, wobei hier häufige Rechtschreibfehler, morphologische Varianten und social-media-spezifische Wörter enthalten sind. Dieses Wörterbuch wurde aus Kundenrezensionen zu primär technischen Produkten generiert und deckt somit die Sparte von elektronischen Produkten ab.

Subjectivity Clues Dictionary Dieses Wörterbuch ist Teil der Anwendung *OpinionFinder* [38] und ist im Netz frei verfügbar. Es umfasst sowohl manuell ausgewählte, als auch automatisch extrahierte Meinungswörter aus Texten. Der Hauptteil des Wörterbuches entstand hierbei durch

die Arbeiten von Riloff und Wiebe in [39].

Hierbei sind folgende Informationen für die Meinungswörter verfügbar:

- **Typ:** Die Stärke der Subjektivität.
- **Wort:** Das Meinungswort selbst.
- **Wortart:** Die Wortart für das jeweilige Meinungswort.
- **Stammform:** Diese Option gibt an, ob die Meinungsausrichtung für alle Wörter mit angegebener Stammform gelten soll oder nicht.
- **Polarität:** Die Ausrichtung der Meinung. Diese kann als positiv, negativ, neutral oder subjektiv angegeben sein.

Dieses Wörterbuch basiert zum größten Teil auf englischsprachigen Nachrichtentexten. Hierbei können nicht immer Aussagen über die Ausrichtung der Meinung getroffen werden, sondern in wenigen Fällen kann nur zwischen Subjektivität und Objektivität unterschieden werden. Abbildung 3.4 zeigt hier die genaue Aufteilung der Meinungswörter.

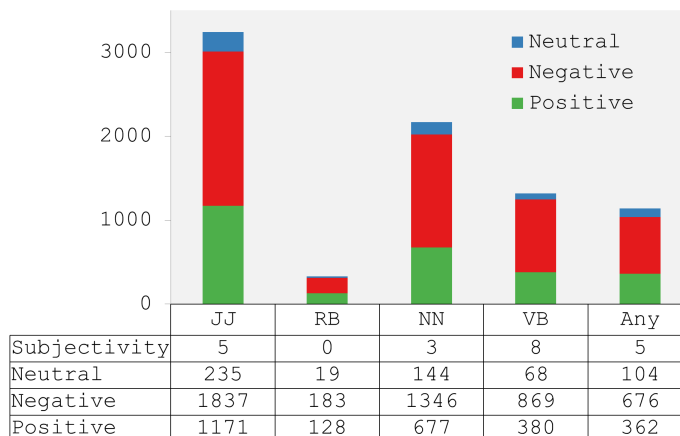


Abbildung 3.4: Das Wörterbuch „Subjectivity Clues“ unterteilt ihre berechneten Meinungswörter nach positiv, negativ, neutral und subjektiv und ist abhängig von der Wortart. Hierbei beschreiben die einzelnen Spalten jeweils eine Wortartklasse (*JJ*=Adjektiv, *RB*=Adverb, *NN*=Nomen, *VB*=Verb, *Any*=nicht spezifiziert).

Financial Dictionary Dieses Wörterbuch beinhaltet 353 positive und 2337 negative Meinungswörter, die von den Autoren Loughran und McDonald manuell erstellt wurden [40]. Hierzu untersuchten sie einen großen Korpus anhand von Worthäufigkeiten und übernahmen diejenigen Wörter, die für ihr Empfinden erheblichen Einfluss in der Finanznachrichtendomäne haben könnten.

Dependencies Zur Bestimmung von Meinungswörtern werden innerhalb dieser Methode die *Abhängigkeitseigenschaften* verwendet. Hierzu soll der Ansatz von Qiu et al. [22], welcher bereits in

Kapitel 2.5.1 erörtert wurde, angewendet werden. Zur Identifikation wird nur ein Teil des Regelkatalogs verwendet, da nur Abhängigkeiten zwischen Meinungswörtern betrachtet werden. Zusätzlich beschränkt sich diese Technik auf das Markieren von Meinungskandidaten als subjektiv, wodurch folglich keine Ausrichtungen berechnet werden.

Zusammenfassung Die Wörterbücher stammen aus unterschiedlichen Dokumentensammlungen und umfassen verschiedene Sparten. Zusätzlich existieren zwei domänenunabhängige Techniken. Die Menge der Meinungswörter innerhalb der Wörterbücher ist ausreichend. Es ist nicht gegeben, dass ein auf den Finanzsektor ausgerichtetes Meinungswörterbuch nur auf diesem Sektor zu guten Ergebnissen führt. Die Evaluation wird folglich dahingegen interessant, wie sich die einzelnen Verfahren ergänzen können und welche Vorteile und Nachteile sie in bestimmten Sparten vorweisen.

3.2.5.2 Merkmale

Produktmerkmalskandidaten werden ähnlich wie im Ansatz von Kisilevich et al. [6] durch bestimmte Abfolgen von Nomen und Verbindungswörter definiert. Diese Auswahl wird durch Zahlen und Possessivpronomina erweitert, da Produktnamen für elektronische Artikel häufig auch Zahlen beinhalten und Bestandteile von Produkten über Possessivpronomina ausgedrückt werden können. Diese Merkmale sind in Abbildung 3.5 in einem endlichen Automaten abgebildet.

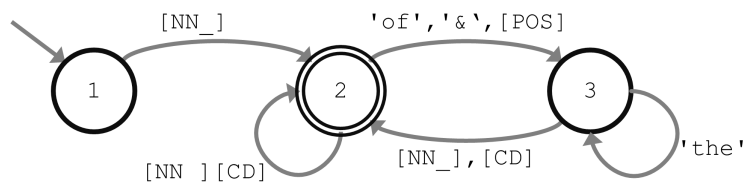


Abbildung 3.5: Der zugrundeliegende Automat kann sowohl Produktmerkmale, wie auch Produktbestandteile identifizieren.

Hierdurch können sowohl einzelne Wörter, Zusammenschlüsse von Wörtern wie auch Produktbestandteile identifiziert werden, wodurch ein hoher Abdeckungsgrad erreicht werden kann. Zusätzlich beschreibt dieses Verfahren eine domänenübergreifende Lösung, da die Merkmale mit Hilfe der Wortartinformationen definiert werden. Abbildung 3.6 gibt hierzu nähere Beispiele anhand von realen Textdaten.

Produktmerkmale: compact-flash card, manual mode, quality, ease of use, picture quality, battery life, point-and-shoot camera, nikon coolpix 5700, custom white-balance setting, nikon coolpix 5700, canon g3, picture quality, durability, camera lcd screen, view mode, compactflash card, selection of shutter speed, uv lens filter

Produktbestandteile: focus of the viewfinder, canon 's battery, adobe 's photoshop elements 2.0 software, lens of the g3, megapixels of the s330, steel heft of the s330, settings of the zoom

Abbildung 3.6

Um die Kenntnisse von Ding, Liu und Yu aus [22] in die Identifikation einbeziehen zu können, können zusätzlich die *Abhängigkeitseigenschaften* betrachtet werden. Dadurch, dass dieses Verfahren ebenfalls domänenübergreifend arbeitet, können unbekannte Produktmerkmale anhand der grammatikalischen Beziehungen zu einem Meinungswort identifiziert werden.

Zusammenfassung Die Identifikation von Produktmerkmalen ist nur auf Nomenketten beschränkt. Dies bedeutet, dass Merkmale, welche durch Verben impliziert werden, nicht gefunden werden können. Die Methode beschreibt folglich nur explizite Produktmerkmale, die als Nomenketten innerhalb der Texte auftauchen. Dies entspricht den Methoden von Ding, Liu und Yu [22] und Kisilevich et al. [6]. Da dieses Verfahren nur auf Wortartinformationen aufbaut, können somit explizite Produktmerkmale domänenübergreifend identifiziert werden. Hierbei wird nicht zwischen Produktmerkmalen, Produkteigenschaften und dem Produkt selbst unterschieden, wodurch der Ansatz von Popescu et al. [31] nicht weiter betrachtet werden soll.

3.2.5.3 Referenzen

Für die Identifikation von merkmalsbezogene Meinungen soll auf das Grundschema von Hu und Liu [27] zurückgegriffen werden. Dieses besagt, dass sich eine Meinung in der Nähe des zu beschreibenden Objektes befindet. Folglich werden die zuvor identifizierten Meinungskandidaten und Produktmerkmalskandidaten über ihre Entfernung in Beziehung gesetzt. Die Berechnungen der Referenzen wird über die in Kapitel 2.5.3 vorgestellten Techniken erreicht, welche nun im weiteren erläutert werden.

Nearest Feature Jeder Meinungskandidat wird mit dem nächstliegenden Produktmerkmalskandidaten verknüpft, welcher sich innerhalb eines Fensters der Größe fünf befindet.

Nearest Feature within delimiter Es wird für jeden Meinungskandidaten der nächstliegende Produktmerkmalskandidat bestimmt, welcher sich im gleichen Haupt- oder Nebensatz befindet.

Dieser Ansatz verwendet somit Zusatzinformationen anhand der Satzstruktur, da Benutzer häufig ihre Meinungen in durch Satzzeichen getrennten Aufzählungen gruppieren. Dabei legt sich diese Methode auf keine Fenstergröße fest, sondern sucht nur innerhalb von Satztrennzeichen.

POS Pattern Diese Methode entspricht dem Ansatz von Kisilevich et al. [6], worin anhand von vorab definierten Wortartmustern merkmalsbezogene Meinungen modelliert werden. Hierfür werden diejenigen Muster, welche mehrere meinugstragende Wörter beinhalten, in jeweils eine Referenz umgewandelt, da eine Referenz in dieser Arbeit nur jeweils ein Meinungswort beinhalten darf.

Dependencies Hier werden die grammatikalischen Abhängigkeiten der Wörter betrachtet. Diese Methode entspricht der Erweiterung des Ansatzes von Qiu et al. [22]. Die dort modellierten Regeln zwischen Produktmerkmalen und Meinungen können ebenfalls als Indiz für die Referenzbestimmung betrachtet werden. Für diese Methode werden folglich zwei Regeln aus diesem Ansatz abgeleitet. *Regel 1* definiert eine Referenz zwischen einem Feature und einem Meinungswort und *Regel 2* definiert diese Beziehung über eine indirekte Relation mit einem weiteren Wort. Dabei werden nur Referenzen übernommen, die sich in einem Fenster der Größe 15 befinden.

Zusammenfassung Die Methoden zur Identifikation von merkmalsbezogenen Meinungen sind abhängig von den zuvor identifizierten Kandidaten. Hierbei beschreibt nicht jeder Kandidat eine tatsächliche Benutzermeinung, wodurch fehlerhafte Referenzen identifiziert werden können. Die Ansätze sind dahingegen domänenunabhängig, da sie zum einen von der domänenunabhängigen Kandidatenidentifikation abhängig sind und zum anderen die Techniken selbst keine Spartenspezifität besitzen. Referenzen haben die höchste Abhängigkeit zu den zuvor berechneten Informationen, wodurch eine deutliche Modellierung der Transparenz erschwert wird.

3.3 Hypothesen

Nachdem die Problemstellungen und deren Lösungsansätze aufgezeigt wurden, kann nun auf die einzelnen Hypothesen dieser Arbeit eingegangen werden. Diese ergeben sich zum einen aus der Modulstrukturmodellierung des Analyseprozesses und zum anderen aus dem Ansatz der Interaktiven Steuerung.

Hypothese 1: *„Die Vereinigung von mehreren Methoden zur Lösung derselben Identifikationsaufgabe ermöglicht ein besseres Ergebnis.“*

Der Analyseprozess besteht aus mehreren Modulen, die wiederum mehrere unterschiedliche Methoden enthalten. Jede Methode entspricht einer Lösung eines Identifikationsproblems. Durch die Vereinigung kann der Umfang der Analyseanwendung erweitert werden, wodurch das Ergebnis besser werden kann.

Hypothese 2: *„Die Priorisierung der Modulentscheidungen anhand von Sicherheitsfaktoren ermöglicht effizientere Korrekturen des Analyseprozesses, welche noch zusätzlich durch automatisch generierte Vorschläge beschleunigt werden.“*

Die Modulentscheidungen werden durch einen Sicherheitsfaktor bewertet. Somit kann der Analyst auf etwaige Unsicherheiten im Analyseprozess hingewiesen werden. Es können zum Beispiel die häufigsten Fehler behoben oder die häufigsten sicheren Entscheidungen korrigiert werden. Dies führt folglich schneller zu einer Verbesserung, als wenn die Korrekturen ohne Priorisierung erfolgen würden. Dabei wird der Analyst zusätzlich durch automatisch generierte Vorschläge unterstützt, wodurch mehrere Entscheidungen gleichzeitig bearbeitet werden können.

Hypothese 3: *„Die Interaktive Steuerung ermöglicht eine effektive und messbare Verbesserung des Ergebnisses.“*

Die Korrekturen des Analysten erreichen eine sichtbare Verbesserung des Analyseprozesses, wobei die positiven Effekte durch mathematische Berechnungen nachgewiesen werden können.

Hypothese 4: *„Die durch die Interaktive Steuerung erzielte Prozessoptimierung lässt sich auch auf andere Textdaten der gleichen Domäne übertragen.“*

Werden die Korrekturen des Analysten gespeichert und auf andere Textdaten der gleichen Domäne übertragen, müssen dort nun weniger Korrekturen durchgeführt werden. Somit kann die Prozessoptimierung auch auf anderen Textdaten weitergeführt werden, wodurch der Analyseprozess sich seinem Optimum annähert.

Kapitel 4

Entwicklung

Dieses Kapitel beschäftigt sich mit der Entwicklung einer Anwendung innerhalb des Meinungsanalyseumfeldes. Es wird ein Werkzeug präsentiert, das in einem Analyseprozess merkmalsbezogene Meinungen identifiziert und Benutzerinteraktionen zur Verbesserung bereitstellt. Die Herausforderungen aus Kapitel 3.1 und die in Kapitel 3.2 aufgezeigten Lösungsansätze bilden hierbei die Grundlagen für die Umsetzung des Meinungsanalysewerkzeuges. Abbildung 4.1 zeigt einen schematischen Überblick über die Bestandteile der im Rahmen dieser Arbeit umgesetzten Anwendung.

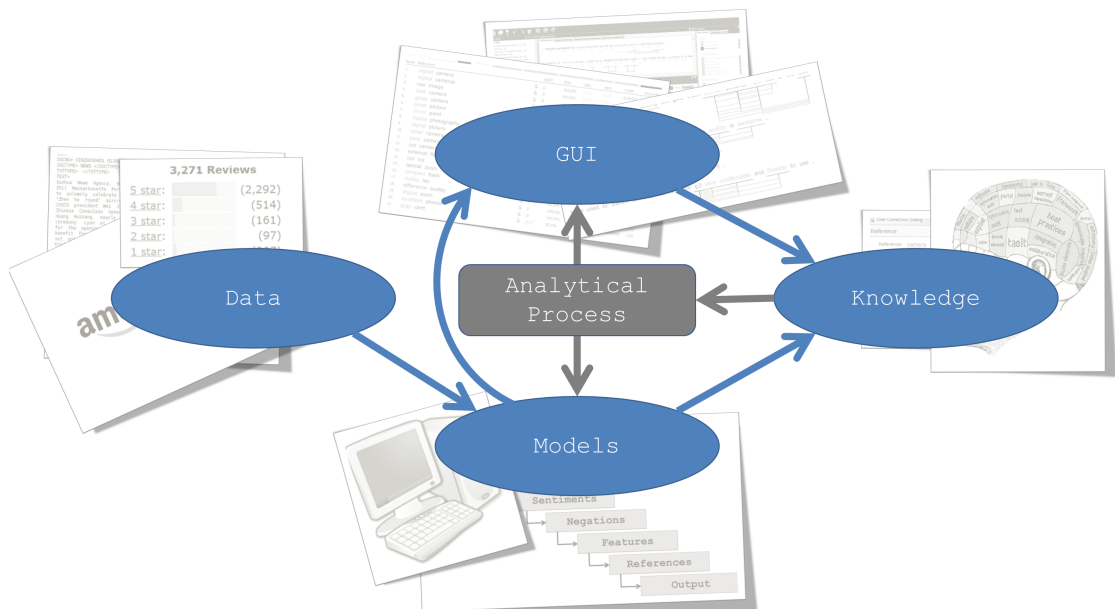


Abbildung 4.1: Der schematische Aufbau der Anwendung ist hierbei in fünf Komponenten unterteilt. Die Anwendung verarbeitet Textdaten aus Kundenrezensionen und speichert berechnete Informationen innerhalb eines Datenmodells. Dieses befindet sich im Wechselspiel mit einem Meinungsanalyseprozess, der seine Berechnungen mittels verschiedener Techniken aus Computerlinguistik und Informatik durchführt und seine Zwischen- und Endergebnisse in einer Oberfläche visuell darstellt. Zusätzlich verfügt die Anwendung über eine Schnittstelle, die das Weltwissen eines Benutzers über die Interaktive Steuerung einbindet und somit dem Menschen die Möglichkeit überlässt, den Analyseprozess zu manipulieren.

Das Projekt wurde innerhalb der Programmierplattform *Eclipse* unter *Java 1.6* entwickelt. Dabei entstanden 221 Klassen, aufgeteilt in 58 Pakete. Im Folgenden werden nun die einzelnen Komponenten aus Abbildung 4.1 herausgegriffen und deren genauer Aufbau und Arbeitsweise erläutert. Zunächst soll der Analyseprozess vorgestellt werden. Dieser beinhaltet die Methoden der Meinungsanalyse und beschreibt den Umgang mit den Textdaten. Hierzu zählen zum Beispiel das Einlesen, die Vorverarbeitung, die Identifikationen von Meinungs- und Produktmerkmalskandidaten, die Berechnung der Referenzen und die Ausgabe der gesammelten Informationen. Daraufhin soll das Datenmodell beleuchtet werden. Dieses speichert das im Verlauf der Analyse berechnete Wissen. Mittels einer Oberfläche wird dem Benutzer die Möglichkeit gegeben, den Analyseprozess zu durchleuchten und Änderungen am Datenmodell sowie dem Analyseprozess selbst vorzunehmen. Welche Interaktionen hierfür durchgeführt werden können, zeigt ein Benutzerszenario am Ende dieses Kapitels.

4.1 Analyseprozess

Der Analyseprozess ist für das Verarbeiten der Textdaten zuständig. Hier werden diejenigen Informationen gesammelt, welche für die Identifikation von merkmalsbezogenen Meinungen benötigt werden. Zunächst soll die Prozessstruktur im Gesamten betrachtet werden, wonach dann genauer auf einzelne Module und deren enthaltenen Methoden eingegangen werden kann.

4.1.1 Analyseprozessstruktur

Die Struktur des Analyseprozesses beschreibt eine Pipeline, welche eine Sammlung mehrerer aufeinander aufbauender Einzelprozesse darstellt. Welche Einzelprozesse für ein Meinungsanalysetool nötig sind, zeigt Abbildung 4.2 in einem groben Überblick.

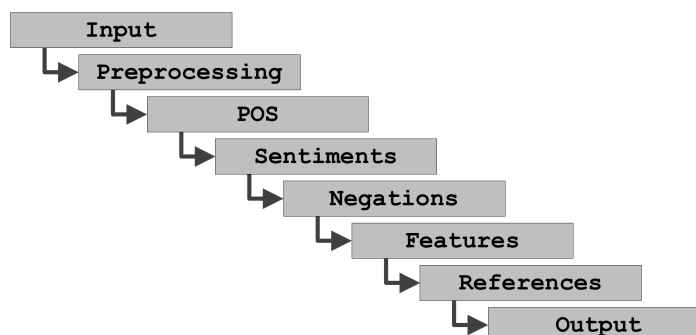


Abbildung 4.2: Um merkmalsbezogene Meinungen identifizieren zu können, müssen Textdaten zunächst eingelesen und vorverarbeitet werden. Nach der Wortarterkennung werden Meinungskandidaten, Negationen und Produktmerkmalskandidaten identifiziert, woraufhin dann die Beziehungen zwischen den Kandidaten berechnet werden. Abschließend können die gesammelten Informationen abgespeichert werden.

Bevor die Analyseanwendung merkmalsbezogene Meinungen ausgeben kann, müssen zuvor Informationen gesammelt werden, die für deren Berechnungen nötig sind. Das folgende Unterkapitel beschreibt hierbei die einzelnen Komponenten des Analyseprozesses umfassend. Zu Beginn werden die Methoden zum Einlesen von Textdaten aus unterschiedlichen Datenformaten vorgestellt (Kapitel 4.1.2.1). In einem Vorverarbeitungsschritt werden diese für die eigentliche Meinungsanalyse normalisiert und aufbereitet (Kapitel 4.1.2.2). Daraufhin können die Identifikationsverfahren für Wortart (Kapitel 4.1.2.3) und Meinungs- sowie Produktmerkmalskandidaten angewendet werden (Kapitel 4.1.2.4 - Kapitel 4.1.2.5). Die daraus resultierenden Informationen fließen anschließend in die Referenzberechnung ein, woraus sich schließlich merkmalsbezogene Meinungen ergeben (Kapitel 4.1.2.7). Diese können dann in verschiedene Ausgabeformate umgewandelt und auf der Festplatte abgelegt werden (Kapitel 4.1.2.8).

4.1.2 Analyseprozesskomponenten

Die in dieser Arbeit verwendeten Methoden finden bereits Einsatz in verschiedenen wissenschaftlichen Publikationen und wurden bereits in Kapitel 2.5 vorgestellt. Ein besonderes Augenmerk ist hierbei auf die Umsetzung der Lösungsansätze aus Kapitel 3.2 gerichtet. Hier sollen im weiteren auch die Berechnungen der Gesamtentscheidungen und Vertrauenswerte der Module erläutert werden, unter anderem auch im Hinblick auf die Forderung nach Transparenz des internen Informationsaustausches.

4.1.2.1 Texteingabe

Die Eingaberoutine stellt die Funktionalität zum Einlesen von beliebigen Textdaten zur Verfügung. Hierfür wurde das Dateiformat einer einfach strukturierten Komma-separierten Tabelle gewählt, welche die verschiedenen Sätze der Benutzer beinhaltet. Zusätzlich können die Sätze, je nach Anwendungsfall, in einzelne Rezensionen gruppiert werden. Nach Ablauf der Texteingabe befinden sich nun die einzelnen Rezensionen mit den dort enthaltenen Sätzen im Speicher und können an die Vorverarbeitung weitergereicht werden.

4.1.2.2 Vorverarbeitung

Die Vorverarbeitung verarbeitet die Sätze im Speicher und reichert sie mit diversen computerlinguistischen Informationen der morphologischen, syntaktischen und semantischen Ebene an. Die Berechnungen werden hierbei über die Algorithmen der *Stanford Core NLP Library*³¹ erreicht, welche sich einfach in die Anwendung integrieren ließ. Dabei gab bereits Kapitel 3.2.4 Aufschluss über

³¹<http://nlp.stanford.edu/software/corenlp.shtml>

die Nutzungsmöglichkeiten dieser Bibliothek. Hierzu zählen zunächst die Segmentierung der Sätze in einzelne Wörter, die Berechnungen deren Grundformen und das Parsen von Abhängigkeitseigenschaften. Die Auflösung von Pronomen ist ebenfalls durch die eingebundene Bibliothek möglich. Anhand von zwei unterschiedlichen Wörterbüchern können Stopwörter markiert werden. Nachdem die computerlinguistischen Informationen extrahiert wurden, liegen nach der Vorverarbeitung nun morphologische und grammatikalische Informationen über einzelne Wörter vor.

4.1.2.3 Wortarten

Zur Bestimmung der Wortarten werden die bereits in Kapitel 2.4 vorgestellten Bibliotheken eingebunden. Dabei bauen die *Stanford Core NLP Library*, die *Maxent Tagger Library* und die *Open NLP Library* allesamt auf dem *Penn Treebank Tag Set* [35] auf, wodurch vergleichbare Wortartentscheidungen entstehen. Diese drei Methoden werden nach Kapitel 3.2 zu einem Modul zusammengeführt. Dieses muss folglich für jeden Term drei Methodenentscheidungen zu einer Gesamtentscheidung verrechnen, welche nun im weiteren betrachtet wird.

In manchen Fällen können sich für Wörter voneinander abweichende Wortarten ergeben. Dies kann zum einen auf die unterschiedlichen Trainingsdaten und zum anderen auf die Algorithmen selbst zurückzuführen sein. Hierbei kann es zu drei unterschiedlichen Fällen kommen.

1. Alle Einzelmethode identifizieren für das jeweilige Wort die gleiche Wortart.
2. Die Wortart zweier Einzelmethode stimmen überein.
3. Alle Wortarten für das jeweilige Wort unterscheiden sich.

Fall 1 beschreibt mit sehr großer Wahrscheinlichkeit die richtige Wortart, wohingegen die anderen Fälle eine Verunsicherung erzeugen. Diese wird größer, je mehr unterschiedliche Wortarten identifiziert werden. Anstelle der Unsicherheit wird der Begriff der Vertrauenswürdigkeit verwendet, welcher sich hierbei gegensätzlich verhält. Ein Vertrauenswert nimmt folglich mit der Anzahl der übereinstimmenden Wortarten zu und ist am größten, wenn sich eine komplette Einigkeit einstellt. Abbildung 4.3 stellt diese Gegebenheit mathematisch dar.

$$(1) \quad \text{conf}_{\text{POS}}(w, t) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1, & t = \text{pos}_i(w) \\ 0 & \text{otherwise} \end{cases}$$

Abbildung 4.3: Die Formel der Vertrauenswertberechnung ähnelt der Berechnung der bedingten Wahrscheinlichkeit. Der Vertrauenswert betrachtet hier, wie wahrscheinlich es ist, einem Wort w die Wortart t zuzuordnen. Der Wert ergibt sich als Faktor zwischen der Anzahl der Methoden, die w mit der Wortart t identifiziert haben und der Anzahl der Methoden, welche sich an der Abstimmung beteiligt haben.

Diese Einzelentscheidungen müssen nun nach Kapitel 3.2 zu einer Gesamtentscheidung verrechnet werden. Diese setzt sich aus allen getroffenen Einzelentscheidungen zusammen. Eine Uneinigkeit

führt folglich dazu, dass einem Term mehrere unterschiedliche Wortarten zugeteilt werden müssen. Somit wird sichergestellt, dass die Wortartinformationen nicht verloren gehen, aber durch einen niedrigeren Vertrauenswert markiert werden. Dies soll nun anhand des Beispiels in Abbildung 4.4 veranschaulicht werden.

	Stanford	Maxent	OpenNLP		POS TAG	Confidence
I	PRP	PRP	PRP	I	PRP	1.0
love	VBP	VBP	VBP	love	VBP	1.0
the	DT	DT	DT	the	DT	1.0
nokia	NN	JJ	JJ	nokia	JJ	0.67
phone	NN	NN	NN		NN	0.33
				phone	NN	1.0

$$\text{conf}_{\text{POS}}(\text{nokia}, \text{JJ}) = \frac{\text{count}(\text{nokia} \cap \text{JJ})}{3} = \frac{2}{3} \cong 0.67$$

Abbildung 4.4: Das Beispiel zeigt für den Satz „I love the nokia phone“ die von den verschiedenen Methoden identifizierten Wortarten. Hier ist ersichtlich, dass das Wort „nokia“ zweimal als Adjektiv und einmal als Nomen deklariert wurde. Die daraus resultierende Gesamtentscheidung beinhaltet somit, dass „nokia“ mit einem Vertrauenswert von 0,67 als Adjektiv zu sehen ist, es aber auch mit einem Vertrauenswert von 0,33 als Nomen gelten kann.

Anhand des Beispiels lässt sich nun auch der Mehrwert der Gesamtentscheidungsberechnung ableiten. Denn „nokia“ als Nomen zu betrachten, wäre für diesen Anwendungsfall korrekt. Dadurch, dass diese Information beibehalten wird, kann davon zum Beispiel die später ansetzende Produktmerkmalserkennung profitieren. Über den Vertrauenswert kann die Entscheidung transparent dargestellt und von späteren Methoden zurückverfolgt werden.

Dem Analyseprozess liegen nun Informationen über die Wortarten der Terme vor. Zusätzlich bestimmt ein Faktor über die Vertrauenswürdigkeit der Wortartwahl.

4.1.2.4 Meinungskandidaten

Das Modul zur Identifikation von Meinungskandidaten umfasst die in Kapitel 3.2.5 vorgestellten Methoden. Somit werden insgesamt fünf Wörterbuchmethoden und das Verfahren der Autoren Qui et al. [22] verwendet. Meinungskandidaten stellen diejenigen Wörter dar, die durch mindestens eine Methoden als Meinungswort identifiziert werden können. Hierbei wird das Wort in den Wörterbüchern gesucht und bei Übereinstimmung die jeweilig resultierende Ausrichtung extrahiert. Kann kein Eintrag gefunden werden, wird dessen Grundform betrachtet. Je nach Methode muss auch die Wortart des jeweiligen Terms mitbetrachtet werden. Die folgende Auflistung gibt diejenigen Informationen an, welche im Datenmodell für einen Meinungskandidaten abgelegt werden können.

- **Ausrichtung:** Die Ausrichtung des Meinungskandidaten kann entweder positiv, negativ oder

neutral ausfallen.

- **Subjektivität:** Sind keine Informationen über die Meinungsrichtungen vorhanden, kann mittels dieser Option der Meinungskandidat als subjektiv markiert werden.
- **Wortart:** Je nach Wortart können sich die Meinungsrichtungen unterscheiden. Somit wird festgehalten, für welchen Wortarttyp der Meinungskandidat steht.
- **Vertrauenswert:** Ein Vertrauenswert gibt an, wie sicher, beziehungsweise unsicher die Entscheidung der Methode über die Meinungsrichtung ist. Der Wert ist hierbei abhängig von der Wortartidentifikation, falls sich die jeweilige Methode auf diese Information stützt. Legt zum Beispiel ein Wörterbuch für das Wort „fair“ fest, dass es als Adjektiv positiv ist, so kann diese Entscheidung nur noch so vertrauenswürdig sein, wie die Entscheidung, dass „fair“ ein Adjektiv beschreibt.

$$(2) \text{ conf}_{\text{SENT},i}(w,t,s) = \begin{cases} 1 & , s = \text{sent}(w,t) \wedge \text{POS independent} \\ \text{conf}_{\text{POS}}(w,t) & , s = \text{sent}(w,t) \wedge \text{POS dependent} \\ 0 & \textit{otherwise} \end{cases}$$

Nachdem die Informationen des Meinungskandidaten aufgezeigt wurden, kann nun auf die Berechnung der Modulentscheidung eingegangen werden. Dies soll zunächst in Abbildung 4.5 beispielhaft veranschaulicht werden.

		GI	BL	FI	SW	SC	DP	RES	AGR
(1)	love	+	+	+	+	+	?	+	6
(2)	damage	-	-	-	-			-	4
(3)	sharp		+		+	-		+	2
(4)	close				+	-	?	?	2
(5)	shoot	-			+			-	1

Abbildung 4.5: Hier sind für Beispielwörter die Entscheidungen der einzelnen Meinungsidentifikationsmethoden abgetragen. Dabei steht + für ein positives und - für ein negatives Meinungswort. Das Zeichen ? steht für Subjektivität. Leere Zellen beschreiben eine neutrale Ausrichtung. Die rechte Tabelle gibt das resultierende Gesamtergebnis (RES) und die Übereinstimmungen der Einzelentscheidungen an (AGR).

Die Bestimmung der resultierenden Meinungsrichtung wird über eine Mehrheitsentscheidung erreicht. Es wird folglich angenommen, dass wenn die Mehrheit der Methoden für eine spezifische Ausrichtung tendiert, diese auch auf den Meinungskandidaten angewendet werden kann. In Abbildung 4.5 trifft dies eindeutig auf die ersten drei Fälle zu. Falls keine Mehrheit vorliegt, so kann wie für Fall 4 lediglich eine Aussage über das Bestehen von Subjektivität getroffen werden. Betrachtet man zusätzlich die Vertrauenswerte der Einzelentscheidungen, so können sich diese ebenfalls auf

die resultierende Ausrichtung auswirken. In *Fall 5* überwiegt die Entscheidung des ersten Wörterbuches, da „shoot“ in Folge eines Wortartidentifikationsfehlers für *SW* weniger vertrauenswürdig ist.

Um nun die getroffene Entscheidung über die Meinungsausrichtung bewerten zu können, soll hieraus ein Vertrauenswert des Moduls abgeleitet werden. Dieser ist abhängig von der resultierenden Meinungsentscheidung und beschreibt die Höhe der Sicherheit einem Meinungskandidaten eine spezifische Ausrichtung zusprechen zu können. Abbildung 4.6 beschreibt diesen Zusammenhang mathematisch.

$$(3) \quad \text{conf}_{\text{SENT}}(w, t) = \frac{1}{n} \sum_{i=1}^n \text{conf}_{\text{SENT},i}(w, t, \text{sent}(w, t))$$

Abbildung 4.6: Der resultierende Vertrauenswert des Moduls wird über das Arithmetische Mittel der Vertrauenswerte der Methodenentscheidungen, welche an der Modulentscheidung beteiligt waren, errechnet.

Dieser Wert beschreibt somit, wie viel Vertrauen durch die Entscheidungen der einzelnen Methoden aufgebaut werden kann. Er ist am höchsten, falls alle Methoden die gleiche Meinungsausrichtung identifizieren und wird geringer, je weniger Überschneidungen es gibt. Hierbei werden die Unsicherheiten der Wortartidentifikation in der Berechnung mitbetrachtet, wodurch schon bestehende Unsicherheiten beibehalten werden und der Vertrauenswert somit transparent nachverfolgt werden kann.

Dem Analyseprozess liegen nun Informationen über Meinungskandidaten vor. Die Identifikationen wurden mittels Vertrauenswerten bewertet und können in anderen Modulen wiederverwendet werden.

4.1.2.5 Produktmerkmalskandidaten

Der Lösungsansatz zur Identifikation von Produktmerkmalskandidaten wurde bereits in Kapitel 3.2.5.2 näher erläutert. Dieser stützt sich auf die Ansätze von Kisilevich et al. [6], indem bestimmte Abfolgen von Nomen und Verbindungswörter in den Texten über einen endlichen Automaten gesucht werden. Ein Mehrwert durch das Verfahren von Ding, Liu und Yu aus [22] kann nicht abgeleitet werden, da die Methode keine zusätzlichen Nomen mehr markiert. Somit enthält dieses Modul lediglich eine einzige Methode, wodurch die Berechnung eines Vertrauenswertes nicht mehr über eine Mehrheitsentscheidung abgeleitet werden kann.

Die identifizierten Produktmerkmalskandidaten müssen folglich anhand ihrer Erscheinung und ihres Aufbaus bewertet werden, um aus diesen lokalen Eigenschaften Rückschlüsse auf die Güte der Kandidaten ziehen zu können. So können zum Beispiel etwaige Unsicherheiten bei der Wortarterkennung die Kandidatenauswahl beeinflussen. Aber auch die Menge von enthaltenen Stopwörtern

oder Sonderzeichen können hierbei Aufschluss geben. Zusätzlich wird die externe Informationsquelle *Wikipedia*, ähnlich zum Ansatz von Kazama et al. [25], zur Validierung verwendet. Die nachfolgende Liste geht hierbei genauer auf die einzelnen Untersuchungen ein und beschreibt die Teilberechnungen, welche später in einen Vertrauenswert eingehen sollen.

- **Wortart:** Da die Identifikation der Produktmerkmale von den vorab berechneten Wortarten abhängen, fließt dieser Vertrauenswert in die Betrachtung mit ein. Denn ein bereits als unsicher markiertes Nomen beschreibt ebenso ein unsicheres Produktmerkmal. Zusätzlich werden hierdurch auftretende Stopwörter und Zahlen bestraft, da sie jeweils unsichere Merkmale ausdrücken können.

$$(4) \quad \text{conf}_{\text{FEATURE},1}(w_1 \dots w_n) = \frac{1}{n} \sum_{i=1}^n \begin{cases} \text{conf}_{\text{POS}}(w_i, t) & , t \in [\text{NN}, \text{NP}, \text{NNS}, \text{NNP}] \\ 0 & \text{otherwise} \end{cases}$$

- **Zeichen:** Ein Produktmerkmal besteht meistens aus alpha-numerischen Zeichen. Oftmals ist eine falsche Taggebung daran beteiligt, dass eine lange Kette von Sonderzeichen, wie „*****“, oder „—“ als Nomen betrachtet wird. Um diese Fälle zu berücksichtigen, wird ein Prozentsatz der alpha-numerischen Zeichen berechnet.

$$(5) \quad \text{conf}_{\text{FEATURE},2}(w_1 \dots w_n) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\text{len}(w_i)} \sum_{j=1}^{\text{len}(w_i)} \begin{cases} 1 & , \text{char}(w_{ij}) \in \{\text{[A - Za - z0 - 9]}\} \\ 0 & \text{otherwise} \end{cases} \right]$$

- **Wikipedia:** Die Wikipedia-Datenbank besteht aus 5.913.728 Einträgen und beinhaltet somit zahlreiche Eigen- und Produktnamen. Die identifizierten Produktmerkmalskandidaten werden nun in beliebige Teilwortketten unterteilt und in der Datenbank gesucht, um die im Kandidaten enthaltene Eigen- oder Produktnamen zu markieren. Für „*canon 's battery*“ können zum Beispiel „*canon*“ und „*battery*“ als Artikel gefunden werden. Dies gilt aber nicht für „*canon 's*“, „*'s battery*“ und „*canon's battery*“. Danach kann ein Prozentsatz der in Wikipedia als Artikel enthaltenen Teilwörter berechnet werden.

$$(6) \quad \text{conf}_{\text{FEATURE},3}(w_1 \dots w_n) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & , w_i \text{ marked as wikipedia substring} \\ 0 & \text{otherwise} \end{cases}$$

Diese unterschiedlichen Teilwerte sollen nun zusammen die lokale Eigenschaft des Produktmerkmalskandidaten beschreiben. Hierfür werden diese mittels der Formel in Abbildung 4.7 verrechnet. Nun sollen aber nicht nur die lokalen Eigenschaften der Produktmerkmalskandidaten betrachtet werden, sondern auch die Auftretungshäufigkeiten im Text, da diese nach [27] ein häufiges Indiz für gute Merkmale beschreiben. Der lokale Vertrauenswert wird folglich mit der Termfrequenz gewichtet, um hieraus einen globalen Vertrauenswert ableiten zu können. Dieser Zusammenhang

$$(7) \quad \text{conf}_{\text{LOCAL_FEATURE}}(w_1 \dots w_n) = \prod_{i=1}^3 (c_i \cdot \text{conf}_{\text{FEATURE},i}(w_1 \dots w_n))$$

Abbildung 4.7: Der lokale Vertrauenswert für einen Produktmerkmalskandidaten ergibt sich aus der Multiplikation der einzelnen Teilwerte, welche über spezifische Faktoren ($c_i \dots c_n$) unterschiedlich gewichtet werden können.

wird in Abbildung 4.8 veranschaulicht. Ein hoher globaler Vertrauenswert entspricht einem häufigen Produktmerkmalskandidaten, der anhand von lokalen Gesichtspunkten wohl geformt ist.

$$(8) \quad \text{conf}_{\text{GLOBAL_FEATURE}}(f) = \text{conf}_{\text{LOCAL_FEATURE}}(f) \cdot \text{tf}(f)$$

Abbildung 4.8: Der bisherige lokale Vertrauenswert wird mit der normierten Termfrequenz multipliziert. Dadurch entsteht eine Gewichtung des Produktmerkmalskandidaten anhand seiner Auftretungshäufigkeit.

Werden die Produktmerkmalskandidaten nach ihrem globalen Vertrauenswert absteigend sortiert, so ergibt sich die Rangfolge der Kandidaten. Dem Analyseprozess liegen nun Informationen über Produktmerkmalskandidaten vor, welche anhand ihrer lokalen und globalen Eigenschaften bewertet wurden.

4.1.2.6 Negationen

Die Identifikation von Verneinungen betrachten den in Kapitel 2.5.4 vorgestellten Grundsatz, dass Negationen die Ausrichtung des nächstgelegenen Meinungswortes invertieren [27]. Die Identifikation erfolgt über eine einfache Heuristik. Hierzu wird in den Texten nach Negationswörtern gesucht und auf das nächstgelegene Meinungswort innerhalb eines Fensters mit bestimmter Größe bezogen. Dies entspricht dem Ansatz von Hu und Liu [27]. Negationswörter sind innerhalb eines Wörterbuches definiert. Dieses umfasst 36 Einträge, wie „not“, „can't“, „don't“, „haven't“, „without“ oder „never“ und berücksichtigt Rechtschreibfehler.

Zusätzlich wird jedes Auftreten der Abhängigkeitseigenschaft „neg“ untersucht, welche grammatikalische Negationen definieren. Hierdurch können auch neutral eingestufte Wörter identifiziert werden, wodurch nach dem Ansatz von Ding, Liu und Yu [12] dieses folglich eine negative Orientierung zugesprochen bekommt.

Die Methoden zur Identifikation von Verneinungen stellen keinen ganzheitlichen Ansatz dar. Es sollen lediglich die eindeutigsten Auftretungsfälle modelliert werden. Die gesammelten Informationen gehen in den späteren Schritt der Referenzbestimmung ein.

4.1.2.7 Referenzen

Zwischen den vorab identifizierten Meinungskandidaten und Produktmerkmalskandidaten können nun Referenzen berechnet werden, welche die eigentliche Meinungsbeziehungen darstellen. Hierfür

stehen der Anwendung vier unterschiedliche Methoden zur Verfügung, welche bereits in Kapitel 3.2.5.3 näher erläutert wurden. Diese Methoden identifizieren auf unterschiedliche Weisen die Beziehungen zwischen Meinungskandidaten und Produktmerkmalskandidaten. Die resultierende Referenz besteht demnach aus folgenden Informationen:

- **Meinungskandidat:** Der Meinungskandidat gibt die zu betrachtende Meinung des Rezensenten an. So beinhaltet er die Ausrichtung der Meinung und einen Vertrauenswert.
- **Produktmerkmalskandidat:** Der Produktmerkmalskandidat enthält diejenigen Wörter, die für das Produkt oder das Produktmerkmal stehen. Er besitzt ebenfalls ein Vertrauenswert.
- **Typ:** Um die Beziehungen der einzelnen Methoden unterscheiden zu können, erhält die Referenz einen bestimmten Typ.
- **Abstand:** Um Aussagen über die Qualität der Referenz treffen zu können, wird der Abstand zwischen Meinungskandidat und Produktmerkmalskandidat nach dem Grundsatz von Hu und Liu in [27] als Zusatzinformation aufgenommen.
- **Negation:** Die Information, dass ein Meinungswort im Satz negiert vorliegen kann, wird ebenfalls in der Referenz gespeichert. Hierdurch wird die Meinungsrichtung der Referenz negiert.
- **Kontextabhängigkeit:** Hier fließen Informationen über die Kontextabhängigkeit des Produktmerkmalskandidaten ein. Dadurch wird gewährleistet, dass die Ausrichtung des Meinungswortes bestehen bleiben kann, die resultierende Ausrichtung der Referenz aber zum Beispiel in Folge eines negativen Kontextes invertiert wird.

Nachdem nun auf unterschiedliche Weisen Referenzen identifiziert wurden, sollen diese anschließend anhand des Grundsatzes aus Kapitel 3.2 bewertet werden. Mittels eines Vertrauenswertes, der sich aus den zugrundeliegenden Informationen ergibt, sollen sich Aufschlüsse über die Güte der Referenzidentifikation ergeben. Eine vertrauenswürdige Referenz enthält hierbei vertrauenswürdige Meinungs- und Produktmerkmalskandidaten, welche einen geringen Abstand zueinander aufweisen. Gewichtet man die Vertrauenswerte entsprechend dem Ansatz von Ding et al. [12], so ergibt sich der in Abbildung 4.9 veranschaulichte Berechnungszusammenhang.

$$(9) \quad \text{conf}_{\text{REF}}(s, f_1 \dots f_n) = \frac{\text{conf}_{\text{SENT}}(s, \text{postagof}(s)) \cdot \text{conf}_{\text{GLOBAL_FEATURE}}(f_1 \dots f_n)}{\text{abs}(\text{distance}(s, f_1 \dots f_n))}$$

Abbildung 4.9: Der Vertrauenswert der Referenz besteht aus dem Produkt zwischen dem Vertrauenswert der Meinung und dem Vertrauenswert des Produktmerkmals und wird nach deren Abstand gewichtet. s entspricht dem Meinungswort und $f_1 \dots f_n$ beschreibt die Wörter des Produktmerkmalskandidaten.

Ähnlich wie bei der Bewertung der Produktmerkmalskandidaten, beschreibt dieser Wert die lokale Eigenschaft der Referenz. Betrachtet man die Auftretungshäufigkeiten der Referenzen, so können hierbei ebenfalls Rückschlüsse auf vertrauenswürdige Meinungen mit Merkmalsbezug getroffen werden. Denn taucht diese öfters an unterschiedlichen Stellen im Korpus auf, so kann dies für eine höhere Relevanz der Rezensenten sprechen. Auf technischer Seite ist eine Referenz vertrauenswürdiger, wenn sie nach dem Grundsatz aus Kapitel 3.2 gleichzeitig von unterschiedlichen Methoden identifiziert wurde. Somit sollen nun auch die Korpuseigenschaften in den Vertrauenswert der Referenzen einfließen.

Hierzu müssen zwei unterschiedliche Auftretungshäufigkeiten betrachtet werden. Zum einen wird gezählt, wie oft die Referenz an unterschiedlichen Stellen im Korpus auftaucht und zum anderen, wie viele verschiedene Methoden diese identifiziert haben. Teilt man diese Häufigkeiten durch jeweils deren Maxima, so ergeben sich normierte Werte, wodurch relevante Referenzen höher gewichtet werden. Abbildung 4.10 stellt die Berechnung des globalen Vertrauenswertes mathematisch dar.

$$(10) \quad \text{conf}_{\text{GLOBAL_REF}}(s, f_1 \dots f_n) = \text{conf}_{\text{REF}}(s, f_1 \dots f_n) \cdot \frac{\text{freq}_{\text{sentence}}(s, f_1 \dots f_n)}{\max(\text{freq}_{\text{sentence}})} \cdot \frac{\text{freq}(s, f_1 \dots f_n)}{\max(\text{freq})}$$

Abbildung 4.10: Der bisherige lokale Vertrauenswert der Referenz wird mit dem normierten Wert dessen Auftretungshäufigkeit und mit dem normierten Häufigkeitswert dessen Identifikation multipliziert. Dadurch wird die Referenz anhand ihres globalen Kontextes gewichtet.

Nach Identifikation der Referenzen ist der automatische Analyseprozess abgeschlossen und es liegen nun Informationen über merkmalsbezogene Meinungen vor, wobei diese anhand von globalen Vertrauenswerten bewertet sind.

4.1.2.8 Ausgabe

Die Ausgabekomponente speichert die berechneten Informationen auf der Festplatte, um diese zum Beispiel in einer Evaluation oder als Eingabe für andere Anwendungen zu verwenden. Der Analyseprozess kann somit festgehalten und zu einem späteren Zeitpunkt erneut aufgerufen werden.

4.2 Datenmodell

Nachdem der Analyseprozess und seine Bestandteile beleuchtet wurden, kann nun auf das Datenmodell eingegangen werden. Dieses enthält diejenigen Objekte, welche im Laufe des Analyseprozesses mit Informationen angereichert werden. Zentrales Element beschreibt hierbei die Klasse *DataContainer*, worin die Terme (*TokenFactory*) und die Dokumente (*ReviewList*) verwaltet werden. Abbildung 4.11 zeigt die Klassenbeziehungen auf Dokumentenseite, während Abbildung 4.12 die Klassenbeziehungen auf Seite der Terme veranschaulicht.

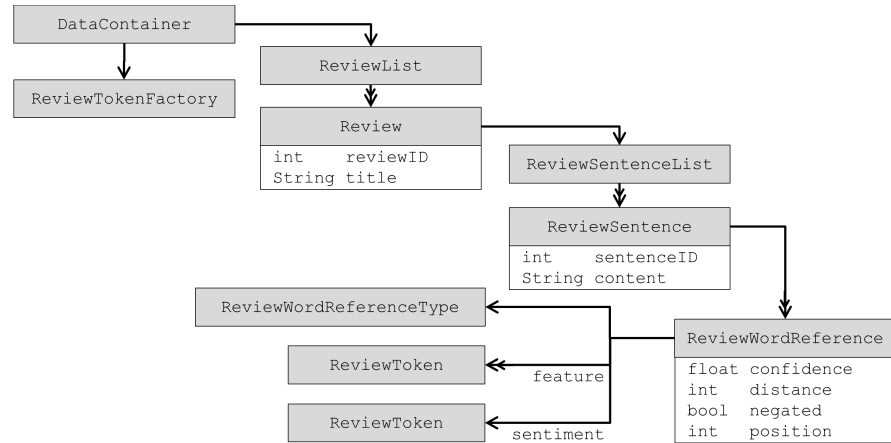


Abbildung 4.11: Das Klassendiagramm zeigt die Beziehungen innerhalb des Datenmodells und die wichtigsten Informationen, welche in den einzelnen Klassen abgelegt werden. Hierbei beschreibt ein doppelter Pfeil eine Assoziation über eine Liste, während ein normaler Pfeil eine einfache Assoziation definiert.

Der Dokumentenkörper besteht aus einer Sammlung von Kundenrezensionen (*Review*), welche ein oder mehrere Sätze (*ReviewSentence*) beinhalten können. Auf Satzebene werden die Informationen der identifizierten Referenzen (*ReviewWordReference*) hinterlegt. Die in den Referenzen enthaltenen Kandidaten verweisen hierbei auf einzelne Terminstanzen (*ReviewToken*) innerhalb der *TokenFactory*-Klasse.

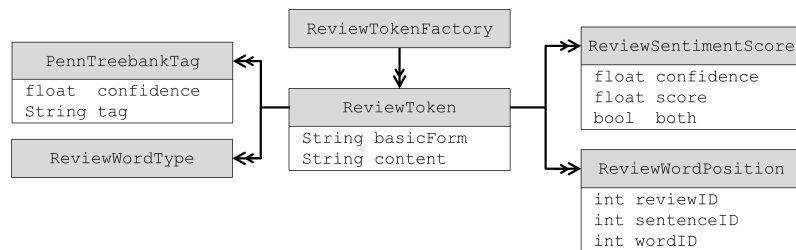


Abbildung 4.12: Das Klassendiagramm zeigt die Beziehungen auf Seiten der Terme und die wichtigsten Informationen, welche in den einzelnen Klassen abgelegt werden. Hierbei beschreibt ein doppelter Pfeil eine Assoziation über eine Liste, während ein normaler Pfeil eine einfache Assoziation definiert.

Die Terme (*ReviewToken*) werden in einer dynamischen Speicherstruktur (*TokenFactory*) abgelegt. Diese ist dafür zuständig, gleiche Terme nur einmalig als Instanz im Speicher bereitzuhalten. Eine Instanz eines Terms ist zum Beispiel abhängig vom Terminhalt selbst, den identifizierten Wortarten (*PennTreebankTag*) sowie den identifizierten Meinungsrichtungen (*ReviewSentimentScore*). Die Struktur ist dahingegen dynamisch, da sich diese Informationen während des Analyseprozesses verändern können. Ist dies der Fall, so teilen sich die Terme in mehrere Instanzen auf. Somit befindet sich jederzeit nur eine Mindestanzahl an Objekten im Speicher, wodurch Ressourcen eingespart und infolge dessen sehr große Datenmengen verarbeitet werden können. Betrachtet man zum Beispiel einen Korpus mit 11524 Wörtern, so ergibt dies zunächst eine Anzahl von 1902 verschiedenen Termen. Nach Identifikation der Wortarten erhöht sich die Menge aufgrund von unterschiedli-

von Benutzerentscheidungen (*ReviewUserSentimentUpdater*) an, welche aus spezifischen Einträgen (*UsersSentimentDictionaryEntry*) innerhalb der Klasse *UsersSentimentDictionary* bestehen. Hier wird zum Beispiel gespeichert, zu welchem Zeitpunkt ein Benutzer eine gewisse Änderung vollzogen hat.

Um nun den Methodenumfang des Meinungsmoduls zu erweitern, wird lediglich eine neue Klasse angelegt, welche innerhalb der Verwaltungsebene (*ReviewSentimentDictionaryList*) registriert werden muss. Abbildung 4.14 demonstriert hierbei deren schematische Grundstruktur.

```
public class ReviewNewSentimentDictionary
    extends ReviewSentimentDictionary
    implements Serializable
{
    public ReviewNewSentimentDictionary( ... ); //initialization details of the method
    public void openOptionDialog();           //override
    public void optionDialogClosed( ... );    //override
    public ReviewList update();               //override
}
```

Abbildung 4.14: Schematische Grundstruktur für das Anlegen einer neuen Identifikationsmethode.

Die neue Identifikationsmethode muss hierbei die Funktionen des Dialogs (*openOptionDialog*, *optionDialogClosed*) sowie die Abarbeitung des Analyseprozesses (*update*) unterstützen. Diese schematische Grundstruktur lässt es zu, dass der Analyseprozess beliebig erweitert und ergänzt werden kann.

4.3 Oberfläche

Die graphische Oberfläche der Anwendung bietet dem Analysten die Möglichkeit den Analyseprozess zu steuern. Hierfür kann er Änderungen am Datenmodell sowie dem Analyseprozess vornehmen. Dieses Unterkapitel beschreibt zunächst die einzelnen Komponenten auf darstellerischer Seite, woraufhin folgend dann auf die Interaktionskonzepte und Dialoge eingegangen werden kann. Daraufhin wird das Zusammenspiel zwischen einzelnen Klassen erläutert. Zuletzt wird anhand eines Benutzerszenarios aufgezeigt, wie genau der Analyseprozess gesteuert werden kann.

4.3.1 Komponenten

Die Oberfläche, welche in Abbildung 4.15 dargestellt ist, besteht im Gesamten aus fünf verschiedenen Bereichen. Bereiche (2), (3) und (4) stellen die berechneten Informationen visuell dar, wohingegen Bereiche (1) und (5) als Interaktionsflächen genutzt werden. Die Bereiche (3), (4) und (5) sind in verschiedene Tabulatoren unterteilt, welche dem Benutzer weitere Zusatzinformationen anbieten.

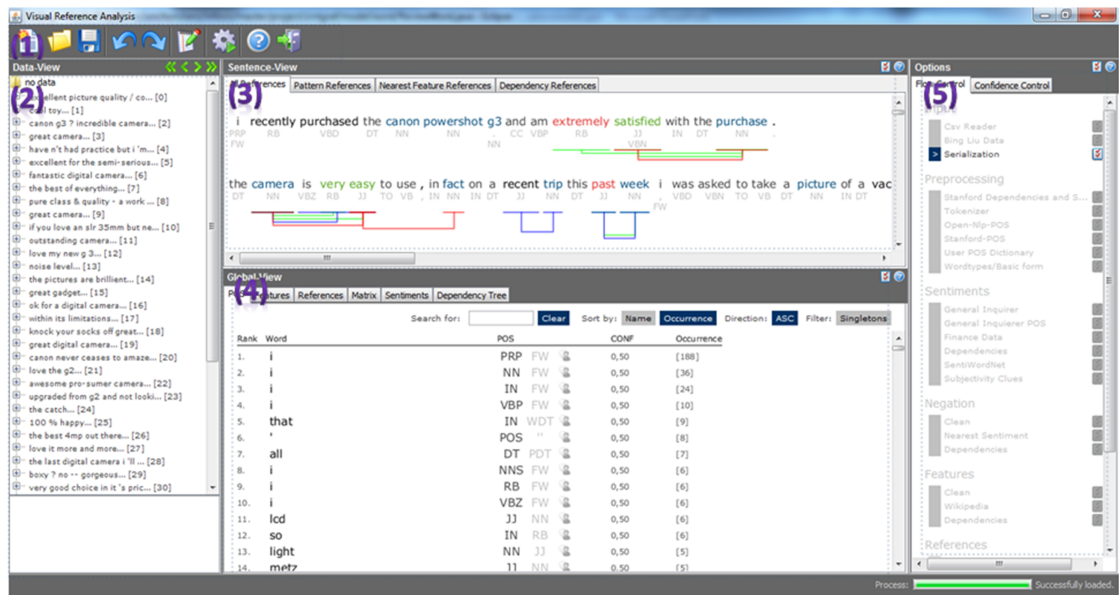


Abbildung 4.15: Die Gesamtübersicht zeigt 5 Hauptteile. (1) Projekt und Menüleiste, (2) Dokumentenübersicht (3) Satzansicht, (4) Globale Ansicht für Wortarten, Meinungskandidaten, Produktmerkmalskandidaten und Referenzen, (5) Analyseprozess und Ansicht der Vertrauenswerte.



Abbildung 4.16: Bereich (1) enthält Funktionen zum Erstellen eines neuen Projekts, Öffnen eines bestehenden Projekts und Speichern des aktuellen Projekts. Zusätzlich können Projekteigenschaften, wie Benutzeränderungen oder Projekttitle eingesehen oder verändert werden. An den letzten Positionen befindet sich die Bedienungselemente zum Starten des Analyseprozesses, zum Aufruf der Hilfe oder dem Beenden der Anwendung.

Die Dokumentenübersicht (2) stellt die einzelnen Kundenrezensionen und die darin enthaltenen Sätze dar, welche zur schnellen Navigation verwendet werden können.

Die Satzansicht (3) in Abbildung 4.17 ermöglicht es, satzbezogene Informationen der Wörter zu veranschaulichen. Hierzu werden die identifizierten Wortarten unterhalb der Terme platziert, wobei sich bei Uneinigkeit bis zu drei verschiedene Wortarten ergeben können. Zusätzlich können Meinungen, Merkmale, Negationen und Stopwörter anhand ihrer Farben unterschieden werden (siehe Abbildung 4.18). Die identifizierten Referenzen, welche sich zwischen den Wörter ergeben, sind als Diagramm dargestellt. Hier hat der Benutzer die Möglichkeit, die farblichen Abhebungen zu verändern oder einzelne Referenztypen zu filtern.

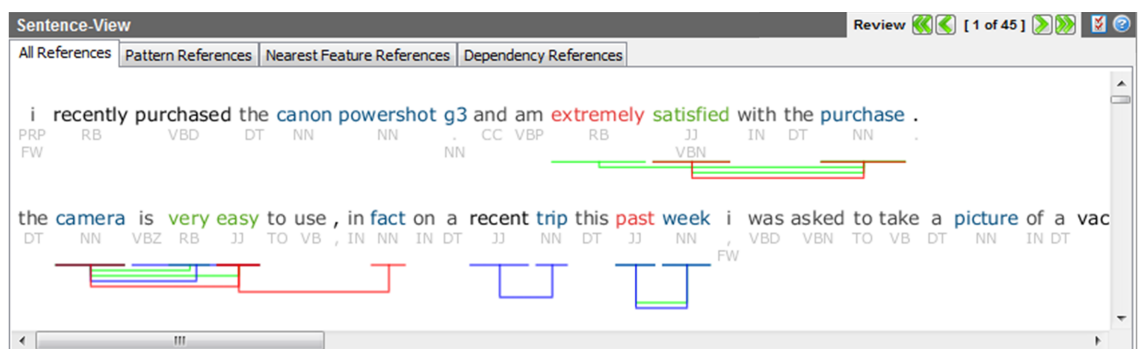


Abbildung 4.17: Die Satzansicht (3) stellt zum einen die verschiedenen Funktionen der Wörter dar, sowie Informationen zu vergebenen Wortarten und identifizierten Referenzen, jeweils in farblicher Kodierung. Diese Ansicht kann vom Benutzer verändert werden, um zum Beispiel Wortartinformationen auszublenden oder Referenzen in einheitlicher Farbe darzustellen. Mittels Tabulatoren lassen sich die unterschiedlichen Referenzmethoden einzeln betrachten. Zusätzlich kann durch die verschiedenen Rezensionen navigiert werden.

	color		rgb
Feature	blue		(0, 75, 124)
Sentiment (positive)	green		(85, 165, 28)
Sentiment (negative)	red		(229, 52, 51)
Sentiment (subjective)	orange		(244, 158, 0)
Negation	purple		(156, 10, 125)
Stopword	grey		(128, 128, 128)
Default	black		(0, 0, 0)

Abbildung 4.18: Die Satzansicht trennt die verschiedenen Informationen der Wörter durch Anwendung unterschiedlicher Farbwerte. Diese werden auch in anderen Darstellungen verwendet, um eine Einheitlichkeit gewährleisten zu können.

Die globale Ansicht (4), wie in Abbildung 4.19 zu sehen, stellt die gesammelten Informationen der Wortarten, Meinungskandidaten, Produktmerkmalskandidaten und Referenzen getrennt voneinander in Listenform dar. Der Benutzer kann die jeweilige Listen umsortieren oder nach spezifischen Termen durchsuchen. Zusätzlich kann er die Terme auch als Filter für die Satzansicht verwenden,

um zum Beispiel die einzelnen Sätze, worin sich die gefilterten Terme befinden, betrachten zu können.

Rank	Word	POS	CONF	Occurrence
81.	the	DT	1,00	[665]
82.	.	.	1,00	[546]
83.	,	,	1,00	[433]
84.	and	CC	1,00	[314]
85.	a	DT	1,00	[262]
86.	to	TO	1,00	[247]
87.	is	VBZ	1,00	[199]
88.	camera	NN	1,00	[180]
89.	of	IN	1,00	[176]
90.	it	PRP	1,00	[173]
91.	you	PRP	1,00	[152]
92.	this	DT	1,00	[133]
93.	for	IN	1,00	[119]
94.	with	IN	1,00	[110]

Abbildung 4.19: Die aktuelle globale Ansicht beschreibt die Darstellung der identifizierten Wortarten. Die einzelnen Listenansichten können über die Tabulatorenumgebung gewechselt werden. Der Benutzer kann die Sortierung verändern oder über Maus-Interaktionen einzelne Terme für die Satzansicht filtern.

In Abbildung 4.20 werden die Analyseprozesskomponenten dargestellt, worin der Benutzer Einstellungen an Einzelmethoden vornehmen oder sie für die Abarbeitung deaktivieren kann. Zusätzlich können die Methodenentscheidungen mittels eines Vertrauenswertes ab- oder aufgewertet werden, wodurch die Identifikationen beeinflusst werden.

Abbildung 4.20: Diese Ansicht besteht aus zwei unterschiedlichen Interaktionsflächen. Auf der linken Seite befindet sich die Übersicht über einzelne Analyseprozesskomponenten. Der Benutzer kann hier einzelne Methoden oder das gesamte Modul deaktivieren oder über eine Schaltfläche einen Dialog aufrufen, worüber sich Einstellungen für die Methoden verändern lassen. Auf der rechten Seite können die Vertrauenswerte der Einzelmethoden modifiziert werden, um zum Beispiel unsichere Methoden abzuwerten.

4.3.2 Schematischer Aufbau

Das Zusammenspiel der Komponenten innerhalb der Oberfläche soll anhand eines Klassendiagramms dargestellt werden. Abbildung 4.21 zeigt hierbei die Beziehungen zwischen den Klassen und deren schematischen Aufbau.

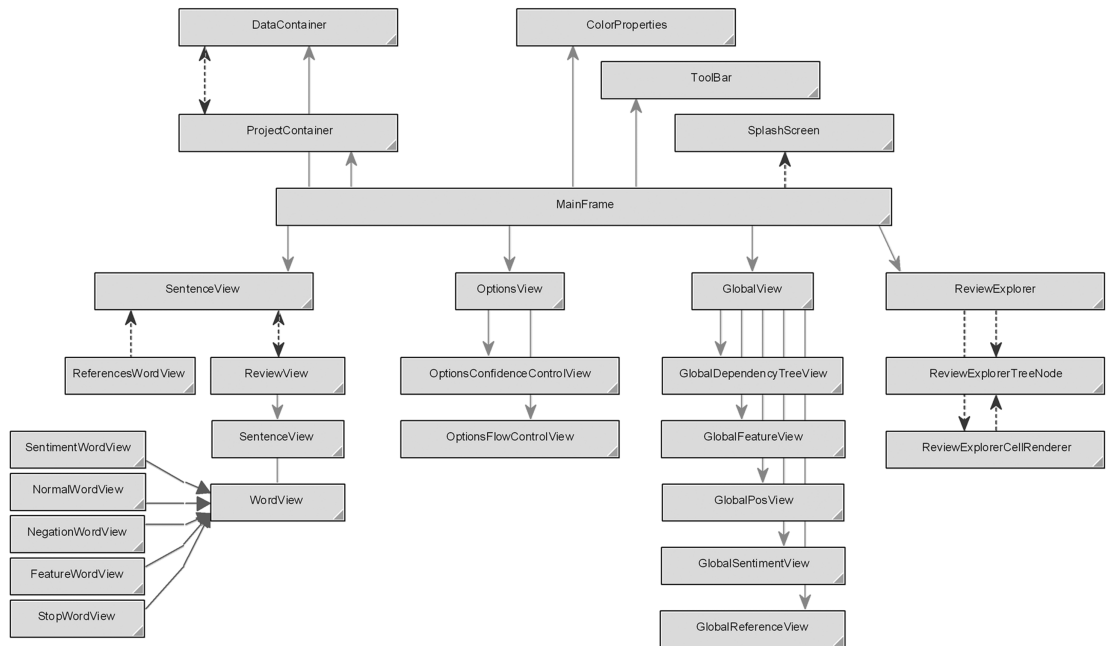


Abbildung 4.21: Ein schematischer Überblick über die verschiedenen Komponenten der Oberfläche.

Hierbei steht die Klasse *MainFrame* im Zentrum des Klassenmodells. Sie verwaltet alle Komponenten, wie die Ansicht der Sätze (*SentenceView*), die Optionen des Analyseprozesses (*OptionsView*), die Darstellung der verschiedenen Identifikatoren (*GlobalView*) und die Übersicht über die Datenkollektion (*ReviewExplorer*). Ebenso verwaltet diese Klasse Aktualisierungen mittels Ereignissen, welche an die einzelnen Komponenten weitergeleitet werden können. Die Anbindung an das Datenmodell ist über die Klasse *DataContainer* modelliert, worauf sich jede Komponente Zugriff verschaffen kann. Hier finden die Manipulationen wie Filterung und Sortierung statt sowie werden dort die Interaktionen des Benutzers verarbeitet.

Zur Oberfläche gehören ebenfalls die Dialoge zur Korrektursteuerung (*UserCorrectionDialog*) und zur Einstellung der Einzelmethode (*ReviewOptionDialog*), welche in Abbildung 4.22 dargestellt sind. Die Erstellung dieser Dialoge ist jeweils durch eine abstrakte Klasse geregelt, worin die Funktionen zum Aufbau der einzelnen Dialogkomponenten implementiert sind. Dadurch kann der Erstellungsaufwand der Dialoge vereinfacht und ein einheitliches Aussehen gewährleistet werden.

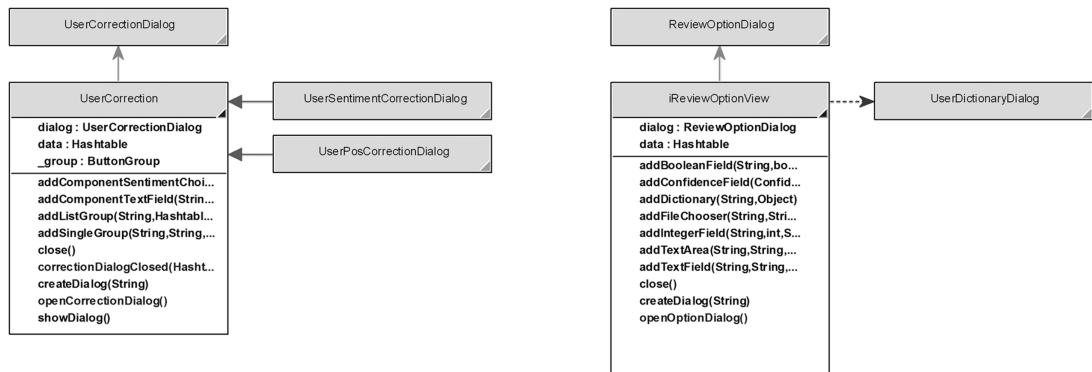


Abbildung 4.22: Die Oberfläche enthält zum einen Dialoge zur Änderung von Identifikationsentscheidungen wie Meinungskandidaten oder Wortarten. Und zum anderen können die Einzelmethode über einen Einstellungssdialog in ihren Eigenschaften wie Schwellenwerte oder Dateipfade verändert werden.

4.3.3 Interaktive Steuerung

Nachdem die Darstellung der Oberfläche aufgezeigt wurde, kann nun auf die interaktiven Komponenten eingegangen werden. Hierzu werden die verschiedenartigen Oberflächenelemente wieder aufgegriffen und beschrieben, wie ein Analyst darin interagieren und den Analyseprozess steuern kann. Hier werden nun die Interaktionstechniken und die verschiedenen Dialogsysteme vorgestellt.

4.3.3.1 Interaktionstechniken

Um den Analysten bei seinem Informationsfindungsprozess unterstützen zu können, sind in der Oberfläche verschiedene Interaktionstechniken implementiert. Hier kann er durch die Daten navigieren, Terme über Suchanfragen auffinden und den Datenraum über Filtertechniken eingrenzen. Zusätzlich hat der Analyst die Möglichkeit, nähere Informationen zu einzelnen Datenpunkten zu erhalten.

Navigation Für die Navigation des Datenraums stehen dem Analysten zwei Bedienelemente zur Verfügung. Die *Dokumentenübersicht* kann dazu verwendet werden, an bestimmte Rezensionen oder Sätze zu springen. Über die Bedienelemente der *Satzansicht* kann zwischen Rezensionen umgeschaltet werden.

Suchanfragen Suchanfragen werden durch reguläre Ausdrücke formuliert und in der *Globalen Ansicht* angewendet. Dabei werden diejenigen Terme, welche nicht zur Suchanfrage passen, aus der Listenansicht ausgeschlossen. Hierdurch kann der Analyst zum Beispiel ähnliche und somit für ihn interessante Terme näher untersuchen.

Filter Um den Datenraum einzugrenzen, können in der *Globalen Ansicht* spezifische Terme ausgewählt werden. Dabei bleiben in der *Satzansicht* nur noch diejenigen Sätze bestehen, welche

mindestens einen der ausgewählten Terme enthalten. Diese Filtertechnik kann aber auch in der *Satzansicht* erfolgen, wobei hier nun die *Globale Ansicht* verändert wird. Dies ermöglicht die Untersuchung auf Teilmengen von Sätzen oder Termen.

Details on Demand Während der Analyse können jederzeit Zusatzinformationen eingesehen werden, wobei diese unterhalb der *Dokumentenübersicht* angezeigt werden. Befindet sich die Maus über einem beliebigen Objekt innerhalb der *Globalen Ansicht* oder der *Satzansicht*, so werden dessen Informationen dargestellt.

4.3.3.2 Dialoge

Die Oberfläche modelliert drei unterschiedliche Dialogarten, welche für verschiedene Zwecke angeboten werden. Einstellungsdialoge nehmen Einstellungen an Methoden vor und Projektdialoge verwalten das Speichern, Öffnen und Bearbeiten von Projekten. Änderungsdialoge ermöglichen die Korrektur von analytischen Fehlern, wodurch der Analyseprozess gesteuert werden kann.

Einstellungsdialoge Die Methoden innerhalb der Analyseprozesskomponenten können unterschiedliche Einstellungen annehmen. Dies können zum Beispiel Parameter zu Dateien oder Schwellenwerte sein, welche die Algorithmen der Einzelmethode beeinflussen. Die Dialoge, wie in Abbildung 4.23, können aus der Darstellung der *Analyseprozesskomponenten* aktiviert werden.

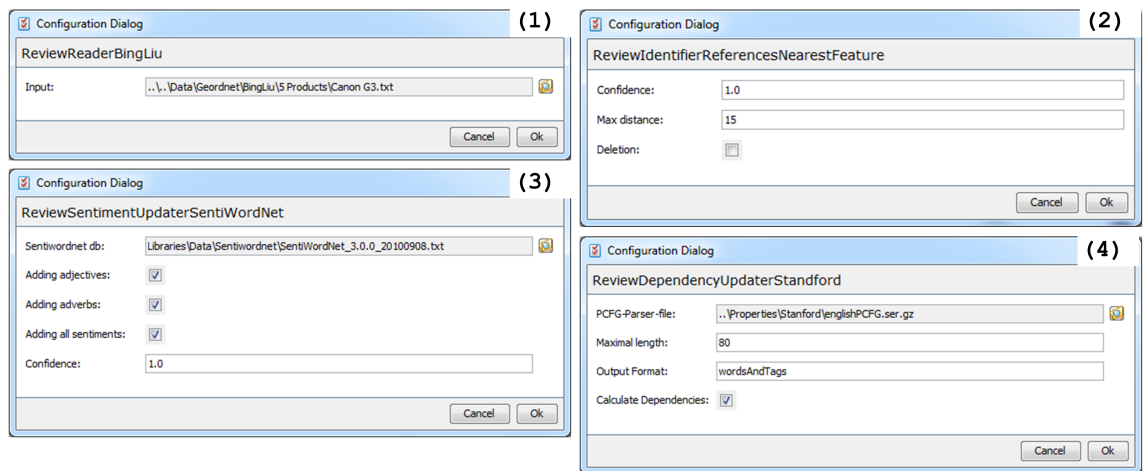


Abbildung 4.23: Dialog (1) enthält ein Parameter zum Öffnen einer Eingabedatei. Dialog (2) ermöglicht Einstellungen zu einer Referenzidentifikationsmethode, in welcher unter anderem eine Maximaldistanz verändert werden kann. Dialog (3) entspricht dem Dialog der Wörterbuchmethode „Sentiwordnet“, worin die Wörterbuchdatei oder Vorabauswahlen zu bestimmten Wortarten definiert werden können. Dialog (4) beinhaltet Einstellung des *Stanford Dependency Parsers*.

Projektdialoge In einer Projektdatei werden der Analyseprozess, die gesammelten Informationen innerhalb des Datenmodells und die Benutzeränderungen gemeinsam abgelegt, um zu einem

späteren Zeitpunkt wiederverwendet werden zu können. Hierzu kann der Benutzer Projekte anlegen, öffnen und speichern. Abbildung 4.24 zeigt die Projektdialoge, in welchen der Benutzer Einstellungen zum Projekt, wie Projektdatei oder Projektbeschreibungen vornehmen und die vom Benutzer getätigten Änderungen einsehen kann.

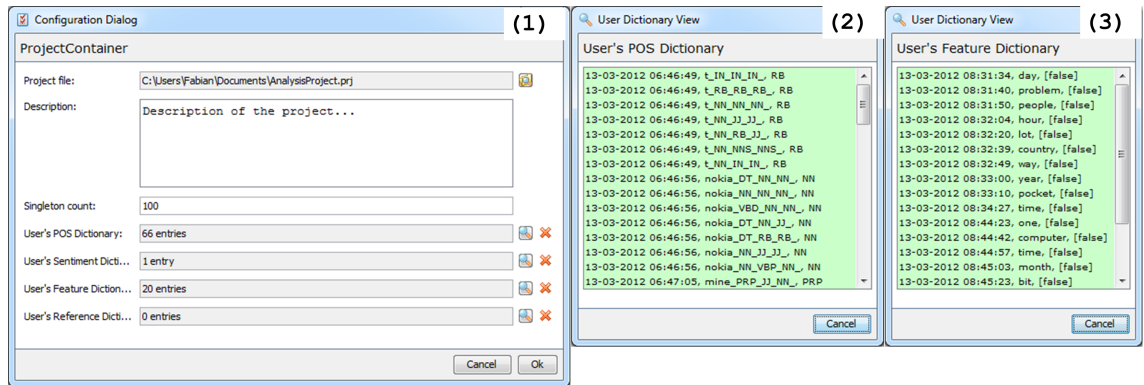


Abbildung 4.24: Dialog (1) zeigt die Übersicht über die Projektinformationen, wie Projektdatei, Projektbeschreibung und die jeweiligen Benutzeränderungen zu Wortart, Meinungskandidat, Produktmerkmalskandidat und Referenz. Dialog (2) gibt die Benutzeränderungen der Wortarten an. Dialog (3) gibt die Benutzeränderungen der Produktmerkmalskandidaten an.

Änderungsdialoge Um den Analyseprozess steuern zu können, kann der Benutzer auf Änderungsdialoge zurückgreifen. Diese modellieren die in Kapitel 3.2 angesprochenen Lösungsansätze und ermöglichen dem Benutzer Korrekturen von Wortarten, Meinungskandidaten, Produktmerkmalskandidaten und Referenzen. Eine lokale Änderung wird im Gegensatz zur globalen Änderung nicht im Projekt gespeichert. Dabei kann die Korrektur durch automatisch generierte Vorschläge beschleunigt werden.

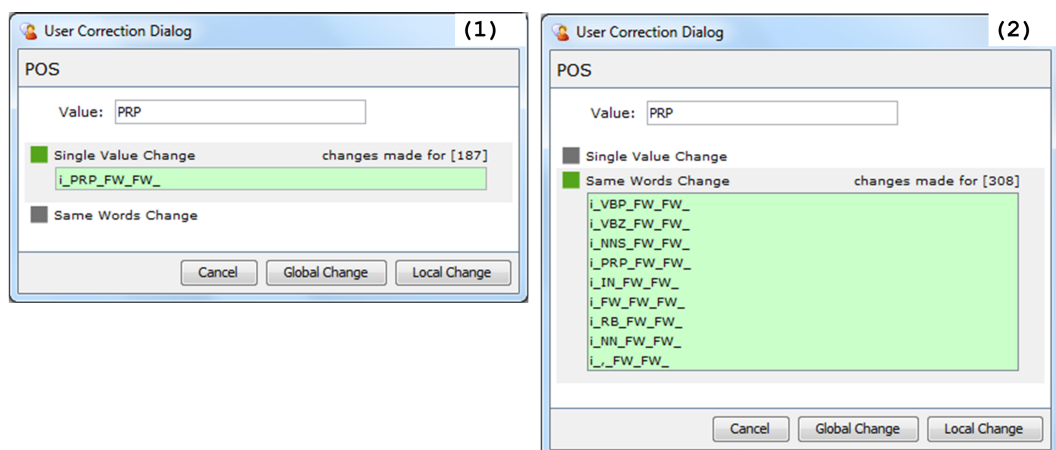


Abbildung 4.25: Hierbei soll der Term „I“ als Personalpronomen („PRP“) markiert werden. Dialog (1) stellt dies für den Einzelfall dar, während Dialog (2) Vorschläge für ähnliche Terme abbildet. Hier macht es Sinn, die Korrektur auf alle Vorschläge anzuwenden.

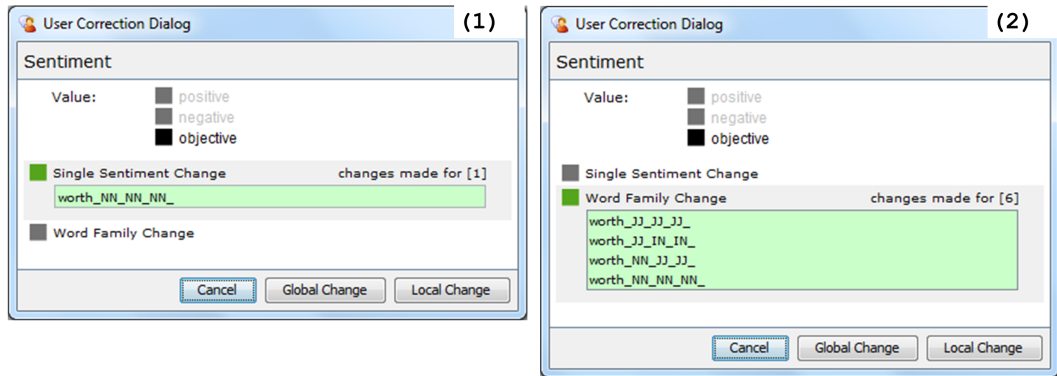


Abbildung 4.26: Hierbei soll der Term „worth“ mit einer neutralen Meinungsausrichtung versehen werden. Dialog (1) würde dies nur auf Nomen anwenden, wohingegen Dialog (2) noch zusätzlich ähnliche Terme vorschlägt.

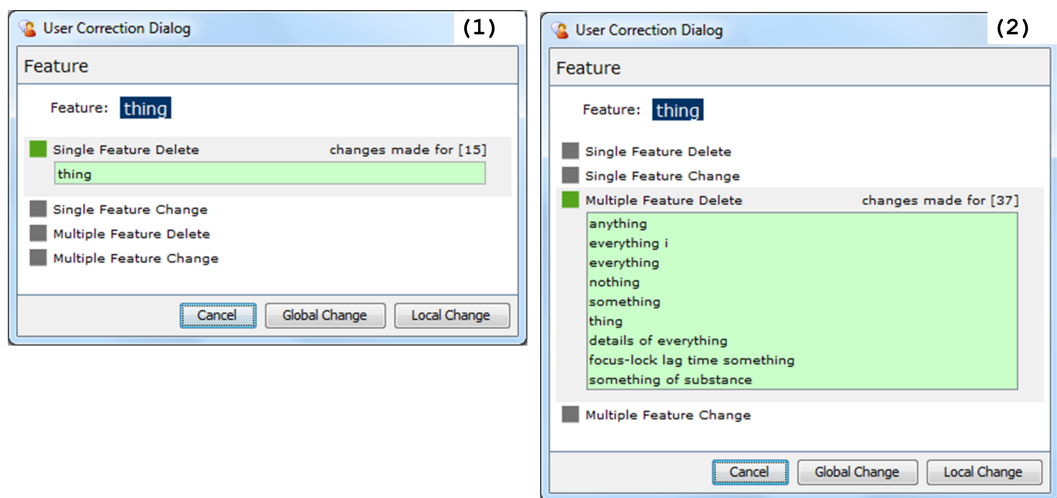


Abbildung 4.27: Hier soll der Produktmerkmalskandidat „thing“ korrigiert werden, da er nur in manchen Fällen ein Merkmal darstellt. Dialog (1) betrachtet hierbei nur einen Auftretungsfall, wohingegen Dialog (2) ähnliche Merkmale darstellt. Hier kann der Analyst spezifische Fälle, wie „anything“ oder „nothing“ markieren, um nur diese aus dem Analyseprozess ausschließen zu lassen.

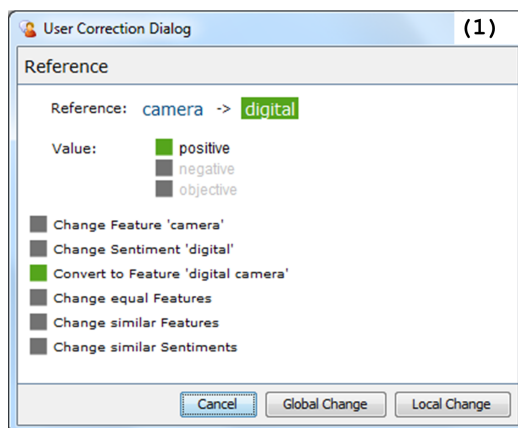


Abbildung 4.28: Für Referenzen existieren unterschiedliche Änderungsfälle. Das gezeigte Beispiel beschreibt einen Fehler, da eine „Digitalkamera“ gemeint ist. Somit muss der Analyst diese Referenz in einen Produktmerkmalskandidaten umwandeln.

4.3.4 Interaktionsszenario

Wofür die Oberflächenelemente nützlich sind und welche Interaktionsmöglichkeiten der Benutzer jeweils ausführen kann, soll nun in diesem Unterkapitel mittels eines Szenarios formuliert werden. Hierbei wird der Benutzer einen Analyseprozess ausführen und eine Korrektur der Wortartinformationen vornehmen.

Start Die Anwendung wird in der Klasse *MainFrame* gestartet. Es erscheint ein *Startbildschirm* bis die Oberfläche und deren Komponenten geladen wurden. Danach wird die eigentliche Anwendung sichtbar, wobei zu Anfang ein leeres Projekt geöffnet wird. Darin befinden sich keine Benutzerentscheidungen, wodurch der Analyseprozess auf kein Benutzerwissen zurückgreifen kann.

Einstellungen Der Benutzer kann nun in der *Ansicht der Analyseprozesskomponenten* Operationen für seine Analyse auswählen. Hierfür wählt er mit der Maus die Methoden und Module, welche er für seinen Analyseprozess verwenden will. Die Eigenschaften der Methoden kann er über die Schaltfläche neben des Methodennamens einsehen, indem sich ein *Konfigurationsdialog* öffnet. Zum Beispiel wählt der Benutzer in Abbildung 4.29 für die Einleseroutine *ReviewReaderBingLiu* eine Eingabedatei, welche Kundenrezensionen zu dem Produkt „Canon G3“ enthält.

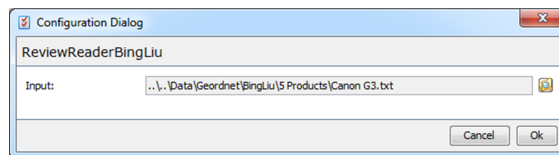
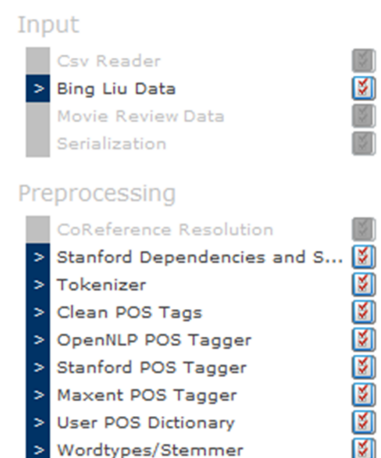


Abbildung 4.29: Der Einstellungsdialog für die Einleseroutine *ReviewReaderBingLiu* verlangt einen Pfad zu einer Eingabedatei, welche Kundenrezensionen beinhaltet. Hierbei hat der Benutzer das Produkt „Canon G3“ gewählt.

Für den Analyseprozess werden nun weitere Methoden ausgewählt, die in der Prozesspipeline abgearbeitet werden sollen. Wie auf der rechten Seite zu sehen, sollen für dieses Szenario nur bestimmte Methoden verwendet werden. Hierzu zählt zunächst die Einleseroutine (*Bing Liu Data*). Zur Vorverarbeitung werden Abhängigkeitseigenschaften (*Stanford Dependencies*), einzelne Wörter (*Tokenizer*), Wortarten (*POS Tagger*), Stopwörter und Grundformen (*Wordtypes/Stemmer*) berechnet.



Berechnung Sind die einzelnen Methoden ausgewählt, so kann der Analyseprozess über die Schaltfläche in der Menüleiste gestartet werden. Da die Eingabedaten mittels der Stanford-Bibliothek vorverarbeitet werden müssen, benötigt die Abarbeitung fünf Minuten Rechenzeit. Danach erscheinen die Visualisierungen und der Benutzer erhält einen Einblick über die berechneten Informationen.

Analyse Der Benutzer kann nun mit der Korrektur von Fehlern beginnen. Hierzu betrachtet er zunächst einmal die identifizierten Wortarten, welche in der globalen Ansicht im ersten Tabulator angezeigt werden. Wie in Abbildung 4.30 ersichtlich, liegen die einzelnen Wörter sortiert nach den häufigsten Fehlentscheidungen vor. Dies ist durch die niedrigen Vertrauenswerte in der letzten Spalte erkennbar. Zusätzlich werden Auftreten mit gleichem Vertrauenswert nach Häufigkeit absteigend sortiert.

Rank	Word	POS	CONF	Occurrence
1.	x	NN LS SYM	0,33	[1]
2.	read	VBD NN VB	0,33	[1]
3.	g3	NN JJ CD	0,33	[1]
4.	g3	CD NNS NN	0,33	[1]
5.	g2	CD JJ NN	0,33	[1]
6.	+	NN IN RB	0,33	[1]
7.	i	PRP FW FW	0,67	[187]
8.	i	NN FW FW	0,67	[38]
9.	i	IN FW FW	0,67	[23]
10.	i	VBP FW FW	0,67	[10]

Abbildung 4.30: Die Tabellenansicht stellt die Wortartidentifikationen dar. Hierbei sind die einzelnen Einträge nach Vertrauenswert aufsteigend sortiert, um dem Benutzer die unsichersten Entscheidungen vorlegen zu können. In dieser Ansicht sind die vier unsichersten Vorkommen des „i“-Phänomens markiert.

Zu analytischen Zwecken kann der Benutzer einzelne Entscheidungen filtern, um nur noch diese Phänomene in der Satzansicht zu betrachten. Hierzu wählt er eine oder mehrere dieser mit der Maus aus, wodurch der Filter zeitgleich aktualisiert wird. In diesem Szenario ist nun ersichtlich, dass sich bei dem Wort „i“ viele Unsicherheiten aus der Wortartidentifikation ergeben. Zum Beispiel erkennt in 187 Fällen nur eine Methode dieses Wort als Personalpronomen. In weiteren Fällen ist selbst die Entscheidung der ersten Methode falsch. Dieses Phänomen kann nun vom Benutzer korrigiert werden.

Korrektur Die Korrektur der Wortarten kann entweder über die Satzansicht oder über die globale Ansicht erfolgen. Hier wählt der Benutzer entweder eine der drei angezeigten Wortarten, oder betätigt mit der Maus die Benutzeränderungsschaltfläche direkt daneben. Daraufhin öffnet sich der Benutzerdialog in Abbildung 4.31, worin nun die Änderung der Wortart erfolgen kann.

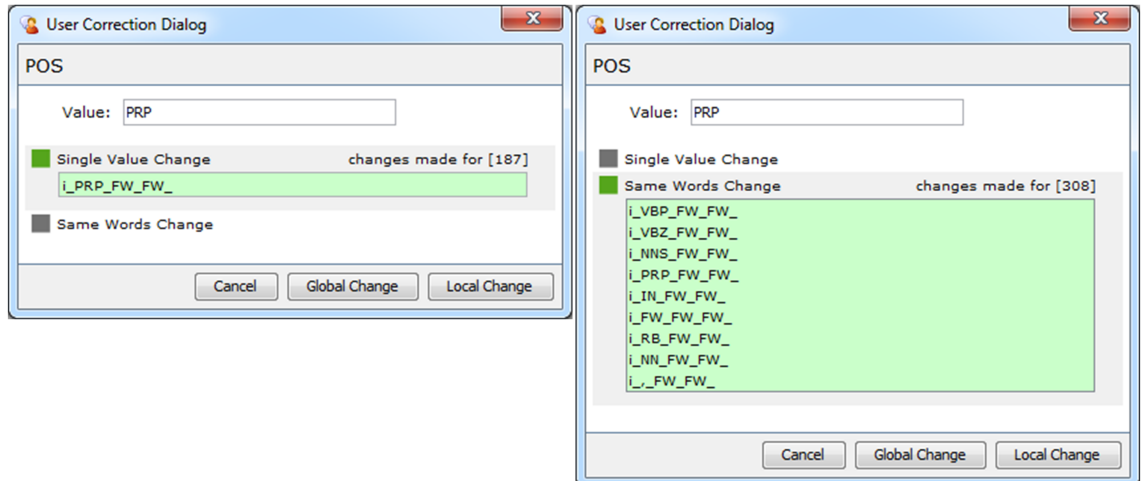


Abbildung 4.31: Der Korrekturdialog der Wortarten unterstützt zwei unterschiedliche Änderungsarten. Die Option *Single Value Change* ändert nur die Wortart für die ausgewählte Kombination von Wortartentscheidungen. Für die Option *Same Words Change* werden alle Phänomene für das zu ändernde Wort zusammengetragen, wobei der Benutzer diese Vorschläge beeinflussen kann.

Innerhalb des Dialogs werden dem Benutzer zwei unterschiedliche Arten von Änderungen angeboten. Zum einen kann er nur die Vorkommen der ausgewählten Kombination von Wortartentscheidungen ändern („*i_PRP_FW_FW_*“), oder es werden alle Phänomene für das Wort „*I*“ zusammengetragen. Hierbei kann er spezifische Fälle auswählen, wodurch sich die Menge der Änderungen verändert. Zusätzlich kann der Benutzer darüber entscheiden, ob die Korrekturen in der Projektdatei (*global*) oder nur zu Korrekturzwecken (*lokal*) gespeichert werden sollen. Für das aktuelle Beispiel entscheidet sich der Benutzer für die Korrektur aller Wörter (308 Fälle) und der globalen Speicherung. Dadurch, dass nahezu jedes Auftreten des Wortes „*I*“ ein Personalpronomen beschreibt, können zum einen diese Fehler behoben werden und zum anderen mittels der Projektdatei für andere Rezensionsdaten wiederverwendet werden.

Fertigstellen Bestätigt der Benutzer diese Korrektur, werden die Änderungsfälle in einem Wörterbuch abgelegt. Für den Analyseprozess bedeutet dies nun, dass sich die Modulentscheidungen verändert haben. Speziell diejenigen Methoden, welche von diesen neuen Entscheidungen abhängig sind, müssen informiert werden. Dies wird über eine abermalige Berechnung des Analyseprozesses erreicht.

Kapitel 5

Erkenntnisse

Dieses Kapitel stellt die Erkenntnisse vor, die sich durch die in dieser Arbeit präsentierten Anwendung ergeben. Hierzu werden zunächst die zur Verfügung stehenden Evaluationsdaten vorgestellt. Anhand der Testspezifikation soll erörtert werden, wie die in Kapitel 3.3 vorgestellten Hypothesen beantwortet werden können. Danach werden die Erkenntnisse vorgestellt, welche durch interessante Funde erweitert und mit Beispielen erläutert werden. Anschließend soll ausblickend gezeigt werden, wie noch bestehende Schwierigkeiten gelöst werden können und welche Erweiterungen denkbar wären.

5.1 Testumgebung

Die Testumgebung besteht aus zwei Komponenten. Die Datengrundlage bilden die zahlreichen Benutzerkommentare aus unterschiedlichen Sparten und Produktplattformen und stellen die Evaluationsgrundlage dar. Die Untersuchungsgegenstände beschreiben die Identifikationen von Meinungskandidaten und Produktmerkmalskandidaten sowie der Referenzen, welche aus der Anwendung extrahiert werden. Welche Informationen für eine Evaluation zur Verfügung stehen und wie diese tabellarisch dargestellt werden, soll nun im Folgenden erläutert werden.

5.1.1 Datengrundlage

Vor der eigentlichen Beschaffung der Daten wurde eine Recherche im Internet durchgeführt. Dabei konnten unterschiedliche Dokumentensammlungen ausfindig gemacht werden, welche für wissenschaftliche Zwecke frei verfügbar sind und zum Teil bereits Einsatz in anderen wissenschaftlichen Arbeiten finden. Diese Sammlungen wurden anhand von Gesichtspunkten bezüglich Effizienz und Effektivität bewertet, wobei die folgende Auflistung diese Kriterien näher erläutert.

- **Effizienz:** Die Textdaten können sich ohne großen Mehraufwand in die Anwendung einpflegen lassen und beinhalten eine genügend große Anzahl an Benutzerkommentaren.
- **Effektivität:** Die Zusatzinformationen der Daten eignen sich für die Evaluation von merkmalsbezogenen Meinungen oder anderen Identifikationsmodulen.
- **Abdeckung:** Die Benutzerkommentare befinden sich in unterschiedlichen, aber vergleichbaren Sparten.

Anhand dieser Kriterien konnten vier Dokumentensammlungen ausgewählt werden, welche in dieser Arbeit Verwendung finden. Diese beinhalten unter anderem Benutzerkommentare über verschiedene Produkte aus Verkaufsplattformen und Filmrezensionen aus Blogs. Abbildung 5.1 zeigt hierbei die Aufteilung der resultierenden Datensätze in die jeweilige Sparte.

	Camera	Mobile phone	Mp3 Player	DVD Player	Software	Router	Printer	Diaper	Movie
Additional Review Data Set	x	x	x		x	x		x	
Customer Review Data Set	x	x	x	x					
Amazon Review Data	x		x				x		
Movie Review Data									x

Abbildung 5.1: Die Tabelle zeigt die in dieser Arbeit verwendeten Dokumentensammlungen und deren Verteilung auf verschiedene Sparten.

Im weiteren sollen nun die einzelnen Dokumentensammlungen vorgestellt werden, speziell mit dem Augenmerk darauf, welche Informationen für eine spätere Evaluation zur Verfügung gestellt werden können und inwiefern diese für die Evaluation der Analyseanwendung effektiv und effizient nutzbar sind.

Additional Review DataSet & Customer Review DataSet Die Autoren Hu, Ding und Liu annotierten Kundenrezensionen über 14 verschiedenen Produkten, welche aus den Verkaufsplattformen *Amazon*³² und *Cnet*³³ stammen. Hierzu wurden für jeden subjektiven Satz diejenigen Produktmerkmale, welche eine Benutzermeinungen ausdrücken, zusammen mit deren Orientierung markiert [27] [12]. Abbildung 5.2 zeigt das Annotationsschema anhand zweier beispielhafter Benutzerkommentare.

³²<http://www.amazon.com/>

³³<http://www.cnet.com/>

the **pictures** are **absolutely amazing**. → **pictures (+1)**
 it fits in a **pocket nicely**. → **size (+1)**

Abbildung 5.2: Die Autoren Hu, Ding und Liu markieren in ihren annotierten Datensätzen jedes meinungstragende Produktmerkmal, jeweils einhergehend mit einer Orientierung der Benutzermeinung. Es gibt explizite Produktmerkmale wie „pictures“ die direkt im Text erwähnt werden. Der zweite Benutzerkommentar beinhaltet ein implizites Produktmerkmal, denn „size“ taucht nicht direkt im Text auf, wird aber durch das Verb „fits“ impliziert.

Die in den Datensätzen annotierten Produktmerkmale müssen nun für die Evaluation angepasst werden, da nicht alle Informationen durch die Konzeption des Analyseprozesses verwertbar sind. Hierbei werden diejenigen Daten eliminiert, welche durch der Analyseanwendung nicht berechnet werden können. Hierzu zählen zum Beispiel implizite Produktmerkmale und Produktmerkmale, welche nicht durch Nomen ausgedrückt werden. Es verbleibt folglich nur noch ein Teil der Ursprungsdaten. Hierbei zeigt Abbildung 5.3 die Verteilung der annotierten Daten auf die einzelnen Fehlerquellen und gibt die Größe der verbliebenen Datenmenge an.

Product	Annotations	Implicit	POS	Remaining
Canon G3	300	22%	13%	194
Canon PowerShot SD500	157	3%	18%	123
Canon S100	229	11%	17%	166
Nikon Coolpix 4300	213	22%	11%	144
MP3 1	889	20%	12%	596
iPod	201	15%	16%	138
MP3 2	629	42%	12%	288
Nokia 6600	482	23%	8%	332
Nokia 6610	361	16%	6%	283
Hitachi router	268	31%	10%	159
Linksys router	234	29%	14%	134
DVD player	454	27%	17%	252
Diaper champ	247	21%	23%	138
Norton	255	16%	20%	162

Abbildung 5.3: Die Tabelle stellt für jedes Produkt die Anteile der spezifischen Fehlergruppen gegenüber. Zunächst besteht der Datensatz aus einer bestimmten Anzahl an annotierten Produktmerkmalen. Nach der Filterung von impliziten Merkmalen und Verben resultiert die Menge, welche als Datengrundlage für die Evaluation zur Verfügung stehen soll.

Zurückgegriffen auf das Beispiel in 5.2, wird das Produktmerkmal „size“, welches durch ein Verb impliziert wird, für die Evaluation nicht weiter betrachtet, wodurch die Ergebnisse genauer interpretiert werden können. Es stehen also nur noch diejenigen Sätze zur Verfügung, in denen das annotierte Produktmerkmal mit einem tatsächlichen Satzbestandteil übereinstimmt und als Nomen repräsentiert ist. Aus diesen Daten resultiert nun eine Tabelle, wie in Abbildung 5.4 schemenhaft dargestellt, welche für jeden Satz angibt, welches Produktmerkmal darin *positiv* oder *negativ* bewertet wurde.

RID	SID	sentence	feature	sentiment
0	0	it is a nice phone.	phone	+1
0	1	a nice phone with nice design.	phone	+1
0	1	a nice phone with nice design.	design	+1
0	2	it has a really bad design.	design	-2

Abbildung 5.4: Die Datentabelle von Hu, Ding und Liu gibt an, in welchem Satz welches Produktmerkmal mit welcher Meinungsrichtung bewertet wurde. Es gibt fünf Spalten. Die ersten zwei Spalten geben die Position des Satzes (*Review-ID* und *Satz-ID*) an. Danach folgt der Inhalt des Satzes selbst und das explizite Produktmerkmal. Die letzte Spalte gibt die Ausrichtung des Merkmals an. Finden sich innerhalb eines Satzes mehrere Merkmale, so wird hierzu jeder Satz einzeln betrachtet.

Diese Datentabelle eignet sich sehr gut zur Evaluation der Identifikation der Produktmerkmalskandidaten, sowie von merkmalsbezogenen Meinungen. Auch die Abdeckung in unterschiedlichen Sparten und mehreren ähnlichen Produkten lassen einen großen Evaluationsspielraum zu.

Amazon Korpus Diese Dokumentenkollektion beinhaltet Benutzerkommentare über verschiedenste Produkte der Verkaufsplattform *Amazon*³⁴. Hierfür wurden 3 Produktgruppen ausgewählt und darin getätigte Kundenrezensionen über verschiedene Produkte innerhalb dieser Sparten heruntergeladen. Nach Säuberung der Kommentare von *HTML*-Zusätzen, wurden die Textdaten anhand ihrer Bewertung gruppiert. Abbildung 5.5 zeigt die Übersicht dieser Dokumentenkollektion und gibt Aufschluss über die Gruppenverteilungen.

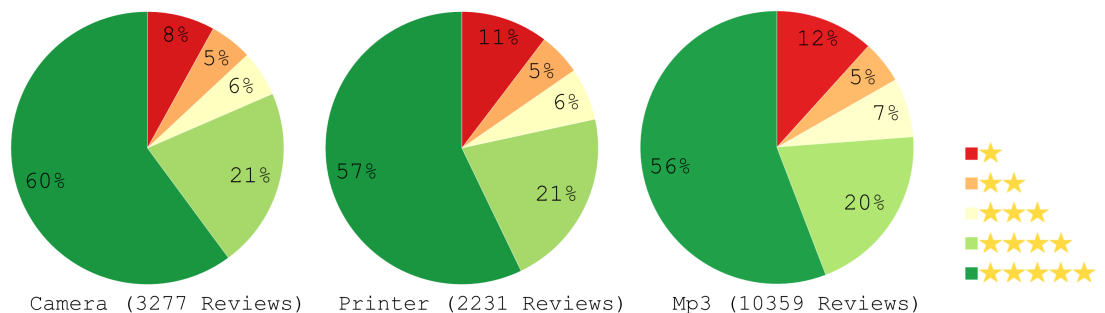


Abbildung 5.5: Die Übersicht der Amazon-Dokumentenkollektion zeigt die Verteilung der verschiedenen Sterne-Bewertungen innerhalb der Produktklassen. Hier steht die Farbe rot für negative Rezensionen, welche entweder mit ein oder zwei Sternen bewertet wurden, und die Farbe grün für positive Rezensionen mit vier oder fünf Sternen.

Für eine Evaluation besteht hier die Möglichkeit, die Benutzermeinungen auf Rezensionsebene zu betrachten, welche in eine von fünf Untergruppen eingeteilt ist. Hierzu können die Identifikationen der Meinungskandidaten oder die Meinungen innerhalb der Referenzen untersucht werden.

³⁴<http://www.amazon.com/>

Filmdaten Diese Kollektion, welche in der Arbeit der Autoren Pang und Lee [41] verwendet wird, besteht aus Benutzerrezensionen der *ImDB-Datenbank*³⁵, wobei diese anhand des entwickelten Klassifikators der Autoren in *positive* und *negative* Rezensionen eingeteilt wurde. Die gesamte Kollektion umfasst insgesamt 1000 positive und 1000 negative Benutzerrezensionen, wobei die negative Klasse 33066 Sätze und die positive Klasse 34420 Sätze beinhaltet. Eine Benutzerrezension besteht hierbei aus durchschnittlich 21 Sätzen. Für die Evaluation steht die Benutzermeinung über einen ganzen Filmbericht zur Verfügung. Dieser Datensatz eignet sich somit für die Untersuchung von Meinungskandidaten und von Meinungen der Referenzen innerhalb der Filmdomäne.

5.1.2 Ausgabe der Anwendung

Die Benutzerrezensionen der Datengrundlage werden nun in die Anwendung eingelesen, verarbeitet und es werden Meinungskandidaten, Produktmerkmalskandidaten sowie deren Beziehungen identifiziert. Diese können dann getrennt aus der Anwendung exportiert werden, um das jeweilige Identifikationsmodul betrachten zu können. So sind je nach Modul auch andere Informationen verfügbar, woraus unterschiedliche Tabellenformate resultieren. Inwiefern sich diese unterscheiden, soll nun näher erläutert werden.

5.1.2.1 Meinungskandidaten

Wird innerhalb eines Satzes ein meinungstragendes Wort identifiziert, so entsteht ein Datentupel aus den Positionsdaten (*Review-ID* und *Satz-ID*), dem Inhalt des Satzes und den eigentlichen Informationen zu diesem Meinungswort. Hierzu zählen das Wort selbst, deren Ausrichtung, Vertrauenswert und den daraus resultierenden Rang. Sätze, welche kein Meinungswort beinhalten, werden ebenfalls in der Ausgabetabelle aufgenommen. Hierdurch können Meinungswörter einzeln oder deren Ausrichtungen auf Satz- oder Rezensionsebene betrachtet werden.

RID	SID	sentence	sentiment	value	confidence	rank
0	0	it is a nice phone.	nice	+1	0.05	1
0	1	a nice phone with nice design.	nice	+1	0.05	1
0	1	a nice phone with nice design.	nice	+1	0.05	1
0	2	it has a really bad design.	bad	-1	0.025	3
1	0	a sentence without sentiment.				

Abbildung 5.6: Die Ausgabetabelle der Meinungskandidaten gibt für jedes Vorkommen an, in welchem Satz es identifiziert wurde (*Review-ID*, *Satz-ID* und Satzinhalt), um welches Wort es sich hierbei handelt und welche Zusatzinformationen die Analyse berechnet hat. Hierbei handelt es sich um die Ausrichtung, einen Vertrauenswert und den resultierenden Rang. Sätze mit mehreren Meinungswörtern werden mehrfach aufgeführt.

³⁵<http://www.imdb.com/>

5.1.2.2 Produktmerkmalskandidaten

Wird innerhalb eines Satzes ein Produktmerkmalskandidat identifiziert, so entsteht ein Datentupel aus den Positionsdaten (*Review-ID* und *Satz-ID*), dem Inhalt des Satzes und dessen eigentlichen Informationen. Hierzu zählen die identifizierte Nomenkette, dessen Vertrauenswert und den daraus resultierenden Rang. Sätze, welche kein Produktmerkmalskandidat beinhalten, werden ebenfalls aufgenommen. Die hieraus resultierende Tabelle enthält alle von der Anwendung berechneten Produktmerkmalskandidaten und wird bei der Evaluation von Produktmerkmalen eingesetzt.

RID	SID	sentence	feature	confidence	rank
0	0	it is a nice phone.	phone	0.05	1
0	1	a nice phone with nice design.	phone	0.05	1
0	1	a nice phone with nice design.	design	0.025	3
0	2	it has a really bad design.	design	0.025	3

Abbildung 5.7: Die Ausgabetable der Produktmerkmalskandidaten gibt für jedes Vorkommen an, in welchem Satz dieser identifiziert wurde (*Review-ID*, *Satz-ID* und Satzinhalt), um welchen Kandidaten es sich hierbei handelt und welche Zusatzinformationen die Analyse berechnet hat, also dessen Ausrichtung, Vertrauenswert und Rang des Kandidaten. Sätze mit mehreren Produktmerkmalskandidaten werden mehrfach aufgeführt.

5.1.2.3 Referenzen

Referenzen beschreiben die Beziehung zwischen einem Meinungskandidaten und einem Produktmerkmalskandidaten. Ein Datentupel besteht hierbei aus den Positionsdaten (*Review-ID* und *Satz-ID*), dem Inhalt des Satzes und den Informationen zum Produktmerkmalskandidaten und zur referenzierenden Meinung. Zuletzt beinhaltet die Tabelle die Informationen der eigentlichen Referenz. Hierzu zählen die Meinungsrichtung der Referenz, dessen Vertrauenswert und einem daraus resultierenden Rangwert. Diese Tabelle stellt die identifizierten Produktmerkmalskandidaten mit den meinungstragenden Kandidaten in Beziehung und kann somit zur Evaluation von merkmalsbezogenen Meinungen herangezogen werden. Da sich innerhalb dieser Tabelle die Informationen zu tatsächlichen Meinungskandidaten und Produktmerkmalskandidaten befinden, können diese zusätzlich auch getrennt voneinander untersucht werden.

5.2 Evaluationspezifikation

Die Evaluationspezifikation beschreibt die Evaluationstechniken der einzelnen Identifikationsmodule. Hierbei wird erläutert, welche Informationen innerhalb der Datengrundlage betrachtet werden müssen, um eine Untersuchung der Module zu ermöglichen. Hierbei sind unterschiedliche Evaluationsarten vorgesehen. Es können sowohl merkmalsbezogene Meinungen, als auch die einzelnen Module untersucht werden. Dabei müssen die Identifikationen entweder einzeln oder auf Rezensenebene betrachtet werden, wobei diese Betrachtungsweise von den Informationen der Datengrundlage abhängig ist. Abbildung 5.8 zeigt einen Überblick über die einzelnen Dokumentensammlungen und die daraus ableitbaren Evaluationsarten.

	Feature	Sentiment	Reference	Reference 's Sentiment	Reference 's Feature	Negations
Additional Review Data Set	x		x		x	x
Customer Review Data Set	x		x		x	x
Amazon Review Data		x		x		x
Movie Review Data		x		x		x

Abbildung 5.8: Die Tabelle stellt die Kollektionen und die daraus ableitbaren Evaluationsarten dar.

Im weiteren werden nun die einzelnen Module betrachtet und näher erläutert welche Berechnungen nötig sind, um Aussagen über deren Genauigkeit treffen zu können.

5.2.1 Meinungskandidaten

Die Evaluation der Meinungskandidaten erfolgt auf dem Amazon- und dem Film-Korpus, wobei deren Charakteristiken bereits in Kapitel 5.1.1 näher vorgestellt wurden. Hierbei beinhaltet der Amazon-Korpus Rezensionen, welche in 5-Sterne-Gruppen eingeteilt sind und der Film-Korpus enthält entweder als positiv oder negativ markierte Rezensionen. Diese unterschiedlichen Bewertungen müssen zunächst vereinheitlicht werden. Hierbei werden die 5-Sterne-Bewertungen des Amazon-Korpus auf die positive oder negative Klasse transformiert. Abbildung 5.9 veranschaulicht diese Transformation für beide Korpora.

	Positive	Negative
Amazon Reviews	> 3 Stars	<= 2 Stars
Movie Reviews	Positive	Negative

Abbildung 5.9: Die Tabelle gibt an, welche Klassenzugehörigkeit sich aus den Bewertungen der Benutzer ergeben. So müssen die 5-Sterne-Gruppen des Amazon-Korpuses auf zwei Klassen aufgeteilt werden. Dabei werden die Rezensionen mit drei Sternen nicht mehr weiter betrachtet.

Für die Evaluation werden nun die identifizierten Meinungskandidaten zur Bestimmung der Klassenzugehörigkeit einer Rezension eingesetzt, welche dann mit der Bewertung der Benutzer verglichen werden kann. Dabei wird davon ausgegangen, dass positiv bewertete Rezensionen häufig mehr positive Meinungswörter und negativ bewertete Rezensionen häufig mehr negative Meinungswörter enthalten [41]. Da neutrale Bewertungen hierdurch nicht eindeutig beschrieben werden können, werden diese für die Evaluation nicht weiter betrachtet. Nun werden die einzelnen Meinungsorientierungen innerhalb einer Rezension aufsummiert, wobei sich hierdurch ein Wert ergibt, welcher die spezifische Klassenzugehörigkeit ausdrücken kann. Hierbei können zusätzlich die Vertrauenswerte der einzelnen Meinungskandidaten einbezogen werden, um zum Beispiel Unsicherheiten innerhalb der Identifikation zu berücksichtigen. Abbildung 5.10 zeigt die zwei unterschiedlichen Formeln zur Berechnung einer Klassenzugehörigkeit.

$$\begin{aligned}
 (1) \text{ sentiment}_r &= \sum_{i=1} \text{subj}(w_{r,i}) \\
 (2) \text{ sentiment}_r &= \sum_{i=1} \text{subj}(w_{r,i}) \cdot \text{conf}_{\text{SENT}}(w_{r,i})
 \end{aligned}$$

Abbildung 5.10: Formel (1) beschreibt die Summe der Meinungsausrichtungen, während Formel (2) diese Ausrichtung zusätzlich mit dem jeweiligen Vertrauenswert des Meinungskandidaten gewichtet.

Die Evaluation der Meinungskandidaten betrachtet die positive und negative Klasse getrennt, da sich die Klassengrößen innerhalb der Datengrundlage sowie innerhalb der einzelnen Meinungswörterbücher deutlich unterscheiden. Um dies zumindest für den ersten Fall umgehen zu können, werden die Klassen der Datengrundlage per Zufallsauswahl in ihrer Größe angeglichen.

Für die Evaluation wird nun die berechnete Klassenzugehörigkeit jeder Rezension mit der Benutzerbewertung aus der Datengrundlage gegenübergestellt. Hieraus lassen sich schließlich Werte berechnen, welche die Genauigkeit und Abdeckung der Meinungskandidatenidentifikation beschreiben. Abbildung 5.11 auf der folgenden Seite gibt hierbei Aufschluss über unterschiedliche Fälle, welche dann mittels der Formeln in Abbildung 5.12 verrechnet werden können.

Sum	Class	Positive	Negative
> 0	POSITIVE	pp	pn
< 0	NEGATIVE	np	nn
= 0	NEUTRAL	op	on

Abbildung 5.11: Je nach Vorzeichen der Summe ergeben sich drei unterschiedliche Klassenzugehörigkeiten (**POSITIVE**, **NEGATIVE** oder **NEUTRAL**), welche mit den Klassen der Datengrundlage (*Positive* oder *Negative*) verglichen werden.

$precision_{(+)} = \frac{pp}{pp + pn}$ $recall_{(+)} = \frac{pp}{pp + np + op}$ $fmeasure_{(+)} = \frac{2 \cdot precision_{(+)} \cdot recall_{(+)}}{precision_{(+)} + recall_{(+)}}$	$precision_{(-)} = \frac{nn}{np + nn}$ $recall_{(-)} = \frac{nn}{pn + nn + on}$ $fmeasure_{(-)} = \frac{2 \cdot precision_{(-)} \cdot recall_{(-)}}{precision_{(-)} + recall_{(-)}}$
--	--

Abbildung 5.12: Die Berechnung der einzelnen Werte erfolgt für jede Klasse einzeln. Dabei beschreibt die *Precision*, wie viele der zum Beispiel als positiv berechneten Rezensionen richtig sind. Der *Recall* gibt an, wie viele Rezensionen der jeweiligen Klasse auch wirklich richtig eingestuft wurden. Zusätzlich werden *F-Measure*-Werte berechnet.

5.2.2 Produktmerkmalskandidaten

Die Evaluation der Produktmerkmalskandidaten erfolgt auf den Datensätzen von Bing, Ding und Liu [27] [12], da hier bereits Annotationen über Produktmerkmale vorliegen. Hierfür werden die Informationen in einer Tabelle gegenübergestellt, welche für jeden Satz die darin befindlichen Merkmale auflistet. Um nun Aufschluss über die Genauigkeit der Identifikationsmethode erlangen zu können, wird innerhalb der Tabellenzeile nach Übereinstimmungen gesucht. Ein Vergleich beider Listen ergibt folglich unterschiedlichen Fälle, welche in die Tabelle in Abbildung 5.13 übertragen werden können. Diese Werte können dann für eine Evaluation herangezogen werden.

- **tp**: Beide Listen enthalten das gleiche Produktmerkmal.
- **fp**: Der Produktmerkmalskandidat wurde nicht annotiert.
- **fn**: Das Produktmerkmal aus der annotierten Liste ist nicht identifiziert worden.

	annotated	not annotated
calculated	tp	fp
not calculated	fn	

$precision = \frac{tp}{tp + fp}$
$recall = \frac{tp}{tp + fn}$

Abbildung 5.13: Die Tabelle gibt die unterschiedlichen Fälle beim Vergleich der Produktmerkmalskandidaten an. Die *Precision* beschreibt folglich, wie viele der identifizierten Produktmerkmalskandidaten auch annotiert wurden, während der *Recall* angibt, wie viele der annotierten Merkmale gefunden werden.

Für den Vergleich der Merkmale können hierbei unterschiedliche Gleichheitsbedingungen in Betracht gezogen werden:

- **Equality:** Merkmale gelten als gleich, wenn sie bis auf ihren Numerus übereinstimmen.
- **Substring:** Merkmale gelten als gleich, wenn eines ein Teilwort des anderen beschreibt.

Die erste Bedingung trifft zum Beispiel auf Wörter, wie „phones“ und „phone“ zu, welche mit großer Sicherheit als gleichwertige Produktmerkmale angesehen werden können. Betrachtet man die Merkmale „phone“ und „speakerphone“, so müsste hier die erste Bedingung gewählt werden, da hier zwei unterschiedliche Produktmerkmale ausgedrückt werden. Auf der anderen Seite können aber auch Produktmerkmalskandidaten wie „radio of the phone“, „nokia 6610 radio“ oder „phone’s radio“ existieren. Diese Beispiele zielen hier auf dasselbe Produktmerkmal ab, welche über die zweite Bedingung modelliert werden müsste.

Hierbei hat eine Vorabuntersuchung gezeigt, dass sich durch die *Equality-Bedingung* weniger Übereinstimmungen ergeben, als durch die *Substring-Bedingung*. Wobei diese primär im hohen Rangbereich greift. Hier befinden sich diejenigen Produktmerkmalskandidaten, welche selten im Text vertreten sind und häufig aus mehreren Wörtern bestehen. Betrachtet man den niedrigen Rangbereich, so erhält man mit der *Equality-Bedingung* nur geringfügig schlechtere Ergebnisse. Dadurch, dass sich durch die geforderte Gleichheit der Merkmale aber eine höhere Nachvollziehbarkeit ergibt, soll nun für den weiteren Verlauf der Evaluation ausschließlich die *Equality-Bedingung* eingesetzt werden.

Um die Rangfolge der Produktmerkmalskandidaten bei der Evaluation miteinbeziehen zu können, wird der Vergleich zwischen berechneten und annotierten Merkmalen jeweils auf eine spezifische Vorabauswahl angestellt. Diese Vorabauswahl wird über einen Schwellenwert definiert, welcher auf die Ränge der einzelnen Produktmerkmalskandidaten angewendet wird. Dieser wird als Laufvariable eingesetzt, wodurch er sich kontinuierlich vergrößert. In den Untersuchungen werden dabei alle Produktmerkmalskandidaten mit einem Rang kleiner oder gleich dem Schwellenwert betrachtet. Bedingt durch die Vergrößerung des Schwellenwertes fließen somit immer mehr Produktmerkmalskandidaten mit höherem Rangwerten und somit kleineren Vertrauenswerten in die Berechnung mit ein.

Dabei ergaben Voruntersuchungen, welche in Abbildung 5.14 dargestellt sind, dass die Produktmerkmalskandidaten der ersten 40 Rängen bereits 40 % des Gesamtvorkommens beschreiben³⁶. Zusätzlich ergab sich daraus, dass im Bereich der hohen Ränge zwar Produktmerkmalskandidaten

³⁶Bei diesem Test wurden die Häufigkeiten der Produktmerkmalskandidaten der ersten 40 Ränge mit den Häufigkeiten aller identifizierten Produktmerkmale verglichen

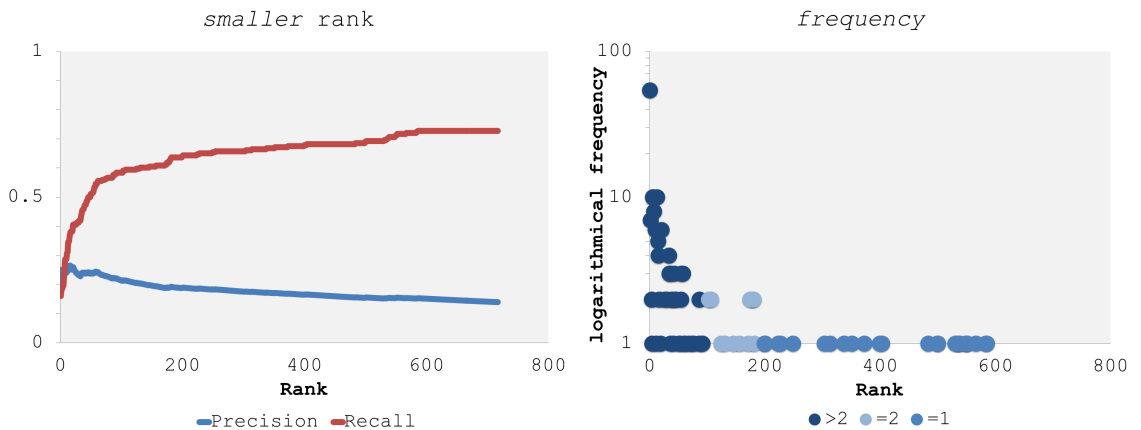


Abbildung 5.14: Das linke Diagramm beschreibt das Verhalten der *Precision*- und *Recall*-Werte der *smaller rank*-Bedingung. Das rechte Diagramm beleuchtet hierbei die Häufigkeiten der Produktmerkmalskandidaten, welche mit den Merkmalen aus der Datengrundlage übereinstimmen. Hier sind verschiedene Häufigkeiten mit anderen Blautönen versehen.

existieren, welche auch annotiert wurden, aber im Vergleich zu den ersten 40 Rängen relativ selten auftauchen. Wegen Gründen der Anschaulichkeit und der Aussagekraft dieses Bereiches werden Aussagen der Evaluation nur noch anhand der ersten 40 Ränge betrachtet, wodurch zumindest ein Großteil der Daten abgedeckt werden kann.

5.2.3 Referenzen

Für die Evaluation der Referenzen stehen unterschiedliche Herangehensweisen zur Verfügung. Dadurch, dass sie tatsächliche Produktmerkmale und tatsächliche Meinungen beinhalten, können zunächst diese Informationen getrennt untersucht werden. Hierzu kann auf die vorhergehenden Evaluationsspezifikationen zurückgegriffen werden. Für die Evaluation der Meinungen kann zum Beispiel nun die Meinungsausrichtung der Referenz betrachtet werden, wodurch die eigentliche Meinungsausrichtung durch Negationen oder Kontext-sensitive Produktmerkmale invertiert werden sein kann.

Die eigentliche Evaluation der Referenzen soll hierbei abermals über die annotierten Daten der Autoren Bing, Ding und Liu [27] [12] erfolgen. Dabei wurde aber nicht direkt markiert, welches Meinungswort sich auf welches Produktmerkmal innerhalb des Satzes bezieht, sondern das Produktmerkmal wurde lediglich mit einer Meinungsausrichtung versehen. Dadurch können keine sicheren Aussagen getroffen werden, dass die Referenz auch tatsächlich mit der Annotation übereinstimmt. Die durch die Anwendung identifizierten neutralen und subjektiven Referenzen, welche erst durch einen Benutzer validiert und korrigiert werden müssen, werden in der Evaluation nicht betrachtet. Denn neutrale Meinungen besitzen keinen relevanten Meinungsgehalt und subjektive Meinungen werden erst durch eine Benutzerkorrektur mit einer Meinungsausrichtung versehen.

Die Evaluation geht hierbei ähnlich wie bei der Untersuchung der Produktmerkmalskandidaten vor. Dabei werden abermals die annotierten Merkmale mit den identifizierten Merkmalen innerhalb der Referenzen gegenübergestellt. Diejenigen Referenzen, welche sich nicht auf ein annotiertes Merkmal beziehen, können hierbei nicht betrachtet werden. Dabei wird abermals die *Equality-Bedingung* eingesetzt, um eine Nachvollziehbarkeit des Ergebnisses gewährleisten zu können. Somit werden nur diejenigen Referenzen überprüft, welche auch als Annotation vorliegen. Ein Vergleich ergibt hierbei unterschiedliche Fälle, welche für eine Evaluation des Moduls herangezogen werden. Diese werden in die Tabelle aus Abbildung 5.15 übertragen, wodurch schließlich *Precision*- und *Recall*-Werte berechnet werden können.

- **tp**: Referenz und Annotation verweisen auf dasselbe Produktmerkmal, wobei die Meinungsrichtungen identisch sind.
- **fp**: Referenz und Annotation verweisen auf dasselbe Produktmerkmal, wobei die Meinungsrichtungen unterschiedlich sind.
- **fn**: Das annotierte Produktmerkmal wurde nicht referenziert.

	correct sentiment	incorrect sentiment
calculated	tp	fp
not calculated	fn	

$$precision = \frac{tp}{tp + fp}$$

$$recall = \frac{tp}{tp + fn}$$

Abbildung 5.15: Die Tabelle gibt die unterschiedlichen Vergleichsfälle der Referenzen an, woraus sich dann *Precision*- und *Recall*-Werte berechnen lassen.

Hierbei kann aus der *Precision* abgeleitet werden, wie viele Meinungsrichtungen der annotierten Referenzen korrekt identifiziert wurden. Der *Recall* betrachtet die Seite der Produktmerkmale und gibt somit an, wie viele der annotierten Merkmale auch referenziert wurden. Diese Werte können somit nur Hinweise auf die Genauigkeiten der Identifikationen von Referenzen liefern, was aber die Evaluation noch zeigen wird.

5.3 Ergebnisse

Dieses Kapitel beschreibt nun die eigentliche Evaluation des Analyseprozesses und der Interaktiven Steuerung. Dabei bilden die in Kapitel 3.3 vorgestellten Hypothesen den Evaluationsrahmen für diese Arbeit. Die Ausgaben der Anwendung sollen mittels der Evaluationstechniken aus Kapitel 5.2 analysiert werden. Dabei werden die jeweiligen Identifikationsmodule getrennt betrachtet, wobei einzelne Untersuchungen aufeinander aufbauen. Die in Kapitel 5.1 vorgestellten Dokumentenkollektionen bilden hierbei die *Ground-Truth*-Daten, worüber eine Evaluation stattfinden kann. Dieses Kapitel beleuchtet zunächst den Analyseprozess und die dahinter liegende Modulstrukturierung. Daraufhin wird die Interaktive Steuerung untersucht und deren Chancen für eine Verbesserung des Analyseprozesses aufgezeigt. Zuletzt folgt die Betrachtung der Übertragbarkeit der Korrekturen auf Daten der gleichen Domäne und abschließende Schlussfolgerungen der Evaluation.

5.3.1 Analyseprozess

Der Analyseprozess besteht aus mehreren Modulen, welche unterschiedliche Aufgaben zur Analyse von Meinungen bewältigen. Dabei wurden Verfahren zur Identifikation von Wortarten, Meinungskandidaten, Produktmerkmalskandidaten und Referenzen in einer Analyseanwendung implementiert. Dabei beinhalten diese Module mehrere Methoden, welche unterschiedliche Herangehensweisen betrachten. Diese liefern Einzelentscheidungen, welche innerhalb des Moduls zu einer Gesamtentscheidung verrechnet werden. Hieraus können Vertrauenswerte abgeleitet werden, welche die jeweiligen Entscheidungen beurteilen.

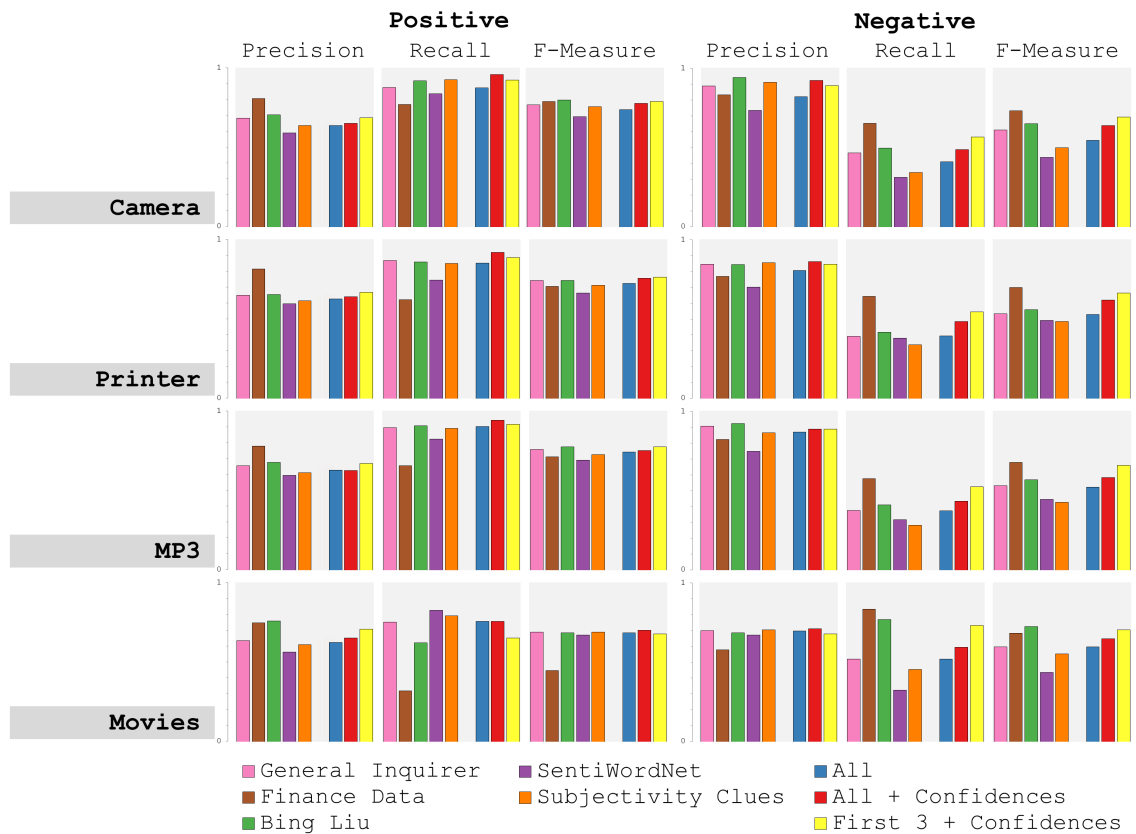
Hypothese: Die folgenden Untersuchungen wollen nun zeigen, dass die Vereinigung von unterschiedlichen Methoden zur Lösung derselben Identifikationsaufgabe den Umfang der Analyseanwendung vergrößert, wodurch sich ein besseres Ergebnis ableiten lässt.

Dabei werden primär die Module zur Identifikation von Meinungskandidaten und Referenzen untersucht, da darin eine Vertrauenswertberechnung modelliert ist, welche auf die Vereinigung mehrerer Methoden aufbauen. Die Identifikation von Wortarten steht hierbei im Hintergrund und soll in diesem Unterkapitel nicht weiter betrachtet werden.

5.3.1.1 Meinungskandidaten

Die Identifikation von Meinungskandidaten kann über sechs verschiedene Methoden realisiert werden, wobei sich die in Kapitel 3.2.5 vorgestellten Wörterbuchmethoden in ihren Inhalten unterscheiden. So enthalten sie zum Beispiel unterschiedliche Mengen an positiven und negativen Meinungswörtern. Die folgende Untersuchung will nun verdeutlichen, inwiefern sich diese Mengenunterschiede in der positiven und negativen Klasse abzeichnen und welche Wirkungen durch die Vereinigung mehrerer Methoden entstehen.

Untersuchung:	Meinungskandidatenidentifikationen [AmazonReviews][MovieReviews]
Beschreibung:	Die Wörterbuchmethoden werden zunächst einzeln und dann im Verbund ausgeführt. Dadurch können deren Unterschiede untersucht werden.
Berechnung:	Precision, Recall, F-Measure aggregiert auf Rezensionsebene Getrennt nach negativer und positiver Klasse
Schritte:	Wörterbuchmethode 1 bis 5 5 Wörterbuchmethoden + Dependency-Methode 5 Wörterbuchmethoden + Dependency-Methode + Vertrauenswertgewichtung 3 Wörterbuchmethoden + Vertrauenswertgewichtung
Darstellung:	Die jeweiligen Werte werden in Balkendiagrammen dargestellt.



Ergebnis: Die Schaubilder zeigen hierbei deutliche Unterschiede zwischen den Methoden. Diese erreichen in der positiven und negativen Klasse sowie in den technischen Domänen und der Film-Domäne verschiedene *Precision*- und *Recall*-Werte, welche sich auf die Inhalte der Wörterbücher zurückführen lassen. Das Finanzwörterbuch beinhaltet zum Beispiel sehr wenig positive Wörter,

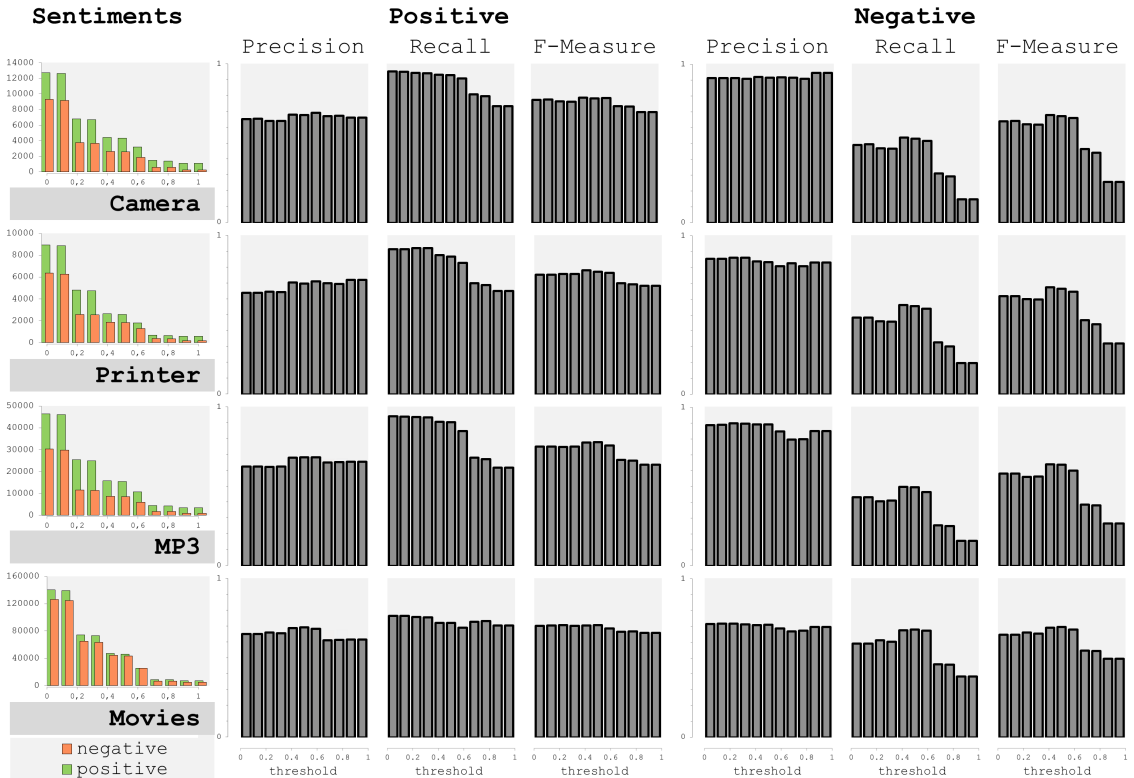
wodurch in der positiven Klasse eine hohe *Precision* mit einem niedrigen *Recall* erzielt wird. Die Betrachtung der negativen Klasse spiegelt dabei das hohe Vorkommen von negativen Wörtern innerhalb dieses Wörterbuches wider. Da die Kollektionen jeweils gleich viele positive, wie negative Rezensionen beinhalten, ergeben sich aus den Mengenunterschieden der Wörterbücher spezifische Effekte innerhalb der Klassen.

Die Summierung der Meinungsrichtungen innerhalb der Rezensionen erzeugen in der negativen Klasse niedrige *Recall*-Werte. Diese verdeutlichen, dass in als negativ markierten Rezensionen durchaus auch mehr positive Wörter vorkommen können. Für Wörterbücher, deren negativen und positiven Klassengrößen vergleichbar sind, kann dies zu einem Nachteil führen, da hierdurch ein falsches Klassifikationsergebnis erzeugt wird. Dies belegt zum Beispiel der herausstechende *Recall*-Wert des Finanzwörterbuches. Zusätzlich fehlt den Meinungskandidaten die Betrachtung von Negationen, welche häufig in der negativen Klasse Verwendung finden, wodurch abermals falsche Klassifikationsergebnisse entstehen.

Vergleicht man die Ergebnisse der technischen Domäne mit den Ergebnissen der Film-Domäne, so ergeben sich in Filmbewertungen weniger gute Resultate. Dies kann zum Beispiel an den Rezensionen liegen, welche aus einem Filmbericht und einer Benutzerbewertung bestehen. Da diese Teile für die Berechnung der Klassenzugehörigkeit zusammen betrachtet werden, wird die Bewertung der Benutzer mit einer Zusammenfassung der Filmhandlung überlagert. Hierdurch wird folglich untersucht, ob es sich um eine eher positive oder negative Handlung handelt. Dabei schneiden Wörterbücher mit ausgewogenen Klassengrößen und Domänen-unspezifischen Meinungswörtern besser ab.

Die Vereinigung der Einzelmethoden erreicht tendenziell bessere Ergebnisse. Die *Precision*- und *Recall*-Werte befinden sich im Mittelfeld zwischen den schlechtesten und besten Werten der Einzelmethoden. Dabei zeigen sie, dass sich die Vorteile und Nachteile der unterschiedlichen Wörterbücher im Zusammenschluss gegenseitig aufheben. So können zum Beispiel Uneinigkeiten erzeugt werden, indem ein Wort unterschiedliche Meinungsrichtungen zugesprochen bekommt. Zusätzlich kann ein als negativ markiertes Wort durch die Summierung der Einzelwerte mit einem positiven Wort aufgehoben werden. Dieses Phänomen kann durch die Gewichtung mit Vertrauenswerten unterdrückt werden, wodurch sich hierbei ein Mehrwert ableiten lässt. Hierdurch profitieren jeweils die *Precision*- sowie die *Recall*-Werte. Greift man zur Identifikation von Meinungskandidaten auf die Auswahl der ersten drei Wörterbuchmethoden zurück, so kann das Ergebnis hiervon ebenfalls profitieren, da weniger Unsicherheiten einbezogen werden.

Untersuchung:	Meinungskandidatenidentifikationen [AmazonReviews][MovieReviews]
Beschreibung:	Die identifizierten Meinungskandidaten werden mittels eines steigenden Schwellenwertes eingegrenzt. Das Ergebnis wird mit immer vertrauenswürdigeren Kandidaten betrachtet. Dadurch kann die Vertrauenswertberechnung untersucht werden.
Berechnung:	Precision, Recall, F-Measure aggregiert auf Rezensionsebene Getrennt nach negativer und positiver Klasse
Schritte:	5 Wörterbuchmethoden + Dependency-Methode + Vertrauenswertgewichtung Schwellenwert erhöht sich um 0.1 im Bereich zwischen 0.0 und 1.0
Darstellung:	Die jeweiligen Werte werden in Balkendiagrammen dargestellt. Ein Balkendiagramm stellt die Meinungskandidaten innerhalb des Bereichs dar.



Ergebnis: Aus der Verteilung der Identifikationen von negativen und positiven Meinungswörtern lässt sich zunächst schließen, dass in technischen Rezensionen mehr positive wie negative Meinungswörter vertreten sind. Dabei nimmt die Anzahl der negativen Meinungswörter bei steigendem Schwellenwert zwischen 0,2 und 0,6 deutlich weniger schnell ab, wie die Anzahl der positiven Meinungswörter. Dies lässt vermuten, dass es entweder stark vertrauenswürdige oder nicht vertrauenswürdige Meinungswörter mit negativer Ausrichtung gibt, wobei dieses Verhältnis für positive Meinungswörter eher angepasster ist.

Aus den rechten Schaubildern lassen sich abermals Unterschiede zwischen der positiven und der negativen Klasse aufzeigen. So nimmt der *Recall* durch Eingrenzung der Meinungskandidaten in der positiven Klasse tendenziell ab, wobei der *Recall* der negativen Klasse für Meinungskandidaten mit einem Vertrauenswert ab 0,4 zunächst steigt und ab einem Wert von 0,7 wieder abfällt. Dieses Phänomen zeigen beide Klassen in ihren *F-Measure*-Werten. Hier wechseln einige der fälschlicher-

weise als positiv berechneten Rezensionen in die negative Klasse, da sowohl die Größenverhältnisse der negativen und positiven Meinungswörter angepasst sind, als auch unsichere Entscheidungen herausgefiltert wurden. Die niedrigen *Precision*-Werte der positiven Klasse zeigen hierbei, dass nur etwa 70% der Berechnungen einer positiven Summe zu einem korrekten Ergebnis kommen. Dies bestärkt die Vermutung, dass nicht immer die als negativ markierten Rezensionen mehr negative als positive Wörter beinhalten.

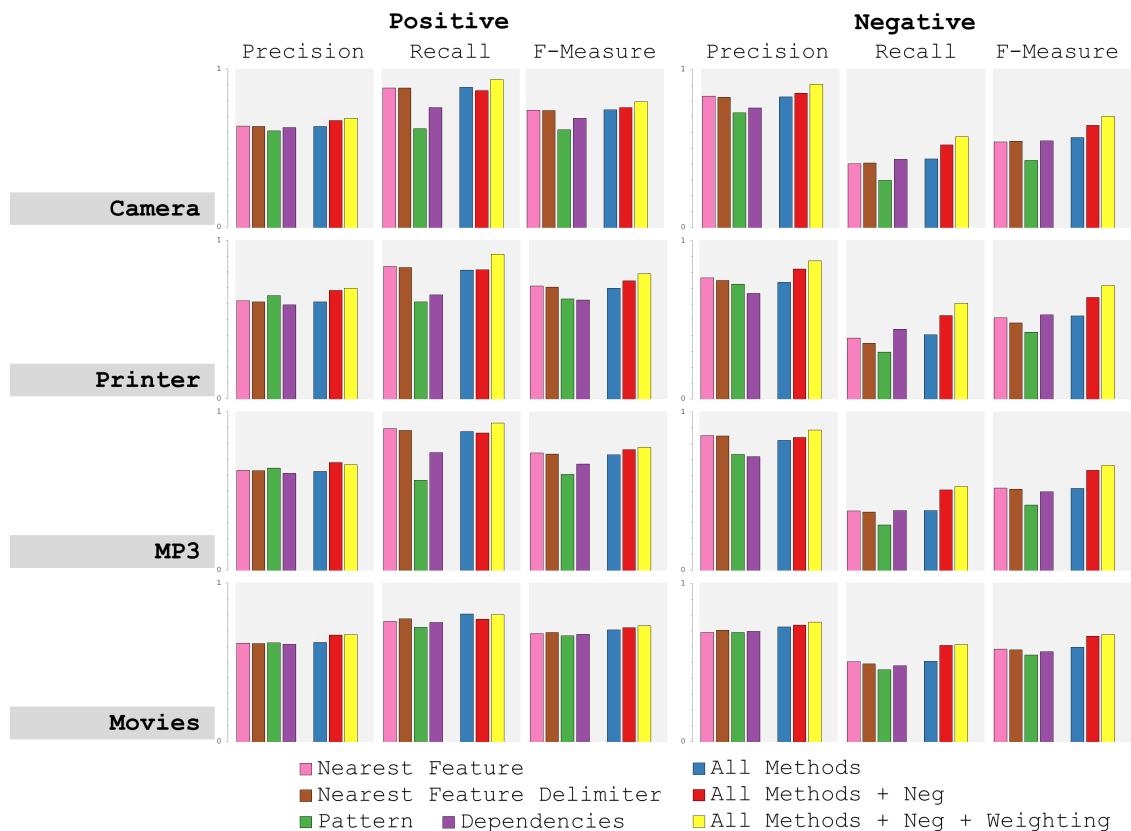
Der Unterschied zu Filmrezensionen besteht hierbei darin, dass diese keine neutralen Rezensionen besitzen und sich somit die falschen Klassifikationsergebnisse auf die einzelnen Klassen verteilen. Beziehen sich die vertrauenswürdigen Meinungen aber zum Beispiel auf den Filmbericht, so besitzen sie keine Meinungsbedeutung für den Rezensenten, werden aber dennoch in die Berechnung der Klassen einbezogen.

Schlussfolgerung: Meinungskandidaten verhalten sich in der positiven und der negativen Klasse unterschiedlich. Dabei können Rezensionen aus technischen Domänen einfacher anhand ihrer enthaltenen Meinungswörter klassifiziert werden, als Rezensionen aus der Film-Domäne, wobei sich jeweils die gleichen Phänomene in den Kollektionen zeigen. Durch die Vereinigung mehrerer Meinungsidentifikationsmethoden lässt sich ein Vertrauenswert ableiten, welcher als Gewicht eingesetzt, die *Precision*- und *Recall*-Werte erhöht. Verlässt man sich auf Meinungskandidaten mit einem Vertrauenswert ab 0,4, so stellt sich ebenfalls tendenziell ein besseres Klassifikationsergebnis ein. Dabei kann das Betrachten von Negationen sinnvoll sein, da vor allem in der negativen Klasse viele positive Meinungswörter verneint werden, was aber durch die bloße Betrachtung von Meinungskandidaten nicht abgedeckt werden kann.

5.3.1.2 Referenzen

Referenzen haben den Mehrwert, dass sie merkmalsbezogene Meinungen modellieren und Negationen einbeziehen. Somit können die Meinungswörter der Referenzen ähnlich wie Meinungskandidaten untersucht werden. Dabei kann auf vier Herangehensweisen zur Identifikation von Referenzen zurückgegriffen werden, welche unterschiedliche Betrachtungsweisen modellieren. So unterscheiden sich die einzelnen Methoden zum Beispiel anhand ihrer Spezifität und ihrer Entfernungseinschränkung zwischen Meinungs- und Produktmerkmalskandidat. Aber auch die Art der Identifikation unterscheidet sich. So werden zum einen Referenzen über Nachbarschaftseigenschaften von Meinungskandidaten identifiziert oder es wird die Umgebung eines Produktmerkmals betrachtet. Die folgende Untersuchung will nun aufzeigen, inwiefern sich das Ergebnis durch die Vereinigung mehrerer Referenzidentifikationsmethoden beeinflussen lässt.

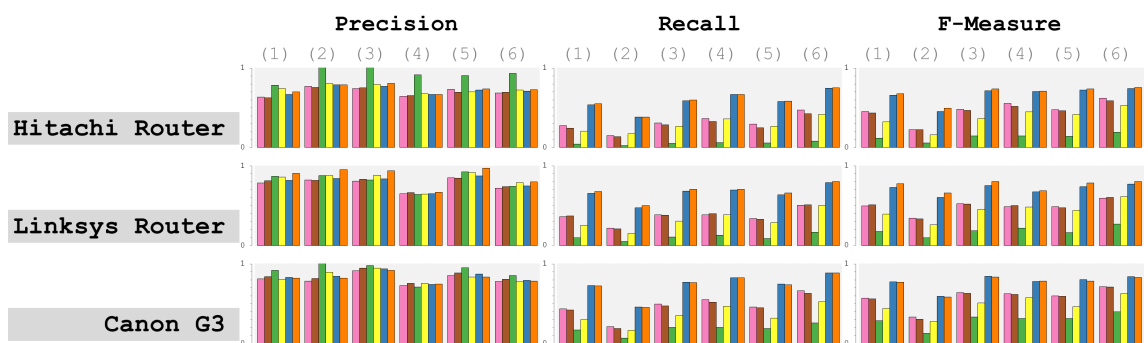
Untersuchung:	Referenzidentifikationen [AmazonReviews][MovieReviews]
Beschreibung:	Die Referenzidentifikationsmethoden werden zunächst einzeln und dann im Verbund ausgeführt. Dabei wurden vorab alle Meinungskandidatenmethoden verwendet und Produktmerkmalskandidaten identifiziert. Hierbei werden lediglich deren Meinungswörter betrachtet, wodurch Unterschiede der Referenzidentifikationsmethoden untersucht werden können.
Berechnung:	Precision, Recall, F-Measure aggregiert auf Rezensionsebene Getrennt nach negativer und positiver Klasse
Schritte:	Referenzidentifikationsmethode 1 bis 4 4 Referenzidentifikationsmethoden 4 Referenzidentifikationsmethoden + Negation 4 Referenzidentifikationsmethoden + Negation + Vertrauenswertgewichtung
Darstellung:	Die jeweiligen Werte werden in Balkendiagrammen dargestellt.



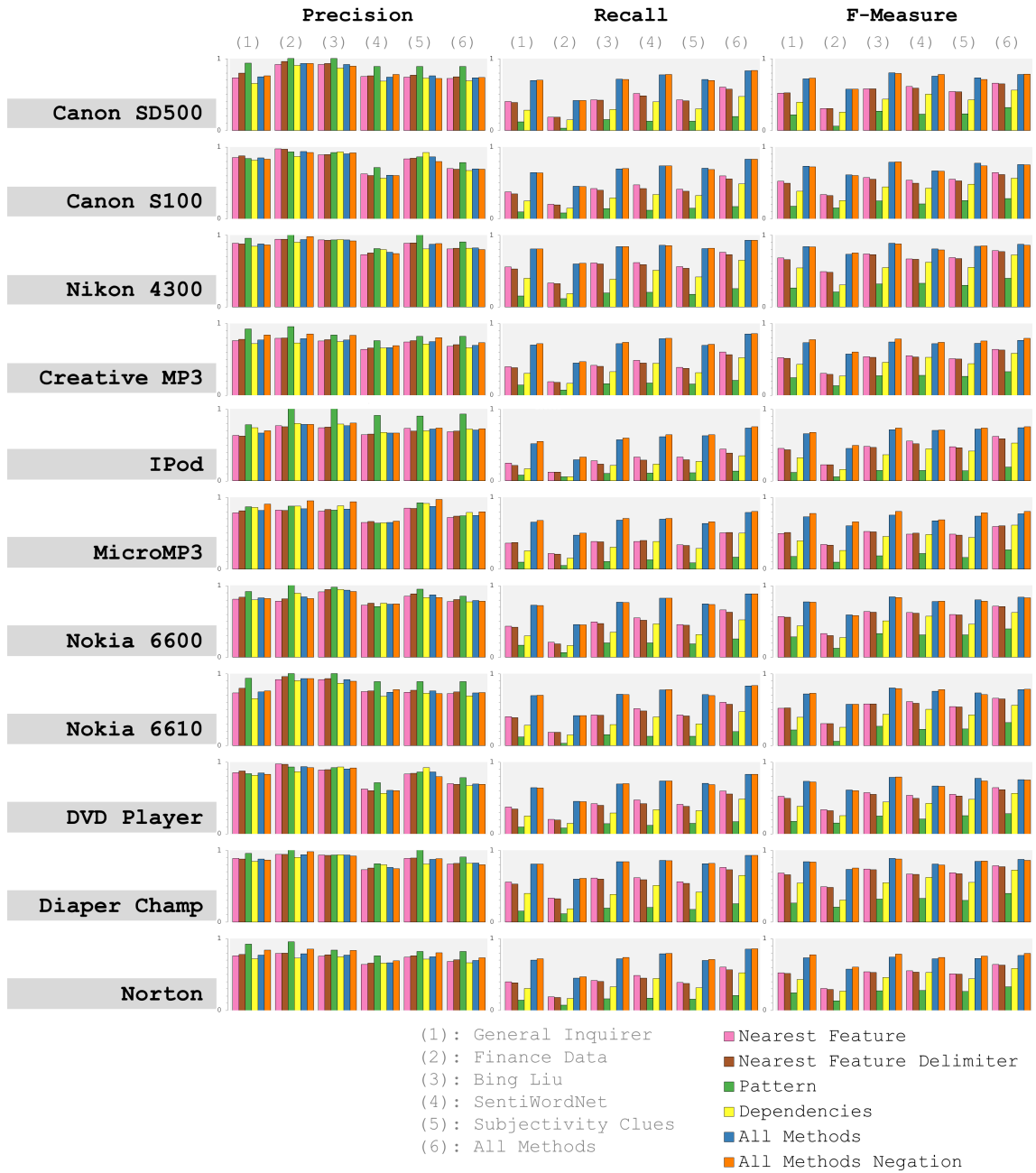
Ergebnis: Das Schaubild zeigt ähnliche Verhaltensweisen wie die vorhergehende Untersuchung der Meinungskandidaten, da die Meinungsausrichtungen der Referenzen von den Meinungskandidaten abhängig sind. Dabei beschreiben die ersten beiden Einzelmethoden recht gute Ergebnisse, da sie Produktmerkmalskandidaten in der Nähe von Meinungskandidaten betrachten und somit nahezu identische Informationen wie die vorhergehende Untersuchung modellieren. Die Methode von Kisilevich et al. [6] erzeugt hierbei das schlechteste Klassifikationsergebnis, wobei dies daran liegen kann, dass diese Methode zum Beispiel nach festgelegten Strukturen in der Nähe eines Produktmerkmalskandidaten sucht und somit nur sehr spezifische und oftmals auch objektive Referenzen identifiziert. Dadurch stehen für die Klassifikation weniger Meinungswörter zur Verfügung und das Ergebnis wird durch Missachtungen von Negationen verzerrt.

Die Vereinigung der Methoden erzeugt hier *Precision*-Werte, welche sich jeweils zwischen den Werten der besten und der schlechtesten Methode befinden. Dies lässt abermals die Vermutung zu, dass sich ähnlich zur Untersuchung der Meinungskandidaten, korrekte Entscheidungen durch falsche Entscheidungen aufheben. Dabei können die *Precision*-Werte durch den Einbezug von Negationen gesteigert werden. Diejenigen Rezensionen, die nun fälschlicherweise als positiv eingestuft wurden, fallen durch Negationen nun in die korrekte Klasse. Dies wirkt sich nur leicht auf den *Recall* der positiven Klasse aus, da nur sehr wenige Rezensionen hinzukommen, die falsch oder neutral eingestuft werden. Das Klassifikationsergebnis lässt sich noch zusätzlich durch eine Gewichtung mit den Vertrauenswerten der Meinungswörter bestärken. Folglich kann durch die Vereinigung der Methoden und mittels der Vertrauenswertgewichtung ein tendenziell besseres Ergebnis erzeugt werden.

Untersuchung:	Referenzidentifikationen [BingLiuReviews]
Beschreibung:	Die Referenzidentifikationsmethoden werden zunächst einzeln und dann im Verbund ausgeführt. Dabei wurden unterschiedliche Kombinationen von Meinungskandidatenmethoden berechnet. Hierdurch können Unterschiede zwischen Referenzidentifikationsmethode(n) und Meinungsidentifikationsmethoden(n) untersucht werden.
Berechnung:	Precision, Recall, F-Measure
Schritte:	Referenzidentifikationsmethode 1 bis 4 x Meinungsidentifikationsmethode 1 bis 5 x 4 Referenzidentifikationsmethoden + Negation + Vertrauenswertgewichtung x 6 Meinungsidentifikationsmethoden
Darstellung:	Die jeweiligen Werte werden in Balkendiagrammen dargestellt.



Fortsetzung auf nächster Seite



Ergebnis: Die einzelnen Referenzmethoden zeigen unterschiedliche Verhaltensweisen. So erreicht die *Pattern*-Methode sehr gute *Precision*-Werte, welche aber auf den verschiedenen Datensätzen variieren. In Verbindung mit dem Finanzwörterbuch besitzt diese Methode in den meisten Fällen die besten *Precision*-Werte. Diese Methode liefert folglich recht wenige Ergebnisse, die aber in Folge einer großen Einschränkung recht genau sind. Als ungenaues Wörterbuch erweist sich *SentiWordNet*, welches in der Kombination mit allen Methoden die niedrigsten *Precision*-Werte erzeugt. Dabei ist aber deren *Recall*-Wert höher als bei anderen Wörterbüchern. Dies lässt darauf schließen, dass bedingt durch die Größe dieses Wörterbuches zwar recht viele, aber dafür unsichere Entscheidungen mit geringer Subjektivität getroffen werden. Hierbei zeigt sich das Wörterbuch

der Autoren Bing und Liu als unauffällig, obwohl es auf diesen Daten generiert wurde. Diejenigen Methoden, welche Referenzen in der Umgebung von Meinungskandidaten betrachten, erreichen höhere *Recall*-Werte, finden aber dadurch auch falsche Referenzen. Die *Dependency*-Methode zeigt sich als eher ungeeignet, da sie die niedrigsten *Precision*- und *Recall*-Werte erzeugt.

Die Vereinigung der Referenzmethoden bestätigt abermals das Phänomen, dass sich hierdurch korrekte Entscheidungen mit inkorrekten mischen können, wodurch sich die *Precision*-Werte wieder im Mittelfeld befinden. Für die Vereinigung der Meinungsidentifikationsmethoden gilt dies genauso, da dort die Menge der Meinungswörter viele falsche Referenzen erzwingen. Dabei können aber mehr Annotationen abgedeckt werden, wodurch die *Recall*-Werte steigen können. Dies unterstreichen die *F-Measure*-Werte, da sich durch die Kombinationen von allen Methoden ein besseres Ergebnis ergibt. Dies wird aber besonders durch den hohen *Recall* bewirkt.

Negationen können hierbei in den meisten Fällen zu einer Verbesserung beitragen, wobei dieser Aspekt von den zugrundeliegenden Daten und den identifizierten Meinungskandidaten abhängig ist. Auch kann hier speziell durch die gewählte Evaluationspezifikation nicht exakt gesagt werden, dass die Referenzen und die eventuell dahinter liegenden Negationen korrekt sind.

Schlussfolgerung: Referenzen hängen von der Berechnung der Meinungskandidaten ab, besitzen aber den Mehrwert, dass hierdurch Negationen betrachtet werden können. Dabei liefern Methoden, welche die Identifikation zu sehr einschränken zwar genauere Ergebnisse, betrachten hierdurch aber nur einen geringen Teil der Daten. Die Evaluation hat gezeigt, dass die Vereinigung mehrerer Referenzidentifikationsmethoden sich hauptsächlich in einem höheren *Recall*-Wert abzeichnet, wobei sich der *Precision*-Wert im Mittelfeld befindet, da sich unterschiedliche Entscheidungen gegenseitig aufheben können. Negationen beeinflussen diese Werte hin zu einem tendenziell besseren Ergebnis, wobei die Evaluation hier nicht zeigen kann, ob diese Negationswörter auch korrekt sind.

5.3.1.3 Hypothese

Die vorhergehenden Untersuchungen konnten zeigen, dass die Vereinigung von unterschiedlichen Methoden den Abdeckungsgrad der Analyseanwendung vergrößern kann. Dabei können die Unterschiede der Einzelmethoden in verschiedenen Domänen vereint werden. Zusätzlich konnte gezeigt werden, dass die Modellierung von Unsicherheiten sinnvoll ist, besonders wenn diese in einer Klassifikation als Gewicht eingesetzt werden.

5.3.2 Interaktive Steuerung

Die Interaktive Steuerung kann unterschiedliches leisten, wobei hier mit wenigen Änderungen viel bewirkt werden soll. Hierzu kann der Analyst auf verschiedene Informationen aus der Anwendung zurückgreifen. Zum Beispiel werden ihm automatisch generierte Vorschläge vorgelegt, die er für seine Korrekturen verwenden kann. Zusätzlich können Identifikationen anhand ihrer Vertrauenswerte priorisiert werden, wodurch zum Beispiel größere Unsicherheiten zuerst betrachtet werden können. Auch kann er die Sätze auf gewohnte Weise kontrollieren und hieraus eigene Korrekturen entwickeln. Diese unterschiedlichen Betrachtungsweisen sollen nun Gegenstand einer Evaluation sein und die Chancen der Interaktiven Steuerung beleuchten. Dabei besteht die Ausgangslage aus den Identifikationen des Analyseprozesses. Der Analyst³⁷ wird diese Informationen korrigieren, wobei er anhand bestimmter Konfigurationen, welche in Abbildung 5.16 dargestellt sind, eingeschränkt ist.

	Occurrences	Priorisations	Suggestions	Sentence Analysis
Setting 1	×			
Setting 2	×	×		
Setting 3	×	×	×	
Setting 4			×	×

Abbildung 5.16: Die Tabelle gibt an, auf welche unterschiedlichen Informationen der Analyst bei seinen Korrekturen zugreifen kann. Darunter befinden sich zum Beispiel die Häufigkeiten der Terme, die Priorisierung anhand der Vertrauenswerte, automatisch generierte Vorschläge und die Analyse der einzelnen Sätze.

Durch diese Einschränkungen muss sich der Analyst an unterschiedlichen Strategien bedienen, um Änderungen am Analyseprozess vornehmen zu können. Diese werden in nachfolgender Auflistung erläutert.

- **Setting 1:** Da nur Häufigkeiten betrachtet werden können, muss sich der Analyst für eine absteigende Sortierung der jeweiligen Identifikationen entscheiden. Dadurch können die häufigsten Fehler, welcher er als solche erkennt, korrigiert werden.
- **Setting 2:** Da nun priorisierte Informationen vorliegen, können die Identifikationen nach Vertrauenswert sortiert werden. Dabei kann sich der Analyst nun entscheiden, entweder häu-

³⁷Hierbei ist der Autor dieser Arbeit als Analyst tätig

fige unsichere, oder häufige sichere Entscheidungen zu korrigieren. Dies hängt primär von den Objekten ab, worauf die Korrektur stattfinden soll.

- **Setting 3:** Diese Strategie entspricht der vorhergehenden Strategie, wobei die Änderungen des Analysten über automatisch generierte Vorschläge erweitert werden. Somit können mehrere ähnliche Entscheidungen beeinflusst werden, wodurch sich schneller ein besseres Ergebnis ergeben kann.
- **Setting 4:** Hierbei werden die Informationen direkt auf den Textdaten analysiert. Hierdurch können Änderungen und Verbesserungen, wie in den vorherigen Konfigurationen erfolgen. Es können sowohl unsichere, wie sichere Entscheidungen identifiziert werden. Durch die Analyse des Satzes können sich aber auch noch nicht identifizierte Funde ergeben, welche mittels einer positiven Korrektur in den Analyseprozess übertragen werden können.

Um die Untersuchungen einheitlich darstellen zu können, wird der Analyst pro Konfiguration auf eine bestimmte Anzahl an Änderungen³⁸ beschränkt. Dabei werden die verschiedenen Identifikationsmodule getrennt korrigiert und betrachtet. Für die Evaluation stehen primär die Korrekturen von Meinungskandidaten, Produktmerkmalen und Referenzen im Vordergrund, daher wird die Korrektur der Wortartentscheidungen nur am Rande betrachtet.

Hypothese: Die Untersuchungen der Interaktiven Steuerung sollen zeigen, dass der Analyst in seiner Korrekturarbeit durch die Berechnungen von Vertrauenswerten und automatisch generierten Vorschlägen unterstützt wird. Hierdurch kann schließlich eine effektive und messbare Verbesserung des Ergebnisses erreicht werden.

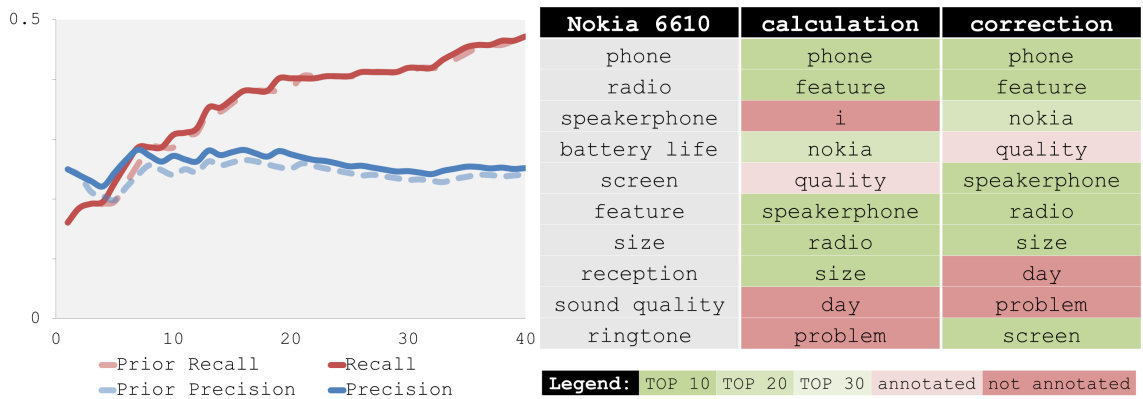
5.3.2.1 Wortarten

Die Untersuchung, inwiefern die Interaktive Steuerung für die Verbesserung von Wortarten geeignet ist, soll in dieser Arbeit nicht im Vordergrund stehen, weshalb hier nur ein kleiner Einblick gewährt wird. Wortartkorrekturen machen zunächst deswegen Sinn, da sich in Rezensionen nicht an Rechtschreibregeln gehalten wird oder Eigennamen und technische Begriffe im Wortschatz der Wortarterkennung nicht enthalten sind. Die Algorithmen zur Erkennung der Wortarten können anhand ihrer Trainingsdaten somit für unbekannte Terme keine eindeutigen Entscheidungen treffen. Diese Arbeit bindet hierbei drei unterschiedliche Algorithmen ein. Dabei müssen diejenigen Entscheidungen korrigiert werden, welche entweder aus unterschiedlichen Wortartentscheidungen bestehen, oder den *PennTreebank-Tag* „FW“, welcher für unbekannte Wortarten vergeben wird,

³⁸Eine Änderung besteht hierbei aus dem Erkennen eines Fehlers, dem Öffnen des Änderungsdialoges, dem Auswählen der Änderung und der eventuellen Überprüfung von automatisch generierten Vorschlägen

enthalten. Die größte Unsicherheit wird dadurch modelliert, wenn die Entscheidungen aller drei Algorithmen unterschiedlich ausfallen, wobei Vortests ergaben, dass dies nur in etwa 1 % der Fälle eintritt. Werden diese Unsicherheiten korrigiert, können davon andere Methoden, welche an diesen Wortartinformationen anknüpfen, profitieren.

Untersuchung:	Wortartkorrekturen [BingLiuReviews]
Beschreibung:	Die Produktmerkmalsidentifikation wird einmal ohne und einmal mit Benutzerverbesserung von 10 Wortartfehlern ausgeführt. Dabei werden die Kandidaten mittels eines steigenden Schwellenwertes eingegrenzt. Das Ergebnis wird somit mit immer mehr Kandidaten eines höheren Ranges betrachtet. Hierdurch kann die Wirkung der Benutzerintervention auf Produktmerkmalskandidaten untersucht werden.
Berechnung:	Precision, Recall
Schritte:	Produktmerkmalskandidaten ohne Korrektur Produktmerkmalskandidaten mit Korrektur
Darstellung:	Liniendiagramme stellen die berechneten Werte dar. Eine Tabelle stellt die ersten 10 Ränge der Annotationsdaten mit den ersten 10 Ränge der Produktmerkmalsidentifikation gegenüber.



Ergebnis: Die Entscheidungen der Wortarterkennung wirkt sich insbesondere auf die Identifikation von Produktmerkmalskandidaten aus. So können hier Fehler, welche zum Beispiel durch das Wort „I“ hervorgerufen werden, ausgeschlossen werden. Dadurch, dass sich dieses Wort nun nicht mehr in der Rangliste befindet, verbessert sich das Ergebnis der ersten 40 Produktmerkmalskandidaten im geringen Maße. Die Korrekturen umfassen meistens Produkt- oder Eigennamen³⁹, welche mit automatischen Vorschlägen angereichert werden können. Diese umfassen alle gleich geschriebenen Terme in der Dokumentenkollektion. In dieser Untersuchung konnten somit mit 10 Wortartkorrekturen 46 unterschiedliche Terme erfasst werden. Dabei betreffen Wortartkorrekturen nur dann die Produktmerkmalsidentifikation, wenn keine der Methoden diesen Term als Nomen interpretiert. Dies ist aber nur in 8 von 46 Termänderungen der Fall, wodurch die geringe Verbesserung erklärt werden kann. Die anderen Änderungsfälle betreffen lediglich die Rangfolge der Produktmerkmalskandidaten.

³⁹10 Wortartkorrekturen: *tri-band, t-zones, t610, pc, sony, bluetooth, internet, samsung, i, nokia*

Schlussfolgerung: Da annähernd alle Methoden auf Wortartinformationen aufbauen, kann es durchaus Sinn machen diese Fehler bereits zu Anfang zu korrigieren. Dies kann auf einem kleinen Teil der Daten geschehen, wodurch sich das Ergebnis gering verbessern lässt. Dadurch kann zum Beispiel die Menge der Produktmerkmalskandidaten vergrößert und Fehler der Wortarterkennung eliminiert werden. Wobei sich besonders die Korrektur von häufigen Nomen auf die Produktmerkmalsidentifikation auswirkt. Dabei kann der Analyst Fehler der Wortarterkennung durch eine Vertrauenswertberechnungen leicht erkennen und gleiche Terme durch automatisch generierten Vorschläge auf einmal korrigieren.

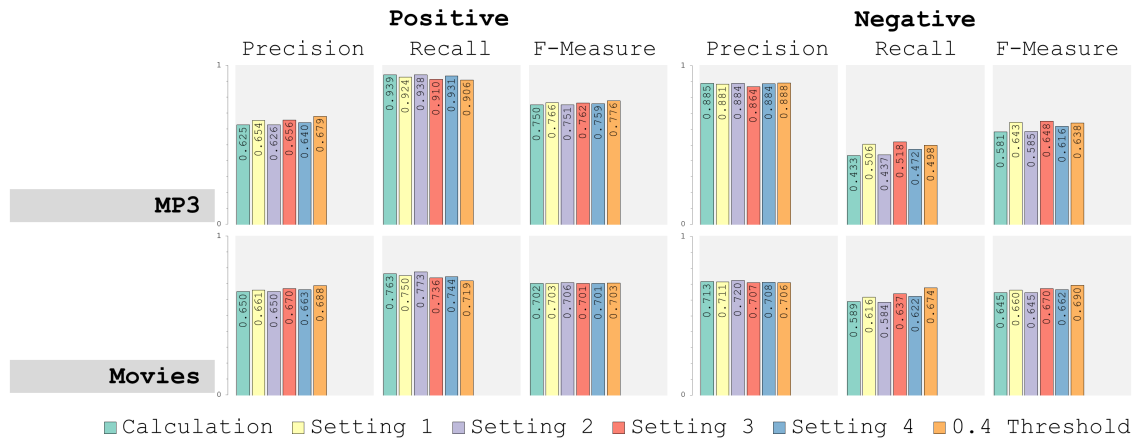
5.3.2.2 Meinungskandidaten

Meinungen beschreiben den primären Untersuchungsgegenstand innerhalb der Meinungsanalyse. Meinungswörter signalisieren folglich subjektive Eindrücke der Rezensenten und werden in dieser Arbeit primär durch Meinungswörterbücher identifiziert. Eine Korrektur dieser Informationen kann hierbei verschiedene Gründe haben. So kann zum Beispiel ein Wort in einer bestimmten Domäne keine oder eine zum Wörterbuch unterschiedliche meinungstragende Bedeutung besitzen. Dies kann durch eine einfache Änderung der Meinungsausrichtung korrigiert werden. Die vorhergehende Untersuchung ergab bereits, dass Meinungskandidaten mit geringem Vertrauenswert das Ergebnis negativ beeinflussen können. Werden diese vom Analysten korrigiert, kann sich hierdurch ein besseres Ergebnis einstellen.

Untersuchung:	Meinungskandidatenkorrekturen [AmazonReviews] [MovieReviews]
Beschreibung:	Die Meinungskandidatenidentifikationen werden mittels verschiedener Interaktionskonfigurationen korrigiert. Hierdurch kann die Wirkung dieser Benutzerinterventionen untersucht werden.
Berechnung:	Precision, Recall, F-Measure aggregiert auf Rezensionsebene Getrennt nach negativer und positiver Klasse
Schritte:	5 Wörterbuchmethoden + Dependency-Methode + Vertrauenswertgewichtung Ergebnis mit 10 Korrekturen von Setting 1 bis Setting 4 Ergebnis mit Meinungskandidaten mit Vertrauenswert ≥ 0.4
Darstellung:	Die jeweiligen Werte werden in Balkendiagrammen dargestellt.



Fortsetzung auf nächster Seite



Ergebnis: Die verschiedenen Korrekturen zeigen lediglich geringe Effekte auf das Ergebnis. Für technische Artikel sind sie weit kräftiger als für Artikel aus der Filmdomäne, was dadurch begründet werden kann, dass Filmrezensionen für einen in dieser Evaluation eingesetzten Klassifikator weniger geeignet sind. In der positiven Klasse kann jede Konfiguration eine geringe Steigerung der *Precision*-Werte erreichen. Dadurch fällt der *Recall* in nahezu allen Fällen ab, bleibt aber über den Werten der Vertrauenswertfilterung. Dabei kann festgehalten werden, dass bestimmte Konfigurationen zu einer Steigerung der *F-Measure*-Werten beitragen, wobei nicht immer die gleiche Konfiguration am besten ist.

In der negativen Klasse verhält sich das Ergebnis gerade umgekehrt. Dort führen Korrekturen primär zu einer Verbesserung des *Recalls* und übertreffen in der technischen Domäne selbst die Werte der Vertrauenswertfilterung. Dabei stellt sich *Konfiguration 3* als eine der besten Änderungsstrategien heraus. Dies kann zum Beispiel daran liegen, dass häufige unsichere Meinungswörter mit positiver Ausrichtung korrigiert werden, wodurch sich nun weniger als falsch eingestufte Rezensionen ergeben. Dabei können die Änderungen aus *Konfiguration 3* durch automatische Vorschläge angereichert werden. In der Kameradomäne konnten hierdurch zum Beispiel 46 Entscheidungen korrigiert werden.

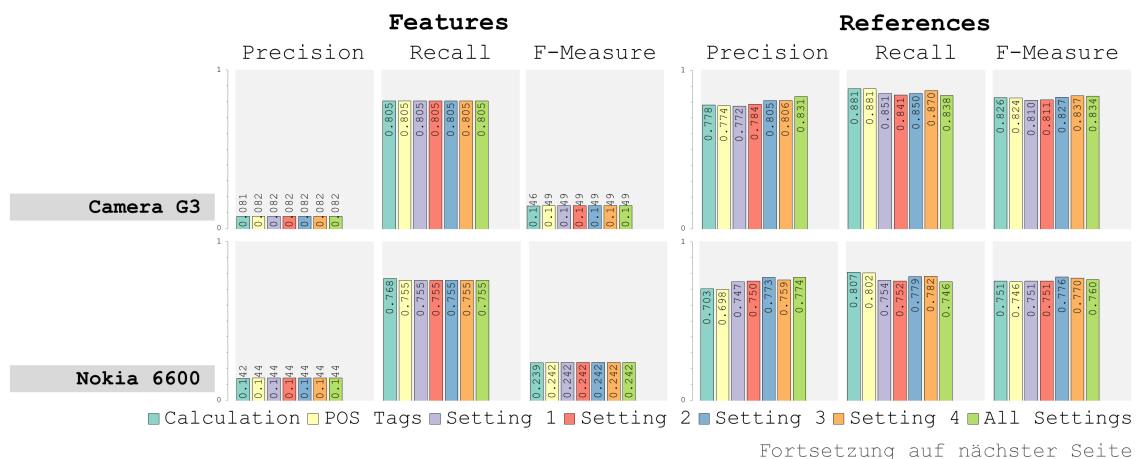
Betrachtet man die Unterschiede zwischen *Konfiguration 1* und *Konfiguration 3*, so lässt sich nicht eindeutig darauf schließen, welche Strategie für Meinungskandidaten geeigneter ist. Dabei können sowohl relativ häufige Meinungswörter ohne Vorschläge korrigiert werden, wie relativ unsichere Meinungswörter mit Vorschlägen. Beide Strategien sind nahezu gleichwertig. Bei *Konfiguration 4* können sich unterschiedliche Effekte auf das Ergebnis einstellen. Dadurch, dass der Analyst keine Informationen zur Häufigkeit der Fehlerfälle hat, werden hierbei auch weniger häufige Kandidaten verbessert. Der Vorteil dieser Strategie besteht darin, dass der Analyst Fehler anhand des Kontextes des Satzes ableiten kann. Hierdurch können auch unbekannte Wortschöpfungen⁴⁰ entdeckt und mit

⁴⁰water-proof, ladylike

einer meinungstragenden Bedeutung versehen werden.

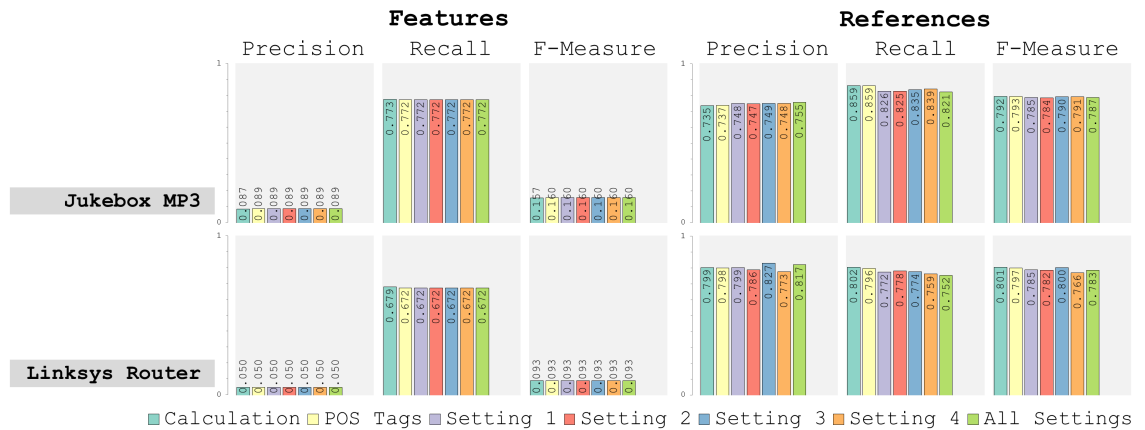
Insgesamt werden in unterschiedlichen Konfigurationen und unterschiedlichen Domänen häufig die gleichen Korrekturen vorgenommen⁴¹, wobei das Wörterbuch *SentiWordNet* die meisten unsicheren Entscheidungen verursacht. Dies spricht zum einen für eine gute Übertragbarkeit der Korrekturen und zum anderen für ein anfälliges Wörterbuch. Dabei können manche Korrekturen aber auch negative Effekte auf das Evaluationsergebnis haben. So werden zum Beispiel Änderungen an Meinungswörtern vorgenommen, welche zwar keine meinungstragende Bedeutung besitzen, aber für einen Klassifikator, wie er in dieser Evaluation eingesetzt wird, sehr ausschlaggebend sein können. Dies trifft zum Beispiel auf die Wörter „not“ und „problem“ zu, welche recht häufig in der negativen Klasse auftauchen. Werden diese dann als neutral eingestuft, sinkt der Genauigkeitswert der negativen Klasse, obwohl es sich hierbei um eine korrekte Änderung handelt. Es müssen folglich recht viele Korrekturen vorgenommen werden, bis sich ein annehmbares Ergebnis einstellen kann. Generell kann gesagt werden, dass sich die Verbesserungen der Meinungskandidaten nicht wie erwartet im Ergebnis widerspiegeln. Dabei liegt aber die Vermutung nahe, dass die positiven Effekte der Interaktiven Steuerung durch die gewählte Evaluationstechnik geschmälert werden.

Untersuchung:	Meinungskandidatenkorrekturen [BingLiuReviews]
Beschreibung:	Die Meinungskandidatenidentifikationen werden mittels verschiedener Interaktionskonfigurationen korrigiert. Dabei soll die Wirkung dieser Benutzerinterventionen auf Produktmerkmalskandidaten und Referenzen untersucht werden.
Berechnung:	Precision, Recall, F-Measure
Schritte:	Ergebnis ohne Korrekturen Ergebnis mit 10 Wortartkorrekturen Ergebnis mit 10 Wortartkorrekturen + Korrekturen von Setting 1 bis Setting 4 Ergebnis mit 10 Wortartkorrekturen + allen Korrekturen aus 4 Settings
Darstellung:	Die jeweiligen Werte werden in Balkendiagrammen dargestellt.



Fortsetzung auf nächster Seite

⁴¹buy, bought, even, as, quality, problem, digital, not



Ergebnis: Eine Änderung der Wortarten spiegelt sich nur gering im Ergebnis wieder. Die *Precision* kann nur steigen, wenn entweder fälschlicherweise als Nomen identifizierte Terme eliminiert⁴² oder nicht als Nomen identifizierte Terme⁴³ hinzugefügt werden. Hier wurden aber primär Unsicherheiten für Nomen behoben, wobei in vielen Fällen bereits mindestens eine Methode den korrigierten Term als Nomen interpretiert hat. Diese Korrekturen können folglich die Produktmerkmalsidentifikation nur im geringen Maße betreffen. Betrachtet man diejenigen Termänderungen, welche unbekannte Nomen hinzufügen, so kann sich dies, bedingt durch die Annotationsweise, auch negativ auf die Ergebnisse der Produktmerkmalsidentifikation auswirken⁴⁴.

Die gewählte Evaluationstechnik kann deswegen keine hohen *Precision*-Werte erreichen, da die Kandidaten bisher noch keinen Meinungsbezug besitzen, aber mit solchen aus den Annotationsdaten verglichen werden.

Die Korrektur von Wortarten zeigen im Schaubild tendenziell negative Effekte auf die Referenzidentifikation. Dies liegt aber primär an der Evaluationspezifikation. So werden zum Beispiel Referenzen, welche sich auf nicht annotierte Merkmale beziehen, nicht betrachtet⁴⁵. Kommen hierbei korrekte Produktmerkmalskandidaten hinzu, betrachtet die *Precision* aber lediglich, ob der Meinungskandidat eine korrekte Ausrichtung zur Annotation besitzt. Das Sinken der *Recall*-Werte ist auf dieselben Gründe zurückzuführen, die auch die fallenden *Recall*-Werte der Produktmerkmalskandidaten bewirken.

Eine Korrektur von Meinungskandidaten kann sich hierbei nicht auf Produktmerkmalskandidaten auswirken, da erst durch die Referenz ein Bezug zwischen ihnen hergestellt wird. Diese bewirken hierbei eine geringe Verbesserungen auf der Seite der *Precision*-Werte. Die Verbesserung bewirkt, dass weniger unsichere Referenzen generiert werden. Die verschiedenen Konfigurationen haben hierbei unterschiedliche Auswirkungen. So variieren die Verbesserungen durch *Konfiguration 1* in

⁴²Eliminierung des Terms „I“, da es sich hierbei um ein Personalpronomen handelt

⁴³Die Terme „samsung“ oder „pc“ wurden in wenigen Fällen als Adjektiv identifiziert

⁴⁴Es wurde zum Beispiel der Term „phone“ annotiert, wobei die Produktmerkmalsidentifikation, bedingt durch die Wortartkorrektur, nun „samsung phone“ identifiziert hat

⁴⁵Zum Beispiel diejenigen Referenzen, welche auf das „i“-Phänomen zurückzuführen sind

verschiedenen Dokumentensammlungen. Dabei ergibt sich für *Konfiguration 3* den größten Zuwachs der *Precision*, da hier die meisten unsicheren Entscheidungen korrigiert werden. Dies kann somit auf die positiven Effekte der Interaktiven Steuerung zurückzuführen sein, wobei diese Vermutung hierbei durch die Hinzunahme mehrerer Korrekturen unterstützt wird.

Durch den Wegfall von Meinungskandidaten werden hierbei weniger Referenzen generiert, wodurch zum Beispiel nicht mehr alle annotierten Merkmale abgedeckt werden können, was den sinkenden *Recall* erklärt.

Die Evaluation spiegelt abermals die Vermutung der vorhergehenden Untersuchung wider. Dabei werden aber nur diejenigen Referenzen betrachtet, welche auch auf ein annotiertes Merkmal weisen. Dabei ist aber nicht gegeben, dass eine Korrektur auch die an der Evaluation beteiligten Referenzen trifft. Die Evaluation zeigt hier folglich nur Auswirkungen auf bestimmte Referenzen, welche auch auf Seite der Annotatoren als merkmalsbezogene Meinungen definiert sind.

Dabei zeigt aber die *Recall*-Seite der Produktmerkmale, dass nur etwa 75% der Produktmerkmal-kandidaten auch wirklich mit den Annotationsdaten übereinstimmen. Ein Vortest ergab hier, dass dies an der Annotationsweise und an der Berechnung der Kandidaten liegt. Dabei sind durch die Annotatoren nur die für sie sinnige Teile der im Text vorkommenden Produktmerkmale erfasst worden. Diejenigen Nomenketten, welche durch das in dieser Arbeit verwendete Verfahren identifiziert wurden, können durch den strikten Vergleich nicht mehr erfasst werden⁴⁶. Zusätzlich wurden nur Meinungsrichtungen der annotierten Merkmale festgelegt, wodurch der genaue Bezug zu einem Meinungskandidaten in der Evaluation nicht exakt betrachtet werden kann.

Schlussfolgerung: Eine Verbesserung des Ergebnisses durch die Korrekturen der Meinungskandidaten kann durch die gewählte Evaluationstechnik nicht eindeutig nachgewiesen werden, was auf eine Schwäche dieser Evaluationstechnik hinweist, wobei sich dies im Nachhinein herausstellte. Dabei zeigte die Vertrauenswertfilterung aber, dass sich durch das Eliminieren von Unsicherheiten durchaus bessere Ergebnisse erzielen lassen. Die Frage, wie viele Korrekturen vorgenommen werden müssten, kann bedingt durch den Umfang dieser Arbeit nicht betrachtet werden. Die Vermutung liegt aber nahe, dass sich mit 10 Korrekturen weniger Verbesserungen einstellen, als dies mit 40 Korrekturen ermöglicht wird. Es konnte aber auf Seite der *Precision* festgestellt werden, dass man auch aus wenigen Korrekturen profitieren kann, wenn diese mit Hilfe von automatisch generierten Vorschlägen erweitert werden.

⁴⁶Dabei werden Nomenketten berechnet, die die Merkmale genauer beschreiben, wie zum Beispiel „*samsung's phone*“, wobei aber nur „*phone*“ annotiert wurde

Ergebnis: Auf Seite der Produktmerkmalskandidaten sollen zunächst die *Precision*-Werte erörtert werden. Diese repräsentieren die Genauigkeit zwischen der Berechnung und den Annotationsdaten, wodurch keine hohen Werte erwartet werden können. Da nicht jedes Vorkommen eines Produktmerkmalskandidaten auch einen Bezug zu einer Meinung aufweist, wurde es folglich nicht annotiert und erzwingt somit niedrige *Precision*-Werte. Dabei stellen sich durch 10 Korrekturen nur geringe Verbesserungen im Bezug auf das Gesamtergebnis ein. Hierbei bewirken *Konfiguration 1* und *Konfiguration 2* die größte Verbesserung, da hierdurch die häufigsten Merkmale verbessert werden. Dabei beschränkte sich der Analyst lediglich auf das Entfernen von bestimmten Merkmalen, welche zwar Nomen sind, aber keine Merkmalsbezüge vorweisen⁴⁷. Dabei können durch automatisch berechnete Vorschläge seltene, aber trotzdem annotierte Merkmale fälschlicherweise eliminiert werden, wodurch die *Recall*-Werte gesenkt werden.

Um die positiven Effekte der Interaktiven Steuerung aufzeigen zu können, wurden so viele Korrekturen ausgeführt, bis nur noch diejenigen Merkmale im Analyseprozess beachtet werden, welche mindestens einmal annotiert wurden. Dies entspricht auf Seite der Evaluation folglich einer perfekten Korrektur. Dabei kann diese Menge entweder über eine positive oder eine negative Korrektur erfolgen. Negative Korrekturen ermöglichen es, Merkmale zu eliminieren, wobei positive Korrekturen Merkmale hinzufügen. In der Kameradomäne können für die Realisierung dieses Ergebnisses entweder 859 negative Korrekturen, oder 63 positive Korrekturen durchgeführt werden. Diese verlangt somit deutlich weniger Aufwand. Im Vergleich zu den 10 Korrekturen des Analysten, stellt sich ein deutlich besseres Ergebnis ein. Dies lässt die Vermutung zu, dass durch mehr Korrekturaufwand ein eindeutigeres Ergebnis erzielt werden kann.

Dabei erreichen die *Recall*-Werte nicht ihr Maximum. Dies kann dadurch begründet werden, dass wenn zwei als Merkmal erfasste Terme nebeneinander stehen, diese im Analyseprozess als eines betrachtet werden⁴⁸. Da aber in diesen Fällen nur ein Teilwort annotiert wurde, wird dies bedingt durch die strikte Evaluationspezifikation als falsch markiert.

Betrachtet man die Effekte der Korrekturen auf Seite der Referenzen, so weckt dies den Anschein einer Verschlechterung. Dies ist aber auf die Evaluationspezifikation zurückzuführen. Die vom Analyst getätigten Korrekturen umfassen den Teil der Produktmerkmalskandidaten, welche ohnehin nicht annotiert wurden und somit für die Evaluation der Referenzen nicht betrachtet werden. Dabei kann sich nun der Effekt einstellen, dass ein alleinstehender Meinungskandidat zum Beispiel einem anderen Produktmerkmalskandidaten in nächster Nähe zugeordnet wird. Befindet sich dieser zufällig in der Liste der annotierten Merkmale, so wird diese Referenz nun in der Evaluation mitbetrachtet. Dabei zeigt die *Precision*, dass es sich hierbei um eine tendenziell falsche Zuord-

⁴⁷ monday, time, year, birthday, parents, review, something, thing

⁴⁸ stitch pictures feature, image software, zoom picture, canon software

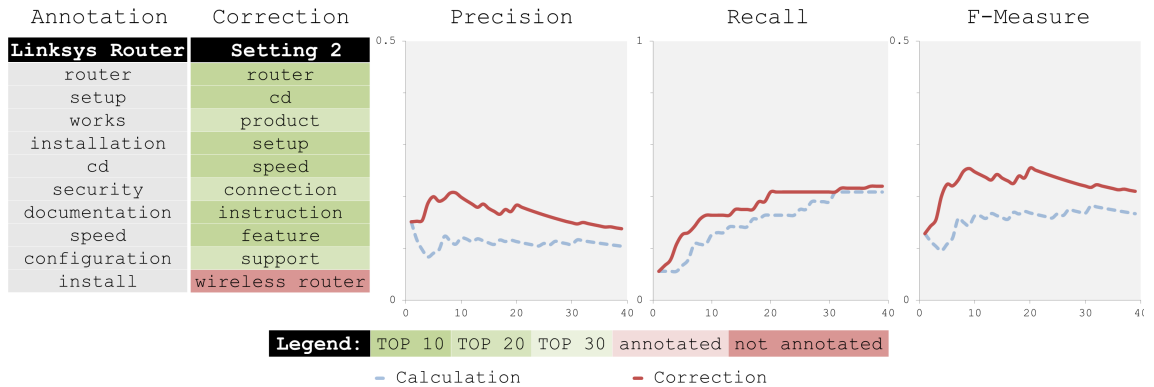
nung handelt. Der *Recall* kann dadurch steigen, da sich neue Referenzen bilden. Die Korrekturen der Produktmerkmale kann folglich durch die *Precision* nicht ausgedrückt werden, da dort die Genauigkeiten der Meinungskandidaten beschrieben werden.

Aus der Untersuchung lässt sich nun ableiten, dass eine Korrektur von Produktmerkmalen sinnvoll ist, da viele falsche Nomenketten identifiziert werden. Dabei müssen aber weit mehr als nur 10 Korrekturen vorgenommen werden, da viele Produktmerkmalskandidaten mit niedrigerem Rang das Ergebnis verfälschen. Dabei zeigt sich, dass eine perfekte Änderung den *Recall* der Referenzen durchaus erhöhen kann. Die verwendete Evaluationstechnik zeigt aber eher Effekte der Meinungskandidaten und kann somit die Korrekturen von Produktmerkmalen nicht widerspiegeln.

Untersuchung:	Produktmerkmalskandidatenkorrekturen [BingLiuReviews]
Beschreibung:	Die Produktmerkmalskandidatenidentifikationen werden mit der zweiten Interaktionskonfiguration korrigiert. Dabei soll die Wirkung dieser Benutzerinterventionen auf die Produktmerkmalskandidaten der ersten 40 Ränge untersucht werden.
Berechnung:	Precision, Recall, F-Measure
Schritte:	Produktmerkmalskandidaten ohne Korrektur Produktmerkmalskandidaten mit Korrektur aus Setting 2
Darstellung:	Linendiagramme stellen die berechneten Werte dar. Eine Tabelle stellt die ersten 10 Ränge der Annotationsdaten mit den ersten 10 Ränge der Produktmerkmalsidentifikation gegenüber.



Fortsetzung auf nächster Seite



Ergebnis: Diese Untersuchung betrachtet hierbei die Effekte einer Korrektur auf den ersten 40 ranghöchsten Produktmerkmalskandidaten. Dabei soll *Konfiguration 2* untersucht werden, da sich die vom Analysten getätigten Korrekturen gerade in diesem Bereich abspielen.

Das Schaubild stellt hierbei das Verhalten der *Precision*-, *Recall*- und *F-Measure*-Werte dar, wobei in deren Berechnungen immer mehr Produktmerkmale mit höherem Rang einfließen. Hieraus können Verbesserungen durch die Korrekturen abgeleitet werden. Die Korrekturen ermöglichen folglich, dass sich mehr vertrauenswürdige Merkmale unter den ersten 40 Produktmerkmalskandidaten befinden.

Dabei sinkt die *Precision*-Kurve, wenn ein Produktmerkmalskandidat nicht annotiert wurde. Dies kann die Interaktive Steuerung verhindern oder mit 10 Korrekturen zumindest hinauszögern. Durch Erhöhung des Schwellenwertes erhöht sich die Menge der Produktmerkmalskandidaten, wodurch der Recall ansteigt. Da die Häufigkeiten der Merkmale zunehmend abnehmen, sinkt allmählich die Steigung der *Recall*-Kurve.

Schlussfolgerung: Betrachtet man nur die Referenzen, so wirkt sich die Evaluationspezifikation eher negativ auf das Ergebnis aus. Dabei konnte aber gezeigt werden, dass diese Art der Betrachtung ungeeignet ist, da sie diese Effekte falsch widerspiegelt. Denn die Wegnahme eines Produktmerkmals erzwingt eine Umorientierung der Meinungen durch die Nächste-Nachbarn-Methoden. Dadurch bleibt das Meinungs Wort für die Berechnung bestehen, wodurch sich das Ergebnis nicht wirklich verbessern kann. Es müssen folglich nicht nur Merkmale, sondern auch Meinungen verbessert werden, um das Ergebnis positiv zu beeinflussen. Dabei kann die Interaktive Steuerung durchaus zu einer Verbesserung beitragen, besonders, wenn man nur die ersten 40 Produktmerkmalskandidaten betrachtet. Über automatisch generierte Vorschläge in *Konfiguration 2* lässt sich am ehesten eine schnelle Verbesserung erreichen. Die Frage, wie viele Verbesserung nötig sind, konnte leider nicht geklärt werden, wobei aber 10 Korrekturen ausreichen können, um zumindest einen Teil der Daten zu verbessern.

5.3.2.4 Referenzen

Referenzen drücken merkmalsbezogene Meinungen aus. Bisher konnte gezeigt werden, dass deren Identifikationen von den Berechnungen der Meinungs- und Produktmerkmalskandidaten abhängig sind. Werden diese im Vorfeld korrigiert, kann sich bereits eine geringe Verbesserung einstellen. Dabei zeigten aber die vorhergehenden Untersuchungen, dass Referenzen durch die vorliegenden Evaluationsdaten nur schwer untersucht werden können. Da der Umfang dieser Arbeit beschränkt ist, konnten keine selbst annotierten Daten erfasst werden. Dieses Kapitel soll nun die Interaktive Steuerung für Referenzen in einem theoretischen Rahmen beleuchten. Dabei werden die unterschiedlichen Änderungsstrategien herausgegriffen und anhand von Beispielen erläutert, wie diese durch den Analysten korrigiert werden können und welche Auswirkungen hierbei entstehen können. Dabei wird auf einen Datensatz zurückgegriffen, der Rezensionen über das Produkt „Canon Camera G3“ enthält. Allgemein wird eine Korrektur von Referenzen durch ein bestimmtes Meinungswort und einem bestimmten Produktmerkmal definiert. Hierdurch können folglich Meinungswörter und Produktmerkmalskandidaten in Kombination ausgeschlossen werden, was durch die vorhergehenden Korrekturen nicht bewerkstelligt werden konnte. Hierdurch kann das Ergebnis auf eine weniger radikale Weise verbessert werden. Dabei hat der Analyst die Möglichkeit den tatsächlichen Kontext zu erfassen. Er kann diese Referenz folglich als sinnvoll oder nicht sinnvoll einstufen und daraufhin Korrekturen veranlassen. Dieses Unterkapitel beleuchtet somit die unterschiedlichen Arten von Korrekturen, welche durch eine Interaktive Steuerung ermöglicht werden sollen.

Konvertierung von Referenzen zu Produktmerkmalen In den Daten kann es oftmals vorkommen, dass sich ein Produktmerkmal aus einem Adjektiv und einem Nomen zusammensetzt⁴⁹. Diese können durch die bisherige Identifikation von Nomenketten nicht erfasst werden. Besteht für dieses Adjektiv nun ein subjektiver Charakter, so wird dieses Phänomen mit großer Wahrscheinlichkeit als Referenz betrachtet. Diese tauchen häufig unter den vertrauenswürdigen Referenzen auf, da sie zum einen einen geringen Abstand zwischen Meinung und Merkmal aufweisen und zum anderen häufig zusammen auftauchen. Somit können diese einfach aufgedeckt und korrigiert werden. Eine solche Korrektur kann folglich zu einem besseren Ergebnis führen. Hierdurch können nun auch Produktmerkmale definiert werden, welche nicht nur aus Nomenketten bestehen, wodurch der Identifikationsumfang erweitert werden kann. Dadurch, dass ein vorher als meinungstragendes Wort nun als Merkmal erfasst ist, können falsche Referenzen eliminiert werden, wodurch sich das Ergebnis verbessert.⁵⁰

⁴⁹digital camera, digital cameras, digital photography, digital zoom, outdoor picture, professional photographer, extra money, white spot, red eye reduction, hard drive

⁵⁰Wird zum Beispiel „digital camera“ als Merkmal erfasst, wobei „digital“ als Meinungswort markiert war, wird es in der Summen-Klassifikation nun nicht mehr betrachtet

Korrekturen von Kombinationen zwischen Produktmerkmalen und Meinungen Oftmals werden Referenzen identifiziert, welche keine merkmalsbezogenen Meinungen modellieren. Dies kann zum einen durch ein falsches Meinungswort, oder durch einen falschen Produktmerkmalskandidaten hervorgerufen werden⁵¹. Die Interaktive Steuerung kann diese Referenzen aus der Identifikation ausschließen. Anhand der Beziehungen lassen sich automatische Vorschläge berechnen, welche entweder aus allen Referenzen mit einem bestimmten Meinungswort oder mit einem bestimmten Produktmerkmal bestehen. Hierdurch ist es möglich, Kombinationen von Meinungen und Produktmerkmalen zu deaktivieren, ohne im Vorfeld Informationen auszuschließen⁵². Damit kann der Analyst das Ergebnis besser beeinflussen, als es mit einer separaten Korrektur von Produktmerkmalskandidaten oder Meinungskandidaten möglich ist, da hierbei der Kontext für den Analysten verständlicher ist. Durch die große Vielfalt an unterschiedlichen Kombinationen kann es sinnvoll sein, dem Analysten eine zusätzliche Darstellung anzubieten, welche alle Referenzen für entweder jedes Meinungswort oder jedes Produktmerkmal zusammenfasst und nach Vertrauenswert sortiert.

Kontextkorrektur Speziell bei unterschiedlichen Domänen können sich die Ausrichtungen von Meinungskandidaten unterscheiden. Dieses Phänomen kann nur über den Kontext oder den Menschen aufgedeckt werden. So existieren zum Beispiel Produktmerkmale, die einen negativen Kontext besitzen wie zum Beispiel das Produktmerkmal „*Tintenverbrauch*“. Dieser negative Kontext beeinflusst folglich diejenigen Meinungswörter, welche sich auf dieses Produktmerkmal beziehen. Die Interaktive Steuerung unterstützt hierbei eine Kontextkorrektur, welche zwischen einer Kombination von Meinungswort und Produktmerkmal definiert werden kann. Diese Phänomene können in verschiedenen Domänen unterschiedlich oft auftauchen. Diese Korrektur kann abermals durch automatisch generierte Vorschläge unterstützt werden, wobei diese durch eine Referenz mit gleichem Produktmerkmal definiert werden. Im Regelfall tauchen diese Kontextänderungen aber eher selten auf, müssen aber dennoch in einem Meinungsanalyseprozess modelliert werden können.

Schlussfolgerung: Korrekturen von Referenzen machen durchaus Sinn, da dort nur Kombinationen von Meinungswörtern und Produktmerkmalen erfasst werden. Dabei kann der Analyst den Kontext der Referenz verstehen, ohne ihn im Text nachlesen zu müssen. Dabei deckt die Interaktive Steuerung viele Möglichkeiten ab, die der Analyst benötigt, um eine Prozessoptimierung durchführen zu können. Hierbei müssen aber recht viele verschiedene Kombinationen untersucht werden, wodurch es sich lohnen kann, zunächst eine Korrektur von Merkmalen vorzunehmen oder

⁵¹Dies kommt zum Beispiel oft in der Kombination von „not“ und einem Produktmerkmal vor

⁵²So kann zum Beispiel die Kombination „*nice weather*“ ausgeschlossen werden, wobei die Kombination „*nice camera*“ bestehen bleiben kann

dem Analysten eine geeignetere Übersicht vorzulegen.

5.3.2.5 Hypothese

Die vorhergehenden Untersuchungen konnten zeigen, dass die Interaktive Steuerung durchaus positive Effekte auf das Ergebnis haben kann. Diese konnten anhand der Evaluationstechnik aber, besonders für Referenzen, nicht eindeutig nachgewiesen werden. Jedoch konnten Vermutungen aufgestellt werden, welche der Interaktiven Steuerung ein großes Potential zusprechen. So können die Ergebnisse insbesondere dann effektiv und effizient verbessert werden, wenn Korrekturen durch automatische Vorschläge erweitert werden, wodurch mehr Fehlerfälle abgedeckt werden können. Dies zeigten primär diejenigen Untersuchungen, welche eine größere Anzahl an Korrekturen simulierten. So ist es folglich sinnvoll, eine durch die Interaktive Steuerung ermöglichte Prozessoptimierung durchzuführen, wobei dies mit deutlich mehr als 10 Korrekturen erfolgen muss.

5.3.3 Übertragung der Prozessoptimierung

Nachdem die Korrekturen bisher nur auf einem Datensatz ausgeführt wurden, soll nun im weiteren der Nutzen einer Übertragung dieser Prozessoptimierung untersucht werden. Da sich die einzelnen Korrekturen auf Terme stützen und diese durch Kombinationen von Wortartentscheidungen definiert sind, sollte die Prozessoptimierung möglichst auf einem großen Datensatz erfolgen, um möglichst viele Vorkommen abdecken zu können. Dies gilt primär für die Korrekturen an Wortarten und Meinungskandidaten, während Korrekturen der Produktmerkmalskandidaten und Referenzen über spezifische Wörter oder Wörterkombinationen definiert werden.

Hypothese: Die Korrekturen, welche durch die Interaktive Steuerung ermöglicht werden, formulieren Wörterbücher, in denen Ausnahmen festgelegt sind. Im Auftretungsfall werden die errechneten Informationen mit den Informationen aus diesem Wörterbuch überschrieben. Dies kann als Prozessoptimierung betrachtet werden, in dem der Analyseprozess auf einen spezifischen Datensatz angepasst wird. Die folgende Evaluation will nun zeigen, dass diese Prozessoptimierung auch in gewissem Maße auf andere Textdaten der gleichen Domäne übertragen werden kann, wodurch dort nun weniger Korrekturen vorgenommen werden müssen.

5.3.3.1 Wortarten

Die Korrekturen der Wortarten umfassen primär die größten Fehler, welche durch Uneinigkeit der Erkennungsmethoden entstehen. So werden zum Beispiel primär nicht als Nomen deklarierte Merkmale hinzugefügt, oder fälschlicherweise als Nomen definierte Terme geändert. Innerhalb einer

Domäne tauchen meist dieselben Terme auf, wodurch die Wortarterkennung zu gleichen Uneinigkeiten kommen kann. Eine Übertragbarkeit macht folglich dann Sinn, wenn es sich um ähnliche Schreibstile handelt und ähnliche Terme verwendet werden. Dies soll hier nicht weiter evaluiert werden, da sich zum einen durch Wortartkorrekturen keine überzeugende Verbesserung einstellt und zum anderen in dieser Arbeit nicht im Vordergrund stehen soll.

5.3.3.2 Meinungskandidaten

Da durch die bisherigen Evaluationen der Meinungskandidaten nur geringe Verbesserungen aufgezeigt werden konnten, liegt die Vermutung nahe, dass die Untersuchung einer Prozessoptimierung hier noch weitaus geringere Resultate vorweisen wird. Dadurch, dass sich diese Evaluationsspezifikation als nicht sinnvoll erwiesen hat, soll hier nun eine Untersuchung aus theoretische Sicht erfolgen, wobei sich diese auf die vorhergehende Untersuchung in Kapitel 5.3.1.1 stützt. Dort wurden Meinungskandidaten mit einem steigenden Schwellenwert eingegrenzt, wobei nur noch diejenigen Kandidaten in die Berechnung eingingen, welche einen Vertrauenswert größer oder gleich dem Schwellenwert besaßen.

Die Untersuchung der Vertrauenswürdigkeiten von Meinungskandidaten lassen somit auch Schlüsse zur Prozessoptimierung zu. Werden die unsichersten Entscheidungen nicht mehr betrachtet, so kann dies zum Beispiel eine Korrektur simulieren. Aus den verschiedenen Werten lässt sich zum Beispiel ableiten, dass sich Korrekturen in verschiedenen Dokumentensammlungen jeweils ähnlich verhalten. Dieses gleichartige Verhalten lässt somit die Vermutung zu, dass diese Effekte mit einer hohen Wahrscheinlichkeit auch in anderen Daten der gleichen Domäne eintreten werden. Dies kann als Indiz für die Möglichkeiten einer Übertragung gelten. Dabei werden umso mehr Korrekturen benötigt, je mehr sich der Wortschatz in den Textdaten unterscheidet. Für große Textmengen können sich dabei eher Überschneidungen ergeben, als für kleine Textmengen.

5.3.3.3 Produktmerkmalskandidaten

Die Prozessoptimierung für Produktmerkmalskandidaten erzeugt ein Wörterbuch, in welchem festgehalten wird, welche Teile einer spezifischen Nomenkette als Produktmerkmal angesehen werden können. Damit können zum einen ganze Nomenketten oder nur Teile von Nomenketten ausgeschlossen werden. Zusätzlich besteht die Möglichkeit, Produktmerkmale aus Referenzen zu generieren oder noch nicht identifizierte Merkmale als solche im Text zu markieren, wodurch ein umfassendes Domänen-spezifisches Wörterbuch entstehen kann.

Untersuchung:	Produktmerkmalskandidatenkorrekturen [BingLiuReviews]
Beschreibung:	Die Produktmerkmalskandidaten werden mit der zweiten Interaktionskonfiguration korrigiert [Nokia 6610]. Diese werden dann auf Textdaten der gleichen Domäne übertragen [Nokia 6600]. Dort werden zusätzliche 10 Korrekturen vorgenommen [Nokia 6600]. Die gesamte Prozessoptimierung wird nochmal auf den Ausgangsdaten angewendet [Nokia 6610].
Berechnung:	Precision, Recall
Schritte:	Produktmerkmalskandidaten ohne Korrektur Produktmerkmalskandidaten mit 10 Korrekturen aus Setting 2 in [Nokia 6610] Produktmerkmalskandidaten mit 10 Korrekturen aus Setting 2 in [Nokia 6600] Produktmerkmalskandidaten mit 20 Korrekturen aus Setting 2 in [Nokia 6600] Produktmerkmalskandidaten mit 20 Korrekturen aus Setting 2 in [Nokia 6610]
Darstellung:	Liniendiagramme stellen die berechneten Werte dar. Eine Tabelle stellt die ersten 10 Ränge der Annotationsdaten mit den ersten 10 Ränge der Produktmerkmalsidentifikation gegenüber.

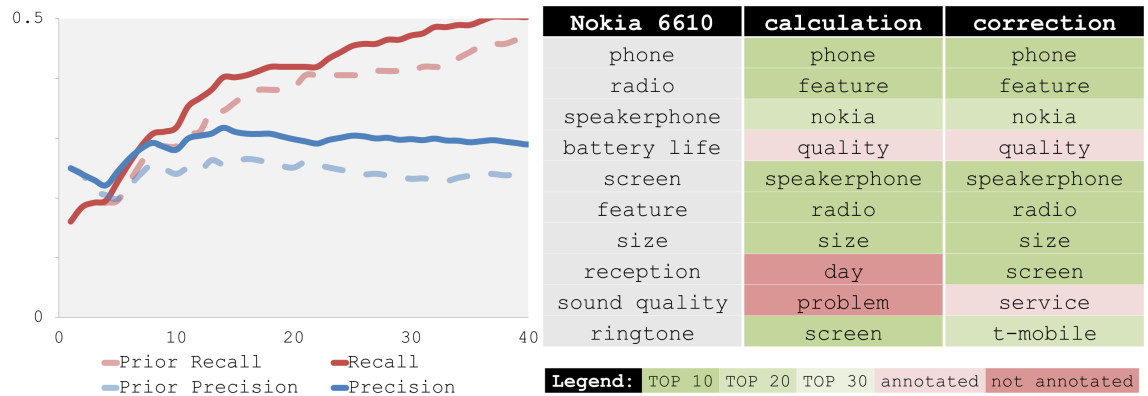


Abbildung 5.17: Der Analyst eliminierte mittels 10 Korrekturen diejenigen Produktmerkmalskandidaten, welche für ihn kein Produktmerkmal darstellen. Hierdurch befinden sich mehr sichere Produktmerkmalskandidaten unter den ersten 40 Rängen, wodurch sich bessere *Precision*- und *Recall*-Werte ergeben.

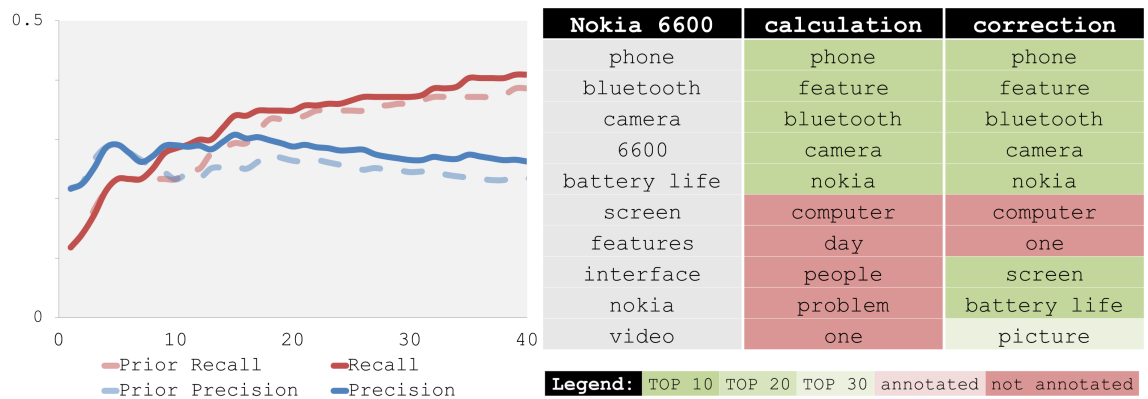


Abbildung 5.18: Werden die Korrekturen auf andere Textdaten angewendet, so entsteht ebenfalls eine Verbesserung, welche aber nicht so hoch ausfällt.

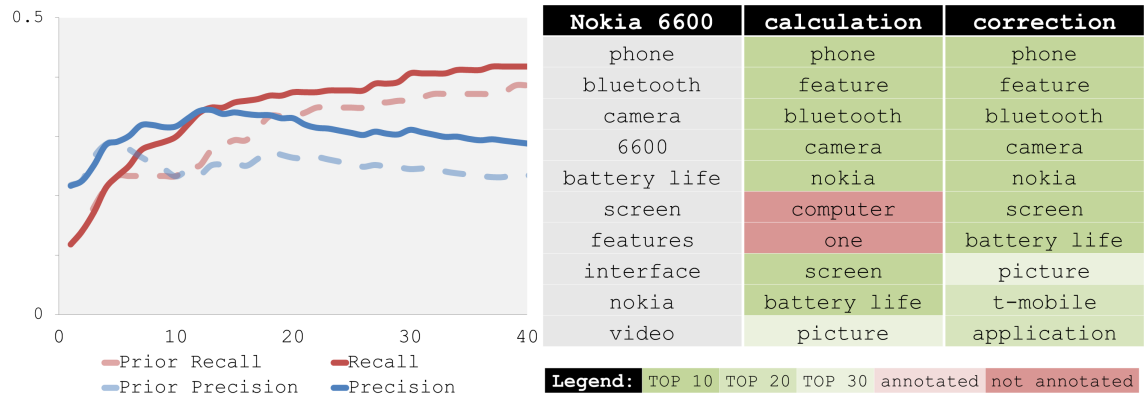


Abbildung 5.19: Werden in diesen Textdaten weitere 10 Korrekturen vorgenommen, verbessern sich die Precision- und Recall-Werte abermals.

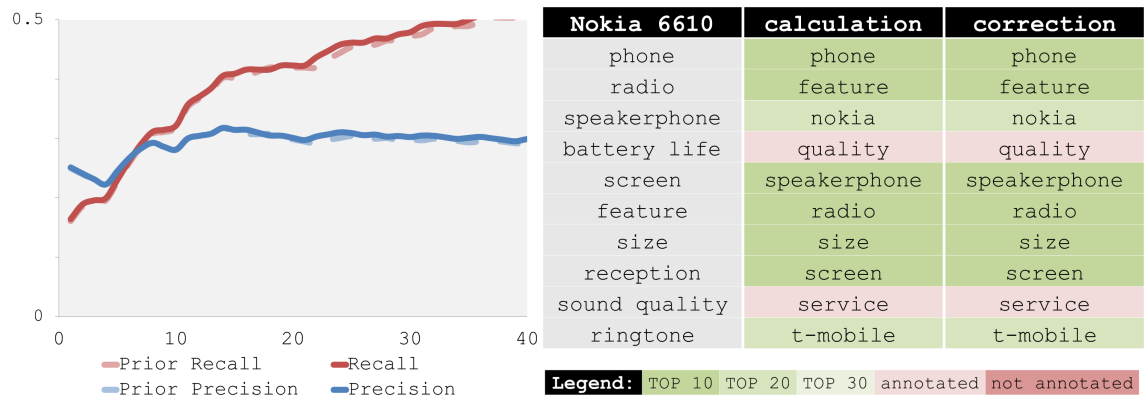


Abbildung 5.20: Wird die aktuelle Prozessoptimierung wieder auf den Ausgangsdaten angewendet, so ergibt sich keine merkliche Verbesserung mehr.

Schlussfolgerung: Eine Prozessoptimierung für Produktmerkmalskandidaten kann auf Textdaten der gleichen Domäne zu einer Verbesserung führen, da diese hier mit großer Wahrscheinlichkeit ebenfalls keine Produktmerkmale darstellen. Dabei konnte festgestellt werden, dass dieser Übertrag durch weitere Korrekturen verfeinert werden muss, da sich selbst Textdaten der gleichen Domäne unterscheiden können, wobei ein Analyst hierbei nun weniger Korrekturen vornehmen muss. Aus dieser resultierenden Prozessoptimierung konnten schließlich beide Kollektionen profitieren. Die Untersuchung hat gezeigt, dass durch mehrmaliges Bearbeiten ein folglich immer spezifischer Wortschatz für eine Domäne entsteht. Engt der Analyst diesen Rahmen zu sehr ein, könnten hierdurch auch Informationen verloren gehen, wodurch sich auch negative Effekte auf das Ergebnis abzeichnen könnten, welche aber in dieser Untersuchung nicht auftraten.

5.3.3.4 Referenzen

Die vorhergehende theoretische Evaluation hat gezeigt, dass die Korrekturen der Referenzen Kombinationen von Meinungswörtern und Produktmerkmalen erfassen, wobei hier der Analyst den

Kontext der Referenz besser verstehen kann. Dabei deckt die Interaktive Steuerung viele Möglichkeiten ab, die der Analyst benötigt, um eine Prozessoptimierung durchführen zu können.

Betrachtet man die Prozessoptimierung auf Seite der Referenzen, so kann diese auch als Kontextanpassung verstanden werden, welche besonders durch Kontextkorrekturen hervorgerufen werden. Dabei werden sensitive Informationen des Kontextes in den Analyseprozess eingebettet, wodurch eine Domänen-spezifische Anpassung erfolgt. Diese kann auf Textdaten der gleichen Domäne übertragen werden, da dort derselbe Kontext vorherrscht. Für diesen Teil der Interaktiven Steuerung ist eine Übertragung folglich sinnvoll.

Dabei kann man für große Textdaten davon ausgehen, dass verbesserte Kombinationen zwischen Meinungen und Produktmerkmalen auch in anderen Dokumentensammlungen existieren. Dort handelt es sich um ähnliche Produkte mit ähnlichen Eigenschaften und merkmalsbezogenen Meinungen. Somit kann der Analyseprozess auch aus diesen Korrekturen profitieren, wodurch in neuen Dokumentensammlungen nun weniger Kontextanpassungen nötig sind.

Hierbei müssen aber recht viele verschiedene Kombinationen untersucht werden, wodurch es sich lohnen kann, zunächst eine Korrektur von Merkmalen vorzunehmen oder dem Analysten eine geeignetere Übersicht vorzulegen.

5.3.3.5 Hypothese

Die Evaluation hat gezeigt, dass ein Übertrag einer Prozessoptimierung auf Textdaten der gleichen Domäne sinnvoll sein kann, da dort zum einen ähnliche Terme vorliegen und derselbe Kontext verwendet wird. Dies zeigt speziell die Untersuchungen der Produktmerkmalskandidaten. Dabei darf der Analyst aber nicht zu spezifisch vorgehen, da hierdurch ein Ausschluss von korrekten Informationen entstehen kann, welcher für das Ergebnis von Nachteil sein könnte.

Wird eine Prozessoptimierung in neuen Textdaten angewendet, so müssen dort zusätzliche Korrekturen ausgeführt werden, da die automatisch generierten Vorschläge von den Termen innerhalb der Dokumentensammlungen abhängig sind. Dies liegt primär an der Konzeption der Interaktiven Steuerung, wobei dieser Nachteil durch eine einfache Modifikation behoben werden kann. Zum Beispiel erfasste eine mit Vorschlägen angereicherte Korrektur diejenigen Merkmale, welche den Term „day“ enthalten. In *Kollektion A* betraf das die zusätzlichen Terme „monday“ und „tuesday“, während die Terme „saturday“ und „sunday“ in *Kollektion B* bisher nicht erfasst werden können. Speichert die Interaktive Steuerung hierbei Korrekturen als eine Art Suchanfrage, so könnte diese bei einem Übertrag auch auf neue Terme angewendet werden, wodurch folglich der Korrekturaufwand noch weiter minimiert werden kann.

5.3.4 Schlussfolgerungen

Die Untersuchungen über den Analyseprozess und der Interaktiven Steuerung ergaben Ergebnisse, welche nur in wenigen Fällen nicht eindeutig nachgewiesen werden konnten. Dies lag primär an der zur Verfügung stehenden Evaluationsgrundlage, welche nur eine indirekte Evaluation der Identifikationen zuließ. Besonders eine konkrete Untersuchung der Referenzen konnte nicht angestellt werden. Die Erstellung von eigenen Annotationsdaten konnte durch den beschränkten Aufwand innerhalb dieser Arbeit nicht durchgeführt werden. Dennoch konnten aber die Untersuchungen zeigen, dass sich aus der Modulstrukturierung des Analyseprozesses ein Mehrwert ableiten lässt, welcher einen positiven Effekt auf das Endergebnis haben kann, was insbesondere durch die Berechnung von Vertrauenswerten erreicht wird. Aber auch die Interaktive Steuerung weist ein hohes Potential auf. Dabei kann auch der Analyst an dem Mehrwert der Vertrauenswertberechnung teilhaben und kann in seinen Korrekturen durch automatisch generierten Vorschlägen unterstützt werden. Dabei zeigten speziell die simulierten Korrekturänderungen, dass sich ein verbessertes Ergebnis ergeben kann. Zusätzlich zeigten Untersuchungen, dass sich aus den Korrekturen eine Prozessoptimierung erreichen lässt, welche auch auf Textdaten der gleichen Domäne übertragen werden kann.

Die Interaktive Steuerung erweist sich somit als geeignetes Werkzeug, einen Analyseprozess zu verbessern, ohne dabei einem Analysten zu viel Korrekturaufwand abzuverlangen. Das Umfeld der Meinungsanalyse kann folglich aus dieser Technik profitieren, wobei durchaus weitere Forschungsfragen angestellt werden können, welche aber den Umfang dieser Arbeit sprengen würden.

5.4 Ausblick

Dieses Kapitel soll nun aufzeigen, wie noch bestehende Schwierigkeiten gelöst werden können und welche Erweiterungen der Anwendung, insbesondere auch im alltäglichen Einsatz, denkbar wären. Zusätzlich werden einzelne Punkte dieser Arbeit herausgegriffen und kritisch hinterfragt.

Methodik Betrachtet man die im Analyseprozess modellierten Identifikationsmethoden, so muss diese hierbei keiner optimalen Lösung entsprechen. In der wissenschaftlichen Literatur können bestimmt weitaus bessere Herangehensweisen existieren, wobei diese durch den Umfang dieser Arbeit nicht betrachtet werden konnten. Hierdurch könnten bestimmt weitaus bessere Einzelergebnisse erzielt werden als es mit den bisherig verwendeten Methoden möglich ist. Ein primäres Erweiterungspotential besteht hierbei für Techniken zur Identifikationen von Negationen und der Auflösung von Anaphern, welche bisher nur als Heuristik eingebunden wurden. Dabei kann aber diese Anwendung Aufschlüsse über die Arbeitsweisen der verschiedenen Methoden geben. Die Evaluation hat

dabei gezeigt, dass eine Gegenüberstellung der Methoden zeigen kann, wo diese ihre Vorteile, beziehungsweise Nachteile haben. Auch hat sich eine Vertrauenswertberechnung als sinnvoll erwiesen, wodurch die Ergebnisse der bestehenden Methoden verbessert werden konnten. An dieser Arbeit können weitere Evaluationen anknüpfen, welche andere Methoden einbinden oder andere Textdaten verarbeiten. Dabei können vielleicht auch andere Annotationsdaten eingesetzt werden, worüber eine genauere Evaluation stattfinden kann, um zum Beispiel auch die Chancen der Interaktiven Steuerung von Referenzen aufzeigen zu können.

Analyseprozess Bisher beschränkt sich die Interaktive Analyse nur auf die Erstellung eines Wörterbuches, welches Ersetzungsregeln beinhaltet. Dabei wurde bereits eine Modifikation vorgeschlagen, welche eine Korrektur durch automatische Vorschläge als Suchanfrage modelliert und somit einen Korrekturaufwand einer Übertragung minimieren kann.

Bisher basieren automatisch generierte Vorschläge nur auf Ähnlichkeiten von Termen. Dabei könnten gerade Korrekturen von Produktmerkmalen durch eine Ähnlichkeit anhand ihrer Wortverwandtschaften oder anhand ihrem Bezug zu Meinungen profitieren. Hierbei könnten externe Informationsquellen oder LSA [42] den Analysten unterstützen.

Eine Korrektur will aber auch eine Verbesserung der Modulentscheidung erreichen. Bisher wurde nur das Ergebnis der Entscheidung beeinflusst, wobei noch zusätzlich die zugrundeliegenden Entscheidungsträger beeinflusst werden könnten. Eine Methode, welche viele Fehler generiert, könnte somit in ihrer Vertrauenswürdigkeit abgestuft werden, wodurch auch andere Entscheidungen betroffen sein können. Hierdurch wird nicht nur eine Prozessoptimierung erreicht, sondern ebenfalls eine Gewichtung der einzelnen Prozesse.

Visualisierung Die visuellen Komponenten innerhalb der Anwendung waren für diese Arbeit bisher eher nebensächlich. Dabei können sie aber eine große Rolle bei der Analyse von Meinungen spielen. So besitzen Visualisierungen zum Beispiel den Vorteil, dass dort Beziehungen und Ähnlichkeiten dargestellt werden können, wodurch ein Analyst anhand von Mustern interessante Bereiche erkennen kann. Im Bereich der Meinungsanalyse existieren bereits eine Vielzahl von anerkannten Techniken zur Analyse von Meinungen, welche die bisherige Tabelle ersetzen können.

Die Analyse könnte hierbei durch eine Beschränkung der Identifikationen anhand von Vertrauenswerten profitieren. Wird dieser Bereich zum Beispiel verkleinert, werden nur noch diejenigen Identifikationen angezeigt, welche Vertrauenswerte in diesem Wertebereich aufweisen. So könnten Unsicherheiten oder Sicherheiten direkt in den Textdaten ersichtlich werden, wodurch der Analyst die dahinter liegenden Prozesse eventuell besser verstehen könnte.

Anwendung als Konsument Diese Anwendung wurde als Expertensystem konzipiert, wodurch ein Konsument hierdurch in seinen alltäglichen Entscheidungen nicht profitieren kann. Dies müssen andere Anwendungen gewährleisten, welche auf den durch die Prozessoptimierung gewonnen Erkenntnissen aufbauen können. Somit werden spezifische Datensätze auf ihr jeweiliges Einsatzgebiet optimiert und in einer anwenderfreundlichen Umgebung präsentiert. Der Konsument könnte zum Beispiel bei seiner Entscheidungsfindung von einer interaktiven Weboberfläche unterstützt werden. Diese stellt die Informationen der merkmalsbezogenen Meinungen mittels den visuellen Ansätzen von Liu et al. dar [37]. Um den Konsumenten nun bei seiner Produktevaluation zu bestärken, kann er Produktmerkmale gemäß seinen Interessen gewichten, wodurch zum einen ein Mehrwert auf Konsumentenseite entsteht und zum anderen auf Seite des Produkthanbieters. Dieser kann nicht nur erfahren, welche Produktmerkmale gut beziehungsweise schlecht ankommen, sondern er kann auch aus den spezifischen Gewichtsentscheidungen ablesen, welche Merkmale seines Produkts auf Interesse seiner potentiellen Kunden stoßen.

Kapitel 6

Zusammenfassung

Diese Arbeit beschäftigte sich mit der Entwicklung einer Anwendung, worin der Analyst mittels einer Interaktiven Steuerung den Analyseprozess beeinflussen kann. Dadurch kann eine Prozessoptimierung erreicht werden, welche die Identifikationen von merkmalsbezogenen Meinungen verbessern kann.

Zu Anfang wurde die Themenstellung dieser Arbeit motiviert. Dabei konnte festgestellt werden, dass heutzutage jeder die Möglichkeit hat, Meinungen zu Produkten im Internet zu veröffentlichen, wobei diese mittlerweile für viele Käufer eine wichtige Entscheidungshilfe darstellen. Aber auch Firmen haben ein hohes Interesse an einer Auswertung dieser Informationen.

Dabei konnte festgestellt werden, dass die von den Produktplattformen zur Verfügung gestellten Bewertungssysteme ungeeignet sind, um die Stärken und Schwächen der Produkte darzustellen.

Um nicht alle Rezensionen selbst durchlesen zu müssen, können diese Identifikationsaufgaben von Techniken der Computerlinguistik und Informatik übernommen werden, welche aber durchaus Schwierigkeiten im Umgang mit natürlich-sprachlichen Texten haben können. So erschweren zum Beispiel Rechtschreibfehler, umgangssprachliche Wortgebräuche oder semantische Mehrdeutigkeiten eine automatische Identifikation.

Dabei zeigte sich, dass der Kontext für eine korrekte Meinungsanalyse betrachtet werden muss, da zum Beispiel Meinungswörter in verschiedenen Sparten unterschiedliche Ausrichtungen besitzen. Ein solches Kontextverständnis kann durch das Weltwissen eines Benutzers ausgedrückt werden. Mithilfe der Interaktiven Steuerung kann dieses in den Analyseprozess integriert werden, um zum Beispiel Domänen untereinander abzugrenzen oder unvermeidliche Fehler korrigieren zu können.

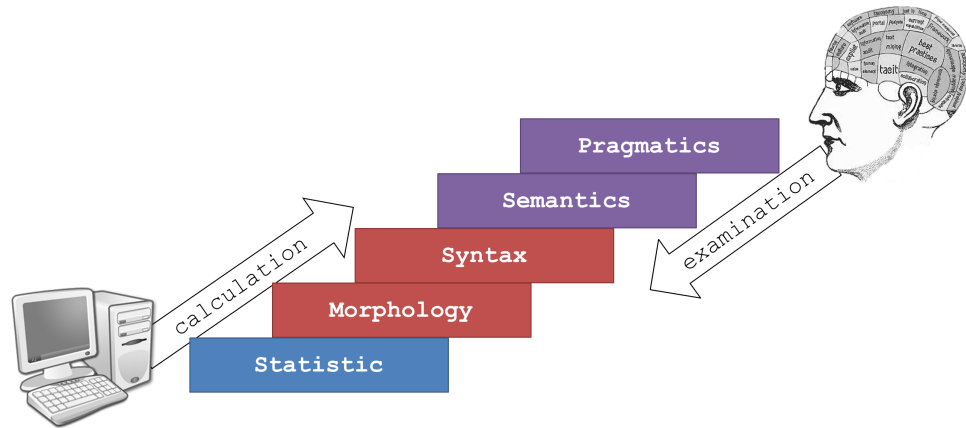


Abbildung 6.1: Diese Arbeit beleuchtet hierbei die Chancen, die durch eine Integration des Benutzers in einen automatischen Meinungsanalyseprozess bewirkt werden können.

Die Grundlagen konnten hierbei die verschiedenen Sichtweisen einer Analyse von Meinungen beleuchten. So wurde unter anderem festgestellt, dass sich der Meinungsbegriff in verschiedenen Textformen durchaus unterscheiden kann. So werden in Verkaufsplattformen primär Produkte und deren Eigenschaften bewertet, wohingegen Foren und Blogs Diskussionen über spezifische Themen beinhalten. Dabei muss der Meinungsbegriff in Nachrichtentexten differenziert betrachtet werden, da hier eine objektive Sicht auf positive, negative oder neutrale Ereignisse geschildert wird.

Diese Textformen enthalten unstrukturierte Daten, welche durch Techniken der Computerlinguistik mit sprachlichem Wissen angereichert werden können, wobei festgestellt werden konnte, dass sich diese in verschiedene sprachlichen Ebenen einstufen lassen. Diese Informationen können durch externe Wissensbasen ergänzt werden, um die Algorithmen der Meinungsanalyse verfeinern zu können.

Aufbauend auf diesem Vorwissen können nun Techniken der Informatik eingesetzt werden, um eine Identifikation von merkmalsbezogenen Meinungen bewerkstelligen zu können. Dabei müssen die Identifikationen von Meinungen, Merkmalen und Referenzen als separate Probleme betrachtet werden. Dabei konnten verschiedenste Lösungsansätze für die gleiche Identifikationsaufgabe vorgestellt werden, wobei sich Unterschiede in deren Herangehensweisen feststellen ließen.

Das Grundlagenkapitel konnte durch die Erläuterung zweier verschiedener Lösungsansätze zur Integration eines Benutzers in einen Analyseprozess abgeschlossen werden.

Anhand der vorgestellten Grundlagen konnte die Problemstellung dieser Arbeit abgeleitet werden. Zunächst wurden die Herausforderungen im Meinungsanalyseumfeld speziell für die Themenstellung dieser Arbeit betrachtet. Hieraus konnten dann Lösungsansätze für die Umsetzung generiert werden.

Unter anderem wurde aufgezeigt, wie sich mehrerer Methoden zur Lösung derselben Identifikationsaufgabe in einen Analyseprozess integrieren lassen und wie aus den Einzelentscheidungen Vertrauenswerte berechnet werden können. Aus diesen Lösungsansätzen ergab sich schließlich das Konzept der Interaktiven Steuerung.

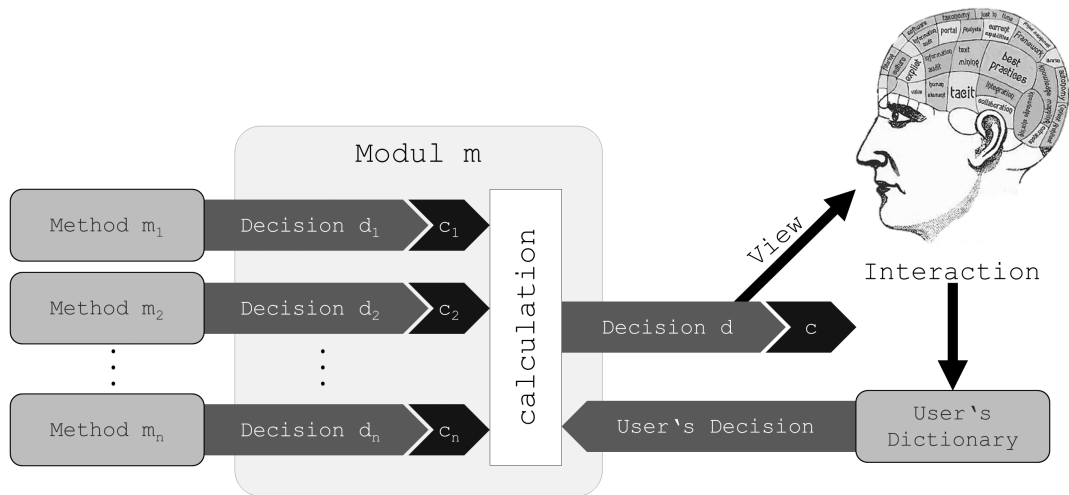


Abbildung 6.2: Das Konzept der Interaktiven Steuerung setzt auf der Modulstrukturierung des Analyseprozesses auf. Dabei können Vertrauenswerte berechnet werden, welche einen Analysten dazu veranlassen können Korrekturen vorzunehmen. Diese werden in Wörterbüchern gespeichert und überlagern die Entscheidungen der automatischen Einzelmethode.

Danach konnte der Analyseprozess beleuchtet werden, welcher Methoden zur Identifikation von Wortarten, Meinungen, Negationen, Produktmerkmalen und Referenzen enthält. Dabei wurden Unterschiede der in der Anwendung verwendeten Einzelmethode aufgedeckt. So unterscheiden sich zum Beispiel einzelne Meinungswörterbücher in ihren Mengen von negativen und positiven Meinungswörtern.

Aus der Problemstellung konnten schließlich vier Hypothesen entwickelt werden, welche unter anderem den Analyseprozess, die Interaktive Steuerung und die Übertragbarkeit der Prozessoptimierung betreffen.

- **Hypothese 1:** „Die Vereinigung von mehreren Methoden zur Lösung derselben Identifikationsaufgabe ermöglicht ein besseres Ergebnis.“
- **Hypothese 2:** „Die Priorisierung der Modulentscheidungen anhand von Sicherheitsfaktoren ermöglicht effizientere Korrekturen des Analyseprozesses, welche noch zusätzlich durch automatisch generierte Vorschläge beschleunigt werden.“
- **Hypothese 3:** „Die Interaktive Steuerung ermöglicht eine effektive und messbare Verbesserung des Ergebnisses.“

- **Hypothese 4:** „Die durch die Interaktive Steuerung erzielte Prozessoptimierung lässt sich auch auf andere Textdaten der gleichen Domäne übertragen.“

Nun konnte auf die eigentliche Entwicklung der Analyseanwendung eingegangen werden. Hierbei wurden die unterschiedlichen Komponenten beleuchtet und aufgezeigt, wie der Analyseprozess aufgebaut ist.

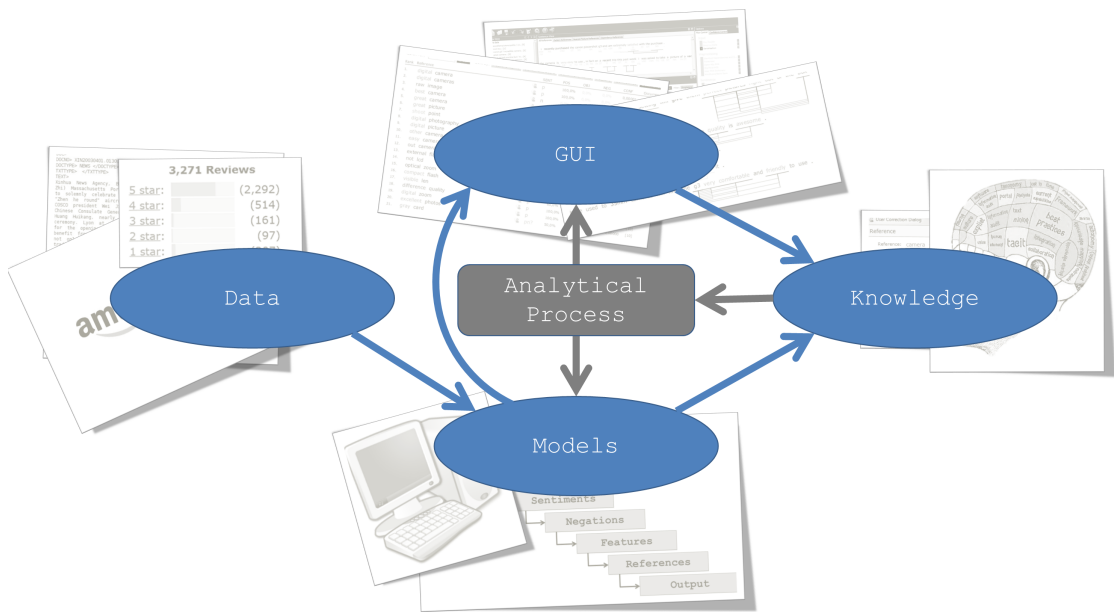


Abbildung 6.3: Die in dieser Arbeit vorgestellte Anwendung beinhaltet unterschiedliche Komponenten. Zunächst werden Daten aus verschiedenen Formaten eingelesen und in einem Datenmodell abgelegt. Dieses wird durch automatische Methoden des Analyseprozesses mit Informationen zu beispielsweise Meinungen und Produktmerkmalen angereichert. Eine Oberfläche enthält hierbei Visualisierungen der unterschiedlichen Informationen und bindet über Interaktionsschnittstellen einen Benutzer ein, welcher mittels der Interaktiven Steuerung den Analyseprozess nun kontrollieren kann.

Der Analyseprozess beinhaltet die Methoden der Meinungsanalyse und beschreibt den Umgang mit den Textdaten. Hierzu zählen zum Beispiel das Einlesen, die Vorverarbeitung, die Identifikation von Meinungs- und Produktmerkmalskandidaten, die Berechnung der Referenzen und die Ausgabe der gesammelten Informationen. Hier wurden die jeweiligen Vertrauenswertberechnungen aufgeführt und deren Abhängigkeiten zu anderen Methoden erörtert. So ergibt sich zum Beispiel für Meinungen ein hoher Vertrauenswert, wenn viele Wörterbuchentscheidungen übereinstimmen. Produktmerkmale bauen hierbei nicht auf Mehrheitsentscheidungen auf, sondern errechnen Vertrauenswerte anhand von Termbewertungen und Häufigkeiten. Referenzen verwenden diese Werte und gewichten sie unter anderem mit dem Abstand zwischen Meinungsword und Produktmerkmal. Zusätzlich konnte aufgezeigt werden, dass die in dieser Arbeit verwendeten Module auf einfache Weise erweitert werden können.

Nachdem der Analyseprozess erläutert wurde, konnte das zugrundeliegende Datenmodell vorgestellt werden, welches die Informationen des Analyseprozesses verwaltet. Hier werden zum Beispiel

die Terme innerhalb einer dynamischen Speicherstruktur abgelegt, um den Speicheraufwand minimieren zu können.

Die in der Anwendung integrierte Oberfläche bildet die Schnittstelle zwischen Benutzer und dem zu steuernden Analyseprozess. Hier konnte erörtert werden, welche Interaktionen den Analyseprozess beeinflussen können und welche Korrekturen mittels der Interaktiven Steuerung möglich sind. Dies konnte anhand eines einfachen Benutzerszenarios nachverfolgt werden.

Am Ende dieser Arbeit wurden die Erkenntnisse vorgestellt, welche sich aus den Evaluationen des Analyseprozesses und der Interaktiven Steuerung ergaben.

Dabei konnten in wenigen Untersuchungen keine eindeutigen Ergebnisse nachgewiesen werden, wobei dies primär an der zur Verfügung stehenden Evaluationsgrundlage lag, welche nur eine indirekte Evaluation der Identifikationen zuließ. Besonders eine konkrete Untersuchung der Referenzen konnte nicht angestellt werden, da hierfür eine Anfertigung eigener Annotationsdaten nötig gewesen wäre.

Andere Untersuchungen konnten aber durchaus positive Effekte vorweisen. Aus der Modulstrukturierung des Analyseprozesses konnte zum Beispiel ein Mehrwert abgeleitet werden, wodurch das Endergebnis positiv beeinflusst werden konnte. Dies wurde insbesondere durch die Berechnung von Vertrauenswerten erreicht.

Aber auch auf Seite der Interaktiven Steuerung konnten positive Effekte nachgewiesen werden. Dabei stellte sich heraus, dass auch der Analyst an dem Mehrwert der Vertrauenswertberechnung teilhaben kann und seine Korrekturen durch automatisch generierte Vorschläge sinnvoll erweitert werden konnten. Dies spiegelte sich ebenfalls in verbesserten Endergebnissen wider, wobei sich durch mehr als 10 Korrekturen eindeutigere Verbesserungen ergaben.

Zusätzlich konnten weitere Untersuchungen zeigen, dass sich aus den Korrekturen eine Prozessoptimierung erreichen lässt, welche auch auf Textdaten der gleichen Domäne übertragen werden konnte.

Diese Arbeit rundete damit ab, dass zuletzt aufgezeigt wurde, wie noch bestehende Schwierigkeiten gelöst werden könnten und welche Erweiterungen der Anwendung noch denkbar wären, um zum Beispiel auch einen alltägliche Einsatz gewährleisten zu können.

Abbildungsverzeichnis

1.1	Beeinflussung der Kaufentscheidung durch Dritte	6
1.2	Rezensionen in einer fiktiven Verkaufsplattform	7
1.3	Meinung, Produktmerkmal und deren Beziehung	8
1.4	Mehrdeutigkeiten in Textdaten	9
1.5	Bedarf einer Einbindung des Weltwissens eines Benutzers	9
1.6	Ergänzung zwischen Mensch und Maschine	10
2.1	Überblick über Rezensionen Dritter bei Amazon.	17
2.2	Morphologische Regeln aus der Computerlinguistik	21
2.3	Auszug aus dem Penn Treebank Tag Set	23
2.4	Auszug aus einem Syntaxparser	24
2.5	Auszug aus einem Dependency-Parser	25
2.6	Beispiel der Anaphorik	26
2.7	Linguistische Regeln von Hazivassilogou und McKeown	29
2.8	Dependency-Regeln von Quiu et al.	29
2.9	Dependency-Regeln von Popescu et al.	30
2.10	Kompaktheit nach Hu und Liu	31
2.11	Redundanz nach Hu und Liu	31
2.12	Dependency-Regeln von Quiu et al.	32
2.13	Referenzgewinnung durch Muster	33
2.14	Problematik bei Kontextabhängigkeit	34
2.15	Auflösung der Kontextabhängigkeit von Ding, Liu und Yu	34
2.16	Negationsregeln von Ding, Liu und Yu	35
2.17	Visuelle Analyse von Textdaten - IN-SPIRE	37
3.1	Modellierung von Sicherheiten	46
3.2	Manipulation durch Weltwissen des Benutzers	48

3.3	Überblick über das Wörterbuch „SentiWordNet“	52
3.4	Überblick über das Wörterbuch „Subjectivity Clues“	53
3.5	Identifikationsautomat für Produktmerkmalskandidaten	54
3.6	Beispiele für Produktmerkmalskandidaten	55
4.1	Schematischer Überblick über die Anwendung	58
4.2	Überblick über die Analyseprozesspipeline	59
4.3	Vertrauenswertberechnung der Wortarten	61
4.4	Gesamtentscheidung der Wortartidentifikation	62
4.5	Gesamtentscheidung für Meinungskandidaten	63
4.6	Vertrauenswertberechnung der Meinungskandidaten	64
4.7	Lokale Eigenschaft des Produktmerkmalskandidaten	66
4.8	Globaler Vertrauenswert des Produktmerkmalskandidaten	66
4.9	Vertrauenswertberechnung der Referenzen	67
4.10	Globale Vertrauenswertberechnung der Referenzen	68
4.11	Klassenbeziehungen des Datenmodells auf Dokumentenseite	69
4.12	Klassenbeziehungen des Datenmodells auf Termseite	69
4.13	Klassenbeziehungen des Meinungsidentifikationsmoduls	70
4.14	Erweiterung des Meinungsidentifikationsmoduls	71
4.15	Gesamtübersicht der Oberfläche	72
4.16	Menüleiste innerhalb der Oberfläche	72
4.17	Satzansicht innerhalb der Oberfläche	73
4.18	Farbtabelle der Oberfläche	73
4.19	Globale Ansicht innerhalb der Oberfläche	74
4.20	Optionsansicht innerhalb der Oberfläche	74
4.21	Klassenbeziehungen innerhalb der Oberfläche	75
4.22	Klassenbeziehungen der Interaktionskomponenten	76
4.23	Einstellungsdialoge innerhalb der Oberfläche	77
4.24	Projektdialoge innerhalb der Oberfläche	78
4.25	Änderungsdialog zu Wortarten innerhalb der Oberfläche	78
4.26	Änderungsdialog zu Meinungen innerhalb der Oberfläche	79
4.27	Änderungsdialog zu Produktmerkmalen innerhalb der Oberfläche	79
4.28	Änderungsdialog zu Referenzen innerhalb der Oberfläche	79
4.29	Einstellungsdialog der Methode ReviewReaderBingLiu	80
4.30	Wortartidentifikationen in der globalen Listenansicht	81

4.31	Benutzeränderung der Wortartinformation	82
5.1	Aufteilung der Dokumentensammlungen in verschiedene Sparten	84
5.2	Annotationsschema von Hu, Ding und Liu	85
5.3	Aufteilung der Produktmerkmale nach Fehlergruppen	85
5.4	Datentabelle von Hu, Ding und Liu	86
5.5	Dokumentensammlung aus Amazon nach Sterne-Klassen gruppiert	86
5.6	Beispiel für eine Ausgabentabelle der Meinungskandidaten	87
5.7	Beispiel für eine Ausgabentabelle der Produktmerkmalskandidaten	88
5.8	Verfügbare Evaluationsarten pro Dokumentensammlung	89
5.9	Evaluationsspezifikation der Meinungskandidaten	90
5.10	Berechnungen der Klassenzugehörigkeiten der Meinungskandidaten	90
5.11	Evaluationstabelle der Meinungskandidaten	91
5.12	Evaluationsberechnungen der Meinungskandidaten	91
5.13	Genauigkeit und Abdeckung für die Evaluation	91
5.14	Vortest der Produktmerkmalskandidaten	93
5.15	Genauigkeit und Abdeckung für die Evaluation	94
5.16	Konfigurationen der Evaluationen der Interaktiven Steuerung	104

Literaturverzeichnis

- [1] Min, H.J., Park, J.C.: Detecting and Blocking False Sentiment Propagation. In: Proceedings of 5th International Joint Conference on Natural Language Processing, Chiang Mai, Thailand, Asian Federation of Natural Language Processing (2011) 354–362
- [2] Sprejz, M.: Meinungsanalyse zu produkten oder services in social media. (2011)
- [3] Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2** (2008) 1–135
- [4] Liu, B.: Sentiment analysis and subjectivity. *Handbook of Natural Language Processing* (2010) 1–38
- [5] Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.: Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th international conference on World Wide Web. WWW '07, New York, NY, USA, ACM (2007) 171–180
- [6] Kisilevich, S., Rohrdantz, C., Keim, D.A.: 'beautiful picture of an ugly place'. exploring photo collections using opinion and sentiment analysis of user comments. In: IMCSIT. (2010) 419–428
- [7] Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. *Computer* (**2010**) 1320–1326
- [8] Devitt, A., Ahmad, K.: Sentiment polarity identification in financial news: A cohesion-based approach. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. (2007)
- [9] Davis, A.K., Piger, J.M., Sedor, L.M.: Beyond the numbers: an analysis of optimistic and pessimistic language in earnings press releases. Technical report (2006)
- [10] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in na-

- tural language processing - Volume 10. EMNLP '02, Stroudsburg, PA, USA, Association for Computational Linguistics (2002) 79–86
- [11] Wiebe, J., Wilson, T., Bruce, R., Bell, M., Martin, M.: Learning subjective language. *Comput. Linguist.* **30** (2004) 277–308
- [12] Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: Proceedings of the international conference on Web search and web data mining. WSDM '08, New York, NY, USA, ACM (2008) 231–240
- [13] Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th international conference on World Wide Web. WWW '03, New York, NY, USA, ACM (2003) 519–528
- [14] Porter, M.F.: An algorithm for suffix stripping. *Program* **3** (1980) 130–137
- [15] Kuhlen, R.: Experimentelle Morphologie in der Informationswissenschaft (DGD-Schriftenreihe ; Bd. 5) (German Edition). Verlag Dokumentation (1977)
- [16] Berger, A.L., Pietra, V.J.D., Pietra, S.A.D.: A maximum entropy approach to natural language processing. *Comput. Linguist.* **22** (1996) 39–71
- [17] Toutanova, K., Klein, D., Manning, C.D., Singer, Y.: Feature-rich part-of-speech tagging with a cyclic dependency network. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1. NAACL '03, Stroudsburg, PA, USA, Association for Computational Linguistics (2003) 173–180
- [18] Klein, D., Manning, C.D.: Accurate unlexicalized parsing. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1. ACL '03, Stroudsburg, PA, USA, Association for Computational Linguistics (2003) 423–430
- [19] Kubler, S., McDonald, R., Nivre, J.: *Dependency Parsing (Synthesis Lectures on Human Language Technologies)*. Morgan and Claypool Publishers (2009)
- [20] Nivre, J.: *Dependency grammar and dependency parsing*. (2005)
- [21] Marneffe, M.c.D., Maccartney, B., Manning, C.D.: Generating typed dependency parses from phrase structure parses. Volume 6., Citeseer (1999) 449–454
- [22] Qiu, G., Liu, B., Bu, J., Chen, C.: Expanding domain sentiment lexicon through double propagation. In: Proceedings of the 21st international joint conference on Artificial intelligence. IJCAI'09, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (2009) 1199–1204

- [23] Lee, H., Peirsman, Y., Chang, A., Chambers, N., Surdeanu, M., Jurafsky, D.: Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In: Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task. CONLL Shared Task '11, Stroudsburg, PA, USA, Association for Computational Linguistics (2011) 28–34
- [24] Nicolov, N., Salvetti, F., Ivanova, S.: Sentiment analysis: Does coreference matter. In: In AISB 2008 Convention Communication, Interaction and Social Intelligence. (2008)
- [25] Kazama, J., Torisawa, K.: Exploiting Wikipedia as External Knowledge for Named Entity Recognition. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. (2007) 698–707
- [26] Hatzivassiloglou, V., Wiebe, J.M.: Effects of adjective orientation and gradability on sentence subjectivity. In: Proceedings of the 18th conference on Computational linguistics - Volume 1. COLING '00, Stroudsburg, PA, USA, Association for Computational Linguistics (2000) 299–305
- [27] Hu, M., Liu, B.: Mining and summarizing customer reviews. In: KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM (2004) 168–177
- [28] Esuli, A., Sebastiani, F.: SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In: 5th Conference on Language Resources and Evaluation. (2006) 417–422
- [29] Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. ACL '02, Stroudsburg, PA, USA, Association for Computational Linguistics (2002) 417–424
- [30] Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics. ACL '98, Stroudsburg, PA, USA, Association for Computational Linguistics (1997) 174–181
- [31] Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. HLT '05, Stroudsburg, PA, USA, Association for Computational Linguistics (2005) 339–346

- [32] Hu, M., Liu, B.: Mining opinion features in customer reviews. In: Proceedings of the 19th national conference on Artificial intelligence. AAAI'04, AAAI Press (2004) 755–760
- [33] Shawn J. Bohn, Augustin J. Calapristi, S.D.B., Nakamura, G.C.: Analytic steering inserting context into the information dialog. (2011)
- [34] Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: Definition, process, and challenges. In Kerren, A., Stasko, J., Fekete, J.D., North, C., eds.: Information Visualization. Volume 4950 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2008) 154–175
- [35] Marcus, M.P., Marcinkiewicz, M.A., Santorini, B.: Building a large annotated corpus of english: the penn treebank. *Comput. Linguist.* **19** (1993) 313–330
- [36] Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilvie, D.M.: *General Inquirer: A Computer Approach to Content Analysis*. The MIT Press (1966)
- [37] Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th international conference on World Wide Web. WWW '05, New York, NY, USA, ACM (2005) 342–351
- [38] Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. HLT '05, Stroudsburg, PA, USA, Association for Computational Linguistics (2005) 347–354
- [39] Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the 2003 conference on Empirical methods in natural language processing. EMNLP '03, Stroudsburg, PA, USA, Association for Computational Linguistics (2003) 105–112
- [40] LOUGHRAN, T., MCDONALD, B.: When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance* **66** (2011) 35–65
- [41] Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. ACL '04, Stroudsburg, PA, USA, Association for Computational Linguistics (2004)
- [42] Landauer, T.K., Foltz, P.W., Laham, D.: *An Introduction to Latent Semantic Analysis*. Discourse Processes (1998) 259–284