RESEARCH

Gerald Petz

Opinion Mining im Web 2.0

Ansätze, Methoden, Vorgehensmodell

Springer Gabler

Opinion Mining im Web 2.0

Gerald Petz

Opinion Mining im Web 2.0

Ansätze, Methoden, Vorgehensmodell

Mit einem Geleitwort von
a. Univ.-Prof. DI Dr. Wolfram Wöß

Springer Gabler

Gerald Petz
Steyr, Österreich

Dissertation Johannes Kepler Universität Linz, 2014

ISBN 978-3-658-23800-1 ISBN 978-3-658-23801-8 (eBook)
https://doi.org/10.1007/978-3-658-23801-8

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Springer Gabler

Springer Gabler ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Geleitwort

Kundenbeziehungen haben in den letzten Jahren für Unternehmen stetig an Bedeutung gewonnen. Das ist an den enormen Werbeausgaben erkennbar, welche insbesondere im World Wide Web bekannten Unternehmen wie beispielsweise Google und Facebook zu hohen Umsatzzahlen und einer bedeutenden Wertschöpfung verholfen haben. Es werden aber auch vermehrt Anstrengungen unternommen, nicht nur personalisierte Werbung über solche Unternehmen an die Kunden zu bringen, sondern direkt aus der in Web 2.0 Anwendungen publizierten Kundenmeinung strukturierte Information abzuleiten, um Produkte und Dienstleistungen zu optimieren. Zusätzlich ist die veröffentlichte Kundenmeinung auch für andere Kunden interessant. In der Informations- und Kommunikationstechnik wird die automatische, strukturierte und zielgerichtete Verarbeitung veröffentlichter Meinungen im Web 2.0 als Opinion Mining (ein Teilgebiet des Text Mining) bezeichnet. Opinion Mining ist mit zahlreichen Herausforderungen behaftet, die teilweise über jene des Text Mining hinausgehen: die Datenmengen und die Aktualisierungshäufigkeit in Web 2.0-Plattformen, umgangssprachliche Äußerungen, Rechtschreib- und Grammatikfehler, Internet-Dialekte und die Bestimmung der Relevanz von Inhalten auf Webseiten.

Dieses Buch beschreibt die Entwicklung eines Vorgehensmodells für das Opinion Mining in Web 2.0-Quellen und dessen Implementierung als Software-Prototyp.

Das entwickelte Vorgehensmodell ist von hoher Relevanz, weil damit die für bestimmte Web 2.0-Anwendungsklassen am besten geeigneten Methoden systematisch zugeordnet und für einen konkreten Anwendungsfall in der Praxis angewendet werden können. Der Fokus wird dabei auf die Perspektive von Unternehmen gelegt, indem benutzergenerierte Meinungen und Stimmungen rund um Produkte und Marken von Unternehmen im Vordergrund stehen, welche extrahiert und analysiert werden.

Die ersten Kapitel des Buches geben einen detaillierten Überblick über den Stand der Forschung im Opinion Mining und dessen Möglichkeiten und Herausforderungen im Kontext von Web 2.0-Quellen, wie grammatikalische Fehler oder Internet-Dialekte. Eine Kategorisierung, Analyse und Evaluierung der Methoden und Algorithmen, die für die effektive Durchführung von Opinion Mining notwendig und geeignet sind, schließt diesen Teil ab.

Im Hauptteil des Buches werden Methoden und Algorithmen des Opinion Mining unter besonderer Berücksichtigung des Anwendungskontexts Web 2.0 in einem durchgängigen Vorgehensmodell systematisiert. Das entwickelte Vorgehensmodell ist eine Kombination von lexikalischen Ansätzen und Machine Learning Ansätzen und besteht aus fünf Phasen, die sequentiell durchlaufen als auch wiederholt durchgeführt werden (Selektion und Extraktion, Generierung der Wissensbasis, Textvorverarbeitung, Sentiment Klassifikation und Aggregation, Visualisierung und Analysen). Das Vorgehensmodell sieht eine Feedbackschleife von der letzten Phase zur ersten Phase vor, um die Qualität des gesamten Ablaufs verbessern zu können. Als Rollen werden Anwender und Domänenexperten definiert. Die Ergebnisse der einzelnen Phasen fließen jeweils in die folgenden Phasen als Eingabe ein. Im Zuge der Evaluierung wurde je nach Verfahren eine Übereinstimmung bei der Berechnung der Stimmungsrichtung von 70-80 % festgestellt.

Am Ende des Buches wird nochmals auf das breite Forschungsfeld Opinion Mining und mögliche Optimierungen hingewiesen. Dazu zählen Entitäten und Aspekte, Sarkasmus, Gesprächsfäden, annotierte Texte, Methoden im Detail sowie Multilingualiät und domänenübergreifende Inhalte mit denen die Verfahren direkt qualitativ beeinflusst und verbessert werden könnten.

Dieses Buch ist ein wichtiger Beitrag im Bereich der Anwendung von Opinion Mining auf Web 2.0-Quellen, um das Forschungsfeld weiter zu entwickeln und das praktische Handeln aus der Unternehmensperspektive zu unterstützten.

a.Univ.-Prof. DI Dr. Wolfram Wöß
Johannes Kepler Universität Linz
Linz, November 2017

Vorwort

Die Idee für die vorliegende Dissertation entstand aus mehreren Forschungsprojekten. „TSCHECHOW – Opinion Mining and Biomedical Information Retrieval“[1] bildete den ersten Rahmen für die vorliegende Arbeit und ermöglichte ein Herantasten an das Forschungsgebiet des Opinion Minings. Die zwei weiteren Forschungsprojekte „OPMIN 2.0 – Opinion Mining im Web 2.0“[2] und „SENOMWEB – Sentiment Extraction and Opinion Mining using Semantic Web Technologies“[3], bei denen der Autor dieser Arbeit jeweils der Projektleiter war, haben wesentlich zu einer Fundierung und Vertiefung in die Thematik geführt und letztlich auch den Anstoß dazu gegeben, diese Dissertation verfassen.

Gerald Petz
Juli 2018

[1] „TSCHECHOW“ war ein FH OÖ basisfinanziertes Forschungsprojekt von 2010 bis 2011

[2] „OPMIN 2.0“ wurde von 2010 – 2013 im Rahmen des Programms „COIN – Aufbau“ gefördert vom BMVIT/BMWFJ (Projekt-Nr. 826793).

[3] „SENOMWEB“ wurde im Rahmen des EU-Programms „Regionale Wettbewerbsfähigkeit OÖ 2007 – 2013 (Regio 13)“ aus Mitteln des Europäischen Fonds für Regionale Entwicklung (EFRE) sowie aus Landesmitteln gefördert.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kurzfassung

Die Art und Weise unserer Kommunikation wurde mit dem Aufkommen des Web 2.0 und seinen vielfältigen Interaktionsmöglichkeiten wesentlich verändert. Konsumenten tauschen sich über Blogs, soziale Netzwerke, Online-Rezensionsportale und Community Sites über Produkte, Marken und Unternehmen aus. Die Auswirkungen von Online-Bewertungen und von elektronischer Mundpropaganda wurden bereits vielfach untersucht und diskutiert. Opinion Mining greift diese Thematik auf und versucht, Meinungen hinsichtlich ihres Meinungsziels und ihrer Stimmungsrichtung zu analysieren. Dabei kommen Methoden aus den Bereichen Web-Crawling, Information Retrieval, Text Mining sowie linguistische und semantische Technologien zum Einsatz. Die im Web veröffentlichten Meinungen sind für viele Anwendungsszenarien interessant: für Käufer, um sich über Produkte zu informieren; für Unternehmen, um die Bedürfnisse von Kunden zu ermitteln; für Politiker, um Stimmungsbilder über politische Themen zu erhalten sowie für das Online Reputationsmanagement.

Hinsichtlich des Opinion Minings für benutzergenerierte Inhalte im Web 2.0 bestehen noch zahlreiche Herausforderungen: die Datenmengen und die Aktualisierungshäufigkeit in Web 2.0-Plattformen, umgangssprachliche Äußerungen, Rechtschreib- und Grammatikfehler, Internet-Slang Begriffe und die Relevanz von Inhalten auf Webseiten. Das Ziel dieser vorliegenden Arbeit ist es daher, ein Vorgehensmodell für das Opinion Mining für Web 2.0-Quellen zu entwickeln und dieses als Software-Prototyp umzusetzen und zu evaluieren.

Zur Erreichung des Ziels werden zunächst Forschungsfragen und eine geeignete Forschungsmethodik festgelegt. Danach werden die Grundlagen des Opinion Minings und der Stand der Forschung aus der Literatur erarbeitet. Anschließend werden mögliche Web 2.0-Anwendungsklassen in Bezug auf die Eignung für Opinion Mining empirisch untersucht und anhand unterschiedlicher Kriterien verglichen. Die spezifischen Charakteristika der einzelnen Web 2.0-Anwendungs-

klassen und der möglichen Konsequenzen für das Opinion Mining werden herausgearbeitet. Im nächsten Schritt werden Methoden, die im Opinion Mining häufig angewendet werden, ermittelt. Anhand verschiedener Kriterien wird evaluiert, inwieweit diese Methoden für benutzergenerierte Texte im Web 2.0 geeignet sind. Zur Konzeption des Vorgehensmodells für das Opinion Mining für Web 2.0 werden Vorgehensmodelle und Ansätze zu Text Mining, Data Mining und Opinion Mining recherchiert und analysiert. Anschließend werden die ermittelten Methoden in einem Vorgehensmodell systematisiert. Das Vorgehensmodell wird auf Basis von experimentellen Prototyping in einen Software-Prototyp implementiert. Der Prototyp wird anhand einer konkreten Web 2.0-Quelle angewendet, um das Vorgehensmodell evaluieren zu können. Die Arbeit schließt mit einer Zusammenfassung und Reflexion der zentralen Ergebnisse und gibt einen Ausblick auf mögliche weitere Anknüpfungspunkte für weitere Forschungsaktivitäten.

Schlagwörter
Opinion Mining, Vorgehensmodell, Web 2.0, Sentiment Analyse, Sentiment Klassifikation

Abkürzungsverzeichnis

AI	Artificial Intelligence
API	Application Programming Interface
C	Kontingenzkoeffizient
CETR	Content Extraction via Tag Ratios
$\chi 2$	Chi-Quadrat-Test
CMMI	Capability Maturity Model Integration
CMS	Content Management System
CRISP-DM	Cross Industry Standard Process for Data Mining
DF-LDA	Dirichlet Forest prior – Latent Dirichlet Allocation
DOM	Document Object Model
EFQM	European Foundation for Quality Management
ERRT	Error Rate Relative to Truncation
EU	Europäische Union
FCE	Frequently Co-occurring Entropy
GET	HTTP-Request-Methode
HMM	Hidden Markov Models
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ITIL	IT Infrastructure Library
IuK	Information und Kommunikation
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases
LDA	Latent Dirichlet Allocation
LaSA	Multilevel Latent Semantic Association
LSA	Latent Semantic Analysis
L-HMM	Lexicalized HMM
ME	Maximum Entropy
MLE	Maximum Likelihood Estimation

MSS	Maximum Subsequence Segmentation
NER	Named Entity Recognition
NLP	Natural Language Processing
OM	Opinion Mining
OWL	Web Ontology Language
PA	Passive-Aggressive Algorithmus
PHP	Hypertext Preprocessor (Skriptsprache zur Erstellung von Webanwendungen)
pLSA	Probabilistic Latent Semantic Analysis
PMI	Pointwise Mutual Information
PMI-IR	Pointwise Mutual Information – Information Retrieval
POS	Part-of-Speech
REST	Representational State Transfer
RUP	Rational Unified Process
PR	Public Relations
PRINCE2	Projects in Controlled Environments
SCRUM	Vorgehensmodell in der Softwaretechnik
SDK	Software Development Kit
SO	Semantische Orientierung (Semantic Orientation)
SPICE	Software Process Improvement and Capability Determination
STTS	Stuttgart-Tübingen Tagset
SVM	Support Vector Machine
SVR	Support Vector Regression
TBL	Transformation Based Learning
TF-IDF	Term Frequency Inverse Document Frequency
TREC	Text Retrieval Conference
TTR	Text-To-Tag Ratio
UGC	User generated content
URL	Uniform Resource Locator
VIPS	Vision-based page segmentation
XML	Extensible Markup Language

XPath	XML Path Language
WSJ	Wall Street Journal
WWW	World Wide Web

1 Einleitung

Im Rahmen des Web 2.0 entwickelte sich das Internet weg von einer unidirektionalen Informationsquelle hin zu einem interaktiven Medium. Damit wurde die Art und Weise unserer Kommunikation maßgeblich gestaltet: Konsumenten tauschen sich über Web 2.0-Anwendungsklassen wie Wikis, Blogs, Community Sites und soziale Netzwerke aus und generieren von sich aus Informationen zu Produkten, Marken und Unternehmen. Dieser geänderten Form der Kommunikation können sich sowohl private Nutzer als auch Unternehmen nicht entziehen.[4] Die von Benutzern generierten Inhalte sind für vielerlei Anwendungsszenarien hochinteressant: für Käufer, um sich über Vor- und Nachteile von Produkten und Erfahrungen von anderen Käufern zu informieren; für Unternehmen, um ihre Kunden und deren Bedürfnisse und Erfahrungen besser zu erkennen; für Politiker, um ein Stimmungsbild über ihre Person oder politische Themen zu erhalten. Weitere Anwendungsszenarien sind beispielsweise das Reputationsmanagement sowie die Verwendung der benutzergenerierten Inhalte für die Erkennung von Trends oder die Vorhersage von Verkäufen.[5] Vielfach stehen bei diesen benutzergenerierten Inhalten nicht nur Fakten im Vordergrund, sondern vor allem auch Meinungen und persönliche Eindrücke zu Marken, Produkten oder Unternehmen.[6,7] Diese benutzergenerierten Inhalte sind zumeist sehr umfangreich; eine computerunterstützte Auswertung und Zusammenfassung ist daher sinnvoll und wünschenswert.

Die Auswirkungen von Online-Bewertungen und elektronischer Mundpropaganda auf das Verhalten von Konsumenten wurden in zahlreichen Publikationen untersucht. Es wurde beispielsweise in einer Studie die Wirkung von Online-Be-

[4] Vgl. Alby, 2008, 15ff.
[5] Vgl. Pang/Lee, 2008, 12ff.
[6] Vgl. Maynard/Bontcheva/Rout, 2012
[7] Vgl. Esuli, 2008, xv

G. Petz, *Opinion Mining im Web 2.0*, https://doi.org/10.1007/978-3-658-23801-8_1

wertungen auf den Umsatz mit neuen Produkten im Bereich Unterhaltungselektronik und Videospiele untersucht. Die Analysen haben gezeigt, dass die Anzahl der Bewertungen einen signifikanten Effekt auf den Umsatz mit neuen Produkten speziell in der Frühphase hat, dieser Effekt aber mit der Zeit abnimmt. Darüber hinaus hat der Anteil der negativen Bewertungen eine größere Wirkung als die der positiven Bewertungen.[8] Die Rezensionen in Reiseinformationsportalen im Web werden von Konsumenten häufig dazu genutzt, um Entscheidungen für oder gegen eine Unterkunft zu treffen, weniger aber für eine Reiseplanung.[9] Eine andere Studie hat ergeben, dass durch Bewertungen – sowohl positive als auch negative – Hotels stärker in das Bewusstsein der Konsumenten treten; positive Bewertungen verbessern aber die Einstellung von Konsumenten zum jeweiligen Hotel.[10]

Ziel des Opinion Minings (in der Literatur finden sich auch andere Begriffe wie „Sentimentanalyse“, „Opinion Extraction“ u.v.m., die teils synonym verwendet werden) ist es daher, diese produkt- oder markenbezogenen Meinungen von Menschen zu finden, zu analysieren und systematisch aufzubereiten.[11] Dabei werden Methoden aus den verschiedensten Bereichen kombiniert: Web-Crawling und Information Retrieval zur Informationsbeschaffung, Text Mining Methoden, linguistische Algorithmen, semantische Technologien sowie Information Retrieval Ansätze zur Inhaltsanalyse und -kondensierung zur Informationsaufbereitung.[12,13]

1.1 Problemstellung und Relevanz

Opinion Mining ist ein Forschungsgebiet, das seit einigen Jahren aus Sicht verschiedenster Disziplinen in zahlreichen Veröffentlichungen behandelt wird. Die ersten wissenschaftlichen Veröffentlichungen, bei denen explizit der Terminus „Opinion Mining“ oder „Sentimentanalyse“ verwendet wurde, finden sich in

[8] Vgl. Cui/Lui/Guo, 2012, 39ff.
[9] Vgl. Gretzel/Yoo, 2008, 35ff.
[10] Vgl. Vermeulen/Seegers, 2009, 123ff.
[11] Vgl. Glance et al., 2005
[12] Vgl. Liu, 2012, S. 8
[13] Vgl. Esuli, 2008, xvi

Nasukawa/Yi[14] bzw. in Dave et al.[15], wobei erste Forschungsergebnisse aber bereits früher veröffentlicht wurden (z.B. Das/Chen[16], Tong[17], Turney[18], Wiebe[19]). Eine Vielzahl von Veröffentlichungen finden sich in Journalen und Tagungsbänden zu den Themen Natural Language Processing (NLP) (z.B. Somasundaran et al.[20], Wiebe/Mihalcea[21], Pang et al.[22]), Data Mining und Web Mining (z.B. Ding et al.[23], Morinaga et al.[24]) und Machine Learning wieder (z.B. Archak[25]). Etliche Publikationen wurden aber auch in den Bereichen E-Commerce (z.B. Ghose/Ipeirotis[26], Hu et al.[27], Park et al.[28]) oder Management Sciences (z.B. Chen/Xie[29], Das/Chen[30], Dellarocas[31]) und vielen anderen Disziplinen veröffentlicht. Die Auswirkungen von Online-Rezensionen auf die Kaufentscheidung wurden ebenfalls vielfach untersucht (z.B. Chen/Xie[32], Chevalier/Mayzlin[33], Lee et al.[34]).

Die Meinungen, die von den Menschen in den verschiedensten Netzwerken und Plattformen im Web hinterlassen werden, stellen letztlich eine wertvolle Informationsquelle für Unternehmen dar. Diese Informationen ermöglichen Analysen über öffentliche Meinungen und über die Reputation von Unternehmen, Analysen von Trends sowie eine neue Art der Marktforschung. Letztlich weist die

[14] Vgl. Nasukawa/Yi, 2003
[15] Vgl. Dave/Lawrence/Pennock, 2003
[16] Vgl. Das/Chen, 2001
[17] Vgl. Tong, 2001
[18] Vgl. Turney, 2002
[19] Vgl. Wiebe, 2000
[20] Vgl. Somasundaran et al., 2009
[21] Vgl. Wiebe/Mihalcea, 2006
[22] Vgl. Pang/Lee/Vaithyanathan, 2002
[23] Vgl. Ding/Liu/Yu, 2008
[24] Vgl. Morinaga et al., 2002
[25] Vgl. Archak/Ghose/Ipeirotis, 2007
[26] Vgl. Ghose/Ipeirotis, 2007
[27] Vgl. Hu/Pavlou/Zhang, 2006
[28] Vgl. Park/Lee/Han, 2007
[29] Vgl. Chen/Xie, 2008
[30] Vgl. Das/Chen, 2007
[31] Vgl. Dellarocas/Zhang/Awad, 2007
[32] Vgl. Chen/Xie, 2008
[33] Vgl. Chevalier/Mayzlin, 2006
[34] Vgl. Lee/Park/Han, 2008

computerunterstützte Analyse von Meinungen im Web 2.0 gegenüber traditionellen Verfahren einige Vorteile auf[35]: eine große Menge an Kundenmeinungen steht zur Verfügung, die Meinungserhebung findet in der natürlichen Kommunikationsumgebung statt und die Meinungen stehen in der Regel kostenlos zur Verfügung. Aufgrund der großen Anzahl an Beiträgen in einer Vielzahl von Plattformen ist eine manuelle Analyse nur bedingt möglich; eine computerunterstützte Analyse ist daher sinnvoll und notwendig.[36]

Obwohl in den letzten Jahren viele Forschungserkenntnisse die Qualität des Opinion Minings immer weiter verbessert haben, gibt es immer noch zahlreiche Herausforderungen:[37]

- Opinion Mining im Web 2.0
  Die Aktualisierungshäufigkeit von benutzergenerierten Inhalten im Web 2.0 ist vergleichsweise hoch und die Meinung der Autoren kann sich rasch ändern. Darüber hinaus können Meinungen auf vielerlei Art und Weisen kundgetan werden: von einem simplen „Like“ in sozialen Netzwerken wie etwa Facebook, über Rezensionen bis hin zu umfangreichen Beiträgen in Weblogs.[38]
- Verarbeitung von umgangssprachlichen Äußerungen
  In Web 2.0-Angeboten werden Meinungen und Beiträge von Benutzern selbst erfasst, oftmals sind diese Texte grammatikalisch falsch, mit umgangssprachlichen oder internet- oder kontextspezifischen Begriffen und Wörtern gespickt. Vielfach setzen Autoren mit ihren Ansätzen aber „saubere“ Texte für die Bewertung voraus.[39] Weiters sind beispielsweise in Microblogging-Plattformen wie Twitter wenig Kontextinformationen vorhanden und es werden Ironie, Sarkasmus, Emoticons und verschiedenste Abkürzungen verwendet, wodurch wiederum das Opinion Mining erschwert wird.[40]

[35] Vgl. Mariampolski, 2001, 7ff.
[36] Vgl. Kaiser, 2009, S. 90
[37] Vgl. Pang/Lee, 2008, 16ff.
[38] Vgl. Maynard/Bontcheva/Rout, 2012
[39] Vgl. Shimada/Endo, 2008, S. 1007
[40] Vgl. Maynard/Bontcheva/Rout, 2012

- Deutsche Sprache
  Ein Großteil der veröffentlichten Forschungsergebnisse fokussiert auf die englische, chinesische oder arabische Sprache; manche der im Opinion Mining verwendeten Algorithmen können aber nicht beliebig auf andere Sprachen übertragen werden.[41,42]
- Relevanz von Inhalt
  Mithilfe von Web Crawlern können Informationen aus dem Web gesammelt werden, aber selbst wenn Web Crawler auf bestimmte Themen oder Domains eingeschränkt werden, bedeutet das nicht zwangsläufig, dass jede einzelne gefundene Seite auch tatsächlich relevant ist bzw. jeder Bereich der Seite relevant ist. Üblicherweise sind Webseiten aus mehreren inhaltlichen Komponenten aufgebaut (wie Navigationselemente, Vorschau auf andere Artikel, Infoboxen für Veranstaltungen, etc.), die aber für das Opinion Mining wenig relevant sind.[43,44]
- Umgang mit Spam
  Meinungen aus den Social Web Plattformen werden immer häufiger für Kaufentscheidungen, für die Gestaltung von Produktdesigns, etc. verwendet; positive Meinungen bedeuten daher häufig bessere Chancen für den Verkauf, für die Kundenakzeptanz, etc. Dieses System kann mit falsch verbreiteten Meinungen ausgenutzt werden, um gezielt Produkte, Services, Organisationen oder Personen zu promoten oder den Mitbewerb durch negative Meinungen zu verunglimpfen. Die automatisierte Erkennung von „Opinion Spam“ ist äußert schwierig.[45,46]

[41] Vgl. Chen/Zimbra, 2010
[42] Vgl. Pang/Lee, 2008, S. 42
[43] Vgl. Maynard/Bontcheva/Rout, 2012
[44] Vgl. Yi/Liu, 2003
[45] Vgl. Jindal/Liu, 2008
[46] Vgl. Jindal/Liu, 2007

1.2 Ziel und Forschungsinteresse

Aus obigen Ausführungen wird evident, dass es noch zahlreiche Herausforderungen im Opinion Mining zu bewältigen gibt, um einen unternehmensrelevanten Einsatz zur Analyse der Meinungen gewährleisten zu können. Um den Herausforderungen systematisch zu begegnen, ist das zentrale *Ziel dieser Arbeit, ein Vorgehensmodell für das Opinion Mining für Web 2.0-Quellen zu entwickeln und dieses als Software-Prototyp umzusetzen und zu evaluieren. Der Fokus wird dabei auf die Sichtweise von Unternehmen gelegt, d.h. es stehen benutzergenerierte Meinungen rund um Produkte und Marken von Unternehmen im Vordergrund; die politische Diskussion sowie deren Behandlung im Opinion Mining wird nicht näher betrachtet.*

Die Forschungsfrage lautet:

Wie kann ein Vorgehensmodell zum Opinion Mining auf Basis einer Analyse von Web 2.0-Quellen sowie von Techniken und Methoden zum Opinion Mining systematisch konzipiert und in einem Software-Prototyp zur Anwendung gebracht werden, um Meinungen und Stimmungen aus Web 2.0-Quellen aus der Perspektive eines Unternehmens zu extrahieren und zu analysieren?

Obige globale Forschungsfrage wird in folgende Fragestellungen untergliedert:

- *Fragestellung 1: Welche Web 2.0-Quellen können als Basis für Opinion Mining herangezogen werden und welche Besonderheiten ergeben sich daraus?*
  Ziel der Fragestellung ist es, zu klären, welche Quellen im Web 2.0 für Meinungen von Web-Benutzern identifiziert werden können und die Besonderheiten dieser Quellen sollen aufgezeigt werden. Als Ergebnis wird ein Überblick über mögliche Web 2.0-Quellen und deren Besonderheiten erwartet.

- *Fragestellung 2: Welche Methoden und Algorithmen sind für die effektive Durchführung des Opinion Minings notwendig?*
  Ziel der Fragestellung ist es, eine umfassende Analyse, Evaluierung und Aus-

wahl von bestehenden Methoden und Algorithmen zur Bewertung des Sentiments in Textdaten durchzuführen. Als Ergebnis wird ein Überblick über Methoden und Algorithmen sowie eine Bewertung der Eignung und Einsetzbarkeit der unterschiedlichen Methoden und Algorithmen in Bezug auf die Sentimentanalyse nach quantitativen und qualitativen Indikatoren erwartet.

- *Fragestellung 3: Wie können die Methoden und Algorithmen des Opinion Minings in einem durchgängigen Vorgehensmodell systematisiert werden?*
  Ziel der Fragestellung ist es, ein Vorgehensmodell zum Opinion Mining und der adäquaten Methoden und Algorithmen zu konzipieren. Als Ergebnis wird ein Vorgehensmodell erwartet.

- *Fragestellung 4: Wie können die identifizierten Methoden und Algorithmen in einem Software-Prototyp implementiert werden?*
  Ziel der Fragestellung ist es, das zuvor konzipierte Vorgehensmodell in einem Software-Prototyp zu implementieren. Das erwartete Ergebnis ist ein Software-Prototyp.

- *Fragestellung 5: Wie kann der Software-Prototyp zur Anwendung gebracht werden und wie kann die Validität des im Prototyp zur Anwendung gebrachten Vorgehensmodells überprüft werden?*
  Ziel der Fragestellung ist es, die Validität des Vorgehensmodells am Beispiel einer konkreten Web 2.0-Quelle zu überprüfen. Als Ergebnis wird eine qualitative Evaluierung des Vorgehensmodells und der eingesetzten Algorithmen und Methoden erwartet.

1.3 Thematische Einordnung und Lösungsmethodik

Die vorliegende Arbeit ist der wissenschaftlichen Disziplin der Wirtschaftsinformatik zuzuordnen. Kerngebiet der Wirtschaftsinformatik ist die Entwicklung und das Management von betrieblichen Informationssystemen. Die Kernaufgaben der

Wirtschaftsinformatik sind die Erklärung und Gestaltung ihres Gegenstandsbereichs „Informationssysteme“.[47] Im anglo-amerikanischen Raum steht die „Information Systems Research“ im Vordergrund, die in der Literatur teilweise gleich zur Wirtschaftsinformatik gesetzt wird[48]; dennoch sind Unterschiede zu erkennen: während in der Wirtschaftsinformatik die technische Seite stärker betont wird, stehen in der Information Systems Research die sozialen Aspekte von Informationssystemen stärker im Vordergrund.[49]

Das Forschungsziel dieser Arbeit ist die Konzeption eines Vorgehensmodells und die Anwendung des Vorgehensmodells in einem Software-Prototypen. Österle et al. halten fest, dass die Ergebnistypen der gestaltungsorientierten Wirtschaftsinformatik unter anderem Konzepte, Modelle, Methoden und Implementierungen von Lösungen als Prototypen sind.[50] Vor dem Hintergrund dieser Feststellung liegt die Zuordnung des Forschungsansatzes dieser Arbeit zur gestaltungsorientierten Wirtschaftsinformatik auf der Hand. Österle et al. führen als typische Forschungsmethoden in der Analysephase Umfragen, Fallstudien und Tiefeninterviews mit Expert/innen an. In der Phase des Entwurfs von Artefakten kommen häufig die Konstruktion von Prototypen und die Modellierung mit Werkzeugen vor, die Evaluierung der Artefakte kann typischerweise mit Hilfe von Laborexperiment, Pilotierung (Anwendung des Prototyps), Feldexperiment oder Prüfung durch Expert/innen vorgenommen werden.[51] Im Rahmen der Design Science Research werden IT-Artefakte konstruiert und evaluiert. Unter IT-Artefakten werden Konstrukte, Modelle, Methoden und Instanziierungen verstanden, um Informationssysteme zu verstehen und entwickeln zu können. Design Science Research erfordert 1) die Identifikation und Beschreibung eines IT-Problems, 2) den Nachweis, dass keine adäquaten Lösungen vorhanden sind, 3) die Konstruktion eines IT-Artefakts, 4) eine Evaluierung des Artefakts, 5) eine Beschreibung des Erkenntnisgewinns und 6) eine Erklärung der Auswirkung auf das IT-

[47] Vgl. Heinrich/Heinzl/Roithmayr, 2007, 15ff.
[48] Vgl. Lehner/Zelewski, 2007, 71f.
[49] Vgl. Stahl, 2009, S. 115
[50] Vgl. Österle et al., 2010, S. 667
[51] Vgl. Österle et al., 2010, S. 668

Management.[52] Hevner et al. schlagen detaillierte Richtlinien und Methoden zur Durchführung von Design Science Research vor.[53] Wilde/Hess untersuchen die Forschungsmethoden der Wirtschaftsinformatik und zeigen, dass argumentativ-, konzeptionell- und formal-deduktive Analysen, Prototyping, Fallstudien und quantitative Querschnittanalysen vorwiegend zum Einsatz kommen.[54]

In der Arbeit werden verschiedene Methoden aus der gestaltungsorientierten Wirtschaftsinformatik eingesetzt, um die Zielstellung zu erreichen. Im Folgenden werden die Struktur und die Lösungsmethodik der Untersuchung vorgestellt, welche aus den Fragestellungen abgeleitet werden.

Fragestellung 1 – Identifikation und Analyse von Web 2.0-Quellen
Zur Identifikation von möglichen Web 2.0-Quellen für Sentimentanalysen wird sowohl bestehende Literatur recherchiert und analysiert als auch eine Web-Recherche und -Analyse durchgeführt. Ziel der Literaturrecherche und -analyse ist die fundierte Erschließung und Aufarbeitung des Forschungskontexts. Um die Web 2.0-Quellen und deren Charakteristika analysieren zu können, müssen Web-Recherchen und -Analysen sowie eine empirische Erhebung durchgeführt werden. In der empirischen Erhebung wird ermittelt, welche Unterschiede zwischen den für das Opinion Mining relevanten Web 2.0-Quellen in Bezug auf auswertungsrelevante Elemente bestehen. Die empirische Erhebung wird im Kapitel 3.3.2 ausführlich dargestellt und erläutert.

Fragestellung 2 – Methoden und Algorithmen zur Durchführung von Opinion Mining
Eine Vielzahl von Methoden und Algorithmen können zur Analyse von Meinungen herangezogen werden. Diese werden in der Literatur recherchiert und anschließend nach quantitativen und qualitativen Kriterien evaluiert, um die Eignung und Einsetzbarkeit in Bezug auf die Besonderheiten unterschiedlicher Web 2.0-

[52] Vgl. March/Storey, 2008, S. 726
[53] Vgl. Hevner et al., 2004, 82ff.
[54] Vgl. Wilde/Hess, 2007

Quellen bzw. der deutschen Sprache beurteilen zu können. Je nach Methode werden unterschiedliche Kriterien angewendet (wie Precision, Recall, Laufzeitperformanz, etc.). Ausgangspunkt für die Literaturrecherche sind die wissenschaftlichen Datenbanken ACM Digital Library, Springer Link, EBSCO und IEEE Xplore Digital Library, scholar.google.com sowie die „klassische“ Bibliothekssuche. Als Suchbegriffe wurden die Begriffe „Opinion Mining“, „Sentiment Analysis“, „Sentiment Classification“ verwendet.

Fragestellung 3 – Erstellung eines Vorgehensmodells

Um ein Vorgehensmodell konzipieren zu können, werden auf Basis einer Literaturrecherche und -analyse zuerst der Begriff sowie die grundlegenden Elemente eines Vorgehensmodells geklärt. Anschließend werden die in Fragestellung 1 identifizierten Web 2.0-Quellen und deren Besonderheiten mit den in Fragestellung 2 ermittelten Algorithmen und Methoden zu einem Vorgehensmodell zur Sentimentanalyse systematisiert. Dies erfordert eine Zusammenführung der gewonnen Erkenntnisse aus der wissenschaftlichen Literatur und der empirischen Erhebung. Das Ergebnis dieser Zusammenführung ist ein Vorgehensmodell zur Durchführung von Opinion Mining für Web 2.0-Quellen. Das Vorgehensmodell beinhaltet dabei Phasen bzw. Aufgaben zum Opinion Mining sowie mögliche, in der jeweiligen Phase einsetzbare Methoden und Algorithmen zur Verarbeitung der gewonnenen Daten.

Fragestellung 4 – Prototyp

Das zuvor konzipierte Vorgehensmodell wird in einem Software-Prototypen abgebildet. Die Vorgehensweise des Prototypings entspricht der des experimentellen Prototypings. Ziel des Prototypings ist es, bestimmte Problemlösungen zu beurteilen und die Tauglichkeit von Architekturmodellen und Lösungsideen für einzelne Systemkomponenten nachzuweisen. Die Vorgehensweise beim experimentellen Prototyping ist wie folgt: Ausgehend von ersten Vorstellungen über das System wird ein Prototyp entwickelt, der es erlaubt, anhand von Anwendungsbeispielen