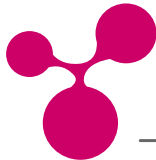


Technische Universität Dresden – Fakultät Informatik
Professur für Multimediaetechnik, Privat-Dozentur für Angewandte Informatik

Prof. Dr.-Ing. Klaus Meißner
PD Dr.-Ing. habil. Martin Engelen
(Hrsg.)



GENEME '10

GEMEINSCHAFTEN IN NEUEN MEDIEN

an der
Fakultät Informatik der Technischen Universität Dresden

mit Unterstützung der

3m5. Media GmbH, Dresden
ANECON Software Design und Beratung GmbH, Dresden
Communardo Software GmbH, Dresden
GI-Regionalgruppe, Dresden
itsax.de | pludoni GmbH, Dresden
Kontext E GmbH, Dresden
Medienzentrum der TU Dresden
objectFab GmbH, Dresden
SALT Solutions GmbH, Dresden
SAP AG, Resarch Center Dresden
Saxonia Systems AG, Dresden
T-Systems Multimedia Solutions GmbH, Dresden

am 07. und 08. Oktober 2010 in Dresden

www.geneme.de
info@geneme.de

D.6 Analyse der Meinungsentwicklung in Online Foren – Konzept und Fallstudie

*Carolyn Kaiser, Freimut Bodendorf
Universität Erlangen-Nürnberg, Lehrstuhl Wirtschaftsinformatik II*

Abstract

Das Web 2.0 ist u. a. auch eine weltweite Plattform für Meinungsäußerungen. Immer mehr Kunden diskutieren online über Produkte und tauschen Erfahrungen aus. Die Analyse der Online Beiträge stellt daher ein wichtiges Marktforschungsinstrument dar. Es wird ein Ansatz zur automatischen Identifikation, Aggregation und Analyse von Meinungen mittels Text Mining vorgestellt und dessen Anwendung an einem Beispiel aus der Sportartikelindustrie aufgezeigt.

1 Motivation

Im Web 2.0 vernetzen sich immer mehr Kunden und diskutieren über Produkte. Dabei beeinflussen sie gegenseitig ihre Meinungen und Kaufentscheidungen. Für Unternehmen stellt die wachsende Anzahl an Kundenrezensionen eine wichtige Informationsquelle dar. Eine riesige Menge an Online Meinungen steht kostenlos und tagesaktuell zur Verfügung und liefert ein kontinuierliches Stimmungsbild. Die Analyse der Meinungsentwicklung ermöglicht es Trends, Chancen und Risiken frühzeitig zu erkennen und entsprechende Marketingmaßnahmen einzuleiten.

Es wird ein Ansatz zur Analyse der Meinungsentwicklung im Internet vorgestellt und an einem Fallbeispiel aus der Sportartikelbranche demonstriert. Hierbei werden Meinungen zu Produkten in Online Foren mittels Text Mining identifiziert und zu einem Gesamtindex aggregiert. Darauf aufbauend erfolgt eine Analyse der Entwicklung des Meinungsbildes. Diese Entwicklung gibt wiederum Aufschluss über die Auswirkung von Werbemaßnahmen und externen Ereignissen. Konkurrenzvergleiche können durchgeführt und der Einfluss auf den Absatz gemessen werden. Nachfolgend werden zunächst themenverwandte Arbeiten vorgestellt. Anschließend wird ein Überblick zu dem Ansatz gegeben und die einzelnen Bestandteile, die Meinungsidentifikation, -aggregation und -analyse näher erläutert.

2 Themenverwandte Arbeiten

Text Mining zielt darauf ab, Muster und Zusammenhänge in Texten zu erkennen. Während früher der Fokus der Text Mining-Forschung auf Fakten lag, stehen heute Meinungen im Zentrum des Interesses. Viele Arbeiten beschäftigen sich mit der Aufdeckung von Meinungen im Web 2.0. Dave et al. 2003, Liu et al. 2005 sowie Popescu und Etzioni 2007 stellen zum Beispiel Methoden vor, die die Identifikation und Analyse von Meinungen in Online Reviews ermöglichen. Hierbei handelt es um eine statische Betrachtung der Meinungen.

Einige Forschungsarbeiten beobachten die zeitliche Entwicklung von Meinungen im Web 2.0. Gruhl et al. 2004, Kempe et al. 2003 sowie Ku et al. 2006 studieren die dynamische Verbreitung von Diskussionsthemen, Ideen und Meinungen im Web 2.0. Externe Ereignisse werden jedoch nicht berücksichtigt.

Viele Forscher befassen sich mit der Aufdeckung von Ereignissen in Texten. Smith 2002, Kumeran und Allan 2004 sowie Yang et al. 1998 entwickeln Methoden, die das automatische Erkennen von neuen Ereignissen in Texten erlauben. Zusammenhänge zwischen Ereignissen und anderen Inhalten im Text werden hierbei nicht untersucht.

Es findet sich eine Reihe von Projekten, die die Korrelation zwischen Online Kommunikation und Aktienkursen analysieren. Choudhury et al. 2008, Antweiler und Frank 2004 sowie Tumarking und Whitelaw 2001 finden einen Zusammenhang zwischen Kommunikationsaktivität und Aktienkursbewegung.

Daneben gibt es Arbeiten, die den Zusammenhang zwischen Online Kommunikation und Absatzzahlen erforschen. Chen et al. 2004 und Gruhl et al. 2005 stellen eine positive Korrelation zwischen der Anzahl von Kundenreviews zu Büchern und Amazon's Absatzranking fest. Tong 2001, Mishne und Glance 2006 sowie Liu et al. 2007 zeigen, dass der Absatz von Kinokarten durch Kundenmeinungen aus Filmreviews prognostiziert werden kann.

Eine umfassende zeitliche Analyse der Meinungsentwicklung unter Berücksichtigung von Werbung, externen Ereignissen, Konkurrenzprodukten und Absatzzahlen, die für die Einschätzung der Meinungsbildung zu einem Produkt entscheidend ist, fehlt jedoch.

3 Ansatz

Der Ansatz zur Analyse der Meinungsentwicklung in Internetforen erfolgt in einem dreistufigen Prozess. Im ersten Schritt werden die in den Posts enthaltenen Meinungen zu den betrachteten Produkten mittels Text Mining als positiv, negativ oder neutral klassifiziert. Anschließend werden die einzelnen Meinungen zu einem Index aggregiert, der die Gesamtmeinung des Forums zu einem Produkt in einem Zeitraum widerspiegelt. Im letzten Schritt wird die zeitliche Entwicklung der Gesamtmeinung analysiert. Hierbei wird die Auswirkung von Werbemaßnahmen und externen Ereignissen auf die Meinungsentwicklung ermittelt, ein Vergleich mit der Meinungsentwicklung des Konkurrenzproduktes durchgeführt und der Einfluss der Meinung auf den Absatz untersucht. Ein kontinuierliches Meinungstracking ermöglicht die Erkennung von Chancen und Risiken sowie die Einleitung von entsprechenden Marketingmaßnahmen. Der Ansatz wird beispielhaft für den Fußballschuh Predator von adidas und das Konkurrenzprodukt Mercurial von Nike aufgezeigt. Hierzu wurden 407 Posts aus dem Jahr 2007 von dem Diskussionsforum fussball-forum.de heruntergeladen.

4 Identifikation von Meinungen

Ziel ist es, Produkte und ihre Bewertungen in Posts zu erkennen. Hierbei werden positive, negative und neutrale Bewertungen unterschieden. Die Identifikation von Produkten und Bewertungen kann als Klassifikationsaufgabe verstanden werden (siehe Abbildung 1). Textpassagen aus Posts werden auf Grund ihrer linguistischen Eigenschaften in Klassen eingeteilt. Wörter werden dahingegen klassifiziert, ob sie Produkte bezeichnen. Sätze werden gemäß ihrer Polarität den Klassen positiv, negativ oder neutral zugeordnet. Durch die Verwendung von Sätzen können sprachliche Phänomene wie Ironie und Verneinung besser berücksichtigt und somit Bewertungen besser klassifiziert werden.

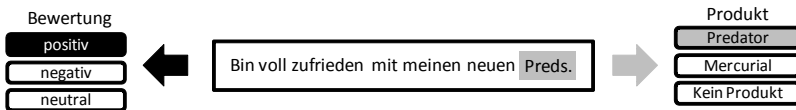


Abbildung 1: Klassifikation von Produkten und Bewertungen

Die Klassifikation basiert auf den linguistischen Attributen der Posts, die bei einer linguistischen Analyse gewonnen werden. Hierbei werden die Posts in Sätze und Worte zerlegt und die grammatikalischen Funktionen der Wörter identifiziert.

Zur Klassifikation von Wörtern und Sätzen wird die Stützvektormethode (Cortes & Vapnik 1995) gewählt, da sie die Verarbeitung einer Vielzahl von Attributen ermöglicht und gute Ergebnisse in ähnlichen Arbeiten (z.B. Pang et al. 2002) geliefert hat.

Die Stützvektormethode erlaubt das Lernen von binären Klassifikationsregeln. Im Fall von mehreren Klassen, wird für jede Klasse eine eigene Klassifikationsregel erzeugt, die angibt, ob eine Textpassage zu dieser Klasse gehört. Um beispielsweise einen Satz nach seiner Polarität zu klassifizieren, werden drei Regeln erlernt: positiv versus nicht positiv, negativ versus nicht negativ und neutral versus nicht neutral. Die endgültige Klasse wird anschließend über einen Mehrheitsentscheid zugeordnet. Die Stützvektormethode benötigt als Input Trainingsdatensätze, bestehend aus Textpassagen mit ihren linguistischen Attributen und den zugeordneten Klassen. Auf Basis der Trainingsdatensätze werden die bestmöglichen Klassifikationsregeln bestimmt. Im zweidimensionalen Raum kann die Klassifikationsregel als Gerade dargestellt werden, die die Textpassagen in zwei Klassen teilt. Abbildung 2 zeigt wie ein Satz auf Basis seiner enthaltenen Wörter in die Klasse positiv eingeordnet wird.

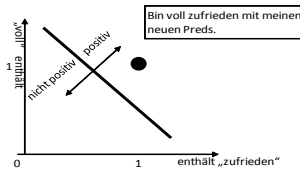


Abbildung 2: Klassifikation mittels Stützvektormethode

Die Evaluierung der erlernten Klassifikationsregeln erfolgt mittels einer fünffachen Kreuzvalidierung. Hierbei wird der Trainingsdatensatz in fünf gleich große Teile gesplittet. In fünf verschiedenen Validierungsdurchläufen werden jeweils vier Teile zum Lernen und ein Teil zum Testen verwendet. Am Ende wird die durchschnittliche Performanz der fünf Durchläufe mittels der Kennzahlen Precision und Recall (Weiss et al. 2005) berechnet. Während die Precision die Genauigkeit der Klassifikationsregeln misst, charakterisiert der Recall die Vollständigkeit.

Im Rahmen dieser Fallstudie wurden 407 Posts bestehend aus 2095 Sätzen aus dem Diskussionsforum fussball-forum.de extrahiert und die darin erwähnten Fußballschuhe Predator und Mercurial mit ihren Bewertungen manuell annotiert. Tabelle 1 zeigt die Performanz, die bei der Anwendung der Stützvektormethode erreicht wurde. Produktbezeichnungen können sehr gut, Bewertungen gut identifiziert werden. Die relativ geringe Anzahl an Wortvarianten zur Bezeichnung der Schuhe vereinfacht die Lernaufgabe. Das Erlernen von Polaritäten ist im Vergleich zu Produktbezeichnungen wesentlich schwieriger. Positive, negative und neutrale Meinungen können mit einer Vielzahl an Sätzen ausgedrückt werden.

Tabelle 1: Klassifikationsperformanz

	Precision	Recall
Produkt	97,56%	94,38%
Polarität	73,26%	73,27%

5 Aggregation von Meinungen

Die Analyse der Meinungsentwicklung setzt eine Aggregation der mit Hilfe von Text Mining identifizierten Einzelmeinungen voraus. Zu diesem Zweck wird ein Meinungsindex m gebildet, der die Anzahl der positiven, negativen und neutralen Meinungsäußerungen in einen Zeitraum t für ein Produkt i zu einer Gesamtmeinung zusammenfasst. Der Index errechnet sich mit folgender Formel:

$$m_i^t = \frac{w * pos_i^t - w * neg_i^t + neu_i^t}{\sum_{i=1}^n (w * pos_i^t - w * neg_i^t + neu_i^t)}$$

Mit:
m: Meinungsindex
pos: Anzahl der positiven Meinungen
neg: Anzahl der negativen Meinungen
neu: Anzahl der neutralen Meinungen
w: Gewichtungsfaktor
i: Produkt
t: Zeitraum

Der Zähler drückt die gewichtete Gesamtmeinung zu einem Produkt *i* im Zeitraum *t* aus. Im Nenner wird die Gesamtmeinung über alle betrachteten Produkten aufsummiert, um eine Vergleichbarkeit zwischen den verschiedenen Produkten zu erreichen. Der Wertebereich des Meinungsindex ist auf das Intervall [-1; +1] normiert. In der vorliegenden Fallstudie wird der Meinungsindex für die Fußballschuhe Predator und Mercurial auf Monatsbasis berechnet. Es wird der Gewichtungsfaktor $w=2$ gewählt, um der höheren Bedeutung von positiven und negativen Meinungen im Vergleich zu neutralen Meinungen für die Gesamtmeinung Rechnung zu tragen.

6 Zeitliche Analyse von Meinungen

Die zeitliche Analyse der aggregierten Meinungen ermöglicht wichtige Erkenntnisse für Unternehmen. So kann die Auswirkung von Werbung und externen Ereignissen auf die Meinungsentwicklung gemessen, ein Konkurrenzvergleich hergestellt und der Einfluss auf den Absatz bestimmt werden.

6.1 Auswirkung von Werbung und externen Ereignissen

Ziel ist es, die Meinungsentwicklung des beworbenen Produktes zu beobachten und die Auswirkungen von Werbekampagnen und externen Ereignissen auf die Meinung zu untersuchen. Ein Meinungstracking ermöglicht es, negative Meinungsentwicklungen aufzudecken und mit Marketingmaßnahmen entgegenzusteuern. Besonders bei wichtigen externen Ereignissen ist es von Bedeutung, die Meinung der Kunden zu erfassen, um auf mögliche Veränderungen reagieren zu können. Darüber hinaus dient es der Erfolgskontrolle von Werbemaßnahmen, so dass mögliche Schwachstellen erkannt und behoben werden können.

Mithilfe von Liniendiagrammen wird die Meinungsentwicklung visualisiert. Hierbei können Veränderungen im Vergleich zu vorhergehenden Zeitpunkten einfach erkannt werden. Abbildung 3 zeigt den Verlauf der Meinungsentwicklung zum adidas Fußballschuh Predator für das Jahr 2007 im Forum fussball-forum.de.

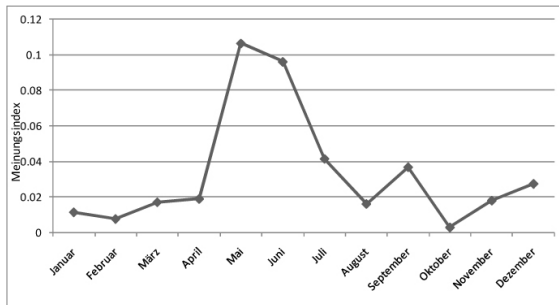


Abbildung 3: Meinungsentwicklung 2007 - Predator

Die Meinungsentwicklung weist einen engen Zusammenhang mit den Ereignissen der deutschen Fußballliga auf. So fällt das Stimmungshoch im Mai mit dem Ende der Bundesliga sowie den Finalspielen des DFB Pokals und der Champions League zusammen. Bemerkenswert hierbei ist, dass die beiden Gewinnermannschaften von DFB Pokal und Champions League jeweils von adidas ausgestattet wurden. Dies könnte die positive Meinung zu adidas' Fußballschuh Predator zusätzlich verstärkt haben. Das Absinken der Stimmung in Juni und Juli geht mit der Fußball-Sommerpause einher. Ab Ende August beginnt die nächste Saison, was sich in einem erneuten Stimmungsanstieg im September niederschlägt.

Auch die Werbeaktivität von adidas liefert einen Erklärungsbeitrag zur Meinungsentwicklung. Der Anstieg der Stimmung im März, Mai und November lässt sich auf Werbekampagnen zurückführen. So startete im März die „Impossible Is Nothing 2007“-Kampagne. Im Mai erreichte die „Predator versus F50“-Kampagne ihren Höhepunkt und im November begann die „DFB Kit Launch“-Kampagne mit großen Auftaktveranstaltungen für die Fußball-Europameisterschaft 2008.

Das Schalten von Werbekampagnen zu besonderen externen Ereignissen kann die Werbewirksamkeit verstärken. Der Stimmungsspeak im Mai lässt sich gut mit der Durchführung der „Predator versus F50“-Kampagne zum Ende der Spielsaison erklären.

Der Einfluss der Fußballereignisse und der Werbekampagnen auf die Meinungsänderung kann durch eine lineare Regressionsanalyse (Hartung et al. 2009) gemessen werden. Sie ermöglicht es, die funktionale Abhängigkeit zwischen Variablen zu bestimmen und mittels Parametern zu quantifizieren. Im vorliegenden Anwendungsfall ergibt sich ein standardisierter Parameterwert von 0,6 für die Fußballereignisse und von 0,49 für die Werbemaßnahmen. Die Parameter sind signifikant mit einer Irrtumswahrscheinlichkeit kleiner als 5%. Das R^2 , das die Güte des Regressionsmodells beschreibt, weist einen Wert von 0,73 auf. Das bedeutet, dass 73% der Variation der Meinungsänderung

durch die Variation der Ereignisse und der Werbung erklärbar sind. Dies zeigt, dass die Ereignisse der Fußballliga und die Werbemaßnahmen einen nicht unwesentlichen Einfluss auf die Meinungsänderung im betrachteten Forum haben.

6.2 Wettbewerbsvergleich

Ziel des Wettbewerbsvergleichs ist es, die Meinungsentwicklung des eigenen Produkts der des Konkurrenzproduktes gegenüberzustellen. Dadurch können zum einen Wettbewerbsvor- und -nachteile erkannt und zum anderen Maßnahmen zur Produktgestaltung oder Werbung abgeleitet werden.

Die Visualisierung der Wettbewerbsanalyse erfolgt mithilfe von Liniendiagrammen. Abbildung 4 stellt die Entwicklung des Meinungsindex für den Fußballschuh Predator von adidas und das Konkurrenzprodukt Mercurial von Nike für das Jahr 2007 dar. Insgesamt gesehen, lässt sich feststellen, dass der Predator mit Ausnahme vom Februar in allen Monaten positiver bewertet wurde als der Mercurial.

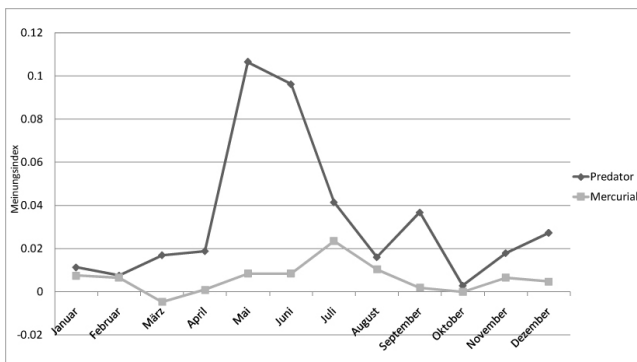


Abbildung 4: Vergleich der Meinungsentwicklung

Es zeigen sich auch wichtige Unterschiede bei der Auswirkung von externen Ereignissen. Während die Stimmung zu adidas' Predator von nationalen Fußballereignissen beeinflusst wird, folgt die Stimmung zu Nike's Mercurial eher internationalen Fußballereignissen. So fällt das Stimmungshoch zum Mercurial im Juli mit der Copa America, der südamerikanischen Fußballmeisterschaft, zusammen. Der Sieg der brasilianischen Mannschaft dürfte die positive Meinung noch zusätzlich verstärkt haben. Denn zum einen wird die brasilianische Mannschaft von Nike ausgestattet und zum anderen ist der brasilianische Nationalspieler Ronaldo Werbeträger für den Mercurial.

6.3 Einfluss auf den Absatz

Zielsetzung der Absatzanalyse ist es, zu untersuchen, ob die Meinung in Foren zu einem Produkt mit dem Absatz zusammenhängt. Damit lässt sich bestimmen, welche Bedeutung das Meinungsbild der virtuellen Welt für den realen Verkaufserfolg eines Unternehmens hat. Besteht zwischen Meinungsbild und Absatz eine Korrelation mit zeitlicher Verzögerung, so können Unternehmen Chancen und Risiken frühzeitig erkennen und entsprechend darauf reagieren.

Die Absatzanalyse erfolgt zunächst durch den visuellen Vergleich der Meinungs- und Absatzkurven mithilfe von Liniendiagrammen. Abbildung 5 stellt die Entwicklung der Meinung und des Absatzes für den Fußballschuh Predator im Jahr 2007 dar. Bei näherer Betrachtung ist festzustellen, dass beide Kurven einen zeitlich verzögerten ähnlichen Verlauf aufweisen. So geht zum Beispiel das Stimmungshoch im Mai/Juni mit einem Absatzhoch im August/September einher. Ebenso führt das Stimmungstief im August zu einem Absatztief im November. Dies legt den Schluss nahe, dass sich die Stimmung im Forum mit einer zeitlichen Verzögerung von drei Monaten im Absatz auswirkt.

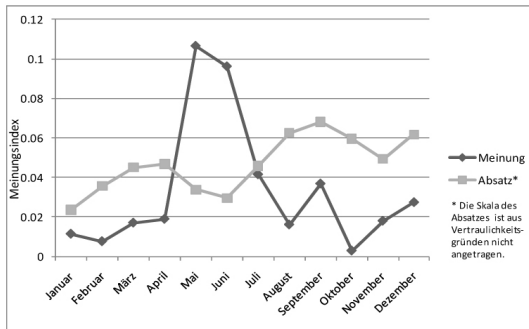


Abbildung 5: Vergleich der Entwicklung von Meinung und Absatz - Predator

Die Abhängigkeit der zeitlichen Verläufe von Meinung und Absatz kann mithilfe der Kreuzkorrelation (Hartung et al. 2009) evaluiert werden. Die Kreuzkorrelation ermöglicht es, eine Zeitreihe als prognostischen Indikator für die Werte einer anderen Zeitreihe zu erkennen. Sie misst die Korrelation der Beobachtungen zweier Zeitreihen bei unterschiedlicher Zeitverschiebung.

Eine Evaluierung des Zusammenhangs mithilfe des Kreuzkorrelationskoeffizienten liefert einen signifikanten (Irrtumswahrscheinlichkeit kleiner als 5%) Wert von 0,61 für den zeitlichen Versatz von drei Monaten. Damit ist für den Predator die positive Korrelation der Meinung zum Absatz mit einem zeitlichen Abstand von drei Monaten bestätigt. Der gleiche Zusammenhang besteht auch für die Meinung und den Absatz des Konkurrenzproduktes Mercurial. Hier ergibt die statistische Auswertung einen signifikanten Kreuzkorrelationskoeffizienten von 0,58 für die zeitliche Differenz von drei Monaten. Die Stimmung im Forum scheint sich mit zeitlicher Verzögerung von drei Monaten auf den Absatz niederzuschlagen.

7 Zusammenfassung und Ausblick

Die steigende Anzahl an Kundenmeinungen im Web 2.0 stellt eine reiche Informationsquelle für Unternehmen dar. Der vorgestellte Ansatz ermöglicht es, durch den Einsatz von Text Mining Meinungen im Web 2.0 zu identifizieren, zu einem Meinungsindex zu aggregieren und dessen zeitliche Entwicklung zu analysieren. Hierbei wird die Auswirkung von Werbemaßnahmen und externen Ereignissen, die Meinung zu Konkurrenzprodukten sowie der Einfluss auf den Absatz berücksichtigt. Die Anwendung auf das Fallbeispiel zeigt die Bedeutung einer dynamischen Meinungsanalyse im Web 2.0 für Unternehmen auf.

Das vorgestellte Verfahren stellt ein Basiskonzept dar, das noch erweitert wird. Ziel ist es, ein Frühwarnsystem zu entwickeln, das Chancen und Risiken auf Basis der Meinung zu einem Produkt, der Meinung zum Konkurrenzprodukt sowie Informationen zu externen Ereignissen, Werbemaßnahmen und Absatz automatisch erkennt und gegebenenfalls Warnungen an Marketing Manager versendet.

Literaturverzeichnis

- Antweiler W., Frank M.Z., Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *Journal of Finance* 59 (3), 2004.
- Cortes C., Vapnik V.N., Support Vector Networks, *Machine Learning* 20, 1995.
- Dave K., Lawrence S., Pennock D. M., Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, *Proc. of the 12th Int'l Conf. World Wide Web, ACM, Budapest 2003*, S. 519-528.
- Chen P.-Y., Wu S.-Y., Yoon J., The impact of online recommendations and consumer feedback on sales, *Proc. of the 25th Int'l Conf. Information Systems*, 2004.
- Choudhury M. D., Sundaram H., John A., Seligmann, D. D. Can Blog Communication Dynamics be Correlated with Stock Market Activity?, *Proc. of the 19th ACM Conf. on Hypertext and Hypermedia*, 2008, Pittsburgh.
- Etzioni O., Cafarella M., Downey D., Popescu A.M., Shaked, T., Soderland S., Weld D.S.; Yates A.: Unsupervised Named-Entity Extraction from the Web: An Experimental Study, *Artificial Intelligence* 165 (1), 2005, S. 191-134.

- Gruhl D., Guha R., Liben D., Nowell J., Tomkins A., Information Diffusion Through Blogspace, Proc. of the 13th int'l conf. WWW, 2004, N.Y.
- Gruhl D., Guha R., Kumar R., Novak J., Tomkins A., The Predictive Power of Online Chatter, Proc. of the 11th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2005, Chicago.
- Hartung J., Elpert B, Klösener K., Statistik, 15. Aufl., Oldenburg, München, 2009.
- Kempe D., Kleinberg J., Tardos E., Maximizing the Spread of Influence through a Social Network, Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, Washington, 2003.
- Ku L.-W., Liang Y.-T., Chen H.-H., Opinion extraction, summarization and tracking in news and blog corpora, AAAI Symposium on Computational Approaches to Analysing Weblogs, 2006, S. 100-107.
- Kumaran G., Allan J., Text Classification and Named Entities for New Event Detection, Proc. of the 25th ann. int'l ACM SIGIR conf., 2004, Sheffield.
- Liu B., Hu M., Cheng J., Opinion Observer: Analyzing and Comparing Opinions on the Web, Proc. of the 14th int'l conf. on World Wide Web, ACM, N.Y., 2005, S. 342-351.
- Mishne G., Glance N., Predicting Movie Sales from Blogger Sentiment, AAAI 2006 Spring Symp. on Comp. Approaches to Analysing Weblogs, 2006.
- Pang P., Lee L., Vaithyanathan S., Thumbs up? Sentiment Classification using Machine Learning Techniques, Proc. of the conf. on empirical methods in natural language processing, ACM, 2002, S. 79-86.
- Popescu A. M., Etzioni O., Extracting Product Features and Opinions from Reviews, Kao A, Poteet S. R. (Hrsg.), Natural Language Processing and Text Mining, Springer, London, 2007, S. 9-28.
- Smith D.A., Detecting and Browsing Events in Unstructured Text, Proc. of the 25th ann. int'l ACM SIGIR conf., 2002, Tampere.
- Tong R., Detecting and tracking opinions in on-line discussions. UCB/SIMS Web Mining Workshop, 2001.
- Tumarkin R., Whitelaw R.F., News or noise? internet postings and stock prices, Financial Analysts Journal, 2001, S. 41–51.
- Yang Y., Pierce T., Carbonell J., A study on retrospective and on-line event detection, Proc. of the 21st ACM Int'l Conf. on Research and Development in Information Retrieval, 1998, S. 28–36.
- Weiss S.M., Indurkha N., Zhang T., Damerau F.J., Text Mining – Predictive Methods for Analyzing Unstructured Information, Springer, N.Y., 2005.