

Technische Universität Dresden
Medienzentrum
Universität Siegen

Prof. Dr. Thomas Köhler
Prof. Dr. Nina Kahnwald
Prof. Dr. Eric Schoop
(Hrsg.)



WISSENS- GEMEINSCHAFTEN 2015

an und mit der Unterstützung der
Technischen Universität Dresden

mit Unterstützung von

BPS Bildungsportal Sachsen GmbH
Campus M21
Communardo Software GmbH
Dresden International University
eScience – Forschungsnetzwerk Sachsen
Gesellschaft der Freunde und Förderer der TU Dresden e.V.
Gesellschaft für Informatik e.V. (GI)
Gesellschaft für Medien in der Wissenschaft e.V.
intecsoft GmbH & Co. KG
Learnical GbR
Landeshauptstadt Dresden
Medienzentrum, TU Dresden
Microsoft Corporation
ObjectFab GmbH
T-Systems Multimedia Solutions GmbH
SQL Projekt AG
Universität Siegen

am 25. und 26. Juni 2015 in Dresden

www.WissensGemeinschaften.org

Bibliografische Information der Deutschen Nationalbibliothek

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

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

ISBN 978-3-95908-010-1

© 2015 TUDpress

Verlag der Wissenschaften GmbH

Bergstr. 70

D-01069 Dresden

Tel.: +49 351 47969720 | Fax: +49 351 47960819

www.tudpress.de

Gesetzt von den Herausgebern.

Druck und Bindung: Sächsisches Digitaldruck Zentrum GmbH

Printed in Germany.

Alle Rechte vorbehalten. All rights reserved.

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt. Jede Verwertung außerhalb der durch das Urheberrecht gesetzten engen Grenzen ist ohne die Zustimmung der Herausgeber unzulässig und strafbar. Das gilt insbesondere für die Vervielfältigung, Übersetzung, Mikroverfilmung und die Einspielung und Bearbeitung in elektronischen Systemen.

3 Integration von Topic Models und Netzwerkanalyse bei der Bestimmung des Kundenwertes

Kai Heinrich

*Technische Universität Dresden, Lehrstuhl für Wirtschaftsinformatik,
Business Intelligence Research*

1 Einleitung

Im Zuge der ungebremsen Ausbreitung des Web 2.0 und der längst eingetretenen Globalisierung der Märkte entwickelt sich das Wissen über die Bedürfnisse und Meinungen von Kunden zum erfolgskritischen Faktor in jedem Unternehmen. Es können jedoch nicht alle Kunden immer profitabel sein, denn nicht jeder Kunde liefert denselben Beitrag zum Unternehmenswert. Neben den offensichtlichen ökonomisch-monetären Kriterien, wie etwa Umsatz oder Deckungsbeitrag, spielen nach (Bruhn, 2002), (Cornelsen, 2000) und (Homburg & Schnurr, 1999) vermehrt auch nicht ökonomische Determinanten, wie Referenz- und Informationspotenziale eine große Rolle. Der Einfluss des Einzelnen wird durch die Weiterentwicklungen im Internet, wie sozialen Netzwerken oder Blogs ermöglicht. Das soziale Netzwerk Facebook verzeichnet über 1.42 Milliarden User weltweit. In den USA sind über 50 Prozent aller Internetnutzer bei Facebook registriert. Der Microblog Service Twitter zählt 288 Millionen Nutzer weltweit (vgl. Statista, 2015). Betrachtet man sich diese Zahlen, so lassen diese keinen Zweifel an der Integration des Web 2.0 in das tägliche Leben und somit auch in das Konsumentenverhalten aufkommen. In der vorliegenden Arbeit werden mit Hilfe der Methoden des Text Mining als Teilbereich der Business Intelligence (BI) und der sozialen Netzwerkanalyse die Referenzpotenziale einzelner Nutzer analysiert, wobei als Quelle das Verhalten der Nutzer bei der Kommunikation in sozialen Netzwerken herangezogen wird. Dabei setzt sich das Referenzpotential nicht ausschließlich aus netzwerkbezogenen Determinanten zusammen, sondern beinhaltet vielmehr auch Komponenten wie Fachwissen. Daher ist eine Integration von Inhalten und Netzwerkstrukturen nötig, um das Referenzpotential vollständig abzubilden. Das Gestaltungsziel der Arbeit besteht aus der Integration der Ansätze der sozialen Netzwerkanalyse und der Ansätze des Text-Mining um eine adäquate Beschreibung des Referenzpotentials mit Hilfe von Struktur- sowie Inhaltsdaten aus sozialen Netzwerken zu ermöglichen.

2 Determinanten des Referenzpotential

Der Kunde gilt in seiner Rolle als Abnehmer von Produkten und Dienstleistungen als wichtigstes Element einer Geschäftsbeziehung. Nach (Tomczak & Rudolf-Sipötz, 2006) kann die Beziehung eines Unternehmens zu einem Kunden als Investitionsobjekt betrachtet werden. Zur Bestimmung des Kundenwertes und des Referenzpotentials

existiert eine Vielzahl von Modellen, welche verschiedene Determinanten berücksichtigen. Aufgrund der Probleme von eindimensionalen Modellen und einigen mehrdimensionalen Vertretern (Bruhn, 2002) wird hier der ganzheitliche Ansatz von (Cornelsen, 2006) als Ausgangspunkt betrachtet. In diesem Ansatz wird der Kundenwert mit Hilfe eines Transaktions-, Informations- und Referenzwertes bestimmt. Unter dem Referenzwert wird nach (Cornelsen, 2006) dabei neben dem monetär zu bestimmenden Referenzvolumen auch das Referenzpotenzial subsumiert. Um eine Aussage über das Referenzpotenzial einer Person treffen zu können, müssen zunächst die Einflussfaktoren näher beschrieben werden.

2.1 Meinungsführerschaft

Nach (Kroeber-Riel & Weinberg, 2006) sind Meinungsführer Personen, die einen stärkeren persönlichen Einfluss ausüben als andere und somit Verhalten und Einstellungen von anderen beeinflussen können. Es handelt sich hierbei allerdings um kein dichotomes Konstrukt mit den Ausprägungen Meinungsführer ja/nein, sondern es wird vielmehr der Grad der Meinungsführerschaft betrachtet (Haseloff, 1986). Nach (Brüne, 1989) gibt es zwei hauptsächliche Einflussfaktoren für den Grad der Meinungsführerschaft einer Person: Zum einen Fachwissen und Involvement und zum anderen die Sozio-Zentralität. Letztere Eigenschaft wird nach (Brüne, 1989) auch mit Bezugspersonen-Einflusspotenzial (BEP) bezeichnet. Das Konzept des virtuellen Meinungsführers beispielweise verkörpert eine Aussage über Personen die ein sehr hohes BEP aufweisen und eine zentrale Rolle in vielen sozialen Kreisen spielen, z.B. prominente Persönlichkeiten. Weiterhin ist entscheidend, dass Meinungsführer meistens auf einen Produktbereich beschränkt sind (King & Summers, 1970) anstatt als globaler Meinungsführer tätig zu sein. Personen mit einem hohen Grad an Meinungsführerschaft verteilen nach dem Modell der zweistufigen Kommunikation (Lazarsfeld, Berelson, & Gaudet, 1944) die erhaltenen Informationen an die Masse. Daher ist der Grad der Meinungsführerschaft ein wichtiger Indikator für das Referenzpotenzial, denn je höher dieser Grad ist, desto höher ist auch das Referenzpotenzial.

2.2 Soziales Netzwerk

Nach (Iacobucci & Hopkins, 1992) ist ein soziales Netzwerk die Komposition einer großen Anzahl von Akteuren, welche durch das Muster der zu Grunde liegenden Interaktion charakterisiert ist. Dabei stehen in dieser Arbeit besonders die Akteure im Mittelpunkt und ihre Verbindung innerhalb ihres sozialen Netzes. (Granovetter, 1982) unterscheidet dabei zwei Arten der Verbindung: schwache Verbindungen (weak ties) und starke Verbindungen (strong ties). Während starke Verbindungen in kleinen kohäsiven Gruppen vorherrschen wie z.B. dem Familienkreis, so ermöglichen schwache Verbindungen die Kommunikation über diese Gruppen hinaus,

beispielsweise im Kreise der Arbeitskollegen. Nach (Brown & Reingen, 1987) ist die Art der Verbindung entscheidend für das Referenzpotenzial. Während strong ties eher bei Low-Involvement-Produkten und Affektkäufen als Referenzgeber fungieren, bieten sich weak ties für rein informative Prozesse an und sind demzufolge für High-Involvement-Produkte besser geeignet. Generell kann festgehalten werden: Umso größer das soziale Netz desto größer die Anzahl potenzieller Referenzgespräche und damit das Referenzpotenzial (Cornelsen, 2000).

2.3 Kundenzufriedenheit

Während die Meinungsführerschaft die Tiefe und das soziale Netz die Breite der Referenzbeziehung charakterisieren, gibt die Kundenzufriedenheit nach (Cornelsen, 2006) die Richtung der Referenz an. An dieser Stelle gibt es allerdings keine klaren Aussagen über den Einfluss der Zufriedenheit auf das Referenzpotenzial. Während (Anderson et al., 1994) beispielsweise eine erhöhte Referenzaktivität nach positiven Erlebnissen feststellten, konnte diese Tatsache von (Schneider, 2007) nicht bestätigt werden. (Duffy, 1994) fand erhöhte Referenzaktivität bei negativen Erlebnissen. Die Kundenzufriedenheit wird im Folgenden nicht betrachtet, obgleich zukünftige Untersuchungen, insbesondere im Bereich Opinion Mining, zeigen müssen inwiefern die Polarität einer Nachricht das Referenzpotential beeinflusst. Diese Untersuchung ist allerdings nicht Gegenstand dieses Beitrages.

3 Social Network Analysis und bisherige Ansätze

Im Rahmen der Social Network Analysis Verfahren haben sich einige Methoden zur Erfassung von Meinungsführern in Weblogs herauskristallisiert. Ein erster Ansatz ist die Übertragung des Konzepts der Network Centrality auf Weblogs (Chin et al., 2007). Ziel dieses Konzepts ist es, Communities in Weblogs zu identifizieren und die Stärke der Beeinflussung zu messen, die innerhalb dieser Communities vorherrscht. Dieser Ansatz vernachlässigt aber die Richtung der Beeinflussung, welche im Rahmen des Kommunikationsverhaltens innerhalb von sozialen Netzwerken eine große Rolle spielt. Das Konzept der Information Novelty nach (Zhang et al., 2002) basiert auf einer inhaltlichen Analyse der Blogeinträge selber. Es wird mittels einer Vergleichsfunktion die Aktualität eines Eintrages in Bezug auf ein abgegrenztes Netz ermittelt. Diese Methode greift erstmals auf inhaltliche Aspekte zurück, vernachlässigt aber die Struktur innerhalb eines sozialen Netzwerks, sowie Besonderheiten in Form von Verstärkungen (z.B. Retweets im Fall Twitter). Abschließend ist an dieser Stelle noch der Page Rank Algorithmus von (Brin et al., 1998) zu nennen. Dieser Algorithmus basiert auf der Idee eines random walk, welcher zu einem stabilen Zustand konvergiert. Anschließend können mit Hilfe der „Google Matrix“ einflussreiche Blogs identifiziert werden. Dieser Ansatz beruht auf der reinen Verlinkungsstruktur und berücksichtigt weder inhaltliche Aspekte noch

die Besonderheiten des zu Grunde liegenden Netzwerks (z.B. 140 Zeichen im Fall Twitter). Im nächsten Abschnitt wird ein Framework vorgestellt welches strukturelle sowie inhaltliche Merkmale und Besonderheit von Netzwerken berücksichtigt.

4 Influence Potential Framework

4.1 Social Network Indicator

Der Social Network Indicator (SNI) ist ein Indikator für die Größe des sozialen Netzes. Aufgrund der verzerrten Selbstrepräsentation innerhalb der meisten virtuellen sozialen Netzwerke ist es allerdings bislang schwierig zwischen strong und weak ties zu unterscheiden. Dieser Problematik kann mit Hilfe des IKI, welcher später beschrieben wird auf inhaltlicher Basis durch einen Verstärkungsfaktor entgegnet werden. Die Größe des sozialen Netzes wird zunächst durch die Anzahl der Beziehungen gemessen:

$$SNI_U \propto N_{Follow}(U) + N_{Bilateral}(U) + N_{Followers}(U)$$

Die Follower-Anzahl bestimmt die jeweiligen eingehenden Beziehungen. Es sei angemerkt, dass an dieser Stelle die Anzahl der einseitigen ausgehenden Beziehungen keine Rolle spielt, da durch diese Beziehungen kein Einfluss ausgeübt werden kann.

4.2 Social Centrality Indicator

Während der SNI die Größe des sozialen Netzes angibt, stellt der Social Centrality Indicator ein Maß für die Position innerhalb dieses Netzes dar. Der SCI gibt damit Aufschluss über die Sozio-Zentralität welche direkt mit dem Grad der Meinungsführerschaft zusammenhängt. Nach dem Modell der zweistufigen Kommunikation von (Lazarsfeld, 1944) bescheinigt auch (Brüne, 1998) Personen die geringen Einflüssen unterliegen aber großen Einfluss ausüben einen hohen Grad der Meinungsführerschaft. Nach (Everett, 2004) besitzen diese Personen ebenfalls eine hohe Sozio-Zentralität. Somit ergibt sich der SCI in Anlehnung an die Kennzahl der Betweenness Centrality als:

$$SCI_U = \sum_{j < k} \frac{g_{jk}(U)}{g_{jk}}$$

Dabei wird das Verhältnis der kürzesten Wege durch den Nutzer U zum Gesamtverhältnis bestimmt, um eine Aussage über die Zentralität des Nutzers zu treffen.

4.3 Involvement Knowledge Indicator

Während der SNI und der SCI rein strukturelle Merkmale ausnutzen um eine Aussage über das Referenzpotenzial zu treffen wird in diesem Abschnitt ein Indikator für das Fachwissen bzw. Involvement vorgestellt. Der Indikator basiert auf der Idee die Statusmeldungen nach Fachinhalten auszuwerten. Dazu muss zunächst ein gewünschtes Thema formuliert werden. Diese Eigenschaft beeinträchtigt die Aussage über das Referenzpotenzial allerdings nicht, da Meinungsführer ohnehin meist auf themen- bzw. produktspezifischer Ebene existieren. Das Thema wird dabei zunächst mit wenigen Schlüsselwörtern (Keywords) formuliert und dann mit Hilfe einer Themenlandkarte angereichert. Dabei wird für einen gewissen Zeitraum jeder Beitrag gespeichert und anschließend mit Hilfe eines Topic-Model-Verfahrens (Blei, Ng, & Jordan, 2003) analysiert. Topic Models repräsentieren dabei ein Dokument als eine Mischung aus verschiedenen Themen (vgl. Abbildung 1). Somit können Dokumente anhand von Themenvektoren verglichen werden. Themenvektoren beinhalten dabei die geschätzten Wahrscheinlichkeiten der Themen in einem Dokument. Die Gesamtheit dieser Vektoren bildet die Themenlandkarte O_j . Dasselbe Verfahren wird auf die Statusmeldungen $A_{u,t}$ des betrachteten Nutzers in einem Zeitraum T_u angewandt.

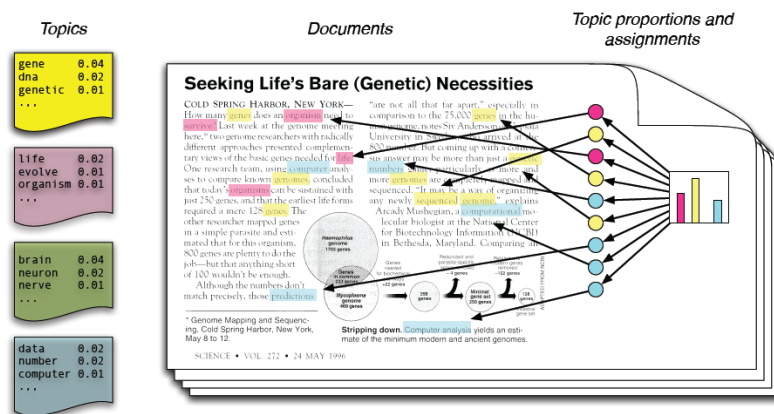


Abbildung 1: Topic Model (Blei et al. 2003)

Anschließend werden die Statusmeldungen mit der Themenlandkarte, unter Anwendung eines Vergleiches durch ein Cosinus-Ähnlichkeitsmaß (Zhang et al., 2002) abgeglichen. Je größer die Übereinstimmung, desto größer der Bezug der untersuchten Person zum Thema. Die Gesamtheit dieser Übereinstimmung gibt an wie hoch das Involvement bzw. Fachwissen einer Person in Bezug auf ein Thema j ist.

Ein zweiter Faktor ist die Anerkennung und Verbreitung der einzelnen Inhalte durch das soziale Netz. Dazu wird die Anzahl der Verstärkungen in Form von Wiederaufgriff des Inhalts durch andere Personen innerhalb des Netzwerkes $r_{u,t}$ berücksichtigt. Durch diese Verstärkung kann im Gegensatz zu den netzwerkbasierten Ansätzen eine Art Gewichtung des Beitrages innerhalb des Netzwerkes vorgenommen werden. Personen welche Beiträge oft von anderen Personen teilen, haben somit eine stärkere Bindung zueinander. Schließlich ist es möglich die Werte multiplikativ mit Parameter α abzuschwächen, falls es sich bei einer Statusmeldung selbst um eine Verstärkung handelt. In diesem Fall nimmt die Funktion $RT_{u,t}$ den Wert 1 an. Andernfalls fällt dieser Term weg, da die Indikatorfunktion bei einem Wert abweichend von 1 stets den Wert 0 annehmen würde. Somit ergibt sich für IKI folgende Form:

$$IKI(U, j) = \sum_{t=1}^{T_U} \text{Cos}(A_{U,t}, O_j) (1 - \alpha 1_{\{1\}}(RT_{U,t})) (1 + r_{U,t})$$

Es sei angemerkt, dass dieser Indikator stets von einem gewählten Thema j abhängt und nicht unabhängig davon betrachtet werden kann.

4.4 Anwendung des Frameworks

Mit Hilfe der einzelnen Indikatoren ist es nun möglich einzelne Personen entsprechend ihrer Werte in ein Portfolio, ähnlich dem in Abbildung 2, einzuordnen. Die einzelnen Indikatoren werden zu einem Referenzpotential-Portfolio integriert. Dabei bezeichnet die Klasse D jeweils die Masse, welche als Baseline zum Vergleich aller Werte dient. Klasse C bezeichnet Personen mit hohem Bezugspersonen-Einflusspotenzial (BEP). Unter Klasse B können alle Experten subsumiert werden, d.h. Personen welche ein hohes Fachwissen bzw. Involvement gegenüber einem Thema j besitzen. Schließlich kombiniert Klasse A einen hohen Grad an Sozio-Zentralität und Fachwissen bzw. Involvement.

Bei Anwendung des Frameworks ist es entscheidend die Indikatoren nebeneinander, anstatt aggregiert zu betrachten, da die Sichtweise je nach Thema bzw. Produkt entsprechend verändert werden muss und eine Gewichtung aufgrund dieser Unterschiede zu Verzerrung führen würde und eine Vergleichbarkeit daher unmöglichen machen würde. Referenzgeber können jeweils themenbezogen unterschiedlich wahrgenommen werden. So werden sicherlich Personen mit einem hohen IKI-Wert eher dazu geeignet sein, eine Referenz für sehr spezifische Fachprodukte, wie Unternehmenssoftware, zu geben, während Personen mit hohen SCI Werten eher als Referenzgeber dienen, wenn es sich um emotionale, Low-Involvement-Produkte wie z.B. Kleidung handelt.

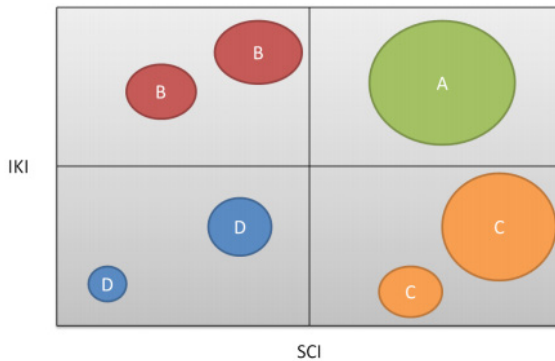


Abbildung 2: Referenzpotential Portfolio

5 Fazit und Ausblick

Der Beitrag hat bekannte Verfahren der sozialen Netzwerkanalyse auf ihre Anwendbarkeit im Rahmen der Bestimmung des Referenzpotentials geprüft. Aufgrund fehlender Anpassungsmöglichkeiten auf die Besonderheiten der inhaltlichen Aspekte von sozialen Netzwerken wurden diese aber als unzureichend zur Beantwortung der Fragestellungen innerhalb dieser Arbeit erachtet. Aus diesem Grund wurden Indikatoren zur Messung des Referenzpotentials in sozialen Netzwerken eingeführt. Da es sich bei der Analyse hauptsächlich um verhaltensorientierte Konstrukte handelt, kann die Operationalisierung mit Hilfe des vorgestellten Frameworks nur als ein Anfang in einer langen Reihe von Forschungsarbeiten angesehen werden. So ist z.B. im Bereich des sozialen Netzes eine Unterscheidung der Gesprächs- oder Mitteilungintensität notwendig, da sich die Aussagen der Wirksamkeit eines Referenzgebers teilweise daran orientieren. Gleichzeitig sollte die Polarität der einzelnen Meldungen mit Hilfe von Opinion Mining untersucht werden, um den Einfluss der Kundenzufriedenheit zu klären und Beiträge ohne empfehlenden Charakter abzuschwächen oder außen vor zu lassen.

Literaturangaben

- Anderson, E., C.Fornell, & Lehmann, D. (1994). Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. *Journal of Marketing* .
- Berghorn, C. (2009). Konzeptualisierung und Ermittlung des Kundenwertes: Am Beispiel einer Volksbank. Diplomica Verlag.
- Blei, David M., Ng, Andrew Y., Jordan, Michael I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks* .

- Brown, J., & Reingen, P. (1987). Social Ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research* .
- Bruhn, M. (2002). Marketing. Grundlagen für Studium und Praxis. Gabler Verlag.
- Brüne, G. (1989). Meinungsführerschaft im Konsumgütermarketing.
- Chin, A., & Chignell, M. (2007). Identifying communities in blogs: roles for social network analysis and survey instruments. *International Journal of Web Based Communities* .
- Cornelsen, J. (2006). Kundenbewertung mit Referenzwerten.
- Cornelsen, J. (2000). Kundenwertanalysen im Beziehungsmarketing.
- Duffy, M. (1994). A Compilation of Three Essays on Modeling Customer Satisfaction.
- Granovetter, M. (1982). The Strength of Weak Ties: A Network Theory Revisited. In P. Marsden, & N. Lin, *Social Structure and Network Analysis*.
- Haseloff, O. W. (1986). Über die Marketingbedeutung von Meinungsführern und Modellpersonen. Realisierung des Marketing .
- Helm, S., & Günther, B. (2006). Kundenwert: Eine Einführung in die theoretischen und praktischen Herausforderungen der Bewertung von Kundenbeziehungen. In S. Helm, & B. Verlag.
- Homburg, C., & Schnurr, P. (1999). Was ist Kundenwert? Universität Mannheim Institut f. Marktorientierte Unternehmensführung (1999).
- Iacobucci, D., & Hopkins, N. (1992). Modeling Dyadic Interactions and Networks in Marketing. *Journal of Marketing Research* .
- Java, A., Finn, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007*.
- King, C., & Summers, J. (1970). Overlap of Opinion Leadership Across Consumer Product Categories. *Journal of Marketing Research* , S. 43–50.
- Lazarsfeld, P. F., Berelson, B. R., & Gaudet, H. (1944). The people's choice: How the voter makes up his mind in a presidential campaign.
- Rudolf-Sipötz, E., & Tomczak, T. (2006). Kundenwert in Forschung und Praxis. THEXIS.
- Schneider, N. C. (2007). Kundenwertbasierte Effizienzmessung: Der Beitrag von Marketingmaßnahmen zur Unternehmenswerterhöhung in der Automobilindustrie.
- Statista 2015, <http://de.statista.com/statistik/daten/studie/181086/umfrage/die-weltweit-groessten-social-networks-nach-anzahl-der-user/> (Abruf: 20.03.2015)
- Zhang, Y., Callan, J., & Minka, T. (2002). Novelty and Redundancy Detection in Adaptive Filtering. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland* .