

# Erhebung und Diagnostik von Markenassoziationsnetzwerken auf Grundlage nutzergenerierter Inhalte

Marc Egger<sup>1</sup>, Gloria Volkmann<sup>1</sup>, und Detlef Schoder<sup>1</sup>

<sup>1</sup> Universität zu Köln, Seminar für Wirtschaftsinformatik und Informationsmanagement, Köln,  
Deutschland  
{egger, volkmann, schoder}@wim.uni-koeln.de

**Abstract.** Bei Konsumententscheidungen greifen Verbraucher auf ihr vorhandenes Wissen über Marken zurück, welches sie im Gedächtnis in Form von Markenassoziationsnetzwerken organisieren und deren Kenntnis es erlaubt Verbraucherentscheidungen besser zu verstehen. Traditionell werden Markenassoziationsnetzwerke mittels reaktiver, befragungsorientierter Instrumente direkt von Verbrauchern erhoben. Gleichzeitig hinterlassen unzählige Verbraucher online in nutzergenerierten Inhalten (UGC) ihre verbal explizierten Wahrnehmungen freiwillig und unangeleitet und erzeugen damit eine neue Datengrundlage zur nichtreaktiven, passiven Erhebung von Markenassoziationsnetzwerken. Traditionelle inhaltsanalytische Verfahren, wie die qualitative Inhaltsanalyse, ermöglichen aufgrund der Quantität, Heterogenität und Erstellungsgeschwindigkeit von UGC entweder nur Stichprobenbetrachtungen, oder sind mit prohibitiv hohen Aufwänden verbunden. Somit erfordert die nichtreaktive Erhebung von Markenassoziationsnetzwerken aus UGC neue, computergestützte Verfahren, zu denen der vorliegende Artikel einen methodischen Beitrag leistet. Auf Grundlage von sprachlichen Mustern traditionell befragungsorientiert erhobener Markenassoziationen wird ein Text Mining Verfahren zur Extraktion von Markenassoziationsnetzwerken aus UGC vorgeschlagen, anhand einer Vergleichsstudie diskutiert und der Einsatz für die Markenführung demonstriert.

**Keywords:** Association Networks, Brand Diagnostics, Text Mining, Network Analysis

13<sup>th</sup> International Conference on Wirtschaftsinformatik,  
February 12-15, 2017, St. Gallen, Switzerland

Egger, M.; Volkmann, G.; Schoder, D. (2017): Erhebung und Diagnostik von Markenassoziationsnetzwerken auf Grundlage nutzergenerierter Inhalte, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 1186-1200

## 1 Einleitung

In Zeiten der Informationsüberlastung und sich zunehmend angleichenden Produkten und Services können Verbraucher diese oft nur schwer unterscheiden [1, 2]. Bei Unsicherheit, sowie in Konsumententscheidungssituationen greifen sie daher auf ihr bereits vorhandenes Wissen über Marken als Entscheidungshilfe zurück [1, 3-5]. Dieser als Markenwissen bezeichnete Beeinflussungsfaktor wirkt auf den kundenorientierten Markenwert [3, 6, 7], wenn Verbraucher mehr oder weniger positiv auf die Elemente des Marketing-Mixes eines markierten- gegenüber eines unmarkierten Produktes oder Services reagieren [1]. Der kundenorientierte Markenwert liefert somit diagnostische Informationen für das Entstehen des monetär messbaren, finanzorientierten Markenwertes, welcher die vom Verbraucher akzeptierte Wertdifferenz zwischen einem unmarkierten und einem markierten aber gleichwertigen Produkt oder Service bezeichnet. Um sich Wettbewerbsvorteile zu verschaffen und den finanzorientierten Markenwert zu erhöhen, müssen Unternehmen daher Strategien zur Steuerung des kundenorientierten Markenwertes entwickeln [1, 2]. Die Ableitung von Strategien setzt allerdings zunächst eine entsprechende Konzeptualisierung sowie die Operationalisierung entsprechender Messgrößen voraus. Gemäß der Konzeptualisierung nach Keller [1] wird der kundenorientierte Markenwert durch das Markenwissen im Gedächtnis des Verbrauchers bestimmt. Dieses setzt sich aus der Markenbekanntheit (Brand Awareness) und dem Markenimage (Brand Image) zusammen [1]. Während die Markenbekanntheit die (Wieder-)erkennungs- und Erinnerungsleistung an die Marke beeinflusst, manifestiert sich das Markenimage in Netzwerken von Markenassoziationen im Gedächtnis der Verbraucher [1]. Diese netzwerkorientierte Sichtweise des Markenimages anhand Markenassoziationsnetzwerken gründet auf der Theorie des Human Associative Memory Models [8] und der Spreading Activation Theorie [4] aus der Kognitionspsychologie. Demnach wird das semantische Gedächtnis des Menschen als Netzwerk aus informationstragenden Knoten (Assoziationen) sowie Kanten unterschiedlicher Stärke verstanden und Aktivierungen einzelner Knoten pflanzen sich entlang der Kanten fort, um weitere Knoten zu aktivieren. Markenassoziationen repräsentieren solche informationstragenden Knoten und entstehen aufgrund direkter (z.B. Verwendung) wie auch indirekter Erfahrungen (z.B. Word-of-Mouth (WoM) der Konsumenten mit dem Leistungsangebot eines Unternehmens [1, 2, 9]. Ein netzwerkorientierter Ansatz Markenassoziationen zu repräsentieren, soll letztlich dazu beitragen, ein besseres Verständnis davon zu erhalten, wie Verbraucher die Marke- sowie das sich aus Menschen, Orten und Nutzungsszenarien zusammensetzende Umfeld der Marke wahrnehmen [10]. Dies schafft eine Grundlage, Präferenzen und Auswahlentscheidungen, und damit die Determinanten des kundenorientierten Markenwertes zu verstehen [1, 10, 11]. Henderson et al. [10, 11] zeigen diese diagnostische Stärke, indem sie aus der Struktur von Markenassoziationsnetzwerken Markenmechanismen wie Markengleichheit, -Verwechselbarkeit oder Verwässerungseffekten erklären können. Keller [1] argumentiert zudem, dass kundenorientierter Markenwert entsteht, wenn Verbraucher starke, vorteilhafte, sowie einzigartige Markenassoziationen besitzen. Letztlich setzen strategische Maßnahmenplanungen zur Schaffung von kundenorientiertem Markenwert jedoch

zunächst die Erhebung von Markenassoziationsnetzwerken als diagnostische Kenngröße voraus. Klassischerweise erfolgt die Erhebung von Markenassoziationen zur Markenimagemessung anhand befragungsorientierter Erhebungsinstrumente zu denen Farsky [12] einen umfassenden Überblick gewährt. Den Erhebungsmethoden ist gemein, dass sie auf reaktiven Befragungen basieren. Aufgrund von Aufwand und Kosten sind Unternehmen jedoch zunehmend an neuen kosteneffizienten Methoden interessiert [13]. Gleichzeitig nutzen heutzutage unzählige Verbraucher das Internet und Social Media Kanäle um freiwillig und unangeleitet, öffentlich zugänglich ihre Einstellungen, Gefühle und Meinungen zu verschiedensten Themen zu verbreiten [14-16]. Diese öffentlichen und häufig in Textform vorliegenden nutzergenerierten Inhalte (User-generated Content : UGC) bieten als qualitative Datengrundlage das Potenzial, Explizierungen von individuellem Markenwissen zu enthalten was eine neue Datenquelle zur nichtreaktiven, passiven Erhebung von Markenassoziationsnetzwerken eröffnen würde. UGC stellt damit eine potentiell kostengünstige und zudem jederzeit verfügbare Alternative zu mittels traditioneller Verfahren erhobener Stichproben dar [13]. Da Markenassoziationen in textbasiertem UGC allerdings durch Worte und Phrasen ausgedrückt werden, müssen diese zunächst identifiziert und extrahiert werden. Als traditionelle Methode kann hier die qualitative Inhaltsanalyse in Betracht gezogen werden, bei welcher ein Analyst Markenassoziationen und Relationen zwischen diesen Assoziationen manuell identifiziert und codiert. Die Menge und Erstellungsfrequenz, sowie das dezentrale Vorkommen von UGC führt jedoch dazu, dass entweder nur eine Stichprobe des insgesamt vorhandenen UGC betrachtet werden kann, oder dass große (manuelle) Aufwände notwendig werden, die zu prohibitiv hohen Kosten der Erhebung führen können. Dies führt zur Notwendigkeit neuer computergestützter Verfahren um den öffentlich vorliegenden UGC zur Ableitung von Markenassoziationsnetzwerken zugänglich zu machen. Der vorliegende Artikel widmet sich dieser Forschungslücke und schlägt ein Verfahren vor, um computergestützt Markenassoziationsnetzwerke aus UGC zu extrahieren. In Abschnitt 2 wird hierzu zunächst ein Überblick über verwandte Arbeiten gewährt und von der vorliegenden Arbeit abgegrenzt. Darauf aufbauend schlägt Abschnitt 3 dann ein neues, automatisiertes Textanalyseverfahren zur Extraktion von Markenassoziationsnetzwerken aus UGC vor. Dieses Verfahren wird dabei induktiv anhand von Sprachmustern aus traditionell erhobenen Markenassoziationen entwickelt. Abschnitt 4 wendet die vorgeschlagene Methodik auf das Beispiel der Marke Google im Kontext des autonomen Fahrens an, und diskutiert den Einsatz der vorgeschlagenen Methodik als diagnostisches Tool für Markenführungsmaßnahmen. Abschnitt 5 schließt mit der Diskussion der Limitierungen und gibt einen Ausblick auf künftige Forschung.

## **2 Verwandte Arbeiten**

Die Extraktion von marketingrelevanten Informationen aus UGC hat in den vergangenen Jahren an Bedeutung gewonnen. Der vorliegende Artikel der eine Methode zur Extraktion von Markenassoziationen aus UGC vorschlägt, grenzt sich wie folgt von anderen Beiträgen ab. Zum einen finden sich Beiträge bei denen Markenassoziationsnetzwerke entweder nicht aus UGC, sondern traditionell über

Befragung erhoben [17-19] werden. In anderen Beiträgen stellt UGC zwar den Untersuchungsgegenstand dar, jedoch werden anstatt Markenassoziationsnetzwerken beispielsweise Marktstrukturinformationen [16] in Netzwerken abgebildet. Weiterhin wurden zu Markenassoziationen ähnliche Konstrukte mittels Text Mining aus UGC extrahiert, jedoch keine Netzwerkstrukturen erhoben, sondern beispielsweise Ranglisten aus Markenassoziationen [20], Produkteigenschaften [21] oder allgemeinen Features [22]. Eine erste Arbeit welche die Extraktion von Markenassoziationsnetzwerken aus UGC und den Nutzen für die Markenführung zeigt [23], ist die Arbeit von Gensler et al. [23]. Im Gegensatz zum hier vorliegenden Beitrag fokussieren Gensler et al. [23] dabei jedoch auf eine spezifische Produktbewertungsplattform als eine bestimmte Ausprägung von UGC und nutzten hier besondere Eigenschaften der dort vorliegenden Produktbewertungen um Markenassoziationen zu extrahieren anstatt den Textinhalt [23]. Dies erlaubt es zwar, das diagnostische Potenzial der Erhebung von Markenassoziationsnetzwerken aus Produktbewertungen zu zeigen, wird jedoch dadurch eingeschränkt, dass die vorgestellte Methodik nicht ohne weiteres auf andere UGC Quellen erweitert werden kann. Damit kann die vorgestellte Methodik nur eine spezifische Stichprobe des verfügbaren UGC auswerten. Der vorliegende Artikel ordnet sich in die bestehende Forschung dahingehend ein, den Fokus von Gensler et al. [23] unter Anwendung zu [22] oder [21] ähnlicher Textanalyseverfahren zu erweitern und damit erstmalig einen Vorschlag zur ganzheitlichen Extraktion von Markenassoziationsnetzwerken aus UGC zu unterbreiten.

### **3 Vorgeschlagenes Verfahren zur Extraktion von Markenassoziationsnetzwerken aus UGC**

Im Folgenden wird das Verfahren zur Erhebung und Konstruktion von Markenassoziationsnetzwerken aus UGC vorgestellt und anhand des Beispiels der Marke Google im Kontext des autonomen Fahrens illustriert. Das vorgeschlagene Verfahren folgt einem 8-stufigen Erhebungsprozess, der sich an das Vorgehen von Gensler et al. [23] anlehnt. Dies sind (1) Festlegung des Untersuchungsobjektes, (2) Datensammlung, (3) Datenbereinigung, (4) Extraktion von Markenassoziationen, (5) Sentimentanalyse, (6) Bestimmung der Assoziationsstärke, (7) Bestimmung der Verbindungen zwischen Assoziationen, sowie (8) die Konstruktion des Markenassoziationsnetzwerks.

#### **3.1 Festlegung des Untersuchungsobjektes**

Im ersten Schritt ist zunächst die Entscheidung für ein Untersuchungsobjekt zu treffen. Hierbei kann es sich um eine Marke, Dachmarke, Produktmarke oder ein Themenumfeld handeln. Im Rahmen dieses Beitrags wird als illustrierende Studie, das Untersuchungsobjekt als die Marke „Google“ innerhalb des Themenfeldes „autonomes fahren“ definiert. Beim autonomen Fahren handelt es sich um eine Innovation im Automobilbereich, bei dem sich Fahrzeuge in naher Zukunft selbstständig fortbewegen ohne das Eingreifen des Fahrers. Im Rahmen dieser Technologie haben

Fahrzeughersteller wie Audi, BMW und Mercedes oder Tesla, jedoch auch vermeintlich branchenfremde Unternehmen wie Google oder Apple Milliarden von Dollar in die Entwicklung investiert. Schlussendlich ist jedoch die Akzeptanz von Verbrauchern maßgeblich für den Erfolg der Markteinführung und Verbraucher treffen ihre Entscheidung auf Grundlage ihres vorhandenen Wissens in Form von Assoziationsnetzwerken. Daher ist die Kenntnis der Assoziationsnetzwerke, insbesondere im Umfeld des autonomen Fahrens, ein wichtiger Schritt für Marken potenzielle Wirkmechanismen im Hinblick auf ihren Markteintritt zu verstehen oder um ihre Marken so zu positionieren dass ein erfolgreicher Markteintritt möglich wird.

### 3.2 Datensammlung

Innerhalb der Datensammlung müssen im Internet solche Texte gesammelt werden, welche von Verbrauchern veröffentlicht wurden und sich thematisch mit dem Untersuchungsobjekt befassen. Im Rahmen des hier beschriebenen Verfahrens wird hierzu die Nutzung von allgemeinen Web-Suchmaschinen wie Google.com oder Bing.com vorgeschlagen. Hierzu muss eine entsprechende boolesche Suchabfrage formuliert und an die Web-Suchmaschine geschickt werden. Die von dieser zurückgelieferten Verweise auf Webseiten (uniform resource locator – URL), sowie die Webseiten selbst werden mittels eines sog. Web-Crawlers abgerufen und gespeichert. Im Rahmen der illustrierenden Studie wurde für diesen Prozess ein Web-Crawler entwickelt, welcher die boolesche Suchabfrage aus Abbildung 1 an die Google Suchmaschine übermittelt, und auf Grundlage der zurückgelieferten URLs die korrespondierenden HTML Webseiteninhalte in einer zentralen Datenbank abspeichert. Insgesamt wurden hierbei 1.1 Millionen englischsprachige HTML Webseiteninhalte von unterschiedlichen URLs gesammelt, welche sich über einen Zeitraum von zehn Jahren zwischen April 2005 sowie April 2015 erstrecken.

Google AND ["autonomous driving" OR "self driving car" OR "self driving cars" OR "driverless car" OR "driverless cars" OR "autonomous vehicle" OR "automated driving" OR "piloted driving" OR "driverless car"]

Abbildung 1: Search Query

### 3.3 Datenbereinigung

Um möglichst relevante UGC Textbeiträge für die Extraktion von Markenassoziationen zu verwenden, wird ein dreistufiger Bereinigungsprozess vorgeschlagen. In der ersten Stufe müssen zunächst Textbestandteile aus den jeweils gesammelten, einzelnen Webseiten erkannt und isoliert werden, da diese üblicherweise nicht nur aus den eigentlichen Textbeiträgen, sondern auch aus anderen Elementen wie z.B. Navigations-, Werbungs-, Kopf- und Fußbereichen bestehen. Diese Elemente werden zusammenfassend als Boilerplate bezeichnet und können möglicherweise Teile der booleschen Suchanfrage beinhalten, obwohl sie inhaltlich keine Informationen tragen (z.B. „autonome Autos“ als Menüpunkt in Navigationselementen). Jede einzelne Webseite kann dabei mehrere irrelevante Boilerplate-Elemente wie auch mehrere

relevante Inhaltselemente beinhalten. Zur Erkennung und Entfernung von Boilerplate-Elementen wird der Ansatz von Kohlschütter [24] empfohlen und verwendet. Dieser basiert auf generischen Eigenschaften von Web-Dokumenten (z.B. link-zu-text Verhältnis) und kann daher als robust gegenüber heterogenen Inputdokumenten angesehen werden, was eine Charakteristik der hier beschriebenen Datensammlung ist. Die zweite Bereinigungsstufe dient der Erkennung und Bereinigung von Duplikaten, welche sonst möglicherweise zu Verzerrungen aufgrund von Mehrfachzählungen - beispielsweise bei der Bestimmung von Markenassoziations- und Vernetzungsstärken - führen können. Duplikate treten beispielweise dann auf, wenn der selbe Inhalt unter unterschiedlichen URLs abrufbar ist (z.B. Druckansichten oder Produktbeschreibungen). Absolute Duplikate können hierbei einfach mittels zeichenweiser Vergleiche erkannt und bereinigt werden. Im letzten Schritt der Datenbereinigung werden nun nicht nutzergenerierte Inhalte (non-UGC) von der Datengrundlage entfernt, da die im Ziel zu erhebende Markenassoziationsnetzwerke die Verbraucherperspektive abbilden sollen und somit ausschließlich von Verbrauchern erzeugte Inhalte (UGC) von Interesse sind. UGC findet sich im Internet verstreut auf sozialen Netzwerken, Foren, Blogs, Produktbewertungsplattformen, E-Commerce Webseiten und letztlich potenziell auf jeder Webseite, die eine Kommentarfunktionalität anbietet. UGC liegt also dezentral vor, und vermischt sich auch mit non-UGC wie z.B. redaktionelle Artikel, Produktbeschreibungen oder Werbung) [25]. Somit ist es zweckmäßig in den gesammelten Dokumenten, non-UGC Dokumente zu bereinigen. Egger et al. [25] haben gezeigt, dass textbasierter UGC inhärente Eigenschaften aufweist, die es erlauben ein binäres Textklassifikationsproblem zu formulieren und mittels Maschinenlernverfahren zu adressieren. Im Rahmen der illustrierenden Studie wird dem Vorgehen von Egger et al. [25] gefolgt und manuell eine Stichprobe von 422 Dokumenten als UGC (214) und non-UGC (208) annotiert um ein sog. Gold Set zu erzeugen. Dieses Gold Set stellt die Grundlage an Dokumenten, welche für das Anlernen eines Klassifizierers verwendet werden und beinhaltet für jedes dort enthaltene Dokument die korrekte Klassifizierung in UGC oder non-UGC. Auf Grundlage dieses Gold Sets wird dann ein auf Support-Vector-Machines (SVM) [26] basierender Klassifizierer unter Verwendung von 10-fold Cross-Validation angelernt. Die 10-fold Cross Validation ist dabei eine Vorgehensweise den Klassifizierer anzulernen und zu validieren. Hierbei wird das Gold-Set in 10 gleichgroße Segmente unterteilt. Von diesen Segmenten werden in 10 Iterationen, jeweils 9 Segmente zum Anlernen des Klassifizierers und eines zur Validierung des Klassifizierers verwendet. Die Validierung erfolgt anhand der Metriken Precision, Recall und Accuracy (Abbildung 2). „true positive“ entspricht hierbei einer gemäß des Gold Sets korrekten Klassifizierung als UGC, „true negative“ einer korrekten Klassifizierung als non-UGC, „false positive“ einer Fehlklassifizierung als UGC und „false negative“ einer Fehlklassifizierung als non-UGC.

$$\begin{aligned}
 \textit{precision} &= \frac{\textit{true positive}}{\textit{true positive} + \textit{false positive}} \\
 \textit{recall} &= \frac{\textit{true positive}}{\textit{true positive} + \textit{false negative}} \\
 \textit{accuracy} &= \frac{\textit{true positive} + \textit{true negative}}{\textit{true positive} + \textit{true negative} + \textit{false positive} + \textit{false negative}}
 \end{aligned}$$

Abbildung 2: Kennzahlen zur Validierung der UGC Klassifikation [25]

Als Ergebnis der Validierung des illustrierenden Beispiels wurde eine Accuracy von 92.29%, eine Precision von 97.18% und Recall von 89.61% bei der UGC Klassifizierung erreicht. Die Anwendung des angelernten Klassifizierers auf die gesammelte Dokumentenkollektion und Bereinigung von non-UGC führt letztlich zu einem Korpus von 33.866 UGC Dokumenten.

### **3.4 Extraktion von Markenassoziationen aus Freitexten**

Zur Konstruktion des Markenassoziationsnetzwerkes sind zunächst die Netzwerkknoten, also die Markenassoziationen zu bestimmen. Bei der vorliegenden Datengrundlage aus textbasierten UGC Dokumenten, manifestieren sich verbal explizierte Markenassoziationen als Textbestandteile in Form von Worten und Phrasen. Um solche Worte und Phrasen in textbasiertem UGC zu finden, die potenziell Markenassoziationen repräsentieren, ist jedoch die Kenntnis darüber notwendig, wie Verbraucher Markenassoziationen verbal explizieren. Nach Lawson [27] werden Markenassoziationen üblicherweise durch Konzepte repräsentiert, die grundsätzliche Eigenschaften, ähnliche Produkte und Themen, Produktverwendungen oder zusammenfassende Bewertungen beschreiben. Nach unserer Kenntnis finden sich jedoch weder in der Opinion Mining noch der Markenführungsliteratur derzeit Beiträge, welche sich der linguistischen Beschreibung von verbalsprachlich explizierten Markenassoziationen widmen. Dies wäre jedoch für die Konstruktion von Text Mining Verfahren zur Aufdeckung von Markenassoziationen von entscheidendem Wert. Als erste Annäherung wird daher im Rahmen dieses Beitrags folgendes induktives Vorgehen vorgeschlagen, um Merkmale von Markenassoziationen zu beschreiben und diese Beschreibungen für die Konstruktion von Text Mining Verfahren nutzbar zu machen. Auf Grundlage einer Studie, die mit einem traditionell zur Erhebung von Markenassoziationen eingesetzten Verfahren durchgeführt wurde, werden sprachliche Regelmäßigkeiten exploriert, die dann Grundlage des zu konstruierenden Text Mining Verfahrens bilden. Dieses induktive Vorgehen erlaubt es zum einen, aus Bekanntem auf Unbekanntes zu schließen und zum anderen, erste Indikationen der Übereinstimmungsvalidität zu messen, indem die Ergebnisse eines etablierten Erhebungsinstrumentes mit dem hier vorgeschlagenen Erhebungsinstrument verglichen werden [28]. Als traditionelles Erhebungsinstrument wird hierzu die Methode der freien Antwort [29] verwendet, welche zur Verfahrensklasse der assoziativen Verfahren [12] zählt. Bei der freien Antwort werden Probanden nach ihren initialen Gedanken bezüglich eines verbalen Stimulus befragt. Im Rahmen der hier durchgeführten Studie wurden dazu 50 Probanden über die Amazon Mechanical Turk Plattform rekrutiert. Die Probanden stammten aus Nordamerika waren zu 64,7% männlich, im Alter zwischen 25 und 40 Jahren, besaßen zur einen Hälfte einen High School oder Collage Abschluss, zur anderen Hälfte einen Bachelor oder Masterabschluss. In der über ein Online-Formular zugänglichen Befragung wurden die Probanden gebeten, 10 Dinge verbal zu explizieren, die ihnen spontan einfallen wenn sie an „autonomes Fahren“ und „Google“ denken. Das Ergebnis der quantitativen Vermessung der erhobenen Assoziationen zeigt, dass 65% (329) der insgesamt 500 erhobenen Assoziationen aus einem bis drei Worten bestehen. 42,6% (213) der Assoziationen bestehen gar aus nur einem-, bzw. 15,4% (77)

aus zwei Worten. Daraus lässt sich schließen, dass Verbraucher ihre Assoziationen offenbar vorwiegend als kurze Informationseinheiten wiedergeben. Um nun verbalsprachliche Muster zum Einsatz für das hier vorgeschlagene Text Mining Verfahren feststellen zu können, werden die erhobenen Assoziationen jeweils in eine Sequenz von Wortarten abgebildet. Hierzu wird ein sogenannter Part-of-Speech Tagger (PoS-Tagger) eingesetzt, der jedem Wort innerhalb der Assoziation ein Part-of-Speech Tag (PoS-Tag) -also eine Wortart zuordnet. Im Rahmen dieser Untersuchung wurde der PoS-Tagger von Schmidt [30] auf die einzelnen Assoziationen angewendet und somit für jede Assoziation eine Repräsentation in Form einer PoS-Tag Sequenz erzeugt. Diese PoS-Tag Sequenzen erlauben es nun, die für Assoziationen üblicherweise verwendeten Wortarten, sowie Gemeinsamkeiten in Wortartenabfolgen zu beschreiben. Im Ergebnis zeigt sich, dass nur 15 PoS-Tag Sequenzen mehr als einmal vorkommen, allerdings in der Lage sind, 61,8% der Assoziationen zu konstruieren. Die durchschnittliche Länge dieser Sequenzen beträgt 2,3 Worte. Die verbleibenden 38,2% der Assoziationen werden hingegen von 189 PoS-Tag Sequenzen konstruiert, die jeweils in dieser Form nur einmal vorkommen. Bei Betrachtung der häufigsten Sequenzen zeigt sich, dass Assoziationen am meisten durch einzelne und zusammengesetzte Nomen (z.B. <technology>, <safety concerns>) konstruiert werden (~30%), gefolgt von einzelnen Adjektiven (~16%, z.B. <efficient>), sowie der Abfolge von <Adjektiv, Nomen> (6%, z.B. <innovative technology>) oder <Nomen, Adposition, Nomen> (2%, z.B. <safety of driver>). Zusammenfassend kann damit geschlossen werden, dass Nomen und Adjektive die wichtigsten Wortarten zur Bildung von Assoziationen darzustellen scheinen. Die Kombination von Adjektiv und Nomen als beschreibende Attribute wie z.B. „innovative technology“ stellt das am häufigsten vorkommende Muster bei Assoziationen dar. Diese Erkenntnisse werden nun genutzt um ein Text Mining Verfahren zu konstruieren. Hierzu wird ähnlich zur oben durchgeführten Charakterisierung von Markenassoziationen ein Vorgehen aus Tokenisierung, Part-of-Speech-Tag (PoS-tag) Sequenzuntersuchung, sowie Sentimentanalyse vorgeschlagen. Zunächst werden dazu die UGC Dokumente in kleinere Bestandteile mittels Tokenisierung [31] segmentiert um für jedes Dokument eine Menge von Sätzen, sowie für jeden Satz eine Menge von Worten erhalten. Da nicht sichergestellt werden kann, dass in den gesammelten UGC Dokumenten Verbraucher ausschließlich über Google und autonomes Fahren sprechen und alle extrahierten Assoziationen sich auf diesen Untersuchungsgegenstand beziehen, wird die Menge der mittels Tokenisierung ermittelten Sätze auf jene reduziert, in denen sich Hinweise auf den Untersuchungsgegenstand finden (z.B. die Nennung von Google). Auf Grundlage dieser selektierten Sätze, wird dann eine PoS-tag-Sequenzuntersuchung zur Erhebung der Markenassoziationskandidaten durchgeführt. Es wird dazu jedem einzelnen Wort eines Satzes ein PoS-Tag mittels des Ansatzes von Schmidt [30] zugewiesen. Dies erlaubt es, für jeden aus Worten bestehenden Satz eine Sekundärrepräsentation zu erstellen, welche allein aus einer Sequenz von PoS-tags besteht. Diese PoS-tag Sequenzen erlauben es nun, Muster zur Extraktion von Markenassoziationen anzuwenden. Auf Grundlage der oben beschriebenen sprachlichen Charakteristika von Markenassoziationen wird hierbei beispielsweise das PoS-tag Muster <Adjektiv, Nomen> verwendet um Wortsequenzen wie <technological innovation> als Markenassoziationskandidaten zu extrahieren und in einer Datenbank zu speichern.



### 3.5 Sentimentanalyse

Bei Markenassoziationen ist die emotionale Orientierung (z.B. positiv, negativ, neutral) von besonderem Interesse. Keller [1] beispielsweise beschreibt das kundenorientierter Markenwert nicht nur durch die im Gedächtnis der Konsumenten vorhandenen starken und einzigartigen, sondern auch vorteilhaften – also positiven - Markenassoziationen entsteht. Daher sollte für die extrahierten Markenassoziationskandidaten die Valenz, also das sog. Sentiment bestimmt werden. Liu [32] liefert hierzu einen umfassenden Überblick über verschiedene Techniken. Da das Forschungsfeld der aspektorientierten Sentimentanalyse hinsichtlich der Erkennungsgenauigkeit schnell voranschreitet, sieht die hier vorgeschlagene Methodik das Einsetzen jeweils aktueller Verfahrensweisen der aspektorientierten Sentimentanalyse vor. Im Rahmen der illustrierenden Studie wird ein einfacher lexikonbasierter Ansatz vorgeschlagen und genutzt, welcher für jede extrahierte Adjektiv-Nomen Kombination ein referenzierendes Lexikon wie z.B. SenticNet [33] zur Bestimmung der emotionalen Orientierung anhand des Adjektivs verwendet. Als Ergebnis der vorangegangenen Schritte wird nun für jedes Dokument eine Menge von Markenassoziationskandidaten (z.B. „leading technology“, „innovative brand“ oder „threatened privacy“) erhalten. Das jeweilige Bezugswort (z.B. das Nomen: „technology“, „brand“, „privacy“) repräsentiert dabei die Markenassoziation, die referenzierenden Worte (z.B. Adjektiv: „leading“, „innovative“, „threatened“) werden als tonalitätstreibende Aspekte verwendet.

### 3.6 Bestimmung der Assoziationsstärke

Die Assoziationsstärke beeinflusst wie schnell und einfach Assoziationen aus dem semantischen Gedächtnis eines Kunden abgerufen werden können [1]. Starke Markenassoziationen, sogenannte „Branded Features“ [10], sind also jene, welche die meisten Kunden mit der Marke in Verbindung bringen und an die sie als erstes im Markenzusammenhang denken. Sie definieren die Marke aus Kundensicht und werden daher auch als „Essenz“ der Marke bezeichnet [34]. Aus diesem Grunde stellt die Stärke von Assoziationen ein wichtiges Attribut zur Beurteilung von Markenassoziationsnetzwerken dar [1]. Angelehnt an traditionelle, qualitative Erhebungsmethoden [34], bei denen die Stärke einer Assoziation anhand der Häufigkeit der Nennung einer Assoziation durch Studienteilnehmer ermittelt wird, soll im Rahmen der hier vorgeschlagenen Methodik die Stärke einer Assoziation anhand ihrer Vorkommenshäufigkeit charakterisiert werden. Hierzu wird ein 2-Mode Netzwerk konstruiert, welches zur Bestimmung der Assoziationsstärke mittels der Gradzentralität [35] dienen soll. Zur Konstruktion des Netzwerkes werden zwei Typen von Knoten unterschieden. Dokumentknoten existieren für jedes UGC Dokument. Als zweiter Typ existiert für jede Markenassoziation die aus einem Dokument extrahiert wurde ein Assoziationsknoten. Jeder Assoziationsknoten eines Dokumentes wird mit jeweils dem korrespondierenden Dokumentknoten über eine Dokument-zu-Assoziation Kante verbunden. Wird eine Assoziation (welche über ihren Namen sowie die emotionale Orientierung zu identifizieren ist) in mehreren unterschiedlichen Dokumenten genannt,

so existiert der Assoziationsknoten dennoch nur einmal, wird jedoch mit jedem Dokumentknoten (und somit dem korrespondierenden Dokument) mit einer Dokument-zu-Assoziationskante verbunden. Abbildung 3 illustriert die Erstellung dieses 2-mode Netzwerk. Hierbei entspricht die Gradzentralität des Assoziationsknoten [35, 36] der Anzahl der Dokumente in denen die jeweilige Assoziation existiert. Um Vergleichbarkeit herzustellen, wird die Gradzentralität mit der Gesamtanzahl von Dokumenten normalisiert.

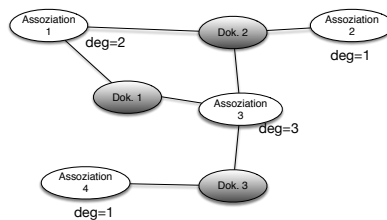


Abbildung 3: 2-mode Netzwerk zur Bestimmung der Assoziationsstärke

### 3.7 Bestimmung der Verbindung zwischen Assoziationen

Während im vorangegangenen Abschnitt das vorgeschlagene Verfahren zur Ermittlung der Stärke von Assoziationen dargestellt wurde, wird nun das Vorgehen erläutert, um Verbindungen zwischen Assoziationen sowie deren Stärke darzustellen. Aus kognitionspsychologischer Sicht wird die Verbindungsstärke zwischen Assoziationen charakterisiert durch die Stärke mit der diese miteinander im Gedächtnis verknüpft werden. Sie beeinflusst damit Aktivierungsausbreitung („spreading activation“) [1], also die Wahrscheinlichkeit das weitere Assoziationen im Gedächtnis des Verbrauchers aktiviert werden, wenn eine einzelne Assoziation z.B. durch Marketing-Mix Aktivitäten aktiviert wird [4, 37]. Im Gegensatz zu traditionellen Consumer-mapping Verfahren [17] können die Verbindungen zwischen Assoziationen im hier vorgeschlagenen Verfahren nicht direkt durch Befragung von Probanden ermittelt werden. Daher wird angelehnt an das von Teichert und Schöntag [38] vorgestellte Verfahren, die Annahme getroffen, dass aus der implizit gemeinsamen Nennung von Assoziationen auf vorhandene Assoziationsnetzwerke geschlossen werden kann. Pro Dokument wird dazu jeweils ein vollständiger Graph aus den dort extrahierten Assoziationen konstruiert (Abbildung 4). Dem unterliegt die Annahme, dass die vom Verbraucher innerhalb eines UGC Dokumentes verbal explizierten Assoziationen im Gedächtnis in einem gemeinsamen Bezug stehen. Über die gesamte Dokumentenkollektion wird dann gezählt, wie oft jeweils ein Paar von gleichen Assoziationen in unterschiedlichen Dokumenten gemeinsam auftritt. Jeweils zwischen diesen Assoziationspaaren wird dann eine gewichtete Kante erzeugt, deren Gewicht dieser Kookkurrenz entspricht – also der Anzahl der Dokumente in denen die beiden Assoziationen gemeinsam auftreten. Da häufig vorkommende Assoziationen in dieser Betrachtung eine höhere Wahrscheinlichkeit haben mit anderen weniger häufig vorkommenden Assoziationen aufzutreten, kann ein Korrekturfaktor einbezogen werden, um solche Verbindungen hervorzuheben die häufiger als zu erwarten wären auftreten.

Statistische Unabhängigkeit vorausgesetzt, entspricht die Wahrscheinlichkeit, dass zwei Assoziationen zufällig gemeinsam auftreten, dem Produkt der relativen Häufigkeit (bzw. normalisierten Gradzentralität) der beteiligten Assoziationen. Diese zu erwartende Wahrscheinlichkeit kann von der relativen Häufigkeit des gemeinsamen Auftretens (Quotient aus absoluter Kookkurrenz zweier Assoziationen und der Vorkommenshäufigkeit der schwächeren Assoziation) abgezogen werden. Abbildung 4 zeigt das resultierende Assoziationsnetzwerk auf Grundlage von zwei einzelner vollständiger Dokumentnetzwerke.

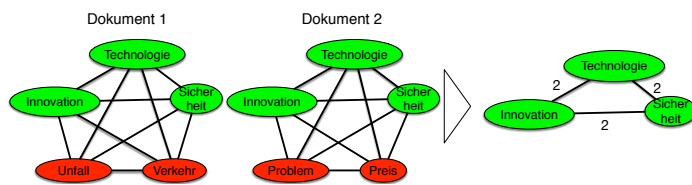


Abbildung 4: Per Dokument und Kookkurrenznetzwerke

### 3.8 Konstruktion des Markenassoziationsnetzwerkes

Zur Konstruktion des Markenassoziationsnetzwerkes werden die Netzwerke aus den Abschnitten 3.6 und 3.7 kombiniert. Hierbei wird das Netzwerk aus Abschnitt 3.6 zum einen dazu verwendet, mittels der normalisierten Gradzentralität die Stärke der Assoziationsknoten zu beschreiben und zum anderen um einen Schwellwert zur späteren Berücksichtigung der Assoziationen im Markenassoziationsnetzwerk festzulegen. Das Netzwerk aus Abschnitt 3.7, welches die Verbindungen zwischen Assoziationen beschreibt, wird dazu um die Information der Assoziationsstärke ergänzt.

Im Bezug zur illustrierenden Studie zeigt Abbildung 5 das Netzwerk der Top-50 Assoziationen für Google im Kontext des autonomen Fahrens. Als Schwellenwert für das gemeinsame Auftreten zweier Assoziationen wurde 18 gewählt. Die Assoziationsknoten wurden gemäß des Anteils positiver (grün), negativer (rot) sowie neutraler (orange) Tonalität eingefärbt. Die Gradzentralität der Assoziationen ist im Zentrum der Assoziationsknoten dargestellt. Insgesamt kann festgestellt werden, dass die im Markenassoziationsnetzwerk in Abbildung 5 vorhandenen Assoziationen inhaltlich ebenfalls bei der befragungsbasierten Studie aus Abschnitt 3.4 wiedergefunden werden können, was eine erste schwache Indikation der Vergleichsvalidität sein könnte. Allgemein sind im Netzwerk in Abbildung 5 zwei schwach miteinander vernetzte Subnetzwerke zu beobachten. Das erste Subnetzwerk beinhaltet Assoziationen die dem Themenfeld des „autonomen Fahrens“ zugeordnet werden können. Diese Assoziationen sind entweder direkt oder in zweiter Ordnung mit der Assoziation „car“ verknüpft. Direkt verknüpfte Assoziationen lassen dabei auf die von Google durchgeführten Prototypentests schließen wie z.B. „testing“, „prototype“, „area“ und „state“. Andere Assoziationen betonen gesellschaftliche, soziale oder persönliche Aspekte. Hierzu zählt beispielsweise die Assoziation „safe“, die stark mit der Assoziation „car“, sowie „driver“ verknüpft ist. Das zweite Subnetzwerk hingegen zeigt Assoziationen, die nicht unbedingt auf das autonome Fahren schließen lassen, jedoch scheinbar starke Markenattribute

darstellen, die von Verbrauchern in kontextunabhängig wiedergegeben werden. Hierzu zählen „computer“, „company“, „services“, „search“, „computer“, „project“, „technology“, „wearables“, „Google Glass“, „devices“ oder „android“. Auffällig ist, dass das erste, dem autonomen Fahren nahe Subnetzwerk, schwach über die Verbindung der zentralen Assoziation „car“ mit der Assoziation „project“ des zweiten Subnetzwerkes verbunden ist, autonomes Fahren also möglicherweise als eines von mehreren Projekten Googles wahrgenommen wird.

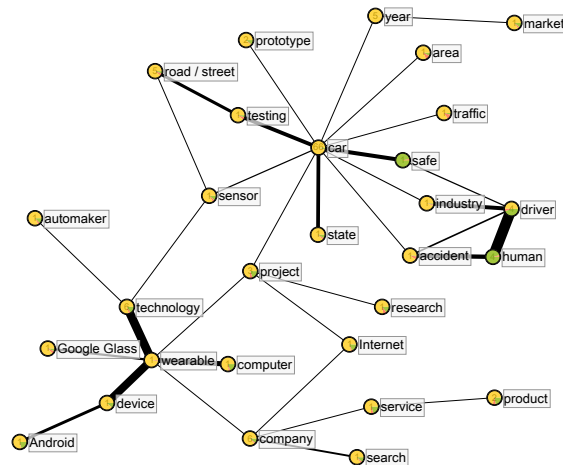


Abbildung 5: Assoziationsnetzwerk Google im Kontext autonomes Fahren

#### 4 Einsatz zur Imagediagnostik

Die Verwendung netzwerkanalytischer Verfahren auf Markenassoziationsnetzwerken und die Interpretation in Bezug auf Markenführungsmaßnahmen wurde insbesondere von Henderson auf Grundlage von durch Befragung erhobenen Markenassoziationsnetzwerken untersucht [10, 11]. Die im Rahmen dieses Beitrags vorgestellte Methodik zur Erhebung von Markenassoziationsnetzwerken aus nutzergenerierten Inhalten können ebenfalls als Grundlage solcher Untersuchungen dienen. Henderson et al. [10] unterscheiden grundsätzlich zwischen Intra-Netzwerkanalysen und Inter-Netzwerkanalysen [10]. In der Intra-Netzwerkanalyse wird auf einzelne Assoziationsknoten fokussiert. Die Gradzentralität wurde hier beispielsweise verwendet um besonders starke Assoziationen sog. „Branded Features“ zu untersuchen. Da diese Branded Features von Verbrauchern als besonders wichtige Markenattribute wahrgenommen werden, sollten diese bei Markenführungsmaßnahmen besonders herausgestellt werden. Henderson et al. [10] verwenden hier beispielsweise die Maße der Gradzentralität (Anzahl der Kanten zu anderen Assoziationsknoten) sowie der Betweenness-Zentralität (Anzahl der kürzesten Wege, die über den Assoziationsknoten zu anderen Assoziationen führen) um die wichtigsten Branded Features zu beschreiben. Im Bezug auf die hier angeführte illustrierende Studie könnte die Gradzentralität Hinweise darauf geben, welche Assoziationen von Verbrauchern mit

Google im Bezug auf das autonome Fahren thematisiert werden und somit wichtige Markenattribute darstellen. Die Verwendung der Betweenness Zentralität könnte verwendet werden, um zu untersuchen wie gem. der Idee des Spreading Activation Networks [4] die Wahrscheinlichkeit besteht, dass eine gewisse Assoziation aktiviert wird, wenn eine andere Assoziation im Markenassoziationsnetzwerk aktiv ist. Marketingmaßnahmen können diese Assoziation dann ansprechen (oder nicht ansprechen) um dafür zu sorgen (oder zu verhindern), dass andere Assoziationen aktiv werden. Neben der Intra-Netzwerkanalyse bietet sich ferner die Möglichkeit unterschiedliche Netzwerke, oder Segmente von Netzwerken zu vergleichen um unterschiedliche Marktsegmente gegenüberzustellen oder Gemeinsamkeiten und Unterschiede in der Verbraucherwahrnehmung herauszuarbeiten. Mit einem geeigneten Informationssystem, welches das hier vorgestellte Verfahren implementiert, könnten beispielsweise Netzwerke anhand der Beiträge einer bestimmten Webseite oder eines Zeitraums segmentiert werden, da für nahezu jedes Dokument die Webseite sowie der Zeitpunkt der Veröffentlichung bekannt ist. [10] führen als Beispiel für Inter-Netzwerkanalysen auch das Maß der Netzwerkdichte als Vergleichskriterium an. Ein sehr dichtes Netzwerk für eine einzelne Marke könnte hierbei eine Indikation auf eine unklare Positionierung geben und somit auf eine mögliche Verwässerung des Markenwertes hinweisen [39]. Gleiches kann angenommen werden, wenn ein Markenassoziationsnetzwerk im Zeitverlauf an Dichte gewinnt [11].

## **5 Diskussion und Ausblick**

Im Rahmen dieses Beitrages wurde ein Verfahren vorgeschlagen, welches potenzielle Markenassoziationen aus UGC extrahiert und somit zur Methodenentwicklung im Bereich der automatisierten Verbraucherforschung beiträgt. Das Verfahren wird hierbei aus der Literatur zur Markenführung und Kognitionspsychologie abgeleitet, ist im Vergleich zu befragender Forschung aufgrund Automatisierung ggf. schneller und potenziell kostengünstiger [13]. Sofern den UGC Beiträgen Metadaten wie Zeitstempel angefügt sind, ist eine Analyse mit dem Verfahren auch rückblickend durchführbar. Den Hauptnutzen besitzt das Verfahren jedoch darin, dass es UGC als eine bisher nicht oder nur schwer holistisch zugreifbare Datenquelle für die Markt- und Verbraucherforschung zugänglich macht. Es lässt sich festhalten, dass mittels unterschiedlicher Intra- sowie Inter-Netzwerkanalysen potenziell Marketing-relevante Informationen abgeleitet werden könnten. Dem vorgelagert stellt sich jedoch zunächst die wichtige Frage welche Gegenstand der folgenden Forschung sein sollte, ob die hier vorgestellten Netzwerke validiert werden können. Ein Ansatz wäre es, die Übereinstimmungsvalidität zu überprüfen: Hierbei wären mit einem „etabliert und als valide geltenden“ Verfahren Assoziationsnetzwerke zu erheben und diese mit den Ergebnissen des hier vorgestellten Verfahrens zu vergleichen. Erste Indikationen wurden hier im Vergleich der erhobenen Assoziationen zu einer befragungsorientierten Studie geliefert. Weiterhin besteht die Frage, wie geeignete Schwellwerte für das Berücksichtigen von erhobenen Assoziationen (z.B. anhand der Gradzentralität) zu setzen sind, da im Vergleich zu befragenden Instrumenten hier nicht 10 oder 20 Assoziationen sowie Relationen

zwischen Assoziationen erhoben werden, sondern Hunderttausende bis zu mehreren Millionen. Letztlich stellt das hier vorgestellte Verfahren einen ersten Ansatz dar, Markenassoziationsnetzwerke aus der digitalen Flut von nutzergenerierten Inhalten als diagnostisches Instrument für die Markenführung verfügbar zu machen. Konkrete Instanzierungen dieses Verfahrens in Form von Informationssystemen (z.B. zum Marketing-Controlling) würden dem unternehmensinternen Methodenkanon zur Marktforschung und Markenführung eine neue Datenquelle erschließen und somit neue, substantielle Erkenntnismöglichkeiten über die Markenwahrnehmung der Verbraucher im Internet eröffnen. Wir hoffen daher, dass andere diesen Ansatz aufgreifen und weitere Forschung im Bereich der textanalysebasierten Erforschung von Markenassoziationsnetzwerken stattfindet.

## References

1. Keller, K.: Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*. 57, 1–22 (1993).
2. Esch, F.-R., Herrmann, A., Sattler, H.: *Marketing-Eine managementorientierte Einführung*. Auflage, Vahlen VI. (2008).
3. Esch, F.-R., Langner, T., Rempel, J.E.: Ansätze zur Erfassung und Entwicklung der Markenidentität. In: [link.springer.com](http://link.springer.com). pp. 103–129. Gabler Verlag, Wiesbaden (2005).
4. Collins, A.M., Loftus, E.F.: A spreading-activation theory of semantic processing. *Psychological review*. 82, 407 (1975).
5. Aaker, D.A.: *Managing Brand Equity: Capitalizing on the Value of a Brand Name*. New York. (1991).
6. Keller, K.L.: Brand synthesis: The multidimensionality of brand knowledge. *Journal of consumer research*. 29, 595–600 (2003).
7. Keller, K.L., Lehmann, D.R.: Brands and Branding: Research Findings and Future Priorities. *Marketing Science*. 25, 740–759 (2006).
8. Anderson, J.R., Bower, G.H.: *Human Associative Memory*. Lawrence Erlbaum (1980).
9. Krishnan, H.S.: Characteristics of memory associations: A consumer-based brand equity perspective. *International Journal of Research in Marketing*. 13, 389–405 (1996).
10. Henderson, G., Iacobucci, D., Calder, B.J.: Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research* 111. 306–327 (1998).
11. Henderson, G.R., Iacobucci, D., Calder, B.J.: Using network analysis to understand brands. *Advances in Consumer Research*. 29, 397–405 (2002).
12. Farsky, M.: *Methoden zur Messung des Markenimages: State of the Art*. Research Papers on Marketing and Retailing, Hamburg (2007).
13. Urban, G.L., Hauser, J.R.: “Listening in” to find and explore new combinations of customer needs. *Journal of Marketing*. 68, 72–87 (2004).
14. Egger, M., Lang, A.: A Brief Tutorial on How to Extract Information from User-Generated Content (UGC). *Künstl Intell*. 27, 53–60 (2013).
15. Decker, R., Trusov, M.: Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*. 27, 293–307 (2010).
16. Netzer, O., Feldman, R., Goldenberg, J., Fresko, M.: Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*. 31, 521–543 (2012).
17. John, D., Loken, B., Kim, K., Monga, AB: Brand concept maps: a methodology for identifying brand association networks. *Journal of Marketing*. 43, 549–563 (2006).
18. Schnittka, O., Sattler, H., Zenker, S.: Advanced brand concept maps: A new approach for

- evaluating the favorability of brand association networks. *International Journal of Research in Marketing*. 29, 265–274 (2012).
19. Brandt, C., de Mortanges, C.P., Bluemelhuber, C., van Riel, A.C.: Associative networks: A new approach to market segmentation. *Int. J. Market Res.* 53, 187–207 (2011).
  20. Crawford Camiciottoli, B., Ranfagni, S., Guercini, S.: Exploring brand associations: an innovative methodological approach. *European Journal of Marketing*. 48, 1092–1112 (2014).
  21. Liu, B., Hu, M., (null): Opinion Observer: Analysing and Comparing Opinions on the Web. Presented at the Proceedings of the 14th international conference on World Wide Web, Chiba, Japan April 1 (2005).
  22. Popescu, A., Etzioni, O.: Extracting product features and opinions from reviews. Presented at the Proceedings of the Conference on Human Language Technology and Empirical Methods in Language Processing (2005).
  23. Gensler, S., Völckner, F., Egger, M., Fischbach, K., Schoder, D.: Listen to Your Customers: Insights into Brand Image Using Online Consumer-Generated Product Reviews. *International Journal of Electronic Commerce*. 20, 112–141 (2015).
  24. Kohlschütter, C., Fankhauser, P., Nejd, W.: Boilerplate detection using shallow text features. Presented at the Proceedings of the third ACM international conference on Web search and data mining (2010).
  25. Egger, M., Lang, A., Schoder, D.: Who Are We Listening to? Detecting User-generated Content (UGC) on the Web. Presented at the ECIS 2015 Completed Research Papers, Münster August 26 (2015).
  26. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* 20, 273–297 (1995).
  27. Lawson, R.: Consumer Knowledge Structures: Networks and Frames. NA - Advances in Consumer Research Volume 25. (1998).
  28. Association, A.P.: Technical recommendations for psychological tests and diagnostic techniques. Part 2 *Psychological Bulletin* (1954).
  29. Olson, J.C., Muderrisoglu, A.: The Stability of Responses Obtained By Free Elicitation: Implications For Measuring Attribute Salience and Memory Structure. *Advances in Consumer Research*. 6, 269–275 (1979).
  30. Schmid, H.: Probabilistic part-of-speech tagging using decision trees. Presented at the Proceedings of the international conference on new methods in language processing (1994).
  31. Feldman, R., Sanger, J.: *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press (2006).
  32. Liu, B.: Sentiment Analysis and Opinion Mining. *Computational Linguistics*. 1–5 (2014).
  33. Cambria, E., Olsher, D., Rajagopal, D.: SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. Presented at the Twenty-eighth AAAI conference on artificial intelligence (2014).
  34. Van Rekom, J., Jacobs, G.: Measuring and managing the essence of a brand personality. *Marketing Letters*. (2006).
  35. Wasserman, S., Faust, K.: *Social network analysis: Methods and applications*. Cambridge University Press (1994).
  36. Scott, J.: *Social network analysis*. Sage (2012).
  37. Anderson, J.R.: A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*. 22, 261–295 (1983).
  38. Teichert, T.A., Schöntag, K.: Exploring consumer knowledge structures using associative network analysis. *Psychol. Mark.* 27, 369–398 (2010).
  39. Barbara Loken, D.R.J.: Diluting Brand Beliefs: When Do Brand Extensions Have a Negative Impact? *Journal of Marketing*. 57, 71–84 (1993).