

Holger Rust

Virtuelle Bilderwolken

Eine qualitative Big Data-Analyse
der Geschmackskulturen im Internet



Springer VS

Virtuelle Bilderwolken

Holger Rust

Virtuelle Bilderwolken

Eine qualitative Big Data-Analyse
der Geschmackskulturen im Internet

Holger Rust
Hannover, Deutschland

ISBN 978-3-658-11885-3 ISBN 978-3-658-11886-0 (eBook)
DOI 10.1007/978-3-658-11886-0

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

Springer VS

© Springer Fachmedien Wiesbaden 2017

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Lektorat: Katrin Emmerich, Katharina Gonsior

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist Teil von Springer Nature

Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Strasse 46, 65189 Wiesbaden, Germany

Inhalt

Einige Gedanken vorweg	1
1 Gesellschaftsbilder in 3W	5
1.1 Die Dinge des Internets	5
1.2 Big Data – modischer Grundbegriff	7
1.3 Big Bullshit, Buzz Word, Mythos?	8
1.4 V-Worte-Chaos	10
1.5 Geheimwaffe Algorithmen	12
2 Anwendungsbeispiele	15
2.1 Kommunikationspartner Waschmaschine	15
2.2 Faszination Bildanalyse	18
2.3 Selfies der Alltagskultur	20
2.4 Zwei Thesen zur Web-Kommunikation	23
2.5 Unvorhersehbare Kaskaden	25
3 Inszenierungen im 3W-Universum	29
3.1 Attraktoren der Inszenierung	29
3.2 Autos und Mode und/oder umgekehrt?	32
3.3 Suchbewegungen im 3W-Universum	38
3.4 Kontextuelle Komparserie	39
3.5 Mustergültige Fokussierung	40
4 Methodologie	45
4.1 Fehleinschätzung der Social Sciences	45
4.2 Methodologie des Projekts	46
4.3 Digital Native Scholars als Experten	52

4.4	Vorschau auf die Ergebnisse	53
4.5	Ergebnisdarstellung	55
5	Ergebnisse	57
5.1	Routinen und Peer Group-Einflüsse	57
5.2	Kommunikationsdynamik in Blogs	71
5.3	Soziokulturelle Attraktoren	78
5.4	Prinzip Kontextualität	84
5.5	Alltags- und Zukunftsmobilität	99
6	Interpretationsansätze	115
6.1	Blick zurück nach vorn	115
6.2	L'homme imaginaire	119
6.3	Kontextuelle Kompetenzen	123
6.4	Cultural Turn of Big Data	125
6.5	Intelligenz versus Rechnerroutinen	127
7	Fünf Lektionen für Zukunftstalente	131
7.1	Ein Bild sagt mehr als tausend Daten	131
7.2	Datenmüll ohne Inhalte	133
7.3	Technizistische Präferenzverfälschung	135
7.4	Narrowing down innovativer Impulse	137
7.5	Einseitige Definition von Talent	139
8	Schluss: Dystopie digitaler Schockstarre	143
	Quellenangaben	147

Einige Gedanken vorweg

Anfang Februar 2016, als das Manuskript dieses Buches eben abgeschlossen war, informierte die Presseabteilung von Googles Mutterkonzern *Alphabet* darüber, dass für die Internet-Suchmaschine in Zukunft ein Spezialist für Künstliche Intelligenz (KI) zuständig sei: John Giannandrea. Die Suchmaschine solle selbständiges Lernen lernen, also darauf getrimmt werden, die Wirklichkeit zu enträtseln, ohne dass die zu diesem Zweck eingesetzte Software vorher entsprechend programmiert werden müsse. Die Kommentatoren stellten diese Personalie in einen größeren Zusammenhang und machten auf den Trend der IT-Giganten aufmerksam, in ihren Produkten verstärkt auf KI zu setzen und hier wiederum die Bilderkennung zentral zu bearbeiten, natürlich auch, um der Konkurrenz die Stirn zu bieten. *Apple* hatte zum Beispiel den KI-Start-Up *Emotient* übernommen. Dessen Technologie soll es erlauben, durch die Analyse von Gesichtsausdrücken Gefühlslagen zu erkennen. Ähnliche „Algorithmen“ sind auch bei anderen Netzwerken im Einsatz, bei *Facebook* zum Beispiel, und in vielen Anwendungen, die auf die Verfeinerung von Werbestrategien durch die Sortierung unstrukturierter Daten ausgerichtet sind. Und die am wenigsten strukturierten Daten verbergen sich nun einmal in Bildern. Auch Giannandrea sprach in den kurzen Statements, die *Google* verbreitete, auf diese Aufgabe an: die Decodierung von Bildern im World Wide Web. Diese Technologie sei, so zitierten deutsche Medien den neuen Chef, von vordringlicher Bedeutung, „um einen wirklich intelligenten Helfer zu erschaffen, der unsere Benutzer mit Informationen und Vorgängen in der echten Welt in Verbindung setzt“.

Bilder aus dem Alltag, Digitalisierung und Echt-Welt: das alles liest sich wie das Abstract einer neuen Erkenntnistheorie, deren Methode eine Hermeneutik ohne menschliche Interpretation ist. Es sind auch die Stichworte des Projekts, das hier nun in seinem sowohl wissenschaftlichen und methodologischen wie pragmatischen Kontext referiert und damit zur Diskussion gestellt werden soll. Das Ziel dieses Projekts und sein Titel war die Kommunikationsdynamik der Entwicklung von „Geschmackskulturen im World Wide Web“ – über Bilder der „echten Welt“. Eine

Gruppe von 60 jungen Leuten, Digital Natives allesamt, durchsuchten mit allen erdenklichen quantitativen und qualitativen Methoden und Techniken die Fotoalben des World Wide Web nach Mustern „alltagskultureller Ausdrucksaktivitäten“. In den letzten Jahren etablierte sich ein ganzes Universum – ein *3W-Universum* – solcher Plattformen, das sich rasend ausdehnt. Weil nun gleichzeitig eine weitere Entwicklung in dieser Zeit rasant an Fahrt aufnahm, die Big Data-Technik, das heißt die unvorstellbare Vergrößerung der Rechnerkapazitäten, Daten zu speichern und in immer kürzeren zeitlichen Distanzen zu ihrer Entstehung zu verarbeiten und eben die automatisierte Bildanalyse ein Ziel der KI sein sollte, wurde dieses Thema Teil des Projekts. Wie genau das vollzogen wurde, ist in den Passagen zur Methodologie beschrieben und in den einschlägigen Kapiteln differenziert ausgeführt.

Was in dieser Analyse nun zutage trat, ist ein seltsames Missverhältnis zwischen dem unfassbaren Aufwand, der weltweit technisch und mathematisch betrieben wird, und dem offensichtlich zentralen Ziel: zu erkennen, was der Kunde will, bevor es selber weiß, und ihm entsprechende Angebote zu unterbreiten. Diese im Grunde genommen triviale Form der Verkaufsförderung bindet ungeheure Potenziale an menschlicher Intelligenz. Um gleich ein Missverständnis zu vermeiden: Dass diese Marketing-Aktivitäten in breiten Bereichen der öffentlichen Wahrnehmungen mit Big Data identifiziert werden, ist eine starke Verkürzung dessen, was die entsprechenden Technologien zu leisten imstande sind. Das wird pointiert in der kritischen Auseinandersetzung der Expertenzirkel deutlich, deren Ausführungen in diesem Projekt (mit Hilfe algorithmischer Techniken) untersucht wurden. *Big Data* zielte in seiner ursprünglichen Widmung in erster Linie darauf, verborgene naturgegebene Routinen zu entschlüsseln, die bislang auf Grund mangelnder Möglichkeiten der Archivierung ungezählter intervenierender Variablen und damit auch der Unmöglichkeit der Identifikation von Zusammenhängen nicht zugänglich waren: Klimatische Prozesse, Wetterphänomene, die Dynamik von Epidemien, tektonische Verwerfungen und ihre Konsequenzen, die Kartografie medizinischer Besonderheiten, astronomische Prozesse, Teilchenbeschleunigung und zahlreiche andere naturwissenschaftliche Herausforderungen, dann aber auch technologisch gestalterische wie das so genannte Internet der Dinge.

Damit übersprang die naturwissenschaftlich-technische Idee eine Grenze und dehnte ihren Geltungsbereich auf das Alltagsleben aus, wo sie nicht nur (wie in einem Teil der neuen Simulationsprojekte einer modernen Soziologie) analytische Fortschritte zeitigen sollte – sie entfaltet auch soziokulturelle Prägekraft. Etwa durch die Erfassung medizinisch bedeutsamer Daten des täglichen Verhaltens, der Kommunikationsakte in den so genannten Social Media und des Kaufverhaltens, das durch eine zunehmende Menge an Sensoren erfasst und archiviert wird. Der Alltag wird damit einer naturwissenschaftlich-technisch begründeten Logik

unterworfen, wobei sich nun eine sehr alte erkenntnistheoretische Frage stellt: Lässt sich eigentlich menschliches Verhalten auf diese Weise identifizieren? Wird es nicht im Prozess der Identifikation so verändert, dass die Identifikation sinnlos wird? Oder ist das Ziel, das echte Leben zu erfassen, am Ende durch ein anderes unausgesprochenes Ziel überlagert: es in der Erfassung gleichzeitig so zu gestalten, dass es erfassbar bleibt?

Das Projekt konzentrierte sich im Kern auf die Bilderwelten, die in den eigens dafür geschaffenen technischen Strukturen inszeniert werden. Sie sind keine echten Welten, sondern *Imaginationen*, also Vorstellungen von echten Welten. Aber was repräsentieren diese Imaginationen, die in Abermilliarden Ausdrucksformen täglich jedem Nutzer des World Wide Web zur Verfügung stehen? Was bedeuten diese Bilderwolken? Diese und viele andere Fragen werden in den sieben Kapiteln dieses Buches gestellt und, so gut es geht, beantwortet. Die Antworten selbst, das zeigt die Relativierung „so gut es geht“, sind wiederum nur Impulse, denkbare Antworten. Diese Zurückhaltung resultiert nicht aus übergroßer Bescheidenheit. Die ist angesichts des gigantischen Materials ohnehin geboten. Wichtiger ist aber, dass sich in der Relativierung eine der wesentlichen Antworten generell offenbart: Je mehr die technische Zurüstung der „echten“ Welt durch die Künstliche Intelligenz voranschreitet, desto weniger differenziert wird sie sich darstellen.

Am Ende könnte sie in einer Art „Data Convenience“ immer nur das darstellen, was ihr in den entsprechenden Vorrichtungen der Datengenese zu aktivieren möglich ist. Was das gesellschaftlich bedeutet, wird breit diskutiert, vor allem unter dem Gesichtspunkt der Datafizierung des individuellen Lebens, dessen Ausdrucksmöglichkeiten verarmen. In dieser düsteren Prognose ist aber interessanterweise gleichzeitig der Hinweis auf eine ökonomische Konsequenz verborgen: Wird nicht am Ende eines immer enger um sich selbst kreisenden Datafizierungsprozesses die Dystopie einer digitalen Schockstarre zu einer Wirklichkeit, die die echte ersetzt, weil keine Impulse von außen mehr kommen? Was bedeutet das für Innovation, Kreativität, Kultur? Und so entsteht eine neue Frage: sollten die Interessenten wirtschaftlicher Nutzbarkeiten aus eben diesen Nutzwerturwägungen nicht besser die Finger aus dem Netz lassen, um zu sehen, was ihnen Impulsen aus der echten Welt geboten wird – sozusagen als visualisierte Big Data, und sie je nach ihrer Kompetenz interpretieren?

1.1 Die Dinge des Internets

Das wachsende Interesse an Alltags-, Mode- und Lifestyle-Themen hat in den letzten Jahren eine vielfältige Szene von Web-Aktivisten inspiriert. Auf ihren Blogs werden Stilfragen verhandelt und Trends verbreitet, Marken diskutiert, Einkaufsmöglichkeiten ausgetauscht und den Sympathisanten, den *Followers*, eine Plattform für Kommentare eingerichtet. Manche von ihnen erwirtschaften sich globale Berühmtheit, Scott Schuman etwa, Fotograf bei Modezeitschriften wie *GQ* und *Vogue*, der mit seinem Blog *thesartorialist.com* einen Maßstab für das Genre insgesamt gesetzt hat. Seine Idee war es, Fotos von modisch interessanten Zeitgenossen und -genossinnen aus dem metropolitanen Alltag zu posten, die dann von den Betrachtern kommentiert werden konnten: „A two-way dialogue about the world of fashion and its relationship to daily life.“

Vielfach sind die Bloggerinnen und Blogger Amateure, die ihren eigenen Stil propagieren und damit Erfolg suchen, mitunter aber auch nur die Kommunikation mit Peers und „Freunden“ pflegen. Sicher ist: Die Szene, in der nach Lust und Laune und nach oberflächlicher Betrachtung anarchisch gepostet wird, wächst. Sie wächst gewaltig, auch deshalb, weil immer wieder neue Möglichkeiten eingerichtet werden, sich auf relativ einfache Art und Weise in dieses Universum einzugliedern. Schumans Blog war zunächst auf *Blogspot* platziert – ein Beispiel nur für die vielfältigen *Blogplattformen*, die es als Open Source-Software jedem ermöglichen, mit eigenen Botschaften, vor allem mit Bildern, individuelle Interessen zu demonstrieren und mit andern zu teilen.

Die Erfolge dieser Initiativen haben die Nutzung von Plattformen attraktiv gemacht. Und die Verlinkung dieser Blogs mit den Social Media trägt dazu bei. Aber sie sind mehr als diese Social Media. Sie sind Foren der Geschmackskulturen, dynamische Rekonstruktionen der in Milliarden Partikel dekonstruierten Welt alltagskultureller Ausdrucksaktivitäten: Sie sind das Arsenal der Dinge des Internets.

Die dynamische Entwicklung der Blog-Szene liegt nicht zuletzt auch darin begründet, dass klassische Medien, Modemagazine zum Beispiel, den Reizwert der oft eher semiprofessionell inszenierten Beiläufigkeiten und Schnappschüsse von anonymen Besuchern am Rande der großen Messen erkannt haben und übernehmen. Auch Unternehmen haben erkannt, dass in diesem Universum eine Möglichkeit besteht, jenseits flacher Werbung in einen bildhaften Diskurs über Geschmacksfragen einzutreten – eine eher heikle Angelegenheit, wie sich später zeigen wird. Denn diese Szene versteht sich immer noch – trotz aller Versuche der Kommerzialisierung – als ein Raum anarchischer Kommunikation, als Spiel mit den Möglichkeiten der analogen Welt.

Zum Beispiel auf der Plattform *Tumblr*, wo zum Beginn des Jahres 2016 etwa 276 Millionen individuelle Blogs mit 128 Milliarden Einträgen gezählt wurden, das sind 80 Millionen jeden Tag. Oder *Instagram*: 300 Millionen monatlich aktive User, von denen über 70 Prozent außerhalb der USA leben und die 70 Millionen Fotos pro Tag hochladen. Dazu *Flickr*, *Pinterest*, neuerdings *SnapChat*, alle mit beeindruckenden Transaktionszahlen unter anderem auch durch die weltweiten passiven Rezipienten, die sich hin und wieder oder auch regelmäßig die eine oder andere Website einfach nur anschauen.

Hunderte von Millionen Digital Natives inszenieren regelmäßig ihr Leben und seine alltagskulturellen Ausdrucksformen auf Blogplattformen des globalen Web-Universums, in Bildern der Produkte, die etwas für sie bedeuten, in stilbewussten Arrangements oder dekonstruiert in Einzelementen, in einer unendlichen, jedem Individuum offenstehenden Kommunikation. Milliarden digitalisierter Statements sind es täglich und global. Die meisten User nutzen mehrere dieser Plattformen gleichzeitig und untereinander verlinkt, aufeinander Bezug nehmend und mit unterschiedlichen Schwerpunkten, dazu die Querverbindungen zu *Facebook*, *Youtube* und *Twitter*, die *RSS-Feeds* und allen erdenklichen sonstigen Kommunikationsmöglichkeiten des mobilen Internets (aktuelle Statistiken zu allen Blogplattformen und Social Media auf <http://expandedramblings.com>).

Wichtig ist also, dass in diesem Universum nicht nur die substanziell erfassbaren Aktivitäten zählen, sondern auch die ungeheure Zahl an Kontakten mit passiven Rezipienten. Denn jeder kann sich den *Sartorialist* oder Zehntausende vergleichbarer Blogs anschauen, jeder kann sich bei den Plattformen anmelden, vielleicht, um einen eigenen Blog zu betreiben, vielleicht aber auch nur um die Postings der globalen Peers zur Kenntnis zu nehmen. Und nicht nur die, sondern auch die Hinweise auf Blogs, die die jeweiligen Bloggerinnen und Blogger selbst interessant finden, mit weiteren Hinweisen auf die Blogs, die die Impulse aufnehmen und vervielfältigen. Schnell wird man sich in einem rasend expandierenden 3W-Universum wiederfinden, konstruiert aus ungezählten Motiven, aus Modulen und Kompositionen der Module

in jeglicher denkbaren Verknüpfung und in unzähligen Kontexten – das alles, ohne dass es ein *digitales Libretto* gäbe. Wie eine Unzahl weiterer unstrukturierter Daten unterschiedlicher Qualität aus undefinierten Quellen wird die schiere Masse dieser Imaginationen, damit überhaupt der Anschein eines ordnenden Prinzips entsteht, zunächst einmal formal in einer Begriffsschublade abgelegt: *Big Data*.

1.2 Big Data – modischer Grundbegriff

Zunächst einmal handelt es sich, wie bereits gesagt, um die Kennzeichnung eines technischen Prozesses, der durch die Vergrößerung und Differenzierung von Rechenkapazitäten entsteht. Bislang chaotisch anmutende Prozesse, die in unglaublicher Geschwindigkeit und in ungeahnten Rückkoppelungsprozessen interner und externer Faktoren verlaufen, lassen sich übersichtlicher und in ihrer unerklärten Systematik darstellen. Die Erhöhung der quantitativen Kapazitäten bei gleichzeitig zunehmender Geschwindigkeit der Rechenoperationen bietet neue Ansätze, die in diesen Domänen unablässig entstehenden und bislang unzugänglichen Daten nach Korrelationen zu durchsuchen. Beides ermöglicht einen technischen Zugriff auf unterschiedlich strukturierte Daten oder auf Daten aus einander artfremden Kontexten, mit dem Ziel, ohne vorwegformulierte Hypothesen zu unentdeckten Zusammenhängen vorzustoßen. Grundlage für die Faszination an dieser Methode ist die Kernthese einer vorgeblich neuen Erkenntnistheorie: Die analoge Welt offenbare ihre innersten Gesetze wie in einem unbewussten Prozess durch die digitale Erfassung ihrer Äußerungen.

Ein zweiter Funktionsbereich entsteht aus der Verknüpfung digitaler Technologien und Maschinen. Sie verstärkt die Transparenz komplexer Prozesse wie Fertigungsketten, Wartungen oder Qualitätskontrollen. Die einschlägig entwickelte Elektronik identifiziert automatisch, was früher nach Gehör, Erfahrung oder Gefühl, im Wesentlichen also in *Trial & Error*-Verfahren, entschieden wurde. Maßnahmen können auf diese Weise auf die beobachtbare Basis digitaler Indikatoren mit prognostischem Potenzial gegründet werden. Der nächste Schritt ist nun, die gesamte Wertschöpfungskette von der Rohstoffgewinnung über die Entwicklung, die Produktionslinien und -verfahren bis hin zum Marketing des Endproduktes und die Reaktionen von Konsumenten zu erfassen und miteinander in Beziehung zu setzen. Vor allem, was den letzten Punkt angeht, die Reaktionen der Konsumenten, erreicht die digitale Technologie der Big Data-Welt eine neue qualitative Dimension: In die zunächst rein technische Prozesskette von Entwicklung, Produktion und Absatz werden menschliche Handlungsoptionen eingefügt,

die sich in Daten niedergeschlagen haben. Längst sind diese Handlungsoptionen daher in die Technologien zur Erfassung integriert. Die Handlungsoptionen, oder um es soziologisch zu definieren, die alltagskulturellen Ausdrucksaktivitäten von Menschen, werden maschinell lesbar gemacht, um auf die gleiche Weise wie bei chaotischen Systemen die geheimnisvolle Logik zu entschlüsseln, nach der Menschen handeln, und Entwicklung, Produktion und Absatz auf diese Logik zu gründen. Kreditkarteneinsätze, RFID-Codes, Telefondaten, Apps und Suchmaschinenanfragen, GPS-Ortungen, Facebook-Einträge, die zunehmend digitalisierten Autos und tausend andere digitale Repräsentationen des analogen Alltagslebens liefern die Daten. Big Data erscheint als die technologische Neuauflage psychoanalytischer Zugriffe auf das Unbewusste.

Es wurde bereits angedeutet: Die Literatur zu Big Data, die in dem hier referierten Projekt mit algorithmischen Akkuratessen analysiert wurde, zeigt denn auch eine weit verbreitete Skepsis, die allerdings im dominierenden Diskurs und der massiven Werbung von Anbietern mitunter verschüttet wird. Diese Skepsis ist keineswegs das verzopfte Produkt verspäteter Maschinenstürmer, sondern bewegt mittlerweile selbst prominente Mitglieder der globalen der Entwicklerszene, wie etwa die im Dezember 2015 gegründete Initiative *Open AI* zeigt.

1.3 Big Bullshit, Buzz Word, Mythos?

Protagonisten dieser Organisation sind Elon Musk (Entwickler des Elektro-Supercars Tesla), Peter Thiel (Erfinder von Pay Pal) und andere Silicon Valley-Größen. Das Ziel: Die Gefahren der künstlichen Intelligenz zu erforschen und öffentlich zu diskutieren. Thiel hatte schon anlässlich einer Buchpräsentation in der Frankfurter Allgemeinen Zeitung vom 21. Januar 2014 überraschend ausgeführt: „Wenn Sie die Worte Big Data oder Cloud Computing hören, sollten Sie so schnell wie möglich weglaufen.“ Nicht minder plakativ und ebenso unverblümt tat Richard Straub, ehemals *Chief Learning Officer* bei IBM, heute vielgesuchter Berater und Key Note Speaker, die Sache ab: „Big Data, Big Bullshit“ (Süddeutsche Zeitung, 13. November 2014). Tom Davenport, Partner bei *Deloitte* und Forscher am *MIT*, verbreitete in einem Report auf der Website des Beratungsunternehmens den Ratschlag, einen anderen Begriff suchen. Warum? „Big data just doesn't describe the phenomenon very effectively“ (Davenport 2014). Drei Stimmen aus einem erstaunlich großen Chor, einhellig: *Big Data* ist ein *Buzz Word*. *Big Data* ist ein *Hype*. *Big Data* wird schon missverstanden, ehe es überhaupt richtig losgeht.

„Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy (Boyd und Crawford: 663). Ein Mythos, eine Illusion. Buzz Words. Bull Shit. Hype. Begriffe wie diese werden häufiger gebraucht. Doch anders als im diffusen Unbehagen einer kulturpessimistischen Technikkritik sind sie mit substanziellen Erfahrungen angereichert.

Was also genau wird kritisiert?

1. Viele Beispiele für den Erfolg der Big Data Analyses stammten nicht aus dem Konsumbereich, sondern aus technologischen, biologischen, meteorologischen geophysikalischen Projekten, aus der Prozesssteuerung, der Genetik oder Geophysik und würden in einem zweifelhaften Analogieschluss einfach auf soziokulturelle Phänomene übertragen.
2. Daraus folgere, dass Big Data nicht als Mittel zur Fokussierung von Fragestellungen diene, sondern als ein automatischer Generator von Antworten missverstanden werde, in dem sich – wenn man nur die richtigen Suchprogramme einsetze – das (analoge) Leben automatisch und ohne dass es dem Kunden bewusst werde selbst zu erkennen gebe.
3. Vor allem aber sei zu beklagen, dass nicht einmal klar werde, was eigentlich *Daten* sind. Jede Branche, jedes Ressort definiere Daten anders oder nutze *andere* Daten oder identische Daten *anders*.
4. Eine Reihe von Problemen entstehe durch diese Konzentration auf *erfassbare* Daten: Zum Beispiel, dass Teile der Welt, die (zur Zeit) weniger Daten produzieren als andere, in der so genannten *Predictive Policy*, also einer Art automatischer Erfolgsvorhersage des Strategischen Managements nicht berücksichtigt würden.
5. Belegbar sind zudem demografische Verzerrungen etwa bei Twitter-Analysen. Sie verweisen grundsätzlich bislang auf ein eher urbanes Nutzermilieu von jüngeren Menschen und mithin auf thematisch eingeschränkte Kommunikationsinhalte.
6. Die Kommunikation über alltagskulturelle Ausdrucksaktivitäten könne aus verschiedenen Gründen nur unzureichend erfasst werden, insbesondere und trotz der hohen Aussagekraft über *Bilddokumente*. Die Unsicherheit bei der Erwartung prognostischer Potenziale entsteht vor allem aus der Tatsache, dass jüngere Nutzer des Internets eine Reihe biografischer Veränderungen vor sich haben.
7. Vor allem aber: Es werde munter drauflos definiert, ohne dass man sich auf eine Bedeutung einigen könnte: aus kommerziellen und wissenschaftlichen, aus ethischen, quantitativen und qualitativen Blickwinkeln, aus der Perspekti-

ve von Branchen und Ressorts, Anbietern von Tools und Datenschützern und App-Entwicklern.

Es mutet in der Tat kabarettistisch an, wenn man die Ergebnisse der Definitionsversuche ordnet. Vor allem, wenn man eine Software nutzt, die wiederkehrende Motive identifiziert. Offensichtlich gab es einen Impuls, der immer dann, wenn man nicht so genau wusste, was man sagen sollte, angewendet wurde. Dieser zunächst originelle Impuls ging 2007 von einer Studie der damaligen finnischen Beratungsfirma META Group aus, in der die sich abzeichnenden Datenfluten der Zukunft mit drei Worten charakterisiert wurden: *Volume*, *Velocity* und *Variety*.

V-Worte.

1.4 V-Worte-Chaos

Geradezu zwanghaft folgten in den nächsten anderthalb Jahrzehnten in unsystematischer Reihenfolge *Veracity*, *Viability*, *Value*, *Visibility*, *Visualization*, *Volatility* und weitere V-Worte, die allerdings wenig zu einer wissenschaftlich gesicherten Zukunftsprognose beitragen konnten.

Trotz der Inflation dieser V-Worte wurde nicht deutlicher, um was es sich eigentlich bei Big Data im Kern handelt. Die Definitionsversuche sind im Einzelnen unzureichend, insgesamt vage, weil sie eine Menge an oft nicht kompatiblen Möglichkeiten für Voraussetzungen und Anwendungen der Datengenese, Datensammlung und -analyse und schließlich -interpretation erfassen. Diese Unschärfe der Definitionen wird von vielen Experten kritisiert (siehe dazu Kapitel 7).

Ein vorweggenommenes Ergebnis der Studie, zu deren Verständnis diese einleitenden Bemerkungen dienen, mag verdeutlichen, was die Kritiker meinen. Wie im Kapitel 4.2 über die Methodologie des Projekts differenzierter ausgeführt wird, haben 12 studentische Mitarbeiterinnen und Mitarbeiter mehr als 260 Dokumente, die sich mit Big Data und seinen Anwendungen beschäftigen, einer Analyse unterzogen. Dabei galt die Aufmerksamkeit auch der Frage, wie denn Big Data definiert sei. Immerhin geht es ja auch darum, klar benennen zu können, *was* untersucht wird. Nur leider war das nicht möglich, wie die folgende Dokumentation einer zufällig ausgewählten Reihe von Definitionsversuchen zeigt. Die Versuche sind in unsystematischer Folge so aufgeführt, wie die automatisierte Analysemethode sie zutage förderte: Big Data ...

- ist in den Unternehmen angekommen. So geben 12 Prozent der Teilnehmer an, Big-Data-Initiativen seien fester Bestandteil ihrer Unternehmensprozesse, bei