

# Social Media Mining in der Praxis

## Was twittert die Bundesregierung und was kann daraus abgeleitet werden?

Durch einen einzigen Tweet erreichte der Regierungssprecher Steffen Seibert im September 2019 circa 922.000 Follower: Unmittelbar und ungefiltert. In diesem Artikel werden ausgewählte Methoden des Social Media Minings vorgestellt und damit die Tweets der Bundesregierung strukturiert und analysiert. Auffällig ist, dass der Regierungssprecher über zahlreiche gesellschaftsrelevante Fragestellungen twittert und zudem den Fokus sehr stark auf die Aktivitäten der Bundeskanzlerin Angela Merkel lenkt.



**Prof. Dr. Martin Vogt**

ist Professor für Business Intelligence, insbesondere Advanced Analytics an der Hochschule Trier. Bevorzugte Forschungsgebiete: Data Science, Risikomanagement.

**Summary:** The government spokesman Steffen Seibert reached about 922,000 followers in September 2019 through a single tweet: direct and unfiltered. In this article selected methods of social media mining are presented and thus the tweets of the federal government structured and analyzed. It is noticeable that the spokesman for the government tweets about numerous socially relevant issues and also places the focus very much on the activities of Chancellor Angela Merkel.

**Stichwörter:** Data Science, Social Media Mining, Text Mining, Twitter, Sentimentanalyse

### 1. Einleitung

Am 3. November 2017 schrieb die *Frankfurter Allgemeine Zeitung* (vgl. Radunski, 2017): „Donald Trumps wichtigste Waffe ist ausgefallen: Nicht das Weiße Haus, nicht das Atomwaffenarsenal, sondern sein Konto bei **Twitter**.“ Hintergrund war, dass sein *Twitter* Konto für einige Minuten durch einen Mitarbeiter von *Twitter* deaktiviert wurde. Gegenüber dem Fernsehsender *Fox* gab *Donald Trump* offen

zu, dass *Twitter* ein Schlüssel für seinen Erfolg ist, da er sofort Millionen Menschen erreicht – unter Umgehung der Medien (vgl. Kanter, 2017). Diese Aussagen verdeutlichen die Schlüsselrolle von **Sozialen Medien** wie *Twitter* für die Verbreitung von Informationen.

Neben *Donald Trump* nutzen zahlreiche weitere Politiker und Unternehmen, darunter der Regierungssprecher der **Bundesregierung** der *Bundesrepublik Deutschland*, *Twitter* zur Übermittlung von Informationen. So wurden zum Beispiel über den *Twitter* Account @RegSprecher des Regierungssprechers *Steffen Seibert* (siehe <https://twitter.com/RegSprecher>) bereits mehr als 11.600 Tweets abgesetzt (Stand 08.09.2019). Jeder neue Tweet erreicht dabei circa 922.000 Follower: Ohne Umwege, direkt und ungefiltert. Es lohnt sich also zu analysieren, über welche Inhalte getwittert wird und wo diese geografisch gelesen werden.

Um die Inhalte der Tweets systematisch untersuchen zu können, werden besondere Verfahren benötigt. In diesem Artikel werden hierzu ausgewählte Methoden des **Social Media Minings** vorgestellt und damit die Tweets der Bundesregierung strukturiert und analysiert. Die Analyse zeigt, dass der Regierungssprecher den Fokus sehr stark auf die Aktivitäten der Bundeskanzlerin *Angela Merkel* legt. Andere Regierungsmitglieder erscheinen nicht in den Top-Platzierungen der am häufigsten genannten Begriffe. Mit Hilfe der vorgestellten Methoden des Social Media Minings lässt sich ebenfalls zeigen, dass die Kanzlerin Merkel häufig in Verbindung mit einer Handlung gebracht wird. Die Auswertung zeigt weiter, dass die Bundesregierung zu zahlreichen gesellschaftsrelevanten Themen twittert. Zu den Themen mit tendenziell negativen Emotionen zählen unter anderem der Kampf gegen Terrorismus, Antisemitismus und Krimi-

nalität, sowie Gewalt gegenüber Frauen und das Ende des Kohlebergbaus. Bei den Themen mit positiven Emotionen wird vor allem über den Europäischen Rat und die (multilaterale) Zusammenarbeit geschrieben. Zudem wurde über die Olympiade in Pyeongchang 2018 getwittert. Weiterführende Verfahren zur Vertiefung der gewonnenen Erkenntnisse, wie etwa eine Analyse der Follower, werden im Ausblick vorgestellt.

Social Media Mining hat eine hohe praktische Bedeutung für zahlreiche betriebswirtschaftliche Fragestellungen, denn es erlaubt Stimmungen und Meinungen in sozialen Netzen zu strukturieren und zu analysieren und somit Trends frühzeitig zu erkennen, die u.a. für Kundenansprachen, Markenpositionierungen oder die unternehmensstrategische Entwicklung von hoher Relevanz sind.

Dieser Artikel ist wie folgt strukturiert: Zunächst werden in Abschnitt 2 Soziale Medien, insbesondere *Twitter* eingeführt und in Abschnitt 3 die theoretischen Hintergründe von Social Media Mining erläutert. Darauf aufbauend werden in Abschnitt 4 die verwendeten *Twitter* Daten vorgestellt und aufbereitet. Anschließend werden in den Abschnitten 5 und 6 die Inhalte der Tweets untersucht: Zunächst in Abschnitt 5 die benutzten Emotionen und anschließend in Abschnitt 6 Wortzusammenhänge. Abschnitt 7 beschäftigt sich mit der Frage, *wo* die Informationen gelesen werden. Hierzu werden die Standorte von 500 Followern analysiert. In Abschnitt 8 werden Anwendungen im Unternehmenskontext diskutiert und schließlich werden die Ergebnisse in Abschnitt 9 zusammengefasst und ein Ausblick gegeben.

## 2. Soziale Medien: Twitter, Facebook und Co

Unter **Sozialen Medien** wird der „Sammelbegriff für Angebote auf Grundlage digital vernetzter Technologien, die es Menschen ermöglichen, Informationen aller Art zugänglich zu machen und davon ausgehend soziale Beziehungen zu knüpfen und/oder zu pflegen“ verstanden (vgl. *Schmidt/Taddicken*, 2017, Seite 8). Es gibt demnach eine Vielzahl Sozialer Medien mit unterschiedlichen Funktionalitäten. Um diese Fülle besser beschreiben zu können, werden diese in vier Gattungen unterteilt: Plattformen, Personal Publishing, Instant Messaging/Chat und Wikis (vgl. *Schmidt/Taddicken*, 2017). Dabei ist eine genaue Abgrenzung einzelner Dienstleister nicht immer möglich, da eventuell Funktionalitäten aus mehreren Gattungen angeboten werden. *Schmidt* und *Taddicken* (2017) kategorisieren die vier Gattungen wie folgt: **Plattformen** bieten registrierten Nutzern die Infrastruktur („Plattform“) für Kommunikation und zum Aufbau sozialer Beziehungen. Der Marktführer unter den Plattformen ist *Facebook*. Andere Beispiele für Plattformen sind zum Beispiel *mySpace* oder *LinkedIn*. Un-

ter der zweiten Gattung **Personal Publishing** werden ähnlich zu Plattformen Anbieter verstanden, die es ermöglichen Inhalte im Internet zu veröffentlichen. Allerdings werden hier eher die einzelnen Autoren betont. Das bedeutendste Beispiel ist der Kurznachrichtendienst *Twitter*. Neben *Twitter* zählen etwa Weblogs zu dieser Gattung. Die dritte Gattung **Instant Messaging/Chat** hat zum Ziel, die synchrone Kommunikation zwischen Nutzern zu ermöglichen. Beispiele hierfür sind etwa *WhatsApp* oder *Snapchat*. Schließlich werden **Wikis**, wie *Wikipedia*, zu der vierten Gattung gezählt.

In diesem Artikel werden *Twitter* Daten ausgewertet. *Twitter* wurde im März 2006 gegründet und ermöglicht Kurznachrichten bis zu 140 bzw. 280 Zeichen, sogenannte Tweets, zu veröffentlichen. *Twitter* besitzt im Durchschnitt 330 Millionen aktive Benutzer pro Monat (vgl. *Clement*, 2019) und spielt somit eine zentrale Rolle in der Verbreitung von Informationen.

## 3. Theoretische Grundlagen des Social Media Minings

**Social Media Daten** unterscheiden sich von klassischen Daten etwa aus Unternehmensdatenbanken: Sie sind sehr umfangreich, Fehler behaftet, unstrukturiert, dynamisch (vgl. *Gundecha/Liu*, 2012) und liegen in verschiedenen Formaten vor: Text, Sprache, Bilder und Videos (vgl. *Injadat et al.*, 2016). Deshalb reichen klassische statistische Methoden zur Auswertung von strukturierten Daten in (relationalen) Datenbanken, wie etwa Regressionsanalysen, häufig nicht aus um Social Media Daten zu analysieren. Je nach Art der Fragestellung sind weitere Methoden aus verschiedenen Feldern wie Informatik, Bildverarbeitung, Linguistik, Machine Learning oder Statistik nötig. Hierzu zählen insbesondere Methoden um Informationen aus Texten (Text Mining) oder Bildern (Image Processing) zu gewinnen, der Netzwerkanalyse (etwa um Follower zu untersuchen) oder zur geografischen Analyse (vgl. *Chen et al.*, 2012).

In diesem Artikel werden Texte – *Twitter* Tweets – mit Methoden des Text Mining ausgewertet. Unter **Text Mining** wird die Trennung von verborgenen und möglicherweise benötigten Informationen von Textdaten verstanden (vgl. *Ghosh et al.*, 2012). Da über 80 % der verfügbaren Informationen als Texte gespeichert sind (vgl. *Ghosh et al.*, 2012), kann Text Mining in zahlreichen praktischen Anwendungen verwendet werden (siehe Abschnitt 8).

Da die meisten Texte jedoch nicht zu dem Zweck geschrieben wurden, um diese mit Hilfe einer Analyse auswerten zu können, sind diese oft unstrukturiert und müssen zunächst aufbereitet werden. Dazu werden etwa alle Wörter kleingeschrieben oder Wörter ohne besonderen Inhalt, wie „der“, „die“, „das“, entfernt (siehe Abschnitt 4). Anschließend

können Informationen aus diesen Texten entnommen werden, etwa die häufigsten Wörter, Wortzusammenhänge oder verwendete Gefühle (siehe Abschnitte 4, 5 und 6). Neben den in diesem Artikel vorgestellten Methoden des Text Minings, gibt es zahlreiche weitere Möglichkeiten, um Informationen aus Texten zu gewinnen. Einen Überblick über Techniken und Anwendungen geben *Preethi und Radha (2017)*. Darüber hinaus stellen *Injadat et al. (2016)* die am häufigsten verwendeten Methoden des Social Media Minings vor und *Hayat et al. (2019)* gehen auf Neuronale Netze ein, welche insbesondere für die Analyse von Bild- und Videodateien genutzt werden können.

#### 4. Die Twitter-Daten der Bundesregierung

Twitter ermöglicht den Zugriff auf ausgewählte Inhalte über eine **Datenschnittstelle**. Auf diese Schnittstelle kann mit der Software R (vgl. *R Core Team, 2019*) zugegriffen werden. Für diese Analyse wurden zum Stichtag 08.09.2019 die letzten 3.200 Tweets des Twitter Kontos @RegSprecher heruntergeladen. Der erhaltene Datensatz enthält insgesamt 88 Variablen. Neben dem Text der Tweets sind dies unter anderem Informationen zur Uhrzeit und zum Ort der Erstellung, sowie der Re-tweeted Status.

Damit die Inhalte der Tweets analysiert werden können, muss der Datensatz zunächst aufbereitet werden. In einem ersten Schritt werden hierzu alle Wörter eines Tweets in eine eigene Zeile und klein geschrieben sowie Satzzeichen und Zahlen entfernt. Diese Transformation erleichtert die spätere Analyse, wie zum Beispiel das Zählen einzelner Wörter. Auf eine Rechtschreibkorrektur und das Zusammenführen von Wörtern mit gleichem Wortstamm (z. B: Mensch und Menschen) wurde hier verzichtet.

Die *Abb. 1* (links) zeigt alle Wörter, die häufiger als 300-mal vorkommen. Auffällig ist, dass die häufigsten Wörter entweder spezifische Internetwörter, wie „t.co“ oder „https“ oder sogenannte **Stoppwörter**: „der“, „die“, „und“ sind. Diese Wörter tragen nicht viel zum Inhalt bei. Deshalb wurden diese in *Abb. 1* (rechts) ebenfalls entfernt. Damit sind die häufigsten Wörter in den Tweets ersichtlich: „Merkel“ und „Kanzlerin“, gefolgt von „Heute“ und „Deutschland“.

Neben der reinen Analyse des Inhalts über die Häufigkeit von Wörtern ist für Politiker und Unternehmen häufig interessant zu verstehen, welche Gefühle dabei ausgedrückt werden, etwa im Zusammenhang mit der eigenen Person, dem Unternehmen oder angebotener Produkte. Dies kann im Rahmen einer Sentiment Analyse geschehen.

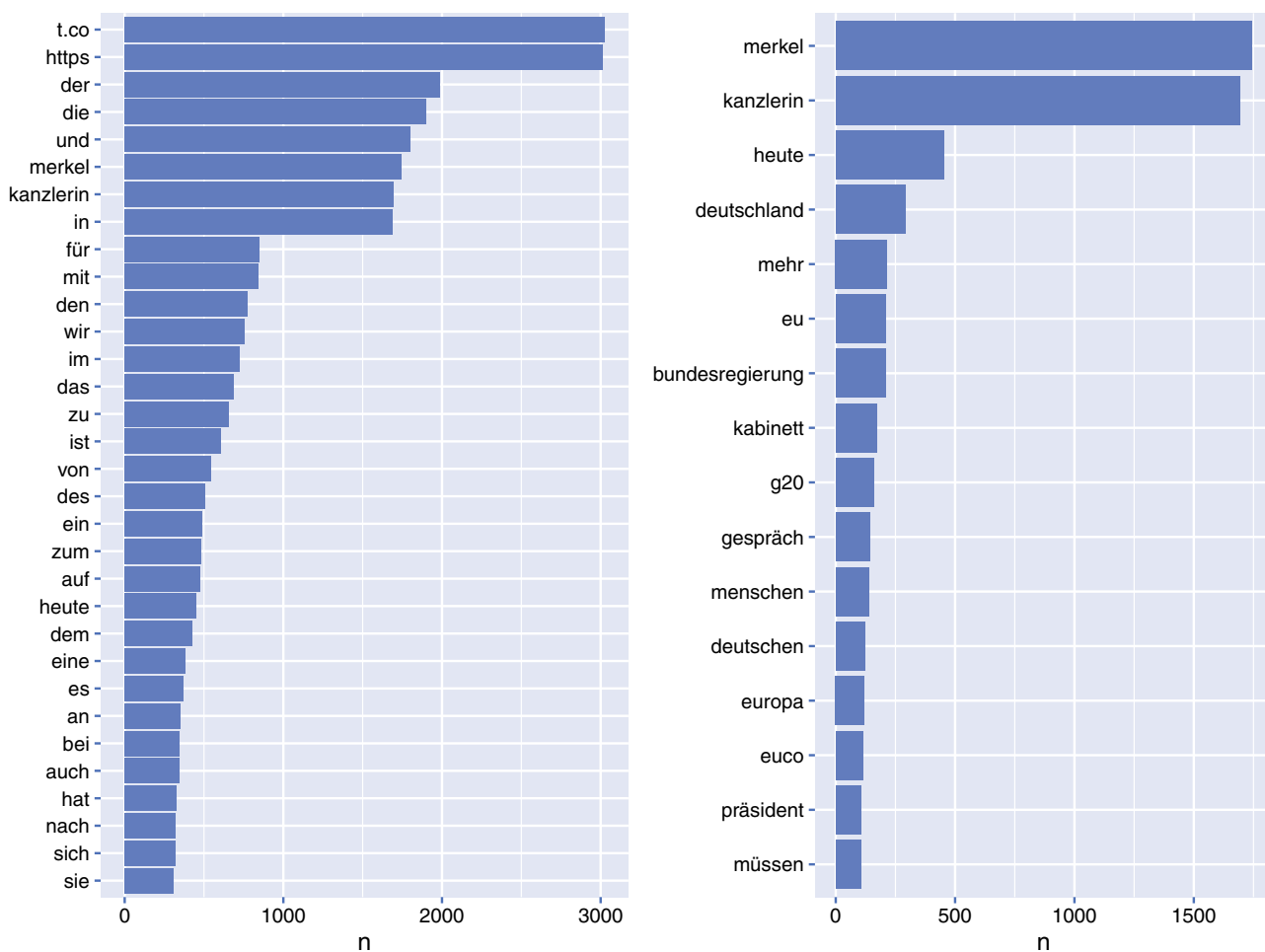


Abb. 1: Worthäufigkeiten mit Stoppwörtern (links) und ohne (rechts)

### 5. Analyse der verwendeten Emotionen

Viele Wörter sind mit Emotionen besetzt, so wird zum Beispiel das Wort „Glück“ meistens mit positiven Gefühlen belegt. Im Folgenden soll im Rahmen einer **Sentiment Analyse** untersucht werden, welche positiven und negativen Wörter die Bundesregierung verwendet.

Hierzu wird jedem Wort eines Tweets entweder das Attribut Positiv oder Negativ zugeordnet, sofern es sich einer dieser Kategorien zuordnen lässt. Zur Zuordnung wird das Lexikon *SentiWS* (**Sentiment Wortschatz**) der positiven bzw. negativen deutschen Wörter verwendet. *SentiWS* ist kostenlos verfügbar und enthält gegenwärtig ungefähr 1.650 positive und 1.800 negative Grundformen (vgl. *Remus et al.*, 2010). Wörter, die nicht in dem Lexikon enthalten sind, werden keiner Kategorie zugeordnet. Hierunter fallen auch Namen wie das am häufigsten in einem Tweet vorkommende Wort „Merkel“. Diese Analyse erfasst ausschließlich die Emotionen positiv und negativ. Auf eine Einteilung der Emotionen in „Graustufen“ und weitere Kategorien wie „Angst“ wird verzichtet.

Abb. 2 zeigt, dass „Kampf“, „Gewalt“ und „Ende“ die häufigsten verwendeten negativen und „Zusammenarbeit“, „Rat“ und „Glückwunsch“ die am häufigsten verwendeten positiven Wörter sind.

Sentiment Analysen können für Politiker und Unternehmen zahlreiche nützliche Informationen liefern, etwa über Emotionen im Zusammenhang mit einer bestimmten Marke. Eine Analyse einzelner Worte kann dabei aber unerwünschte Ergebnisse liefern. So ist das Wort „Ende“ in dem verwendeten Lexikon *SentiWS* negativ belegt. Dies ist zum Beispiel in Ausdrücken wie „Ende des Friedens“ oder „Ende des Lebens“ angemessen. Wird aber vom „Ende des Krieges“ gesprochen, so kann „Ende“ durchaus auch positiv belegt sein. Um die Inhalte besser einschätzen zu können, werden deshalb im nächsten Abschnitt Wort-Zusammenhänge analysiert.

### 6. Analyse der Wort-Zusammenhänge

Bisher wurden Wörter einzeln betrachtet. Zahlreiche Wörter wie Kampf, Gewalt, Glückwunsch und Ende stehen im Kontext aber nicht alleine, sondern in einem inhaltlichen Zusammenhang mit anderen Wörtern. Gegen was wird gekämpft? Gewalt wem gegenüber? Was ist zu Ende? Was wird beglückwünscht? Um diese Fragen zu beantworten werden im Folgenden die **Wort-Zusammenhänge** analysiert.

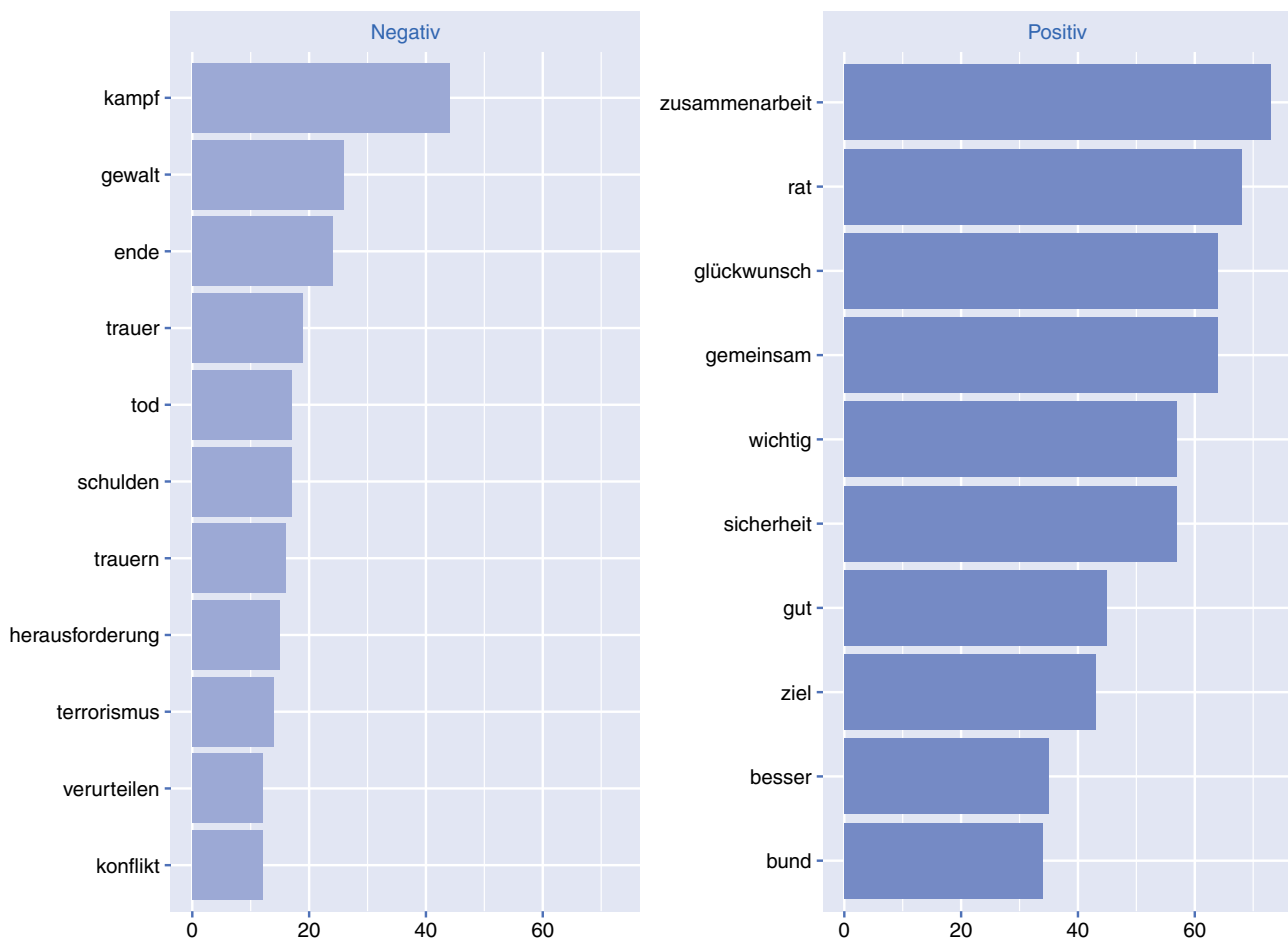


Abb. 2: Emotionen in den Tweets der Bundesregierung

### 6.1. Analyse der Wortkorrelationen

Zunächst soll untersucht werden, welche Wörter häufig gemeinsam in einem Tweet verwendet werden. Hierzu werden die Wörter und Tweets in einer sogenannten **Document-Term-Matrix** (DTM) dargestellt. Die DTM enthält die 3200 Tweets („Dokumente“) als Zeilen und alle in einem der Tweets vorkommenden Wörter („Terms“) als Spalten, insgesamt 3.200 Zeilen und 14.577 Spalten. Die Einträge der Matrix geben die Häufigkeit des Vorkommens des entsprechenden Wortes in dem jeweiligen Tweet an. Kommt ein Wort in einem Tweet vor, enthält die Matrix eine 1, ansonsten eine 0. Durch diese Darstellung kann die Korrelation der Wörter mit dem normalen **Korrelationskoeffizienten** berechnet werden.

Die *Tab. 1* enthält die häufigsten negativen und positiven Wörter sowie die zugehörigen Wörter mit den höchsten Wortkorrelationen. Der Korrelationskoeffizient ist in der Klammer angegeben. Damit können die oben aufgeworfenen Fragen beantwortet werden.

Bei den negativ belegten Wörtern geht es um den Kampf im Zusammenhang mit Terrorismus, Antisemitismus und Kriminalität. In den Tweets mit dem Thema Gewalt wird häufig von Frauen und sexueller Gewalt gesprochen. Beim Begriff Ende ist die Zuordnung nicht so eindeutig. Zum einen wird über das Ende des Weltkriegs und zum anderen über das Ende des Kohlebergbaus (DankeKumpel, ProsperHaniel) getwittert. Je nach Perspektive kann damit das Wort „Ende“ als positiv oder negativ interpretiert werden. Bei den positiven Worten wird die „Zusammenarbeit“ zusammen mit den Wörtern „multilateral“ und „vertrauensvoll“ verwendet. Besonders häufig wird hierbei das Land „Italien“ genannt. Im Zusammenhang mit „Rat“ wird über die Europäische Kommission, europäisch und Brüssel geschrieben. Schließlich enthalten die Korrelationen Hinweise darauf, dass „Herzliche“ „Glückwünsche“ für einen „Sieg“ bei der Olympiade 2018 in Pyeongchang getwittert wurden.

**Wort-Korrelationen** dienen dazu Wörter in einen Zusammenhang zu setzen. Dadurch kann analysiert werden, welche Wörter häufig zusammen in einem Tweet verwendet werden. Manchmal reicht dies aber nicht aus, denn bei der Berechnung der Wort-Korrelationen wird außer Acht gelassen, dass manchmal die direkte Wortfolge von Interesse ist.

So kann das Wort „Gewalt“ entweder dazu verwendet werden, um zur Gewalt aufzurufen („Gewalt gegen Frauen!“) oder um diese zu verhindern, etwa durch die Aussage „Keine Gewalt gegen Frauen!“. Auch die Aussagen „Ich liebe dich“ und „Ich liebe dich nicht“ unterscheiden sich fundamental voneinander, auch wenn die betrachteten Wortkorrelationen recht hoch sein können. In diesem Fall helfen N-Gramme weiter.

### 6.2. Analyse der Wort-Zusammenhänge mit N-Grammen

**N-Gramme** sind Worttupel, also Wortpaare der Länge N. Der in einem der Tweets verwendete Satz „In Deutschland wird es immer wärmer“ enthält zum Beispiel die folgenden 2-Gramme: „In Deutschland“, „Deutschland wird“, „wird es“, „es immer“ und „immer wärmer“. Durch diese Darstellung kann die direkte Wortfolge zweier (oder mehrerer) Wörter untersucht werden. *Abb. 3* zeigt häufig vorkommende 2-Gramme der untersuchten Tweets. Die Richtung der Pfeile geht vom erstgenannten zum zweitgenannten Wort. Anhand der Abbildung (vgl. *Silge/Robinson, 2019, Kapitel 4*) ist etwa zu erkennen, dass „Merkel“ häufig vor einer Tätigkeit verwendet wird: „Merkel“ spricht, reist, empfängt, begrüßt, gratuliert oder besucht. Dagegen folgt „Merkel“ häufig auf die Worte Angela, Chancellor und Kanzlerin.

Bisher wurden Informationen aus den Inhalten der Tweets untersucht. Für viele Politiker und Unternehmen ist häufig zusätzlich interessant zu wissen, *wo* Tweets gelesen werden, also die Frage: Wo ist der Standort der Follower des Regierungssprechers?

## 7. Geografische Analyse der Follower

Die verfügbaren *Twitter* Daten enthalten neben den Tweets auch Informationen über den **Standort**. Für das Konto des Regierungssprechers ist dies etwa der Standort Berlin. *Twitter* ermöglicht zudem den Zugriff auf die **Follower** eines Kontos und sogar auf den Standort dieser Follower, sofern der Standort durch den Benutzer freigegeben wurde. Damit kann eine geographische Analyse der Follower durchgeführt werden. Hierzu wurden die ersten 500 Follower heruntergeladen.

Tab. 1: Wortkorrelationen der häufigsten emotionalen Wörter

WORT	WÖRTER MIT DER HÖCHSTEN KORRELATION
KAMPF	Terrorismus (0,24), Antisemitismus (0,21), Kriminalität (0,21)
GEWALT	Hilfetelefon (0,48), sexueller (0,41), Frauen (0,40)
ENDE	Weltkrieg (0,30), dankekumpel (0,29), prosperhaniel (0,29)
ZUSAMMENARBEIT	Italien (0,21), Multilaterale (0,21), Vertrauensvolle (0,17)
RAT	EUCO (0,53), europäischen (0,52), Brüssel (0,28)
GLÜCKWUNSCH	Herzlichen (0,7), Pyeongchang2018 (0,38), Sieg (0,29)

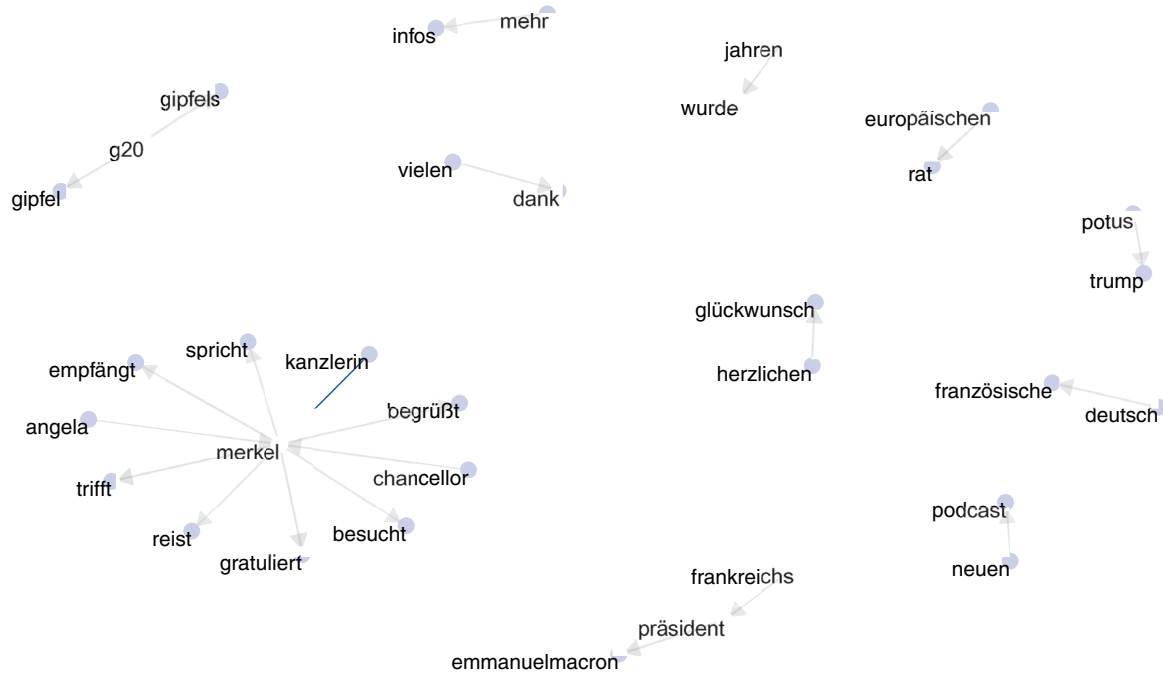


Abb. 3: 2-Gramme

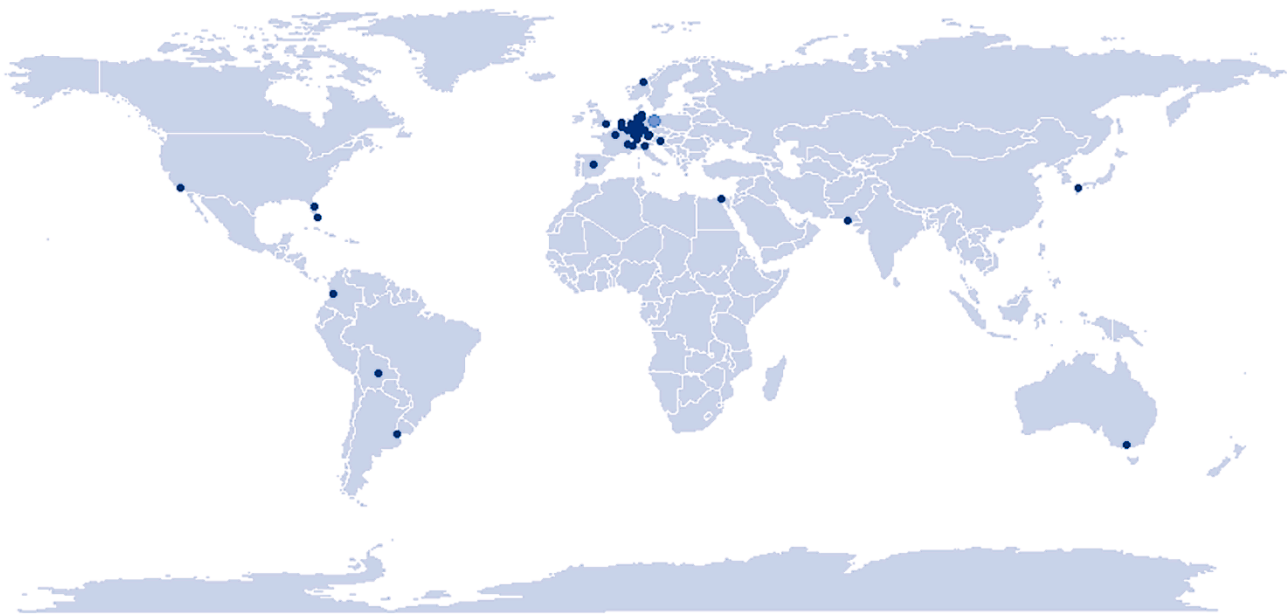


Abb. 4: Follower (Dunkelblau) und Bundesregierung (Mittelblau)

Die Abb. 4 (vgl. Leek, 2011) zeigt diejenigen der ausgewählten 500 Follower der Bundesregierung, welche den Standort freigegeben haben. Aus der Abbildung ist ersichtlich, dass der Schwerpunkt der Follower in Deutschland und Europa zu finden ist, es aber auch vereinzelt Follower auf anderen Kontinenten gibt.

### 8. Anwendungen im Unternehmenskontext

Social Media Mining bietet zahlreiche Anwendungsmöglichkeiten im Unternehmenskontext, etwa in den Bereichen Marketing, strategische Unternehmensführung und Risikomanagement.

Im Bereich Marketing und strategische Unternehmensführung kann etwa eine regionale Analyse der Follower helfen die folgenden Fragen zu beantworten: Wo erreiche ich meine (potentiellen) Kunden? In welchen Regionen besitzt mein Unternehmen wenige/viele Follower? Welche Infrastruktur, welches Bildungsniveau und welche Bevölkerungsdichte haben die Regionen mit den meisten Followern? Darüber hinaus können die Follower der Follower mit einbezogen werden und die Informationsverbreitung im Netzwerk untersucht werden. Dies ist etwa für **Influencer Marketing** interessant.

Zusätzlich können insbesondere Verfahren des **Machine Learning** verwendet werden, um Voraussagen zu tätigen:

Welche Follower werden am ehesten positiv auf ein bestimmtes Produkt reagieren? Welche Follower retweeten am ehesten einen bestimmten Tweet? Welche Inhalte werden am meisten geteilt? Wie wird mein Produkt im Vergleich zur Konkurrenz wahrgenommen? Hierzu können neben den Inhalten in den Texten auch die Inhalte in den verwendeten Bildern untersucht werden, um den Markt besser zu verstehen: Wieviele Personen sind auf den Bildern mit meinem Produkt im Durchschnitt zu erkennen? Wo wurden die Bilder aufgenommen?

Social Media Mining kann zudem im Rahmen eines Reputationsrisikomanagements helfen, geäußerte Kritik bis hin zu einem Shit-Storm schnell zu erkennen und ermöglicht somit frühzeitig Gegenmaßnahmen einzuleiten.

## 9. Zusammenfassung und Ausblick

*Twitter* hat jeden Monat im Durchschnitt 330 Millionen aktive Nutzer. Einer davon ist das *Twitter* Konto @Reg-Sprecher des Regierungssprechers der Bundesregierung. Unter diesem Konto wurden bereits mehr als 11.600 Tweets veröffentlicht. Um Rückschlüsse auf die verwendeten Wörter und Emotionen, aber auch über Wortzusammenhänge zu ziehen, können Methoden des Social Media Minings verwendet werden. Am Beispiel der Bundesregierung konnte gezeigt werden, dass über verschiedene Themen mit positiven und negativen Emotionen getwittert wurde. Hierzu zählen unter anderem die Olympiade 2018, das Ende des Weltkriegs, der Kampf gegen Terrorismus, Antisemitismus und Kriminalität sowie Gewalt gegenüber Frauen.

Neben dem reinen Text der Tweets sind in den *Twitter* Daten zahlreiche weitere Variablen enthalten. Diese ermöglichen tiefergehende Analysen, etwa zur Struktur der Follower. Am Beispiel des Regierungssprechers konnte gezeigt werden, dass der Standort der meisten der untersuchten Follower in Deutschland bzw. Europa liegt, es aber auch vereinzelt auf anderen Kontinenten weitere Follower gibt.

Wie viele Verfahren, birgt auch das Social Media Mining, neben der Möglichkeit detaillierter Analysen, ein gewisses Risiko zum Missbrauch, wie der Fall von *Cambridge Analytica* eindrucksvoll gezeigt hat (vgl. etwa *Davis*, 2015). Denn neben einer reinen Analyse der Inhalte können die Ergebnisse in einem zweiten Schritt verwendet werden, um Kampagnen durchzuführen und gezielt selektive Informationen zu verbreiten. Dadurch kann versucht werden, aktiv die Stimmung bzw. die Meinung in einer Bevölkerung zu verändern. Im Rahmen des **Cambridge Analytica Skandals** wurde deutlich, dass genau dies im Rahmen des Präsidentschaftswahlkampfes von *Donald Trump* (vgl. *Reinbold/Schnack*, 2016), aber auch der *VoteLeave* Kampagne im Vorfeld des *Brexit*s, getan wurde (vgl. *Cadwalladr*, 2019).

Ein verantwortungsvoll durchgeführtes Social Media Mining bietet jedoch zahlreiche faszinierende Möglichkeiten der Informationsgewinnung für Politiker, Unternehmen und Wissenschaft. Gerade im Rahmen der Forschung gibt es eine Vielzahl interdisziplinärer Fragestellungen, welche Fachwissen, etwa aus den Disziplinen Data Science, Soziologie, Informatik, Betriebswirtschaftslehre oder Volkswirtschaftslehre, benötigen.

## Literatur

- Cadwalladr, C.*, The Great British Brexit robbery how our democracy was hijacked, Online, URL: <https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy> (07.05.2019).
- Chen, H., Chiang, R.H.L., V.C. Storey*, Business Intelligence and Analytics: From Big Data To Big Impact, 2012, Mis Q. 36, Seiten 1165–1188.
- Clement, J.*, Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 (in millions), Online, URL: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/> (14.08.2019).
- Davies, H.*, „Ted Cruz campaign using firm that harvested data on millions of unwitting Facebook users“, Online, URL: <https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data> (11.12.2015).
- Eugster, M. J. A., T. Schlesinger*, osmar: OpenStreetMap and R, in The R Journal Vol. 5/1, June 2013, Online, URL: <https://journal.r-project.org/archive/2013/RJ-2013-005/RJ-2013-005.pdf> [2013].
- Ghosh, S., S. Roy, S.K. Bandyopadhyay*, A tutorial review on Text Mining Algorithms, 2012, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 4, Seiten 223–233.
- Gundecha, P., H. Liu*, Mining Social Media: A Brief Introduction, *Infoms* 2012, Online, URL: <http://dx.doi.org/10.1287/educ.1120.0105>
- Hayat et al.*, „Towards Deep Learning Prospects: Insights for Social Media Analytics.“, 2019, *IEEE Access*, Vol. 7, Seiten 36958–36979.
- Injadad, M. A. M., Salo F., A.B. Nassif*, Data Mining Techniques in Social Media: A Survey, 2016, *Neurocomputing*, Volume 214, Seiten 654–670.
- Kanter, O.*, Bing-bing-bing, Online, URL: <https://www.spiegel.de/politik/ausland/donald-trump-erklart-seine-tweets-bing-bing-bing-a-1174096.html> (22.10.2017).
- Leek, J.*, An R function to map your Twitter Followers, Online, URL: <https://simplystatistics.org/2011/12/21/an-r-function-to-map-your-twitter-followers/> (12.12.2011).
- Preethi, M., P. Radha*, A Survey Paper on Text Mining -Techniques, Applications And Issues, 2017, *IOSR Journal of Computer Engineering*, Seiten 46–51.
- R Core Team*, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Online, URL: <https://www.R-project.org/>, 2019.
- Radunski, M.*, Twitter-Mitarbeiter hat Trumps Account abgeschaltet, Online, URL: <https://www.faz.net/-hbi-93eh> (03.11.2017).
- Reinbold, F., T. Schnack*, Ich ganz alleine habe Trump ins Amt gebracht, Online, URL: <https://www.spiegel.de/netzwelt/netzpolitik/donald-trump-und-die-daten-ingenieure-endlich-eine-erklarung-mit-der-alles-sinn-ergibt-a-1124439.html> (06.12.2016).
- Remus, R., U. Quasthoff, G. Heyer*, SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In: Proceedings of the 7th International Language Resources and Evaluation (LREC'10), Seiten: 1168–1171, 2010.
- Schmidt, J.-H., M. Taddicken*, Handbuch Soziale Medien, 2017, Springer Fachmedien Wiesbaden.
- Silge, J., D. Robinson*, Text Mining with R, Online, URL: <https://www.tidytextmining.com/> (02.09.2019).