

Bachelorarbeit

Nadja Behn

Social Media Monitoring: Konzeption und Entwicklung
eines leichtgewichtigen Tools für kleine und
mittelständische Unternehmen

Nadja Behn

Social Media Monitoring: Konzeption und Entwicklung eines leichtgewichtigen Tools für kleine und mittelständische Unternehmen

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung
im Studiengang Bachelor of Science Wirtschaftsinformatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Olaf Zukunft
Zweitgutachter: Prof. Dr. Ulrike Steffens

Eingereicht am: 11. Februar 2020

Nadja Behn

Thema der Arbeit

Social Media Monitoring: Konzeption und Entwicklung eines leichtgewichtigen Tools für kleine und mittelständische Unternehmen

Stichworte

Social Media, Social Media Monitoring, Twitter, Python, Elasticsearch, Soziale Medien, Sentiments

Kurzzusammenfassung

Social Media Monitoring ermöglicht es, sich einen Überblick über die Online-Präsenz eines Unternehmens zu verschaffen. Das Ziel dieser Arbeit war es, ein leichtgewichtiges Social Media Monitoring Tool zu konzipieren und zu entwickeln. Die Evaluation zeigt, dass die definierten Anforderungen größtenteils umgesetzt werden konnten. Für das Einsetzen des Tools in einem Unternehmen sollte die Software weitergehend getestet und ausgebaut werden.

Title of Thesis

Social Media Monitoring: Design and implementation of a lightweight tool for small and medium-sized businesses

Keywords

social media, social media monitoring, Twitter, Python, Elasticsearch, sentiment analysis

Abstract

Social Media Monitoring allows companies to get an overview of the online presence of their business. The goal of this bachelor thesis was to design and implement a lightweight social media monitoring tool. The evaluation of this tool showed that the defined requirements could mostly be implemented. To use the tool in a company the tool should be tested even more and it should be extended by at least the requirements that were not implemented yet.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Zielsetzung	1
1.2	Abgrenzung	2
1.3	Aufbau der Arbeit	2
2	Grundlagen	3
2.1	Kleine und mittelständische Unternehmen	3
2.2	Social Media	4
2.2.1	Definition	4
2.2.2	Entstehung	6
2.2.3	Aufbau, Nutzen und Kritik an Social Media	8
2.2.4	Twitter	14
2.3	Social Media Monitoring	15
2.3.1	Social Media Monitoring für kleine und mittelständische Unternehmen	16
2.3.2	Twitter im Social Media Monitoring	16
2.4	Elastic Stack	16
2.4.1	Elasticsearch	17
2.4.2	Kibana	17
3	Analyse	18
3.1	Anforderungsanalyse	18
3.1.1	Funktionale Anforderungen	18
3.1.2	Nicht-funktionale Anforderungen	20
3.2	Existierende Vergleiche	21
3.3	Existierende Software	22
4	Konzeption	25
4.1	Unterscheidung fachliche und technische Konzeption	25

4.2	Fachliche Konzeption	25
4.2.1	Beschreibung des Tools	26
4.3	Technische Konzeption	28
4.3.1	Komponenten	29
4.3.2	Big Data Lebenszyklus	31
5	Realisierung und Test	33
5.1	Verwendete Technologien und Tools	33
5.1.1	Twitter	33
5.1.2	Python	33
5.1.3	Elasticsearch	34
5.1.3.1	Kibana	35
5.2	Implementierung	36
5.3	Erhobene Daten	37
5.3.1	Diagramme und Auswertung	38
5.3.1.1	Sentiments	38
5.3.1.2	Erwähnungen	38
5.3.1.3	Länder	40
5.3.1.4	Suche	41
5.3.1.5	Tweets des Unternehmens	42
5.4	Tests	42
6	Evaluation	44
6.1	Anforderungsabgleich	44
6.1.1	Funktionale Anforderungen	44
6.1.2	Nicht-funktionale Anforderungen	45
6.2	Kritische Betrachtung	46
7	Zusammenfassung und Ausblick	48
7.1	Zusammenfassung	48
7.2	Ausblick	49
	Abbildungsverzeichnis	50
	Tabellenverzeichnis	51
	Literaturverzeichnis	52

A Inhalt der CD-ROM	57
Selbstständigkeitserklärung	58

1 Einleitung

Täglich werden unzählige Postings in sozialen Netzwerken veröffentlicht, dieses Ausmaß kann kaum von Menschenhand überschaut und analysiert werden. Doch die Wichtigkeit als Unternehmen den Überblick über Erwähnungen, Kritik und Lob zu behalten ist enorm. Negative Postings, zum Beispiel von Influencern, können dazu führen, dass Aktienwerte sinken, das Image verschlechtert wird und vieles mehr. Zu diesem Zweck existieren immer mehr Social Media Monitoring Tools, die die Online-Kanäle eines Unternehmens beobachten und Daten speichern. Dies ist besonders für kleine und mittelständische Unternehmen wichtig, da die Menge der Interaktionen potentiell nicht so hoch ist wie bei großen Unternehmen, weshalb einzelne Postings mehr herausstechen.

In dieser Arbeit soll daher ein prototypisches Social Media Monitoring Tool für kleine und mittelständische Unternehmen konzipiert und entwickelt werden. In diesem Kapitel werden Zielsetzung und Abgrenzung definiert und danach der Aufbau der Arbeit dargestellt.

1.1 Zielsetzung

Ziel dieser Arbeit ist es, ein prototypisches leichtgewichtiges Social Media Monitoring Tool für kleine und mittelständische Unternehmen zu konzipieren, zu entwickeln und anschließend zu evaluieren. Des Weiteren soll auf den Nutzen eines solchen Tools für kleine und mittelständische Unternehmen eingegangen werden. Wesentliche Fragestellung hierbei ist, wie und ob sich aus den gesammelten (Beispiel-) Daten nützliche Informationen extrahieren lassen, um diese Erkenntnisse dann auf kleine und mittelständische Unternehmen substituieren zu können.

1.2 Abgrenzung

Ziel dieser Arbeit ist es nicht, ein einsatzbereites Social Media Monitoring Tool zu entwickeln. Aufgrund des beschränkten Zeitrahmens und der beschränkten Ressourcen werden Themen wie zum Beispiel die (technische) Integration in ein Unternehmen sowie das Testen auf verschiedenen Betriebssystemen und mit mehreren Benutzern nicht betrachtet.

1.3 Aufbau der Arbeit

Diese Arbeit ist wie folgt aufgebaut:

In Kapitel 1 wird in das Thema der Bachelorarbeit eingeführt sowie die Zielsetzung und Abgrenzung definiert. Außerdem wird der Aufbau der Bachelorarbeit dargelegt.

In Kapitel 2 werden grundlegende Themen wie kleine und mittelständische Unternehmen erläutert, Social Media definiert und beschrieben, Social Media Monitoring beleuchtet und der Elastic Stack vorgestellt.

In Kapitel 3 erfolgt zunächst die Anforderungsanalyse, danach werden existierende Software und Vergleiche betrachtet.

In Kapitel 4 wird ein Konzept auf Grundlage der in Kapitel 3 durchgeführten Analyse erstellt.

In Kapitel 5 folgt eine Betrachtung der verwendeten Technologien und Tools, sowie eine Beschreibung der Implementierung. Im Anschluss werden die erhobenen Daten genauer betrachtet und analysiert und zuletzt folgt ein Test der Software.

Abschließend fasst Kapitel 7 die Inhalte dieser Arbeit zusammen.

2 Grundlagen

In diesem Kapitel soll ein grundlegendes Verständnis für die Themen kleine und mittelständische Unternehmen, Social Media, Social Media Monitoring und dem Elastic Stack geschaffen werden.

2.1 Kleine und mittelständische Unternehmen

Diese Arbeit soll vor allem kleinen und mittelständischen Unternehmen dienen, die mit Hilfe von Social Media Monitoring beispielsweise ihre Reichweite erhöhen, den Bekanntheitsgrad erweitern und/oder Kampagnen planen wollen. Zunächst muss die Frage geklärt werden, was kleine und mittelständische Unternehmen ausmacht, beziehungsweise wie diese definiert sind. Zu diesem Zweck wird die Definition der Europäische Kommission herangezogen, die kleine und mittelständische Unternehmen wie folgt definiert: eine Kombination von Beschäftigtenzahl und Jahresumsatz oder Jahresbilanzsumme [24].

Unternehmensgröße	Beschäftigtenzahl	Jahresumsatz	Jahresbilanzsumme
klein	bis 49	bis 10 Millionen	bis 10 Millionen
mittel	bis 249	bis 50 Millionen	bis 43 Millionen

Tabelle 2.1: Definition kleine und mittelständische Unternehmen

Charakteristisch für kleine und mittelständische Unternehmen im Kontext dieser Bachelorarbeit sollte sein, dass diese Social Media Kanäle haben oder aufbauen wollen.

2.2 Social Media

Im Folgenden wird anhand verschiedener Quellen eine Definition für Social Media erarbeitet, anschließend wird auf die Entstehung von Social Media eingegangen. Des Weiteren wird der grundlegende Aufbau und der Nutzen von Social Media Plattformen aufgezeigt. Darauf folgt eine Betrachtung einiger kritischer Stimmen. Außerdem wird das soziale Netzwerk Twitter gesondert vorgestellt, da das Hauptaugenmerk dieser Arbeit auf diesem als Datenquelle liegt.

2.2.1 Definition

Social Media, soziale Medien oder soziale Netzwerke gibt es sowohl im digitalen als auch im realen Leben. Ein soziales Netzwerk im realen Leben könnte zum Beispiel der Freundes- oder Kollegenkreis sein. Im Folgenden sind digitale soziale Netzwerke gemeint, wenn die Rede von sozialen Netzwerken ist.

Für Social Media gibt es viele Definitionen, nachfolgend werden einige Definitionen betrachtet und auf diese wird dann im weiteren Verlauf dieser Arbeit Bezug genommen.

Kreutzer definiert soziale Netzwerke als Plattformen, auf denen es möglich ist „Beziehungen zu Business-Partnern und/oder Privatpersonen auf- und auszubauen“. Ergänzend können Personen mit den gleichen Interessen gefunden werden [25, S. 61].

Decker erklärt Social Media in zwei verschiedenen Formen. Wenn man den Ausdruck in der Einzahl betrachtet, stellt Social Media eine Plattform bereit, auf der Benutzer „miteinander interagieren, dabei eigene Inhalte kreieren (sogenannten User-Generated-Content) und diese miteinander austauschen.“ In der Mehrzahl benennt *Decker* Social Media als „die damit verbundenen digitalen Kommunikations-Dienste und Technologien, um die oben erwähnte Interaktion zwischen den Nutzern zu ermöglichen.“ Er stuft zusätzlich Begriffe wie: Social Media, soziale Medien, soziale (Anwendungs-) Plattformen, Social-Media-Kanäle (oder -Plattformen) und Social Web als synonym ein [11, S. 50].

Böker beschreibt Social Media ähnlich wie *Decker*, allerdings unterscheidet *Böker* nicht zwischen Einzahl und Mehrzahl. *Böker* legt außerdem besonderes Augenmerk darauf, dass verschiedene Nutzer Inhalte gemeinschaftlich erweitern können. Ferner verdeutlicht er, dass Social-Media-Plattformen „erst durch die Beiträge, die hochgeladenen Filme,

Kommentare und Profile der Nutzerinnen und Nutzer“ individualisiert werden, zuvor stellen sie „nur ein leeres Gerüst dar“ [9, S. 9 f.].

Zahn charakterisiert Social Media ähnlich wie *Böker*: Benutzer teilen Inhalte, Wissen, Erlebnisse und vieles mehr. Auch sie betont die gemeinsame Gestaltung der Inhalte und die hiermit verbundene Kontaktknüpfung. Ergänzend hinzukommt, dass „die Grenze zwischen Produzent und Konsument verschwimmt“ [48, S. 9].

Gabriel und Röhrs definieren soziale Netzwerke und Social Media getrennt. Ein soziales Netzwerk ist eine zum Austausch dienende „lose Verbindung von Menschen“ in einer Online-Gemeinschaft. Social Media hingegen definieren sie ähnlich wie *Zahn* [33, S. 12].

Ellison und Boyd benennen drei Indikatoren für soziale Netzwerke:

- die Erstellung eines Nutzerprofils in einem System,
- die Kontaktknüpfung mit anderen Benutzern und
- das Studieren der entstandenen Verbindungen [13, S. 2].

Der *Bundesverband Digitale Wirtschaft* (BVDW) fügt zu den bisher genannten Punkten hinzu, dass der Austausch in sozialen Netzwerken verschiedenartig möglich ist: „One-to-one- sowie One-to-many-Kommunikation“ [7, S. 3].

Bruhn und Hadwich vermitteln knapp, dass die „Beziehungen“ zwischen den Benutzern Social-Media-Plattformen ausmachen [5, S. 3].

Babka unterscheidet wie *Decker* Social Media in der Einzahl und Mehrzahl. Sie benennt soziale Medien in der Einzahl als „das Phänomen Social Media“ und in der Mehrzahl als „soziale Netzwerke“ [2, S. 1].

Aus diesen Definitionen lässt sich schließen, dass die Begriffe *Social Media*, *soziale Medien* und *soziale Netzwerke* sehr eng miteinander zusammenhängen bzw. synonym verwendet werden. *Kreutzer* beschreibt in seiner Definition sogar den Begriff *soziales Netzwerk* mit dem Ausdruck *Social-Media-Plattform*. Auch in *Babkas* Definition werden die beiden Ausdrücke miteinander beschrieben. Getrennt definiert werden die Terme von *Roland und Röhrs*, allerdings ist deren Definition für Social Media sehr ähnlich zu *Ellison und Boyd* und *BVDWs* Definitionen für soziale Netzwerke. Daher entscheide ich mich in dieser Arbeit dafür, die Begriffe synonym zu behandeln.

Kurz zusammengefasst ist die Kernaussage, die alle Definitionen der herangezogenen Literatur vereint, dass soziale Medien es ihren Benutzern erlauben Inhalte, Meinungen und vieles mehr zu teilen und zu gestalten, ein Profil zu erstellen und sich mit anderen Benutzern dieser Plattform zu verbinden.

2.2.2 Entstehung

Die Entstehung von Social Media begann im Jahre 1969, als Computer zum ersten Mal über weite Distanzen miteinander verbunden wurden. In den siebziger Jahren wurden erstmals elektronisch schwarze Bretter implementiert, auf denen Informationen ausgetauscht werden konnten. Der Austausch über die elektronischen schwarzen Bretter war allerdings nur per Telefon und mit wechselseitigem Ausschluss¹ möglich.

Massentauglich wurde das Internet, als 1989 das World Wide Web entwickelt wurde. Zuvor war das Internet eher für universitäre, staatliche und militärische Zwecke eingesetzt worden. Im Jahre 1996 wurde die Messaging-Plattform ICQ online geschaltet und erfreute sich schnell an Beliebtheit. Ein Jahr später wurde Google.com registriert und entwickelte sich, wie wir heute wissen, zu einem der größten Unternehmen weltweit.

2003 wurde das soziale Netzwerk MySpace mit dem Schwerpunkt Musik gegründet sowie LinkedIn und Xing, auf denen Geschäftskontakte geknüpft und gepflegt werden können. Des Weiteren wurde auch Skype in diesem Jahr gegründet. Skype war in dem Sinne besonders, als dass es erstmals möglich war kostenlos über das Internet zu telefonieren. Mit der Gründung von Facebook im Jahr 2004 spielte auch das Prinzip „Web als Plattform“ eine wichtige Rolle. Für die breite Masse und mobilen Geräte wurde Facebook 2006 geöffnet und wuchs bis Ende des Jahres 2006 rasant auf 12 Millionen Mitglieder heran. Nach und nach wurden Webseiten offener und es wurden öffentliche Schnittstellen (APIs) angeboten, sodass aus „Walled Gardens“² Plattformen wurden. In dem Zuge öffnete z. B. Facebook seine Tore für Webseiten, damit die Benutzer von Facebook dort mit ihrem Facebook-Account kommentieren konnten. Der Meilenstein von zwei Milliarden aktiven Nutzern wurde am 27. Juni 2017 erreicht³.

¹Es ist nur einem Benutzer zur Zeit möglich schreibend auf die Daten zuzugreifen

²Geschlossene Plattform, kein Zugriff auf die Daten von außen

³Siehe: Abbildung 2.1



Abbildung 2.1: Erreichen der zwei Milliarden aktiven Nutzer (Post von Mark Zuckerberg)

Natürlich hat nicht nur Facebook zur Weiterentwicklung und Verbesserung von sozialen Netzwerken beigetragen, sondern maßgeblich auch viele von den noch heute existierenden Netzwerken. Eine Plattform, die es heute nicht mehr gibt, ist die VZ-Gruppe. Aus dieser Netzwerk-Gruppe war StudiVZ die erste Plattform und wurde 2005 gegründet. Die komplette VZ-Gruppe wurde allerdings, größtenteils von Facebook, abgelöst und musste im September 2017 Insolvenz anmelden. Heute noch relevant und vielleicht relevanter als je zuvor sind YouTube und Twitter. YouTube startete 2005 mit dem Motto „Broadcast Yourself“⁴. Seit 2006 ist YouTube eine Tochtergesellschaft vom Internetgiganten Google. Auf YouTube können Videos hochgeladen, kommentiert, geliked und angesehen werden. Zukunftsweisend ist auch das soziale Netzwerk Twitter durch das, mit seinen 140 Zeichen⁵ Nachrichten („Tweets“) eine neue Art der Interaktion entstanden ist. Diese Art eines sozialen Netzwerks nennt sich Microblogging-Plattform. Auch der Sofortnachrichten-Dienst WhatsApp ist nicht mehr aus dem Alltag wegzudenken. Die Gründung von WhatsApp fand im Jahr 2009 statt und im Jahr 2014 wurde WhatsApp dann von Facebook übernommen. Mit WhatsApp ist es möglich Videos, Fotos, Textnachrichten und viele andere Dateien zu verschicken. Täglich werden über WhatsApp ca. 55 Milliarden Nachrichten zwischen mehr als einer Milliarde Nutzern⁶ verschickt[3].

⁴zu Deutsch: „Sende dich selbst“.

⁵Im November 2017 auf 280 Zeichen angehoben.

⁶Siehe: Abbildung 2.2



Abbildung 2.2: WhatsApp in Zahlen

Das auf Fotos basierende soziale Netzwerk Instagram wurde 2010 gegründet und wurde, wie auch Whatsapp, von Facebook übernommen, allerdings schon zwei Jahre früher. Populär wurde Instagram vor allem dadurch, dass es möglich war Filter über Fotos zu legen. Auch Google probierte mehrfach, sich im Bereich Social Media zu etablieren und 2010 wurde Google+ live geschaltet. Bei dem sozialen Netzwerk Google+ gibt es ähnlich wie bei Facebook eine eigene Profilseite. Viel erfolgreicher war Google+ mit dem Benutzerprofil, welches eine automatische Anmeldung auf allen zu Google gehörenden Webseiten mit dem sogenannten „Single-Log-In“ und gleichzeitig eine bessere Art der Datenspeicherung ermöglicht. Hierdurch können die Nutzer nun auch persönlich und individuell angesprochen werden. Das Single-Log-In führte dazu, dass die Anzahl der registrierten Nutzer von Google+ extrem anstieg, wobei die Anzahl der aktiven Nutzer minimal ist [11].

2.2.3 Aufbau, Nutzen und Kritik an Social Media

Aufbau Mit einer klaren Definition für Social Media wird nun geklärt, wie der Ablauf der Nutzung dieser Plattformen ist. Der Beitritt in eine Social Media Plattform ist grundsätzlich ähnlich. Um sich anzumelden ist der erste Schritt die Registrierung,

meist mit Hilfe einer E-Mail-Adresse. Im zweiten Schritt wird ein Profil erstellt und mit Informationen und Fotos versehen. Danach folgt im dritten Schritt die Vernetzung mit anderen Personen. Zu diesem Zweck werden bei vielen sozialen Netzwerken Personen vorgeschlagen, die man kennen könnte⁷ oder Seiten, die einen interessieren könnten [25, S. 63].



Abbildung 2.3: Facebook: „Personen, die du kennen könntest“

Nutzen Für Unternehmen sind soziale Netzwerke wichtig, weil sie auf diesen die Konsumenten und Interessenten ihrer Produkte und Dienstleistungen erreichen beispielsweise durch gezielte Kampagnen. Außerdem können Unternehmen auf sozialen Netzwerken mit ihrer Zielgruppe kommunizieren. Der persönliche Austausch ist wichtig für die Bindung zwischen Unternehmen und Kunden/Interessenten. Die Kommunikation über Social Media ist individueller als über andere Werbekanäle. Um die Aufmerksamkeit der Kunden zu erwecken, ist es hilfreich Content (Inhalte) zu teilen, der unterhaltsam und/oder nützlich für den Kunden ist [25, S. 4]. Die amerikanische Fast-Food-Kette *Taco Bell* teilt zum

⁷Siehe: Abbildung 2.3

Beispiel Inhalte ihrer Follower. Dies schafft den Anreiz, dass Follower etwas posten, in der Hoffnung, dass Taco Bell den eigenen Inhalt teilt.⁸



Abbildung 2.4: Twitter: @tacobell teilt den Inhalt von @marharf

Auch für die Entscheidungsfindungen ist die Kommunikation und das Einbinden der Follower wichtig, zum Beispiel über Umfragen⁹ [25, S. 64].

⁸Siehe: Abbildung 2.4

⁹Siehe: Abbildung 2.5



Abbildung 2.5: Twitter: Umfrage an Follower von @nailogical

Die Beeinflussung findet nicht nur einseitig statt, auch Privatpersonen haben durch Social Media eine ganz neue Art des Einflusses auf Unternehmen in der Öffentlichkeit. Jede Person kann zur Meinungsbildung über ein Unternehmen beitragen, beispielsweise über Blogs, Rezensionen aber auch auf der betrachteten Plattform Twitter [1, S. 586]. Auch Unternehmen können eine neue Reichweite erzielen. Die potentiellen Kunden müssen nicht mehr die Webseite des Unternehmens besuchen, sondern können einfach dem Social-Media-Account folgen, um auf dem neuesten Stand zu bleiben. Mit dieser neuen Art der Öffentlichkeit und der Datenverbreitung entstehen täglich immense Datenmengen. 90% der Gesamtdaten sind innerhalb der letzten zwei Jahre¹⁰ entstanden. Social Media Seiten wie Twitter sammeln jegliche Daten, die bei der Nutzung ihrer Produkte und Services entstehen. Diese werden dann ausgewertet. Die Auswertung macht es möglich zum Beispiel gezieltere Werbung¹¹ zu schalten [22].

¹⁰Weltweit verfügbare Daten, stand 2013 [44]

¹¹siehe: Abbildung 2.6



Abbildung 2.6: Twitter: Promoted Tweet

Kritik Die meiste Kritik bekommen soziale Medien in Bezug auf das Thema Datenschutz. Larry Sanger, der Mitgründer von Wikipedia, sagte in einem Interview, dass die sozialen Netzwerke die Daten ihrer Nutzer benutzen um Profit zu machen und dabei die Privatsphäre und Datensicherheit ihrer Nutzer missbrauchen. Zusätzlich kritisiert er die Algorithmen stark: „They can shape your experience, they can control what you see, when you see it and you become essentially a cog in their machine“ [36]. Der Verbraucherschutz bemängelt beim Thema Datenschutz zusätzlich, dass die Datenverarbeitung intransparent ist und es dem Nutzer erschwert wird den Überblick und die Gewalt über die eigenen Daten zu behalten [21]. Soziale Netzwerke wie Facebook und Twitter regen den Nutzer stetig dazu an Daten preiszugeben, zum Beispiel in Form von Statusupdates.¹²



Abbildung 2.7: Twitter: What's happening?

¹²siehe: Abbildung 2.7

Diese Aufforderungen erwecken den Eindruck, dass es jemanden interessiert, was gerade im Leben des Benutzers passiert [16, S. 144]. So nehmen Soziale Medien Einfluss auf die Informationsbeschaffung, die Art das eigene Leben zu teilen und vieles mehr. Schmidt beschreibt, dass „in den und durch die sozialen Medien zahlreiche Belanglosigkeiten zirkulieren.“ [35, S. 288]. Ein großes Thema, gerade in den Generationen Y¹³ und Z¹⁴, ist auch die Angst etwas zu verpassen¹⁵, wenn man nicht dauerhaft online ist [30, S. 1841]. Viele Benutzer haben das Gefühl, auf allen Social-Media-Kanälen gleichermaßen aktiv sein zu müssen, weil zum Beispiel in der Online-Welt mehrere „Persönlichkeiten“ von dem Benutzer existieren. Diese „Persönlichkeiten“ haben dann unterschiedliche Follower, Freunde oder Abonnenten, die separat mit Inhalten gefüttert werden. Aktuell existiert die „Dolly-Parton-Challenge“¹⁶, die zeigt, wie unterschiedlich Personen sich auf verschiedenen sozialen Netzwerken verhalten und darstellen. Zugleich werden alle Social-Media-Kanäle beobachtet, um umgehend auf Neuigkeiten reagieren zu können [35, S. 288].



Abbildung 2.8: Dolly-Parton-Challenge

¹³Generation Y sind Menschen die zwischen 1980 und 2000 geboren wurden

¹⁴Generation Z sind Menschen die ab 2000 geboren wurden

¹⁵„Fear of missing out“

¹⁶siehe: Abbildung 2.8

2.2.4 Twitter

Das Hauptaugenmerk dieser Arbeit liegt auf dem sozialen Netzwerk Twitter. Wie bereits in der Entstehung beschrieben, ist Twitter eine Microblogging-Plattform. Der Unterschied zu einem Blog ist, dass die Zeichenzahl begrenzt ist. Twitter ist weltweit die bekannteste Microblogging-Plattform und wird in über 40 Sprachen angeboten. Bei Twitter kann in Echtzeit kommuniziert werden. Die abgeschickten Statusmeldungen nennen sich „Tweets“ und diese können andere Nutzer „Retweeten“ (zu ihrer eigenen Seite hinzufügen und mit den eigenen Followern teilen). Um Tweets einer bestimmten Themengruppe zu finden oder hinzuzufügen, können sie mit einem „Hashtag“ (#) versehen werden.



Abbildung 2.9: Twitter: #HAW

Eine weitere Besonderheit an Twitter ist, dass die Tweets von jedem gelesen werden können, auch ohne die Erstellung eines Profils [18].

Datenstruktur Twitter speichert die Tweets als JSON Objekte. JSON Objekte sind semi-strukturierte Daten, weil sie keiner allgemeinen Struktur unterliegen. Die Datenfelder können trotzdem größtenteils gleich sein, aber auch je Datensatz variieren. Diese

Flexibilität geht mit einem Mehraufwand bei dem Auslesen der Daten einher [19, S. 13].

Beispiel Tweet Objekt [42]:

```
{
  "created_at": "Wed Oct 10 20:19:24 +0000 2018",
  "id": 1050118621198921728,
  "id_str": "1050118621198921728",
  "text": "To make room for more expression , we will now
count all emojis as equal.. ",
  "user": {},
  "entities": {}
}
```

2.3 Social Media Monitoring

Social Media Monitoring ist die kontinuierliche Überwachung bedeutsamer Themen und Erwähnungen in sozialen Netzwerken [1, S. 587]. Die hierdurch gesammelten Daten werden aufbereitet und können beispielsweise einen Überblick über die Markttrends liefern [6, S. 82].

Das Angebot an Social Media Monitoring Tools ist groß. Im Internet finden sich mit einer einfachen Suche verschiedenste Tools mit unterschiedlichen Funktionalitäten. Mit Hilfe dieser Tools lassen sich Entscheidungen von Unternehmen erleichtern, Sentiments¹⁷ bestimmen, Kritik und Lob analysieren und vieles mehr. Die Daten können dann benutzt werden, um die (potentiellen) Kunden zu beeinflussen und deren Verhalten zu analysieren [44, S. 775 f.]. Der Nutzen von Social Media Monitoring Tools variiert stark zwischen verschiedenen kleinen und mittelständischen Unternehmen. Einige von ihnen besitzen keine Online-Präsenz, somit hat Social Media Monitoring für diese gar keine Relevanz. Diese Bachelorarbeit soll einen Nutzen für diejenigen kleinen und mittelständischen Unternehmen bieten, die verstärkt auf Social Media Monitoring setzen und ihre Social Media Kanäle pflegen.

¹⁷Empfindung oder Stimmung zu einem Thema.

Zieldefinition Bevor mit dem Monitoring begonnen wird, sollten die Ziele des Monitorings definiert werden, denn nur mit konkreten Zielen kann der Erfolg gemessen werden.

Erfolgsmessung In sozialen Netzwerken geht es meist um die Anzahl der Interaktionen (Likes, Retweets, Kommentare). Diese Anzahl verglichen mit der Gesamtreichweite ist messbar [1, S. 588]. Außerdem ist es beispielsweise möglich, die Reichweite der Tweets vor und nach einer Kampagne zu vergleichen.

2.3.1 Social Media Monitoring für kleine und mittelständische Unternehmen

Social Media Monitoring ist für kleine und mittelständische Unternehmen besonders interessant, da es die einzige Möglichkeit ist Feedback (positiv und negativ) in Echtzeit zu erheben. Es kann durch den Einsatz von Social Media Monitoring beispielsweise die Zielgruppe im Auge behalten oder die Konkurrenz beobachtet werden. Wichtig hierbei ist die oben erwähnte Zieldefinition, um einen Nutzen aus den Daten ziehen zu können.

2.3.2 Twitter im Social Media Monitoring

Wenn Unternehmen zum Beispiel über Twitter Angebote, Rabatte und Aktionen teilen, bekommt der (potentielle) Kunde das Gefühl, dass er bevorzugt wird, weil er diesem Account folgt. Die Kaufbereitschaft steigt durch diese semi-persönliche Ansprache. Twitter lässt sich in Echtzeit beobachten, was es Unternehmen ermöglicht direkt zu reagieren. Ein weiterer Vorteil bei Twitter ist, dass nicht nur die Person die sich geäußert hat die Reaktion des Unternehmens sieht, sondern auch alle anderen Personen Zugriff auf diese haben [10, S. 82].

2.4 Elastic Stack

Die zu sammelnden Daten sollen gespeichert und visualisiert werden, hierfür hat sich nach einigen Abwägungen, die im Folgenden beschrieben werden, der Elastic Stack als besonders passend herausgestellt. Der Elastic Stack ist ein Zusammenspiel von mehreren Tools, die die Firma Elastic anbietet. Im Folgenden wird auf Elasticsearch und Kibana eingegangen.

2.4.1 Elasticsearch

Elasticsearch ist ein Tool, welches große Datenmengen speichern und durchsuchen kann. Außerdem können Daten visualisiert¹⁸ und analysiert werden. Bei Elasticsearch kann mit Volltextabfragen gearbeitet werden [29, S. 181]. Des Weiteren ist Elasticsearch fehlertolerant und skalierbar. Die API von Elasticsearch erlaubt den Zugriff mit vielen Programmiersprachen [34, S. 171]. Mit Elasticsearch ist es möglich, in strukturierten sowie in unstrukturierten Daten zu suchen und diese zu analysieren. Es wird außerdem mit der Schnelligkeit der Datenabfragen geworben, welche durch Indexierung möglich ist. Überdies arbeitet Elasticsearch mit Repliken, sodass die Daten sicher vor Verlusten und stets verfügbar sind [12, <https://www.elastic.co/de/what-is/elasticsearch>]. Der Nachteil von Elasticsearch ist, dass jeder Benutzer auf alle Daten zugreifen kann [40]. Im Vergleich zu einem RDBMS ist Elasticsearch schemafrei, das heißt zwei Dokumente des gleichen Typs können verschiedene Felder haben (es wird kein Speicher für Null-Werte verschwendet). Im Vergleich zu Solr ist Elasticsearch weniger komplex und detailliert [41].

2.4.2 Kibana

Mit Hilfe von Kibana ist es möglich, die in Elasticsearch gespeicherten Daten zu visualisieren und beispielsweise durch Diagramme aufzubereiten [12, <https://www.elastic.co/de/products/kibana>]. Für die Visualisierung stehen viele Möglichkeiten bereit, aus denen ausgewählt werden kann. Der Konkurrent Grafana schneidet insofern schlechter ab, als dass er nicht dafür ausgelegt ist (Voll-Text-)Suchanfragen zu formulieren. Allerdings erlaubt Grafana es Daten aus vielen unterschiedlichen Quellen zu beziehen, während Kibana nur in Kombination mit Elasticsearch funktioniert. In puncto Sicherheit schneidet Grafana klar besser ab, denn auf die mit Kibana generierten Dashboards hat erstmal jeder Zugriff [47].

¹⁸mit Kibana

3 Analyse

Die Analyse ist in drei Abschnitte unterteilt. Im Abschnitt 3.1 werden die Anforderungen analysiert. Darauf folgt in Abschnitt 3.2 eine Ansicht von existierenden Vergleichen und anschließend bezugnehmend auf diese in Abschnitt 3.3 die Betrachtung existierender Software.

3.1 Anforderungsanalyse

Als Grundlage für die Konzeption und Entwicklung wird als Beispielszenario der bisher ausschließlich auf Instagram vertretene Blog „nagelgepinsel“ gewählt. Dieser steht somit stellvertretend für kleine Unternehmen, die eine neue Internetpräsenz auf der Microblogging-Plattform Twitter aufbauen und beobachten möchten. Die im Folgenden genannten funktionalen und nicht-funktionalen Anforderungen müssen mess- und überprüfbar sein.

3.1.1 Funktionale Anforderungen

Die funktionalen Anforderungen an das zu entwerfende Tool geben den unmittelbaren Nutzen der Software wieder, also *was* die Software tun soll. Zu diesem Zweck werden User Stories und deren Akzeptanzkriterien definiert.

User Stories Sie repräsentieren eine Anforderung des Kunden an die zu implementierende Software. Außerdem sind sie Kommunikationsgrundlage zwischen Product Owner und Entwicklern [46, S. 49f.].

FA-01 Als Anwender kann ich Suchanfragen (wie viele Tweets gibt es zu dem Namen des Unternehmens, wie oft hat Benutzer XY das Unternehmen retweeted/geliked/erwähnt) starten, um Analysen auf den zurückgegebenen Daten zu machen.

AK-FA-01 Wenn der Anwender eine Suchanfrage startet, soll eine Anzahl an passenden Ergebnissen zurückgegeben werden.

FA-02 Als Anwender kann ich Diagramme erhalten (zum Beispiel Balken- oder Kreisdiagramme), um mir einen Überblick über die Erwähnungen des Unternehmens zu verschaffen.

AK-FA-02 Wenn der Anwender den Button „Visualisierung erstellen“ anklickt, soll ein zu den ausgewählten Daten passendes Diagramm angezeigt werden.

FA-03 Als Anwender kann ich die Menge der Interaktionen sehen (Retweets, Likes), um aufbauend auf diesen Entscheidungen für zukünftige Beiträge zu treffen.

AK-FA-03 Wenn der Anwender nach Interaktionen zu Beiträgen sucht, sollen diese vollständig zurückgegeben werden.

FA-04 Als Anwender kann ich mir Sentiments anzeigen lassen, um das Stimmungsbild des Unternehmens einzuschätzen.

AK-FA-04 Wenn der Anwender Sentiments der gesuchten Daten sehen möchte, sollen positive, negative und neutrale Sentiments angezeigt werden.

FA-05 Als Anwender kann ich einen bestimmten Begriff beobachten, um über Erwähnungen des Unternehmens informiert zu werden.

AK-FA-05 Wenn der Anwender einen Begriff beobachtet, soll bei jedem Vorkommen dieses Begriffs eine E-Mail an den Anwender gesendet werden.

3.1.2 Nicht-funktionale Anforderungen

Die nicht-funktionalen Anforderungen an das zu entwerfende Tool geben wieder, *wie* das System benutzt wird und *wie* deren Funktionen erbracht werden sollen. Die nicht-funktionalen Anforderungen orientieren sich an ISO 9126 und der Kategorie nicht-funktionale Anforderungen von Robertson [31].

Oberflächenanforderungen Unter diesem Punkt wird das Aussehen der Software beschrieben, zum Beispiel, dass es attraktiv für junge Leute sein soll. Für diesen Punkt werden in dieser Arbeit keine Anforderungen definiert.

Benutzbarkeit Unter dieser Anforderung werden die Kriterien für die Benutzung verstanden. Inbegriffen sind hier zum Beispiel: *Verständlichkeit*, *Erlernbarkeit* und *Bedienbarkeit*

NFA-01 Das Tool soll ohne Schulungen von einem Informatiker benutzbar sein.

NFA-02 Ein neuer Nutzer muss innerhalb von einer Stunde Suchanfragen formulieren und das Tool verstehen können.

Performanz Unter diesen Punkt fallen zum Beispiel die *Antwortzeit* und *Verfügbarkeit* des Systems.

NFA-03 Das Tool muss Suchanfragen innerhalb von 60 Sekunden bearbeiten.

Operationale Anforderungen Diese Anforderungen beschreiben die Umwelt in der das Tool funktionieren muss.

NFA-04 Die Umgebung auf der das Tool läuft muss einen Zugang zum Internet haben, um Tweets zu laden.

Wartbarkeit Unter dem Punkt Wartbarkeit werden die Forderungen an die *Erweiterbarkeit*, *Instandhaltung* und *Fehlerbehebung* verstanden.

NFA-05 Das Tool muss auf verschiedenen Betriebssystemen (macOS Catalina, Windows 10) ausführbar sein.

NFA-06 Das Tool soll die Behebung von Fehlern mit vertretbarem Aufwand (zwei Arbeitstage) ermöglichen.

NFA-07 Das Tool soll Erweiterungen und Änderungen mit einem Aufwand von höchstens fünf Arbeitstagen ermöglichen.

Zugriffsschutzanforderungen Der Zugriff auf die Daten unterliegt verschiedenen Anforderungen für unterschiedliche Nutzer, unter diesem Punkt wird auch die Integrität der Daten sowie die Privatsphäre beleuchtet. Wie auch für die Kategorie *Oberflächenanforderungen* werden hier keine Anforderungen definiert.

Kulturelle, politische und rechtliche Anforderungen Auch diese Punkte werden in der Arbeit nicht explizit betrachtet.

3.2 Existierende Vergleiche

Die existierenden Tools wurden schon in vielen wissenschaftlichen Arbeiten von verschiedenen Seiten beleuchtet. Das Fraunhofer Institut hat verschiedene Paper („An approach for evaluation of social media monitoring tools“ [39] und „Marktstudie Social Media Monitoring Tools IT-Lösungen zur Beobachtung und Analyse unternehmensstrategisch relevanter Informationen im Internet“ [23]) zu dem Thema veröffentlicht, welche auch als Quellen herangezogen wurden. Im nächsten Abschnitt wird sich auf diese bezogen und sie werden näher erläutert. Das Semantic Technology Institute Innsbruck hat eine Präsentation veröffentlicht, in der es um das Thema Social Media Monitoring geht. Außerdem werden einige der auf dem Markt verfügbaren Tools verglichen [15]. Im Paper „Weak signals detection: Criteria for social media monitoring tools“, welches im Rahmen der

„5th ISPIM Innovation Symposium“ vorgestellt wurde, werden SMM-Tools im Kontext von sogenannten „schwachen Signalen“¹ verglichen [45].

3.3 Existierende Software

Durch die steigende Wichtigkeit von Social Media für Unternehmen gibt es eine große Auswahl an Social Media Monitoring Tools. Um sich über diese einen Überblick zu verschaffen, wird eine Auswahl an Tools nun in einer Tabelle² dargestellt und eingeordnet [39, S. 62][23, S. 33ff.][43]. Die betrachteten Kriterien sind:

- Datenquellen und -arten: Ist es möglich, verschiedene *Webseiten*, *Social Media Plattformen* und *Blogs* als Quellen zu verwenden und können verschiedene Datenarten wie *Fotos*, *Videos* oder *Text* verarbeitet werden?
- Funktionen: Dazu gehört die Suche in den Daten und das Auswerten von Sentiments.
- Historische Daten: Diese dienen zum Vergleich mit aktuellen Auswertungen sowie für Langzeitbeobachtungen (zum Beispiel von Kampagnen).
- Kosten: Hier werden die monatlichen Kosten betrachtet, die anfallen, um das Tool zu verwenden.
- Dashboard: Stellt eine Darstellung der Daten bereit mit der Möglichkeit, Daten aus verschiedenen Quellen zusammen in Grafiken abzubilden. Zusätzlich sollte es möglich sein das Dashboard sowie die Grafiken benutzerspezifisch anzupassen.

Software	Datenquellen	Funktionen	Historische Daten	Kosten ³	Dashboard

¹Schwache Signale: Informationen über beginnende Trends, die eventuell Auswirkungen auf das Unternehmen haben.

²✓ = Kriterium erfüllt, ✗ = Kriterium nicht erfüllt, Zelle leer gelassen = keine ausreichenden Informationen

³€ = unteres Preissegment (bis 1.000€), €€ = mittleres Preissegment (bis 2.000€), €€€ = oberes Preissegment (≥ 5.000€)

Brandwatch	✓	✓	✓	€	✓
Tweetdeck	✗	✓		kostenlos	✗
Alterian - SM2	✓	✓	✓	€	✓
Converseon	✓	✓		€	✓
evolve24 Mirror	✓	✓		€€€	✓
Meltwater Buzz	✓	✓	✓	€	✓
NM Incite My BuzzMetrics	✓	✓	✓	€	✓
Radian6	✓	✓	✓	€	✓
Sysomos	✓	✓	✓	€	✓
Visible Technologies Intelligence	✓	✓	✓	€€	✓

Tabelle 3.1: Einordnung existierender Software

Nach dieser allgemeinen Vorstellung werden im Folgenden zwei Tools genauer betrachtet und in *intern* und *extern* kategorisiert. Das erste *Brandwatch Analytics* ist ein externes SMM-Tool. Externe SMM-Tools sind dadurch gekennzeichnet, dass eine externe Firma beauftragt wird, sich dem Monitoring anzunehmen. Folglich liegt hier die Expertise nicht im Unternehmen und ist somit für Unternehmen zu empfehlen, die keine eigenen Kapazitäten für diesen Bereich haben. Das zweite *TweetDeck* ist ein internes SMM-Tool und kann intern von allen Mitarbeitern angewendet werden. Interne SMM-Tools sind häufig dadurch gekennzeichnet, dass nur ein Social Media Kanal beobachtet werden kann. Hinzu kommt, dass die Expertise bei internen SMM-Tools in der Firma liegt und somit für Unternehmen zu empfehlen ist, die langfristig einen bestimmten Kanal beobachten wollen [8]. Die beispielhaft betrachteten Tools werden erst kurz beschrieben, danach folgt ein Vergleich anhand folgender Kriterien: *Kosten*, *(Daten-)Quellen* und *Funktionen*.

Brandwatch Analytics Brandwatch ist ein im Jahr 2010 gegründetes Unternehmen mit über 500 Mitarbeitern⁴ auf vier verschiedenen Kontinenten [4]. Die monatlichen **Kosten** betragen zirka zwischen 1.000€ und 2.000€. Als **Quellen** zur Datenbeschaffung können die verschiedensten sozialen Netzwerke benutzt werden, aber auch Pressedienste und viele weitere Webseiten kommen in Frage. Die Daten können auch verschiedenartig sein, das heißt es können unter anderem Formate wie Foto, Video oder Text analysiert

⁴stand 2018

werden. Des Weiteren werden zahlreiche verschiedene **Funktionen**, wie das Durchsuchen, Filtern und Beobachten der Daten angeboten. Zusätzlich steht ein Dashboard zur Verfügung mit dem die Daten visualisiert werden können. Dieses Dashboard kann der Benutzer individuell anpassen. Bei besonderen Ereignissen können Alerts per E-Mail an die Nutzer geschickt werden [23, S. 60ff.].

TweetDeck TweetDeck ist eine im Jahr 2008 erstmals veröffentlichte Desktop-Applikation, die im Jahr 2011 von Twitter gekauft wurde. Um TweetDeck nutzen zu können, wird lediglich ein (Twitter-)Profil benötigt und es fallen keine **Kosten** an. Dies geht mit einer geringeren Varietät an **Quellen** einher, es kann nur Twitter als Datenquelle benutzt werden. Hier können aber dann die verschiedenen Datenformen, die Twitter anbietet, analysiert werden. Die wichtigsten **Funktionen**, wie zum Beispiel Tweets (und Hash-tags) erstellen, planen, durchsuchen, in Sammlungen zusammenfassen und beobachten, werden auch von TweetDeck abgedeckt. Allerdings stößt es bei großen Datenmengen an seine Grenzen und ist somit eher für Anfänger, kleine und mittelständische Unternehmen, die eine kostenlose Software bevorzugen, geeignet [14].

4 Konzeption

Auf Grundlage der in Kapitel 3 beschriebenen Analyse wird in diesem Kapitel ein Konzept für ein Social Media Monitoring Tool entwickelt.

4.1 Unterscheidung fachliche und technische Konzeption

Im Folgenden wird zwischen der fachlichen und der technischen Konzeption unterschieden. Diese orientiert sich an Siederslebens Unterscheidung zwischen „Anwendungsarchitektur“ und „Technikarchitektur“ [38, S. 146].

Anwendungsarchitektur (A-Architektur) Diese beschreibt die Software aus Sicht der Anwendung, es sind keine technischen Details des Systems bekannt. Die A-Architektur wird in der fachlichen Konzeption erstellt.

Technikarchitektur (T-Architektur) Zu dieser zählen die einzelnen Komponenten der Software, die das System technisch realisieren. Die T-Architektur wird in der technischen Konzeption erstellt.

4.2 Fachliche Konzeption

In diesem Abschnitt wird die Außensicht des Tools beschrieben, so wie es von potentiellen Kunden wahrgenommen wird. Die folgende Beschreibung orientiert sich an den in Kapitel 3 formulierten Anforderungen.

4.2.1 Beschreibung des Tools

Die steigende Beliebtheit von sozialen Medien und der damit verbundene Austausch von Informationen und Meinungen führt dazu, dass Unternehmen nicht mehr alle Erwähnungen, Beschwerden und Lob mitbekommen. Mit Hilfe eines Social Media Monitoring Tools ist es Unternehmen möglich den Überblick zu behalten und schnellstmöglich auf zum Beispiel negative Erwähnungen zu reagieren. Die Wichtigkeit einer schnellen Reaktion kann beispielhaft durch diesen Tweet¹ von Kylie Jenner belegt werden, in dem sie sich über Snapchat äußert und der Aktienwert des Unternehmens danach um 8% sank.



Abbildung 4.1: Twitter: Kylie Jenner Tweet Snapchat

Das zu entwickelnde Tool soll es möglich machen, solche negativen Auswirkungen zu verhindern, indem das Unternehmen über Erwähnungen (FA-05) frühzeitig informiert wird und auf diese bei Bedarf reagieren kann. Dies ist besonders für kleine und mittelständische Unternehmen von Bedeutung, da durch eine geringe(re) Anzahl an Interaktionen negative Postings potentiell besonders viel Aufmerksamkeit erlangen könnten, weil sie nicht in der Masse untergehen. Durch die Echtzeit-Auswertung könnten beispielsweise potentielle Shitstorms² abgewendet werden. Außerdem verhilft das Sammeln der Daten in Echtzeit zu einer repräsentativen Menge an Daten. Ferner soll es, wie in FA-01 beschrieben, eine Suchfunktion geben, die gesammelt alle Daten (Tweets) zu dem Suchbegriff zurückgibt. Diese Daten sollen dann, mit Hilfe von Kibana, in verschiedenen Diagrammen abgebildet werden können (FA-02). Des Weiteren soll es möglich sein, detailliertere Informationen aus den Daten zu extrahieren, wie zum Beispiel Sentiments (FA-04) oder die Anzahl der Interaktionen (FA-03).

¹siehe: Abbildung 4.1

²ein Sturm von negativer Kritik im Internet

Fachliches Datenmodell Das folgende fachliche Datenmodell³ zeigt unabhängig von der Implementierung die Bestandteile des Tools. Ein Benutzer kann beliebig viele Feeds durchsuchen, diese Feeds können verschiedene Daten beinhalten, beispielhaft zu nennen wären Beiträge und deren Kommentare. Aufbauend auf diesen Daten können Analysen gestartet werden. Wenn man beim Beispiel der Beiträge und deren Kommentare bleibt, könnte in einem Zeitraum betrachtet werden, wie die Sentiments sich entwickeln oder zu welchem Anteil diese positiv, negativ oder neutral sind. Diese Analyse kann anschließend in einem oder mehrern Diagramm(en) abgebildet werden, der Zeitverlauf könnte in einem Balkendiagramm abgebildet werden und die Anteile von positiven, negativen und neutralen Kommentaren kann in einem Kreisdiagramm abgebildet werden. Es ist Benutzern möglich beliebig viele dieser Diagramme zu erhalten und sie zu einem (oder mehreren) Dashboard(s) zusammenfügen.

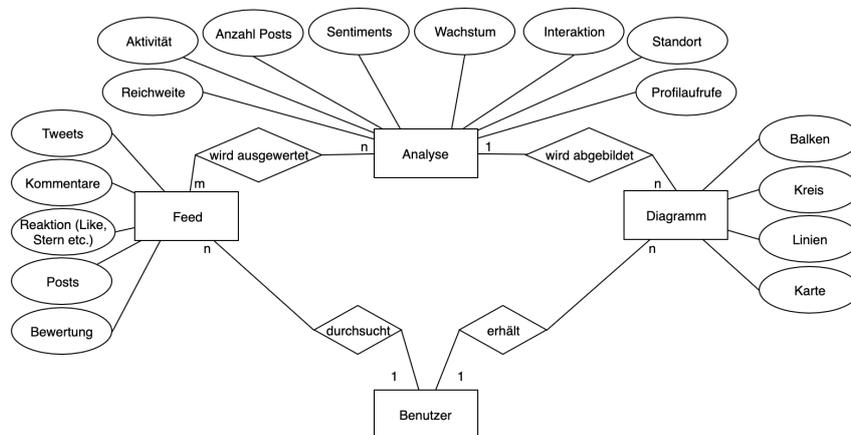


Abbildung 4.2: Fachliches Datenmodell

Graphische Benutzungsoberfläche Um sich das Tool besser visuell vorstellen zu können, wird im Folgenden ein Wireframe skizziert⁴. Zunächst soll der Startbildschirm nur aus einer Suche im oberen Drittel des Fensters bestehen. Sobald der Button „Suche“ gedrückt wird, sollen die Suchergebnisse vorerst in einer Liste (in textueller Form) dargestellt werden. Durch einen Klick auf den Button „als Diagramm darstellen“, werden die Daten als Diagramm dargestellt, die Liste der gesuchten Daten bleibt trotzdem sichtbar. Unter dem Diagramm (default: Balkendiagramm) gibt es die Möglichkeit zwischen verschiedenen Diagrammtypen zu wechseln. Wenn ein Tweet ausgewählt wird, können

³siehe: Abbildung 4.2⁴siehe: Abbildung 4.3

Sentiments angezeigt werden. Außerdem soll es möglich sein, einen bestimmten Begriff zu beobachten.

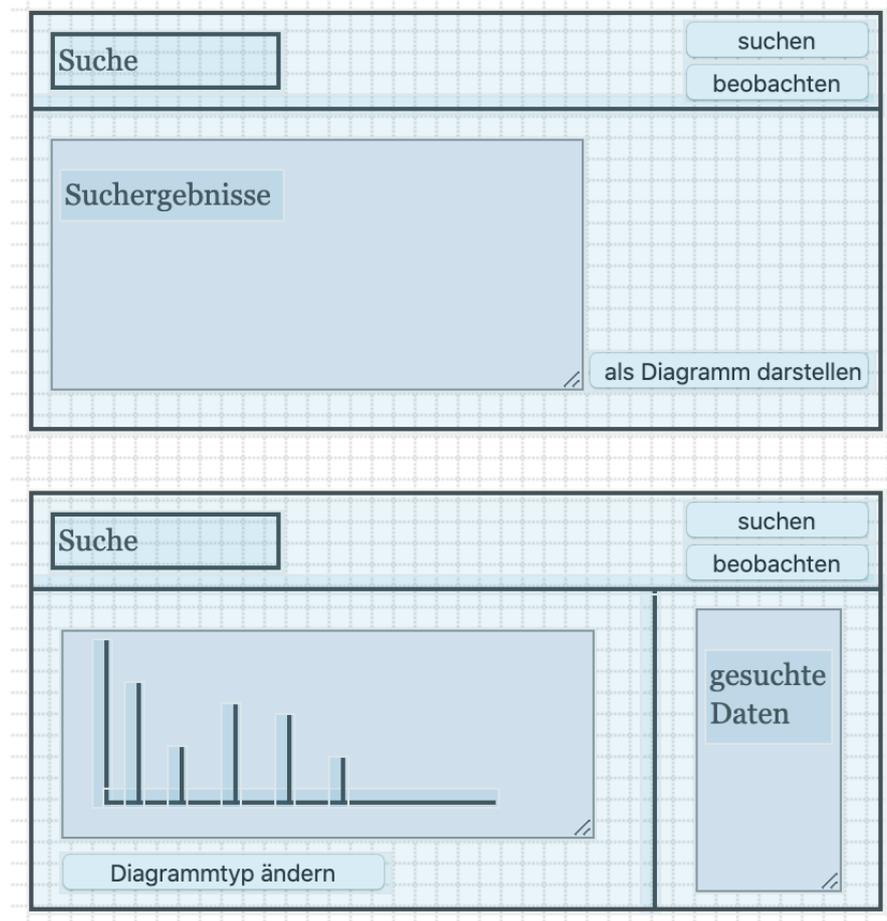


Abbildung 4.3: Wireframe: Entwurf der GUI

4.3 Technische Konzeption

Die in der fachlichen Konzeption und in den funktionalen Anforderungen (3.1.1) beschriebenen Aspekte sollen nun mit einer technischen Konzeption untermauert werden. Software kann in ihre einzelnen Komponenten unterteilt werden. Für diese Arbeit orientiere ich mich an dem Komponentenansatz von Siedersleben [38].

4.3.1 Komponenten

Komponenten sind ein wesentlicher Bestandteil der Softwareplanung. Sie stellen eine (oder mehrere) Schnittstelle(n) bereit, sind austauschbar, verstecken ihre Implementierung und sind wiederverwendbar [38, S. 42f.].

Datenbeschaffung Hierfür kann jedes soziale Netzwerk verwendet werden. In dieser Arbeit wird zu Testzwecken Twitter verwendet und es besteht die Möglichkeit dies beliebig zu erweitern. Die Postings dienen als zentrale Quelle zur Auswertung und Darstellung der Daten. Um die Daten von den sozialen Netzwerken zu beschaffen wird in der Realisierung die Programmiersprache Python benutzt.

Datenspeicherung Die durch Twitter beschafften Daten sollen in einer Datenbank gespeichert werden. Wie in den Grundlagen beschrieben, haben die von Twitter beschafften Daten das Format JSON. Für diese Arbeit entscheide ich mich für Elasticsearch. Für die Datenspeicherung stehen auch noch andere Datenbanken, wie zum Beispiel die NoSQL Datenbank mongodb zur Verfügung. Mit mongodb ist es, wie auch mit Elasticsearch, möglich, Daten im JSON Format sowie große Datenmengen zu speichern. Die Entscheidung fiel auf Elasticsearch, da es bei Elasticsearch möglich ist, weiterführende Suchen in den Daten durchzuführen.

Datensuche Um die beschafften Daten zu durchsuchen und dann in der Datenabbildung darzustellen soll es eine Datensuchkomponente geben. Es soll unter anderem möglich sein Volltext-Suchen und Indexsuchen zu starten. Die zurückgelieferten Ergebnisse können benutzt werden, um Entscheidungen zu treffen. Wie im Punkt Datenspeicherung erläutert, wird Elasticsearch für die Suche in den Daten verwendet.

Datenabbildung Um die Daten in Beziehung zueinander zu bringen und besser interpretieren zu können, soll es möglich sein sie in verschiedenen Diagrammen abzubilden. Die Diagramme sollen auch eine *Sentiment-Analyse* ermöglichen. Die Sentiment-Analyse ist ein Bestandteil des Natural Language Processing (NLP) und hilft bei der Einordnung von Texten. Die Sentiments spiegeln die Meinung/Stimmung des Verfassers eines Posts wider und können entweder in zwei Kategorien (positiv, negativ) oder in drei (positiv, negativ oder neutral) eingeteilt werden [32, S. 16f.]. Diese Analyse soll genutzt werden, um

mit Hilfe von Posts ein Meinungsbild über ein Unternehmen zu bestimmen. Um die Daten abzubilden gibt es verschiedene Tools auf dem Markt. Grafana zum Beispiel ermöglicht es, Daten aus verschiedensten Quellen zu visualisieren. Da ich mich bei der Datenspeicherung und Suche für Elasticsearch entschieden habe, macht es Sinn das dazugehörige Tool Kibana für die Visualisierung zu verwenden.

Komponentendiagramm Um das Zusammenspiel der einzelnen Komponenten darzustellen, wurde folgendes Diagramm⁵ erstellt. Das Diagramm zeigt die oben beschriebenen Komponenten: die Datensuche, Datenspeicherung, Datenabbildung und Datenbeschaffung sind ein geschlossenes System. Durch einen Adapter ist die Datenbeschaffung mit sozialen Netzwerken, wie zum Beispiel Facebook, Twitter und Youtube, verbunden.

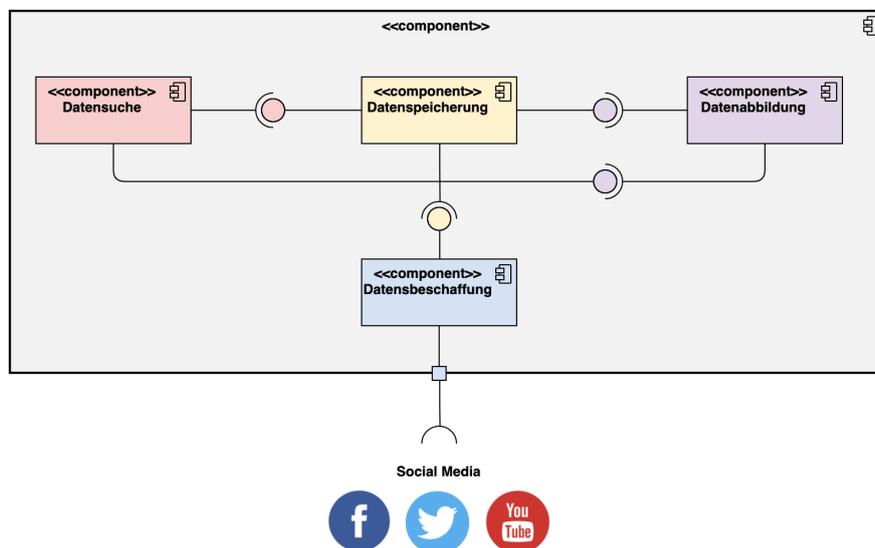


Abbildung 4.4: Komponentendiagramm

⁵siehe: Abbildung 4.4

4.3.2 Big Data Lebenszyklus

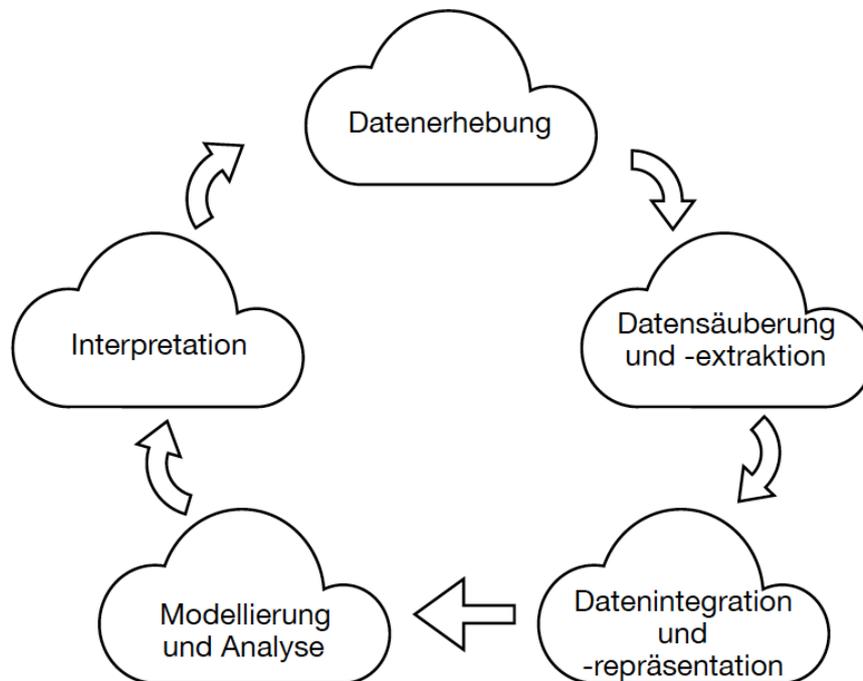


Abbildung 4.5: Big Data Lebenszyklus

Aufbauend auf den Komponenten lässt sich der Big Data Lebenszyklus⁶ als Prozess für den Nutzen und die Verwendung des Tools beschreiben. Im ersten Schritt *Datenerhebung* werden die Daten aus den verschiedenen Social Media Kanälen beschafft. Die erhobenen Daten bieten verschiedene Chancen, gerade für kleine und mittelständische Unternehmen. Sie können ihre Produkte und Services durch die in einem späteren folgenden Schritt *Analyse* besser und effizienter an die Bedürfnisse und Wünsche der Kunden anpassen. Da die Datenmengen enorm groß sein können, ist es sinnvoll diese zu filtern. Beim Filtern dürfen keine wichtigen Informationen verloren gehen, darum ist es entscheidend, dass geeignete Filter definiert werden, auch um eventuelle Muster in den Daten erkennen zu können. Dadurch, dass diese Daten aus den verschiedensten Quellen kommen, müssen sie im zweiten Schritt *gesäubert* und *extrahiert* werden, damit sie alle das gleiche Format haben und ihnen Informationen entnommen werden können. Hierfür ist ein Social Media Monitoring Tool für kleine und mittelständische Unternehmen besonders wertvoll, da diese häufig nicht die Kapazitäten haben sich umfassend mit dem

⁶siehe: Abbildung 4.5

Säubern und Extrahieren der Daten zu beschäftigen. Im dritten Schritt *Datenintegration und -repräsentation* wird ein finales, homogenes Datenset erstellt, indem die Daten aus den verschiedenen Quellen zusammengeführt werden. Wenn die Daten alle den gleichen Aufbau haben, ist es einfacher sie zu analysieren. Im vierten Schritt der *Modellierung und Analyse* werden die Daten durchsucht und in Zusammenhang zueinander gesetzt. Dafür können die Daten beispielsweise in Diagrammen abgebildet oder Sentiments erstellt werden. Laut einer Studie des Fraunhofer-Instituts aus dem Jahre 2012 sind kleine und mittelständische Unternehmen, was die Nutzung von Big Data angeht, noch nicht optimal ausgebaut, es fehlen interne Ressourcen und die personelle Verantwortung ist unklar [37, S. 8ff.]. Dafür bietet sich ein Social Media Monitoring Tool an, mit dem im Idealfall viele verschiedene Mitarbeiter umgehen können. Im fünften Schritt erfolgt die *Interpretation* der Daten. Es werden Entscheidungen getroffen, ob zum Beispiel Kampagnen gestartet werden sollen. Das Fraunhofer-Institut stellt außerdem fest, dass Big Data die Effizienz der Unternehmensführung steigert. Beispielsweise können im Einzelhandel genauere Prognosen getroffen werden, welche Produkte besonders häufig verkauft werden. Ein Social Media Monitoring Tool hilft hier bei der Auswertung der Daten. Durch die Interpretation kann die Sprache der Kunden angenommen werden und somit zum Beispiel eine engere Bindung an Kunden erfolgen [37, S. 10f.]. Die Interpretation erfolgt immer von einem „Experten“, da die Computersysteme Fehler machen oder die Daten fehlerhaft sein könnten. Weil die Möglichkeit besteht, dass die Ergebnisse des Computers wieder neue Fragen aufwerfen, beginnt der Zyklus häufig von vorne [20].

5 Realisierung und Test

In diesem Kapitel wird das in Kapitel 4 vorgestellte Konzept für ein Social Media Monitoring Tool realisiert und im Anschluss getestet. Zu diesem Zweck werden im ersten Abschnitt dieses Kapitels die verwendeten Technologien und Tools vorgestellt, darauf folgt eine Beschreibung der Implementierung. Ferner werden die erhobenen Daten untersucht und am Ende des Kapitels einige Tests der Software gemacht.

5.1 Verwendete Technologien und Tools

In diesem Abschnitt werden die für die Realisierung verwendeten Technologien und Tools vorgestellt.

5.1.1 Twitter

Um auf die Daten von Twitter zugreifen zu können muss der Twitter Account zunächst zu einem sogenannten Entwickler-Account erweitert werden. Nachdem man eine App erstellt, werden diesem Account dann Keys und Tokens zugewiesen, mit denen es möglich ist auf Tweets zuzugreifen.

5.1.2 Python

In dieser Arbeit wird Python (Version 3.7.2) als Programmiersprache verwendet, um auf Twitter zuzugreifen und die von Twitter beschafften Daten in Elasticsearch zu speichern. Die Vorteile von Python sind, dass die Sprache vergleichsweise schnell erlernt werden kann und viele nützliche Programmbibliotheken existieren, wie zum Beispiel folgende:

Tweepy (Version 3.8.0) Tweepy ist eine Python-Bibliothek, um auf die Twitter API zuzugreifen. Neben Tweepy gibt es noch weitere Bibliotheken wie zum Beispiel *Twython* und *TwitterAPI*. Für Tweepy spricht, dass es am weitesten verbreitet zu sein scheint und die Dokumentation strukturiert und einfach zu verstehen ist.

TextBlob (Version 0.15.3) TextBlob ist eine Python-Bibliothek die Natural Language Processing (NLP) ermöglicht. TextBlob basiert auf der Bibliothek Natural Language Tool KIT (NLTK), welche eine der größten NLP-Bibliotheken ist[28]. Da TextBlob geeignet ist, um Sentiments zu analysieren und relativ leicht zu verstehen ist, habe ich mich in dieser Arbeit dafür entschieden. Für die Sentimentanalyse gibt es bei TextBlob eine Wortsammlung, in der Wörter mit ihrer positiven, negativen oder neutralen Belegung aufgelistet sind. Des Weiteren gibt es zwei Parameter, die Gegensätzlichkeit bzw. *polarity* und die Subjektivität bzw. *subjectivity* die zur Einordnung der Tweets dienen. Die *polarity* liegt zwischen -1 und +1, wobei alle Werte kleiner als null *negativ*, gleich null *neutral* und größer als null *positiv* sind. Die *subjectivity* liegt zwischen 0 und +1, null bedeutet, dass das Wort objektiv benutzt wurde, alles darüber gibt den Grad der Subjektivität an. Ein und das selbe Wort kann verschieden gedeutet werden, beispielsweise ist das Wort „hard“ 12 Mal in der Wortsammlung aufgeführt. Je nach zugeordnetem Sinn wird es anders bewertet, in folgender Tabelle ein Auszug aus der Wortsammlung für das Wort *hard*[26].

sense	polarity	subjectivity
resisting weight or pressure	0.0	0.0
not easy	-0.5	1.0
very strong or vigorous	0.5	0.5

Tabelle 5.1: *Polarity* und *Subjectivity* des Wortes *hard* bewertet anhand des Sinns

5.1.3 Elasticsearch

Um schnell auf die Daten zuzugreifen werden diese mit Indizes versehen. Diese bestehen intern mindestens aus einem *shard*, können aber auch in mehrere *shards* aufgeteilt werden. Für einen *shard* können beliebig viele *Replicas* der Daten angefertigt werden, mit dem Hintergrund, dass die Verfügbarkeit der Daten selbst bei einem Systemabsturz gewährleistet ist. Die *replicas* und *shards* können je nach Größe auf mehrere *nodes* (Server)

aufgeteilt werden. Diese horizontale Skalierung dient der Verringerung der Last für die einzelnen Server. Eine Gruppe von *nodes* wird als *cluster* bezeichnet[17]. In dieser Arbeit wird Elasticsearch in Version 7.5.1 verwendet.

5.1.3.1 Kibana

Nachdem die Daten mit Hilfe des Python-Codes über einen Index in Elasticsearch gespeichert wurden, kann Kibana (Version 7.5.1) mit diesem Index verbunden werden.

Use Elasticsearch data Connect to your Elasticsearch index

Abbildung 5.1: Kibana mit Daten aus Elasticsearch verbinden

Danach können mit den Daten zum Beispiel verschiedenste Diagramme und Dashboards erstellt sowie Queries zur Datensuche gestartet werden. Die folgende Query gibt die unter dem Index *bachelor* gespeicherten Daten wieder, die im Feld „sentiment“ *positive* als Argument haben.



Abbildung 5.2: Kibana Datensuche Möglichkeit 1: Query

Zurückgegeben wird unter anderem beispielsweise folgende JSON Datei.

```
{
  "_index" : "bachelor",
  "_type" : "text-type",
  "_id" : "h7nvr28B4QQWZXKEl5nn",
  "_score" : 1.0798894,
  "_source" : {
```

```

"author" : "lisalabarge169",
"date" : "Thu Jan 16 19:58:27 +0000 2020",
"message" : ""RT @jeffsessions: Thank you, President Trump,
for taking another strong step to protect religious liberty.
"polarity" : 0.21666666666666665,
"subjectivity" : 0.49166666666666664,
"sentiment" : "positive"
}
}

```

Die zweite Möglichkeit Daten in Kibana zu suchen ist durch Filter.

EDIT FILTER [Edit as Query DSL](#)

Field	Operator
sentiment.keyword	is

Value

positive

Create custom label?

Cancel Save

Abbildung 5.3: Kibana Datensuche Möglichkeit 2: Filter

5.2 Implementierung

Die Implementierung verbindet die einzelnen Komponenten miteinander. Zu Testzwecken wurde in dieser Arbeit das soziale Netzwerk Twitter benutzt. Die Implementierung umfasst verschiedene Python-Klassen. Die zentrale Klasse greift mit Hilfe von Tweepy und der Klasse *StreamListener* auf Twitter zu, die JSON Dateien werden dann in Python-Objekte umgewandelt. Danach wird der Text des Tweets einem TextBlob Objekt zugewiesen, um so das Sentiment des Tweets zu bestimmen. Nachdem das Datum so modifiziert wurde, dass es in Elasticsearch verwendbar ist, wird der Tweet mit beliebig

vielen Parametern des Tweets und unter Angabe eines Indizes in Elasticsearch gespeichert. In der Start-Methode wird unter anderem ein Stream-Objekt erstellt, welches dann durch die Methode *filter()* die gewünschten Zeichenketten findet und speichert. Die zweite zentrale Klasse, die für die Speicherung der Tweets von, in diesem Fall Donald Trump, benötigt wird, ist wie folgt aufgebaut: Es wird in der wesentlichen Methode eine Liste initialisiert, in die dann mittels einer Tweepy Methode die neusten Tweets (maximal 200 auf einmal) der Zielperson oder des Unternehmens gespeichert werden. Danach werden die Tweets, ähnlich wie in der zuvor beschriebenen Klasse unter der Angabe eines Indizes in Elasticsearch gespeichert. In der Start-Methode wird beim Aufruf der Methode als Parameter der Benutzername, der Person oder des Unternehmen übergeben, dessen Tweets gesammelt werden sollen. Abweichend vom Entwurf wurde das Beobachten eines bestimmten Begriffs (in Echtzeit) nicht implementiert.

Duplikate Wie in der Konzeption unter dem Punkt Big Data Lebenszyklus beschrieben wurde müssen die Daten gesäubert werden. In diesem Fall heißt das, die erhobenen Daten auf Duplikate zu untersuchen und diese dann gegebenenfalls zu löschen, um die Daten nicht zu verfälschen. Zu diesem Zweck wurde eine Python-Klasse erstellt, welche aus angegebenen Datenfeldern einen Hashwert berechnet. Jedem Dokument eines Elasticsearch Indizes wird ein Hashwert zugeordnet. Danach wird über die Dokumente mit dem selben Hashwert iteriert und es werden alle Duplikate gelöscht, sodass nur noch ein Dokument zu jedem Hashwert existiert[27].

5.3 Erhobene Daten

Das in der Analyse beschriebene Beispielszenario bietet keine Möglichkeit auf Twitter Daten zu sammeln, da für *nagelgepinsel* noch keine ausreichende Reichweite auf Twitter zur Verfügung steht. Darum wird die Implementierung mit der Beobachtung von Folgenden Zeichenketten, Hashtags und Benutzernamenerwähnungen getestet und die passenden Tweets gespeichert: „*Donald Trump*“, *#donaldtrump*, *#trump*, *#potus*¹ und *@realDonaldTrump*. Außerdem wurden Tweets von Donald Trump selbst gesammelt und gespeichert.

¹Dieser Hashtag wird für den Präsidenten der USA benutzt.

5.3.1 Diagramme und Auswertung

Die gesammelten Daten können in Diagrammen abgebildet, durchsucht und ausgewertet werden. Außerdem lassen sich Dashboards erstellen auf denen mehrere Diagramme zusammengefasst sind.

5.3.1.1 Sentiments

Abbildung 5.4 lässt sich entnehmen, dass die Gesamtstimmung zum Thema Donald Trump überwiegend neutral mit einer Tendenz zum positiven liegt, obwohl diese Daten gesammelt wurden, während das Impeachment-Verfahren gegen Präsident Trump lief. Für kleine und mittelständische Unternehmen könnte beispielsweise auf den Sentiments aufbauend analysiert werden, wie gut eine Kampagne ankam. Hierfür wäre es sinnvoll ab dem Start der Kampagne die Postings 24 Stunden am Tag zu verfolgen.

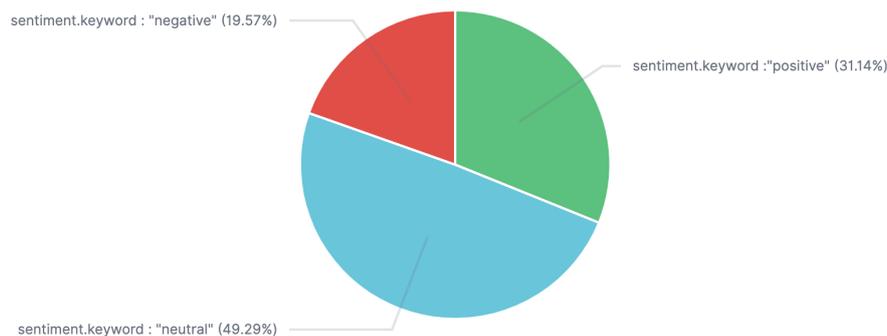


Abbildung 5.4: Sentiments Diagramm

5.3.1.2 Erwähnungen

Die Erwähnungen eines Unternehmen oder in diesem Fall von Donald Trump können auf verschiedene Arten in Verbindung gesetzt werden. In Abbildung 5.5 wird die Anzahl der Erwähnungen² an verschiedenen Tagen aufgeteilt in Sentiments gezeigt. Da aufgrund des Aufwands nicht jeden Tag Daten gesammelt wurden und an den Tagen, an denen Daten gesammelt wurden das Tool nicht immer für den gleichen Zeitraum lief, sind diese Daten nicht aussagekräftig. Würde das Tool allerdings über einen Zeitraum von 24 Stunden am

²siehe oben Links in der Abbildung, der Filter auf Tweets gesetzt die den Benutzer *@realDonaldTrump* erwähnen

Tag laufen, könnten hier Aussagen darüber getroffen werden, an welchen Tagen besonders positive, neutrale oder negative Tweets verfasst werden.

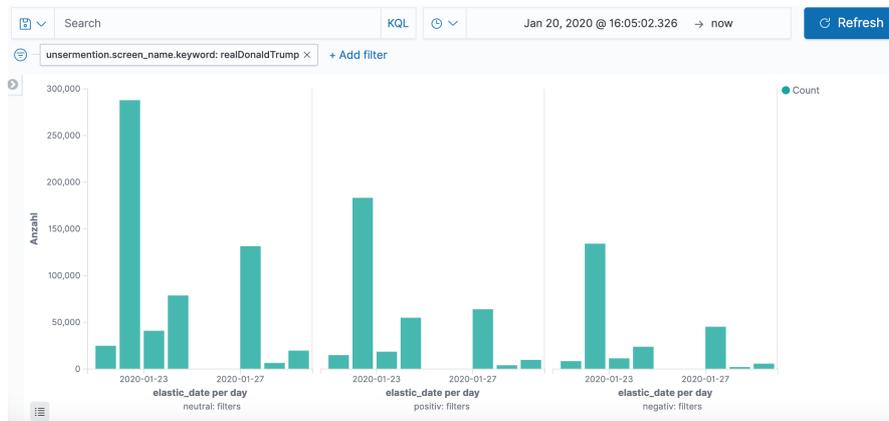


Abbildung 5.5: Erwähnungen 01: Donald Trump und Sentiments Diagramm

Die Daten können beispielsweise auch im Zusammenhang mit Benutzern, die das Unternehmen bzw. Donald Trump am häufigsten erwähnen, untersucht werden. In Abbildung 5.6 sind die 10 Benutzer aufgeführt, die *@realDonaldTrump* am häufigsten in ihren Tweets erwähnt haben. Diese Daten könnten zum Beispiel für Kampagnen benutzt werden, in denen die aktivsten Anhänger belohnt werden. So können kleine und mittelständische Unternehmen die Bindung mit den Kunden vertiefen.

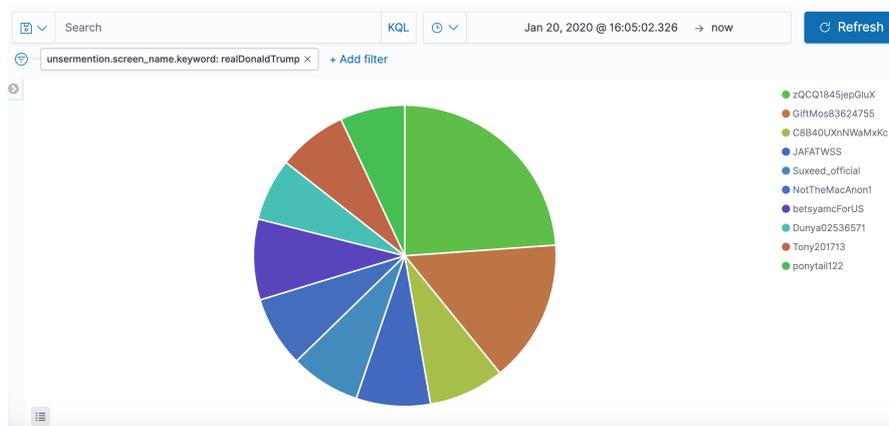


Abbildung 5.6: Erwähnungen 02: Top 10 Benutzer die Donald Trump am häufigsten erwähnen

Die Erwähnungen könnten noch mit vielen weiteren Parametern in Zusammenhang gebracht werden, die vorgestellten zeigen exemplarisch mögliche Nutzungen dieser Daten.

5.3.1.3 Länder

Manche Tweets haben auch das Land in dem sie verfasst wurden als Parameter. Auf der Karte in Abbildung 5.7 ist zu sehen in welchen Ländern³ die Tweets, die dieses Parameter haben, verfasst wurden. Solche Karten können zum Beispiel von Unternehmen mit unterschiedlichen Standorten genutzt werden, um zu analysieren, an welchen Standorten besonders viel Interaktion mit Kunden stattfindet. Darauf aufbauend kann entschieden werden, wo zum Beispiel vermehrt Werbung geschaltet werden sollte.



Abbildung 5.7: Karte Standorte Tweets

Die Länder lassen sich beispielsweise auch in folgender Form abbilden: Je größer der Name des Landes geschrieben ist, desto mehr Tweets wurden hier verfasst. Abbildung 5.8 ist mit keinem Filter versehen, es wäre allerdings denkbar hier nach Sprache zu filtern.

³Beschränkt auf 50 Länder die am häufigsten vorkamen

5.3.1.5 Tweets des Unternehmens

Durch die Anzahl der Interaktionen, die Beiträge eines Unternehmens haben, kann die Reichweite abgeschätzt werden. In Zukunft kann das Unternehmen dann mehr Beiträge dieser Art erstellen. Abbildung 5.10 zeigt die maximalen Retweets der (gesammelten) Beiträge von Donald Trump. Der erfolgreichste Tweet, gemessen an der Retweetzahl, hat über 94 Tausend Retweets. Diese Anzahl an Retweets kann kaum bei kleinen und mittelständischen Unternehmen erwartet werden, die Grafik soll nur exemplarisch zeigen, dass es möglich ist, solche Daten auszuwerten.

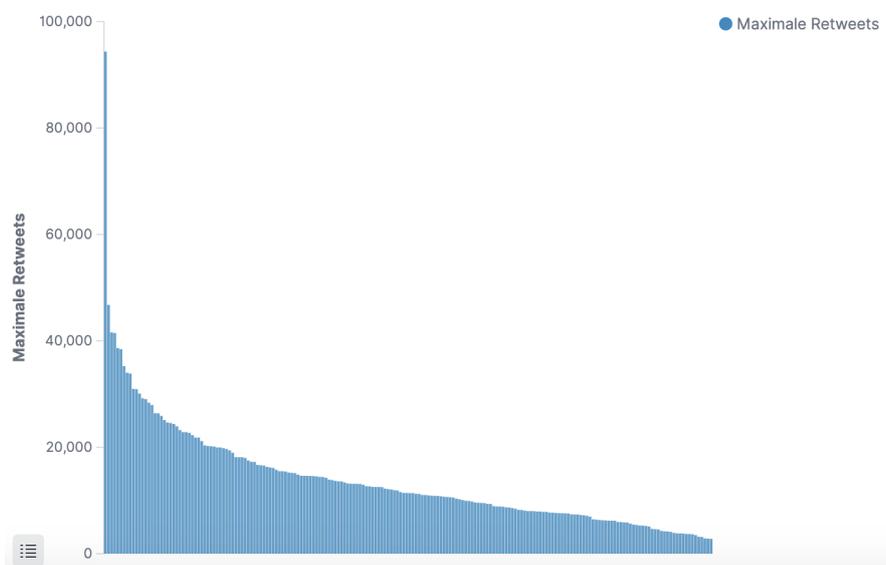


Abbildung 5.10: Anzahl der Interaktionen am Beispiel von Retweets

5.4 Tests

Durchlaufzeit der Suchanfragen Im Folgenden soll die nicht-funktionale Anforderung NFA-03 getestet werden. Hierfür werden verschiedene Suchanfragen definiert und durchlaufen, um zu schauen ob die Anforderung erfüllt ist.

Suche	Zeit in ms ⁴
GET /bachelor/_search?q=sentiment:positive	18
GET /bachelor/_search?q=language:en	38
POST /bachelor/_search <pre>{ "query": { "wildcard": { "author": "t*" } }, "_source": ["author"] }</pre>	33
POST /bachelor/_search <pre>{ "query": { "match": { "verified": "true" } }, "_source": ["verified", "author"] }</pre>	8

Tabelle 5.2: Suchanfragen um NFA-03 zu testen

Die erste Suchanfrage gibt aus dem Index *bachelor* die Objekte zurück, die in dem Feld *sentiment* den Wert *positiv* haben. Die zweite Suchanfrage sucht ebenfalls im Index *bachelor* alle Objekte deren Sprache Englisch ist. Die dritte Anfrage gibt aus dem selben Index jene Objekte wieder in denen ein *t* im Namen des Autors vorkommt. Die letzte Suchanfrage gibt die Objekte zurück deren *verified* Parameter den Wert *true* hat.

⁴Millisekunden

6 Evaluation

In diesem Kapitel soll das in Kapitel 4 konzipierte und in Kapitel 5 realisierte Tool kritisch betrachtet und evaluiert werden. Hierfür werden zunächst die Anforderungen aus Kapitel 3 mit dem implementierten Tool verglichen und danach eine kritische Betrachtung vorgenommen.

6.1 Anforderungsabgleich

In diesem Abschnitt soll überprüft werden, ob die in der Analyse an ein leichtgewichtiges Social Media Monitoring Tool geforderten Anforderungen in der Realisierung erfüllt wurden.

6.1.1 Funktionale Anforderungen

In Tabelle 6.1.1 werden die Akzeptanzkriterien zu den funktionalen Anforderungen aus Kapitel 3 (Abschnitt: Funktionale Anforderungen 3.1.1) dargestellt. In der Spalte *erfüllt* wird beleuchtet, ob die Kriterien und somit die Anforderung durch das Tool erfüllt wurde.

Akzeptanzkriterium	erfüllt
AK-FA-01: Wenn der Anwender eine Suchanfrage startet, soll eine Anzahl an passenden Ergebnissen zurückgegeben werden.	✓
AK-FA-02: Wenn der Anwender den Button „Visualisierung erstellen“ anklickt, soll ein zu den ausgewählten Daten passendes Diagramm angezeigt werden.	✓
AK-FA-03: Wenn der Anwender nach Interaktionen zu Beiträgen sucht, sollen diese vollständig zurückgegeben werden.	✓

AK-FA-04: Wenn der Anwender Sentiments der gesuchten Daten sehen möchte, sollen positive, negative und neutrale Sentiments angezeigt werden.	✓
AK-FA-05: Wenn der Anwender einen Begriff beobachtet, soll bei jedem Vorkommen dieses Begriffs eine E-Mail an den Anwender gesendet werden.	✗

Tabelle 6.1: Abgleich der Akzeptanzkriterien für die funktionalen Anforderungen mit dem entworfenen Tool

Die oben stehende Tabelle zeigt, dass alle funktionalen Anforderungen bis auf die Beobachtung eines bestimmten Begriffs in der Realisierung erfüllt wurden. Diese Anforderung wurde auf Grund des beschränkten Zeitrahmens nicht mehr umgesetzt, könnte allerdings nachträglich eingebaut werden. Die Realisierung deckt sonst alle elementaren funktionalen Anforderungen ab.

6.1.2 Nicht-funktionale Anforderungen

In Tabelle 6.1.2 werden die nicht-funktionalen Anforderungen aus Kapitel 3 (Abschnitt: Nicht-funktionale Anforderungen 3.1.2) aufgelistet. In der Spalte *erfüllt* wird überprüft, ob die Anforderung durch das Tool erfüllt wurde.

Nicht-funktionale Anforderungen	erfüllt
NFA-01: Das Tool soll ohne Schulungen von einem Informatiker benutzbar sein.	✓
NFA-02: Ein neuer Nutzer muss innerhalb von einer Stunde Suchanfragen formulieren und das Tool verstehen können.	✓
NFA-03: Das Tool muss Suchanfragen innerhalb von 60 Sekunden bearbeiten.	✓
NFA-04: Die Umgebung auf der das Tool läuft muss einen Zugang zum Internet haben, um Tweets zu laden.	✓
NFA-05: Das Tool muss auf verschiedenen Betriebssystemen (macOS Catalina (Version 10.15), Windows 10) ausführbar sein.	✓
NFA-06: Das Tool soll die Behebung von Fehlern mit vertretbarem Aufwand (zwei Arbeitstage) ermöglichen.	

NFA-07: Das Tool soll Erweiterungen und Änderungen mit einem Aufwand von höchstens fünf Arbeitstagen ermöglichen.	
---	--

Tabelle 6.2: Abgleich der nicht-funktionalen Anforderungen mit dem entworfenen Tool

Die nicht-funktionalen Anforderungen NFA-01, NFA-02 und NFA-05 wurden mit Hilfe einer (Wirtschafts-)Informatikerin, die das Tool auf einem Computer mit dem Betriebssystem Windows 10 getestet hat, überprüft und sind erfüllt. Durch die Eingabe vieler verschiedener Suchanfragen konnte auch NFA-03 erfolgreich getestet werden (siehe Tabelle 5.4) und ist erfüllt. Die Anforderung NFA-04 muss für die funktionale Anforderung FA-05 erfüllt sein. Da Anforderung FA-05 nicht umgesetzt wurde benötigt das Tool nicht zwingend Internet, denn die bereits erhobenen Daten können ohne einen Internetzugang in Verbindung gesetzt und analysiert werden. Die Anforderungen NFA-06 und NFA-07 sind in diesem Rahmen erfüllt, allerdings wurden diese nicht im realen Betrieb des Tools getestet und gelten deshalb weder als erfüllt noch als nicht erfüllt.

6.2 Kritische Betrachtung

Im Rahmen der prototypischen Realisierung wurde aufgrund der Fokussierung auf die Implementierung die Verteilung des Tools auf mehreren Computern und die Verwendung von mehreren Benutzern zur selben Zeit nicht getestet. Bei der Weiterentwicklung sollten diese Punkte vor der Auslieferung des Tool aber unbedingt getestet werden. Ebenso kritisch zu betrachten ist die Sentiment-Analyse, da es nicht Ziel dieser Arbeit war eine vollständig Text-Mining-Komponente zu erstellen. Deshalb sollte die Richtigkeit der Sentiments in Frage gestellt werden, vor allem in Bezug auf Ironie und Sarkasmus. Für das nächste Mal würde es sich deshalb anbieten, die auf dem Markt verfügbaren NLP-Tools genauer zu evaluieren und zu betrachten. Außerdem sollte für eine Einführung in ein Unternehmen die Echtzeit-Abfrage hinzugefügt werden, um benachrichtigt zu werden, wenn das Unternehmen zum Beispiel negativ erwähnt wird. Bewährt hat sich die Programmierung mit der Sprache Python aufgrund der zahlreichen Bibliotheken die zur Verfügung stehen. Auch das soziale Netzwerk Twitter hat sich bezüglich des einfachen Zugriffs auf die Daten als praktisch erwiesen. Donald Trump als beobachtete Person hat sich als nur teilweise sinnvoll herausgestellt. Das hohe Tweetaufkommen ist für kleine und mittelständische Unternehmen kaum repräsentativ, für den Zeitraum, der für die Sammlung der Daten

zur Verfügung stand, wären bei der Beobachtung eines kleinen oder mittelständischen Unternehmens aber voraussichtlich nur wenig Daten zusammengekommen und dadurch wären die Analysen vermutlich nicht aussagekräftig gewesen. Besonders wichtig für kleine und mittelständische Unternehmen wäre es daher, die Echtzeit-Verarbeitung der Daten umzusetzen, um den vollen Nutzen des Social Media Monitoring Tools ausschöpfen zu können. Überdies hat sich die Arbeit mit dem Elasticstack als unproblematisch und ziel erfüllend herausgestellt. Kibana ist einfach zu bedienen und könnte nach einer kurzen Einführung auch von Nicht-Informatikern verwendet werden. Das ist besonders für kleine und mittelständische Unternehmen von Bedeutung, da die personelle Zuständigkeit von Big Data Analysen hier häufig ungeklärt ist [37], durch die einfache Bedienung von Kibana können dann viele Mitarbeiter eingebunden werden. Diese Optimierungsmöglichkeiten könnten dabei helfen, das Tool irgendwann in einer Produktiv-Umgebung zu benutzen.

7 Zusammenfassung und Ausblick

In diesem Kapitel wird der Inhalt dieser Arbeit zusammengefasst sowie ein Ausblick auf mögliche Vertiefungen und Weiterentwicklungen geliefert.

7.1 Zusammenfassung

In dieser Arbeit wurde gemäß dem anfänglich definierten Ziels ein prototypisches Social Media Monitoring Tool für kleine und mittelständische Unternehmen konzipiert und entwickelt.

Kapitel *eins* wurde dazu genutzt das Ziel dieser Arbeit zu definieren und abzugrenzen.

Kapitel *zwei* diente der Einführung in das Thema sowie dazu, Begrifflichkeiten wie zum Beispiel „Social Media“ und „kleine und mittelständische Unternehmen“ und Grundlagen für den weiteren Verlauf der Arbeit zu klären.

In Kapitel *drei* wurden im ersten Abschnitt funktionale und nicht-funktionale Anforderungen an das Tool definiert. Im zweiten Abschnitt wurden existierende Software und Vergleiche inspiziert.

Kapitel *vier* beinhaltet aufbauend auf der Analyse aus Kapitel 3 ein in fachlich und technisch unterteiltes Konzept, in dem das Tool und seine Funktionsweise beschrieben wurden.

In Kapitel *fünf* wurde das auf der Analyse aufbauende Konzept realisiert und getestet. Zu diesem Zweck wurden zunächst die verwendeten Technologien und Tools vorgestellt und darauffolgend die Implementierung beschrieben. Anschließend wurden die durch die Implementierung erhobenen Daten betrachtet und ausgewertet. Letztlich wurden noch Tests durchgeführt.

In Kapitel *sechs* wurde das prototypische Tool unter Heranziehung der Anforderungen aus Kapitel 3 evaluiert und folgend kritisch betrachtet.

Die Evaluation hat gezeigt, dass die definierten Anforderungen größtenteils umgesetzt werden konnten. Allerdings müsste für eine Benutzung in kleinen und mittelständischen Unternehmen noch weiter an dem Tool gearbeitet werden. Die prototypische Realisierung hat zum gegebenen Zeitpunkt dennoch einen Mehrwert für Unternehmen und es sollte auf jeden Fall von kleinen und mittelständischen Unternehmen in Betracht gezogen werden, Social Media Monitoring zu betreiben. Das Potential von Big Data und deren Auswertung mit Hilfe von Social Media Monitoring Tools ist für kleine und mittelständische Unternehmen sehr groß und der Nutzen ist vielseitig.

7.2 Ausblick

Bei der Erstellung dieser Bachelorarbeit lag das Hauptaugenmerk auf der Konzeption und Entwicklung des Tools. Die Untersuchung der Daten stellte sich als äußerst interessant dar und könnte (auch durch Erweiterungen des Tools) noch tiefergehend fortgeführt werden. Vor allem die Weiterentwicklung und Untersuchung der Sentiments könnte sehr aufschlussreich sein. Des Weiteren wäre es gewinnbringend weitere Social Media Plattformen als Datenquellen heranzuziehen, um einen gesamtheitlichen Überblick über die Onlinepräsenz eines Unternehmens zu erlangen.

Abbildungsverzeichnis

2.1	Erreichen der zwei Milliarden aktiven Nutzer (Post von Mark Zuckerberg)	7
2.2	WhatsApp in Zahlen	8
2.3	Facebook: „Personen, die du kennen könntest“	9
2.4	Twitter: @tacobell teilt den Inhalt von @marharf	10
2.5	Twitter: Umfrage an Follower von @nailogical	11
2.6	Twitter: Promoted Tweet	12
2.7	Twitter: What’s happening?	12
2.8	Dolly-Parton-Challenge	13
2.9	Twitter: #HAW	14
4.1	Twitter: Kylie Jenner Tweet Snapchat	26
4.2	Fachliches Datenmodell	27
4.3	Wireframe: Entwurf der GUI	28
4.4	Komponentendiagramm	30
4.5	Big Data Lebenszyklus	31
5.1	Kibana mit Daten aus Elasticsearch verbinden	35
5.2	Kibana Datensuche Möglichkeit 1: Query	35
5.3	Kibana Datensuche Möglichkeit 2: Filter	36
5.4	Sentiments Diagramm	38
5.5	Erwähnungen 01: Donald Trump und Sentiments Diagramm	39
5.6	Erwähnungen 02: Top 10 Benutzer die Donald Trump am häufigsten er- wähnen	39
5.7	Karte Standorte Tweets	40
5.8	Namen der Länder in denen Tweets verfasst wurden.	41
5.9	Volltextsuche in den Daten	41
5.10	Anzahl der Interaktionen am Beispiel von Retweets	42

Tabellenverzeichnis

2.1	Definition kleine und mittelständische Unternehmen	3
3.1	Einordnung existierender Software	23
5.1	<i>Polarity</i> und <i>Subjectivity</i> des Wortes <i>hard</i> bewertet anhand des Sinns . . .	34
5.2	Suchanfragen um NFA-03 zu testen	43
6.1	Abgleich der Akzeptanzkriterien für die funktionalen Anforderungen mit dem entworfenen Tool	45
6.2	Abgleich der nicht-funktionalen Anforderungen mit dem entworfenen Tool	46

Literaturverzeichnis

- [1] ASSMANN, S. ; PLEIL, T.: *Handbuch Unternehmenskommunikation Strategie – Management – Wertschöpfung*. Springer Gabler, 2014. – 585 ff. S. – ISBN 978-3-8349-4543-3
- [2] BABKA, S.: *Social Media für Führungskräfte Behalten Sie das Steuer in der Hand*. Springer Gabler, 2016. – 1 S. – ISBN 978-3-658-05767-1
- [3] BRANDT, M.: *WhatsApp in Zahlen*. online. 2017. – URL <https://de.statista.com/infografik/10550/wathsapp-in-zahlen/>. – Zugriffsdatum: 12.08.2019
- [4] BRANDWATCH: *Brandwatch 2018: Unser Jahresrückblick*. online. 2018. – URL <https://www.brandwatch.com/de/blog/brandwatch-2018-unser-jahresrueckblick/>. – Zugriffsdatum: 19.10.2019
- [5] BRUHN, M. ; HADWICH, K.: *Einsatz von Social Media für das Dienstleistungsmanagement*. Springer Gabler, 2015. – 3 S. – ISBN 978-3-658-07234-6
- [6] BUCHENAU, P. ; FÜRTBAUER, D.: *Chefsache Social Media Marketing Wie erfolgreiche Unternehmen schon heute den Markt der Zukunft bestimmen*. Springer Gabler, 2015. – ISBN 978-3-658-07508-8
- [7] BVDW: *Glossar: Social Media*. online. 2016. – URL <https://www.bvdw.org/themen/publikationen/detail/artikel/glossar-social-media/>. – Zugriffsdatum: 03.09.2019
- [8] BVDW: *SOCIAL MEDIA MONITORING IN DER PRAXIS*. online. 2017. – URL https://www.bvdw.org/presseserver/Publikationen/BVDW_Leitfaden_Social_Media_Monitoring.pdf. – Zugriffsdatum: 24.10.2019
- [9] BÖKER, K-H. ; DEMUTH, U. ; THANNHEISER, A. ; WERNER, N.: *Social Media - Soziale Medien? Neue Handlungsfelder für Interessenvertretungen*. Hans-Böckler-Stiftung, 2013. – 9–10 S. – ISBN 978-3-86593-180-1

- [10] CEYP, M. ; SCUPIN, J-P.: *Erfolgreiches Social Media Marketing Konzepte, Maßnahmen und Praxisbeispiele*. Springer Gabler, 2013. – ISBN 978-3-658-00035-6
- [11] DECKER, A.: *Der Social-Media-Zyklus Schritt für Schritt zum systematischen Social-Media-Management im Unternehmen*. Springer Gabler, 2019. – ISBN 978-3-658-22872-9
- [12] ELASTIC: online. – <https://www.elastic.co/de/products/elasticsearch> and <https://www.elastic.co/de/what-is/elasticsearch> and <https://www.elastic.co/de/products/kibana> and <https://www.elastic.co/de/what-is/elk-stack>
- [13] ELLISON, N. B. ; BOYD, D. M.: Social Network Sites: Definition, History, and Scholarship. In: *Journal of Computer-Mediated Communication* (2007), S. 2. – URL <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- [14] EVERTZ, K.: *21 Tweetdeck Tipps fürs bessere Twittern*. online. 2016. – URL <https://www.monitoringmatcher.de/2016/03/21-tweetdeck-tipps/>. – Zugriffsdatum: 04.10.2019
- [15] FENSEL, D. ; LEITER, B. ; STAVRAKANTONAKIS, I.: *Social Media Monitoring*. online. 2008. – URL <https://oc.sti2.at/sites/default/files/SMM%20Handouts.pdf>. – Zugriffsdatum: 05.11.2019
- [16] FRITZ, Johannes: Wie sozial sind die Sozialen Medien wirklich? In: *Gesellschaft - Wirtschaft - Politik (GWP)* (2013), S. 144 ff.
- [17] GORMLEY, C. ; TONG, Z.: *Elasticsearch: The Definitive Guide*. O'Reilly Media, Inc., 2015. – ISBN 978-1-449-35854-9
- [18] HAEWOON, K. ; LEE, C. ; PARK, H. ; MOON, S.: What is Twitter, a Social Network or a News Media? In: *Department of Computer Science, KAIST* (2010), S. 591
- [19] HOLLAND, H. ; ROSSA, P.: BIG-DATA-MARKETING: CHANCEN UND HERAUSFORDERUNGEN FÜR UNTERNEHMEN. In: *Update 19 WS 14/15 Forschung + Wirtschaft* (2014), S. 13
- [20] JAGADISH, H.V. ; GEHRKE, J. ; LABRINIDIS, A. ; PAKONSTANTINOY, Y. ; M.-PATEL, J. ; RAMAKRISHNAN, R. ; SHAHABI, C.: Big Data and Its Technical Challenges. In: *Communications of the ACM* (2014)

- [21] JAKOBS, H.: *Verbraucherschützer beklagen Datenschutzmängel bei sozialen Medien.* online. 2018. – URL <https://www.handelsblatt.com/politik/deutschland/dsgvo-verbraucherschuetzer-beklagen-datenschutzmaengel-bei-sozialen-medien/23160794.html?ticket=ST-11611526-ENJigt0ApnPR6eQnjWOn-ap5>. – Zugriffsdatum: 30.09.2019
- [22] KAHL, M.: Big Data-Revolution, Überwachung und soziale Medien: Gefahr für die Demokratie? In: *SOZIALPOLITIK.CH* (2018), S. 4
- [23] KASPER, H. ; DAUSINGER, M. ; KETT, H. ; RENNER, T.: *Marktstudie Social Media Monitoring Tools IT-Lösungen zur Beobachtung und Analyse unternehmensstrategisch relevanter Informationen im Internet.* Fraunhofer-Institut für Arbeitswirtschaft und Organisation, 2010. – ISBN 978-3-8396-0194-5
- [24] KOMMISSION, Europäischen: *KMU-Definition der Europäischen Kommission.* online. 2005. – URL <https://www.ifm-bonn.org/definitionen/kmu-definition-der-eu-kommission/>. – Zugriffsdatum: 27.01.2020
- [25] KREUTZER, R.: *Social-Media-Marketing kompakt - Ausgestalten, Plattformen finden, messen, organisatorisch verankern.* Springer Gabler, 2018. – 61–62 S. – ISBN 978-3-658-21147-9
- [26] LORIA, S.: *textblob.* online. 2014. – URL <https://github.com/sloria/TextBlob/blob/eb08c120d364e908646731d60b4e4c6c1712ff63/textblob/en/en-sentiment.xml>. – Zugriffsdatum: 17.01.2020
- [27] MARQUARDT, A.: *deduplicate-elasticsearch.py.* online. 2018. – URL <https://github.com/alexander-marquardt/deduplicate-elasticsearch/blob/master/deduplicate-elasticsearch.py>. – Zugriffsdatum: 21.01.2019
- [28] PALAKOLLU, S.: *Top 5 Natural Language Processing Python Libraries for Data Scientist.* online. 2019. – URL <https://medium.com/@srimanikantapalakollu/top-5-natural-language-processing-python-libraries-for-data-scientist-32463d36feae>. – Zugriffsdatum: 22.01.2020
- [29] PINGKAN, L. P. I. ; WIDYAWAN, W. N.: An Evaluation of Twitter River and Logstash Performances as Elasticsearch Inputs for Social Media Analysis of Twitter. In: *2015*

- International Conference on Information, Communication Technology and System (ICTS)* (2015), S. 181
- [30] PRZYBYLSKI, A. K. ; MURAYAMA, K. ; DEHAAN, C. R. ; GLADWELL, V.: Computers in Human Behavior - Motivational, emotional, and behavioral correlates of fear of missing out. In: *elsevier* (2013)
- [31] ROBERTSON, J. ; ROBERTSON, S.: Volere Requirements Specification Template. (2007), S. 30 ff.
- [32] ROELEN-BLASBERG, T.: *Automatisierte Präferenzmessung Extraktion und Evaluation von Produktattributen auf Basis von Online-Rezensionen*. Springer Gabler, 2019. – 16 f. S. – ISBN 978-3-658-23831-5
- [33] ROLAND, G. ; RÖHRS, H-P.: *Social Media Potenziale, Trends, Chancen und Risiken*. Springer Gabler, 2017. – 12 S. – ISBN 978-3-662-53991-0
- [34] SAI DIVYA, M. ; GOYAL, S. K.: ElasticSearch An advanced and quick search technique to handle voluminous data. In: *COMPUSOFT, An international journal of advanced computer technology* (2013), S. 171 ff.
- [35] SCHMIDT, J.H. ; TADDICKEN, M. ; PSCHIDA, D.: *Handbuch Soziale Medien*. Springer VS, 2017. – ISBN 978-3-658-03765-9
- [36] SCHULZE, E.: *Wikipedia co-founder slams Mark Zuckerberg, Twitter and the „ap-palling“ internet*. online. 2019. – URL <https://www.cnbc.com/2019/07/05/wikipedia-co-founder-larry-sanger-slams-facebook-twitter-social-media.html>. – Zugriffsdatum: 30.09.2019
- [37] SCHÄFER, A. ; KNAPP, M. ; MAY, M. ; VOSS, A.: BIG DATA – Vorsprung durch Wissen. (2012), S. 8 ff.
- [38] SIEDERSLEBEN, J.: *Moderne Softwarearchitektur Umsichtig planen, robust bauen mit Quasar*. dpunkt.verlag, 2004. – ISBN 3-89864-292-5
- [39] STAVRAKANTONAKIS, I. ; GAGIU, A.-E. ; KASPER, H. ; TOMA, I. ; THALHAMMER, A.: An approach for evaluation of social media monitoring tools. In: *1st International Workshop on Common Value Management CVM2012* (2012), S. 52 ff.
- [40] TAKASE, W. ; NAKAMURA, T. ; WATASE, W. ; SASAKI, T.: A solution for secure use of Kibana and Elasticsearch in multi-user environment. In: *Cornell University* (2017), S. 2

- [41] TALAS, A. ; POP, F. ; NEAGU, G.: Elastic Stack in Action for Smart Cities: Making Sense of Big Data. In: *Computer Science Department, Faculty of Automatic Control and Computers, University „Politehnica“ of Bucharest, Bucharest, Romania* (2017), S. 471 ff.
- [42] TWITTER: *Tweet Object*. online. 2018. – URL <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object>. – Zugriffsdatum: 03.10.2019
- [43] TWITTER: *About advanced TweetDeck features*. online. 2019. – URL <https://help.twitter.com/en/using-twitter/advanced-tweetdeck-features>. – Zugriffsdatum: 12.11.2019
- [44] VENZKE-CAPRARESE, S.: Social Media Monitoring Analyse und Profiling ohne klare Grenzen? In: *DuD - Datenschutz und Datensicherheit* (2013), S. 775 – 779
- [45] WELZ, K. ; BRECHT, L. ; PENGL, A. ; KAUFFELDT, J. V. ; SCHALLMO, D. R. A.: Weak signals detection: Criteria for social media monitoring tools. (2012)
- [46] WIRDEMANN, R. ; MAINUSCH, J.: *Scrum mit User Stories*. Hanser, 2017. – 50 S. – ISBN 978-3-446-45077-6
- [47] YIGAL, A.: *Grafana vs. Kibana: The Key Differences to Know*. online. 2018. – URL <https://logz.io/blog/grafana-vs-kibana/>. – Zugriffsdatum: 30.09.2019
- [48] ZAHN, A-M.: *Status Quo: Wo steht die Social Media Forschung heute? Überblick zu Instrumenten, Anforderungen und Anwendungspotentiale der Social Media Forschung*. online. 2013. – URL <http://docplayer.org/2221376-Status-quo-wo-steht-die-social-media-forschung-heute-ueberblick-zu-instrumenten-anforderungen-und-anwendungspotentiale-der-social-media-forschung.html>. – Zugriffsdatum: 05.09.2019

A Inhalt der CD-ROM

Dieser Arbeit liegt eine CD-ROM mit folgendem Inhalt bei:

- *thesis.pdf*: Diese Arbeit im PDF Format.
- *Literatur.zip*: Beinhaltet die Literatur die für diese Arbeit herangezogen wurde.
- *code.zip*: Der Quellcode für das prototypische Tool.
- *ReadMe.rtf*: Information für die Installation und Benutzung.

Erklärung zur selbstständigen Bearbeitung einer Abschlussarbeit

Gemäß der Allgemeinen Prüfungs- und Studienordnung ist zusammen mit der Abschlussarbeit eine schriftliche Erklärung abzugeben, in der der Studierende bestätigt, dass die Abschlussarbeit „– bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit [(§ 18 Abs. 1 APSO-TI-BM bzw. § 21 Abs. 1 APSO-INGI)] – ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt wurden. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich zu machen.“

Quelle: § 16 Abs. 5 APSO-TI-BM bzw. § 15 Abs. 6 APSO-INGI

Erklärung zur selbstständigen Bearbeitung der Arbeit

Hiermit versichere ich,

Name: _____

Vorname: _____

dass ich die vorliegende Bachelorarbeit – bzw. bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit – mit dem Thema:

Social Media Monitoring: Konzeption und Entwicklung eines leichtgewichtigen Tools für kleine und mittelständische Unternehmen

ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

Ort Datum Unterschrift im Original