

Institut für Visualisierung und Interaktive Systeme

Universität Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

Bachelorarbeit

Online-Studie zur Verständlichkeit von
integrierten wortgroßen
Visualisierungen

Miriam Awad-Mohammed

Studiengang:	Informatik
Prüfer/in:	Prof. Dr. Thomas Ertl
Betreuer/in:	Tanja Blascheck, Petra Isenberg, Franziska Huth, Johannes Knittel
Beginn am:	9. März 2020
Beendet am:	25. November 2020

Kurzfassung

Auf der Social-Media-Plattform Twitter werden täglich etwa 500 Millionen Tweets gepostet. Häufig werden dabei Bilder, Videos oder Zeitungsartikel verlinkt. Gerade wenn es um kontroverse und emotionale Themen geht, sind die Meinungen, die in den Tweets geäußert werden, sehr unübersichtlich. In solchen Fällen fällt es den Nutzern schwer, die Themenvielfalt richtig einzuschätzen und zu interpretieren. Um diese Problematik zu lösen, soll im Rahmen dieser Arbeit eine Visualisierung entwickelt werden, die Aufschluss über die verschiedenen Themen geben soll.

Da es sich bei Tweets um kleine Online-Texte handelt, bietet es sich an, wortgroße Visualisierungen als Darstellungsform zu nutzen. Für die Entwicklung der Visualisierung wird auf bekanntes Wissen zurückgegriffen und aufgebaut. So wird neben der Erläuterung von grundlegenden Begriffen unter anderem auch auf Theorien des Text-Bild-Verständnisses eingegangen. Im Rahmen dieser Arbeit werden außerdem verschiedene Visualisierungstypen vorgestellt und ihre Vor- und Nachteile abgewogen.

Als Datengrundlage für die Visualisierung dienen Tweets, die am 02.06.2020 von der Online-Plattform extrahiert wurden. Diese wurden dann mittels des K-Means++ Clustering gruppiert und später reduziert. Der Datensatz wurde von den Betreuern zur Verfügung gestellt.

Um zu untersuchen, ob die Visualisierung das Leseverhalten der Nutzer positiv beeinflusst, wurde eine Online-Studie geplant und durchgeführt. In der Studie wurden drei unterschiedliche Versuchsbedingungen getestet: Für die Teilnehmer der ersten Bedingung wurde eine Darstellung der Tweets mit integrierten Visualisierungen gewählt. Bei der zweiten Bedingung wurden die Tweets je nach Thema unterschiedlich eingefärbt und bei der dritten wurden die Tweets mit einem textuellen Vermerk auf das Thema versehen. Hierdurch sollte untersucht werden, ob und inwiefern die Komplexität der Darstellungsform Einfluss auf das Leseverhalten der Nutzer hat. Die Ergebnisse der Studie zeigten jedoch, dass die gemessenen Unterschiede zwischen den Versuchsbedingungen nicht signifikant genug waren, um Aussagen diesbezüglich treffen zu können.

Inhaltsverzeichnis

1	Einleitung	9
2	Grundlagen	11
2.1	Wortgroße Visualisierungen	11
2.2	Text-Bild-Verständnis	12
2.3	Online-Studien	14
3	Verwandte wissenschaftliche Arbeiten	15
3.1	Design und Platzierung von wortgroßen Visualisierungen	15
3.2	Interaktionslevel der Visualisierungen	16
3.3	Einfluss von wortgroßen Visualisierungen auf das Leseverhalten	17
4	Problemstellung	19
5	Implementierung	21
5.1	Funktionalität	21
5.2	Design	21
5.3	Verwendete Programmiersprachen und Bibliotheken	24
5.4	Datensatz	24
6	Studie	27
6.1	Studiendesign	27
6.2	Probanden	27
6.3	Apparatus	29
6.4	Verlauf	29
6.5	Auswertung	32
7	Diskussion	41
8	Zusammenfassung und Ausblick	43
	Literaturverzeichnis	45

Abbildungsverzeichnis

2.1	Visualisierungspipeline nach Card et al. [CMS99]	12
2.2	Kognitive Theorie des Multimedialen Lernens nach Mayer [May09]	13
3.1	Drei Platzierungsmöglichkeiten der Visualisierung bezüglich der dazugehörigen Entität	16
5.1	Darstellung als Kreisdiagramm	22
5.2	Darstellung als Treemap	22
5.3	Darstellung als Säulendiagramm	22
5.4	Implementierte Visualisierung	23
5.5	Visualisierung mit Originalfarben	23
5.6	Visualisierung in Bezug auf Rotschwäche (Protanomalie)	23
5.7	Visualisierung in Bezug auf Rotblindheit (Protanopie)	23
5.8	Visualisierung in Bezug auf Grünschwäche (Deuteranomalie)	24
5.9	Visualisierung in Bezug auf Grünblindheit (Deuteranopie)	24
6.1	Verteilung der Twitternutzung	28
6.2	Vertrautheit mit verschiedenen Visualisierungstypen. Bewertungen wurden auf einer Skala von 1 bis 5 abgegeben.	28
6.3	Beispiele aus dem verwendeten Twitter Feed (Versuchsbedingung: Visualisierung)	30
6.4	Beispiele aus dem verwendeten Twitter Feed (Versuchsbedingung: Farbe)	31
6.5	Beispiele aus dem verwendeten Twitter Feed (Versuchsbedingung: Text)	31
6.6	Histogramm für die benötigte Studienzeit	32
6.7	Einfache Boxplots der benötigten Studienzeiten	33
6.8	Säulendiagramm für die benötigte Aufgabenzeit	33
6.9	Durchschnittswert und Standardabweichung für einzelne Versuchsbedingungen (benötigte Aufgabenzeit)	34
6.10	Test für Normalverteilung	34
6.11	Ergebnisse des Kruskal-Wallis Test	34
6.12	Antworten auf Frage 1	35
6.13	Antworten auf Frage 2	35
6.14	Antworten auf Frage 3	36
6.15	Chi-Quadrat Test für alle Versuchsbedingungen (nur für Frage 1)	36
6.16	Chi-Quadrat Test für die Versuchsbedingungen <i>Visualisierung</i> und <i>Farbe</i> (nur für Frage 1)	37
6.17	Chi-Quadrat Test für die Versuchsbedingungen <i>Visualisierung</i> und <i>Text</i> (nur für Frage 1)	37
6.18	Chi-Quadrat Test für die Versuchsbedingungen <i>Farbe</i> und <i>Text</i> (nur für Frage 1)	38
6.19	Antworten auf Frage 4	38
6.20	Mann-Whitney-U-Test für die Versuchsbedingungen <i>Visualisierung</i> und <i>Farbe</i>	40

6.21	Mann-Whitney-U-Test für die Versuchsbedingungen <i>Visualisierung</i> und <i>Text</i> . . .	40
6.22	Mann-Whitney-U-Test für die Versuchsbedingungen <i>Farbe</i> und <i>Text</i>	40

1 Einleitung

Auf Social-Media-Plattformen wie Twitter werden häufig Bilder, Videos oder Links geteilt und kommentiert. Gerade wenn es dabei um kontroverse und emotionale Themen geht, verliert der Leser bzw. die Leserin schnell den Überblick über die diskutierten Themen. Während es im ersten Moment zum Beispiel um einen Impfstoff in Zeiten der Pandemie geht, werden in den nächsten Tweets politische Beschlüsse diskutiert und später Möglichkeiten, regionale Restaurants in dieser schwierigen Lage zu unterstützen. Zumeist lässt sich eine solche thematische Abstufung in der Realität nicht wirklich erkennen. Die Themenvielfalt führt schnell zu einer unübersichtlichen Menge, die die Nutzer nur mit Mühe einschätzen und interpretieren können.


Clusteringverfahren setzen genau an diesem Punkt an. Sie bieten nämlich die Möglichkeit, Angaben über die Themenvielfalt zu machen, indem sie ähnliche Kommentare gruppieren. Diese Angaben können dann wiederum visuell dargestellt werden, sodass die Nutzer die Informationen möglichst einfach aufnehmen können. Gerade wenn es um kleine Online-Texte wie beispielsweise Tweets geht, ist es von Vorteil, wortgroße Visualisierungen zu nutzen, da diese sich besonders leicht in den Textfluss integrieren lassen.

Mit einer Online-Studie soll untersucht werden, ob und inwieweit die entwickelten Visualisierungsmöglichkeiten Leser dabei unterstützen, die Themen- und Meinungsvielfalt einzuschätzen und verschiedene Kommentare bzw. Tweets dementsprechend einzuordnen. Hierbei soll auch untersucht werden, ob die Komplexität der verschiedenen Visualisierungsmöglichkeiten Auswirkungen auf das Ergebnis hat.





Die Arbeit befasst sich zunächst mit der Erläuterung wichtiger grundlegender Begriffe und Theorien und widmet sich danach verwandten wissenschaftlichen Texten. Hier wird ein besonderer Fokus auf das Design und die Platzierung wortgroßer Visualisierungen gesetzt. Aufbauend auf den Erkenntnissen dieser Arbeiten wird die Visualisierung entwickelt. Im Kapitel Implementierung wird auf das Design und die Funktionalität der Visualisierung eingegangen, aber auch auf die verwendeten Programmiersprachen und Bibliotheken sowie auf den verwendeten Datensatz. Um die Visualisierung zu testen, wird eine Online-Studie geplant und durchgeführt. Die daraus resultierenden Ergebnisse werden zusammengefasst und diskutiert. Zum Schluss wird noch ein Ausblick über mögliche Verbesserungen bezüglich der Visualisierung gegeben.

2 Grundlagen

2.1 Wortgroße Visualisierungen

Der Begriff „wortgroße Visualisierungen“ (engl.: word-scale visualizations) ist eine Weiterentwicklung des Konzeptes „Sparklines“. In seiner Theorie beschreibt Edward Tufte Sparklines als „datenreiche, design-simple, wortgroße Grafiken“ [Tuf06]. Im Allgemeinen bezeichnen diese Sparklines oft kleine Linien- oder Balkendiagramme (), die sich in unmittelbarer Nähe zu dem Wort oder der Entität befinden, auf die sie sich beziehen. [BW17]

In einer Erweiterung dieses Begriffes wurde der Term „wortgroße Grafiken“ eingeführt. Diese neue Definition beinhaltet eine größere Variation an Grafiken. Hierbei macht es keinen Unterschied, ob die Visualisierung die Größe eines einzelnen Buchstabens, Wortes oder eines Paragraphen hat. [GBWI17]

Im Allgemeinen unterscheidet man zwischen zwei Arten von wortgroßen Grafiken: Grafiken mit einer datengesteuerten Kodierung (z.B. , ) und Grafiken ohne eine datengesteuerte Kodierung (z.B. , ). Während Ersteres sich auf quantitative Informationen bezieht, wie beispielsweise eine Länge oder einen Prozentwert, repräsentiert die zweite Art von wortgroßen Grafiken allgemeine Konzepte bzw. Ideen. [GBWI17]

Wenn wir in dieser Arbeit von wortgroßen Visualisierungen sprechen, handelt es sich hierbei immer um die oben genannten Grafiken mit einer datengesteuerten Kodierung.

Die Nutzung von wortgroßen Visualisierungen unterstützt meist eins von drei Zielen. Die erste Möglichkeit ist es, den Inhalt eines Textes zu verdeutlichen. Hierbei können Daten beispielsweise direkt aus dem Text genommen werden, visuell dargestellt werden (z.B. Kreis- oder Balkendiagramm) und dann als unterstützendes Element in den Text eingefügt werden. Solche Abbildungen können als Vermittler komplexer Daten dienen oder sogar Zusammenfassungen des Inhalts darstellen. [GBWI17]

Ein zweites Ziel kann sein, den Inhalt des Textes zu erweitern bzw. ihn in Kontext zu setzen. Wenn man beispielsweise von einem Aktienwert eines Unternehmens spricht und verdeutlichen möchte, wie sehr sich dieser innerhalb eines bestimmten Zeitraums verändert hat, ist es hilfreich den Aktienkurs insgesamt darzustellen. Hierbei unterstützt die Nutzung externer Daten das Verständnis des Textes. [GBWI17]

Ein weiteres Ziel des Autors kann sein, Informationen zu teilen, die über den Inhalt des Textes hinausgehen. Diese Möglichkeit wird insbesondere dann genutzt, wenn es darum geht, Leser auf einen anderen gegensätzlichen Standpunkt hinzuweisen. Auf diese Art und Weise werden die Leser motiviert, sich kritisch mit dem Inhalt des Textes auseinanderzusetzen und sich so eine eigene Meinung zu bilden. [GBWI17]

In Kapitel 3 wird darauf eingegangen, wie man wortgroße Visualisierungen gestalten und positionieren kann und was für einen Einfluss dies auf das Leseverhalten hat.

2.2 Text-Bild-Verständnis

Bevor genauer auf Theorien des Text-Bild-Verständnisses eingegangen wird, setzt sich diese Arbeit mit den Zielen der Informationsvisualisierung auseinander. Card et al. [CMS99] definieren Informationsvisualisierung als „die Nutzung computergenerierter, interaktiver, visueller Repräsentationen von abstrakten, nicht-physikalischen Daten zur Verstärkung des Erkenntnisgewinns“.

Im Allgemeinen dient die Informationsvisualisierung also dazu, Daten auf eine visuelle und aussagekräftige Weise darzustellen, damit Nutzer sie besser verstehen können. Um dieses Ziel zu erreichen, sind einige Schritte notwendig.

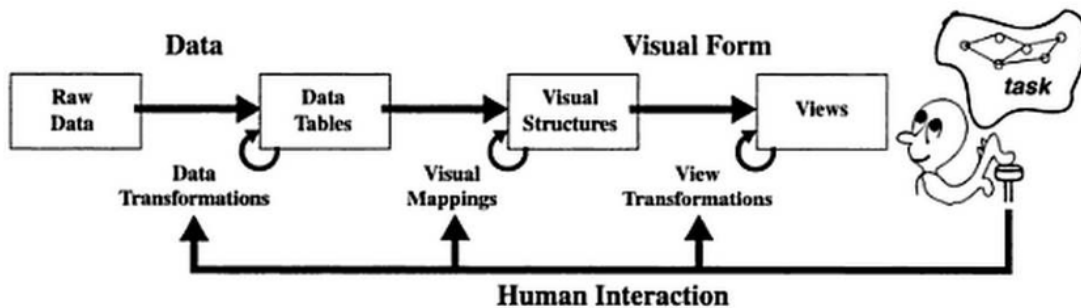


Abbildung 2.1: Visualisierungspipeline nach Card et al. [CMS99]

Abbildung 2.1 zeigt ein Modell, das die einzelnen Schritte in einer Visualisierungspipeline zusammenfasst. Zunächst werden die Rohdaten eingelesen und aufbereitet, sodass die Datenmenge auf die wichtigsten Daten reduziert wird. Im Anschluss werden visuelle Strukturen wie Form oder Farbe für die Daten erstellt. Die Ansichtstransformation bestimmt schließlich Eigenschaften wie die Position und den Maßstab der visuellen Strukturen.[Tat15]

Zu den Zielen der Informationsvisualisierung gehören das Entdecken von neuen Zusammenhängen zwischen den präsentierten Daten, eine einfachere Erforschung des Informationsraums und eine zuverlässige Entscheidungsfindung. [PD10]

Da die in Kapitel 2.1 beschriebenen Grafiken nicht immer einen riesigen Datensatz repräsentieren, stellt sich die Frage, warum es überhaupt sinnvoll ist, Grafiken in Texte einzubinden, wenn diese doch schon einen bestimmten Sachverhalt hinreichend wiedergeben.

Die Antwort darauf ist der sogenannte Multimedia-Effekt. Dieser beschreibt den positiven Einfluss, den eine solche duale Darstellungsform hat. Im Folgenden wird dieser Einfluss mittels theoretischer Modelle erklärt.

Die *Kognitive Theorie des Multimedialen Lernens* [May09] stützt sich auf die Annahme, dass das kognitive System in drei Speicher unterteilt ist. Abbildung 2.2 zeigt diese Dreiteilung durch die Unterscheidung von *Sensorischem Speicher*, *Arbeitsgedächtnis* und *Langzeitgedächtnis*.

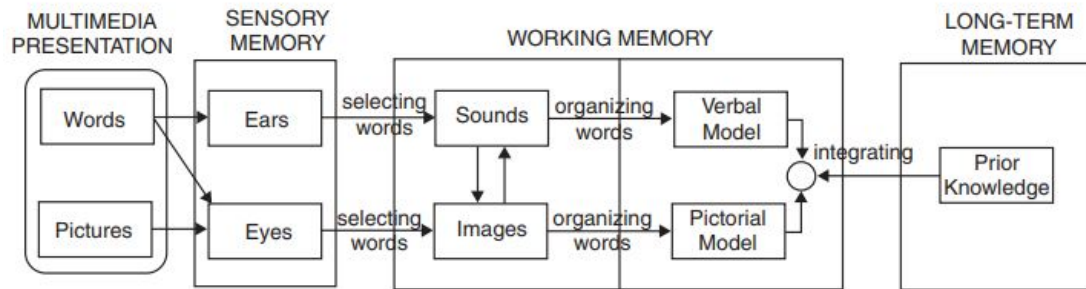


Abbildung 2.2: Kognitive Theorie des Multimedialen Lernens nach Mayer [May09]

Die Theorie verfolgt die Annahme, dass die Informationsverarbeitung in einem auditiv-verbalem und in einem visuell-bildhaften Kanal stattfindet. Die Unterteilung in diese zwei Kanäle beginnt schon beim sensorischen Speicher. Auditive Informationen werden demnach im auditiv-verbalem Kanal verarbeitet. Visuelle Informationen werden im visuell-bildhaften Kanal verarbeitet. Wie in Abbildung 2.2 beschrieben, können die Informationen aber auch zwischen den Kanälen wechseln. So wird geschriebener Text beispielsweise zunächst im visuellen Teil des visuell-bildhaften Kanals und dann im verbalen Teil des auditiv-verbalem Kanals verarbeitet. [Sch10]

Im beschriebenen Modell spielt das Arbeitsgedächtnis für das Text- und Bildverständnis eine zentrale Rolle. So kann eine duale Informationsdarstellung dazu führen, dass das Arbeitsgedächtnis durch die parallele Text- und Bildverarbeitung optimal ausgelastet wird und auf diese Weise bessere Lernergebnisse erzielt werden. Andererseits betont die *Cognitive Load Theorie* (CLT, Sweller 1999,2005), dass eine Überlastung des Arbeitsgedächtnis möglich ist. Diese lässt sich der Theorie zufolge auf drei Faktoren zurückführen.

- **intrinsic load** Die Komplexität der Aufgabe entscheidet über die Beanspruchung des Arbeitsgedächtnisses.
- **extraneous load** Eine schlechte Darstellung und Gestaltung des Lernmaterials führt zu einer höheren Beanspruchung des Arbeitsgedächtnisses.
- **germane load** Dieser beschreibt die kognitive Belastung, die für den Lernprozess notwendig ist.

Der *Cognitive Load Theorie* zufolge muss trotz „intrinsic load“ und „extraneous load“ genug Arbeitsspeicherkapazität zur Verfügung stehen, damit der Mensch sich neues Wissen aneignen kann. Da der „intrinsic load“ jedoch nicht reduziert werden kann, muss der „extraneous load“ reduziert werden. Dies bedeutet, dass der Lerninhalt optimal gestaltet werden muss, um bestmögliche Ergebnisse erzielen zu können.

Die Theorien zeigen, dass es nicht immer sinnvoll ist, zusätzliche Informationen (beispielsweise durch Bilder) in den Text einzubauen. Wenn es sich aber um instruktionale Bilder handelt, also um Abbildungen, die eine erklärende oder organisierende Funktion besitzen, ist es unter bestimmten Bedingungen sinnvoll, diese einzubauen. [BS12]

Diese Bedingungen lauten wie folgt:

- Zwischen dem Text und dem Bild muss ein kohärenter Zusammenhang bestehen.

- Die Informationen sind zu komplex bzw. das Vorwissen der Lesenden ist zu gering. Hier hilft das Bild dem Lernenden dabei, sich eine Vorstellung bilden zu können.
- Die Lernenden besitzen die nötigen kognitiven Fähigkeiten, das Bild zu verarbeiten.

2.3 Online-Studien

Es gibt immer mehr Möglichkeiten, das Verständnis komplexer Inhalte mit Hilfe von Visualisierungen zu erleichtern. Aufgrund dieser Visualisierungsvielfalt ist es jedoch auch umso wichtiger, zu erforschen, wie Nutzer mit solchen Visualisierungen interagieren können. [ED16]

Zwar werden viele solcher Nutzerstudien immer noch im Labor vor Ort durchgeführt, dennoch steigt die Anzahl von online durchgeführten Studien. Dabei wird oft die „Crowdsourcing“-Methode verwendet. Crowdsourcing beschreibt die Auslagerung verschiedener Teilaufgaben an Freiwillige. Bei Nutzerstudien zielt der Crowdsourcing-Ansatz vor allem darauf ab, möglichst viele unterschiedliche Nutzer auf die Studie aufmerksam zu machen.

Obwohl diese Art von Studien verschiedene Vorteile bringt, wie beispielsweise eine größere Anzahl an Nutzern, geringeren Zeitaufwand und geringere Kosten, bergen Online-Studien auch manche Risiken.

Man sollte sich nämlich bewusst sein, dass einige Nutzer nur des Geldes wegen an den Studien teilnehmen und sich deshalb auch einfach nur zufällig durchklicken bis die Studie beendet ist. Um diesen „Random-Clickers“ entgegenzuwirken, muss man sich also Mechanismen überlegen, die diese Nutzer filtern und so eine Verfälschung der Ergebnisse verhindern. [KYY12]

Da man außerdem in Online-Studien nicht direkt mit den Nutzern kommunizieren kann, um gegebenenfalls nochmals Sachen zu erläutern, sind verständliche Erklärungen und Versuchsaufgaben entscheidend.

Je nachdem, was untersucht werden soll, ist es darüber hinaus wichtig, bestimmte Logging-Mechanismen einzubauen. So kann man beispielsweise sehen, was die Nutzer angeklickt haben oder wie viel Zeit sie für eine bestimmte Aufgabe benötigt haben. [ED16]

3 Verwandte wissenschaftliche Arbeiten

In diesem Kapitel werden wissenschaftliche Arbeiten hinsichtlich des Designs und der Positionierung wortgroßer Visualisierungen vorgestellt.

3.1 Design und Platzierung von wortgroßen Visualisierungen

In ihrer Arbeit setzen sich Goffin et al. [GBWI17] mit dem Gestaltungsraum wortgroßer Visualisierungen auseinander. Hier wird zunächst einmal zwischen einer statischen und dynamischen Integration unterschieden.

Dynamische Integration bedeutet, dass die Visualisierung als Antwort auf eine bestimmte Nutzereingabe dargestellt wird. So kann eine Grafik beispielsweise sichtbar werden, wenn sich die Nutzer mit dem Mauszeiger über der Grafik befinden. Der Vorteil einer dynamischen Integration besteht darin, dass die Nutzer entscheiden, wie viel sie von den Visualisierungen sehen möchten. Auf diese Weise kann sicher gestellt werden, dass die Text-Visualisierung-Ratio nicht zu unübersichtlich für die Nutzer wird.

Die dynamische Integration hat zum Nachteil, dass sie nicht in Papierform genutzt werden kann. Auch erschwert sie zumeist die Möglichkeit eines direkten Vergleichs zweier Grafiken, da man je nach Integrationsmodell nicht beide Grafiken gleichzeitig sehen kann. Darüber hinaus ändert sich gegebenenfalls der Textfluss bei jedem Erscheinen oder Verschwinden einer Visualisierung, was unter anderem das Lesen erschweren kann.

Eine statische Integration der wortgroßen Visualisierungen kann dagegen in Papierform genutzt werden, da die Grafiken hier dauerhaft dargestellt werden. Ein anderer Vorteil, der sich hieraus ergibt, ist die Möglichkeit des direkten Vergleichs zwischen den Visualisierungen. Da die Leser bei einer statischen Integration jedoch nicht mehr entscheiden, wie viel sie von den Visualisierungen sehen, könnte es zu einer Informationsüberflutung kommen, die wiederum das Leseverständnis beeinträchtigen kann.

Der Autor muss hier also die einzelnen Vor- und Nachteile abwägen und das Integrationsmodell auswählen, das am besten zu seinen Anforderungen passt.

Goffin et al. stellen unter anderem auch verschiedene Platzierungsmöglichkeiten für wortgroße Visualisierungen vor.

Die Arbeit von Goffin et al. konzentriert sich dabei auf drei Hauptpositionen (vgl. Abbildung 3.1) im Bezug auf die Entität:

1. Es gibt die *top* Position, in welcher die linke, untere Ecke der Visualisierung auf der linken, oberen Ecke der Entität liegt - die Visualisierung befindet sich also direkt über der Entität.

2. Es gibt die *right* Position, in welcher die linke, untere Ecke der Visualisierung auf der rechten, unteren Ecke der Entität liegt - die Visualisierung befindet sich also direkt neben der Entität.
3. Es gibt die *baseline* Position, in welcher die linke, untere Ecke der Visualisierung auf der linken, unteren Ecke der Entität liegt - Sie liegen also direkt aufeinander.

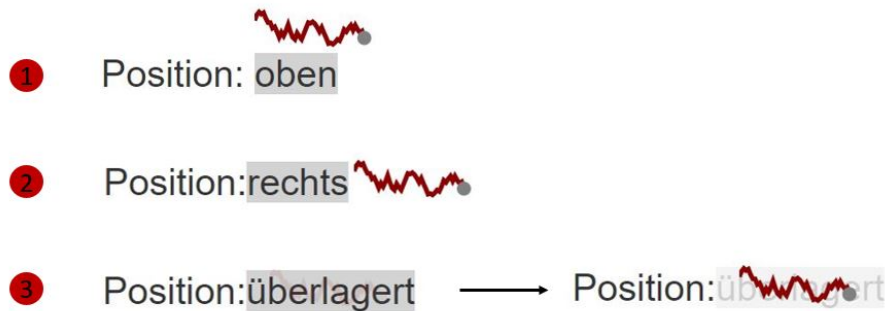


Abbildung 3.1: Drei Platzierungsmöglichkeiten der Visualisierung bezüglich der dazugehörigen Entität

3.2 Interaktionslevel der Visualisierungen

Der Designraum von Visualisierungen beinhaltet auch verschiedene Interaktionslevel, die von überhaupt nicht interaktiv bis sehr interaktiv reichen. In ihrer Arbeit unterscheiden Beck und Weiskopf [BW17] zwischen drei Ebenen:

- Keine Interaktion (Level 0): Wenn die Visualisierungen nicht übermäßig komplex und unübersichtlich ist, kann eine wortgroße Visualisierung in ihrer Gestalt selbst zusätzliche wichtige Informationen geben. Bei kleinen, statischen Visualisierungen müssen Autoren jedoch meist Erklärungen im umgebenden Text geben.
- Lokale Interaktion (Level 1): Hier liefert die Interaktion mit der Grafik zusätzliche Informationen, indem beispielsweise ein Detailsfenster beim Hovern erscheint, Teile der Visualisierung vergrößert werden oder relevante Textinhalte hervorgehoben werden.
- Globale Interaktion (Level 2): Globale Interaktionen beschreiben Verbindungen, die über den lokalen Rahmen hinausgehen. Diese Verbindungen können zum Beispiel zwischen verschiedenen Grafiken bestehen und einen Zusammenhang zwischen diesen aufzeigen.

Alle in dieser schriftlichen Arbeit verwendeten wortgroßen Visualisierungen sind nicht interaktiv und gehören zu Level 0.

3.3 Einfluss von wortgroßen Visualisierungen auf das Leseverhalten

In einer anderen Arbeit befassen sich Goffin et al. [GWBI15] mit dem Effekt, den wortgroße Visualisierungen auf das Leseverhalten haben. Sie analysierten dabei, ob und inwiefern die Platzierung von kleinen Visualisierungen Einfluss auf die Lesegeschwindigkeit und das Gedächtnis der Probanden hat.

Die untersuchten Platzierungsmöglichkeiten der Grafiken entsprechen den vorgestellten Positionen aus Abbildung 3.1.

Ergebnisse der Studie zeigten, dass obwohl die Lesegeschwindigkeit bei der Visualisierung mit der Platzierung *right* größer war, die Unterschiede insgesamt nicht signifikant waren.

Auch konnte die Studie nicht bestätigen, dass die Visualisierung das Gedächtnis besser beeinflusst als der Text ohne Visualisierung.

Was die Studie aber zeigte, war dass die Mehrheit der Probanden die Informationen eher aus der Visualisierung als aus den Texten holte. Dies konnte untersucht werden, indem den Lesern ein Satz mit integrierter Visualisierung gezeigt wurde, wobei Text und Grafik widersprüchliche Informationen lieferten.

Die Ergebnisse eines Fragebogens der Studie legten nahe, dass keine der einzelnen Positionen als die beste Position angesehen werden konnte.

4 Problemstellung

Social-Media-Plattformen werden immer häufiger genutzt um, auf verschiedene Themen aufmerksam zu machen. Dabei steht vor allem Twitter im Mittelpunkt. Hier werden täglich Tausende von Bildern, Videos und Links geteilt und kommentiert.

Je nachdem wie kontrovers oder emotional das Thema ist, können Leser schnell den Überblick über die diskutierten bzw. angesprochenen Themen verlieren. Clusteringverfahren bieten die Möglichkeit, ähnliche Kommentare zu gruppieren und damit Angaben über die Themenvielfalt zu machen.

Um den Lesern diese Informationen am besten zur Verfügung zu stellen, bietet es sich an, die Angaben visuell darzustellen. Da es sich gerade bei Tweets um relativ kurze Online-Texte handelt, sollten die Darstellungsformen möglichst einfach und klein gehalten werden, da diese sich besonders leicht in den Textfluss integrieren lassen.

Der Kern dieser Arbeit befasst sich damit, eine geeignete Visualisierungsform zu entwickeln, sie zu implementieren und zu untersuchen, ob und inwieweit die Komplexität der Darstellungsform Einfluss auf das Leseverständnis hat.

5 Implementierung

In diesem Kapitel wird auf Details der Implementierung eingegangen. Dazu zählt zum einen, welches Design und welche Farben für die Darstellung der Visualisierung genutzt wurden. Zum anderen werden die genutzten Programmiersprachen, Bibliotheken und Frameworks erläutert. Zum Ende des Kapitels hin soll noch auf den verwendeten Datensatz eingegangen werden.

5.1 Funktionalität

Im Folgenden geht es darum, eine Visualisierungsform zu finden, die Informationen darüber geben kann, wie groß die Themenvielfalt in einem bestimmten Twitter Feed ist.

Die Visualisierung sollte dabei möglichst einfach und klein sein, da sie später bei jedem Tweet eingebunden wird. Sie soll bereits ohne Interaktion der Nutzer Aufschluss darüber geben können, wie viele Themen es gibt und in welchem Größenverhältnis sie zueinander stehen. Außerdem soll den Lesern durch die Visualisierung ersichtlich sein, in welchem Cluster sich der jeweilige Tweet befindet. So sehen die Nutzer auf den ersten Blick, wie vielfältig die Tweets sind und kann auch bei jedem Tweet einschätzen, wie sehr die Meinung dieses Tweets im Allgemeinen vertreten ist.

Genauere Informationen soll die Visualisierung nur geben, wenn das vom Nutzer erwünscht ist und er aktiv mit ihr interagiert. Diese Informationen umfassen dann die Größe des gesamten Twitter Feeds, die Größe des jeweiligen Clusters und Schlüsselwörter, die das Thema des jeweiligen Clusters repräsentieren. Da diese Informationen nur dann erscheinen, wenn die Nutzer aktiv mit der Visualisierung interagieren, wird gewährleistet, dass die Leser nicht mit zu viel unnötigen Informationen überfordert werden. Sie entscheiden, ob und wann sie genauere Angaben sehen möchten.

5.2 Design

5.2.1 Visualisierungstyp

Bei der Auswahl des Visualisierungstypen stand die Einfachheit im Mittelpunkt. Die Kuchendiagramm-Darstellung wurde dabei von vornherein ausgeschlossen, da es Menschen zum Teil schwer fällt, Winkel richtig abzuschätzen [Bar86]. Außerdem müsste das Kreisdiagramm bei vielen verschiedenen Clustern möglichst groß dargestellt werden, um gewährleisten zu können, dass die einzelnen Kreisabschnitte gut erkennbar sind. Dies widerspricht aber unserer Anforderung, die Visualisierung möglichst klein zu halten.

Eine ähnliche Argumentation ist auch bei der Treemap-Darstellung möglich: Die Größenverhältnisse kann man manchmal nicht gut einschätzen und außerdem können zu viele Cluster dazu führen, dass die Nutzer den Überblick verlieren und beispielsweise nicht mehr wissen, mit welchen der Treemap-Rechtecke sie schon interagiert haben.

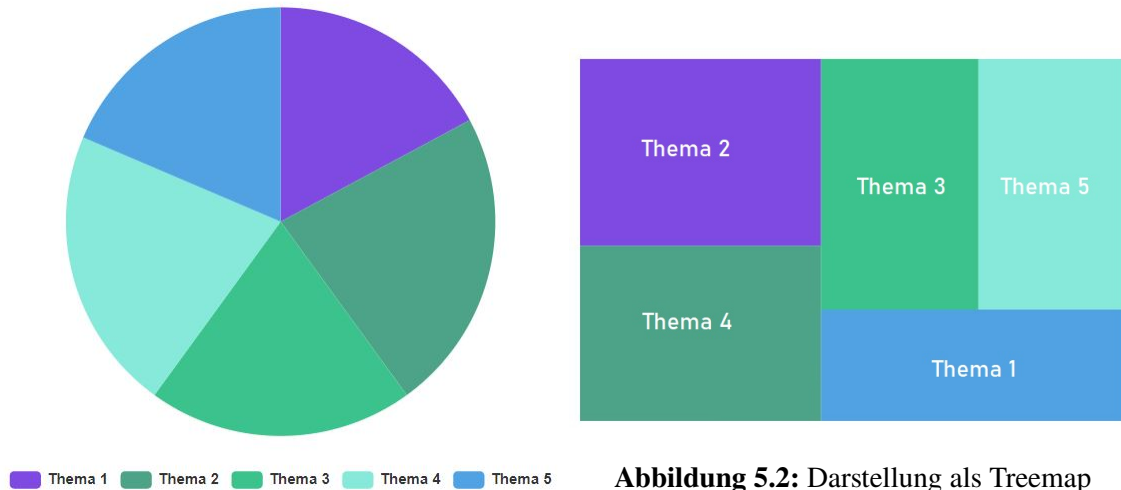


Abbildung 5.2: Darstellung als Treemap

Abbildung 5.1: Darstellung als Kreisdiagramm

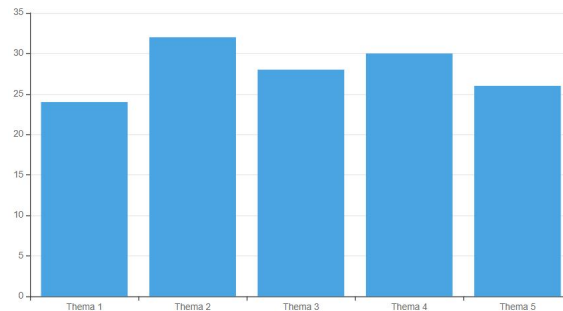


Abbildung 5.3: Darstellung als Säulendiagramm

Die Abbildungen 5.1 bis 5.3 zeigen, dass Kreisdiagramme und Treemaps Informationen nicht immer gut wiedergeben können. In beiden Darstellungsformen ist nicht unbedingt ersichtlich, welches Thema das größte ist. Wirft man aber einen Blick auf das Säulendiagramm darunter, sieht man sehr leicht, dass Thema 2 am größten ist.

Da es bei der Visualisierung darauf ankommt, sie möglichst kompakt zu halten, wurde der Darstellungsform gestapeltes Balkendiagramm mit nur einem Balken gewählt (siehe Abbildung 5.4). Dies bietet mehrere Vorteile: Egal wie viele Cluster man darstellen muss, der Platz, der in Y-Achsenrichtung benötigt wird, bleibt konstant. Außerdem können die Nutzer mit einer konstanten Rechtsbewegung des Cursors sequentiell mit allen Clustern interagieren. Ein Vorteil ist auch, dass der gesamte verfügbare Platz in X-Achsenrichtung genutzt wird.

Punkt 1 in Abbildung 5.4 zeigt einen Marker in der Visualisierung. Dieser gibt den Nutzern an, in welchem Cluster bzw. Themenbereich sich der jeweilige Tweet befindet. Die Größe des jeweiligen Rechtecks verdeutlicht, in welchem Größenverhältnis der jeweilige Cluster zu den anderen Clustern

steht. Bei Punkt 2 sieht man, dass ein schwarzes Informationsfenster erscheint, wenn man über die einzelnen Rechtecke des Balkens hovert. Im Fenster werden nähere Angaben bezüglich des Themenbereichs gemacht: So erfahren die Nutzer, wie viele Tweets zum jeweiligen Themenbereich gehören, was die Schlüsselwörter des jeweiligen Themas sind und wie viele Tweets es insgesamt im Twitter Feed gibt.

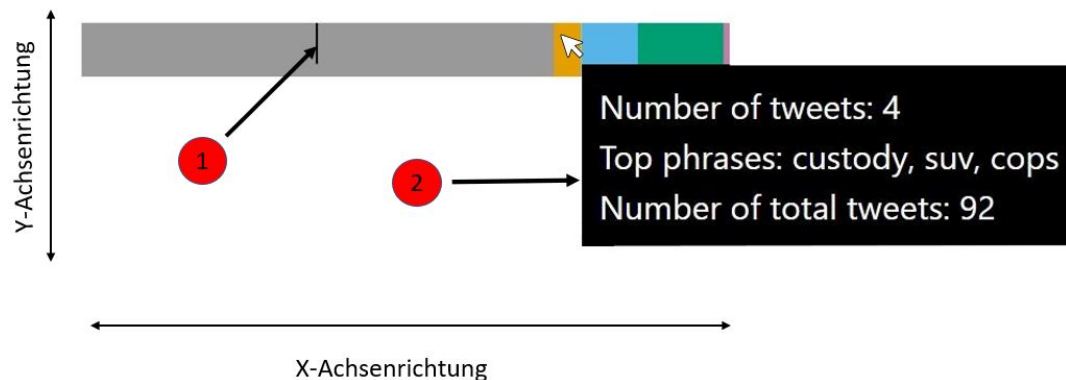


Abbildung 5.4: Implementierte Visualisierung

5.2.2 Farbauswahl

Die einzelnen Cluster bzw. Themenbereiche werden verschieden eingefärbt. Als Grundlage für die Farbpalette wurden verschiedene qualitative Farbpaletten von *colorbrewer* und die „Colour schemes and templates“-Webseite von Paul Tol verglichen. Bei der Auswahl der Farben wurde darauf geachtet, die Visualisierung möglichst freundlich für Menschen mit einer Sehschwäche zu gestalten. Die ausgewählte Farbpalette enthält 5 verschiedene Farbtöne und wurde mit *Coblis*, einem Farbenblindheitssimulator, getestet. Auch wurde eine Farblegende in den Fragekatalog eingebaut, wenn Fragen bezüglich eines bestimmten Clusters/Themas gestellt wurden.

Unten wird die Visualisierung mit ihren Originalfarben dargestellt und darüber hinaus auch eine mögliche Farbwahrnehmung von Menschen mit einer Farbsehschwäche (vgl. Abbildungen 5.5 - 5.8)



Abbildung 5.5: Visualisierung mit Originalfarben



Abbildung 5.6: Visualisierung in Bezug auf Rotschwäche (Protanomalie)



Abbildung 5.7: Visualisierung in Bezug auf Rotblindheit (Protanopie)



Abbildung 5.8: Visualisierung in Bezug auf Grünschwäche (Deuteranomalie)



Abbildung 5.9: Visualisierung in Bezug auf Grünblindheit (Deuteranopie)

5.3 Verwendete Programmiersprachen und Bibliotheken

Die Studie wurde im Single-Page-Design entwickelt, was bedeutet, dass alle Daten beim ersten Seitenaufruf geladen wurden. Die Implementierung der Seite erfolgte mit den Sprachen HTML5, CSS, Javascript und PHP.

Außerdem wurden für die Studie und Visualisierung folgende Bibliotheken und Frameworks genutzt:

FROE: Da die Studie auf der Online-Plattform Prolific laufen sollte, wurde FROE¹ genutzt. Dieses Framework wurde für eine einfachere Realisierung von Online-Studien mit Prolific entwickelt.

D3.js: Data Driven Documents (D3²) ermöglicht die effiziente Manipulation von Dokumenten auf der Grundlage von Daten. Bei minimalem Overhead ist D3 extrem schnell und unterstützt große Datensätze und dynamische Verhaltensweisen für Interaktion und Animation.

jQuery: jQuery³ ist eine kleine und funktionsreiche JavaScript-Bibliothek. Sie macht Dinge wie das Durchqueren und Manipulieren von HTML-Dokumenten, Ereignisbehandlung und Animation viel einfacher.

5.4 Datensatz

Der Datensatz, der dieser Studie zugrunde liegt, stammt von Tweets, die am 02.06.2020 gepostet wurden und wurde von den Betreuern zur Verfügung gestellt.

Die Daten wurden danach gruppiert, zu welcher URL sie sich jeweils äußern. Wenn sich beispielsweise zwei Tweets zu einem bestimmten Video-Link äußern, gehören sie zur selben Gruppe.

Von der gesammelten Datenmenge wurden die 100 größten Gruppen extrahiert, auf welchen dann ein Clusteringverfahren angewandt wurde. Bei dem Clusteringverfahren handelt es sich um das K Means ++ Clustering⁴, das in unserem Fall bis zu zehn verschiedene Cluster zurückgab.

Bei der Untersuchung der entstandenen Gruppen fiel auf, dass sehr viele Tweets den gleichen Inhalt enthalten (Retweets). Da wir für die Studie jedoch einen möglichst kompakten Twitter Feed wollten als einen mit über 2000 Tweets, wurde der Datensatz so bearbeitet, dass er nur einzigartige Tweets enthielt.

¹<https://github.com/yvonnejansen/FROE>

²<https://d3js.org/>

³<https://jquery.com/>

⁴<https://en.wikipedia.org/wiki/K-means%2B%2B>

Aus den 100 Gruppen wurde dann derjenige gewählt, der den folgenden Kriterien am meisten entsprach: Erstens, der Datensatz besitzt unter sieben Cluster und zweitens, die Cluster zeigen eine klare thematische Abgrenzung untereinander auf. Die Kriterien wurden deshalb so gewählt, da die Anzahl der Themen bei der Studie beschränkt werden sollte und die Probanden einen wirklichen Unterschied zwischen den Themen feststellen können sollten.

Der ausgewählte Datensatz enthielt insgesamt 92 einzigartige Tweets und bezog sich auf ein Video, das zeigt, wie ein SUV am 01.06.2020 in Buffalo (New York) in eine Menge von Polizisten und Militär fährt.

Die Informationen bezüglich der Tweets sind in sechs verschiedenen JSON-Dateien gegeben. In der ersten JSON Datei (*clusterinfo.json*) befinden sich allgemeine Eigenschaften und Informationen zu den Clustern. Listing 5.1 zeigt den Aufbau dieser Datei.

```
{
  "Link": "https://video.twimg.com/ext_tw_video/1267637503990824963/pu/vid/700x1280/
QMq2qaxMdRT1hEub.mp4?tag=10", #der geteilte Link bzw. das Datenfragment
  "NumClusters": 5, #Anzahl der Cluster
  "TotalNumTweets": 8508, #Anzahl aller Tweets in der Gruppe
  "TotalNumUniqueTweets": 92, #Anzahl der einzigartigen Tweets in der Gruppe
  "Clusters": [ #genauere Informationen bezüglich der einzelnen Cluster
    {
      "Id": 0, #Clusternummer, in unserem Fall von 0 bis 4
      "NumTweets": 5850, #Anzahl der Tweets in diesem Cluster
      "NumUniqueTweets": 67, #Anzahl der einzigartigen Tweets in diesem Cluster
      "TopPhrases": [ #Wörter oder Phrasen, die beschreibend für dieses Cluster sind
        {
          "Content": "buffalo", #Phrase
          "Weight": 0.7255007 #Gewichtung (je höher, desto wichtiger)
        },
        . . .
        {
          "Content": "dead",
          "Weight": 0.055667926
        }
      ],
      "ClosestSentences": [ # Tweets, die dem Clusterzentrum am nächsten sind
        {
          "Content": "a protest in a car run over a group of cops in buffalo ny some of them
might be dead", # Inhalt des Tweets
          "Weight": 0.23364776 # Distanz (je kleiner, desto näher)
        },
        . . .
        {
          "Content": "just happened a protest in a car run over a group of cops in buffalo ny
some of them might be dead 2020",
          "Weight": 0.26476336
        }
      ]
    },
  ],
}
```

```

    . . . .
  ]
}

```

Listing 5.1: Ausschnitt aus der Datei *clusterinfo.json*

Zu jedem Cluster gibt es zusätzlich eine separate Datei *cluster[Nr].json*, die alle Tweets mit den zugehörigen Metadaten enthält. Listing 5.2 zeigt einen Ausschnitt der Datei *cluster3.json*.

```

[
  {
    "Item1": { #gibt genauere Informationen zum jeweiligen Tweet
      "CreatedAt": "2020-06-02T02:51:06Z", #Uhrzeit der Erstellung des (Re)Tweets (in GMT+0)
      "Id": 1267649917868052480, #Eindeutige Tweet ID
      "Text": "VIDEO: Domestic terrorist appears to run over police in New York State. ", #
      Inhalt des Tweets
      "UserId": 125128723, #Eindeutige Nutzer ID
      "UserName": "Raheem Kassam", #Nutzername
      "UserScreenName": "RaheemKassam",
      "UserIsVerified": true,
      "UserFollowersCount": 195835,
      "UserFriendsCount": 2065,
      "UserListedCount": 1343,
      "UserFavouritesCount": 27456,
      "UserStatusesCount": 10451,
      "UserCreatedAt": "2010-03-21T19:58:27Z",
      "UserProfileImageUrlHttps": "",
      "Language": "en",
      "PossiblySensitive": false,
      "IsRetweet": false,
      "EntitiesHashTags": [],
      "EntitiesUserMentions": [],
      "EntitiesSymbols": [],
      "EntitiesUrl": [],
      "EntitiesMedia": [
        {
          "Url": "",
          "ExpandedUrl": "",
          "ContentUrl": ""
        }
      ]
    },
    . . . .
  ]
}

```

Listing 5.2: Ausschnitt aus der Datei *cluster3.json*. Die für die Studie wichtigen Keywords sind rot markiert.

6 Studie

6.1 Studiendesign

Für die Studie wurde das „Between-Groups“ Design genutzt. Dabei wurde nur der Einfluss einer einzigen unabhängigen Variable getestet. Diese Variable umfasst drei Versuchsbedingungen: Darstellung mittels Visualisierung, Darstellung mittels Farbhintergrund und Darstellung mittels textueller Angabe. Das „Between-Groups“ Design garantiert hier, dass unterschiedliche Bedingungen von unterschiedlichen Teilnehmern getestet werden. Teilnehmer konnten also nicht mehrere Versuchsbedingungen testen.

Die zu testende Bedingung war die einzige Variable, die systematischen Einfluss auf das Experiment nehmen sollte. Durch eine zufällige Zuteilung der Bedingungen an die Teilnehmer sollte sichergestellt werden, dass andere äußere Einflüsse (Alter der Teilnehmer, Geschlecht, Sehschwäche etc.) nicht systematisch einer Personengruppe zugeordnet werden können. So wird nämlich auch nur der Einfluss der zu testenden Versuchsbedingung betrachtet.

Bei der Untersuchung ging es in erster Linie darum, den Zusammenhang zu folgenden abhängigen Variablen zu untersuchen: die Zeit, die für die Aufgabe benötigt wurde, die Korrektheit der ersten drei Fragen und die subjektive Beantwortung der vierten Frage.

6.2 Probanden

Insgesamt nahmen 66 Probanden an der Studie teil. Bei der Auswertung der Ergebnisse wurden jedoch vier Probanden ausgeschlossen (siehe Kapitel 6.5). Daher wird im Folgenden nur auf die verbliebenen 62 Teilnehmer eingegangen.

Circa 63% der Probanden waren männlich, der Rest weiblich. Die Mehrheit (82,3%) bestand aus 18 bis 27 Jährigen, 14,5% gehörten zur Altersgruppe 28 bis 38 Jahre und der Rest zur Altersgruppe 38 bis 47 Jahre. Zwei Probanden gaben an, dass sie unter einer Rot-Grün-Schwäche leiden und eine Person, dass sie eine Blau-Gelb-Schwäche hat.

Da die wortgroßen Grafiken in Tweets integriert werden sollten, wurden die Teilnehmer nach ihrer Twitternutzung befragt. Dabei konnten sie einer der Möglichkeiten *niemals*, *monatlich*, *wöchentlich*, *täglich* auswählen. Das Säulendiagramm aus Abbildung 6.1 zeigt die Verteilung dazu.

43,5% der Probanden gaben an, dass sie Twitter täglich nutzen. 14,5% waren wöchentliche Twitter-Nutzer und der Rest nutzte die Plattform entweder monatlich oder niemals.

Außerdem sollten die Probanden angeben, wie vertraut sie mit bestimmten Visualisierungstypen sind. Die Angaben sollten sie auf einer Skala von 1 bis 5 machen, wobei 1 *überhaupt nicht vertraut* und 5 *sehr vertraut* bedeutet.

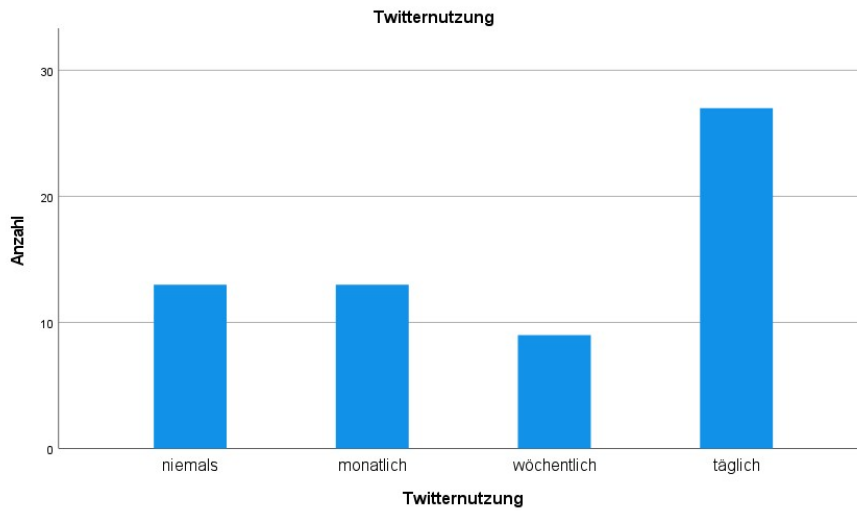


Abbildung 6.1: Verteilung der Twitternutzung

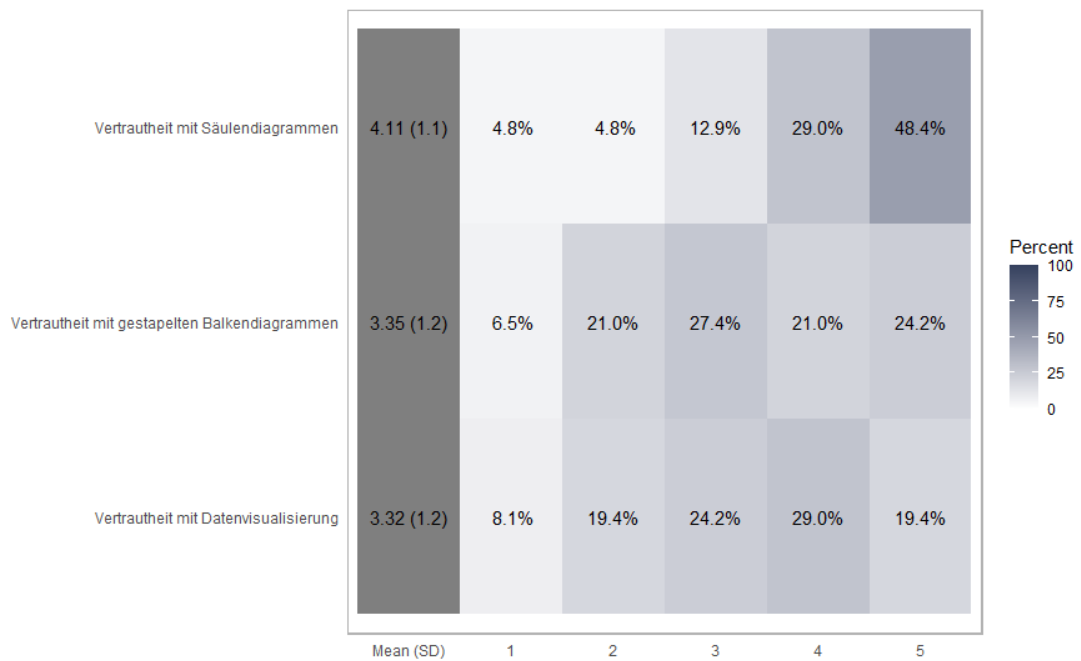


Abbildung 6.2: Vertrautheit mit verschiedenen Visualisierungstypen. Bewertungen wurden auf einer Skala von 1 bis 5 abgegeben.

Während die Probanden allgemein angaben, mit Datenvisualisierung vertraut zu sein, kann man der Heatmap aus Abbildung 6.2 entnehmen, dass ihre Erfahrungen eher auf Säulendiagrammen basieren als auf gestapelten Balkendiagrammen.

6.3 Apparat

Die Online-Studie wurde auf Prolific¹ gehostet. Prolific ist eine Crowdsourcing-Plattform, die eine schnelle und zuverlässige Datenerhebung ermöglicht, indem sie Forschende mit Zielpersonen auf der ganzen Welt verbindet.

In den Einstellungen der hier durchgeführten Prolific-Studie wurde festgelegt, dass die Probanden zwei Kriterien erfüllen müssen, um zugelassen zu werden: Zum einen sollten sie die englische Sprache fließend beherrschen, damit die späteren Ergebnisse nicht durch eventuelle Sprachbarrieren verfälscht werden. Zum anderen sollten sie die Studie über einen Desktop durchführen, da die Darstellung der Studie und der Visualisierung hier am meisten getestet wurde.

Einer der wichtigsten Punkte beim Hosten der Studie war die Erfassung der Prolific ID. Die ID besteht aus 24 alphanumerischen Zeichen und dient der Pseudonymisierung der Teilnehmer. Damit die Durchführung richtig geloggt werden konnte, wurde die Prolific ID in den URL Parametern der Studie übergeben und verarbeitet. Außerdem musste ein Bestätigungslink generiert und in der Studie integriert werden, damit die Studie im Prolific-System als beendet gekennzeichnet wird. Nur auf diese Weise kann gewährleistet werden, dass die Probanden für ihre Teilnahme bezahlt werden.

Der letzte Schritt vor der Veröffentlichung der Studie war die Angabe der gewünschten Teilnehmerzahl und des Lohns, den die Probanden erhalten sollten.

6.4 Verlauf

In diesem Abschnitt wird der Ablauf der Studie erklärt. Nach einer kurzen Studienbeschreibung und der Datenschutzerklärung gelangen die Teilnehmer zum ersten Fragenkatalog. Neben demographischen Daten wie Geschlecht oder Alter sollten sie Informationen darüber geben, wie oft sie auf der Social-Media-Plattform Twitter aktiv sind und ob sie gegebenenfalls an einer Sehschwäche leiden.

Im zweiten Fragenkatalog ging es um die Vertrautheit der Teilnehmer mit Datenvisualisierung, insbesondere der Darstellung mittels Säulendiagrammen und gestapelten Balkendiagrammen.

Im Anschluss dazu wurde eine kurze Einführung in die Darstellungsform der jeweiligen Bedingung und die Aufgabenstellung eingeblendet. Bevor die Probanden zur richtigen Aufgabe gelangten, sollten sie eine Übungsaufgabe machen. Diese ist einfacher als die richtige Aufgabe und umfasst einen Twitter Feed mit lediglich sechs Tweets. Die Aufgabe dient gleichermaßen als Einführung der Darstellungstypen als auch als mögliches Ausschlusskriterium bei der späteren Ergebnisauswertung.

Den Teilnehmern wurde dann das Thema des nächsten Twitter Feeds vorgestellt: Die Tweets beziehen sich alle auf ein Video, das zeigt wie ein SUV am 01. Juni 2020 in eine Menge von Polizisten fährt. Da das eigentliche Video sensible Inhalte enthält, wurde nur ein Screenshot gezeigt. Der Link für die Quelle wurde mit einem Hinweis auf möglicherweise verstörende Inhalte hinzugefügt.

¹<https://www.prolific.co/>

Nun erschien die Seite der eigentlichen Aufgabe. Die Seite war in zwei Hälften unterteilt: Links befand sich der Twitter Feed, der aus dem Datensatz von Kapitel 5.4 generiert wurde. Auf den ersten Blick bestand der Feed aus 21 Tweets, jedoch konnten sich die Probanden optional mehr Tweets anzeigen lassen, indem sie auf den Button „More Tweets“ klickten. Insgesamt standen ihnen Daten von 92 Tweets zur Verfügung.

Da in der Studie drei unterschiedliche Versuchsbedingungen untersucht werden, sieht der Twitter Feed bei den Probanden unterschiedlich aus (siehe Abbildungen 6.3 - 6.5).

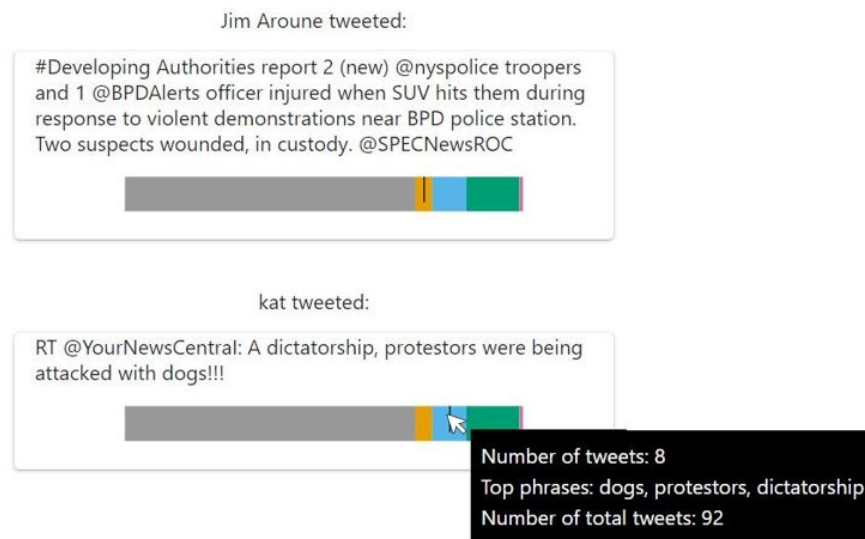


Abbildung 6.3: Beispiele aus dem verwendeten Twitter Feed (Versuchsbedingung: Visualisierung)

Abbildung 6.3 zeigt einen Ausschnitt des Twitter Feeds, der bei Bedingung 1 verwendet wurde. Alle Tweets wurden mit der integrierten Visualisierung angezeigt. Der einzige Unterschied, der zwischen den Tweets bestand, war der Marker, der angibt, zu welchem Thema bzw. Cluster der jeweilige Tweet gehört. So sieht man im Beispiel, dass der erste Tweet zum zweiten Cluster gehört und der zweite Tweet zum etwas größeren dritten Cluster.

Bei der zweiten Bedingung waren die Tweets unterschiedlich eingefärbt. Die Hintergrundfarbe zeigte zu welchem Thema der jeweilige Tweet gehört. Im Beispiel in Abbildung 6.4 sieht man, dass die angezeigten Tweet unterschiedliche Dinge diskutieren.

Abbildung 6.5 bildet einen Ausschnitt des Twitter Feeds von Bedingung 3 ab. Hier wurde bei jedem Tweet in der unteren rechten Ecke angegeben, zu welchem Thema der jeweilige Tweet gehört. Das Beispiel oben zeigt, dass der erste Tweet zum fünften Thema gehört und der zweite Tweet zum Thema 1.

Die rechte Hälfte der Aufgabenseite bestand aus den Aufgaben, die die Teilnehmer lösen sollten. Alle Aufgaben konnten mit den primären 21 Tweets gelöst werden oder auch mit den Tweets, die bei jedem Drücken des „More Tweets“-Buttons hinzugefügt wurden.

Heavy Lifting ★★ ★ tweeted:

@FBI @TheJusticeDept Just happened: a protest in a car run over a group of cops in Buffalo NY... Some of them might be dead #Riots

kat tweeted:

RT @YourNewsCentral: A dictatorship, protestors were being attacked with dogs!!!

Abbildung 6.4: Beispiele aus dem verwendeten Twitter Feed (Versuchsbedingung: Farbe)

Black Lives Matter tweeted:

RT @tiawdnaeulav: #WorldLeadersAgainstTrump Americans need you!!! Speak out against the human rights violations #Bl...

topic 5

Fancyfaceb796 tweeted:

@JoeBiden Peaceful protests? BREAKING: Car runs into police officers during protest in Buffalo, New York; at least 2 injured

topic 1

Abbildung 6.5: Beispiele aus dem verwendeten Twitter Feed (Versuchsbedingung: Text)

Bei der ersten Aufgabe sollten die Teilnehmer angeben, wie viele unterschiedliche Themen im Twitter Feed besprochen wurden. Danach sollten sie einschätzen, wie viele Tweets prozentual zum größten Thema gehörten. Bei der nächsten Aufgabe wurden den Teilnehmern verschiedene Schlüsselwörter gezeigt und sie sollten entscheiden, welche davon das Thema 3 (die blau eingefärbten Tweets) am besten beschreiben. Außerdem sollten die Probanden angeben, welche zwei Themen ihrer Meinung nach am widersprüchlichsten sind.

Nach jeder dieser Aufgaben sollten die Teilnehmer den Schwierigkeitsgrad auf einer Skala von 1 bis 5 einschätzen.

Sobald die Probanden mit der Hauptaufgabe fertig waren, sollten sie in einem letzten Fragebogen angeben, wie sicher sie sich mit ihren Antworten fühlten und wie sehr ihnen die Darstellungsform der verschiedenen Themen gefiel.

Außerdem konnten sie zum Schluss noch optional Feedback zur Studie, den Aufgaben oder der Darstellungsform geben.

6.5 Auswertung

Im Folgenden wird auf die Ergebnisse der Online-Studie eingegangen. In den Grafiken wird zum besseren Verständnis von den drei Bedingungen *Visualisierung*, *Farbe* und *Text* gesprochen. Hierbei steht *Visualisierung* für die Darstellung mittels der entwickelten Visualisierung, *Farbe* für die Darstellung mittels der Hintergrundfarbe des jeweiligen Tweets und *Text* für die Darstellung mittels textueller Angabe im Tweet.

Wie bereits in Kapitel 6.2 erwähnt, wurden vier Teilnehmer bei der Auswertung ausgeschlossen. Das Ausschlusskriterium bestand aus zwei Punkten: Zum einen mussten die Teilnehmer auffällig wenig Zeit für die Studiendurchführung aufgewandt haben, und zum anderen mussten sie die Fragen der Übungsaufgabe falsch beantwortet haben. Nur wenn beide Punkte zutrafen, wurden Probanden ausgeschlossen.

Da vier Probanden das Ausschlusskriterium erfüllten, wurden nur die Ergebnisse der verbliebenen 62 Teilnehmer ausgewertet.

Für die Auswertung wurde die Softwareplattform IBM® SPSS® Statistics genutzt. Alle nachfolgenden Grafiken wurden mit dieser Software erstellt.

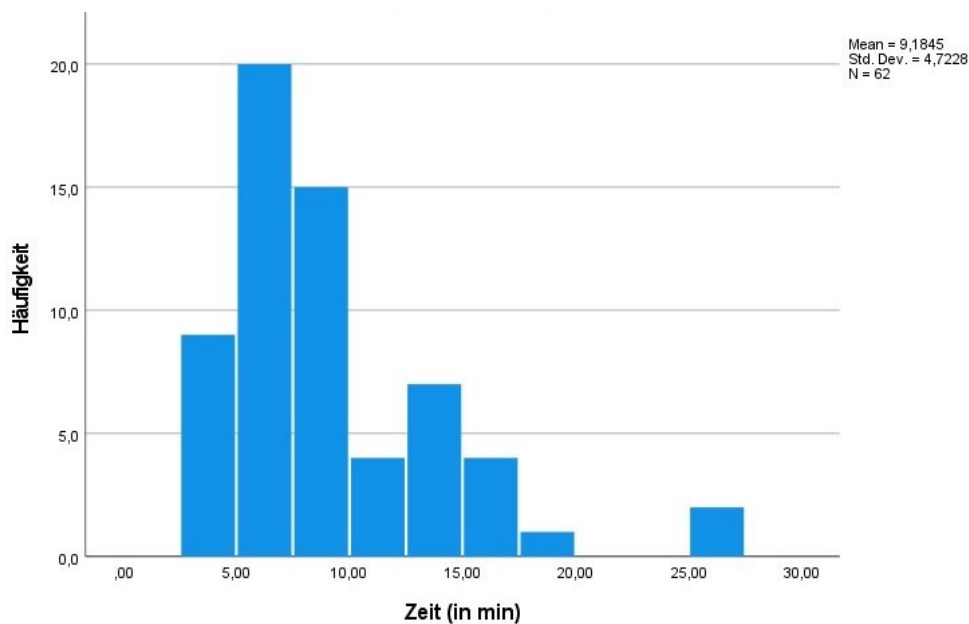


Abbildung 6.6: Histogramm für die benötigte Studienzeit

Im Schnitt benötigten die Probanden ca 9 Minuten für die gesamte Studie. In Abbildung 6.6 sieht man die Verteilung der gemessenen Zeiten.

In Abbildung 6.7 sind die verschiedenen Boxplots der Zeiten abgebildet. Während man hier schon erkennen kann, dass die Mediane der gemessenen Zeiten bei allen Versuchsbedingungen recht ähnlich sind, ist auch auf den ersten Blick ersichtlich, dass es genau zwei Ausreißer gibt. Diese gehören zur Bedingung *Visualisierung* und benötigten jeweils 27 und 25 Minuten für die gesamte Studie.

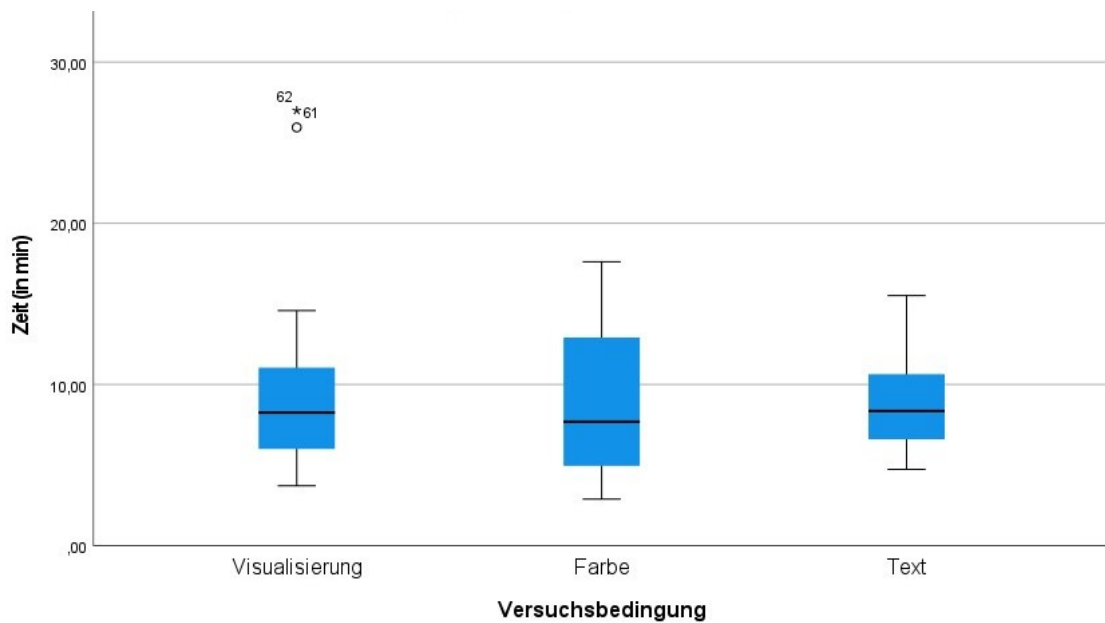


Abbildung 6.7: Einfache Boxplots der benötigten Studienzeiten

Um entscheiden zu können, ob die Ergebnisse der Ausreißer von der Auswertung ausgeschlossen werden sollen, wurde ein näherer Blick auf die Teilzeiten der zwei Probanden geworfen. Proband 1 verbrachte 17 Minuten auf der Startseite und Proband 2 17 Minuten auf der Datenschutz-Seite. Für den Rest benötigten beide durchschnittlich lange. Aus diesem Grund wurden die Teilnehmer nicht ausgeschlossen.

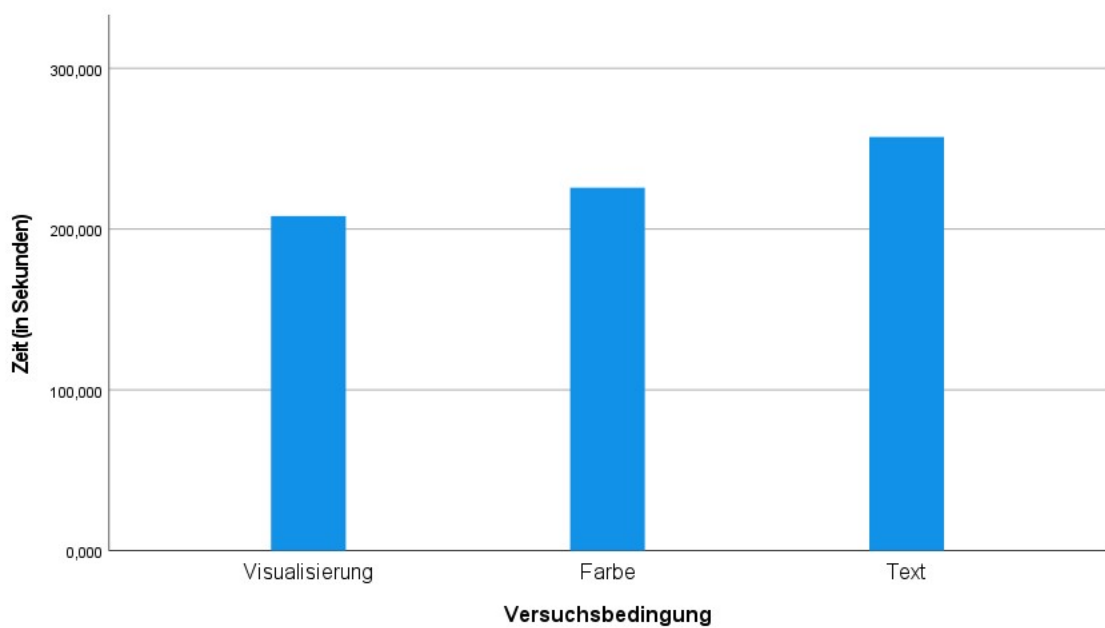


Abbildung 6.8: Säulendiagramm für die benötigte Aufgabenzeit

Report

timeTaskinsec	Durchschnitt	Anzahl	Standardabweichung
Versuchsbedingung			
Visualisierung	208,12332	19	67,694973
Farbe	225,74567	18	143,567310
Text	257,33736	25	108,615911
Gesamt	233,08385	62	110,404200

Abbildung 6.9: Durchschnittswert und Standardabweichung für einzelne Versuchsbedingungen (benötigte Aufgabenzeit)

Abbildung 6.8 zeigt die Zeit, die die einzelnen Bedingungen im Durchschnitt für die Aufgabe benötigten. Die Darstellung mit der Visualisierung schneidet hier unter den Teilnehmern besser ab als die Bedingungen *Farbe* und *Text*.

Um Aussagen über die Signifikanz der Unterschiede treffen zu können, muss ein Test durchgeführt werden. Hierfür muss aber zunächst geklärt werden, ob es sich bei der Verteilung der Aufgabenzeiten um eine Normalverteilung handelt.

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Studienzeit	,158	66	,000	,858	66	,000

a. Lilliefors Significance Correction

Abbildung 6.10: Test für Normalverteilung

Die Ergebnisse aus Abbildung 6.10 zeigen, dass dies nicht der Fall ist. Daher wurde entschieden, den nicht-parametrischen Kruskal-Wallis Test durchzuführen.

Hypothesis Test Summary				
	Nullhypothese	Test	Sig. ^{a,b}	Entscheidung
1	Die Verteilung von Studienzeit ist über die Kategorien von Versuchsbedingung identisch.	Kruskal-Wallis-Test bei unabhängigen Stichproben	,910	Nullhypothese beibehalten.

a. Das Signifikanzniveau

b. Asymptotische Signifikanz werden angezeigt.

Abbildung 6.11: Ergebnisse des Kruskal-Wallis Test

Da der Signifikanzwert des Kruskal-Wallis Tests über dem typischen Alphaswert von 0.05 liegt, wird die Nullhypothese beibehalten. Dies bedeutet, dass bei den gemessenen Zeiten keine signifikanten Unterschiede zwischen den einzelnen Versuchsbedingungen erkennbar sind.

Im Folgenden soll nun auf die Beantwortung der Aufgaben eingegangen werden.

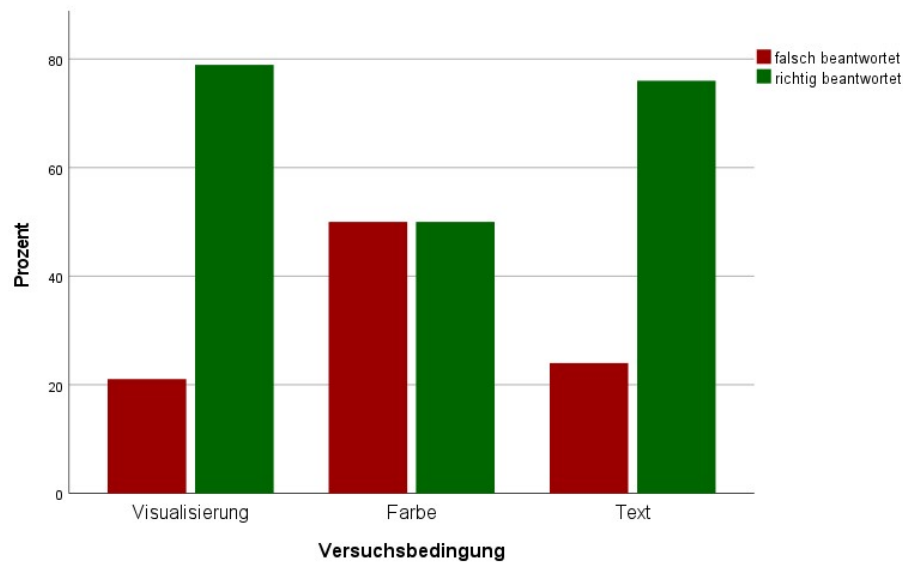


Abbildung 6.12: Antworten auf Frage 1

Bei der ersten Aufgabe sollten die Probanden angeben, um wie viele Themen es im angezeigten Twitter Feed geht. Abbildung 6.12 zeigt, dass *Farbe* die höchste Fehlerquote erzielte mit 50%, während die anderen Bedingungen in etwa ähnliche Werte erreichten (*Visualisierung* 21% Fehlerquote, *Text* 24% Fehlerquote).

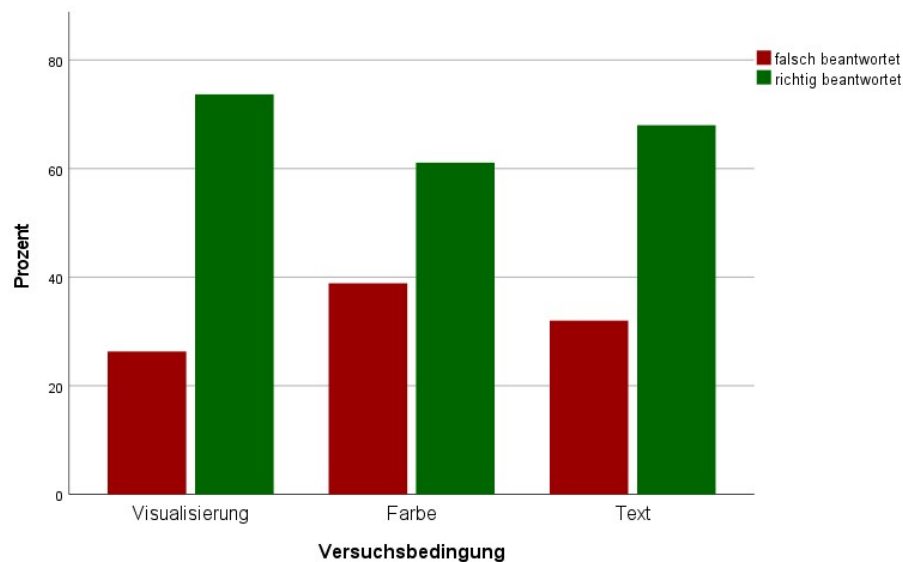


Abbildung 6.13: Antworten auf Frage 2

In der zweiten Aufgabe ging es um den prozentualen Anteil des größten Themas. Auch hier schnitt *Farbe* am schlechtesten ab mit einer Fehlerquote von 39%. Die Bedingung *Visualisierung* erzielte den besten Wert mit einer Fehlerquote von 26% und *Text* erreichte eine Fehlerquote von 32%.

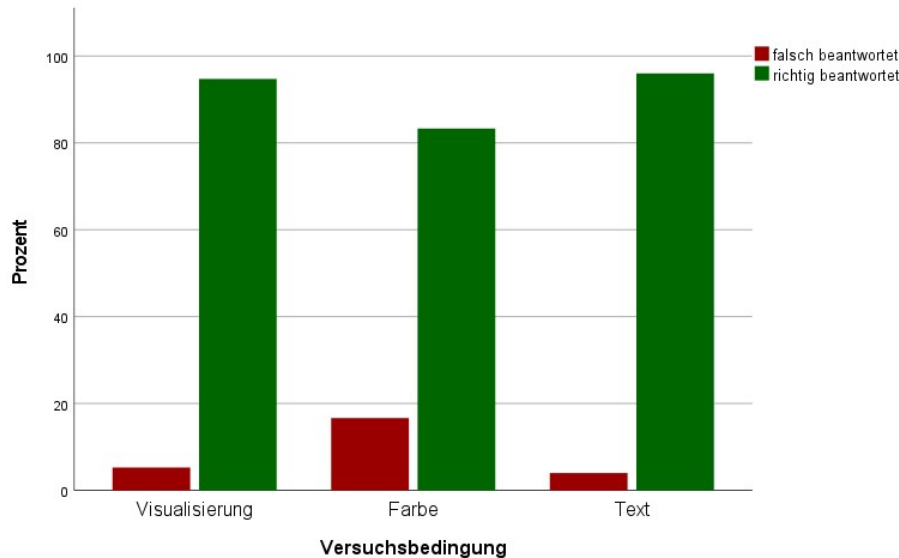


Abbildung 6.14: Antworten auf Frage 3

Bei der Benennung des Themas von Cluster 2 erzielten alle Versuchsbedingungen recht gute Ergebnisse. *Text* hatte die geringste Fehlerquote mit 4%, dicht gefolgt von Bedingung *Visualisierung* mit einer Fehlerquote von 5%. *Text* erzielte eine Fehlerquote von 17%.

Um Aussagen darüber treffen zu können, ob diese unterschiedlichen Fehlerquoten signifikant sind, wurde ein Chi-Quadrat Test bei allen Fragen angewandt. Der Test wurde jeweils einmal für alle Versuchsbedingungen durchgeführt und dann paarweise mit zwei Bedingungen.

Abbildung 6.15 bis Abbildung 6.18 zeigen die Ergebnisse der Tests für die erste Frage. Weder bei der ersten Frage noch bei den anderen zwei Fragen war der Wert der asymptotischen Signifikanz unter 0.05. Sowohl der Chi-Quadrat Test als auch der Fisher's Exakt Test zeigten, dass kein signifikanter Zusammenhang zwischen den Fehlerquoten und den jeweiligen Versuchsbedingungen besteht.

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)
Pearson-Chi-Quadrat	4,515 ^a	2	,105
Likelihood-Quotient	4,349	2	,114
Zusammenhang linear-mit-linear	,003	1	,958
Anzahl der gültigen Fälle	62		

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5.
Die minimale erwartete Häufigkeit ist 5,52.

Abbildung 6.15: Chi-Quadrat Test für alle Versuchsbedingungen (nur für Frage 1)

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Sig. (zweiseitig)	Exakte Sig. (einseitig)
Pearson-Chi-Quadrat	3,399 ^a	1	,065		
Kontinuitätskorrektur ^b	2,247	1	,134		
Likelihood-Quotient	3,463	1	,063		
Exakter Test nach Fisher				,091	,066
Zusammenhang linear-mit-linear	3,307	1	,069		
Anzahl der gültigen Fälle	37				

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 6,32.

b. Wird nur für eine 2x2-Tabelle berechnet

Abbildung 6.16: Chi-Quadrat Test für die Versuchsbedingungen *Visualisierung* und *Farbe* (nur für Frage 1)

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Sig. (zweiseitig)	Exakte Sig. (einseitig)
Pearson-Chi-Quadrat	,053 ^a	1	,817		
Kontinuitätskorrektur ^b	,000	1	1,000		
Likelihood-Quotient	,054	1	,817		
Exakter Test nach Fisher				1,000	,556
Zusammenhang linear-mit-linear	,052	1	,819		
Anzahl der gültigen Fälle	44				

a. 1 Zellen (25,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 4,32.

b. Wird nur für eine 2x2-Tabelle berechnet

Abbildung 6.17: Chi-Quadrat Test für die Versuchsbedingungen *Visualisierung* und *Text* (nur für Frage 1)

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)	Exakte Sig. (zweiseitig)	Exakte Sig. (einseitig)
Pearson-Chi-Quadrat	3,114 ^a	1	,078		
Kontinuitätskorrektur ^b	2,075	1	,150		
Likelihood-Quotient	3,111	1	,078		
Exakter Test nach Fisher				,109	,075
Zusammenhang linear-mit-linear	3,042	1	,081		
Anzahl der gültigen Fälle	43				

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 6,28.

b. Wird nur für eine 2x2-Tabelle berechnet

Abbildung 6.18: Chi-Quadrat Test für die Versuchsbedingungen *Farbe* und *Text* (nur für Frage 1)

Bei der letzten Frage sollten die Probanden die zwei Cluster bestimmen, die sich von den Themen her am meisten widersprechen. Da diese Aufgabe zum Teil subjektiv zu lösen ist, kann hier nicht von Fehlerquoten gesprochen werden. Wenn man sich aber die Antworten in Abbildung 6.19 anschaut, sieht man, dass bei *Visualisierung* zwei Antwortmöglichkeiten besonders oft gewählt wurden, während es bei den anderen Bedingungen mehr als zwei sind.

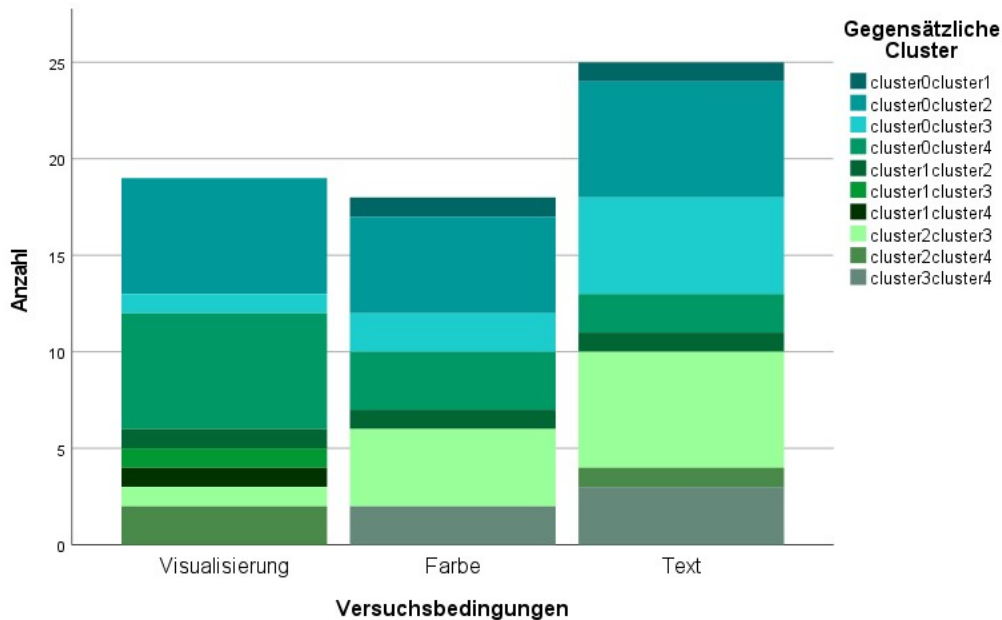


Abbildung 6.19: Antworten auf Frage 4

Frage	Visualisierung	Farbe	Text	Gesamt
Frage 1	2.11	2.61	2.16	2.27
	<i>1.15</i>	<i>1.46</i>	<i>1.07</i>	<i>1.22</i>
Frage 2	2.42	2.78	2.92	2.73
	<i>0.96</i>	<i>1.17</i>	<i>1.00</i>	<i>1.04</i>
Frage 3	2.16	2.00	1.72	1.94
	<i>1.07</i>	<i>1.09</i>	<i>0.84</i>	<i>0.99</i>
Frage 4	3.58	3.78	3.48	3.60
	<i>1.22</i>	<i>1.26</i>	<i>1.12</i>	<i>1.18</i>

Tabelle 6.1: Angegebene Schwierigkeitsgrade der einzelnen Fragen auf einer Skala von 1 bis 5. Die durchschnittlichen Werte sind in Druckschrift angegeben und die Standardnormalabweichung in kursiver Schrift. Alle Werte wurden auf zwei Nachkommastellen gerundet.

Frage	Visualisierung	Farbe	Text	Gesamt
Wie sicher waren Sie sich bei der Beantwortung der Fragen?	3.26	3.61	3.60	3.50
	<i>0.99</i>	<i>0.78</i>	<i>0.87</i>	<i>0.88</i>
Wie zufrieden waren Sie mit der Darstellung der unterschiedlichen Themen?	3.26	3.89	3.48	3.53
	<i>1.04</i>	<i>1.13</i>	<i>0.96</i>	<i>1.05</i>

Tabelle 6.2: Ergebnisse des letzten Fragenkatalogs. Bewertungen wurden auf einer Skala von 1 bis 5 abgegeben. Die durchschnittlichen Werte sind in Druckschrift angegeben und die Standardnormalabweichung in kursiver Schrift. Alle Werte wurden auf zwei Nachkommastellen gerundet.

Nach jeder der Fragen sollten die Probanden angeben, wie schwer sie die Beantwortung der Frage fanden. Die Einschätzung wurde auf einer Skala von 1 bis 5 angegeben, wobei 1 für *nicht schwierig* und 5 für *sehr schwierig* steht. Die Ergebnisse wurden in Tabelle 6.1 zusammengetragen. Bei den ersten zwei Fragen schneidet die Visualisierung am besten ab, bei Frage 3 am schlechtesten und bei der letzten Frage ungefähr durchschnittlich. Insgesamt fanden die Teilnehmer die Beantwortung von Frage 3 am einfachsten und die Beantwortung von Frage 4 am schwierigsten.

Nach der Hauptaufgabe sollten die Probanden zwei letzte Fragen beantworten. Auch hier sollten sie ihre Einschätzung auf einer Skala von 1 bis 5 geben, wobei 1 für *sehr unsicher* bzw. *sehr unzufrieden* steht und 5 für *sehr sicher* bzw. *sehr zufrieden*. Tabelle 6.2 zeigt die Ergebnisse der Auswertung. Bei beiden schnitt *Farbe* am besten ab und *Visualisierung* am schlechtesten.

Um zu untersuchen, ob die Ergebnisse statistisch signifikant sind, wurde der nicht-parametrische Mann-Whitney-U-Test auf jeweils zwei Bedingungen angewandt (siehe Abbildungen 6.20 bis 6.22). In keinem der Fälle war der Wert in den Zeilen *Asymp. Sig* bzw. *Exact Sig* kleiner als 5%. Damit wurde gezeigt, dass kein signifikanter Unterschied zwischen den einzelnen Versuchsbedingungen besteht.

Teststatistiken^a

	Schwierigkeit Frage 1	Schwierigkeit Frage 2	Schwierigkeit Frage 3	Schwierigkeit Frage 4	Sicher bei Beantwortung	Zufrieden mit Darstellung
Mann-Whitney-U-Test	140,000	139,500	154,000	156,000	129,500	110,500
Wilcoxon-W	330,000	329,500	344,000	327,000	319,500	300,500
Z	-,979	-,993	-,542	-,480	-,1338	-,1,910
Asymp. Sig. (2-seitig)	,328	,321	,588	,631	,181	,056
Exakte Sig. [2*(1-seitige Sig.)]	,358 ^b	,343 ^b	,620 ^b	,663 ^b	,210 ^b	,066 ^b
Exakte Sig. (2-seitig)	,326	,340	,604	,643	,196	,056
Exakte Sig. (1-seitig)	,163	,167	,300	,332	,095	,027
Punkt-Wahrscheinlichkeit	,005	,020	,006	,033	,007	,003

a. Gruppenvariable: condition

b. Nicht für Bindungen korrigiert.

Abbildung 6.20: Mann-Whitney-U-Test für die Versuchsbedingungen *Visualisierung* und *Farbe*

Teststatistiken^a

	Schwierigkeit Frage 1	Schwierigkeit Frage 2	Schwierigkeit Frage 3	Schwierigkeit Frage 4	Sicher bei Beantwortung	Zufrieden mit Darstellung
Mann-Whitney-U-Test	228,000	171,000	220,000	182,500	185,500	213,500
Wilcoxon-W	418,000	361,000	545,000	507,500	375,500	403,500
Z	-,235	-,1,644	-,432	-,1,389	-,1,304	-,595
Asymp. Sig. (2-seitig)	,814	,100	,666	,165	,192	,552
Exakte Sig. (2-seitig)	,825	,100	,683	,184	,198	,564
Exakte Sig. (1-seitig)	,417	,053	,346	,092	,101	,282
Punkt-Wahrscheinlichkeit	,004	,007	,005	,012	,010	,013

a. Gruppenvariable: condition

Abbildung 6.21: Mann-Whitney-U-Test für die Versuchsbedingungen *Visualisierung* und *Text*

Teststatistiken^a

	Schwierigkeit Frage 1	Schwierigkeit Frage 2	Schwierigkeit Frage 3	Schwierigkeit Frage 4	Sicher bei Beantwortung	Zufrieden mit Darstellung
Mann-Whitney-U-Test	187,500	209,500	184,000	198,000	222,000	165,500
Wilcoxon-W	512,500	380,500	509,000	523,000	547,000	490,500
Z	-,956	-,397	-,1,046	-,716	-,079	-,1,525
Asymp. Sig. (2-seitig)	,339	,691	,296	,474	,937	,127
Exakte Sig. (2-seitig)	,345	,703	,303	,482	,949	,129
Exakte Sig. (1-seitig)	,174	,351	,153	,241	,490	,063
Punkt-Wahrscheinlichkeit	,009	,022	,004	,001	,035	,004

a. Gruppenvariable: condition

Abbildung 6.22: Mann-Whitney-U-Test für die Versuchsbedingungen *Farbe* und *Text*

7 Diskussion

Das Ziel der Online-Studie war es, herauszufinden, ob und inwieweit integrierte wortgroße Grafiken das Leseverhalten der Nutzer beeinflussen.

Die Ergebnisse zeigten, dass die Teilnehmer der Bedingung *Visualisierung* weniger Zeit für die Durchführung der Aufgabe benötigten. Der Unterschied zwischen den gemessenen Zeiten war jedoch nicht signifikant genug, um Aussagen darüber treffen zu können, ob die Nutzung von integrierten Grafiken die Lesegeschwindigkeit erhöht bzw. die Bearbeitungszeit reduziert.

Bei den Fragen „Wie viele Themen gibt es?“ und „Wie groß ist das größte Thema?“ schnitten die Teilnehmer der Bedingung *Visualisierung* am besten ab, dicht gefolgt von den Teilnehmern der Bedingung *Text*. Die Frage „Um welches Thema geht es in Cluster 2?“ erzielte bei allen Teilnehmern gute Ergebnisse. Jedoch schnitt *Text* hier etwas besser ab als die Bedingung *Visualisierung*. *Farbe* erzielte bei jeder der drei Fragen die schlechtesten Ergebnisse. Wie auch bei den vorher gemessenen Zeiten konnte bei den Fehlerquoten kein signifikanter Unterschied festgestellt werden.

Die Ergebnisse aus der Abbildung 6.19 zeigten, dass sich die Teilnehmer der Bedingung *Visualisierung* einiger waren, wenn es darum ging, gegensätzliche Themen festzustellen. Während die anderen Probanden größtenteils drei Möglichkeiten angaben, waren es bei der *Visualisierung* nur zwei.

Insgesamt konnten unsere Annahmen, dass die integrierten Visualisierungen zu einer kürzeren Bearbeitungsdauer und geringeren Fehlerquoten führen, nicht bestätigt werden.

Die Ergebnisse können verschiedene Gründe haben. Einer der Gründe kann sein, dass die Probanden nicht genug Zeit hatten, sich mit der Funktionalität der Visualisierung auseinanderzusetzen. Menschen sind normalerweise nicht täglich mit Visualisierungen konfrontiert und benötigen daher eine bestimmte Einarbeitungszeit. Auch führt die Vielfalt der Visualisierungstypen und Funktionen dazu, dass sich Nutzer nicht immer auf bisheriges Wissen verlassen können. Die Bedingung *Text* baut dagegen auf bekanntes Wissen auf: Wir sind jeden Tag rein textuellen Informationen ausgesetzt und sind geübt darin, diese zu verarbeiten. Schon von klein auf wurde unser Leseverständnis entwickelt und verbessert. Durch diese Übung fällt es uns einfacher Informationen aus Texten zu entnehmen und auszuwerten.

Eine Erklärung dafür, warum *Farbe* im Vergleich zu den anderen Bedingungen so schlecht abschneidet, ist die folgende: Da hier die Tweets nur unterschiedlich eingefärbt waren, enthielt diese Darstellungsform einerseits nicht so viele Informationen wie die Bedingung *Visualisierung* und war andererseits weniger direkt als der textuelle Themenvermerk in der Bedingung *Text*.

Ein anderer Grund für die Ergebnisse kann der unterschiedliche Zeitaufwand zwischen den verschiedenen Versuchsbedingungen sein. Da die Teilnehmer der Bedingung *Text* im Schnitt etwa 30 bzw. 50 Sekunden länger mit der Aufgabe verbrachten als Teilnehmer der Bedingung *Farbe*

oder *Visualisierung*, kann man das Argument anführen, dass sie sich genauer mit den Inhalten auseinandergesetzt haben. Die sorgfältigere Bearbeitung der Aufgaben könnte zum Teil der Grund für die guten Ergebnisse der Versuchsbedingung *Text* gewesen sein.

Nichtsdestotrotz zeigten die Ergebnisse, dass es Verbesserungsmöglichkeiten für die Visualisierung gibt. So ist es vorteilhafter, wenn die Visualisierung statt der Anzahl der Tweets im jeweiligen Cluster eher den prozentualen Anteil angibt. Dies unterstützt die Lesenden dabei, sich ein besseres Bild über die Größenverhältnisse zu machen, ohne Rechenoperationen durchführen zu müssen.

Was bei der Implementierung der Visualisierung aber auch schnell klar wurde, war dass die theoretischen Vorstellungen nicht immer der Realität entsprachen. Gerade wenn es nämlich um kontroverse Themen ging, erhofften wir uns klare Abgrenzungen zwischen den diskutierten Meinungen. Was wir jedoch vorfanden, war etwas anderes: Viele Kommentare wurden in verschiedene Cluster bzw. Themenbereiche unterteilt, da sie unterschiedliche Phrasen benutzten, um die gleiche Meinung wiederzugeben. Tweets beinhalteten zum Teil mehrere Themen und konnten nicht wirklich einem einzigen Cluster zugeordnet werden. Sarkastische Botschaften enthielten teilweise Satzstrukturen gegensätzlicher Meinungen und wurden deshalb dem gleichen Cluster zugeordnet. Insgesamt verschwammen die Grenzen zwischen den einzelnen Themen und waren meist nicht so „eindeutig“ wie anfangs erwartet.

8 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde eine Visualisierung entwickelt, die Twitter Nutzern dabei helfen soll, sich einen Überblick über diskutierte Themen zu verschaffen. Die Visualisierung kann einzelne Datensätze veranschaulichen und zeigt den Zusammenhang zwischen unübersichtlichen Datenmengen auf. Durch sie erfahren die Nutzer, wie groß ein Twitter Feed ist, wie viele Themen in diesem diskutiert werden, wie groß die einzelnen Themen sind und welche Phrasen für die Themen charakteristisch sind. Außerdem wissen die Nutzer bei jedem Tweet, zu welchem Thema dieser gehört.

Nach der Erläuterung wichtiger grundlegender Begriffe und Theorien, widmete sich diese Arbeit verwandten wissenschaftlichen Texten und legte dabei einen Fokus auf das Design und die Integration von wortgroßen Visualisierungen. Aufbauend auf diesem Wissen wurde eine Visualisierung entwickelt, die sich in Tweets integrieren lässt. Im Verlauf dieser Arbeit wurde auf das Design und die Funktionalität der Visualisierung eingegangen, aber auch auf die verwendeten Programmiersprachen und Bibliotheken sowie auf den Datensatz, der ihr zugrunde liegt.

Um die Visualisierung zu testen, wurde eine Online-Studie durchgeführt. Für die Studie wurde das „Between-Groups“ Design genutzt, welches in unserem Fall drei verschiedene Versuchsbedingungen testete: In der ersten Darstellung wurde die Visualisierung in die Tweets eingebunden, in der zweiten waren die Tweets je nach Thema unterschiedlich eingefärbt und in der dritten wurde das jeweilige Thema bei jedem Tweet textuell eingefügt. Die Studie untersuchte dabei, ob Nutzer der Visualisierung Themenfragen besser und schneller beantworten können als Nutzer der anderen Darstellungsformen. Die Ergebnisse dieser Studie zeigten jedoch, dass die gemessenen Unterschiede zwischen den Versuchsbedingungen nicht signifikant genug waren, um Aussagen diesbezüglich treffen zu können.

Ausblick

Die Ergebnisse zeigen, dass die Visualisierung überarbeitet werden muss, um das Leseverhalten der Nutzer nachweislich zu verbessern. Zum einen sollte die Visualisierung statt der absoluten Anzahl der Tweets in jedem Cluster den prozentualen Anteil angeben, damit die Nutzer sich ein besseres Bild über die Größenverhältnisse machen können. Zum anderen muss ein Weg gefunden werden, das Clusteringverfahren so zu verbessern, dass die Tweets semantisch besser gruppiert werden. Darüber hinaus sollte gerade bei sehr kleinen Clustern darauf geachtet werden, dass sie in so einer Weise dargestellt werden, dass die Nutzer sie nicht übersehen.

Außerdem könnte man die Visualisierung so erweitern, dass die Nutzer mehr Möglichkeiten haben, mit ihr zu interagieren. Auf diese Weise kann man den Fokus auf das Interesse der Anwender setzen. Sie könnten beispielsweise entscheiden, ob die Visualisierung bei jedem Tweet erscheint oder nur bei dem Tweet, den sie gerade lesen. Auch könnten sie zwischen einer thematisch gruppierten

Darstellungsform und einer Darstellung des zeitlichen Verlaufs entscheiden. Diese würde nämlich zeigen, auf welche Weise sich das Thema verändert hat. Zuletzt sollte auch eine größere Farbauswahl zur Verfügung stehen. So hätten Nutzer die Möglichkeit, je nach Sehschwäche oder Vorliebe zwischen verschiedenen Farbdarstellungen zu wählen.

Literaturverzeichnis

- [Bar86] A. J. Barret. „The Elements of Graphing Data. W. S. Cleveland. Wadsworth Advanced Books and Software, Monterey, CA. 1985. 323 pp. Illustrated. 18.95.“. In: *The Aeronautical Journal (1968)* 90.893 (1986), S. 112–112. DOI: [10.1017/S0001924000015475](https://doi.org/10.1017/S0001924000015475) (zitiert auf S. 21).
- [BS12] C. Baadte, W. Schnotz. „Das Verstehen von Texten mit Bildern“. In: *Weiterbildung, 2012*, 6, 35-37 (Nov. 2012) (zitiert auf S. 13).
- [BW17] F. Beck, D. Weiskopf. „Word-Sized Graphics for Scientific Texts“. In: *IEEE Transactions on Visualization and Computer Graphics* 23.6 (2017), S. 1576–1587. DOI: [10.1109/TVCG.2017.2674958](https://doi.org/10.1109/TVCG.2017.2674958) (zitiert auf S. 11, 16).
- [CMS99] S. K. Card, J. D. Mackinlay, B. Shneiderman, Hrsg. *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999. ISBN: 1558605339 (zitiert auf S. 12).
- [ED16] O. Eltayeb, W. Dou. „A Survey on Interaction Log Analysis for Evaluating Exploratory Visualizations“. In: (Okt. 2016), S. 62–69. DOI: [10.1145/2993901.2993912](https://doi.org/10.1145/2993901.2993912) (zitiert auf S. 14).
- [GBWI17] P. Goffin, J. Boy, W. Willett, P. Isenberg. „An Exploratory Study of Word-Scale Graphics in Data-Rich Text Documents“. In: *IEEE Transactions on Visualization and Computer Graphics* 23.10 (2017), S. 2275–2287. DOI: [10.1109/TVCG.2016.2618797](https://doi.org/10.1109/TVCG.2016.2618797) (zitiert auf S. 11, 15).
- [GWBI15] P. Goffin, W. Willett, A. Bezerianos, P. Isenberg. „Exploring the Effect of Word-Scale Visualizations on Reading Behavior“. In: (Apr. 2015). DOI: [10.1145/2702613.2732778](https://doi.org/10.1145/2702613.2732778) (zitiert auf S. 17).
- [KYY12] S.-H. Kim, H. Yun, J. S. Yi. „How to filter out random clickers in a crowdsourcing-based study?“ In: *ACM International Conference Proceeding Series* (Okt. 2012). DOI: [10.1145/2442576.2442591](https://doi.org/10.1145/2442576.2442591) (zitiert auf S. 14).
- [May09] R. E. Mayer. *Multimedia Learning*. 2nd. USA: Cambridge University Press, 2009. ISBN: 0521514126 (zitiert auf S. 12, 13).
- [PD10] B. Preim, R. Dachsel. *Interaktive Systeme: Band 1 Grundlagen, Graphical User Interfaces, Informationsvisualisierung*. ger. eXamen.press. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2010. ISBN: 9783642054020 (zitiert auf S. 12).
- [Sch10] A. Schüler. „Interferieren (geschriebene) räumliche Textinhalte mit der Bildverarbeitung im visuell-räumlichen Arbeitsgedächtnis? Testung einer erweiterten kognitiven Theorie multimedialen Lernens“. Diss. Universität Tübingen, 2010. URL: <http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-52830> (zitiert auf S. 13).
- [Tat15] M. Tatzgern. „Situating Visualization in Augmented Reality“. Diss. Juni 2015 (zitiert auf S. 12).

[Tuf06] E. R. Tufte. *Beautiful Evidence*. Graphics Press, 2006. ISBN: 978-0-9613921-7-8 (zitiert auf S. 11).

Alle URLs wurden zuletzt am 18. 11. 2020 geprüft.

Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Ort, Datum, Unterschrift