



Universität Regensburg

Entwicklung und Evaluation eines Tools zur lexikonbasierten Sentiment Analysis für die Digital Humanities

Bachelorarbeit im Fach Medieninformatik am
Institut für Information und Medien, Sprache und Kultur (I:IMSK)

Vorgelegt von: Johanna Dangel
Adresse:
E-Mail (Universität):
E-Mail (privat):
Matrikelnummer:
Erstgutachter: Prof. Dr. Wolff
Zweitgutachter: Dr. Raphael Wimmer
Betreuer: Thomas Schmidt
Laufendes Semester: WS 2019/20
Abgegeben am: 26.03.2020

Inhalt

1	Zusammenfassung/Abstract.....	7
2	Einleitung	8
3	Related Work.....	9
3.1	Literaturrecherche	9
3.2	Text Mining.....	10
3.2.1	Definitionen und Begriffe	11
3.2.2	Untergebiete des Text Mining.....	12
3.3	Tools in den Digital Humanities	13
3.3.1	Etablierung der Softwarewerkzeuge	13
3.3.2	Maßnahmen.....	14
3.3.3	Beispiele an Tools in den Digital Humanities	15
3.4	Sentiment Analysis.....	16
3.4.1	Definitionen und Begriffe	17
3.4.2	Aufgaben der Sentiment Analysis	19
3.4.3	Analyseebenen	19
3.4.4	Subjektivität-Klassifizierung und Sentiment-Klassifizierung	20
3.4.5	Meinungszusammenfassung	22
3.4.6	Probleme von Sentiment Analysis.....	22
3.4.7	Sentiment Analysis in den Geisteswissenschaften	23
3.5	Lexikonbasierte Sentiment Analysis.....	25
3.5.1	Vorgehensweise der lexikonbasierten Sentiment Analysis	26
3.5.2	Lexika.....	26
3.5.3	Valence shifters	29
3.6	User-Centered Design.....	30
4	Arbeitsagenda	32
5	Anforderungsanalyse.....	33
5.1	Anforderungserhebung	33
5.1.1	Interviews	34
5.1.2	Konkurrenzanalyse.....	39
5.2	Anforderungsanalyse.....	48
5.2.1	Import der zu analysierend Texte.....	48

5.2.2	Export der Ergebnisse.....	49
5.2.3	Natural language processing	49
5.2.4	Lexika.....	50
5.2.5	Sentiment Analysis, deren Ergebnisse und Nachvollziehbarkeit .	51
5.2.6	Usability.....	52
5.3	Anforderungsspezifizierung	53
6	Entwurf und Implementierung der Software.....	57
6.1	Lexikonbasierter Ansatz und Schema des Lexikons	57
6.2	Ermitteln der Testdateien für die Implementierung	58
6.3	Implementierung und technischer Hintergrund	59
6.3.1	Wahl des Webframeworks und Aufbau.....	59
6.3.2	Back-End: Sentiment Analysis	60
6.3.3	Front-End.....	66
6.3.4	Einbinden weiterer Sprachen zur Sentiment Analysis.....	67
6.4	Designprozess und Funktionalitäten.....	68
6.4.1	Erster Designprozess	68
6.4.2	Formative Evaluation	74
6.4.3	Summative Evaluation	85
6.4.4	Funktionalität	93
7	Diskussion.....	107
8	Limitations	108
9	Future Work	109
10	Ausblick.....	109
	Literaturverzeichnis	110
	Elektronische Daten	119
	Erklärung zur Urheberschaft.....	120
	Erklärung zur Lizenzierung und Publikation dieser Arbeit.....	121

Abbildungsverzeichnis

Abbildung 1: Ausschnitt SentiWS.....	27
Abbildung 2: Ausschnitt GermanPolarityClues	28
Abbildung 3: Ausschnitt Berlin Affective Word List - Reloaded	29
Abbildung 4: Screenshot der Anwendung SentiStrength	41
Abbildung 5: Screenshot der Anwendung SentiStrength - Ergebnis	42
Abbildung 6: Screenshot der Anwendung Sentiment Analyzer	43
Abbildung 7: Screenshot der Anwendung Stanford Sentiment Analysis	44
Abbildung 8: Screenshot der Anwendung Stanford Sentiment Analysis - Ergebnis	45
Abbildung 9: Lingmotif – Eingabe (Moreno-Ortiz, 2017, p. 74)	46
Abbildung 10: Lingmotif – Ergebnis (Moreno-Ortiz, 2017, p. 75)	46
Abbildung 11: Screenshot der Anwendung Voyant - Eingabe.....	47
Abbildung 12: Screenshot der Anwendung Voyant - Ergebnis.....	47
Abbildung 13: Taskanalyse eines Nutzers	54
Abbildung 14: Use-Case-Diagramm - Julia.....	55
Abbildung 15: Use-Case-Diagramm - Michael	56
Abbildung 16: Aufruf von Model und View im Controller.....	66
Abbildung 17: Medium-Fidelity Prototypen - Home	69
Abbildung 18: Medium-Fidelity Prototypen - Upload.....	70
Abbildung 19: Medium-Fidelity Prototypen - Result-Page	70
Abbildung 20: High-Fidelity Prototypen - Information-Page.....	71
Abbildung 21: High-Fidelity Prototypen - Upload-Page	72
Abbildung 22: High-Fidelity Prototypen - Result-Page	72
Abbildung 23: Neues Layout - Upload-Page	73
Abbildung 24: Neues Layout - Result-Page mit normalisierten Scores	73
Abbildung 25: Information-Page	74
Abbildung 26: Time per Task.....	83
Abbildung 27: Design der klickbaren Panels (Issue Nr. 12) - Davor.....	84
Abbildung 28: Design der klickbaren Panels (Issue Nr. 12) – Danach.....	84
Abbildung 29: Aufgabenstellung von Task 5.....	87

Abbildung 30: Benötigte Zeit pro Aufgabe aller Nutzer.....	88
Abbildung 31: Vorgefundene Fehler der Nutzer pro Aufgabe	88
Abbildung 32: Erfolgsrate pro Aufgabe und Nutzer.....	89
Abbildung 33: Wordcloud mit häufigen Wörtern.....	90
Abbildung 34: Charakterbereiche der Anwendung.....	91
Abbildung 35: Mittleren Ausprägungen der Dimensionen.....	91
Abbildung 36: SUS.....	91
Abbildung 37: About-Page.....	94
Abbildung 38: Documentation-Page	94
Abbildung 39: Tooltip.....	95
Abbildung 40: Note.....	95
Abbildung 41: Note für BAWL-R	95
Abbildung 42: Upload-Page mit erweiterten Optionen.....	95
Abbildung 43: Pflichtfeldprüfung	96
Abbildung 44: Falsches Dateiformat für die Textdatei.....	96
Abbildung 45: Result-Page.....	97
Abbildung 46: Dokumente und Ordner	97
Abbildung 47: Close-Reading des Textes „Der Freigeist“	98
Abbildung 48: Überschrift und Legende.....	98
Abbildung 49: SBW ohne Negation.....	99
Abbildung 50: SBW mit Negation.....	99
Abbildung 51: Sentiment bearbeiten	99
Abbildung 52: Polaritätsstärke ist keine Zahl	99
Abbildung 53: Sentiment eines Textausschnittes	99
Abbildung 54: Visualisierung - Document Score	100
Abbildung 55: Visualisierung - Polaritätsgewichtung	101
Abbildung 56: Visualisierung - Verhältnis aller Wörter	101
Abbildung 57: Visualisierung - Stärkste negative SBWs (Wordcloud).....	102
Abbildung 58: Tooltip für das Wort "Angst" in der Wordcloud	102
Abbildung 59: Visualisierung - Stärkste positive SBWs (Balkendiagramm).....	102
Abbildung 60: Visualisierung – Verhältnis der SBSs.....	103

Abbildung 61: Visualisierung - Timeline des Sentiments eines Dokuments.....	104
Abbildung 62: Visualisierung – Sätze mit der größten Polaritätsstärke	104
Abbildung 63: Close-Reading des Ordners „Lessing“	105
Abbildung 64: Ergebnisse herunterladen.....	105
Abbildung 65: Vergleich der Polaritätsgewichtungen	106
Abbildung 66: Beispiele	108

1 Zusammenfassung/Abstract

In dieser Bachelorarbeit wird eine Webanwendung implementiert, mit Hilfe derer der Nutzer eine lexikonbasierte Sentiment Analyse durchführen kann. Im Rahmen des User-Centered-Design Prozess wurde hierbei Experten des Studienggebiet der Digital Humanities und potenzielle Nutzer der Anwendung miteinbezogen. Auf diese Weise wurden Funktionalitäten herausgearbeitet, mit derer der Nutzer die Analyse anpassen und mit dem Tool interagieren kann.

In this bachelor thesis a web application is implemented, which allows the user to perform a lexicon-based sentiment analysis. In the context of the user-centered design process, experts in the field of digital humanities and potential users of the application were involved. In this way, functionalities were worked out with which the user can customise the analysis and interact with the tool.

2 Einleitung

Die vorliegende Arbeit thematisiert die Entwicklung und Evaluation einer Webanwendung zur Sentiment Analyse für Wissenschaftler der Digital Humanities. In der zu entwickelten Anwendung wird hierbei die Gefühlslage eines Dokuments oder mehreren Texten untersucht. Hierbei wird sich der Methode des lexikonbasierten Ansatzes bedient. Dieser Ansatz greift auf ein Lexikon zurück, in dem meinungstragende Wörter mit ihrer semantischen Orientierung und Stärke festgehalten sind (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011, p. 268). Aktuell gibt es kein frei verfügbares online Tool für die Digital Humanities, welches die Anliegen der Wissenschaftler befriedigt (Flüh, 2019). Dazu gehört sowohl die Schwierigkeit der Adaption der zu untersuchenden Domäne als auch, dass sich die Polarität eines Wortes im Laufe der Zeit verändern kann (Flüh, 2019; Schmidt, Burghardt, & Wolff, 2018). Aus diesem Grund beschäftigt sich die vorliegende Arbeit mit der Entwicklung einer Webanwendung, die speziell auf die Bedürfnisse von Forschern in den Digital Humanities eingeht.

Die Abschlussarbeit gliedert sich hier in vier Teile. Zuerst wird im dritten Kapitel die Thematik von Text Mining, dort angesiedelten Anwendungen, die Sentiment Analyse, sowie das User-Centered-Design aufgearbeitet und wichtige Begrifflichkeiten definiert und erläutert. Im Kapitel 4 der Arbeit wird auf die Erhebung der Anforderung des Tools eingegangen. Hierfür wurden Interviews mit Experten und potenzielle Nutzer geführt sowie eine Konkurrenzanalyse angewandt. Zusätzlich wurden Steckholder identifiziert und anhand von Methoden des User-Centered-Design die Anforderungen spezifiziert und in einem Anforderungsdokument festgehalten. Nachdem die Anforderungen erhoben wurden, wurde im fünften Kapitel mit dem Entwurf und der Implementierung der Anwendung begonnen. Dabei wurde im Kontext des User-Centered-Design Prozesses die Anwendung einer formativen und einer summativen Usability Studie unterzogen. Abschließend findet sich eine Diskussion und ein Kapitel zu den Limitierungen der Anwendung. Zuletzt werden Future Work und ein Ausblick des Tools erläutert.

3 Related Work

Im folgenden Kapitel wird das Studienggebiet beschrieben, auf das die Methodik der Anwendung zurückgreift. Zunächst wird auf Text Mining anhand relevanter Literatur eingegangen. Hierbei wird auf die Problematik der wachsenden Anzahl an digitalen Dokumenten und die dadurch entstehenden Probleme betrachtet, Begriffe und Definitionen erläutert, sowie bekannte Methoden in diesem Forschungsfeld thematisiert. Als Nächstes wird die Etablierung von Tools in den Digital Humanities dargelegt und einige Anwendungen aus diesem Themengebiet vorgestellt. Anschließend ermöglicht die Arbeit einen Einblick in die Sentiment Analysis und dem User-Centered-Design Prozess.

3.1 Literaturrecherche

Um passende Literatur für die Abschlussarbeit zu recherchieren, wurden diverse Plattformen und Techniken angewandt. So bietet Google Scholar einen guten Überblick an existierenden wissenschaftlichen Arbeiten. Zudem wurde die ACM Digital Library, der Regensburger Katalog und die AAAI Digital Library verwendet. Darüber hinaus boten Referenzen von Werken Hinweise auf einschlägige Literatur. Sollte ein gesuchtes Werk in den digitalen Bibliotheken oder in der Bibliothek der Universität Regensburg nicht zugänglich gewesen sein, so wurde auf ResearchGate, der Suchmaschine Google und auf die Fernleihe der Universität zurückgegriffen. Zudem wurde als Suchbegriffe sowohl deutsche als auch englischsprachige Begriffe, sowie Synonyme und eine erweiterbare Suche angewandt. Auf diese Weise wurde neben Sentiment Analyse ebenfalls nach Sentiment Analysis, Opinion Mining oder nach der Wortgruppe „Digital Humanities“ gesucht. Da neben dem Auffinden von Paper das Identifizieren von passenden Tools der Digital Humanities und der Sentiment Analysis darstellt, lieferte die Suche „digital humanities tools“ mit der Suchmaschine Google bereits Ergebnisse auf Webseiten, die Applikationen für die Digital Humanities sammeln, vorstellen und ordnen. Dazu gehört die University of Arizona Libra-

ries¹, die Seite digihum.de², forText³ und TAPoR⁴, auf welches die University of Arizona Libraries ebenfalls verweist. Hierbei waren vor allem die University of Arizona Libraries und TAPoR hilfreich, um die passenden Tools ausfindig zu machen. Zudem wurde im Oberseminar auf eine verwandte Anwendung verwiesen. Um geeignete Ressourcen für die Lexika zur Implementierung der Anwendung herauszufinden, wurde neben einschlägiger Literatur auch die Webseite der Interest Group on German Sentiment Analysis ⁵ (IGGSA) hinzugezogen.

3.2 Text Mining

In den letzten Jahren nahm die Digitalisierung immer mehr zu. Hierbei lässt sich ein Zuwachs an wissenschaftlichen Literatur, etwa über das Google Books Projekt, als auch eine wachsende Akzeptanz von kollaborativer Technologien verordnen (Blake, 2011, p. 123). Dies führte dazu, dass die Anzahl von Textdateien und online verfügbaren Dokumenten rasant zunahm (Lemke & Wiedemann, 2016, p. 1; Mehler & Wolff, 2005, p. 1). Dem Information Retrieval, seinen gängigen Prozessen und unsere kognitiven Verarbeitungskapazitäten werden diese große Anzahl an Texten und der damit einhergehenden „Informationsflut“ nicht mehr gerecht, was unter dem Begriff „Informationsüberflutung“ bekannt ist (Blake, 2011, p. 123; Mehler & Wolff, 2005, p. 1) .

Um diese Überflutung an Informationen zu verhindern, werden computer-gestützte Prozesse benötigt, die es ermöglichen, Unmengen an Texten semantisch zu analysieren und, automatisch bzw. semi-automatisch, zu strukturieren (Heyer, Quasthoff, & Wittig, 2006, p. 3). Die hierbei verwendeten Prozesse beinhalten sowohl statistische als auch (computer-) linguistische Methoden (Hippner & Rentzmann, 2006, p. 287), womit, vergleichbar mit dem Bergbau (Heyer et al., 2006, p. 4), neues und potenziell nützliches Wissen aus den unter-

¹ Verfügbar unter <https://libguides.library.arizona.edu/dighumantools> (zuletzt aufgerufen am 10.03.20)

² Verfügbar unter <https://digihum.de/tools/> (zuletzt aufgerufen am 21.03.20)

³ Verfügbar unter <https://fortext.net/tools> (zuletzt aufgerufen am 21.03.20)

⁴ Verfügbar unter <http://tapor.ca/home> (zuletzt aufgerufen am 21.03.20)

⁵ Verfügbar unter <https://sites.google.com/site/iggsahome/downloads> (zuletzt aufgerufen am 21.03.20)

suchten unstrukturierten oder semi-strukturierten Textdokumenten gewonnen wird (Hippner & Rentzmann, 2006, p. 287). Um dieser vagen Beschreibung des Konzepts von Text Mining (Text Mining) sammelt sich ein interdisziplinäres Forschungsgebiet unterschiedlicher Communities, welche sich durch verschiedene Aspekte und eine Namensvielfalt auszeichnen und in Anwendungen aus verschiedenen Bereichen aufgegriffen wird (Aggarwal & Zhai, 2012c, p. 8; Mehler & Wolff, 2005, p. 2). Unter den verwandten Forschungsgebieten sind hierbei Information Retrieval (IR), Natural Language Processing (NLP) und Information Extraction (IE) zu verordnen (Hotho, Nürnberger, & Paaß, 2005, p. 2). Nachfolgend werden die einzelnen Namensgebungen deren Definitionen beschrieben. Außerdem werden Untergebiete des Text Mining aufgezeichnet.

3.2.1 Definitionen und Begriffe

Es gibt diverse Definitionen des Text Mining und Begriffe, unter denen Text Mining zu finden ist. So kann Text Mining als ein im Kontext von Datenbanken, als einen Teilschritt des *Knowledge Discovery in Databases* (KDD) gesehen werden (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Dagegen wird unter *Knowledge Discovery in Text* (KDT) das Auffinden von Wissen in innerhalb eines Text beschrieben (Kodratoff, 1999, p. 5). Eine weitere Benennung des Text Mining ist das *Textual Data Mining* (TDM). Während sich KDD auf Daten beruft, wird beim TDM textuelle Einheiten exploriert und ein mehrstufiger Prozess beschrieben, der neben KDD auch IR, IE und die Präsentation der Daten beinhaltet (Losiewicz, Oard, & Kostoff, 2000). Im Gegensatz zu dem gerade definierten TDM grenzt Hearst (1999) *Text Data Mining* von IR ab. Aber er begründet die Metapher des „mining“ als ungenügend, da das Schürfen auch wertloses Gestein zu Tage fördert, was in Text Mining jedoch nicht der Fall ist (Hearst, 1999, p. 3). Hierbei wird die David Hand, Heikki Mannila und Padhraic Smyth aufgestellte Definition zu Data Mining

“Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.”

genutzt (Hand, Mannila, & Smyth, 2001, Chapter 1). Und die dabei verwendeten Methoden des Data Mining auf die auf unstrukturierte textuelle Einheiten angewandt (Solka, 2008, p. 96). Eine weitere Bezeichnung von Text Mining ist *Text Knowledge Engineering*. Hierbei dienen unstrukturierte natursprachliche Texte als Grundlage, um mithilfe von Wissensmodellierungsprozessen textbasierte Wissenserfassungsmethoden zu generieren (Hahn & Schnattinger, 1998). Neben den oben erläuterten Begriffen sind auch die Begriffe *Knowledge Discovery in Textual Databases* (Feldman & Dagan, 1995), sowie im Kontext des Web die Bezeichnungen *Web Mining*, *Web Content Mining*, *Web Structure Mining* und *Web Usage Mining* (Mehler & Wolff, 2005, p. 7 f.) bekannt.

Anstelle Text Mining auf Methoden einzugrenzen und zu definieren, ordneten Hotho und Kollegen (2005) Text Mining anhand der Perspektive von verwandten Forschungsgebieten in drei Bereiche ein. Zum einen verstehen sie Text Mining als Informationsextraktion, wobei man mithilfe in den wesentlichen Informationen gewinnt (Hotho et al., 2005). Als zweite Gruppe lässt sich Text Mining als Text Data Mining verordnen. Dabei werden mithilfe von Algorithmen und Methoden in Texten nützliche Muster gefunden (Hotho et al., 2005). Zuletzt werden Strömungen, in denen Text Mining in einem Prozess als Teilschritt angewendet wird, in der Gruppe der KDD-Prozesse gesammelt (Hotho et al., 2005).

3.2.2 Untergebiete des Text Mining

Bei der Anwendung von Text Mining können eine Vielzahl von Methoden und Algorithmen Einsatz finden. So kann man mithilfe von verschiedenen Algorithmen, aus Texten *Informationen extrahieren* (Blake, 2011, p. 129). Dies wird benötigt, um so Rückschlüsse auf das im Text zu explorierende Wissen zu erhalten (Aggarwal & Zhai, 2012c, p. 4). Ein weiteres Themengebiet ist das *reduzieren der Dimensionalität*. Haben mehrere Wörter gleiche Bedeutungen, so lässt sich mit diesem Methodenfeld diese Begriffe zusammenfügen und so auch Ausdrücke, die über mehrere Bedeutungen verfügen erkennen (Crain, Zhou, Yang, & Zha, 2012, p. 130). Zwei der bekanntesten Techniken sind hierbei das Latent semantic indexing, um latenten semantische Bereiche zu identifizieren, und

Themenmodelle, um eine Dimensionsreduktion des Textes zu erreichen (Crain et al., 2012, p. 131). In diesem Bereich sind auch Stemming (das Reduzieren eines Begriffs auf dessen Wortstamm) und das identifizieren von Stopwörtern (Wörter die häufig auftreten, jedoch keine Relevanz haben) angesiedelt (Blake, 2011, p. 134). Darüber hinaus ist es möglich, mithilfe von verschiedenen Methoden den Text *zusammenzufassen* und so wichtige Inhalte zu erfassen (Nenkova & McKeown, 2012, p. 43). Zugleich lassen sich mithilfe von *Clustern* (Aggarwal & Zhai, 2012b, p. 78) und *Klassifikationen* (Aggarwal & Zhai, 2012a) Texte einordnen. Ein weiteres Themengebiet behandelt sich mit der Stimmungserkennung. Dieses Gebiet ist als *Sentiment Analysis* bekannt. Hierbei werden die Stimmung, Meinung, Emotionen und Einstellungen der jeweiligen Person untersucht und erkannt (Liu & Zhang, 2012, p. 415).

3.3 Tools in den Digital Humanities

Sucht man nach Tools und kleineren Anwendungen, die Digital Humanities (DH) Verwendung finden, so stößt man auf frei verfügbaren interaktiver Softwarewerkzeuge, die die Arbeit in den DH erleichtern. Ebenso bietet etwa digi-hum.de⁶, die Simon Fraser University⁷ und die University of Arizona⁸ bereits eine große Übersicht an verwendbaren Tools.

3.3.1 Etablierung der Softwarewerkzeuge

Damit sich Applikationen erfolgreich etablieren können und genutzt werden, müssen diese sowohl dem Anspruch der User Experience, dem Nutzungserlebnis, und der Usability, der Benutzerfreundlichkeit, gerecht werden (Burghardt & Wolff, 2014, p. 1). Allerdings ist genau ebendieses nicht passiert, und die entwickelten Anwendungen sind, trotz erheblicher Investitionen, am Rande der geisteswissenschaftlichen Forschung geblieben (Gibbs & Owens, 2012, p. 1). Burghardt und Wolff nennen hierbei als Gründe, die schlechte Etab-

⁶ Verfügbar unter <http://digi-hum.de/tools/> (zuletzt aufgerufen am 21.03.20)

⁷ Verfügbar unter <https://www.lib.sfu.ca/help/publish/dh/dh-guide/tools> (zuletzt aufgerufen am 21.03.20)

⁸ Verfügbar unter <https://libguides.library.arizona.edu/dighumantools> (zuletzt aufgerufen am 21.03.20)

lierung der erforderlichen Entwicklungsmethoden und dass weiterhin der intellektuelle Input des Forschenden von zentraler Bedeutung ist, auch wenn viele Arbeitsschritte werkzeuggestützt ablaufen (2014, p. 1). Gibbs und Owens begründen die mangelnde Akzeptanz der Tools damit, dass nicht klar war, was die Wissenschaftler von den Werkzeugen erwarten, wie sie mit verschiedenen Arten von Werkzeugen umgehen wollen und wie sie ihren gegenwärtigen und zukünftigen Nutzen wahrnehmen (2012, p. 1). Diese Erkenntnisse über das Nutzerverhalten sind jedoch elementar, um eine gute Usability und User Experience der Anwendungen zu garantieren. Dadurch scheinen die Tools an den Nutzer vorbei entwickelt zu sein. Es stellte sich heraus, dass die Anwendungen nicht ausgefeilter sein müssen, vielmehr ist es wichtig Transparenz und Benutzerfreundlichkeit zu optimieren, um Frustration seitens der Anwender zu vermeiden und sich auch „traditionelle“ Historiker sich mit der Anwendung solcher Technologien sicher fühlen (Gibbs & Owens, 2012, p. 2). Dieses vernachlässigen des Benutzers wird verdeutlicht durch die Tatsache, dass teilweise kaum oder nur bedingt auf das Interface Design eingegangen wurde und was dazu führte, dass beispielsweise ein umfangreiches Softwarepaket ausschließlich über die Kommandozeile angesprochen werden kann, was wiederum für einige Sprachwissenschaftler ein Hindernis darstellt (Burghardt & Wolff, 2014, p. 1 f.).

3.3.2 Maßnahmen

Um dem Fernbleiben der Benutzerfreundlichkeit entgegen zu wirken, schlagen Burghardt und Wolff die Etablierung von, aus der Medieninformatik stammenden, Usability-Methoden vor, um sowohl die weniger Technikaffineren den Einstieg zu erleichtern und das Werkzeug attraktiver zu machen (2014, p. 2). Ein weiterer Vorschlag ist es, um die Akzeptanz der Tools zu verbessern, mithilfe der Benutzeroberfläche und der Dokumentation den Nutzer über die Fähigkeiten des Tools zu informieren, ihn zu ermutigen für ihn neue Werkzeuge zu nutzen und die Usability zu erhöhen (Gibbs & Owens, 2012, p. 12).

3.3.3 Beispiele an Tools in den Digital Humanities

Wie bereits erwähnt, haben DH-Forscher Zugang zu vielen Anwendungen, die sie in ihrer Arbeit einbinden können. Nachfolgend werden einige dieser Tools vorgestellt. Hierbei wurde darauf geachtet, dass die präsentierten Applikationen frei zugänglich sind.

*TAPoR (Version 3)*⁹ dient als Portal für Textanalyse- und Textabfrage-Tools, die bei den DH Forschung Verwendung finden. Es wird von Geoffrey Rockwell, Stéfán Sinclair und Milena Radzikowska geleitet und wurde von Compute Canada Calcul unterstützt (Rockwell, Sinclair, & Radzikowska, n.d.). Hier kann der Nutzer nach bestimmten Werkzeugen und Funktionen suchen und auf einer detaillierten Seite die wichtigsten Informationen der Anwendung einsehen, sowie über eine Verlinkung zu der Webseite des Tools weitergeleitet werden¹⁰. Das webbasierte und open-source Textanalyse-Tool *Voyant Tools*¹¹ wird von Stéfán Sinclair und Geoffrey Rockwell geleitet und bietet eine Vielzahl an Visualisierungsmöglichkeiten der Ergebnisse. Das Tool analysiert den übergebenen Text unter anderem nach den meist vorkommenden Worten, wie sehr zwei Begriffe miteinander korrelieren, wie viele Begriffe ein Satz durchschnittlich beinhaltet oder auch die relative Häufigkeit einzelner Wörter in einem Absatz (Sinclair & Rockwell, 2016). Mithilfe von *Wrangler*¹² können Forscher der DH ihre Daten säubern, transformieren und das Ergebnis als Script (Python oder JavaScript) oder als Daten (beispielsweise JSON oder CSV) exportieren (Stanford Visualization Group, 2011). Durch die *Stanford Sentiment Analysis*¹³ lässt sich die Stimmung einer Eingabe analysieren. Dabei berücksichtigt dieses Tool durch eine Sentiment Treebank den Satzbau und die Reihenfolge der Wörter, wodurch die Genauigkeit der Vorhersage gegenüber anderen Modellen erhöht wird (Socher et al., 2013). Neben verschiedenen Applikationen gibt es sehr viele Softwarepakete und Programmbibliotheken, die den Forschern zu Verfügung

⁹ Verfügbar unter <http://tapor.ca/home> (zuletzt aufgerufen am 21.03.20)

¹⁰ Siehe beispielsweise <http://tapor.ca/tools/341> (zuletzt aufgerufen am 21.03.20)

¹¹ Verfügbar über <https://voyant-tools.org/> (zuletzt aufgerufen am 21.03.20)

¹² Verfügbar über <http://vis.stanford.edu/wrangler/> (zuletzt aufgerufen am 21.03.20)

¹³ Verfügbar unter <http://nlp.stanford.edu:8080/sentiment/rntnDemo.html> (zuletzt aufgerufen am 21.03.20)

stehen und in ihrer Arbeit einbinden können. So bietet das *Natural Language Toolkit*¹⁴ (NLTK) und das *Machine Learning for Language Toolkit*¹⁵ (MALLET) umfangreiche Möglichkeiten, um mit menschliche Sprachdateien zu arbeiten, um Sprachanalysen durchzuführen und technische Anwendungen zu implementieren, welche linguistische Aufgaben beinhalten (Bird, Klein, & Loper, 2009). Um Daten interaktiv zu visualisieren, stehen den Forschern zahlreiche Programm-bibliotheken, wie etwa *D3.js*¹⁶ und *Kartograph*¹⁷, zur Verfügung.

3.4 Sentiment Analysis

Sentiment Analysis, welches ein Untergebiet des oben beschriebenen Text Mining darstellt, und auch unter den Begriffen Opinion Mining und Sentiment Detection verbreitet ist, beschäftigt sich mit der Stimmungsanalyse. Dieses Studienggebiet untersucht die Emotionen, Gefühle, Meinungen, sowie Einstellungen, Bewertungen und Einschätzungen von Personen gegenüber Entitäten (Liu, 2012, p. 7). Liu argumentiert hierbei, dass vor dem Jahre 2000 kaum in der Linguistik und in dem natural language processing (NLP) über dieses Themengebiet geforscht. Dieses Wachstum an dem Interesse dieses Studienggebiets führt er auf mehrere Gründe zurück. Zum einen, dass es einen Anstieg von kommerziellen Anwendungen gäbe, die die Sentiment Analysis vorantrieben, was auch eine Motivation für die Forschung darstelle. Als Zweites nennt Liu, dass dieses Fach viele interessante Forschungsfragen beinhalte, welche davor noch nie untersucht wurden. Zuletzt sieht er auch einen Zusammenhang mit der zunehmenden Zahl an sozialen Medien, die diese umfangreiche Forschung erst ermöglichen. Dieser Anstieg führte dazu, dass es viele Daten von sozialen Medien im Web existieren, welche eine Meinung implizieren. Durch die Tatsache, dass sich Menschen von der Auffassung anderer beeinflussen lassen, sieht Liu die Sentiment Analysis auch einen wichtigen Einfluss in die Sozial-, Wirtschafts-, und Politikwissenschaften, sowie das Forschungsfeld im Mittelpunkt

¹⁴ Verfügbar unter <https://www.nltk.org/> (zuletzt aufgerufen am 21.03.20)

¹⁵ Verfügbar unter <http://mallet.cs.umass.edu/index.php> (zuletzt aufgerufen am 21.03.20)

¹⁶ Verfügbar unter <https://d3js.org/> (zuletzt aufgerufen am 21.03.20)

¹⁷ Verfügbar unter <http://kartograph.org/> (zuletzt aufgerufen am 21.03.20)

in der Forschung von sozialen Medien (2012, p. 7 f.). Überdies wird Sentiment Analysis bereits von vielen Firmen verwendet. So können wendigen Unternehmen Sentiment Analysis an, um auf diese Weise ihre Marktstrategien zu verbessern oder die Kundenmeinung bezüglich ihrer Produkte untersuchen (Cambria, Schuller, Xia, & Havasi, 2013, p. 16).

3.4.1 Definitionen und Begriffe

Nachfolgend werden die einzelnen Definitionen und Begriffe, die in Verbindung mit Sentiment Analysis stehen, definiert und erläutert. So repräsentiert eine *Entität* das zu untersuchende Zeigerobjekt (Liu & Zhang, 2012, p. 417). In der Literatur findet man hierbei häufig folgende Definition (Liu & Zhang, 2012, p. 417):

“An entity e is a product, service, person, event, organization, or topic. It is associated with a pair, $e : (T, W)$, where T is a hierarchy of components (or parts), sub-components, and so on, and W is a set of attributes of e . Each component or sub-component also has its own set of attributes.”

Die Autoren Liu und Zhang (2012, p. 417) führen auf, dass sich somit eine baumartige oder hierarchische Entität vorliege, wobei die Wurzel dieser Entität der Name ist und jeder Knoten, der kein Wurzelknoten dieser Entität ist, eine Segment oder Subkomponente davon sei, deren Links eine Teil-von Relation darstellen würden. Wird diese Darstellung vereinfacht, so erhält man auf der ersten Ebene die Entität selbst und auf der zweiten Ebene alle Komponenten und Attribute der Entität, welche Aspekte genannt werden. Nach Liu und Zhang präsentierte *Subjektivität* keine faktischen Informationen und drückt im Kontext von Meinung die persönliche Sicht, Überzeugungen und Gefühle aus. Subjektive Gedanken und Empfindungen werden hierbei als *Emotion* definiert (Liu & Zhang, 2012, p. 420). Unter einer *Stimmung*, dem Sentiment, versteht man eine Bewertung, ein Gefühl, eine Einstellung oder auch ein Emotion, die mit einer Meinung in Verbindung gebracht wird (Liu, 2015, p. 20). Hierbei könne man zwischen einer rationalen Stimmung, welche keine Emotionen beinhaltet und rational getroffen wurden, und einer emotionalen Stimmung, welche nicht greifbar seien und eine emotionale Reaktion ausdrücken, unterschei-

den (Liu, 2015, p. 20 f.) Liu (2012, p. 17 f.) merkt an, dass die Stimmung auf verschiedene Arten beschrieben werden könne. So sei es zum einen möglich, anhand eines Wertes auf einer numerischen Skale die Intensität der Stimmung festzustellen. Eine weitere Möglichkeit sei es, mithilfe der Begriffe positiv, neutral und negativ, die Stimmung widerzuspiegeln (Liu, 2012, p. 17 f.). Diese Gemütslage ist auch, sofern sie in positiv, negativ und neutral ausgedrückt wird, ebenfalls unter den Begriffen Polarität, Meinungsorientierung, Stimmungsrichtung oder semantische Orientierung bekannt (Liu & Zhang, 2012, p. 418). Ein stimmungstragendes Wort unter hierbei unter dem Begriff sentiment-bearing word (SBW) geläufig. Nach Ding, Liu und Yu (2008, pt. 3) ist ein *Meinungsinhaber* die Person oder Organisation, die die bestimmte Meinung vertritt. Um eine *Meinung* zu definieren wird häufig auf die von Liu (2012, p. 19 f.) aufgestellte Definition zurückgegriffen (Liu, 2015, p. 22; Liu & Zhang, 2012, p. 418; Petz, 2019, p. 16), wobei die Benennung der einzelnen Parameter abweichen können. Hierbei stelle eine Meinung das Quintupel $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ dar. Dabei referenziere e_i auf den Namen der Entität. Mithilfe von a_{ij} definierte Liu den Aspekt der Entität e_i und überdies ließe sich durch s_{ijkl} die Stimmung des Aspektes a_{ij} bestimmen. Hierbei könne das Gefühl sowohl positiv, negativ oder neutral sein, sowie über verschiedene Stärken oder Levels verfügen. Um den Meinungsinhaber zu auszudrücken, werde der Parameter h_k verwendet. Der letzte Parameter t_l des Quintupels definiere den Zeitpunkt, bei dem die Meinung von dem Meinungsinhaber ausgedrückt wurde. Beziehe sich die Meinung auf eine Entität als Ganzes, so würde der Aspekt GENERAL verwendet (Liu, 2012, p. 19 f.). Um Fehler und Unstimmigkeiten zu vermeiden, sei es wichtig, dass zum einen alle fünf Informationen einander entsprechen müssen, unerlässlich sind. Darüber hinaus sei diese Begriffsbestimmung eine Transformierung von unstrukturierten Daten in strukturierten Daten, weshalb man das Quintupel als Datenbankschema sehen könne, welche jedoch nicht alle Facetten einer Meinung beschreiben könne. Zuletzt merkt der Autor an, dass diese Definition von Meinung nur die reguläre Meinung definiert (Liu, 2012, p. 20 f.). Gibt man seine Meinung durch Vergleiche zwischen Entitäten kund, so handle es sich um

eine vergleichenden Meinung (Liu, 2012, p. 110). Oftmals werde in der Forschungsliteratur aus Vereinfachungsgründen die Meinung als reguläre Meinung verstanden (Liu & Zhang, 2012, p. 417). Neben der Differenzierung von Meinung zwischen der Regulären und der Vergleichenden, könne man eine Meinung auch in explizite Meinung, wobei der subjektive Standpunkt klar erkennbar ist, und implizite Meinung, die einen objektiven Blickwinkel impliziert, unterscheiden (Liu, 2012, p. 26). Die nachfolgende Arbeit konzentriert sich auf das Identifizieren von SBW, während das Konzept des Quintupels $(e_i, a_{ij}, s_{ijkl}, h_k, t_i)$ eher verworfen wird.

3.4.2 Aufgaben der Sentiment Analysis

Die interdisziplinäre und komplexe Aufgabe der Sentiment Analysis beinhaltet mehrere separate Aufgaben (Kumar & Sebastian, 2012, p. 4). Die Forscher Kumar und Sebastian (2012, p. 4) benennen diese als Klassifizierung der Subjektivität und als Klassifizierung der Stimmung. Zwischen diesen beiden Schritten gäbe es noch zwei, nicht immer notwendige, Vorgänge. Dies seien die Extraktion des Meinungsinhaber und die Extraktion der Entität oder des Features (Kumar & Sebastian, 2012, p. 4). Wesentlich genauer erläutert Liu die individuellen Teilschritte einer Sentiment Analysis, anhand der oben definierte Quintupel einer Meinung (Liu, 2012, p. 21). Die einzelnen Schritte haben das Ziel, jeweils einen bestimmten Bestandteil des Tupels zu generieren und das Tupel zu formulieren (Liu, 2012, p. 21). In einer späteren Publikation erweitert Liu (2017, p. 21 f.) diese Aufgaben um die Extraktion und Lösung von Meinungsgründen, sowie die Extraktion und Auflösung von Meinungsqualifizierungen.

3.4.3 Analyseebenen

Nach Varghese und Jayasree kann man eine Sentiment Analysis bei einem natürlich-sprachlichen Text auf drei Ebenen durchführen (2013, p. 312 f.). Hierbei könne man die Emotionen auf der *Dokumentenebene*, auf der *Satzebene* oder auf der *Phrasenebene* analysieren (Varghese & Jayasree, 2013, p. 312 f.). Die beiden Forscher (2013, p. 312) äußerten, dass die Sentiment Analysis auf der Dokumen-

tenebene nur für eine einzelne Revision über ein Thema anwendbar sei und dadurch die Gefahr bestehe, dass dem zu explorierenden Dokument mehrere Entitäten miteinander verglichen werden könne, wodurch Vergleichsätze entstehen würden. Diese Klassifikation der Ebene biete jedoch den Vorteil, dass sich ein Gesamtgefühl gegenüber der Entität feststellen ließe, allerdings ist es nicht möglich, verschiedene Emotionen über unterschiedliche Merkmale einer Entität nicht zu ermitteln (Varghese & Jayasree, 2013, p. 312). Bei der Zweiten von Varghese und Jayasree (2013, p. 312 f.) beschriebene Ebene wird für jeden Satz die Polarität einzeln berechnet. Zuletzt stellen sie die *Phrasenebene* vor. Wobei hierbei Phrasen, die SBWs beinhalten, finden und auf der Phrasenebene klassifizieren würde (Varghese & Jayasree, 2013, p. 313). Jedoch berücksichtige diese Ebene keine Langstreckenabhängigkeiten und führe zu Schwierigkeiten, wenn die Negationen weit von ihren SBW entfernt lägen (Varghese & Jayasree, 2013, p. 313). Weitere Ansätze sind hierbei noch die von Gerald Petz (2019, p. 86) erläuterte *Eigenschaftsebene*, auch bekannt als die von Liu (2012, p. 11) erklärte *Entitäts- und Aspekteebene*, welche davon ausgeht, dass eine Meinung aus einem Gefühl und einem Meinungsziel bestehe und das Ziel habe, dieses Empfinden über die Entität oder Aspekt zu extrahieren (Liu, 2012, p. 11). Die letzte hier vorgestellte Analyseebene ist die *Wortebene* da. Die Aufgabe hierbei sei es, herauszufinden, welche Polarität das Wort beinhaltet und stellt ein wichtiger Arbeitsschritt der Sentiment Analysis dar (Kumar & Sebastian, 2012, p. 5).

3.4.4 Subjektivität-Klassifizierung und Sentiment-Klassifizierung

Liu führt (2015, p. 72 f.) auf, dass das Konzept der Subjektivität für Verwirrung bei den Forschern sorgte, weshalb man in vielen Beiträgen Sentiment und Subjektivität als identisch angesehen werden. Jedoch sei es wichtig darzulegen, dass dies nicht der Fall ist, auch wenn diese eine große Überschneidung beinhalten. Für einige Analyseebenen kann es wichtig sein, zu identifizieren, ob dieser Satz relevant ist oder nicht. Hierbei wird zwischen der Objektivität, die tatsächlich existiert, und der Subjektivität, welche in dem Kopf einer Person

entsteht, unterscheiden (Liu, 2015, p. 72). Liu (2015, p. 72) definierte die Subjektivität eines Satzes folgendermaßen:

“An objective sentence states some factual information, whereas a subjective sentence expresses some personal feelings, views, judgments, or beliefs. “

Varghese und Jayasree (2013, p. 313) legen dar, dass sich in den subjektiven Sätzen Informationen für eine Sentiment Analysis verbergen, während objektive Texte nicht relevant sind.

Während sich mithilfe der Klassifizierung der Subjektivität aussagen kann, ob dieser Satz oder diese Einheit für die Erhebung der Stimmung gegenüber einer Entität relevant ist. Lässt sich mithilfe der Klassifizierung des Sentiments aussagen, welche Stimmung vorherrscht. Um die Stimmung zu klassifizieren bestehen eine Vielzahl von Ansätzen. Eine detaillierte Übersicht der verwendeten Methoden und Ansätze werden in den Arbeiten von Petz (2019, p. 158 ff.), Medhat und Kollegen (2014, p. 1096 f.) und von Kumar und Sebastian (2012, p. 8 f.) vorgestellt. In der Literatur wird die Klassifizierung des Sentiments oftmals in die einzelnen Analyseebenen unterteilt (Kumar & Sebastian, 2012, p. 5 f.; Liu, 2010b, p. 10 ff., 2012, p. 30 ff., 2015, p. 47 ff.; Liu & Zhang, 2012, p. 422 ff.; Petz, 2019, p. 58 ff.), da dies jedoch zu umfangreich wird diese Unterscheidung hier vermieden. Grundsätzlich kann man zwischen zwei Ansätzen unterscheiden. Zum einen gibt es den *Maschine Learning Ansatz* und zum anderen gibt es den *lexikonbasierten Ansatz*. Aufgrund des Mangels an annotierten Korpora, wird in der Praxis häufig auf die Methoden des lexikonbasierten Ansatzes zurückgegriffen (Schmidt, Burghardt, & Wolff, 2018; Sun, Luo, & Chen, 2017). Der Maschine Learning Ansatz kann in die überwachten und unüberwachten Lernmethoden unterteilt werden. Ist es schwierig benannte Trainingsdokumente zu finden, so wird auf die unüberwachten Lernmethoden zurückgegriffen (Medhat et al., 2014, p. 1098). Die beaufsichtigten Lernmethoden begründen sich auf das Vorhandensein dieser Trainingsdokumente (Medhat et al., 2014, p. 1098). Bekannte Methoden der beaufsichtigten Lernmethoden sind Naive Bayes, Bayesian Network und Neural (Medhat et al., 2014, p. 1099 f.). Der lexikonbasierende Ansatz verwendet ein Wörterbuch, welches SBWs beinhaltet, und gleich dieses

mit den Daten zur Bestimmung der Stimmung und Polarität ab (D'Andrea, Ferri, Grifoni, & Guzzo, 2015, p. 30). Ein weiterer Ansatz, der neben dem lexikonbasierten Ansatz und dem Maschine Learning Ansatz existiert, ist das Kombinieren beider Ansätze. Was als hybriden Ansatz bekannt ist (D'Andrea et al., 2015, p. 30).

3.4.5 Meinungszusammenfassung

Nach Liu (2017, p. 23 f.) dient die Meinungszusammenfassung (Opinion Summarization, OS) dazu, mehrere Meinungen zusammenzufassen. Er führt auf, dass, aufgrund der Subjektivität, es nötig ist eine große Menge an Auffassungen zu analysieren, um handeln zu können. Hierbei könne, im Gegensatz zu einer traditionellen Zusammenfassung eines oder mehrerer Dokumente, welche sich je nach verwendetem Algorithmus, unterschiedlich ausfallen können, auf das Quintupel einer Definition zurückgegriffen werden. Was wiederum die Auswertung erleichtern würde. So sollen alle OS-Algorithmen darauf ausgelegt sein, immer die gleich Zusammenfassung zu erzeugen, welche jedoch, aufgrund der abweichenden Genauigkeiten der Algorithmen, möglicherweise noch unterschiedliche endgültige Zusammenfassungen erstellen (Liu, 2017, p. 23 f.). Petz (2019, p. 49 f.) führt bezüglich der OS auf, dass die traditionellen Zusammenfassung, mithilfe der Extrahieren von wichtigen Sätzen erstellt würden, während sich die Zusammenfassung einer Meinung sich auf Entitäten und ihre Aspekte und Stimmungsrichtungen fokussiere. Hierbei identifiziere er die OS auf der Aspekte Ebene, die das Quintupel der Meinung als Ausgangsbasis nutzt, und die kontrastive OS, welche gegensätzliche Meinungen gegenüber stellt, als zwei Forschungsrichtungen der OS (Petz, 2019, p. 51 ff.).

3.4.6 Probleme von Sentiment Analysis

In der Literatur wird vielfach auf das Problem der Sentiment Analysis verwiesen. Hierbei steht das Bestimmen einer Opinion, sowie die verschiedenen Arten dieser Meinung, als auch die Definierung der OS im Vordergrund (Liu, 2015, p. 16 ff.). Hierbei führt Liu (2010a, pt. 2) die Identifizierung des Objektes, die Klassifizierung der Stimmungsrichtung sowie die Extrahierung der Eigenschaften

und das Gruppieren von Synonymen als technische Schwierigkeiten auf, auf die man in der Sentiment Analysis tritt. Als weitere Schwierigkeiten in der Sentiment Analysis nennen Varghese und Jayasree (2013, p. 313 f.) die Auflösung von Ko-Referenzen, die in Vergleichssätzen vorhanden sein können, die Extrahierung der Relation zwischen den Wörtern in einem Satz, sowie die Domänenabhängigkeit, bei der eine Stimmung in verschiedenen Domänen unterschiedliche Bedeutungen beinhalten kann. Zusätzlich spielt die Erkennung von Opinion Spam in der Sentiment Analysis eine wichtige Rolle. Opinion Spam wird in der Literatur häufig mit dem Web 2.0 und die dortige Bewertungen von Produkten und den Missbrauch einer solchen Beurteilung in Verbindung gebracht (Jindal & Liu, 2008, p. 219; Petz, 2019, p. 54).

3.4.7 Sentiment Analysis in den Geisteswissenschaften

Häufig wird die Sentiment Analysis in Kombination mit der Emotion Analysis aufgegriffen und beide Analysen sind eng miteinander verwandt. Deshalb finden sich in diesem Kapitel sowohl Forschungsarbeiten der Sentiment Analysis als auch der Emotion Analysis im geisteswissenschaftlichen Feld. So wird die Methodik Sentiment Analysis im Kontext von Medien, beispielsweise bei Tweets (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011; Kouloumpis, Wilson, & Moore, 2011; Zhang, Ghosh, Dekhil, Hsu, & Liu, 2011) durchgeführt. So analysieren Agarwal und Kollegen (2011) das Sentiment von Micoblogging anhand von Twitter Daten. In dieser Studie werden auf die Eigenschaften die Daten eingegangen. So kann eine Meinung durch Akronyme (z. B. gr8t → great) und Emoticons (:-) → positive) transportiert werden. Zudem ist es bei dieser Art von Daten möglich, mithilfe von Hashtags Themen anzudeuten und mithilfe von ‚@‘ andere Nutzer als Ziel zu markieren (Agarwal et al., 2011, p. 32). Ein wichtiger Bestandteil dieser Studie war das Einführen eines Emoticon- Wörterbuchs und eines Akronym-Wörterbuch um die Tweets entsprechend Vorverarbeiten zu können. Anschließend untersuchten die Forscher die Genauigkeit zweier Klassifikationsaufgaben auf verschiedenen Modellen (Agarwal et al., 2011). Sie kamen zu dem Ergebnis, dass die Klassifizierung mithilfe der Polaritäten positive, negativ und neutral die besten Ergebnisse hervorbrachte und

das entworfene Modell mit Baumkernen die anderen Modelle übertraf. Ein weiterer Bereich, in dem die Sentiment Analysis in Studien angewendet wird ist anhand von Nachrichten (Godbole, Srinivasaiah, & Skiena, 2007; Kim & Hovy, 2006). So untersuchten die Forscher anhand von online Nachrichten auf das Identifizieren einer Meinung, dessen Meinungsinhaber und das zugehörige Thema (Kim & Hovy, 2006, p. 1). Eine Motivation von Kim und Hovy für diese Studie war, dass ein besseres Verständnis von Beziehungen zwischen Ländern und Organisationen erreicht werden könne, indem man Meinungsinhaber mit unterschiedlichen Positionen zu verschiedenen sozialen und politischen Themen gruppiert (2006, p. 1). Der hierfür gewählte Ansatz beinhaltete 723 SBWs (Verben und Adjektive) und 20133 Sätze, welche sich auf einen meinungstragenden Frame beziehen, und wobei die semantische Struktur zur Identifikation genutzt wurde (Kim & Hovy, 2006, p. 3 f.). Laut beiden Autoren konnte das entwickelte System eine Verbesserung gegenüber dem Ausgangspunkt erzeugen. Jedoch ließen die Ergebnisse darauf schließen, dass die Identifizierung von Meinungsinhabern und Themen eine anspruchsvolle Aufgabe darstellt (Kim & Hovy, 2006, p. 7). Zudem finden sich wissenschaftliche Publikationen zur Sentiment Analysis in Verbindung mit literarischen Textgattungen und historischen Texten (Buechel, Hellrich, & Hahn, 2016; Sprugnoli, Tonelli, Marchetti, & Moretti, 2016). So erzeugten Sven Buechel, Johannes Hellrich und Udo Hahn ein Sentiment-Wörterbuch, indem die Forscher ein zeitgenössisches Lexikon erweiterten und an die Domäne für deutschsprachige Texte des 18. und 19. Jahrhunderts anpassten (Buechel et al., 2016, p. 54). Sie konnten langfristigen Entwicklungen von Emotionen in Textform, sowie Hinweise auf kontrastierende Muster in akademischen Texten und in der Belletristik aufzeigen (Buechel et al., 2016, p. 59). Außerdem finden sich Studien von literaturwissenschaftlichen Texten anhand von diversen Textgattungen. Dazu gehören etwa Dramen (Nalisnick & Baird, 2013; Schmidt & Burghardt, 2018), Novellen und Märchen (Mohammad, 2011; Volkova, Mohler, Meurers, Gerdemann, & Bühlhoff, 2010; Zehe et al., 2016). Schmidt und Burghardt 2018 führten eine Studie zur Bewertung von Techniken der lexikon-basierten Sentiment Analyse durch. Wobei die

Forscher hierbei als Grundlage die Stücke von Gotthold Ephraim Lessing wählten (Schmidt & Burghardt, 2018, p. 139). Zunächst erstellten die Forscher einen Gold Standard mit 200 manuell annotierten Reden (Schmidt & Burghardt, 2018, p. 140). Um verschiedene Techniken zu evaluieren, wählten Schmidt und Burghardt verschiedene Parameter (2018, p. 141 ff.). So wurden über 400 verschiedene Konfigurationen von Stoppwortlisten, Lexika, die Unterscheidung zwischen Groß- und Kleinbuchstaben, linguistischen Varianten, Sentiment-Berechnung und das Verwenden von Lemmatisierungsverfahren erstellt und durchgeführt (Schmidt & Burghardt, 2018, p. 141 ff.). Anhand dieser Studie konnten die Forscher wertvolle Erkenntnisse gewinnen. Nachfolgend wird nur exemplarisch einige Feststellungen aufgezählt. So übertreffen generell Lexika, welche mit Polaritätsstärken ausgezeichnet sind, die Berechnungen des Sentiment-Score gegenüber Lexika mit dichotomen Polaritäten. Zudem liefert. Außerdem empfehlen sie das Nutzen von Stoppwortlisten, sofern Stoppwörter durch zusätzliche NLP-Prozesse erzeugt werden oder im verwendeten Lexikon enthalten sind. Die letzte hier aufgeführte Schlussfolgerung ist, dass das Lexikon SentiWS die beste Leistung im Vergleich der anderen Lexika mit denselben gewählten Parameter hervorbrachte (Schmidt & Burghardt, 2018, p. 144 f.).

3.5 Lexikonbasierte Sentiment Analysis

Da die zu implementierende Anwendung auf den Ansatz der lexikonbasierten Sentiment Analysis zurückgreift, wird in diesem Kapitel auf diesen Ansatz detaillierter eingegangen. Die Motivation für den gewählten Lösungsansatzes ist, dass sich je nach Domäne der DH das Vokabular und die SBWs unterscheiden können oder nicht im Lexikon vorhanden sind (Flüh, 2019; Schmidt, Burghardt, & Wolff, 2018). Aus diesem Grund ist es wichtig, dass der Nutzer die Sentiment Analysis an seine Domäne anpassen kann. Somit bietet sich eine lexikonbasierte Sentiment Analysis an, wobei der Benutzer die in dem Lexikon enthaltenen SBWs anpasst.

3.5.1 Vorgehensweise der lexikonbasierten Sentiment Analysis

Der lexikonbasierte Ansatz ist, vereinfacht gesagt, ein Wortabgleich mit einem Sentiment-Lexikon. Dabei skizzierte Manasee Godsay (Godsay, 2015, p. 27) nachfolgenden Prozess. Zunächst werde zu Beginn der Analyse, das Ziel dieser Sentiment Analysis und der Inhalt des zu untersuchenden Textes festgelegt. Anschließend werde der Text in das System geladen und vorverarbeitet, was beispielsweise das umwandeln von Symbolen in Textinhalten und das Löschen von unerwünschten Worten impliziert. Sobald dies getan ist, werde der Text geparkt und mit POS-Tags versehen. Der vierte Teilschritt beinhalte ebenfalls NLP, bei dem der Text durch das Auffinden von Synonymen und Stoppwörter der Text verfeinert wird. Der letzte Schritt stehe für das Analysieren und Bewerten des Textes, wobei mithilfe von Scoring die Intensität des Sentiments ermittelt werde (Godsay, 2015, p. 27).

3.5.2 Lexika

Nach D'Andrea und Kollegen (2015, p. 30) gebe es drei Techniken, um ein solches Wörterbuch zu generieren und werden in zwei Gruppen von verschiedenen Lösungswegen kategorisiert. Beim wörterbuchbasierenden Ansatz werde eine kleine Anzahl an Wörtern manuell eingeordnet. Anschließend werde die Liste durch Synonymen und Antonymen erweitert. Dieser Prozess werde so lange wiederholt, bis eine Iteration keine neuen Synonymen und Antonymen zu dem Wörterverzeichnis hinzufügt hat, wodurch ein domainunspezifisches Wörterbuch entsteht (Kim & Hovy, 2004, Chapter 2.1.1; Kumar & Sebastian, 2012, p. 5 f.). Der Corpus-basierte Ansatz löse die Schwierigkeit der Bildung von SBW mittels von kontextspezifischen Orientierungen. Während die statistischen Techniken mithilfe von statistische Verfahren Meinungsäußerungen festzustellen, nutzen semantische Techniken, um die Polarität eines eng verwandten Wortes zu identifizieren (Medhat et al., 2014, p. 1102 ff.). Die erstellten Lexika können hierbei unterschiedliche Informationen über die Wörter beinhalten und in verschiedenen Datenformaten vorliegen. Da in dieser Arbeit sich auf die Analyse von deutschsprachigen Texten beruft, werden nachfolgend etablierte Sentiment-Lexika vorgestellt.

Das deutschsprachige und frei verfügbare Wörterbuch *SentimentWortschatz*, kurz *SentiWS*, beinhaltet 3468 SBWs, die in der Grundform dargestellt werden (Remus, Quasthoff, & Heyer, 2010). Hierbei wird die Polarität eines SBW in einer Skala von -1 (maximal negativ) bis +1 (maximal positiv) ausgedrückt und auf vier Nachkommastellen gerundet (Remus et al., 2010, p. 1170). Die einzelnen SBWs des Lexikons sind mit einem POS-Tag versehen und können so in Adjektive, Adverbien, Nomen und Verben unterteilt werden. Zudem sind zu den einzelnen Grundformen die verschiedenen Flexionsvarianten des Wortes gespeichert (Remus et al., 2010, p. 1168). Zählt man die Grundformen und dessen Flexionen zusammen, so erhält man ein Sentiment-Wörterbuch, welches 16406 positive Sentiment tragende Wörter und 16328 negativ geladenen Wörter dokumentiert. Das Wörterbuch ist hierbei so strukturiert, dass zuerst die Grundform eines Wortes gelistet ist. Daraufhin folgt dessen POS-Tag und dessen Flexionsvarianten (Remus et al., 2010, p. 1168). Nachfolgend wird ein Ausschnitt von *SentiWS* präsentiert (Abbildung 1).

```
Fleiß|NN 0.3113 Fleißes,Fleißs
Flexibilität|NN 0.0040
Fortschritt|NN 0.0040 Fortschritte,Fortschrittes,Fortschritts,Fortschritten
Freiheit|NN 0.0040 Freiheiten
Freispruch|NN 0.0040 Freisprüchen,Freisprüche,Freispruches,Freispruchs,Freispruche
Freude|NN 0.6502 Freuden
Freund|NN 0.0116 Freunden,Freundes,Freunde,Freunds
Freundlichkeit|NN 0.0913 Freundlichkeiten
Freundschaft|NN 0.2059 Freundschaften
```

Abbildung 1: Ausschnitt SentiWS

Ulli Waltinger (2010) erzeugte mithilfe durch den Ansatz einer Kombination von einer halb-automatischen Übersetzungsmethode und einer manuellen Bewertung und Erweiterung des Begriffs das *GermanPolarityClues* Lexikon (Version 0.2, April 2012). Auf diese Weise konnten 10141 Sentiment tragende Wörter (Flexionsvarianten und Grundformen) identifiziert werden. Im Gegensatz zu *SentiWS*, bei dem die einzelnen Flexionsvarianten am Ende jeder Zeile gesammelt stehen, wird für jedes Feature beim *GermanPolarityClues* eine extra Zeile verfasst (Waltinger, 2010). Im Schema des Lexikons wird zunächst das Feature dokumentiert. Anschließend folgen die Angaben über das Lemma und der POS-Tag ebenjenes Feature. Schlussendlich wird das Wort mit der zugehörigen

Polarität (positiv, negativ, neutral) versehen. Eine Besonderheit dieses Lexikons ist es, dass jedes Wort noch mit drei Wahrscheinlichkeiten versehen sind. Diese Polaritätswahrscheinlichkeiten sagen aus, wie wahrscheinlich es ist, dass dieses Wort in einer positiven/neutralen/negativen Beziehung während der Evaluation auftritt (Waltinger, 2010, p. 1649). Zudem sind einige Sonderzeichen (zum Beispiel ,?', ,!', ,(, ,)' und ,*') und die Zahlen von Null bis Neun vermerkt. Bei der Inspektion der Daten fällt auf, dass für einige Wörter über keine der drei Polaritätswahrscheinlichkeiten verfügen (Abbildung 2).

Heiligtum	Heiligtum	NN	positive	0.0040/-/-
Heiligtumes	Heiligtum	NN	positive	0.0040/-/-
Heiligtums	Heiligtum	NN	positive	0.0040/-/-
Heiligtümer	Heiligtum	NN	positive	0.0040/-/-
Heiligtümern	Heiligtum	NN	positive	0.0040/-/-
Heilklima	Heilklima	NN	positive	-/-/-
Heilkraft	Heilkraft	NN	positive	-/-/-
Heilung	Heilung	NN	positive	0.0040/-/-
Heilungen	Heilungen	NN	positive	0.0040/-/-

Abbildung 2: Ausschnitt GermanPolarityClues

Die *Berlin Affective Word List - Reloaded* (BAWL-R) wurde von VÖ und Kollegen (2009) erstellt und umfasst 2107 Nomen, 504 Verben und 291 Adjektive und repräsentiert positive, negative und neutrale Wörter. Dabei wurde im Rahmen einer Studie mit Studenten die einzelnen Wörter anhand ihrer emotionalen Valenz, ihrer Arousal and Darstellbarkeit gewertet. Auf diese Weise wurde die emotionale Valenz auf einer Skala von -3 (maximal negativ) und +3 (maximal positiv) definiert. Die Darstellbarkeit wird in diesem Lexikon durch einen numerischen Wert von +1 (minimale Darstellbarkeit) bis +7 (maximale Darstellbarkeit) definiert. Aufgrund der sexuellen Assoziation mit dem deutschen Begriff „Erregung“, wurde, um die Arousal von einem Spektrum von +1 bis +5 zu werten, auf die abstrakte Figuren zurückgegriffen (Vö et al., 2009, p. 535). Die erhobenen Informationen des BAWL-R lassen sich in verschiedenen Kategorien klassifizieren. Es gibt generelle Informationen über das SBW (Vö et al., 2009, p. 536). Dazu gehört das Wort in Großbuchstaben und in Kleinbuchstaben (Abbildung 3). Sowie die Wortklasse, welche in Nomen, Verben und Adjektive unterteilt werden kann. Zudem existieren Subjektive Wertungen: die emotiona-

le Valenz, die Arousal und die Darstellbarkeit (Võ et al., 2009, p. 536). Neben dem Durchschnitt der Wertungen findet sich hierbei auch die Standardabweichung. Zuletzt werden einige psycholinguistische Indizes aufgeführt (Võ et al., 2009, p. 536). Darunter finden sich die Anzahl der Silben, die Anzahl der Buchstaben, die Anzahl der Phoneme im Wort und die Häufigkeit der orthographischen Nachbarn (Anzahl der Wörter, die durch ändern eines Buchstabens gebildet werden können).

WORD	WORD_LOWER	WORD_CLASS	EMO_MEAN	EMO_STD	AROUSAL_MEAN	AROUSAL_STD
AAL	aal	N	-0,5	0,707106781	2,380952381	1,244033379
AAS	aas	N	-2,1	1,100504935	2,631578947	1,422459775
ABART	abart	N	-1,6	0,699205899	3,277777778	1,017815166
ABBAU	abbau	N	-1	1,1697953	3	1,297771369
ABBAUEN	abbauen	V	-0,8	0,92	2,105263158	1,242521488
ABBILD	abbild	N	-0,2	0,632455532	2,105263158	0,809302638
ABBRUCH	abbruch	N	-0,7	1,159501809	2,904761905	1,04425868
ABDANKEN	abdanken	V	-0,4	0,84	2,666666667	1,084652289
ABDRUCK	abdruck	N	-0,1	0,316227766	2,235294118	0,9701425

Abbildung 3: Ausschnitt Berlin Affective Word List - Reloaded

3.5.3 Valence shifters

Neben den SWBs, die eine Stimmung in sich tragen, existieren Wörter, die dieses Sentiment verändern können. So können Amplifiers, Downtoners und Negatoren, beispielshalber „very“ und „not“, sowie Irrealis markers (Irrealismus-Marker), die auf eine Unzuverlässigkeit des Sentiment eines Satzes hindeuten, Einfluss auf das Sentiment nehmen (Taboada et al., 2011). Die Autoren führen hierbei auf, dass beispielsweise das Wort „sleazy“ einen Sentiment-Wert von -3.0 beinhaltet, während das Begriff „somewhat“ einen Modifizier von -30% enthält. Tritt nun der Begriff „somewhat sleazy“ auf, impliziert dieser einen Sentiment-Wert von -2.1. Um mit Negationen umzugehen stellen Taboada und Kollegen zwei Varianten vor. So gibt es den Ansatz, der das Sentiment umkehrt (switch Negation). Wodurch ein Wort mit einem +3 Sentiment in Kombination mit „not“ zu einem -3 Sentiment-Wert umgewandelt wird. Die Problemantik hierbei ist, dass beispielsweise „not excellent“ weniger positiv ist als „not good“. Aus diesem Grund wählten die Autoren für ihre Arbeit eine zweite Methode, bei der das Sentiment des Wortes um einen festen Betrag, etwa 4, in Richtung der entgegengesetzte Polarität verschoben wird (Taboada et al., 2011).

3.6 User-Centered Design

Die Implementierung der Web-Anwendung orientiert sich an dem User-Centered Design (UCD) um eine optimale Benutzbarkeit des Tools zu garantieren. Aus diesem Grund wird dieser Design Prozess an dieser Stelle erläutert und einen Einblick in das zukünftige Vorgehen ermöglicht. Die Methodik des UCD entstammen aus der Human-Computer-Interaction und umfasst die Entwicklung und das Design von Software, mit dem Ziel den Bedürfnissen der Nutzer zu erfüllen und diesen bei der Entwicklung miteinbeziehen (Lowdermilk, 2013, pt. 2). Darüber hinaus zeichnet sich durch seine „kontinuierliche und interaktive Vorgehensweise“ (Eberhard-Yom, 2010, p. 122) aus. Die ISO-Norm des User-Centered Design DIN EN ISO 9241-210 „Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme“ legt hierbei die wesentlichen Aspekte fest (Jacobsen & Meyer, 2017, p. 83). Die einzelnen Phasen der Entwicklung beinhalten das Analysieren der Anforderungen der Nutzer, beispielsweise durch Zielgruppenbefragungen, die Konzeption von Design Lösungen, etwa mit Hilfe von Scribbels oder Wireframes, die Umsetzung und das Testen von Prototypen, wie etwa durch sogenannte Usability Tests oder Eyetracking mit den Nutzern, sowie die Implementierung der Anwendung (Eberhard-Yom, 2010, p. 123 f.). Weitere bekannte Praktiken um den Nutzer und seine Anforderungen zu untersuchen schlagen Still und Crane Interviews, Tagebuchstudien, Fokusgruppen, sowie durch beobachtete Methoden und das Analysieren von bestehenden Daten vor (Still & Crane, 2017, p. 67 ff.). Um Design Lösungen zu finden bieten so sich etwa das Entwickeln von Use Cases, Aufgabenanalysen, das Anfertigen von User Matrizen, das Erstellen von Nutzerprofilen und Personas sowie das Anfertigen eines Storyboard an (Still & Crane, 2017, p. 86 ff.). Die gewonnenen Kenntnisse werden anschließend in Prototypen umgesetzt. Hierbei lassen sich zwischen verschiedenen Stadien der Prototypen unterscheiden. Die einfachste und günstigste Version eines Prototyps stellen die Papierprototypen dar, die in der Gruppe der Low-Fidelity-Prototypen einzuordnen ist. Hierbei werden aus Papier Prototypen gebaut, und mit dem man klickbare Elemente und Interaktionen überprüfen kann (Still & Crane, 2017, p. 166 f.). Die

nächste Stufe stellen die Medium-Fidelity Prototypen dar. Durch Wireframes wird eine realistischere Umgebung als mit einem Low-Fidelity Prototypen geboten. Hierbei lassen sich bereits Designentscheidungen, das Layout, der Aufbau sowie die Grundfunktionen ablesen (Still & Crane, 2017, p. 172). Die letzte Gruppe der Prototypen stellen die High-Fidelity Prototypen dar. Diese Prototypen sind in ihrer Funktionalität und ihrem Design dem Endprodukt schon sehr nah, jedoch ist das Back-End noch nicht komplett und es werden keine Live Daten übertragen (Still & Crane, 2017, p. 174). Um das entwickelte Design zu überprüfen schlagen Still und Crane zwei Arten von Usability Tests vor. Zum einen das A/B Testen, bei dem man zwei Prototypen oder Produkte miteinander vergleicht, und zum anderen das formative Kontrollen, welches auch als iterative Testen oder als Guerilla-Test bekannt ist und dabei eine kleine Gruppe von Teilnehmern das Design in jeder Iteration testet (Still & Crane, 2017, p. 192 f.).

4 Arbeitsagenda

In diesem Kapitel werden die nächsten Schritte erläutert und motiviert. Wie bereits in der Related Work aufgearbeitet wurde, ist es für die Etablierung von Softwarewerkzeugen wichtig, dass diese an die Bedürfnisse des Nutzers angepasst sind und dessen Anforderungen entsprechen. Aus diesem Grund finden sich in den kommenden Arbeitsschritten Methoden des User-Centered-Design Prozess zurückgegriffen. Um eine Anwendung zu implementieren wird in der ersten Phase zunächst die Anforderungen an das System zu erheben und festzuhalten. Hierbei wurden Stakeholder identifiziert, Interviews mit Experten und potenziellen Nutzern der Sentiment Analysis geführt und eine Konkurrenzanalyse verschiedener Tools durchgeführt. Anhand der gewonnenen Anforderungen wurden diese in einem weiteren Prozess analysiert, spezifiziert und in einem Anforderungsdokument festgehalten. Dieses Dokument dient als Grundlage um die Anwendung zu entwerfen und zu implementieren. Um den Nutzer hierbei einzubeziehen, Usability-Issues aufzudecken und dessen Meinung über die Anwendung einzuholen werden ein formativer Usability-Test im Labor anhand von Mitarbeitern und Studenten der Universität Regensburg und eine summative Studie in Form eines Remote Usability-Test anhand von Experten und potentielle Nutzer angewandt. Die gewonnenen Kenntnisse der Usability Test fließen anschließend im Design und in der Funktionalität der Anwendung ein.

Arbeitsagenda

1. Anforderungsanalyse
2. Erster Designprozess
3. Formative Evaluation
 - 3.1. Evaluation
 - 3.2. Verbesserungen
4. Summative Evaluation
 - 4.1. Evaluation
 - 4.2. Verbesserung

5 Anforderungsanalyse

In diesem Kapitel werden die Anforderungen an die Web-Anwendung erfasst, analysiert und beschrieben. Im Rahmen des USD Prozessen werden hierbei bereits die Zielgruppe der Applikation miteinbezogen. Um die Anforderungen an die Anwendung zu definieren, wird das von IEEE erstellte Konzept zur Anforderungsanalyse durchgeführt. Dabei werden zuerst in der Anforderungserhebung (Abran, Moore, Bourque, & Dupuis, 2004, p. 2–4) die Anforderungen gesammelt. In der Anforderungsanalyse werden diese anschließend analysiert (Abran et al., 2004, p. 2–6). In der Anforderungsspezifikation wird, abhängig von der Komplexität des Systems, das System und dessen Anforderungen, so- wie die Anforderungen an die Software definiert und in einem (elektronischen) Dokument festgehalten, welche systematisch überprüft, bewertet und genehmigt werden kann (Abran et al., 2004, p. 2-7 f.). Zunächst werden die Stakeholder identifiziert. Hierbei richtet sich das Tool an DH-Forscher als Zielgruppe, und Personen, die für den Support der Webanwendung zuständig sind. Zudem gibt es Menschen, die für die Wartung und den Support der Anwendung verantwortlich sind. Die definierten Anforderungen können zwischen funktionalen und nicht-funktionalen Anforderungen unterschieden werden. Während funktionale Anforderungen definieren, was ein System aus einem fachlichen Blickwinkel leisten soll und sowohl anwendungsorientiert als auch domänenspezifisch sind (Hammerschall & Beneken, 2013, p. 82), sammeln sich unter dem Begriff der nicht-funktionale Anforderungen alle Anforderungen, die nicht zu den funktionalen Anforderungen fallen (Hammerschall & Beneken, 2013, p. 88).

5.1 Anforderungserhebung

Um die Anforderungen an das Tool zu erheben, wurden Interviews mit mehreren Personen der Zielgruppe durchgeführt. Somit erfährt man von den Vorstellungen und Anforderungen der User an die Anwendung, sowie die Eigenschaften und Erfahrungen, die der Nutzer bereits im Anwendungsbereich von Webtools und Sentiment Analysis mitbringt. Darüber hinaus wurde eine Konkur-

renzanalyse von ähnlichen Webanwendungen durchgeführt, um so die Stärken und Schwächen von anderen Applikationen herauszufinden und wie man sich von diesem Abheben könnte.

5.1.1 Interviews

Da die Sentiment Analysis ein nicht alltägliches Anwendungsgebiet ist, wurde für die Befragung DH-Forscher im Rahmen von semistrukturierten Interviews herangezogen, welche bereits über Kenntnisse in diesem Gebiet verfügen oder potenzielle Nutzer darstellen. Je nach Präferenz der Interviewpartner wurde das Gespräch persönlich, telefonisch oder über Skype geführt. So ergaben sich am Ende zwei persönliche Interviews, eines mithilfe des Telefons und drei über Skype. Bezüglich der Strukturierung der Befragung wurde ein semistrukturiertes Interview gewählt. Dies bietet den Vorteil, dass sich mithilfe dieses Strukturierungsgrads eines Interviews die Interviewbeteiligten zusätzliche Bedenken und Probleme äußern, sowie bisher unbekannte Schwierigkeiten aufdecken können (Wilson, 2013, pt. 2). Der hierfür entwickelte Leitfaden beinhaltet zunächst einige Fragen zur Person und zu den Kenntnissen in dem Anwendungsgebiet. Anschließend folgen Fragen zu dem zu untersuchenden Text, den Sentiment Lexika, auf denen der Analyseansatz basiert, darüber hinaus Fragen zur Analyse, zu den Ergebnissen der Analyse und deren Visualisierung. Nachfolgend werden die interviewten Personen vorgestellt und Ergebnisse der Befragungen zu jedem Themenbereich gesammelt erläutert, um eine ausschweifende Zusammenfassung jedes einzelnen Interviews zu vermeiden. Zu den interviewten Personen gehören zwei DH-Forscher, welche tiefgreifende Erfahrungen mit Sentiment Analysis gesammelt haben und auch schon entsprechende Tools programmisch umgesetzt (P1; P5). Darüber hinaus fokussierte sich eine weitere befragte DH-Forscherin auf Anwendungen in den DH und deren pädagogischer Vermittlung (P3). Hinzu kommen zwei Literaturwissenschaftler/innen mit Interesse an den DH. Diese setzen sich aus einer Person, die über viel Erfahrung in der Sentiment Analysis verfügt (P4), und eine Person, welche sich mit dieser Methodik vertraut gemacht hat (P6), zusammen. Eine Person stellt hierbei der Projektkoordinator einer virtuellen Forschungsumgebung zur digitalen Erfor-

schung von Texten dar (P6). Die letzte Teilnehmerin ist eine Studentin in den DH, welche über fortgeschrittene Erfahrungen in der Sentiment Analysis vorweisen kann und bereits selbst ein Tool in diesem Forschungsfeld entwickelt hat (P2).

5.1.1.1 Zu untersuchenden Texte

Um herauszufinden, welche Texte die Forscher untersuchen, wurden in den Interviews Fragen bezüglich der Strukturierung des Textes, sowie der Umfang und auch das Dateiformat und dessen Import gestellt. Hierbei wurden sehr unterschiedliche Antworten gegeben. Aufgrund der verschiedenen Anwendungsgebiete und der heterogenen Datengrundlage, ließe sich so keine einheitliche Strukturierung des zu untersuchenden Textes feststellen. So können historische Branchenzeitungen in Zeitungsartikel und Headings unterteilt werden (P1). Dramen können in Titel und Untertitel, Personen, und beispielsweise in Akte und Szenen strukturiert werden (P4) oder in Überschriften und Kapitel gegliedert sein (P1). Darüber hinaus kann der Text als Fließtext mit einzelnen Sätzen und ohne Überschriften (P2). Weiter kann der Text bereits so vorverarbeitet sein, dass dieser keine Strukturelemente aufweist, normalisiert wurde und Absätze, sowie Überschriften, entfernt wurden (P3). So unterschiedlich die Texte strukturiert sind, so inhomogen sind auch deren Umfang. Jedoch sind diese Texte, im Vergleich mit anderen Gebieten eher kurz und auch die Größe der Korpora sind vergleichsweise gering (P5; P6). Hierbei lassen sich Unterschiede zwischen der genauen Größenordnung der Korpora feststellen. So schwanken die Angaben zwischen maximal 100 (P6) und mehrere 100 (P1) Texten in einem Korpus. Bei anderen Datengrundlagen sind die Texte hingegen auf eine gewisse Zeichenanzahl begrenzt (P2). Damit die Anwendung einen Text analysieren kann, muss dieser in das System importiert werden. Hierbei herrschte Einigkeit aller befragten Personen, dass der Nutzer den Text als Datei in die Anwendung hochladen kann (P1; P2; P3; P4; P5; P6). Die Bereitstellung eines Inputfeldes wird von einer Person als überflüssig angesehen (P1), wobei dies ein kleiner Showcase sein kann, indem eine kleine Eingabe in Echtzeit analysiert wird und so der Nutzer eine Vorstellung der Analyse bekommt, die an seinen hochgela-

denen Dateien durchgeführt wird (P5). Um die Texte hochzuladen und dort weiter verarbeiten zu können, ist es wichtig, zu wissen, in welchem Dateiformat die Dokumente vorliegen werden. Hierbei wurde besonders häufig eine reine Textdatei (.txt) als Format genannt (P1; P2; P3; P4; P5; P6). Ebenfalls wurde eine XML-Datei (.xml) als ein häufig zu verwendetem Format angesehen (P1; P3; P4; P5; P6). Wobei dieses Format primär für Anwender mit Erfahrung geeignet sei (P3). Weitere genannte Dateiformate ist JSON (.json) (P5), sowie PDF (.pdf) und das von Microsoft entwickelte Dateiformat. docx (P3). Bezüglich der Möglichkeit den Text hochzuladen, wurde angemerkt, dass die Interoperabilität wichtig ist (P6). Dabei ist gemeint, dass der Nutzer den Text von einem anderen System in die Anwendung importieren kann, ohne dass diese Datei angepasst werden muss. Dies wird auch von einem skizzierten Workflow verdeutlicht, bei dem der Nutzer einen Text von einem Repository herunterlädt, diesen auf dem Desktop speichert und in die Anwendung importiert (P3). Darüber hinaus ist anzunehmen, dass die Texte in einer UTF-8 Kodierung vorliegen (P5).

5.1.1.2 Natural language processing

Als Nächstes wurde auf die Verarbeitung des Textes eingegangen. Hierbei finden oft NLP Prozesse Gebrauch. Dabei stellen die Lemmatisierung und das Stemming häufig angewendete Arbeitsschritte dar (P1; P2; P3; P5; P6). Bei der Lemmatisierung wird das Wort in sein Lemma zurückgeführt, während Stemming das Begriff auf seinen Wortstamm herunterbricht. Zwar nimmt das Lemmatisieren eine wichtige Stelle für den Abgleich zwischen Wort und dem verwendeten Sentiment-Wörterbuch ein (P3), jedoch ist dies technisch schwierig umzusetzen (P5). Ein weiterer bekannter Prozess ist die Named-entity-recognition, welche potenziell sinnvoll für die Anwendung ist (P1; P6). Weitere Prozesse ist den Text in Kleinbuchstaben zu konvertieren oder Stoppwörter herauszufiltern, deren Liste man eventuell anpassen kann (P5). Darüber hinaus gehören das Part-of-speech Tagging und die Tokenisierung zu den Standardprozessen des NLP (P1). Je nach Know-how gibt es unterschiedlich Präferenzen bezüglich der Kontrolle der durchzuführenden NLP-Prozesse. Während Menschen mit weniger Erfahrung sich wünschen, dass das Programm sehr viele

Prozesse dem Nutzer abnimmt und in den Hintergrund verlagert, bevorzugen Personen mit mehr Wissen die Möglichkeit den Text selbst entsprechend vorzubereiten und so die Führung über die Analyse zu erhalten (P1; P2; P3; P4; P6).

5.1.1.3 Sentiment Analysis

Darüber hinaus sollte ersichtlich sein, auf welcher Ebene die Anwendung das Sentiment berechnet, und es wurde angemerkt, ob es nicht eine Einstellung gäbe, bei der man die Texte in Einheiten aufteilen könne (P1; P4). Überdies wurde erwähnt, dass es interessant ist, welche Basisemotion vorherrscht und in welcher Intensität die Emotionen und was genau bewertet wird auftreten (P3; P4). Abgesehen davon birgt die Analyse die Schwierigkeit, dass Emotionen in verschiedene Repräsentationsformen, beispielsweise non-verbal und körperliche Zustände, vorliegen (P3) und auch die Sprache, in der der Text verfasst wurde, zu Herausforderungen führen kann (P1; P2). In den Befragungen wurde von Personen angemerkt, dass es SBWs gibt, die domänenspezifisch sind oder im Laufe der Zeit ihr Sentiment verändert haben (P1; P3; P6). Hierbei wurde als Exempel für ein domänenspezifisches Meinungswort „Schulden“ genannt, welches ohne besonderen Kontext negativ belastet ist, jedoch in der Finanzbranche neutraler gesehen wird (P1). Ein weiteres Beispiel war die Anwendung eines medizinischen Sentiment-Lexikons auf eine literarische Textsorte, wodurch die Aussagefähigkeit der Sentiment Analysis und die Zitierwürdigkeit der Ergebnisse leiden würden (P3). Für ein SBW, welches im Laufe der Zeit seine Bedeutung veränderte, wurde das Wort „toll“ dargelegt, welches heutzutage ein positives Sentiment besitzt, während es im historischen Kontext so viel wie „verrückt“ bedeutet und somit im geschichtlichen Zusammenhang mit einem negativen Sentiment behaftet ist (P6). Aus diesem Grund ist es für diese Personen sehr wichtig, dass der Nutzer die Möglichkeit hat, ein eigenes Wörterbuch hochzuladen und zu verwenden oder ein bestehendes anzupassen (P1; P3; P6). Dabei wurde erwähnt, dass beispielsweise die Literaturwissenschaftler ihre Texte sehr gut kennen (P6). Für die Nutzung eines eigenen Wörterbuchs verwendet eine befragte Person eine CSV-Datei, in der in der einen Spalte die SBWs stehen und in der anderen Spalte die Polaritätsstärken der Wörter gespei-

chert sind (P1). Die meist genannten Sentiment-Wörterbücher sind das SentiWS und das NRC Lexikon (P2; P3; P4).

5.1.1.4 Ergebnisse der Sentiment Analysis

Zuletzt wurden die Personen bezüglich der erwarteten Ergebnisse und deren Visualisierungen befragt. Hierbei gab es sehr viele Ideen. Häufig wurde ein Wert gewünscht, der die Polarität des Sentiment über den gesamten Text aussagt (P1; P2; P5; P6). Darüber hinaus lässt sich der Sentiment Score auch für jeden einzelnen Satz oder des gesamten Korpus abbilden (P1; P5), sowie die Anzahl der Wörter die ein positives, negatives und neutrales Sentiment beinhalten (P2; P5). Sehr häufig wurde eine Art Verlaufskurve gefordert, die den Sentiment Score im Verlauf des Textes visualisiert (P1; P2; P3; P4; P5). Ebenfalls wurde häufig gewünscht, dass der Nutzer die Evaluation näher untersuchen kann. So wurde ein „Blick zurück“ zum „Ursprung“ des Sentiments (P2; P4) gefordert und eine Aufschlüsselung der Wörter, damit man die Auswertung nachvollziehen kann. Was auch für die Transparenz der Analyse wichtig ist (P3). Allgemein sollte der Nutzer die Möglichkeit haben, nachzuvollziehen, welches Lexikon einbezogen wurde, welche Parameter gewählt, was genau analysiert und wie der Aspekt der Datensicherheit gehandhabt wird (P3). Zudem könne die Analyse untersuchen, welche Schlüsselwörter mit Sentiment belastet sind (P4) und die Option bieten, Texte zu vergleichen und unter mehreren Perspektiven zu betrachten (P4; P5; P6). Die Ergebnisse der Exploration des Sentiments kann man hierbei auf unterschiedliche Weise darstellen. Hierbei ist es wichtig, dass auf einem Blick die Auswertung erkennbar ist (P2). So bieten sich Histogramme, Balkendiagramme, Kuchendiagramme und Wortclouds an, um die Daten zu visualisieren (P2; P4; P5). Darüber hinaus ist das Einbinden von primitiven Suchfunktionen möglich (P5). Das Layout der Result-Page soll hierbei eine Mischung zwischen Distant-Reading und Close-Reading bieten und die Texte in einer Reihenfolge anordnen, dass wenn diese mit dem Schema „Jahreszahl_Titel“ benannt wurden, auch in dieser Abfolge angezeigt werden (P5). Zudem sollte der Nutzer auch die Texte nach dem ermittelten Sentiment Score sortieren lassen können und den Text in Gruppen einzuordnen (P5). Eine weite-

re Möglichkeit, die der Verwender haben sollte, ist das Manipulieren der Ergebnisse. Auf diese Weise kann er in die Daten eingreifen und Änderungen vornehmen, beispielsweise wenn der Analyse ein Fehler unterlaufen ist, oder ganz simpel die Farben der Visualisierungen nach seinen eigenen Bedürfnissen anzupassen (P6). Es wurde die Möglichkeit, die Informationen interaktiv im Browser darzustellen als Bedürfnis genannt (P6), sowie das Herunterladen der Ergebnisse. Hierbei wurden vor allem die Dateiformate CSV (P1; P2; P4; P5; P6), XML (P1; P3; P5) und für das Herunterladen der Grafiken das Format JPEG (P2; P3; P4) genannt. Weitere genannte Formate sind TXT (P3) JSON (P2), PDF (P3) und EXEL (P4). Hierbei ist es wünschenswert, dass das XML CATMA kompatibel ist, damit dort die Auswertung weiterbearbeitet werden kann (P6).

5.1.1.5 Allgemeines zur Anwendung

Eine Person begründet, die Nichtnutzung von Anwendungen, welche eine Sentiment Analysis anbieten, damit, dass es zum einen kein einsteigerfreundliches Tool auf dem Markt gibt, welches einfach zu bedienen ist. Das Schema der lexikonbasierten Analyse ist dabei bekannt, allerdings ist der manuelle Abgleich der Wörter schwer und eine passende Software ist, bis auf LIWC, nicht vorhanden. Zusätzlich müsse man oft die Kommandozeile für die Benutzung des Tools verwenden oder die Anwendung bietet nur ein Eingabefeld mit einer begrenzten Anzahl von Zeichen, wobei der Text jedoch länger wäre (P3). Besonders ist es für die Community wichtig, dass die Anwendung über ein grafisches User Interface verfügt und eine gute Usability aufweist (P6).

5.1.2 Konkurrenzanalyse

Neben den Interviews mit potenziellen Nutzern und Experten auf diesem Fachgebiet, wurden darüber hinaus eine Konkurrenzanalyse als Anforderungsquelle hinzugezogen. Hierbei wurde darauf geachtet, dass eine Vielfalt von Anwendungen untersucht wurden. Zunächst wurden Tools „SentiStrength“, „Sentiment Analyzer“ und das Stanford Tool „Sentiment Analysis“ ausgewählt, die eine Webanwendung mit einem grafischen User Interface darstellen und Sentiment Analysis anbieten. Einige dieser Anwendungen stellen eine kostenfreie

Web-Demo für ihre kostenpflichtige API zur Verfügung. Hierbei wurde ausschließlich die Webanwendung betrachtet. Zudem existiert die Desktopanwendung „Lingmotif“, die eine Sentiment Analysis für die DH anbietet. Hierbei wurde nach einer Lizenz für die Untersuchung angefragt, jedoch wurde diese Anfrage nicht beantwortet, weshalb nur oberflächlich auf das Tool eingegangen wird. Die Anwendungen, welche eine Sentiment Analysis unterstützen, werden dabei in Bezug auf deren Funktionalitäten sowie der technische Hintergrund beleuchtet. Anschließend deren Usability anhand die von Jakob Nielsen 1994 aufgestellten Usability-Heuristiken für User Interface Design geprüft (Nielsen, 1994). Darüber hinaus gibt es eine Reihe von Webtools, mit denen DH-Forscher ihre Daten untersuchen und bearbeiten können. Da eine Analyse dieser Applikation bezüglich ihrer Usability und User Experience für die Anforderungen der zu entwickelnden Tools ergiebig sein könnte, wurde „Voyant“, welches eine Textanalyse bietet, betrachtet.

5.1.2.1 SentiStrength

SentiStrength¹⁸ ist ein, für die universitäre Forschung, frei verfügbares Tool, welches sowohl als online Anwendung zugänglich ist als auch heruntergeladen werden kann und in verschiedene Sprachen bereitstellt. Für die Untersuchung wird die online Version herangezogen (Abbildung 4). SentiStrength nutzt den lexikonbasierenden Ansatz, welches über 2310 SBWs verfügt, und dessen Algorithmus eine Rechtschreibkorrektur einbezieht. Hierbei werden auch Wörter und Emotionen berücksichtigt, die die Stimmung verstärken oder negieren. Darüber hinaus existieren Regeln für negativen Fragen, Ausrufezeichen, doppelte Interpunktion und Buchstaben. Die Skala für die positive und das negative Sentiment umfasst jeweils den Bereich von Eins bis Fünf, wobei Fünf die stärkste Ausprägung des Sentiments impliziert. Dabei wird eine negative Stimmung durch negative Zahlen repräsentiert. Die Anwendung unterstützt viele Sprachen wie Englisch, Deutsch und Spanisch (Thelwall, 2017; Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010). Der Nutzer hat hierbei die Möglichkeit für verschiedene Aspekte der Analyse, wie etwa anhand einer spezifischen

¹⁸ Verfügbar unter <http://sentistrength.wlv.ac.uk/> (zuletzt aufgerufen am 23.03.20)

Domäne, einen zu analysierenden Text in eines der vielen kleinen Textfelder einzugeben. Die vielen Inputs wirken jedoch ziemlich unübersichtlich und längere Texte ragen über die Felder hinaus und werden so nicht vollständig bei der Eingabe angezeigt. Während der Durchführung einer Analyse ist schwer zu erkennen, was der aktuelle Systemstatus ist. Da sich die Webseite vor und nach einer Analyse auf dem ersten Blick kaum verändert. Im oberen Bereich wird das Ergebnis der Analyse eingeblendet (Abbildung 5). Zusätzlich bekommt der Nutzer hierbei die Auswertung der einzelnen Sätze als Begründung der Klassifikation mitgeliefert. Dadurch lässt sich das Ergebnis nachvollziehen. Das System gewährt dem Nutzer viele Freiheiten für die Analyseeinstellungen. Jedoch wirkt dies leicht überfordernd, da das Anwendung dem Benutzer keine Hilfestellungen bietet. Ein weiteres Problem stellt die Fehlervermeidung dar, da der Nutzer eine leere Eingabe analysieren kann. Das GUI des online Tools wirkt sehr überladen. Durch den geringen Weißraum sowie die inkonstante Darstellung der GUI-Elemente, wie beispielsweise die Größe der Buttons und Input-Felder scheint alles unübersichtlich und unstrukturiert. Was sich auch auf die Lesbarkeit des Inhaltes und die Ästhetik auswirkt. Darüber hinaus wird der Inhalt des Tools nicht nach Semantik geordnet. Vielmehr befindet sich der ganze Sachverhalt, bis auf die Frequently Asked Questions, auf einer einzigen Seite. So verliert der Nutzer schnell den Überblick.

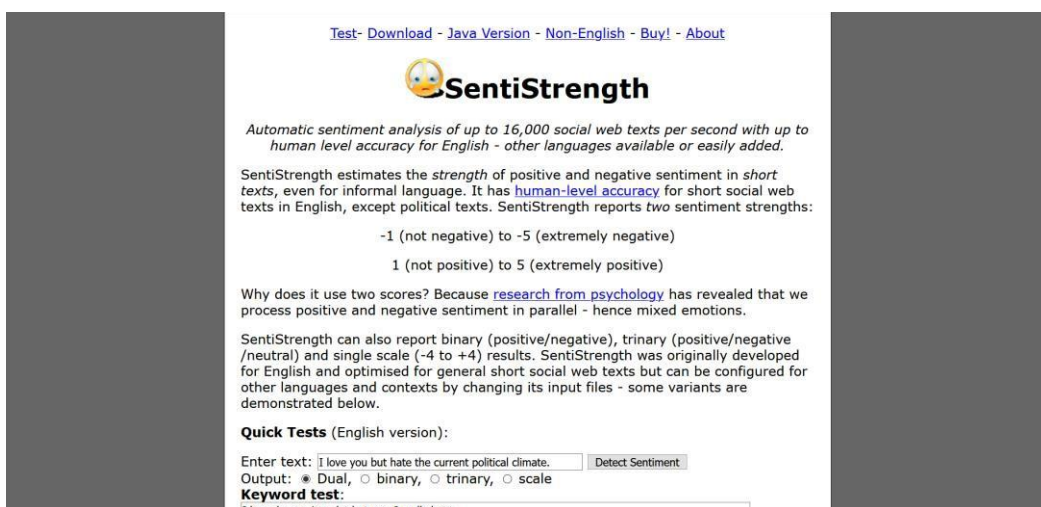


Abbildung 4: Screenshot der Anwendung SentiStrength

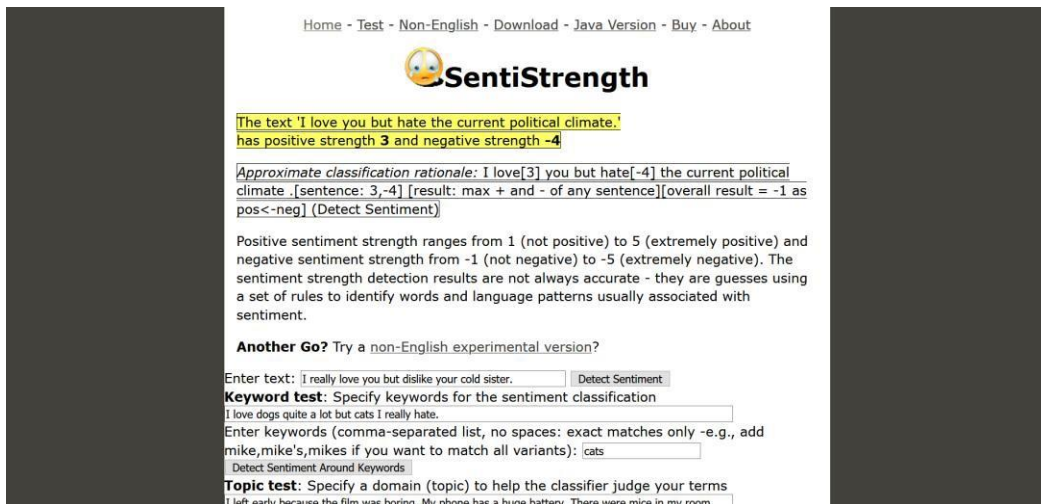


Abbildung 5: Screenshot der Anwendung SentiStrength - Ergebnis

5.1.2.2 Sentiment Analyzer

Der Sentiment Analyzer¹⁹ bietet den Nutzern ein frei verfügbares online Tool, mit dem man eine Texteingabe anhand einer Sentiment-Skala von -100 bis 100 analysieren kann, wobei -100 für das negativste Sentiment und 100 für das positivste Sentiment steht (Abbildung 6). Hierbei ist zu beachten, dass der ermittelte Wert, das Sentiment für die gesamte Eingabe repräsentiert und die Analyse nicht auf einem bestimmten Themenbereich ausgerichtet ist und am genauestens für englische Texte, welche nach 1990 verfasst wurden, funktioniert (Soper, n.d., sec. About) . Ruft der Nutzer die Seite auf, wird sogleich ein Beispieltext in einem Textfeld analysiert. Oberhalb des Textfeldes sind Hilfestellungen zur Nutzung der Applikation für den Benutzer zu finden. Da der Nutzer wenig Freiheiten bezüglich der Handhabung der Anwendung besitzt, gibt dies ihm ein Gefühl der Sicherheit hinsichtlich der Benutzung des Tools vor. Jedoch wird das Analysieren von leeren Eingaben nicht verhindert und das Untersuchen von sehr langen Inputs ruft einen Fehler hervor. Der ermittelte Sentiment-Wert wird durch ein Diagramm in Form eines Messgeräts visualisiert. Bei diesem Tool ist anzumerken, dass es dem User nicht möglich ist, zu sehen, wie das Sentiment berechnet wurde. Auch ist die Webanwendung sehr konsistent gestaltet, da der Nutzer sich immer auf derselben Seite befindet, deren Design sich nie ändert.

¹⁹ Verfügbar unter <https://www.danielsoper.com/sentimentanalysis/default.aspx> (zuletzt aufgerufen am 22.03.20)

Durch diese Übersichtlichkeit und einfache Bedienung der Texteingabe wird eine gute Usability erzeugt. Jedoch wirkt das Design der Webseite nicht besonders zeitlos.

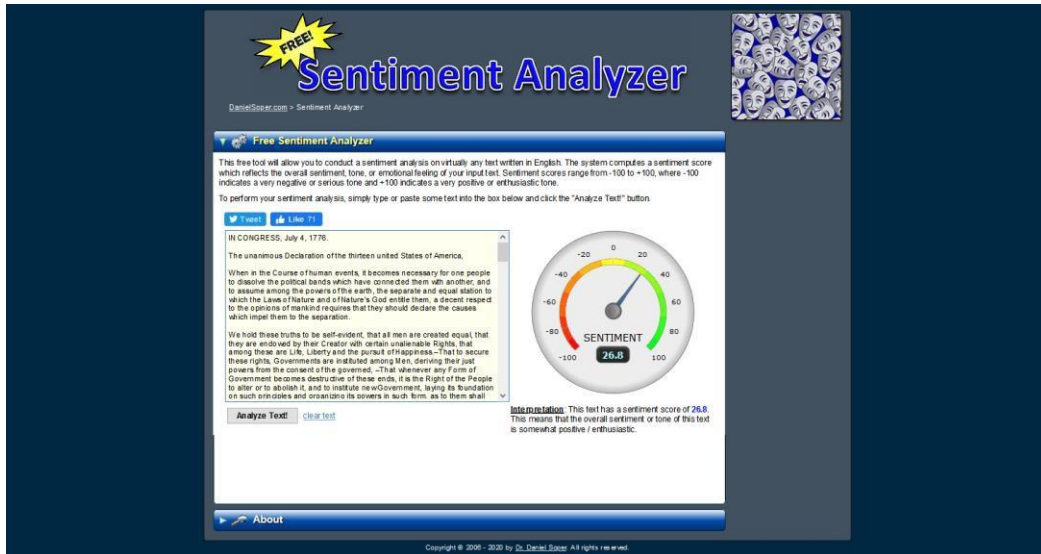


Abbildung 6: Screenshot der Anwendung Sentiment Analyzer

5.1.2.3 Stanford Sentiment Analysis Tool

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng und Christopher Potts der Stanford Natural Language Processing Group entwickelten mithilfe von rekursiven neuronalen Netzen der Stanford Sentiment Treebank, dessen Korpus auf Filmbewertungen basiert, ein Sentiment Analysis²⁰ Tool (Socher et al., 2013) (Abbildung 7). Der Nutzer hat hierbei die Möglichkeit, seinen Text in ein Textfeld einzugeben oder diesen als Datei hochzuladen. Allerdings ist beim Hochladen die Textlänge auf 200 Zeilen begrenzt. Diese Funktionalität des Hochladens eines Textes ist sehr positiv, sollte jedoch noch weiter ausgebaut werden. Zusätzlich hat der Nutzer die Möglichkeit, durch eine Checkbox, die erstellten Bäume in Binärer form darstellen zu lassen. Sobald der Text analysiert ist, erschien auf der Result-Page die Treebank der einzelnen Sätze der Eingabe, sowie die Polarität der dieser und deren Bestandteile (Abbildung 8). Hierbei kann der Verwender der Anwendung eine Rückmeldung geben, ob alle Benennungen korrekt waren. Zusätzlich ist ein

²⁰ Verfügbar unter <http://nlp.stanford.edu:8080/sentiment/rntnDemo.html> (zuletzt aufgerufen am 22.03.20)

Link vorhanden, mit dem der Nutzer die Ergebnisse herunterladen kann. Bezüglich der Heuristiken nach Nielsen, ist es dem User nicht möglich, einen leeren Text zu analysieren. Hierbei erscheint eine entsprechende Fehlermeldung. Durch einige Auswahlmöglichkeiten, ob der Nutzer einen Text eingibt oder eine Datei hochlädt, sowie die Darstellung der Treebank, erhält der Anwender ein Gefühl der Freiheit, ohne dass dieser dabei die Kontrolle verliert. Ebenfalls ist es möglich mithilfe von Shortcuts den Text in das Textfeld einzufügen. Im unteren Bereich der Demo gibt es ein Kommentarfeld, in dem der Nutzer ein Kommentar hinterlassen kann und so mit anderen Nutzern in Kontakt treten kann. Darüber hinaus finden sich auf der Webseite zahlreiche Informationen zu dem Projekt, wie etwa die Verantwortliche und das zugehörige Paper, sowie technische Hintergründe. Die Visualisierung der Anwendung ist sehr modern und ästhetisch. Es existiert genügend Weißraum, was die Applikation optisch aufwertet und die einzelnen Bereiche der Webseite sind klar strukturiert.

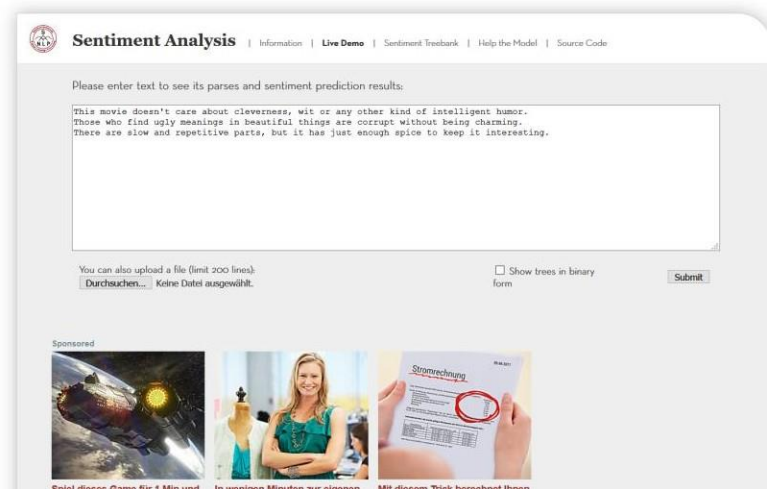


Abbildung 7: Screenshot der Anwendung Stanford Sentiment Analysis

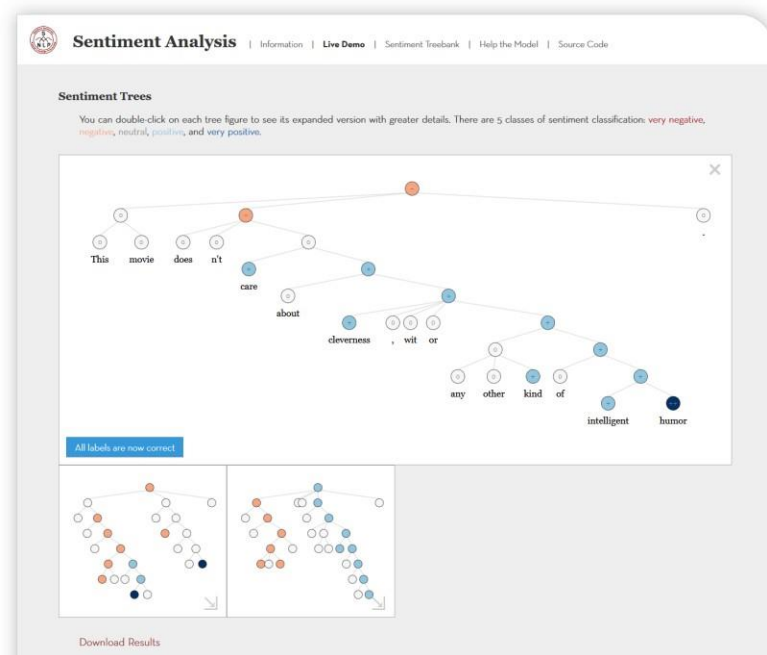


Abbildung 8: Screenshot der Anwendung Stanford Sentiment Analysis - Ergebnis

5.1.2.4 Lingmotif

Lingmotif (Moreno-Ortiz, 2017) ist eine Desktop-Anwendung, um eine Sentiment Analysis für die DH durchzuführen (Abbildung 9). Hierbei nutzen sie eine Scala zwischen -5 und 5, wobei hierbei auch der Nutzer ein eigenes Lexikon verwenden kann und zwischen verschiedenen Sprachen wählen kann. Um den Text einzugeben, hat der Nutzer die Möglichkeit einen Text in ein Textfeld einzugeben oder mehrere Texte hochzuladen. Lädt der Nutzer einen Text hoch, so erhält dieser Informationen über die Text Analyse als auch über die Sentiment Analysis. Zudem werden SBW im Text farblich in Rot, Grün und Gelb hervorgehoben, und je nach Ansicht, auch erweiterte Informationen des SBW zur Verfügung gestellt (Abbildung 10). Des Weiteren bekommt der Nutzer, vergleichbar mit dem Sentiment Analyzer, den Sentiment Score mithilfe von zwei Messgeräten visualisiert. Werden mehrere Dateien analysiert, so kann der Nutzer zwischen mehreren Modi, beispielsweise parallel oder zusammengefügt, wählen.

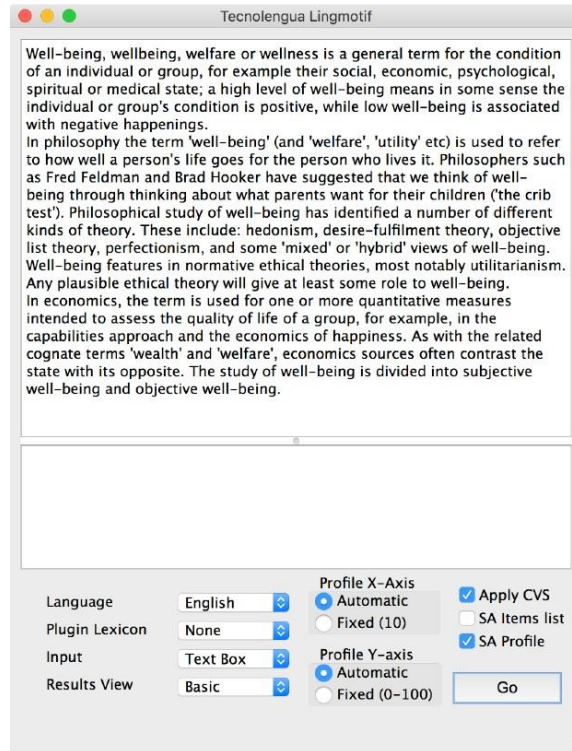


Abbildung 9: Lingmotif – Eingabe (Moreno-Ortiz, 2017, p. 74)

War is a state of **armed conflict** [val:-4; Type:MW; Lex:general; POS:NN] between societies.

It is generally characterized by **extreme collective aggression** [val:-7; Type:CVS; Lex:general; POS:JJ], **destruction** [val:-4; Type:S
W; Lex:general; POS:NN], and **usually high mortality** [val:-3; Type:CVS; Lex:general; POS:RB].

An **absence_of_war** [val:0; Type:CVS; Lex:general; POS:NN] is usually called " **peace** [val:4; Type:SW; Lex:general; POS:NN]".

Abbildung 10: Lingmotif – Ergebnis (Moreno-Ortiz, 2017, p. 75)

5.1.2.5 Voyant

Das Webtool Voyant²¹ bietet den Nutzern die eine Umgebung, um Texte zu lesen und zu analysieren. Hierbei stellt das Tool eine umfangreiche Dokumentation der Anwendung und deren Benutzung zur Verfügung. Dies ist vor allem bei der erstmaligen Verwendung sehr hilfreich. Möchte man einen Text analysieren, so hat der Nutzer die Möglichkeit, diesen in ein Textfeld einzugeben, ihn hochzuladen oder einen vordefinierten Korpus zu Analyse auszuwählen (Abbildung 11). Das Design der zugehörigen Seite ist hierbei sehr übersichtlich und mit viel Weißraum gestaltet. Nachdem der Nutzer seine Eingaben abgeschickt hat, gelangt man auf der Result-Page der Anwendung. Diese nimmt den kompletten Bildschirm ein, und die Inhalte sind in ein Raster eingegliedert.

²¹ Verfügbar unter <https://voyant-tools.org/> (zuletzt aufgerufen am 22.03.20)

5.2 Anforderungsanalyse

Nachfolgend werden die Ergebnisse der Anforderungserhebung analysiert. Hierbei werden Problemfelder identifiziert, notwendige und wünschenswerte Features extrahiert und Kernaussagen herausgearbeitet. Zunächst lässt sich festhalten, sich die einzelnen Datengrundlagen der Forscher in Umfang und Struktur sehr unterscheiden. Dies wirkt sich auch auf die Präferenzen punkto der Analyse aus.

5.2.1 Import der zu analysierenden Texte

Es herrscht Einigkeit darüber, dass ein Upload der zu untersuchenden Dateien gewünscht ist. Die hierbei primär vertretenden Dateiformate stellen reine Textdateien, sowie XML-Dateien dar. Während der Upload und das Analysieren der Textdateien sich einfacher gestalten lässt, könnte das Auswerten der XML-Dateien Schwierigkeiten bergen. Das liegt daran, dass in einem XML-Element der Aufbau der Datei wesentlich umfangreicher ist. Zwar besteht die Möglichkeit, nur den Text des gesamten Dokuments herauszulesen, jedoch könnte das jeweilige Schema bereits mit NLP Prozessen vorverarbeitet sein. So bietet das Deutsche Textarchiv etwa schon bearbeitete XML Dateien an, in denen das Lemma als Attribut gespeichert ist. Hierbei wäre ein entsprechender Workflow, in dem die Eigenschaft erhalten bleibt, von Nöten. Dieser birgt jedoch die Anforderung, dass dieser bei unterschiedlichen XML Schemas zuverlässig funktioniert. Es wichtig, dass das System mit den von anderen Anwendungen oder Ressourcen zur Verfügung gestellten Texten kompatibel ist. Also dass der Nutzer beispielsweise ein Text, welcher er vom TextGrid Repository heruntergeladen hat, genauso weiterverwenden kann. Die untersuchten Sentiment Analysis-Webanwendungen unterstützen die Eingabe nur durch ein Textfeld oder ein stark begrenzter Upload dessen geboten. Zum einen ist dies sehr aufwendig, zum anderen Verhindertes, dass mehrere Dokumente gleichzeitig analysiert und miteinander verglichen werden können. Darüber hinaus bieten die meisten der untersuchten Sentiment Analysis Anwendungen keine Möglichkeit,

eigene Lexika hochzuladen, was jedoch von den befragten Personen größtenteils gefordert wurde.

5.2.2 Export der Ergebnisse

Darüber hinaus ist es gewünscht, die Ergebnisse auch exportieren zu können. Häufig wurden hierbei die Datenformate CSV und XML, sowie JPEG genannt. Eine Person hat sich das Exportieren als PDF gewünscht, dies ließe sich umsetzen. Jedoch können sich alle darin befindenden Informationen auch von den anderen Dateiformaten dargestellt werden. Wie es gefordert ist, dass das System die Texte aus Anwendungen unverändert verwenden kann. So würde erwünscht, dass eine entsprechende Kompatibilität in die gegensätzliche Richtung ebenfalls existieren sollte. Hierbei wurde von einer Person gewünscht, dass das Heruntergeladene Ergebnis weiter in dem Annotationstool CAMTA weiterverarbeitet werden soll. Sollte diese Fiktionalität umgesetzt werden, muss in Erfahrung gebracht werden, was die Anforderungen an das Format sind und wie diese das Ergebnis sich in dieses integrieren lässt.

5.2.3 Natural language processing

Um die Analyse durchzuführen, sind bestimmte NLP Prozesse von Nöten. Einige Texte sind mit diesen NLP Prozessen bereits verarbeitet. Sei es durch den Nutzer selbst oder das zu explorierende Dokument wird von der Quelle bereits in dieser Form bereitgestellt. Um redundante Verarbeitungsschritte zu vermeiden und dem Anwender die Kontrolle über diese Prozesse zu geben, ist eine entsprechende Einstellung oder Auswahlmöglichkeit zu Beginn hilfreich, welche festlegt, ob das System bestimmte NLP Prozesse durchführen soll oder nicht. Sollte Lemmatisierung sich als einen benötigten Arbeitsschritt herausstellen, so wird eine genaue Betrachtung erforderlich sein, ob dies nicht der Benutzer oder ein speziell dafür ausgelegtes Programm übernimmt. Dies liegt daran, dass je nach Datengrundlage, dass unter anderem unterschiedliche Schreibweisen desselben Wortes existieren (Althoff, 1978, p. 100; Patzold, 2013; Schlaps & Schares, 2008) und dieser Verarbeitungsschritt zeitintensiv ist. Darüber hinaus stellt das NLTK den WordNet Lemmatizer ausschließlich für englischsprachige

Texte zur Verfügung und ein vergleichbarer Lemmatisierer für deutschsprachige Texte ist seitens des NLTK nicht vorhanden. Damit der Abgleich des Wortes mit dem Lexikon erfolgreich ist, kann es hilfreich sein, sowohl der nachzuschlagende Begriff als auch alle Schlüsselwörter in dem Lexikon in Kleinbuchstaben umzuwandeln. Auf diese Weise wird vermieden, dass ein SBW, welches im Wörterbuch in Kleinbuchstaben geschrieben wird, im Text am Beginn eines Satzes steht und damit nicht gefunden wird. Da die Nutzung einer Stopppword Liste die Performance verbessern könnte, ist eine Integrierung ebenjener sicher von Vorteil. Dies liegt daran, dass die Stopppwörter sehr häufig in Texten vorkommen und in der Regel keinen Mehrwert bieten. Das aufwendige Nachschlagen dieser überflüssigen Wörter in einem umfangreichen Sentiment-Lexikon, sollte deshalb vermieden werden. Es kann der Anwendungsfall auftreten, dass im Wörterbuch SBWs vorkommen, die in einem bestimmten Kontext ignoriert werden sollen. So ist das Anpassen oder erweitern einer Stopppword Liste hilfreich. Auf diese Art kann, der Nutzer bestimmte SBWs ignorieren, ohne ein gesamtes Lexikon zu erstellen.

5.2.4 Lexika

Vor allem wird von dem System eine gewisse Personalisierbarkeit bezüglich des zu verwendeten Lexikons erwartet, da je nach Domäne und Kontext, die SBW mit unterschiedlichem Sentiment behaftet sind. Aus diesem Grund ist es für die Nutzer sehr wichtig, dass diese die Möglichkeit haben, ein eigenes Sentiment-Lexikon zu verwenden oder ein bestehendes anzupassen. Hierbei ist das Ziel, dass die verwendeten Lexika alle dasselbe Schema verfolgen. Deshalb eignet sich eine CSV Datei, in denen in der linken Spalte die mit Sentiment behaftete Wörter stehen und in der rechten Spalte die zugehörigen Sentiments. Durch dieses simple Schema werden nur die nötigsten Informationen, die für die Analyse benötigt werden, gesammelt und so auch die Erstellung eines eigenen Lexikons vereinfacht.

5.2.5 Sentiment Analysis, deren Ergebnisse und Nachvollziehbarkeit

Ein weiterer bedeutsamer Punkt stellt die Nachvollziehbarkeit der Analyse dar. Dies ist wichtig, damit die Resultate zitierfähig sind, und es gibt auch dem Nutzer ein Gefühl der Kontrolle. Dabei wollen die Personen erkennen, wie die Anwendung zu den Ergebnissen der Analyse gekommen ist, was dahintersteckt und welche Parameter gewählt wurden. So können Nutzer auf diese Weise diese Resultate prüfen. Dies kann durch das Bereitstellen der gewählten Parameter und durch nähere Informationen der Analyse selbst geschehen. Bezüglich der Visualisierung dieser Ergebnisse konnten sowohl Gemeinsamkeiten als auch Unterschiede zwischen den einzelnen Personen festgestellt werden. Oft ist eine Art Verlauf gewünscht, um die Entwicklungen des Sentiments im Text zu visualisieren. Auf diese Weise lässt sich feststellen, ob es eine Entwicklung der Stimmung innerhalb des Textes, beispielsweise zu Beginn und am Ende des Dokuments, vorliegt. Eine Schwierigkeit könnte hierbei sein, den Text hierfür in passende Einheiten aufzuteilen. Wobei von einer Person gewünscht wurde, dass der Nutzer die Aufteilung selbst festlegen kann. Jedoch sollte hierbei im Laufe der Entwicklung genauer betrachtet werden, wie diese Einteilung aussehen sollte. Beispielsweise alle 15 Sätze oder alle 20% des Dokuments. Für die Analyse wird darüber hinaus ein Sentiment Score gewünscht. Welcher aussagt, wie das Sentiment über den gesamten Text ausgerichtet ist. Sowie einen Sentiment Score, der das Sentiment aller Texte zusammen repräsentiert. Ein weiterer Aspekt, ist das Untersuchen einer Basisemotion, wie etwa Zorn, Angst und Liebe, des Dokumentes. Untersucht man den Text nach einem Sentiment, welche als Valenz in Form einer Zahl repräsentiert wird, und nach einer Basisemotion, wobei das Sentiment in mehrere Emotionskategorien eingeteilt wird, so ist die Nutzung von zwei Lexika von Nöten oder ein Lexikon, welches sowohl die Valenz als auch die Emotionen abbilden kann. Somit muss jedes Wort in zwei unterschiedlichen Wörterbüchern abgefragt werden, was sich auf die Performance und Übersichtlichkeit der Evaluation auswirken kann. Zudem handelt es sich bei diesem Tool um eine Anwendung der Sentiment Analysis, und nicht der Emotions Analysis. Weiter zu analysierenden Daten ist beispielsweise die An-

zahl der positiven und negativen Wörter. Einige äußerten den Wunsch, als Funktion eine vergleichende Perspektive zu implementieren. Hierbei ist es von Nöten, dass mehrere Texte hochgeladen werden können. Auf der Result-Page sollen die Ergebnisse dann so geordnet sein, dass wenn der Dateiname mit einer Jahreszahl beginnt, nach dieser sortiert wird. Darüber hinaus sollte es möglich sein, die Texte nach ihrem Sentiment Score zu ordnen. Neben dieser Sortierung bietet sich auch an, die Texte in Gruppen einzuteilen und einzelne Dokumente oder Gruppen anschließend miteinander zu vergleichen. Hierbei sollen die Ergebnisse aus einer Mischung aus Distant-Reading und Close-Reading präsentiert werden. Sodass die Person zwischen mehreren Texten übersichtlich wechseln kann, um diese anschließend genauer zu betrachten. Dies ist sehr wichtig, da es zur Übersichtlichkeit beiträgt, wenn man nicht alle detaillierten Informationen aller Texte auf einmal angezeigt bekommt, sondern nur aktuell relevante Daten. Eine weitere mögliche Funktionalität ist das Durchsuchen der Inhalte auf der Result-Page, beispielsweise nach einem bestimmten SBW.

5.2.6 Usability

Ein weiterer wichtiger Aspekt ist die Usability, was sowohl in den Interviews als auch in der Related Work herausgearbeitet wurde. Auch wenn der Aspekt der Usability für die Methodiken der Medieninformatik etabliert sind, wo wird hier nochmals auf die Thematik eingegangen. Für die Usability ist ein guter Anhaltspunkt die Jakob Nielsen aufgestellten Usability-Heuristiken für User Interface Design. So sollte die Anwendung intuitiv bedienbar sein und lieber weniger Funktionen zur Verfügung stellen, anstatt dass der Nutzer zu viele Möglichkeiten geboten wird, was diesen überfordern könnte. Bei der Untersuchung der Tools fällt auf, dass detaillierte Informationen, beispielsweise eine Übersicht der SBWs, schwer zu finden sind. Vielmehr ist es nötig, in den entsprechenden Research Paper nach den gesuchten Auskünften zu suchen. Dies trägt nicht zu die von der Zielgruppe geförderte Nachvollziehbarkeit bei. Dabei würde sich anbieten auf der Webseite eine Seite einzurichten, die alle technische Hintergründe und Informationen übersichtlich darstellt. Dies wird vom Stanford Sentiment Analysis Tool gut umgesetzt. Auch der Sentiment Analyzer bie-

tet einen klar ersichtlichen Bereich (About) in dem die relevanten Informationen nachzulesen sind. Beim SentiStrength sind hierbei jedoch alle Hinweise auf einer einzigen Seite. Gerade die ersten beiden Applikationen sind nicht besonders ästhetisch gestaltet, was sich auf die Attraktivität des Tools auswirkt. Jedoch ist erkennbar, dass es sehr übersichtlich ist, wenn die einzelnen Funktionsbereiche, beispielsweise die Analyse an sich und technische Hintergründe, auf unterschiedlichen HTML-Seiten dargestellt werden. So findet sich der Nutzer gut zurecht und kann einfacher auf der Webseite navigieren. Bei zwei Tools wird auf der Result-Page die genaue Analyse der Sätze visualisiert. Dies ist praktisch und gibt dem Nutzer Kontrolle. Darüber hinaus ist das Verwenden eines Rasters um Informationen dazustellen und zu ordnen sicher hilfreich, da jede Kategorie ihr eigenes Feld besitzt und so von den anderen Informationen abgegrenzt werden.

5.3 Anforderungsspezifizierung

In diesem Kapitel werden die Nutzeranforderungen spezifiziert und in einem Anforderungsdokument festgehalten. Hierbei wird sich den Methodiken des UCD bedient (Eberhard-Yom, 2010). Zunächst wird eine hierarchische Taskanalyse (Abbildung 13) einer Sentiment Analysis angefertigt. Hierbei werden die wichtigsten Nutzerziele zu identifiziert und die Aufgaben in Subziele zu unterteilt. Auf diese Weise wird das Vorgehen des Nutzers während der Analyse untersucht. Als Erstes werden vom Anwender die betroffenen Daten hochgeladen. Dabei kann der Nutzer einen oder mehrere Texte auf seinem Endgerät auswählen. Zudem stehen ihm verschiedene Einstellmöglichkeiten zur Verfügung, mit der sich die Analyse nach seinen Bedürfnissen anpassen kann. Dazu gehört das Wählen eines vordefinierten oder eigenen Lexikons, das erweitern einer Stopword Liste sowie das Aufspüren von Negationen im Text. Nachdem der User alle benötigten Daten hochgeladen hat, führt das Programm im Hintergrund die Sentiment Analysis anhand der übergebenen Dateien und Parameter durch. Sobald diese Analyse abgeschlossen ist, werden dem Nutzer die Resultate auf einer Result-Page präsentiert. Auf dieser Seite erforscht der Anwender die Ergebnisse der Sentiment Analysis und dies im Anschluss nach Wunsch

bearbeitet. Dazu könnten beispielsweise das Bilden von Gruppen oder das Modifizieren der Evaluation eines Textes zählen. Möchte der Benutzer die Ergebnisse persistent speichern, so hat dieser die Möglichkeit, die Resultate in verschiedenen Dateiformaten herunterzuladen.

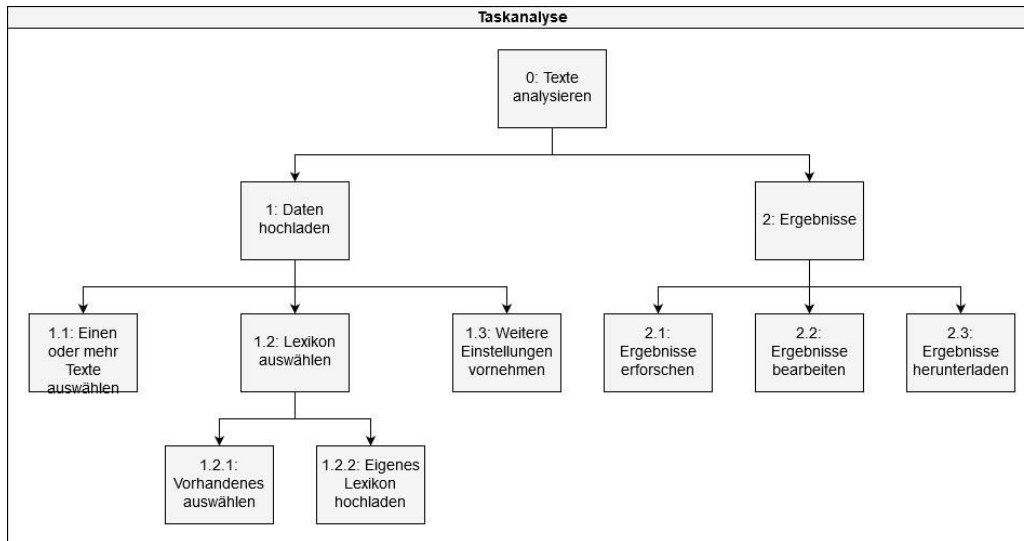


Abbildung 13: Taskanalyse eines Nutzers

Im zweiten Schritt wurden zwei Personas sowie der dazugehörige User Scenarios, je eine User-Story, sowie Use-Case-Diagramme angefertigt. Diese Personas stellen fiktional Benutzer der Anwendung dar, welcher über konkrete Eigenschaften und Bedürfnissen eines Nutzers verfügen. Auf diese Weise können die verschiedenen Wünsche der Nutzergruppen identifiziert werden.

Die erste Persona stellt Julia dar. Sie ist 28 Jahre alt und wohnt in Augsburg. Privat interessiert sie sich für neue technische Entwicklungen und geht gerne in der Natur spazieren. Nachdem sie ihren Master in Informatik letztes Jahr erfolgreich abgeschlossen hat fängt sie bald an beruflich als wissenschaftliche Mitarbeiterin an einer Universität im Bereich automatischer Sprachverarbeitung zu arbeiten und wird auch Kurse zum Thema Text Mining veranstalten. Da sie in ihrem zukünftigen Arbeitsumfeld ebenfalls mit der Sentiment Analysis in Berührung kommen wird, möchte sie sich schon davor mit dieser Methodik vertraut machen. So erhofft sie sich, einen guten ersten Eindruck bei ihren Kollegen zu hinterlassen. Bislang hat Julia noch keine Erfahrungen mit Sentiment Analysis gesammelt. Deshalb ist es ihr als Anfänger wichtig, dass die Anwen-

dung keine großen Voraussetzungen bezüglich der Methodik erfordert. Auf der Suche nach einer passenden Applikation ist sie auf dieses Tool gestoßen.

„Als Laie, möchte ich Erläuterungen über die Sentiment Auswertung, damit ich die Methodik der Sentiment Analysis einfacher verstehe.“

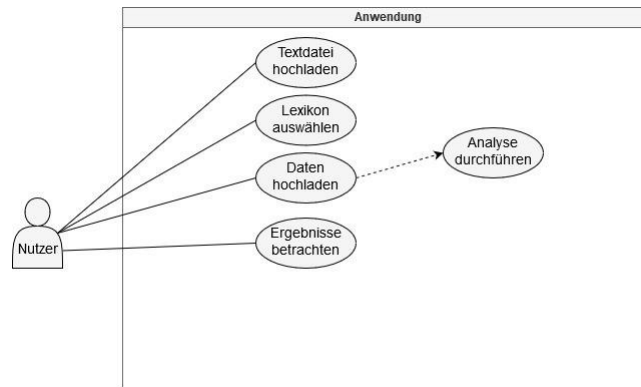


Abbildung 14: Use-Case-Diagramm - Julia

Die zweite Persona ist Michael, ein DH-Forscher und 45 Jahre alt. Er forscht und lehrt an einer Universität in Hamburg. Dort lebt er auch mit seiner Frau und seinen zwei Kinder am Stadtrand. Aufgrund seiner Arbeit ist Michael bereits mit der Methodik der lexikonbasierten Sentiment Analysis vertraut. Da er nicht besonders technikaffin ist, hat er diese Analyse bisher immer manuell durchgeführt. Nun möchte er es jedoch wagen und diese Sentiment Analysis computer-gestützt durchführen. In seinem aktuellen Forschungsprojekt untersucht Michael Tagebucheinträge eines bestimmten Autors, dessen Ergebnisse er auf einer Konferenz nächstes Jahr vorstellen möchte. Für die Analyse hat er sich mit den Texten bereits vertraut gemacht und diese inzwischen mit NLP Prozesse vorbereitet. Um die Sentiment Analysis durchzuführen, hat Michael ein eigens Sentiment-Lexikon für diese Einträge angelegt. Nun möchte er mithilfe des Tools die Analyse durchführen.

„Als Forscher, möchte ich die Grafiken herunterladen, um diese später in meiner Arbeit einbauen zu können.“

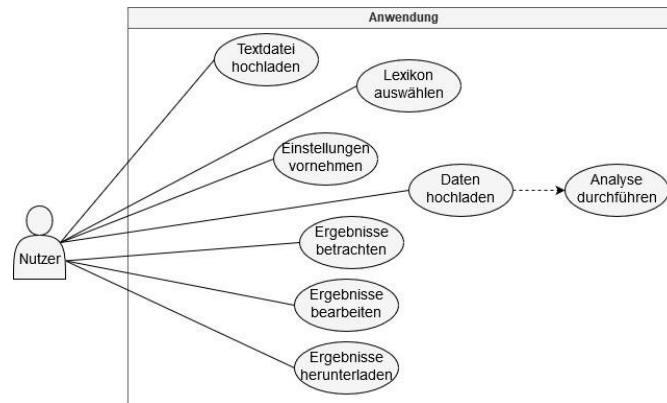


Abbildung 15: Use-Case-Diagramm - Michael

Nachdem die Nutzeranforderungen untersucht wurden, wurden alle Erkenntnisse in einer Anforderungsdokumentation festgehalten²². Hierbei wurden als Erstes die Anforderungen in funktionale und nicht-funktionale Anforderungen aufgeteilt. Anschließend wurden die Anforderungen je eine Priorität zugeordnet. Hierbei stehen Anforderungen mit Priorität Eins für Anforderungen, über die sich der Nutzer freut, wenn diese vorhanden sind. Anforderungen der zweiten Priorität sind, wurden oft von mehreren Personen genannt oder würden die Anwendung in ihren Fähigkeiten (funktional oder nicht-funktional) bereichern. Die letzte Einordnung bezüglich des Vorrangs sind Anforderungen der dritten Priorität. Diese Anforderungen sind unerlässlich für die Anwendung und der Zufriedenheit der Nutzer.

²² Die Anforderungsdokumentation findet sich in den elektronischen Daten unter /7_Anforderungsanalyse

6 Entwurf und Implementierung der Software

Nachfolgend wird die Implementierung der Software beschrieben. Hierbei wurde zunächst ein Design Entwurf der Anwendung entworfen, welcher anschließend mit dem Webframework Flask umgesetzt wurde. Dieses Kapitel ist in mehreren Teilbereichen aufgeteilt. Zunächst werden die getroffenen Annahmen und Vorbereitungen für die Implementierung der Applikation erläutert. Zum besseren Verständnis werden Designprozess der Benutzeroberfläche und technische Umsetzung des Tools voneinander getrennt dargestellt, obwohl sowohl die Verwirklichung als auch das Design der Anwendung ein iterativer Prozess repräsentiert, der während der gesamten Implementierung des Systems stattfindet. Um die Anwendung zu implementieren, wurden zwei Annahmen festgelegt, auf die auch Taboada und Kollegen (2011, p. 270) zurückgegriffen haben. Dabei wird angenommen, dass eine semantische Orientierung als Zahlenwert ausgedrückt werden kann und diese semantische Orientierung unabhängig zum Kontext steht.

6.1 Lexikonbasierter Ansatz und Schema des Lexikons

Da die Anwendung für viele unterschiedlichen Domänen verwendbar sein sollte, wurde sich für den lexikonbasierten Ansatz entschieden. Wobei die Möglichkeit bestehen soll, ein eigenes Lexikon für die entsprechende Domäne verwenden zu können. Dies hat der Grund, dass aktuelle Sentiment Lexika, welche auf die Gegenwartssprache ausgerichtet sind, nicht das Vokabular der zu untersuchenden Domäne ausgelegt sind und auch ein Mangel an annotierter Korpora existiert (Schmidt, Burghardt, & Wolff, 2018, Chapter 3). Aufgrund der getroffenen Annahmen ist eine Voraussetzung für die verwendeten Lexika, dass die dort gespeicherten SBWs über Polaritätsstärken verfügen. Häufig finden sich in den Lexika auch Informationen über die jeweilige Wortart. Diese Information wird jedoch nicht übernommen, da zum einen diese Benennung zwischen den einzelnen Lexika nicht einheitlich sind. Darüber hinaus wird, die Nutzung eines eigenen Lexikons komplizierter ist, da der Nutzer das Part-of-

speech-Tagging (POS-Tagging) selbstständig für sein Lexikon durchführen müsste. Um die zur Verfügung gestellten Lexika einheitlich zu strukturieren und eine einfache Verwendung zu gewährleisten, wurde folgendes CSV Schema in der Anforderungsanalyse festgelegt:

```
Sentiment_ bearing_word_1; Polaritätsstärke_1  
Sentiment_ bearing_word_2; Polaritätsstärke_2  
Sentiment_ bearing_word_3; Polaritätsstärke_3
```

Um positiv behaftete Wörter von negativen geladenen Worten zu unterscheiden, wurde festgelegt, dass negative SBWs auch über einen negativen numerischen Wert als Polaritätsstärke verfügen. Die Trennung von Vorkommastellen und Nachkommastellen kann hierbei durch ein Komma oder einem Punkt geschehen. Somit ist 1.58 gleichbedeutend mit 1,58. Um die einzelnen Datensätze zu trennen wird ein Zeilenumbruch verwendet. Dagegen wird für die Trennung der einzelnen Datenfeldern ein Semikolon genutzt. Für das BAWL-R Lexikon wurden zur Erstellung des vorgefertigten Lexikons der Begriff in Kleinbuchstaben und dessen durchschnittlichen Sentiment-Wert genommen. Für das SentiwS wurden die Wörter, dessen Flexionsvarianten und den Sentiment-Wert für die Definierung eines Lexikons genutzt.

6.2 Ermitteln der Testdateien für die Implementierung

Um die Anwendung zu Testen werden unterschiedliche Dateien für den Upload und für das Lexika verwendet. Dabei wurden XML- und Textdateien aus verschiedenen Ressourcen hinzugezogen, die auch der Zielgruppe zur Verfügung sehen. Zum einen wurde als Textquelle das von TextGrid zugängliche Repository ²³ und zum anderen das Deutsche Textarchiv ²⁴ ausgewählt. Aus TextGrid wurden fünf Dramen von Gotthold Ephraim Lessing ausgewählt. Hinzu kommen drei Prosa aus dem Deutschen Textarchiv von Franz Kafka. Diese liegen sowohl als einfache Textdatei vor, als auch als XML Dateien, wobei jeder Token ein Lemma Attribut besitzt. Zusätzlich stammen aus dem Deutschen Textarchiv 3 Briefe von Auerbach an Sanders, sowie zwei Novellen von

²³ Verfügbar unter <https://textgridrep.org/repository.html> (zuletzt aufgerufen am 22.03.20)

²⁴ Verfügbar unter <http://www.deutschestextarchiv.de/> (zuletzt aufgerufen am 22.03.20)

Leo Goldhammer und zwei Dramen Arthur Schnitzler. Zusätzlich wurde für das Testen bestimmter Aspekte passende Textdateien erstellt. Beispielsweise „Lea ist mir unsympathisch“ für das Kontrollieren des Negations-Präfixes „un“.

6.3 Implementierung und technischer Hintergrund

Im Fokus dieses Kapitels steht die Implementierung der Software, dessen technischer Hintergrund, sowie die dafür getroffenen Vorüberlegungen erläutert. Hierbei ist zu beachten, dass die beschriebene Implementierung den End-Stand der Anwendung zeigt.

6.3.1 Wahl des Webframeworks und Aufbau

Zu Beginn der Implementierung wurde überlegt, mit welchen Technologien die Webanwendung umgesetzt werden soll. Da sich die Verarbeitung des natürlichsprachigen Textes sich das NLTK anbietet und Python aufgrund der fehlenden Variablendeklarationen, Semikolons und Blockklammern das Schreiben des Codes besonders schnell von statten geht, wurde der Fokus daraufgelegt, dass das Back-End die Programmiersprache Python unterstützt. Um spätere Interaktionen mit dem Nutzer zu ermöglichen, wurde darüber hinaus JavaScript für die Umsetzung des Front-End bestimmt. So fiel die Wahl des passenden Webframeworks auf Flask. Dies hat mehrere Gründe. Zum einen ist es ein sehr bekanntes WSGI Webframework und bietet eine umfangreiche Dokumentation. Darüber hinaus ist es ein Mikroframework. Diese sind nicht so umfangreich wie Full Stack Frameworks und bieten beispielsweise keine Autorisierung. Da die Anwendung aber sehr minimalistisch in seiner Architektur gehalten wird, keine Datenbanken benötigt werden und auch keine Rollen zum Login und zur Administration vergeben werden, ist der Umfang eines Mikroframeworks für das System ausreichend. Somit besteht die Applikation aus mehreren Komponenten. Zum einen gibt es die .py Dateien, welche im Backend aufgerufen werden. Hierbei gibt es die settings.py. In dieser Datei sind alle Einstellungen gesammelt. Beispielsweise die maximale Dateigröße des File-Upload und die erlaubten Dateiformate. In dieser Datei wird ebenfalls definiert, was Stoppwörter, Interpunktionen und Negationen sind. In main.py sind die Routen zu den ein-

zelenen templates gespeichert, welche sich wiederum aufgrund des MVC Ansatzes in einem separaten Ordner names „templates“ befinden. In dieser Python Datei findet ebenfalls Back-End die Verarbeitung des natürlichsprachigen Textes statt, wobei hier auf die Klassen in der Datei `token_class.py` zurückgegriffen werden. Im Ordner „template“ befinden sich alle templates der Anwendung. Hierbei existiert eine `base.html`, deren Inhalt mit dem jeweiligen child template der einzelnen Page der Webapplikation gefüllt wird. Zuletzt sind im static Ordner alle statischen Dateien gesammelt. Hier findet sich die vorverarbeiteten Lexika, auf die in `main.py` zugegriffen werden können. Darüber hinaus sind in diesem Ordner alle JavaScript-Dateien, CSS-Dateien, Bilder und Plugins, welche sowohl JavaScript als auch CSS sein können, gespeichert. Das Programm wurde in Python 3.7.1 und in Flask 1.1.1, sowie der JavaScript Version ECMAScript 6, CSS und HTML5 umgesetzt.

6.3.2 Back-End: Sentiment Analysis

Im nachfolgenden Kapitel wird der für die Sentiment Analysis entwickelte Algorithmus, sowie die dafür getroffene Vorüberlegungen beschrieben. Für eine Sentiment Analysis werden häufig auch Valence shifters, wie etwa „sehr“, berücksichtigt. Da jedoch kein passender Datensatz mit deutschen Valence shifters gefunden wurde, wurden diese in der implementierten Sentiment Analysis nicht beachtet.

Vorüberlegungen

Im ersten Schritt wurden Vorüberlegungen bezüglich der Verarbeitung des Textes und der eingesetzten NLP Prozessen getroffen. Hierbei wird angenommen, dass der Text bereits in einer vorverarbeiteten Form vorliegt, und beispielsweise unerwünschte Textabschnitte entfernt oder Symbole in Wörter umgewandelt wurden. Dies hat den Grund, dass die Nutzer ihre Texte kennen und oftmals die Texte dem Benutzer schon vorverarbeitet zur Verfügung gestellt werden. Diese Annahmen stammen aus den geführten Interviews und der Inspektion der Testdateien. Sollten diese Schritte nicht durchgeführt werden, so hat der Nutzer die Möglichkeit im Front-End die Evaluation des Textes anzupassen.

Darüber hinaus wurden Überlegungen bezüglich Negationen, deren Erkennung und Auswirkung auf SBWs durchgeführt. Die hierfür verwendeten Negationen stammen von Harald Weinrich (2007, Chapter 8 f.) und beinhalten 22 Negationsmorpheme, 18 Negationspräfixe und 5 Suffixnegationen für deutschsprachige Texte. Um die Wörter zu definieren, auf die sich die Negation auswirkt, wurde auf eine traditionelle Methode zurückgegriffen, wobei eine Negation innerhalb eines festgelegten Bereiches erkannt wird und sich der Begriff zwischen der Negation und der ersten darauffolgenden Interpunktion befindet. Viele Studien zu diesem Thema untersuchen das statische Fenster in einem Rahmen zwischen zwei und fünf Begriffen und anhand verschiedener Algorithmen und Regeln, wobei sie sich überwiegend auf englischsprachige Texte fokussieren (Dadvar, Hauff, & de Jong, 2011; Farooq, Mansoor, Nongaillard, Ouzrout, & Qadir, 2017; Jia, Yu, & Meng, 2009). Deshalb wurde auf die Arbeit von Wiegand, Wolf und Ruppenhofer (2018) zurückgegriffen, die einen Bereich von vier Wörtern in beide Richtungen identifizierten. Bei der Ermittlung des Negationsbereichs spielen Interpunktionen eine entscheidende Rolle, wurden diese aus dem Python Modul „string“ entnommen und anschließend erweitert. Dieses Modul beinhaltet nicht alle Interpunktionen, beispielsweise die deutschen Anführungszeichen. Deshalb wurde die Liste im Laufe der Entwicklung weiter angepasst. Da auch nebenordnende Konjunktionen zwei Hauptsätze miteinander verbinden, werden hierbei die Wörter „und“, „oder“ und „weil“ wie Interpunktionen gehandhabt. Zuletzt wurde definiert, wie sich die Negation auf das Sentiment auswirkt. Hierbei wurde auf den Ansatz der switch Negation zurückgegriffen, da die unterschiedlichen Lexiken über eine unterschiedliche Spannweite der Polaritätsstärken verfügen. Weshalb sich das Verschieben des Sentiments um einen festgelegten Betrag nicht anbietet. Darüber hinaus werden in der Analyse einige Besonderheiten der Negation erkannt. Da „Nichts als X“ und „Nichts anderes/anderem als“ als Synonyme für „nur“ und „ausschließlich“ verstanden werden, werden diese Formulierungen für die Erkennung von Negationen nicht berücksichtigt. So wird die Negation in „Nichts als Klagen“ ignoriert, da dies ein Synonym für „Nur Klagen“ ist. Zusätzlich wird eben-

falls die nebenordnende Konjunktion „weder X noch Y“ erkannt (Weinrich, 2007, Chapter 8 f.). Für das Beispiel „weder schön noch intelligent“ wird sowohl „schön“ negiert als auch „intelligent“.

Um die Dokumente zu analysieren, ist es von Nöten, die Texte zunächst in einzelne Sätze und anschließend erst in Wörter zu segmentieren. Auf diese Weise lassen sich Untersuchungen auf der Satzebene umsetzen. Neben der Tokenisierung wird auch die Lemmatisierung als einen sinnvollen NLP Prozess für die Sentiment Analysis angesehen, da es sich gezeigt hat, dass dieser Vorgang die Leistung der Analyse verbessert (Schmidt, Burghardt, & Dennerlein, 2018). Hierbei wird das Wort in dessen Grundform zurückgeführt. Jedoch vermindert dieser Prozess die Performance der Analyse, wodurch sich, dessen Laufzeit verlängert. Aus diesem Grund erscheint es von Nöten, das Berücksichtigen von Stoppwörtern während der Sentiment Analysis, da diese über wenig Relevanz verfügen. Eine Liste deutscher Stoppwörter entstammt der Stoppwortliste des NLTK, wobei Verneinungen hierbei entfernt wurden, da diese separat erkannt werden. Somit ergibt sich eine Liste von deutschen Stoppwörtern mit einer Länge von 223 Wörtern. Man kann bei einer Sentiment Analysis die Groß- und Kleinschreibung berücksichtigen oder nicht. Da die Groß- und Kleinschreibung in der deutschen Sprache jedoch Einfluss auf die Bedeutung der Formulierung haben kann, gut zu sehen an dem Beispiel „Der gefangene Floh“ und „er Gefangene floh“, sollte dieser Vorgang als optionale Möglichkeit zu Verfügung gestellt werden. Allerdings ist es nicht möglich, dieser Hergang komplett auszusparen. Da, sollte eine Lemmatisierung durchgeführt werden, da wie sich zeigte, die Gefahr besteht, dass Verben nominalisiert werden („lief“ → „das Laufen“). Zudem sind alle SBWs des Lexikons BAWL-R in Kleinbuchstaben notiert, weshalb hier auch die Möglichkeit bestehen muss, Groß- und Kleinschreibung im Text nicht zu beachten, da sonst beispielsweise Nomen im Lexikon nicht gefunden werden. Ebenfalls wird die Überprüfung auf Negationen und Stoppwörter automatisch die Groß- und Kleinschreibung nicht beachtet, da diese bereits alle in Kleinbuchstaben vorliegen.

Für die Sentiment Analysis und deren Auswertung ist es notwendig, dass die Applikation erkennt, ob das vorliegende Token ein Wort darstellt. Diese Information wird beispielsweise für das darstellen von Anteilen in den Kreisdiagrammen, für Nutzerinteraktionen und für das Normalisieren der Scores anhand der Wortanzahl im Text benötigt. Für das Identifizieren eines Tokens als Wort wird hierbei auf den Regulären Ausdruck zur Erkennung eines Begriffes zurückgegriffen. Um ein Wort als dieses zu erkennen, muss das Element aus mindestens einem Wordzeichen bestehen.

Zusätzlich können aufgrund des Datentyps zur Darstellung von Fließkommazahlen geringfügige Unregelmäßigkeiten zur Berechnung von Sentiment-Werten auftreten. Aus diesem und aus Darstellungsgründen werden alle Fließkommazahlen der Applikation grundsätzlich auf fünf Nachkommastellen gerundet.

Algorithmus zur Sentiment Analysis

Die Sentiment Analysis wird zentral in `main.py` durchgeführt. Hierfür werden in der `settings.py` vorgenommene Definierungen von Negationen, dem Negationsbereich und Interpunktionen importiert. Zusätzlich wird auf die in `token_class.py` zugegriffen, in der die Basisklasse `Token`, sowie die davon abgeleiteten Klassen `PunctuationToken`, `NegationToken` und `SentimentToken` befinden. Der Algorithmus zur Sentiment Analysis der Texte beinhaltet mehrere Schritte. Der nachfolgende Code zeigt den schematischen Aufbau des Codes, mithilfe die, an die Result-Page zu übergebenen Daten ermittelt werden. Zuerst werden die Parameter bestimmt (Schritt 1) und anhand derer das gewählte Lexikon sowie die Stoppwörter initialisiert (Schritt 2). Wobei hierbei auch die manuelle Anpassung der Stoppwort-Liste berücksichtigt wird. Im nächsten Arbeitsschritt wird mit einer Schleife die Sentiment Analysis für jedes einzelne Dokument durchgeführt (Schritt 3) und das Ergebnis einem Objekt-Array beigelegt (Schritt 4). Zuletzt wird die Result-Page gerendert und dabei sowohl der erzeugte Objekt-Array als auch die vom Nutzer definierten Parameter übergeben. Im Front-End werden die Informationen dieses Objekte-Arrays und die

gewählten Parameter weiterverarbeitet, um anschließend dem Anwender die gewünschten Funktionalitäten zu bereitzustellen.

1. Parameter ermitteln (Upload Optionen)
2. Lexikon und Stoppwörter initialisieren
3. Sentiment Analysis für jedes einzelne Dokument durchführen
Ergebnis für jedes Dokument: Objekt mit Dateinamen, Id und Auswertung (Array mit einem HTML-String für jeden Satz)
4. Alle Sentiment Analysis-Ergebnisse werden in einem Objekt-Array gespeichert
5. Rendern der Result-Page mit den Gewählten Optionen und den explorierten Dateien als übergebenes Parameter-Objekt

In der Sentiment Analysis (Schritt 3) wird der Text mit Hilfe des NLTK zuerst in Sätze tokenisiert. Diese werden anschließend nacheinander betrachtet, wobei der Satz nochmals in einzelne Wörter sequenziert wird. Jeder Token wird daraufhin nach seinen Eigenschaften untersucht. Dieser Vorgang beinhaltet mehrere Schritte, welcher sich je nach gewählten Optionen sich leicht unterscheiden kann. Zunächst wird festgestellt, ob das Wort einer Interpunktion oder Negation darstellt und sollte dies der Fall sein, eine Instanz mit der entsprechenden Subklasse PunctuationToken oder NegationToken erzeugt. Alternativ wird der Token mit den Stoppwörter abgeglichen. Wenn der Token nicht als Stoppwort identifiziert wurde, wird er mit dem initialisierten Wörterbuch verglichen. Mögliche Schritte können hierbei das lowercasen des Wortes oder die Lemmatisierung, oder das Untersuchen der morphologischen Struktur auf negierende Präfixe oder Suffixe sein. Sollte ein entsprechendes SBW im Lexikon vorhanden sein, so wird eine Instanz der Subklasse SenitmentToken erzeugt, welche über Attribute des Lexikonschlüssels, den Sentiment-Wert und, sollte es eine Negation aufgrund seiner morphologischen Struktur beinhalten, den zu negierenden Token verfügt. Jedes dieser Elemente wird in einem initialisierten Array gespeichert. Sollte das identifizieren von Negationen gewünscht sein, so wird dieser

Array anschließend durchlaufen, wobei NegationToken und die in den Vorüberlegungen beschriebenen Sonderfälle identifiziert werden. Für jede erkannte Negation wird die Polaritäten der sich darauf beziehende SentimentToken angepasst. Dabei darf zwischen dem NegationToken und dem SentimentToken kein PunctuationToken liegen. Anschließend wird der Array, in dem sich der Token befinden in einen einzelnen HTML-String umgewandelt. Dieser String stellt nun wieder einen kompletten Satz dar und jedes Token verfügt dort über explorierte Meta-Daten. So ergibt sich ein Array mit HTML-Strings für jeden Satz. Nachdem ein Text analysiert wurde, wird ein Objekt erzeugt, welches über den Dateinamen, die Id der Datei und dem Array mit den explorierten HTML Sätzen verfügt (Ende der Sentiment Analysis). Die Id der Datei setzt sich aus „file_“ und dem Index der Datei zusammen.

Eingesetzte Bibliotheken

1. NLTK

Diese Bibliothek wird für die Verarbeitung des natürlichsprachigen Textes hinzugezogen. Sie findet in der Datei „main.py“ Einsatz und ist unter anderem für die korrekte Tokenisierung in Sätze und Wörter verantwortlich. Darüber hinaus wurde aus dieser Bibliothek die Stoppwortliste entnommen und für die Software angepasst.

2. textblob_de

Um den Nutzer die automatische Lemmatisierung des Textes zu ermöglichen, wurde diese Bibliothek ausgewählt. Die Auswahl einer passenden Bibliothek stellte sich als nicht sehr trivial heraus, da diese Bibliothek deutschsprachige Texte unterstützen muss und die Performance nicht zu stark beeinflussen. Somit wurde diese Bibliothek ausgewählt.

3. Pandas und csv

Beide Bibliotheken finden im Kontext der Verarbeitung des Lexikons Einsatz und werden in Abhängigkeit davon genutzt, ob der Nutzer ein persönliches Lexikon verwendet oder ein vom System bereitgestelltes Lexikon nutzt.

6.3.3 Front-End

Im Front-End werden im Rahmen des MVC-Musters die Funktionalitäten in mehrere Dateien aufgeteilt. Hierbei werden in `sentimentAnalysis.js`, dem Controller, Nutzeraktionen abgefangen und dementsprechend mit View und dem Model agiert (Abbildung 16). Im Model, zu finden in der Datei `sentimentAnalysis_model.js`, werden die explorierten Ergebnisse gehandhabt. Hier sind alle Sentiment Analysis Resultate, sowie Operatoren, um diesen Daten zu manipulieren, zu verwalten und mit ihnen zu arbeiten gespeichert. Dabei wird auf die Basisklasse `SentimentEntity` und ihre Subklassen `SentimentFolder` und `SentimentFile` zurückgegriffen. Für jede analysierte Datei wird ein `SentimentFile` und für jeden erstellen Ordner ein `SentimentFolder` instanziiert. Diese Klassen sind in der Datei `class_sentiment_entity.js` zu finden. Um einzelne Sätze in einer `SentimentFile` zu definieren, wird hierbei die Klasse `Sentence` genutzt. Diese Klasse ist in `class_sentence.js` zu finden und beinhaltet Informationen der Sentiment Analysis auf Satzebene. Darüber hinaus werden die Grafiken mithilfe `sentimentAnalysis_graph.js` erstellt. Hierbei geben `get`-Methoden ein entsprechendes `svg`-Element, anhand der übergebenen Parameter, zurück. In der View, welche sich in `sentimentAnalysis_view.js` befindet, werden die vom Controller empfangene Anweisungen in das DOM eingespeist. Zentral Konfigurationen des Front-End sind in der Datei `sentimentAnalysis_config.js` dokumentiert. Hierbei sind etwa die Farben für Grafiken, die festgelegte Anzahl der Nachkommastellen und Benennungen der Export-Dateien festgehalten. Somit ist eine schnelle und einfache Anpassung, auch von Personen, welche nicht mit dem Code vertraut sind, gewährleistet.

```
let newNormScoreFolder = model.updateFolder(folderId, filesIds);
view.displayFolderScore(folderId, newNormScoreFolder);
```

Abbildung 16: Aufruf von Model und View im Controller

Eingesetzte Bibliotheken

Nachfolgend werden die genutzten Bibliotheken zur Implementierung des Front-Ends beschrieben. Hierbei werden die einzelnen Bibliotheken je nach Verwendung gruppiert

1. jQuery und jQuery MultiSelect

Einige Bibliotheken und Funktionalitäten die JS Bibliothek „jQuery“ miteingebunden. Zudem wurde für Mehrfacheingaben die Bibliothek „jQuery MultiSelect“ verwendet.

2. Font Awesome, Popper und Tippy

Da die Benutzeroberfläche verschiedene Symbole beinhaltet, werden diese mithilfe von „Font Awesome“ angezeigt. Diese bietet den Vorteil, dass sie sehr einfach einzubinden und anzupassen sind. Aus der gleichen Motivation wurden Tooltips, die dem Nutzer an verschiedenen Stellen Informationen bereitstellen oder mit denen der Benutzer interagieren kann, mithilfe von „Tippy“ umgesetzt. Diese Tooltip- und Popover-Bibliothek, unterstützt von Popper, bietet dem Entwickler die Möglichkeit auf simple Weise Tooltips zu erstellen und nach seinen Wünschen

3. D3 und D3-cloud

Beide Bibliotheken werden dazu genutzt, um Grafiken zu visualisieren. Während D3-cloud speziell für das Erstellen von Wordclouds benötigt wird, werden alle restlichen Visualisierungen mit D3 umgesetzt. Diese werden in `sentimentAnalysis_graphs.js` für die Erstellung der Diagramme eingesetzt.

4. canvg, JSZip und FileSaver

Diese Bibliotheken werden verwendet, um die erstellten Grafiken zu exportieren. Canvg parst hierbei die SVG in ein Canvas Element. So ist es möglich Grafiken als PNG (mithilfe der data URI) in einer Zip Datei zu speichern und herunterzuladen. Hierbei wurde PNG aufgrund dessen transparenten Hintergrund als Dateiformat gewählt.

6.3.4 Einbinden weiterer Sprachen zur Sentiment Analysis

Die Sentiment Analysis wurde auf eine Art implementiert, sodass es in Zukunft sehr einfach sein wird, weitere Sprachen zur Verarbeitung von natürlichsprachigen Texten einzupflegen. Das nachfolgende Codebeispiel zeigt hierbei beispielhaft das Definieren der zu verwendeter Sprache und der Stoppwörter.

```
settings.py:
```

```
LANGUAGE_LEXICON = {"german": ["SentiWS", ... ]}
STOPWORDS = {'german': ['aber', 'alle', ... , 'zwischen']}

main.py:
from settings import LANGUAGE_LEXICON, STOPWORDS
...
# define language
for key_lang, values_lexicon in LANGUAGE_LEXICON.items():
    if selected_lexicon in values_lexicon:
        lang = key_lang
...
# stopwords
stopwords = STOPWORDS[lang]
...
stopwords.extend(personal_stopwords)
```

6.4 Designprozess und Funktionalitäten

In diesem Kapitel werden das Design und die Funktionalitäten der Anwendung beschrieben. Dabei wurde im Rahmen des User-Centered-Design der Nutzer durch formative und einer summative Usability-Tests mitebezogen. Dieses Kapitel gliedert sich in vier Teile. Zunächst wird der erste Designprozess, beginnen von Scribbles bis hin zur ersten Implementierung, beschrieben. Anschließend folgen die formative und summative Evaluation mit dessen Ergebnissen und daraus resultierenden Verbesserungen. Zuletzt wird auf die Funktionalität der entwickelten Anwendung eingegangen.

6.4.1 Erster Designprozess

Zunächst wurden mithilfe von Scribbles Low-Fidelity Prototypen der Anwendung angefertigt. Aufgrund der Anforderung, dass die Web-Anwendung über mehrere Seiten strukturiert sein soll, wurde eine Menüleiste im oberen Areal angefertigt. Die Nutzung einer solcher Leiste hat einige Vorteile. Durch die farbliche Hervorhebung des aktivierten Menüpunktes, weiß der Nutzer sofort, wo er sich befindet. Darüber hinaus bietet das Menü auf den ersten Blick bereits einen klaren Überblick über die Strukturierung der Seite. Anschließend wurde

das Design mithilfe der Web-Anwendung Moqups in einen Medium-Fidelity Prototypen²⁵ umgesetzt. Um Personen, die der deutschen Sprache nicht mächtig sind, die Anwendung benutzbar zu gestalten, wurde das GUI in Englisch verfasst.

Strukturierung der Webseite

- Home-Page (Abbildung 17): kurze Vorstellung der Anwendung, sowie Kontaktinformationen
- Information-Page: Informationen zur Benutzung der Anwendung, Ziel ist, dass der Nutzer mithilfe dieser Page einschätzen kann, ob die Applikation für sein Anwendungsfall der Analyse brauchbar ist und wie sie verwendet wird
- Upload-Page (Abbildung 18): Zugang über den Menüpunkt „Sentiment Analysis“, hochladen der Texte, Auswahl des Lexikons, anpassen der Analyse für fortgeschrittene Nutzer (NLP Prozesse, erstmal verdeckt)
- Result-Page (Abbildung 22): Ergebnis der Sentiment Analysis, Nutzer wird von der Upload-Page zur Result-Page weitergeleitet, quantitative Übersicht aller Dokumente, qualitative und detaillierte Darstellung der Texte

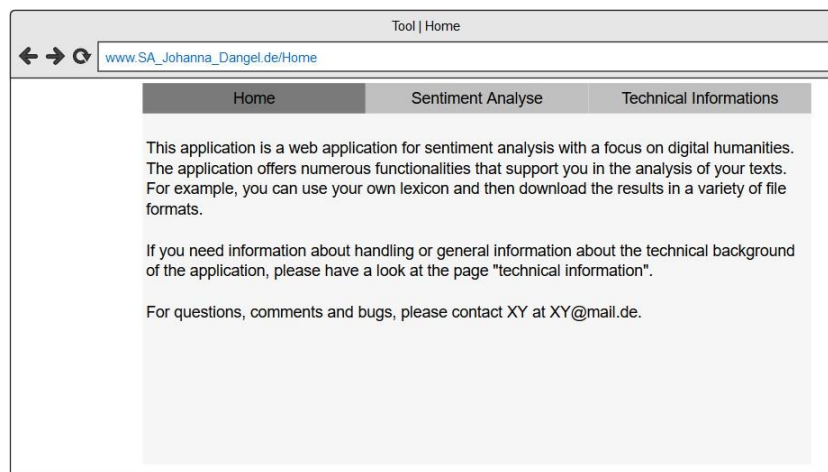


Abbildung 17: Medium-Fidelity Prototypen - Home

²⁵ Klickbarer Prototyp verfügbar unter

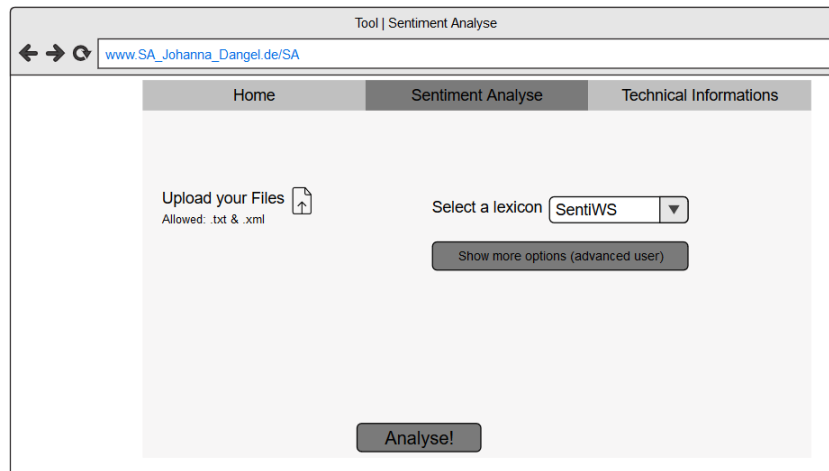


Abbildung 18: Medium-Fidelity Prototypen - Upload

Ist die Analyse durchgeführt worden, so gelangt der Anwender auf eine Seite, auf der die Resultate angezeigt werden. Die Grundidee war, den das Close- und Distant-Reading in jeweils einen eigenen Bereich einzuordnen. Zusätzlich soll die Close-Reading Zone zwischen den Visualisierungen und der Auswertung des Textes unterteilt sein. Alle Bereiche sollen unabhängig voneinander scollbar sein. Zudem wurden die gesetzten Parameter der Analyse als Notizzettel am linken Rand dargestellt.

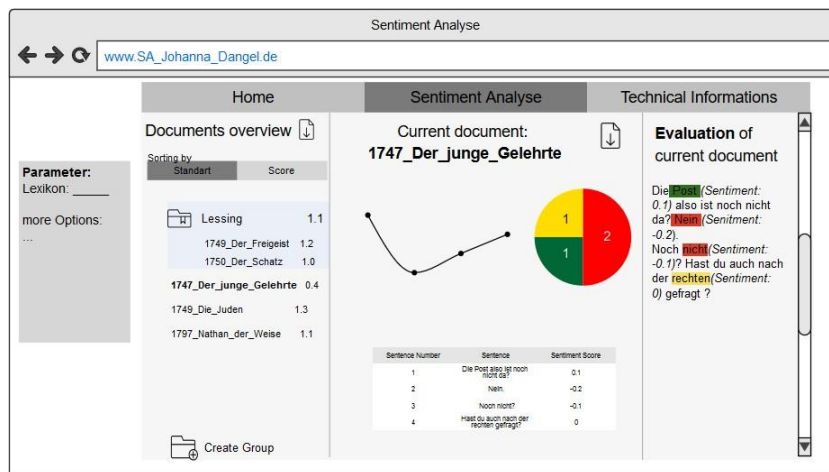


Abbildung 19: Medium-Fidelity Prototypen - Result-Page

Im nächsten Schritt wurde das Design nochmals verfeinert und in einem High-Fidelity Prototypen²⁶ umgesetzt, sodass bereits ein grundlegendes Design erkennbar ist. Als Erstes wurde sich mit der Auswahl der Hausfarbe und die Farbe der Visualisierungen beschäftigt. Damit auch Nutzer mit einer Rot-Grün-

²⁶ Klickbarer Prototyp verfügbar unter

Sehgeschwäche die Anwendung uneingeschränkt nutzen können, werden positive Sentiments blau, neutrale Sentiments gelb und negative Sentiments rot visualisiert. Im Allgemeinen war das Ziel, dass nicht zu viele Farben verwendet werden, da ein sehr buntes Design den Nutzer ablenken kann. Als Hausfarbe wurden deshalb Blautöne gewählt. Als erstes wurde die „Home“ in „About“ umbenannt. Dies hat der Grund, dass die Seite allgemeine Informationen über die Anwendung bereitstellt. Der Menüpunkt „Technical Information“ wurde in Information abgekürzt, weil dort nicht nur technische Informationen zu der Applikation vermerkt sind (Abbildung 20). Vielmehr soll der Nutzer hier auch Informationen zur genauen Benutzung der Anwendung finden.

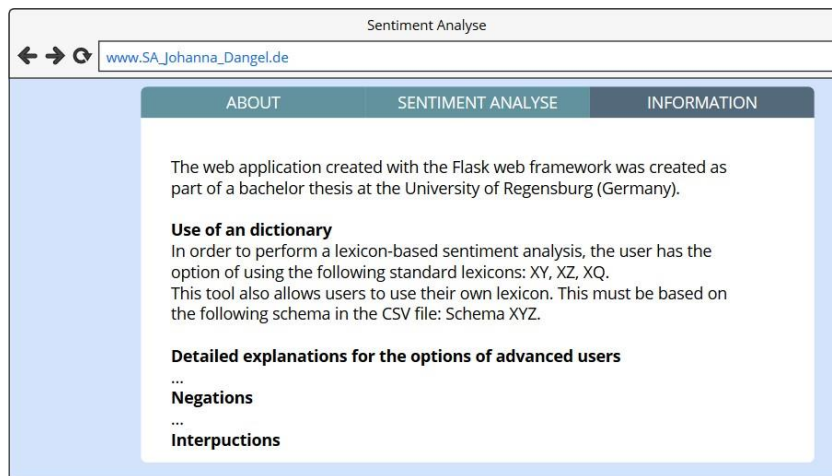


Abbildung 20: High-Fidelity Prototypen - Information-Page

Ebenfalls wurde die Upload-Page (Abbildung 21) und die Result-Page (Abbildung 22) überarbeitet. Hier wurde sowohl die Farben und Schriften der Interfaces angepasst als auch ein Tooltips eingebaut. In der Result-Page wurde die Unterteilung zwischen Distant-Reading und Close-Reading nochmals durch eine stärkere und farblich hervorgehobene Trennungslinie verstärkt. Auch die Darstellung von Ordnern wurde leicht angepasst. Hierbei wird in der ersten Ebene, dort wo der Ordnername steht, farbig hinterlegt. Alle Elemente, die sich in diesem Ordner befinden, werden durch einen linken Balken, der die gleiche Farbe besitzt wie die Hintergrundfarbe des Ordnernamens, gekennzeichnet. Ebenfalls wurde der Close-Reading Bereich angepasst. Neu ist ein Header, in dem der Dateiname steht und die ganze Breite des Close-Reading Areal einnimmt. Darunter wird der Bereich in zwei Spalten eingeteilt. Auf der linken

6. Entwurf und Implementierung der Software

Seite werden die Grafiken und Ergebnisse dargestellt. Auf der rechten Seite hat der Nutzer die Möglichkeit, sich die Evaluation des Textes anzusehen und diese gegebenenfalls zu bearbeiten. Tut er dies, so soll die linke Seite anhand der neuen Daten angepasst werden.

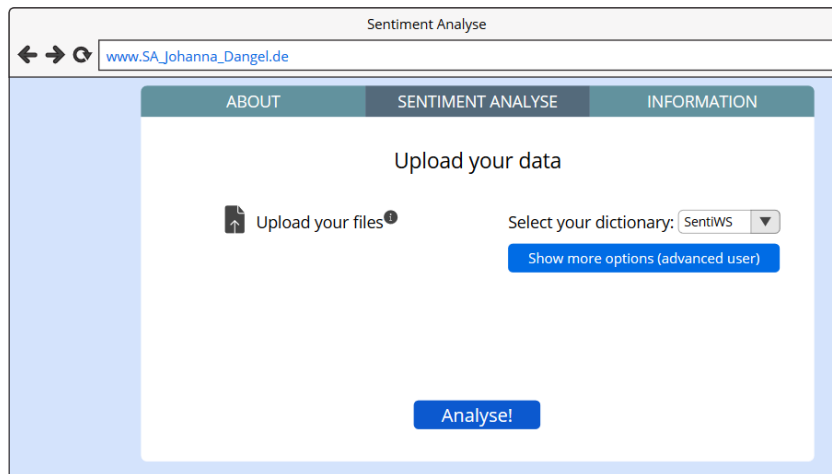


Abbildung 21: High-Fidelity Prototypen - Upload-Page

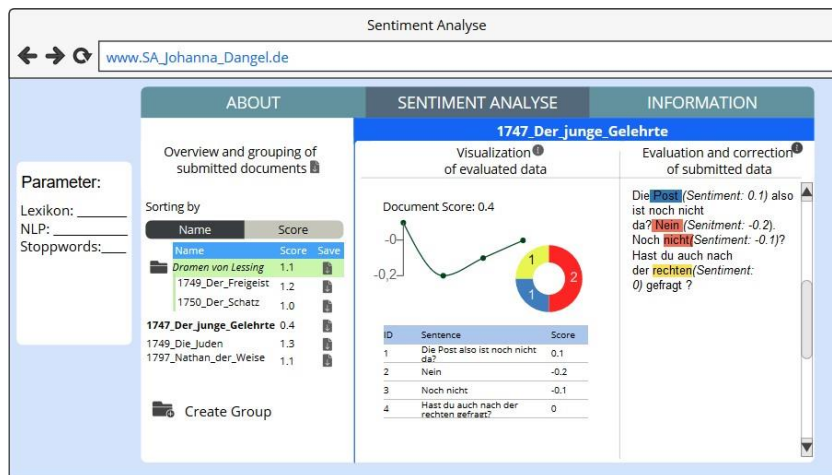


Abbildung 22: High-Fidelity Prototypen - Result-Page

Während der Implementierung des Entwurfs in eine reale Webanwendung stellte sich heraus, dass der Interaktionsbereich der Anwendung sehr gering ist. Deshalb wurde für die Upload-Page und die Result-Page die Navigationsleiste in den linken oberen Bereich verschoben (Abbildung 18). Ebenfalls wurde auf der Result-Page die Visualisierungen der Daten in einzelne Panels platziert. So kann der Nutzer interaktiv mit der Anwendung interagieren und nicht relevante Inhalte verdecken (Abbildung 19). Da es für den Nutzer verwirrend sein

6. Entwurf und Implementierung der Software

kann, dass zwischen den Sentiment Score der einzelnen Dokumente eine große Differenz vorliegen kann, wurden die Werte normalisiert. Dies tritt vor allem auf, wenn die Texte sehr unterschiedlich lang sind. Hierbei wurde, je nach Analyse Ebene der Visualisierung die Normalisierung anhand aller Wörter oder anhand aller SBWs in dem Dokument oder Satzes normalisiert.

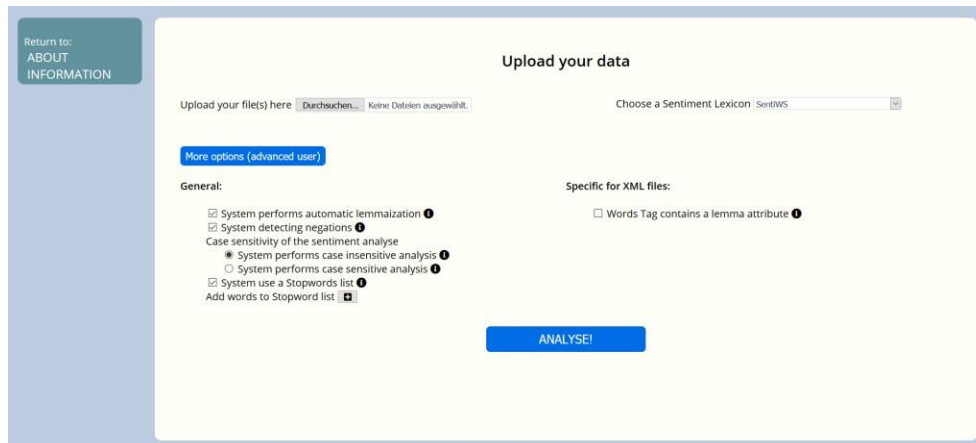


Abbildung 23: Neues Layout - Upload-Page

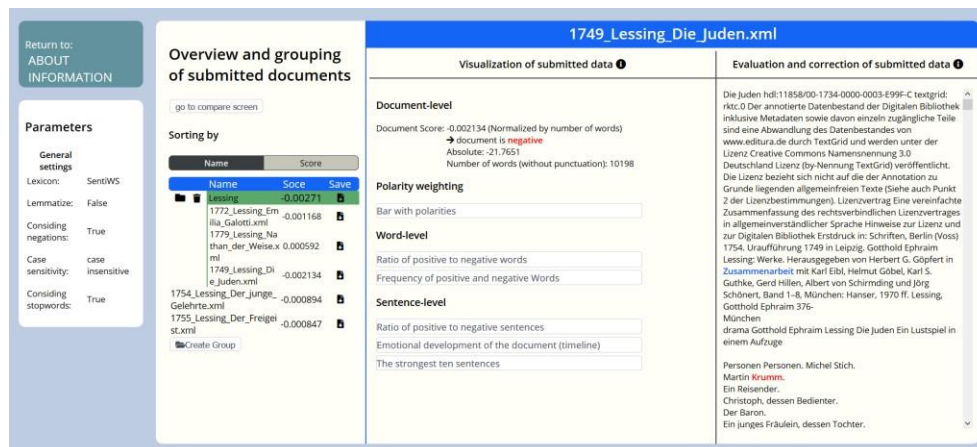


Abbildung 24: Neues Layout - Result-Page mit normalisierten Scores

Zusätzlich verfügt die Information-Page über eine Navigation, um die Interaktion zwischen den einzelnen Überschriften zu verbessern. Klickt der Nutzer hierbei auf ein Element der Navigation, so scrollt die Anwendung automatisch in den gewünschten Bereich (Abbildung 25).

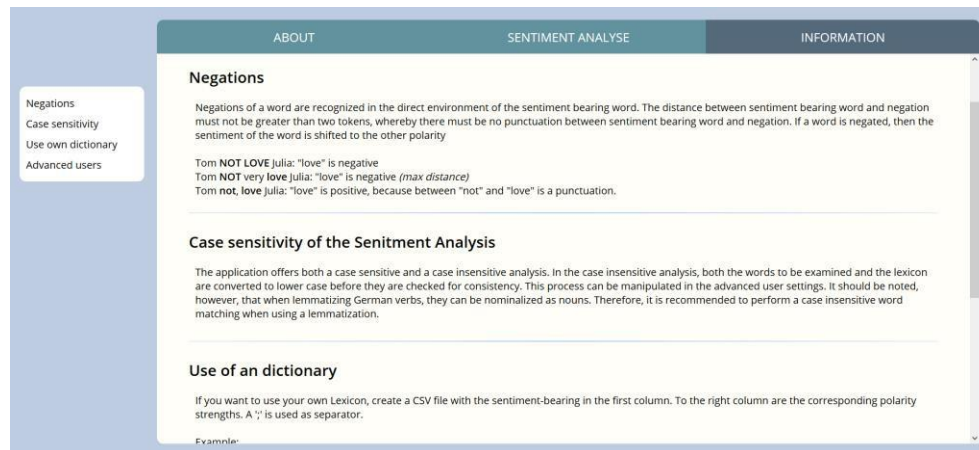


Abbildung 25: Information-Page

6.4.2 Formative Evaluation

In diesem Unterkapitel wird die formative Evaluation und die durchgeführten Verbesserungen aufgrund dessen Ergebnisse beschrieben.

6.4.2.1 Evaluation

Im nächsten Schritt wurde eine formative Studie zum Testen der Usability des aktuellen Designs durchgeführt, wobei die laborgestützte Methode (LAB) angewandt wurde. Hierbei wurde zunächst ein Testplan erstellt, welcher zwei explorative und sieben spezifische Tasks beinhaltet. Spezifische Aufgaben waren das Suchen einer Information und das Analysieren von Texten anhand verschiedener Parameter dar. Zudem wurden dem Nutzer spezifische Tasks über das Nachbearbeiten, Gruppieren, Vergleichen und Exportieren der analysierten Dateien gestellt. Für diese Usability-Studie wurden zwei weibliche und acht männliche Teilnehmer im Alter zwischen 21 und 30 Jahren ($M = 24.6$, $SD \approx 2.95$) rekrutiert. Wobei alle an der Universität Regensburg arbeiten oder studieren. Alle Usability-Tests fanden hierbei als Benutzertests in einem ruhigen Raum an der Universität Regensburg statt. Im Rahmen der Studie mussten die Personen zu Beginn Fragebögen ausfüllen, die die demografischen Daten, sowie Informationen über deren Kenntnisse in Sentiment Analysis beinhalten. Von diesen zehn Teilnehmer hatten vier Testpersonen über keinerlei Vertrautheit mit dem Konzept der Sentiment Analysis. Fünf Personen besaßen über etwas Know-how

und ein Beteiligter über gute Erfahrung mit der Methodik der Sentiment Analysis. Darüber hinaus haben von all diesen Teilnehmer eine Person einmalig eine Sentiment Analysis durchgeführt. Eine Weitere hat eine Sentiment Analysis bereits mehrmals angewendet. Ausgehend von diesen beiden Personen haben beide eine Analyse anhand eines Programms verarbeitet. Zudem hat eine der beiden Teilnehmer eine Sentiment Analysis manuell und mithilfe der Implementierung eines passenden Codes geleistet. Während des Usability-Tests wurden mehrere Aspekte untersucht. Hierbei wurden Informationen über das Verhalten und der Meinung des Nutzers gesammelt. Um die Auffassung eines Teilnehmers zu identifizieren, wurden Kommentare, die der Nutzer im Rahmen der Thinking aloud Methodik während der Aufgaben von sich gab, notiert. Zudem wurde das Verhalten mithilfe mehrerer Metriken festgehalten. Dadurch, dass der Nutzer während der Interaktion mit der Anwendung keine Wahlfreiheit hat, ist hierbei besonders die Effektivität und die Effizienz der Applikation entscheidend. So wurden Fehler notiert, um die Error Frequency zu erhalten. Hinzu kommen Metriken für die Task Success und die Task Time. Um die Task Success zu dokumentieren, wurde eine Skala mit den Werte 0, 0.5 und 1 angewendet. Hierbei steht 0 für einen nicht erfolgreich abgeschlossenen Task. 0.5 für einen teilweise abgeschlossenen Task und der Wert 1 für einen vollständig erfolgreichen vollendeten Task. Um die Task Time zu berechnen, würde die Zeit gemessen, die die Person brauchte, um eine Aufgabe abzuschließen. Dabei startet die Zeitmessung ab dem Zeitpunkt, sobald die Testperson die Aufgaben verstanden hat und erste Interaktionen mit der Anwendung beginnt. Die Messung der Zeit endet ab dem Punkt, sobald die Aufgabe erfüllt wurde oder der Nutzer meinte, er wäre am Ende der Aufgabe angelangt. Als letzte Metrik, um das Effektivität der Applikation zu messen, wurde die Error Frequency während der Aufgaben berechnet. Aussagen, die von den Nutzern während des Tests geäußert wurden, werden in Klammer mit der zugehörigen Nummer dokumentiert. So verweist P1 auf die erste Person, die an dem Usability-Test teilgenommen hat. Für Explorationen wurde das Verhalten der Nutzer nicht berücksichtigt.

6. Entwurf und Implementierung der Software

In der ersten Aufgabe wurde die Startseite exploriert, wofür bis zu 03:01 Minuten dafür aufgebracht wurden (M = 01:11 min., SD = 02:51 min.). Hierbei ist anzumerken, dass viele gleich auch die Upload-Page und die Information-Page exploriert haben. Häufig wurde erwähnt, dass auf der rechten Seite der About-Page der Abstand größer ist als auf der linken Seite (P2), sodass man denkt, dass auf der linken Seite was fehlt (P4). Zudem wurde angemerkt, dass das Aside auf der linken Seite auf der Information-Page dasteht und auf der About-Page nicht (P8). Um diese Inkonsistenz auszubessern, sollte die About-Page ebenfalls ein Aside erhalten. In diesem kann dann der Link zur Information-Page sowie die Kontaktmöglichkeiten stehen. Da dieser Issue jedoch nur von kosmetischer Natur ist, erhält dieser als geringfügig eingestuft. Auf der About-Page wurde die Übersichtlichkeit sehr gelobt (P1; P5; P7; P9), wobei der Aufbau der Seite ganz Basic ist (P6) und Beschreibungen über die Applikation (P3; P6; P7; P10), sowie Kontaktinformationen (P3; P9) zur Verfügung stellt. Jedoch wurde hierbei angemerkt, dass in der Beschreibung zwar steht, dass die Anwendung viele Funktionalitäten beinhalte, aber nicht genau welche (P8), diese sollte man schnellstmöglich hinzufügen, da auf dieser Seite der Nutzer einen ersten Eindruck der Applikation bekommt und diese Seite des Tools bestmöglich präsentieren soll. Zudem wurde die angenehme Farbwahl erwähnt (P9) und dass ein Link zu den Informationen hervorgehoben wird (P9). Außerdem führe es zu Verwirrung, dass beim Wechseln zur Upload-Page sich ein extra Tab öffnet (P4), was sich jedoch einfach beheben lässt und die Funktionalität der Anwendung nicht einschränkt. Ein Teilnehmer war nicht direkt klar, was unter den Reiter „Information“ verstanden wird (P6). Hier könnte eine Umbenennung sinnvoll sein. Eine weitere Person meinte, dass die Navigationselemente für die Information-Page, aufgrund des geringen Umfangs unnötig seien, der Plan ist es jedoch, diese im finalen Design zu erweitern (P4).

Identifizierte Usability-Issues:

1. Links sehr viel Rand (About-Page)
2. Beschreibung möglicher Funktionalitäten (About-Page)
3. Automatisches öffnen eines neuen Tabs (Upload-Page)

4. Ungenaue Bezeichnung des Reiters „Information“

In der zweiten Aufgabe ging es herauszufinden, ob das System eine Negation in einem vorgegebenen Satz erkennt. Hierbei gibt es zwei Lösungswege. Zum einen kann der Nutzer die entsprechende Information auf der Information-Page finden. Darüber hinaus kann er eine eigene Textdatei anlegen und diese analysieren lassen, und im Result-Page zur gesuchten Lösung gelangen. Zum Lösen dieser Aufgabe benötigten die Personen zwischen 00:15 Minuten und 05:42 Minuten ($M = 02:55$ min., $SD = 01:41$ min.). Wobei insgesamt zwei Fehler unterliefen ($M = 0.02$, $SD \approx 0.42$) und eine Task Success Rate von 80% erreicht wurde. Um diese Aufgabe zu lösen wurde eine eigene Textdatei erstellt (P2; P3; P7; P9; P10). Hierbei wurde besonders häufig ein Eingabefeld gewünscht, in dem der Nutzer einen Text direkt eingeben kann (P3; P5; P6; P10), da dies praktischer ist (P10). Aus diesem Grund ist ein kleines Textfeld sinnvoll, in dem man stichprobenartig einige Formulierungen überprüfen kann, dessen Umsetzung über eine hohe Priorität verfügt. Zudem herrschte Unsicherheit, welche Textdateien erlaubt sind (P2; P9) und woher der Unterschied zwischen dem normalisierten und absoluten Score herkommt (T10). Die Benutzeroberfläche sollte deshalb an den entsprechenden Stellen über die passenden Hinweise verfügen. So kann man auch vermeiden, dass die Nutzer die Ergebnisse falsch interpretieren und anstatt der Scores der einzelnen Wörter zu untersuchen, sich an der Polarität des Dokument Score orientierten.

Identifizierte Usability-Issues:

5. Eingabefeld erwünscht
6. Unklarheit, welche Dateien erlaubt sind (Upload-Page)
7. Falsches interpretieren der Scores (Result-Page)

Die nächste Aufgabe stellt das Ausführen einer Sentiment Analysis anhand vorgegebener Dateien dar. Hierbei benötigten die Teilnehmer zwischen 00:30 Minuten und 01:14 Minuten ($M = 00:43$ min., $SD = 00:13$ min.). Wobei hierbei keinerlei Fehler vorgekommen sind und die Task Success Rate bei 100% lag. Bei dieser Aufgabe war nicht genau klar ob es möglich ist, mehrere Dokumente gleichzeitig zu analysieren kann (P1; P4; P7; P8; P9; P10), weshalb ein entspre-

chender Hinweis beim Upload unabdingbar ist und schnell umgesetzt werden muss. Zudem hatten Personen, die in der vorherigen Aufgabe bereits eine Analyse durchgeführt hatten, Schwierigkeiten, zu Upload-Page zurückzugelangen, da ein passender Link auf der Result-Page fehlte, sodass die die Browser-Navigation nutzen oder über eine der anderen beiden Seiten zur Upload-Page gelangten. Um die Navigation zu erleichtern, bietet sich in diesem Fall an, bei den Links, welche sich oben Links auf der Result-Page befinden, ein entsprechender Link zur Upload-Page hinzuzufügen.

Identifizierte Usability-Issues:

8. Hinweis zum Upload mehrere Dateien (Upload-Page)
9. Button für neue Analyse (Result-Page)

Die dritte Aufgabe stellt das Explorieren der Result-Page dar. Hierbei wurden schnell die wichtigsten Funktionalitäten gefunden (P1; P2). Positiv sind hierbei hover-Effekte der detaillierten Darstellung des Textes (P1; P2; P8; P10) aufgefallen. Jedoch war nicht genau klar, was der Anwendungszweck einer Sentiment Analysis ist, weshalb die Anwendung dem Nutzer Kontextinformationen über die Sentiment Analysis zur Verfügung stellen sollte (P7). Dies sollten gerade unterfahrene Nutzer (P4) die Interaktion mit dem Tool und dessen verständlich der Ergebnisse unterstützen. Aus diesem Grund sind auch einige Begriffserklärungen, wie den Sentiment Score, von Nöten (P3; P6; P8). Zudem war den Nutzern nicht direkt ersichtlich, dass die Panels, in denen sich die Diagramme verbergen, klickbar sind („Kann ich das [Panel für die Grafiken] anklicken?“ P6). So äußerte eine Person den Eindruck, dass man darin etwas sortieren kann (P4). Um die Problematik zu beheben, bietet es sich an, die Darstellung der Panels zu überarbeiten. Sofern die Nutzer die Grafiken exploriert haben, fallen diese den Nutzern positiv auf. So werden diese als „leicht zu verstehen und recht cool“ (P2) und als „eindrucksvoll“ (P5) beschrieben. Aber hier fallen ein paar Schwächen auf. Die Wordclouds, welche die stärksten Begriffe in Abhängigkeit ihrer Frequenz und Polaritätsstärke darstellen, werden zwar als schön erachtet (P2). Jedoch wurden diese von einer Person falsch interpretiert, sodass diese nur die meist genannten Wörter (P8) abbildet. Ein Teilnehmer war nicht gleich klar, was

diese Clouds darstellen (P5). Zudem hat niemand mit der Timeline agiert, um zu der gewünschten Textstelle zu scrollen und einer Person war sich nicht genau sicher, was diese Grafik bedeutet (P2). Bei beiden Diagrammen ist eine entsprechende Notiz von Nöten oder eine passende Benennung des zugehörigen Panels. Darüber hinaus tauchten einige Schwierigkeiten bei der Visualisierung des Textes auf. So war die Farbcodierung der einzelnen Wörter des Textes nicht sofort schlüssig (P5; T6) und eine Legende wurde von einer Person vermisst (T2). Um dieses Problem zu beheben, ist eine entsprechende Legende im oberen Bereich des Textes sinnvoll und sollte zeitnah umgesetzt werden. Ungeachtet dessen sollte man an dieser Stelle festhalten, dass auf der anderen Seite die Farbcodierung an sich gut und intuitiv ist (P7) und diese relevanten Wörter direkt ins Auge fallen (P9). Eine weitere Problematik war die Darstellung des absoluten Sentiment Scores, da die Nutzer nicht wussten, ob der Punkt für Dezimalzahlen oder zur Zifferngruppierung dient (P6; P7). Weiter war nicht genau klar, ob nur der aktuelle Text oder alle auf einmal angezeigt werden (P4). Derzeit befindet sich der Dateinamen der gewählten Datei in einem Header. Diesen könnte man noch deutlicher hervorheben, damit direkt klar ist, dass nur ein Text betrachtet wird. Bezüglich der Informationsflut herrscht Ausgeglichenheit. Eine Person merke an, dass die Seite super aussieht (P1) während eine andere die vielen Informationen bemängelte (P3). Zuletzt noch einige positive Äußerungen der Explorationen. Die Nutzer haben schnell das Schema des Layouts verstanden. Sodass ganz auf der rechten Seite die Parameter der Analyse stehen (P6; P8; P9), der Nutzer zwischen den einzelnen Texten in der Übersicht wechseln kann (P2; P4; P10) und der aktuelle Text im rechten Bereich abgebildet wird (P2; P4; P5; P10), welchen man anschließend näher betrachten kann. Bei dieser Aufgabe wurden durchschnittlich 03:41 Minuten dafür aufgewendet (SD = 01:27min.).

Identifizierte Usability-Issues:

10. Kontextinformationen zur Sentiment Analysis
11. Begriffserklärungen
12. Design der klickbaren Panels

13. Erklärung der Wordcloud und der Timeline
14. Legende über die Farbkodierung im Text
15. Dezimalzahl beim absoluten Score kennzeichnen
16. Header des Close-Reading Bereich

Die nächste Aufgabe beinhaltete das Anpassen des Sentiments eines Wortes. Dieser Task wurde von allen Personen recht schnell erfolgreich und fehlerfrei mit einer durchschnittlichen benötigten Zeit 00:50 Minuten (SD = 00:23min.) erfüllt. Hierbei wurde, um das Wort zu finden, teilweise die Browser- Suchfunktion genutzt, was sehr gut funktionierte (P2; P4). Jedoch war nicht unmittelbar ersichtlich, ob bereits das entsprechende Dokument angezeigt wird (P4), was sich aber durch das anpassend es Headers im Close-Reading Bereich lösen sollte (16. Usability-Issue). Allerdings traten hierbei zwei Problematiken auf. Erstens war vielen Personen unklar, ob die Dezimalzahl mit Komma oder mit einem Punkt geschrieben werden muss (P2; P5; P8) und zum zweiten springt das geänderte Wort zwischen zwei Zahlen (P1; P6; P10), wodurch sich die Interaktion mit dem Wort als schwierig gestaltet. Auch wenn es für die Funktionalität der Anwendung von keinerlei Belangen ist, ob der Begriff mit einem Punkt oder einem Komma geschrieben wird, so gibt ein passender Place- holder dem Benutzer trotzdem die Nötige Sicherheit, weshalb dieser Issue auf jeden Fall umgesetzt werden muss. Die Nutzer halfen sich in dieser Situation, indem sie die Schreibweise bei anderen SBWs kontrollierten (P9) oder hofften, dass ihre Eingabe korrekt war. Insgesamt wurde die Funktionalität schnell gefunden und einfach angewendet (P1). Zudem sieht der Nutzer, dass sich die Auswertung sofort anpassen (P3).

Identifizierte Usability-Issues:

17. Placeholder beim Sentiment Input
18. Springendes Wort beim hover-Effekt

Der fünfte Task beinhaltet das Erstellen eines Ordners und dem hinzufügen die passenden Elemente im Benutzerinterface. Diese Aufgabe schafften alle problemlos. Jedoch hatte eine Person leichte Schwierigkeiten beim Drag and Drop, was das loslassen im entsprechenden Bereich betrifft. Hierbei bietet es sich an,

die Linienhöhe des Ordnernamens zu vergrößern, damit der Bereich, in dem der Nutzer loslassen muss, groß genug ist (P10). Zudem hat eine Person bemängelt, dass die roten Rahmen, der den Zielbereich kennzeichnet, etwas verwirrend (P3). Insgesamt ist das Erstellen eines Ordners, und das Hinzufügen der Dateien sehr intuitiv (P2; P6; P7). Sowie an der Realität orientiert (P7). Zudem hat der erstellte Ordner direkt einen eigenen Sentiment Score (P2; P10). Für diese Aufgabe benötigten die Nutzer im Mittel 00:34 Minuten (SD = 00:23). *Identifizierte Usability-Issues:*

19. Drag and Drop verbessern

20. Farbe des Zielbereichs beim Drag and Drop ändern

In der nächsten Aufgabe mussten die Nutzer die einzelnen Dokumente anhand zweier Merkmale vergleichen. Dies fällt sowohl anhand der Anzahl von insgesamt acht Fehler (M = 0.8, SD ≈ 0.92) als auch der Task Success Rate von 65% auf. Wobei hierbei die Nutzer bis zu 04:10 Minuten (M = 01:50 min, SD = 01:23) benötigten. Primäres Problem war hierbei, dass die Teilnehmer den Vergleichsbildschirm nicht nutzen. Grund dafür war, dass der Button, der die Nutzer zum Vergleichsbildschirm leitet, nicht sehr auffällig dargestellt wurde. Zudem wurde nicht immer die komplette Lösung oder nur ein teilweise richtiges Ergebnis abgegeben, was an der resultierenden Arbeitsweise gelegen haben könnte. Es ist sehr wichtig, dass dieser neu gestaltet wird, da die Nutzer sonst keinen Zugang zu dieser Funktion haben. Vielmehr versuchten sie, zwischen den einzelnen Texten hin und her zu wechseln. Umso zu Lösung zu gelangen (P3; P4; P7; P8; P9; P10). Hierbei ist positiv aufgefallen, dass die Grafiken geöffnet bleiben, während der Nutzer zwischen den einzelnen Dokumenten wechselt (P7; P10). Auch bei dieser Aufgabe kam die Frage auf, ob beim absoluten Score eine Dezimalzahl sei (P10).

Identifizierte Usability-Issues:

21. Auffälliger Button zum Vergleichsbildschirm

Als in der sechsten Aufgabe wurden die Teilnehmer aufgefordert, eine neue Analyse mit XML Dateien durchzuführen. Hierbei hatten die Dokumente die Besonderheit, dass die einzelnen Wörter bereits über ein Attribut besitzen, wel-

ches das Lemma beinhaltet. Dieses Lemma sollten die Nutzer nun für ihre Sentiment Analysis nutzen. So mussten die Teilnehmer den Attributen Namen in ein entsprechendes Input-Feld der "advanced-User Options" eingeben. Einige Nutzer führten eine automatische Lemmatisierung durch, anstatt das Attribut zu verwenden (P4; P7; P8). Zudem füllten einige Nutzer zuerst das entsprechende Feld nicht aus, da die sich darin befindeten Placeholder den Anschein machten, dass das Feld bereits vorausgefüllt ist (P1; P2). Hierbei bekommt der Anwender die Meldung, dass der Input noch nicht ausgefüllt wurde. Um dieses Verhalten zu vermeiden, sollten die Placeholder angepasst werden. Des Weiteren haben zwei Personen auf der Information-Page Informationen über die Funktion eingeholt (P2, P8). Es wurde festgestellt, dass die Analyse nun etwas mehr Zeit beansprucht (P10). In dieser Aufgabe wurden insgesamt 4 Fehler gemacht ($M = 0.4$, $SD \approx 0.42$) und eine Task Success Rate von 80% erreicht und durchschnittlich 02:00 Minuten ($SD = 01:20$ min.) benötigt.

Identifizierte Usability-Issues:

22. Placeholder im Input des Lemma-Attribut nicht eindeutig

Die letzte Aufgabe stellte das Herunterladen der erstellten Grafiken dar. Hierbei suchten zwei Personen zuerst die Darstellungen im Text (P2; P7). An dieser Stelle tritt wieder die Problematik auf, dass die Panels, in denen sich die Grafiken befinden, sehr unscheinbar sind. Deshalb brauchte eine Person sehr lange bis diese, die zu speichernde Visualisierungen gefunden hatte (P3). Einige probierten die Grafiken zu speichern, indem sie mit Rechtsklick versuchen die Bilder herunterzuladen (P5; P7; P10). Eine Implementierung der Funktion, dass man die Diagramme per Rechtsklick herunterladen kann, ist an dieser Stelle sinnvoll und intuitiv. Aus diesem Grund sollte die Fiktionalität baldmöglichst umgesetzt werden. Zudem kam die Anzahl der zur Verfügung gestellten Grafiken und die Benennung derer gut an (P7; P10). Hierbei wurden insgesamt vier Fehler gemacht ($M = 0.4$, $SD = 0.7$) und durchschnittlich 01:06 Minuten ($SD = 01:02$ min.) gebraucht. Alle Teilnehmer haben die Aufgabe erfolgreich beendet.

Identifizierte Usability-Issues:

23. Speichern der Grafiken mithilfe eines Rechtsklick

Während der Testphase konnten insgesamt 23 Usability-Issues identifiziert werden. Diese wurden alle in einer Excel-Datei eingetragen, beschrieben, Lösungsansätze Notierung und einer Priorität zugeordnet. Insgesamt wurden 22 Fehler gemacht und eine Task Success Rate von 89% (exklusive Explorations) erreicht. Mit den beiden Explorations-Aufgaben wurde eine Task Success Rate von 92% ermittelt. Die nachfolgende Grafik zeigen die Boxplots der Time per Tasks. Aufgaben, welche Exploationen repräsentieren, sind die Aufgaben 0 und 3.

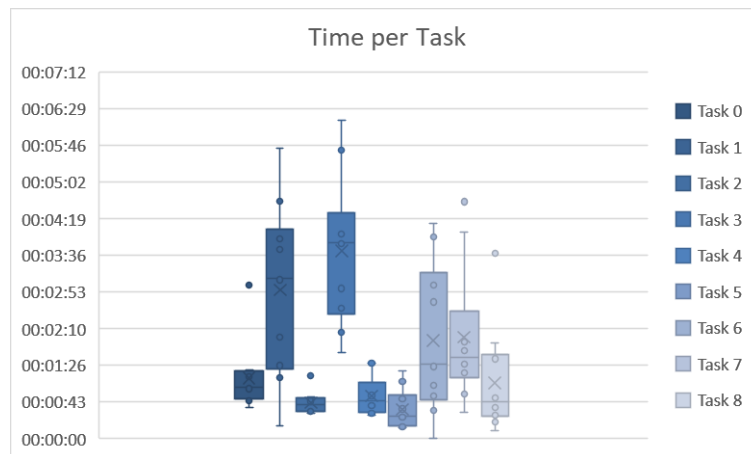


Abbildung 26: Time per Task

6.4.2.2 Verbesserungen

Nachfolgend werden die einzelnen Tasks mit den gewonnenen Erkenntnissen und aufgetretenen Usability Problemen in einer Tabelle²⁷ dokumentiert. Hierbei werden die Usability-Probleme mit der von Nielsen entworfenen Severity Rating gewertet. Insgesamt konnten 23 Usability-Issues werden. Dabei werden die einzelnen Issues von einer Skala bis zu einer Skala von vier bewertet. Je höher die Bewertung ist, desto dringlicher ist die Ausbesserung des Problems. Zusätzlich beinhaltet jeder festgehaltener Issue die dazugehörige Verbesserung. Anbei ist ein Ausschnitt dieser Dokumentation zu finden (Tabelle 1).

²⁷ Verfügbar unter \Studien\usability_issues.xlsx

Nr.	Titel	Beschreibung	Lösung	Priorität
2	Beschreibung möglicher Funktionalitäten (About-Seite)	Es wird in der Beschreibung nicht genau genannt, welche Funktionalitäten die Anwendung bietet	Funktionalitäten der Beschreibung beifügen.	3
3	Automatisches öffnen eines neuen Tabs (Upload-Seite)	Das automatische öffnen eines neuen Tabs, wenn man zur Upload-Seite wechselt, führt zur Verwirrung.	Entfernen des Attributs im Link, der dieses Verhalten definiert.	2

Tabelle 1: Ausschnitt Dokumentation der Usability-Issues

Von den 23 entdeckten Usability-Issues konnten 22 ausgebessert werden. Die nachfolgenden beiden Grafiken zeigen die klickbaren Panels, in denen sich die einzelnen Diagramme befinden, während der Usability-Studie (Abbildung 27) und nach deren Überarbeitung (Abbildung 28).

Word-level

Ratio of positive to negative words

Frequency of positive and negative Words

Abbildung 27: Design der klickbaren Panels (Issue Nr. 12) - Davor

Charts of Word-level

Pie chart: Ratio of positive to negative words

Cloud & Bar chart: Most strongest words

Abbildung 28: Design der klickbaren Panels (Issue Nr. 12) – Danach

6.4.3 Summative Evaluation

In diesem Unterkapitel wird die summative Evaluation und die durchgeführten Verbesserungen aufgrund dessen Ergebnisse beschrieben.

6.4.3.1 Evaluation

Nachdem eine formative Usability Studie durchgeführt wurde und die aufgetretenen Usability-Issues dokumentiert und ausgebessert wurden, wurde eine summative Studie angewandt. Die nachfolgende Dokumentation orientiert sich nach der ISO/IEC 25062:2006. Während die formative Studie das Ziel hatte, Usability-Issues aufzudecken, wurde in dieser Studie der Fokus auf die Gesamteinschätzung der Applikation gelegt. Ziel dieser Studie ist zu prüfen, ob die Anwendung, die im Kontext der DH die festgestellten Anforderungen erfüllt und hierbei das System eine Gebrauchstauglichkeit garantiert. Um dies zu testen wurden vier Teilnehmer rekrutiert, wobei drei Personen bereits an der Anforderungsanalyse teilgenommen haben. Bei den vier Testpersonen handelt es sich um je zwei Frauen und Männern im Alter zwischen 25 und 31 Jahren ($M = 28.75$, $SD \approx 2.87$), wobei zwei Personen etwas Erfahrung und zwei Teilnehmer viel Erfahrung mit der Methodik der Sentiment Analysis besitzen. Dabei hat eine Person noch nie eine Sentiment Analysis durchgeführt, während die anderen dies schon mehrmals getan haben. Von diesen drei Personen hat eine Person die Analysen manuell ausgeführt, eine Testperson hat die Sentiment Analysis mithilfe eines Programms und die Implementierung eines passenden Codes geleistet. Die dritte Person hat Erfahrungen in allen drei vorgeschlagenen Varianten der Durchführung der Sentiment Analysis. Das für das hierfür verfasste Szenario stellte das exemplarische Analysieren von ausgewählten Dramen dar, um sich mit der Sentiment Analysis vertraut zu machen. Die für die Studie erstellten Aufgaben beinhalteten das Explorieren der Seite, eine Informationssuche, das Hochladen von Dateien, das Untersuchen und Bearbeiten der Sentiment Analysis, sowie das Vergleichen und Exportieren der Sentiment Analysis-Ergebnisse.

Aufgaben:

0. Exploration
1. Informationssuche
2. Durchführen der Sentiment Analysis
3. Exploration der Result-Page
4. Negation erkennen
5. Sentiment korrigieren
6. Ergebnisse vergleichen
7. Ergebnis exportieren

Um die Studie durchzuführen, wurde ein synchroner Remote Usability-Test (RS) angewandt. Dies bietet den Vorteil, dass die Studie an keinen Ort gebunden ist und die Anwendung direkt in der natürlichen Umgebung der Teilnehmer getestet werden kann, um so umgebungsspezifische Issues aufzudecken. Zudem zeigte eine frühere Studie, dass die Anzahl der gefunden Usability-Issues zwischen LAB und RS vergleichbar sind (Andreasen, Nielsen, Schröder, & Stage, 2007). Um die Studie durchzuführen, wurde die Webanwendung und eine Webseite, welche den Nutzer durch die Studie führt, deployt. Auf dieser Webseite sind die Links zu den Fragebögen, der Applikation und die einzelnen Tasks (Abbildung 29) gespeichert. Um der einzelne Test durchzuführen, wurde als Software Skype benutzt. Hierbei teilten die Teilnehmer ihren Bildschirm, sodass der Moderatorin die Interaktion des Nutzers verfolgen konnte. Ebenfalls wurde der Test für die spätere Analyse aufgezeichnet. Um die Tasks durchzuführen, wurden Texte vorbereitet, welche die Nutzer im Vorfeld übermittelt wurden. Da bei einer Person Skype nicht zugänglich war, wurde die Studie vor Ort verlagert, jedoch mit derselben Umsetzung wie bei den restlichen Teilnehmern.

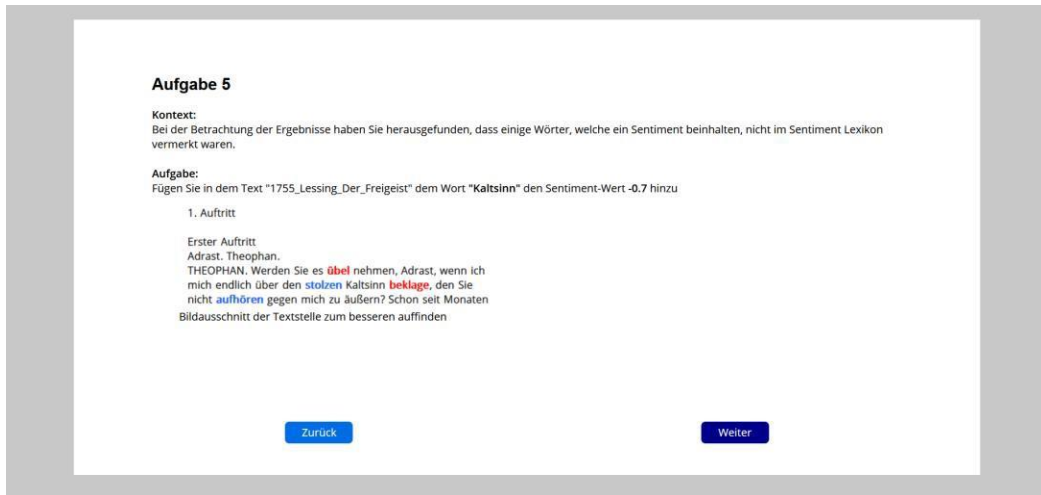


Abbildung 29: Aufgabenstellung von Task 5

Wie in der formativen Studie wurden für jede Aufgabe die Effizienz mithilfe der Time per Task ermittelt. Die Messung der Time per Task beginnt ab dem Zeitpunkt, wenn der Nutzer zur Anwendung wechselt. Die Messung wurde beendet, sobald der Nutzer zu den Aufgaben erfüllt hatte. Um die Effektivität der Anwendung zu prüfen wurden die Error Frequency und die Tasks Success dokumentiert, wobei diese wie in der formativen Studie gemessen wurden. Um die Meinung des Nutzers festzuhalten, wurden die Kommentare der Testpersonen notiert und der Teilnehmer nach der Studie noch nach dessen Eindruck zu dem Tool befragt. Zusätzlich wurden vorab die demografischen Informationen der Person erfasst. Nach der Studie wurden die Nutzer aufgefordert die deutsche Version des System Usability Scale (SUS) (Brooke, 1996) sowie den AttrakDiff (Hassenzahl, Burmester, & Koller, 2003) zu beantworten. Während der SUS Fragebogen auf die Usability der Applikation misst, geht es bei AD um die „Messung [der] wahrgenommener hedonischer und pragmatischer Qualität“ (Hassenzahl et al., 2003, p. 1), wobei die User-Experience anhand der Attraktivität eines Produktes gemessen wird. Dabei definieren Hassenzahl und Kollegen die beiden Qualitäten folgendermaßen: „Ist ein interaktives Produkt zur Manipulation der Umwelt geeignet, und wird auch von seinen Benutzern so wahrgenommen, besitzt es „pragmatische“ Qualität. Erweitert ein interaktives Produkt hingegen durch neue Funktionen die Möglichkeiten des Benutzers, stellt neue Herausforderungen, stimuliert durch visuelle Gestaltung und neuar-

tige Interaktionsformen oder kommuniziert eine gewünschte Identität [...] besitzt es „hedonische“ Qualität“ (Hassenzahl et al., 2003, p. 2). Der AttrakDiff wurde als Messinstrument ausgewählt, um die User-Experience und die Attraktivität zu erfassen. Zum einen ist die User-Experience für die Etablierung von Anwendungen wichtig. Zum anderen spielt die Ästhetik einer Anwendung und dessen minimalistisches Design eine Wichtige Rolle nach den Usability Heuristiken nach Nielsen (Nielsen, 1994). Nachdem alle Nutzertests durchgeführt wurden, wurden die Tests analysiert und deren Ergebnisse dokumentiert. Da es in dieser Studie sich nicht um das Auffinden von Issues fokussiert, sondern einen Gesamteindruck der Anwendung untersucht wird. Werden die Resultate hier ohne das spezielle Vorgehen der Nutzer zu den einzelnen Tasks beschrieben.

Als erstes wird auf das Verhalten des Nutzers, anhand der erhobenen Metriken eingegangen. Durchschnittlich brauchten die Nutzer 01:39 Minuten (SD = 2:11 min.), um eine Aufgabe zu erledigen. Hierbei ist jedoch zu beachten, dass vor allem die Exploration der Result-Page viel Zeit für sich beanspruchte. Das nachfolgende Diagramm veranschaulicht die Verteilung der benötigten Zeiten anhand der einzelnen Aufgaben (Abbildung 30). Anschließend wurden die vorgefundenen Fehler gezählt (Abbildung 31). Hierbei kam heraus, dass sowohl der zweite als auch der vierte Teilnehmer keine Fehler produzierten. Insgesamt sind vier Fehler aufgetreten (M ≈ 0.1, SD ≈ 0.4).

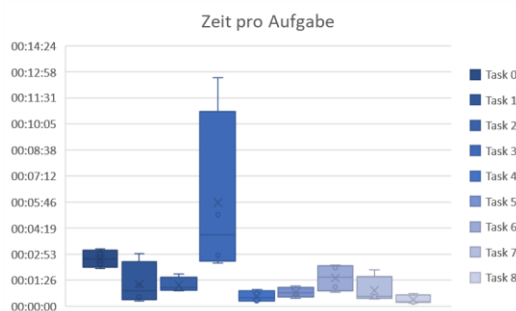


Abbildung 30: Benötigte Zeit pro Aufgabe aller Nutzer

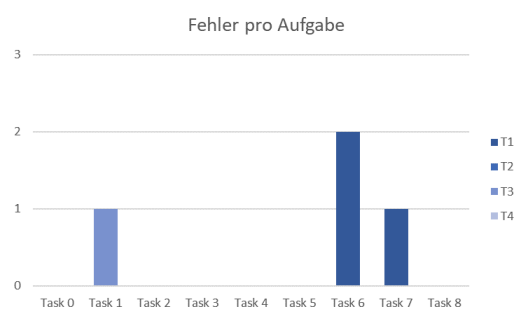


Abbildung 31: Vorgefundene Fehler der Nutzer pro Aufgabe

Darüber hinaus wurde ebenfalls erfasst, ob der Nutzer die spezifische Aufgabe erfüllt hat und wie hoch die durchschnittliche Task Success Rate insgesamt ist (Abbildung 32). Die mittlere Erfolgsrate lag bei 97%, rechnet man die Evaluati-

onen heraus, so gelangt man auf eine Task Success Rate im Mittel von 96%. Beide Aufgaben sind darauf zurückzuführen, dass die Nutzer die Aufgaben nicht korrekt gelesen haben. Auf ihre Weise jedoch die Aufgabe korrekt erfüllt haben, weshalb das Ergebnis nur teilweise richtig ist.

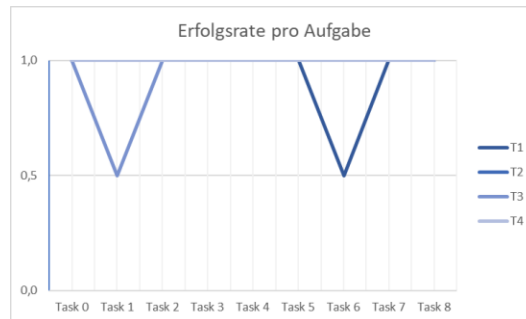


Abbildung 32: Erfolgsrate pro Aufgabe und Nutzer

Das Feedback bezüglich der Anwendung, deren Umsetzung und Funktionalitäten ist positiv ausgefallen. So wurde positiv angemerkt, dass umfangreiche Informationen über die Parameter zur Verfügung gestellt werden (P1), eine Demonstration der Sentiment Analysis zugänglich (P2; P4) ist und die Applikation sehr übersichtlich gestaltet ist (P1; P3; P4). Zudem ist die Anwendung auch für Personen mit einer Rot-Grün-Blindheit benutzbar (eine Teilnehmende hatte ebendiese Sehschwäche). Auf der Ergebnisseite fielen die umfangreichen Funktionalitäten positiv aus, sowie die gelungene Balance zwischen dem Close-Reading und dem Distant-Reading (P2). Die Optik des Tools wirke sehr professionell (P2). Zudem seien die Anforderungen an die Anwendung erfüllt, es ist eine Transparenz vorhanden, man kann Metadaten nutzen und auch ein Lexikon hinzufügen, was der betreffenden Person sehr wichtig war (P1). Darüber hinaus sei das Tool super nutzbar für die DH (P1). Und auch die Information-Page ist sehr gut angekommen (P1; P2). Die nachfolgende Grafik stellt Wörter dar, die besonders häufig vorkommen ²⁸ (Abbildung 33). Hierbei wurden Stoppwörter sowie Füllwörter entfernt. Es ist zu erkennen, dass viele positiv gestimmte Wörter wie „schön“, „toll“, „super“ und „cool“ vorkommen.

²⁸

Interaktiv verfügbar unter

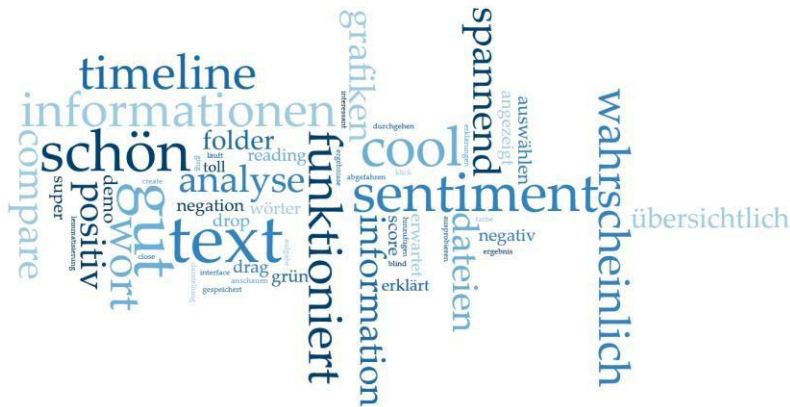


Abbildung 33: Wordcloud mit häufigen Wörtern

Anschließend werden die Ergebnisse der beiden Fragebögen untersucht und ausgewertet. Als erstes wird auf den AttrakDiff Fragebogen eingegangen. Anschließend wird der SUS näher untersucht. Der AttrakDiff bietet mehrere Diagramme, mithilfe derer der Nutzer sein System bezüglich der beschriebenen Qualitäten einordnen kann. Wie man erkennen kann, kann die Anwendung anhand des Konfidenz-Rechtecks der pragmatischen Qualität zugeordnet werden (Abbildung 34). Somit unterstützt das Produkt den Nutzer optimal. Dies spiegelt sich auch in der Ausprägung der Dimensionen wider (Abbildung 35). Hierbei ist sowohl die pragmatische Qualität (PQ) als auch die Attraktivität (ATT) im überdurchschnittlichen Bereich. Jedoch befindet sich sowohl die Identität und die Stimulation der hedonischen Qualität im durchschnittlichen Bereich. So bietet sich an, diese Punkte nochmals genauer zu betrachten.

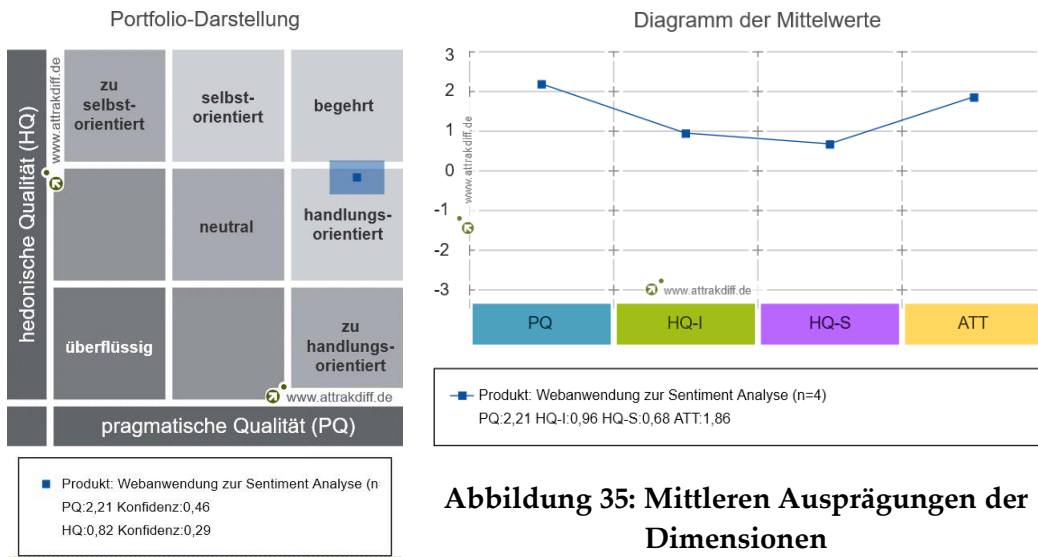


Abbildung 35: Mittleren Ausprägungen der Dimensionen

Abbildung 34: Charakterbe- reiche der Anwendung

Im nächsten Schritt wird der SUS untersucht. Hierbei stellte sich heraus, dass sich ein Ausreiser befindet. Bei genauerer Untersuchung des zugehörigen Tests und die Antworten der Person im AttrakDiff in der Dimension der pragmatischen Qualität, lässt sich daraus vermuten, dass die entsprechende Person die Skala falsch interpretiert hat. Dieser Teilnehmer hat alle Aufgaben fehlerfrei und erfolgreich erfüllt. Auch die pragmatische Qualität der Anwendung wurde von dieser Person sehr gut bewertet (beispielsweise bei dem Wortpaar kompliziert-einfach). Aus diesem Grund wurden die Antworten dieser Person bei der Auswertung des SUS nicht berücksichtigt. Wertet man den SUS aus, somit erhält man einen Score von 92,5 (SD ≈ 9). Anhand dieses Scores lässt sich aussagen, dass die Usability der Anwendung exzellent ist (Abbildung 36).

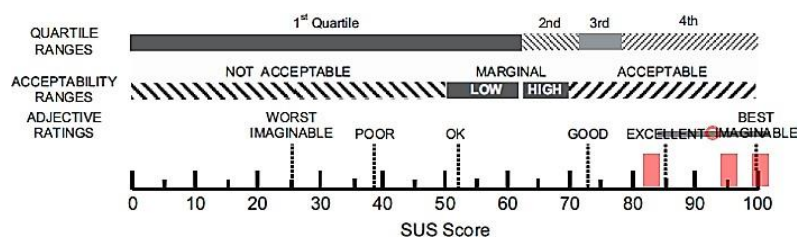


Abbildung 36: SUS

6.4.3.2 Verbesserungen

Während der Test sind wenige Aspekte aufgefallen, die noch etwas verbessern werden, können. Zudem wurden einige wünschenswerte Funktionalitäten seitens der potenziellen Nutzer geäußert. So wurden folgende Änderungen vorgenommen:

1. Wenn ein Nutzer mit der Maus über einen Link schwebt, so wird dieser unterstrichen. Aufgrund unterschiedlicher Bildschirmbreiten stellte ich beim Test heraus, dass einige Links beim hovern zwischen den Zeilen springen. Da sich die Strichstärke erhöhte. So ist der Effekt auch besser erkennbar.
2. Auf der About-Page ist die Auflistung der Funktionalitäten mit entsprechenden Links zu dem passenden Bereich in der Information-Page versehen werden.
3. Da eine Person nach der Quelle der Stoppwörter gefragt hat, sind nun auf der Information-Page die Quellen der Stoppwörter und Negationen angeben.
4. Auf der Result-Page sind nun die Sätze der Tabelle mit den stärksten positiven und negativen Sätzen klickbar. Wenn ein Nutzer auf einen Satz klickt, scrollt das System automatisch zu der entsprechenden Stelle im Text.
5. Außerdem ist dem Nutzer nun möglich, Textausschnitte mit gedrückter Maustaste zu markieren und so die Sentiment Informationen des gewählten Textausschnittes zu erhalten (Abbildung 53).
6. Die Interaktion beim Vergleich wurde überarbeitet. Nun muss der Nutzer nicht mehr auf einen separaten Button klicken um die aktuelle Konfiguration anzuzeigen. Sobald sich ein Parameter ändert, wird sofort ein neuer Vergleich gestartet.
7. Auf den Vergleichsbildschirm befindet ein Button, mit dem der Nutzer zur letzten Datei oder zum letzten Ordner zurückkehren kann.

8. Öffnet sich die Ergebnisseite, so ist das oberste Panel der Visualisierungen automatisch geöffnet
9. Möchte der Nutzer einen Text hochladen, so öffnet sich ein Fenster, indem sich all dessen Dateien befinden. Nun werden nur noch die Dateien angezeigt, welche auch die korrekte Dateierweiterung besitzen.
10. Es hat sich eine falsche Bezeichnung eingeschlichen. Das Wort „Analyse“ wird durch „Analysis“ ersetzt.
11. Der Reiter „Information“ wird in „Documentation“ umgewandelt, da diese Seite schon einer kleinen Dokumentation gleicht.
12. Zudem werden ein paar Benennungen/Labels umbenannt.

6.4.4 Funktionalität

Im nachfolgenden Kapitel werden die Funktionalität und die Benutzeroberfläche beschrieben.

6.4.4.1 *Navigationselemente*

Je nach der aktuellen Page, auf die der Nutzer gerade aufgerufen hat, kann dieser mithilfe einer horizontalen Navigation Bar (About-Page und Documentation-Page) oder einer vertikalen Navigation Bar (Upload-Page und Result-Page) navigieren.

6.4.4.2 *About-Page*

Die About-Page (Abbildung 37) verfügt über Kontextinformationen der Sentiment Analysis und des gewählten Ansatzes. Zudem wurde auf dieser Seite eine Liste mit den genauen Funktionalitäten, sowie eine Demoversion für das deutsche Sentiment-Lexikon SentiWS integriert. Da diese Demo nur einen kleinen Einblick in die Funktionalität der Anwendung beinhaltet, wurde die Zeichenzahl auf 200 Zeichen begrenzt und die einfachste Darstellung unterstützt. Auf diese Weise soll der Nutzer animiert werden, eine Sentiment Analysis durchzuführen, um die ganze User Experience zu erfahren. Um das Design zu vereinheitlichen, besitzt diese Seite nun ebenfalls ein Aside-Element, welches Kontaktinformationen, sowie den Hinweis über die Dokumentation enthält.

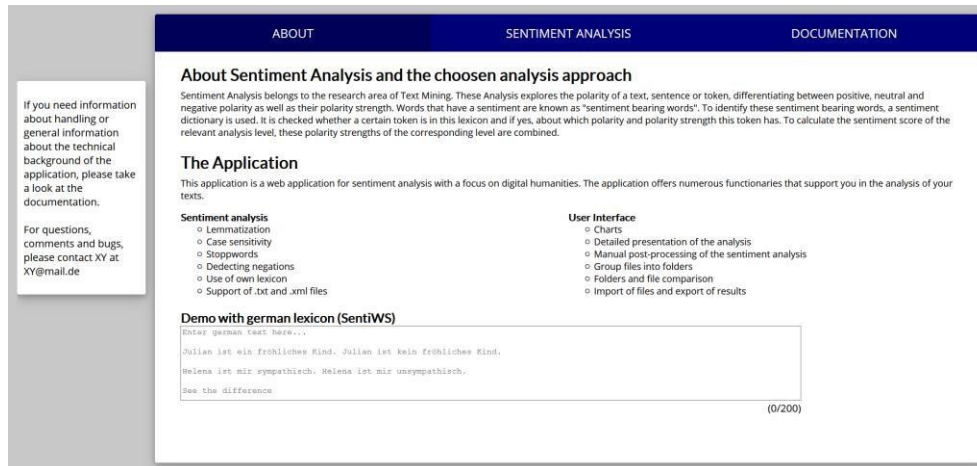


Abbildung 37: About-Page

6.4.4.3 Documentation-Page

Die Documentation-Page (Abbildung 38) ist mit den aktuellen Informationen ausgestattet und beinhaltet nun eine umfangreichere Navigation innerhalb dieser Seite. Die Informationen decken sowohl allgemeine Informationen über Teilaspekte der Sentiment Analysis, sowie konkrete Informationen über die Funktionalität der Anwendung ab.

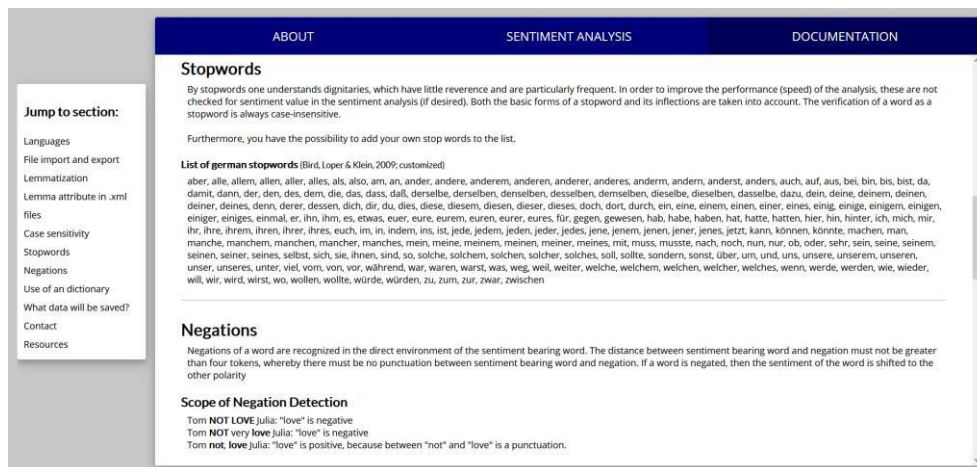


Abbildung 38: Documentation-Page

6.4.4.4 Upload-Page

Neben Tooltips (Abbildung 39), welche dem Nutzer wichtige Hilfestellungen leisten können, gibt es nun auch Hinweise (Abbildung 40), die von besonderer Priorität sind. Diese Informationen werden direkt und dauerhaft angezeigt. So kann der Nutzer direkt erkennen, ob beispielsweise mehrere Dateien hochgeladen werden können und welche Dateiformate erlaubt sind.

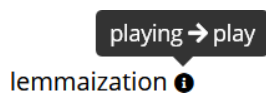


Abbildung 39: Tooltip

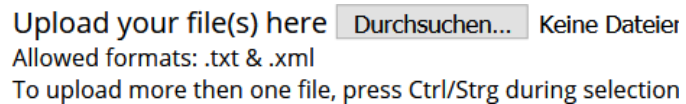


Abbildung 40: Note

Sowohl für das Lexikon BAWL-R (Abbildung 41) als auch persönliche Lexikon stehen dem Nutzer nun Notes zur Verfügung, die nach ausgewähltem Lexikon automatisch eingeblendet werden.

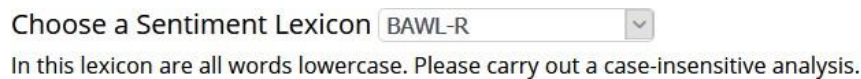


Abbildung 41: Note für BAWL-R

Unter „More Options“ finden sich Einstellungen, mithilfe der Nutzer seine Analyse an seine Bedürfnisse anpassen kann (Abbildung 42).

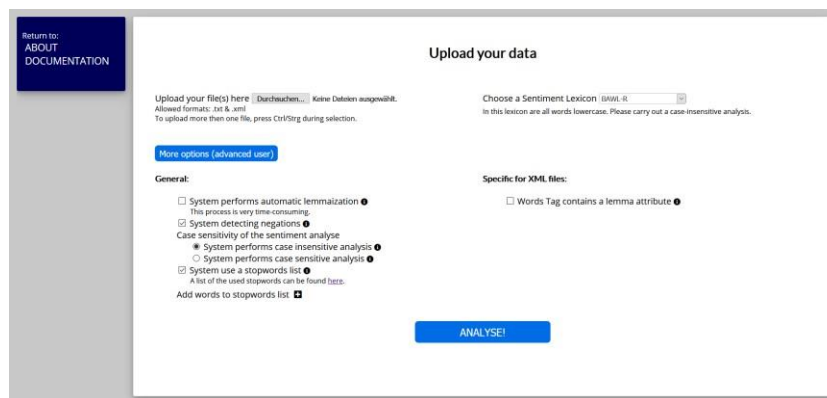


Abbildung 42: Upload-Page mit erweiterten Optionen

Um der User vor Fehler zu bewahren, gibt es mehrere Validierungen, die vor und während der Analyse stattfinden. Sollte der Nutzer vergessen, ein erforderliches Feld auszufüllen, wird eine browsereigene Fehlermeldung (Abbildung 43) bei dem zugehörigen Pflichtfeld angezeigt. Ist während der Analyse eine Eingabe nicht valide, kehrt der Benutzer zur Upload-Page zurück und ihm wird eine entsprechende Benachrichtigung (Abbildung 44) visualisiert. Dies kann bei falschen Dateiformaten, ungültigen Polaritätssärken (Sting ist nicht in Float konvertierbar) und einer Leerzeicheneingabe des Lemma-Attributs geschehen.

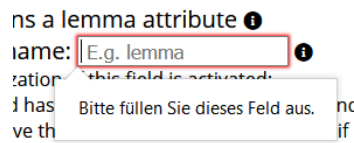


Abbildung 43: Pflichtfeldprüfung

Allowed text file types: txt, xml (file: 1749_Lessing_Die_Juden.json)



Abbildung 44: Falsches Dateiformat für die Textdatei

6.4.4.5 Result-Page

Die Result-Page (Abbildung 45) stellt die komplexeste aller Pages der Applikation dar. Aus diesem Grund wird hierbei nochmals auf das Layout eingegangen. Ganz links befindet sich ein Aside, in dem sich zum einen die Navigation und zum anderen die gewählten Parameter für die Sentiment Analysis zu sehen ist. Daneben ist sich der Distant-Reading Bereich. Hier kann der Nutzer zwischen den einzelnen Dokumenten wechseln und diese in Gruppen mithilfe Drag and Drop gruppieren (Abbildung 46). In diesem Gebiet ist es dem Benutzer möglich, die Dateien zu sortieren. Überdies kann er mithilfe eines Buttons die hochgeladenen Texte und erstellen Ordner miteinander vergleichen. Auf der rechten Seite befindet sich der Close-Reading Bereich. Je nachdem, ob der Benutzer derzeit ein Element, sei es eine Textdatei oder ein Ordner, betrachtet oder mehrere miteinander vergleicht, passt sich der Text des Close-Reading Headers, sowie dessen Hintergrundfarbe an den aktuellen Fall an. Betrachtet der Nutzer ein einzelnes Element, so steht im Header der Name des Elements und besitzt die Hintergrundfarbe blau. Dagegen erscheint im Vergleich „Compare files and folders“ auf einem grünen Grund. Nacheinander wird auf die Visualisierung von einzelnen Textdateien und Ordner eingegangen. Anschließend auf die Ansicht, in der man einzelne Elemente miteinander vergleichen kann.

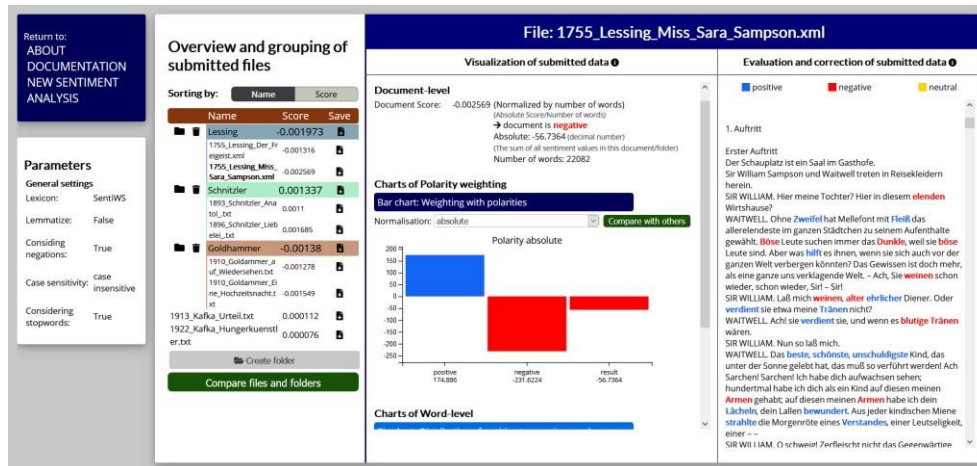


Abbildung 45: Result-Page

Name	Score	Save
Lessing	-0.001973	
1755_Lessing_Der_Fr eigeist.xml	-0.001316	
1755_Lessing_Miss_S ara_Sampson.xml	-0.002569	
Schnitzler	0.001337	
1893_Schnitzler_Ana to!,txt	0.0011	
1896_Schnitzler_Lieb ele!,txt	0.001685	
Goldhammer	-0.00138	
1910_Goldammer_a uf_Wiedersehen.txt	-0.001278	
1910_Goldammer_Ei ne_Hochzeitsnacht.t xt	-0.001549	
1913_Kafka_Urteil.txt	0.000112	
1922_Kafka_Hungerkuenstl er.txt	0.000076	

Create folder

Abbildung 46: Dokumente und Ordner

Untersucht der Nutzer eine Textdatei, so wird der Close-Reading Bereich in zwei Areale aufgeteilt (Abbildung 47). Auf der rechten Seite hat der Anwender die Analyse des Textes genauer zu betrachten und auf der linken Seite findet der Nutzer Visualisierungen der explorierten Informationen.

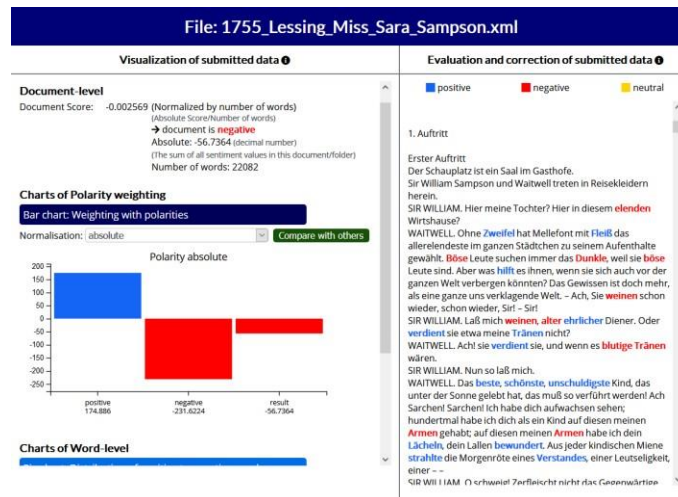


Abbildung 47: Close-Reading des Textes „Der Freigeist“

Bei der Darstellung des untersuchten Textes befindet sich im oberen Bereich eine passende Überschrift, sowie eine Legende (Abbildung 48), welche die Farbcodierung der SBWs visualisiert. Da auch die Möglichkeit besteht, dass in einem Lexikon neutrale Wörter vorkommen, werden Begriffe, die im Lexikon existieren, jedoch neutral sind, in gelber Schrift dargestellt.

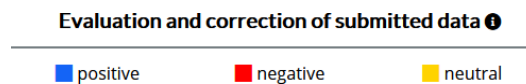


Abbildung 48: Überschrift und Legende

Jedes der identifizierte SBWs ist mit einem Tooltip versehen. Hovort der Nutzer über das entsprechende Wort, so kann diese genaueren Informationen zu dem SBW aus dem Tooltip entnehmen. In diesem Tooltip ist gespeichert, welches der Lexikon-Schlüssel ist, also welches Wort mit dem Lexikon überprüft wurde, was die dazugehörige Polaritätsstärke ist, einen Wahrheitswert, obder Begriff eine Negation beinhaltet, und das Endergebnis des Sentiments (Abbildung 49). Wird das Wort negiert, wo wird zusätzlich noch die Negation hervorgehoben (Abbildung 50).

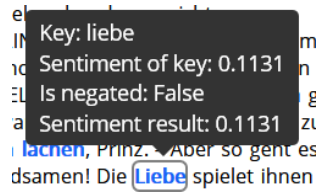


Abbildung 49: SBW ohne Negation

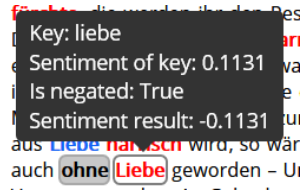


Abbildung 50: SBW mit Negation

Möchte der Nutzer nach der Analyse ein SBW bearbeiten oder einem Wort einen Sentiment-Wert hinzufügen, so hat dieser die Möglichkeit. Hierbei öffnet sich bei Klick auf das zu korrigierende Begriff ein Fenster, bei dem der Anwender die Polaritätsstärke und eine Negation eingeben kann (Abbildung 51). Bei der Bestätigung wird geprüft, ob der Sentiment-Wert eine Zahl ist. Und wenn nicht, eine entsprechende Fehlermeldung zurückgegeben (Abbildung 52). Ist der eingegabene Wert korrekt, werden alle Informationen und Grafiken direkt aktualisiert.

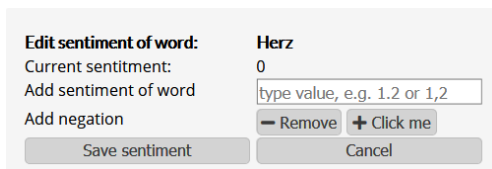


Abbildung 51: Sentiment bearbeiten

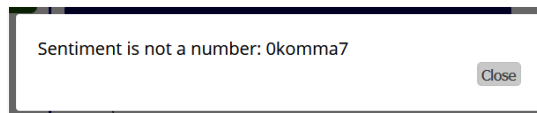


Abbildung 52: Polaritätsstärke ist keine Zahl

Darüber hinaus kann der Nutzer einen Textausschnitt mit gedrückter Maustaste markieren, um so an spezifische Details des Ausschnittes zu gelangen (Abbildung 53).

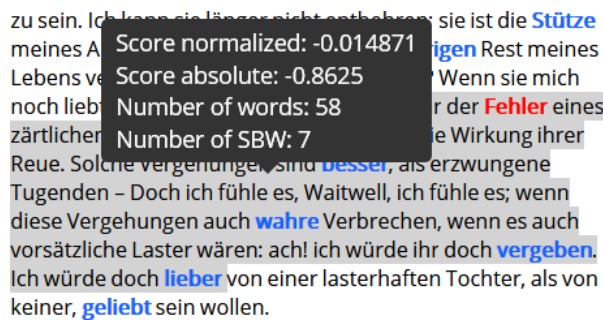


Abbildung 53: Sentiment eines Textausschnittes

Auf der linken Seite des Close-Reading Bereichs sind Informationen und Diagramme über den Sentiments innerhalb des Dokuments dargestellt und nach Analyse-Ebene geordnet, wobei die Diagramme sich in klickbaren Panels befinden. Alle nachfolgenden Informationen entstammen aus der Sentiment Analysis

der Novelle „Auf Wiedersehen!“ von Leo Goldhammer unter den Standard-einstellungen der Sentiment Analysis.

Auf der Dokumenten-Ebene befindet sich Informationen über den Document Score. Hierbei erhält der Nutzer Informationen über den normalisier-ten und absoluten Sentiment Score, die Polarität des Dokuments, sowie die An- zahl der Wörter, anhand denen der absolute Score normalisiert wurde (Abbildung 54). Zudem findet der Nutzer hier kleine Hinweise, die ihm bei dem Verständnis der Informationen helfen sollen.

```
Document Score: -0.001278 (Normalized by number of words)
(Absolute Score/Number of words)
→ document is negative
Absolute: -7.3873 (decimal number)
(The sum of all sentiment values in this document/folder)
Number of words: 5780
```

Abbildung 54: Visualisierung - Document Score

Darunter befindet sich ein Balkendiagramm, welches die Polaritätsgewichtung detailliert visualisiert. Hierbei hat der Nutzer die Möglichkeit, die Gewichtung als absoluten Wert oder als normalisierten Wert, anhand aller Wörter oder nur anhand der SBWs, aufzeigen zu lassen. Dabei wird die positive Gewichtung blau und die negative Gewichtung rot dargestellt. Das Ergebnis dieser Gewich- tung wird in der zugehörigen Polarität, blau oder rot, eingefärbt. Darüber hin- aus beinhaltet das Diagramm einen Titel, wobei hierbei dokumentiert wird, um welchen Wert die Polaritätsgewichtungen normalisiert wurden. Die untere Gra- fik zeigt die beispielhafte Darstellung der Polaritätsgewichtung anhand des normalisierten Scores. Unter jedem Balken steht jeweils der Aspekt (positiv, negativ oder neutral), sowie die darzustellende Polaritätsgewichtung (Abbildung 55). Neben den Diagrammen befindet sich ein grüner „Compare with others“ Button. Über diesen kann der Nutzer direkt zum Vergleich dieses Diagramms aller ausgewählten Elemente gelangen, ohne dass er selbst noch entsprechende Einstellungen vornehmen muss.

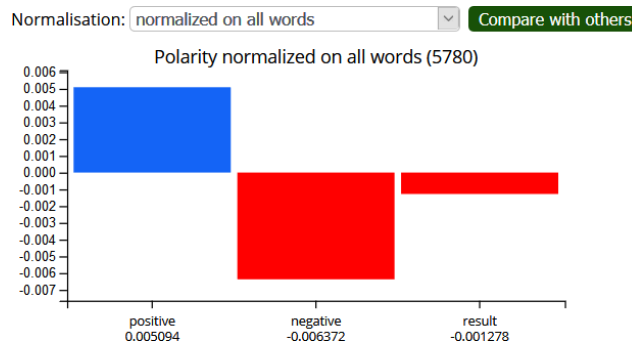


Abbildung 55: Visualisierung - Polaritätsgewichtung

Auf der Word-Ebene werden die einzelne SBWs betrachtet. Hier stehen dem Nutzer verschiedene Visualisierungsmöglichkeiten zur Auswahl. Mithilfe eines Kreisdiagramms kann der dieser die Verhältnisse zwischen positiven und negativen SBWs untersuchen. Zusätzlich hat der Nutzer hierbei die Auswahl die Verteilung von allen Wörtern anhand ihrer Polarität anzuzeigen, wie im Beispiel zu sehen ist. Jeder Kreissektor beinhaltet ein Label, das sowohl die Polarität als auch die absolute Anzahl und der prozentuale Anteil zum Grundwert darstellt. Jedes Diagramm besitzt einen Titel, in dem die entsprechende Verteilung angegeben ist (Abbildung 56).

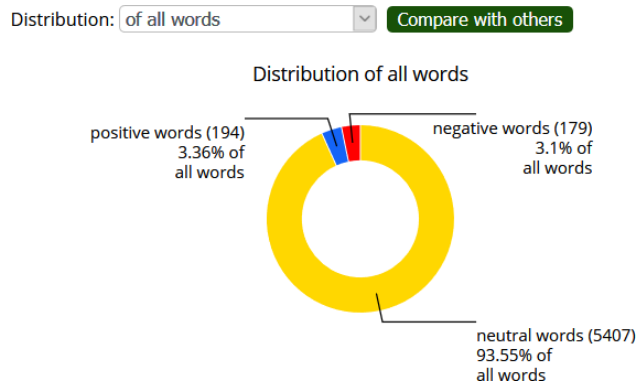


Abbildung 56: Visualisierung - Verhältnis aller Wörter

Eine weitere Visualisierungsmöglichkeit ist, das Untersuchen der einzelnen SBWs in Abhängigkeit ihre Polarität und Häufigkeit (Polaritätsstärke * Frequenz im Text). In der Wordcloud werden die SBWs angezeigt, wobei mithilfe der Schriftgröße und der Schriftfarbe sich eine Ordnung festlegen lässt (Abbildung 57). Der Nachteil einer Wordcloud ist es allerdings, dass die Fläche, welcher anhand des Scores errechnet wird, kaum vom Auge erfasst wird und auch längere Wörter größer erscheinen (Holtz, 2018). Für genauere Informatio-

nen über die Stärke eines Begriffs in einer Wordcloud, sind diese mit Tooltips versehen (Abbildung 58). Aufgrund der beschriebenen Problematik einer Wordcloud, wird dem Nutzer als alternative ein Balkendiagramm visualisiert (Abbildung 59). Dieses Diagramm stellt aufgrund dessen Darstellungsform weniger SBW dar. Wird ein Wort negiert, so wirkt diese Negation sich auf die darzustellenden Diagramme aus. Weshalb in positiven Graphen, negierte negative Wörter zu finden sind. Dasselbe gilt für negierte positive Wörter in negative Graphen.



Abbildung 57: Visualisierung - Stärkste negative SBWs (Wordcloud)



Abbildung 58: Tooltip für das Wort "Angst" in der Wordcloud

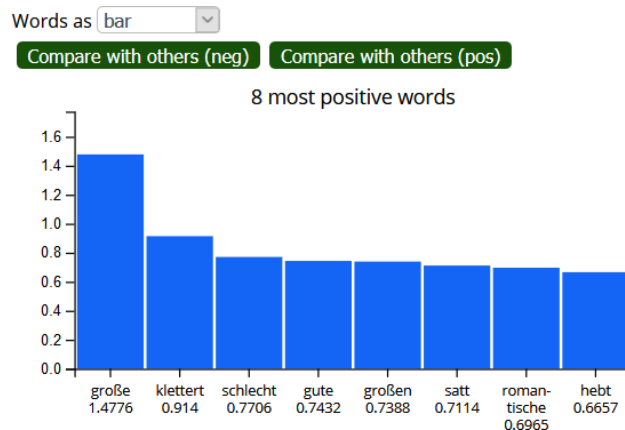


Abbildung 59: Visualisierung - Stärkste positive SBWs (Balkendiagramm)

Zuletzt stehen dem Nutzer Graphen auf der Satzebene zur Verfügung. Hierbei hat der Nutzer, wie auf der Wordebene, die Möglichkeit die Verteilung von positive, negativen und neutralen Sätzen in Form eines Kuchendiagramms zu visualisieren. Das folgende Beispiel zeigt das Verhältnis von Sentiment bearing sentences (SBSs) in der untersuchten Novelle (Abbildung 60).

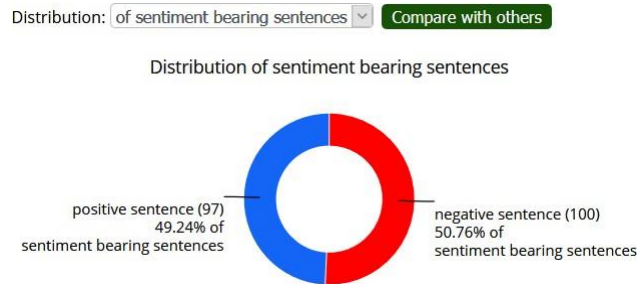


Abbildung 60: Visualisierung – Verhältnis der SBSs

Hier wird die Entwicklung des Sentiments in Form einer Timeline dargestellt (Abbildung 61). Dabei hat der Nutzer die Möglichkeit, sowohl den absoluten Score eines Satzes als auch dessen Normalisierung zu betrachten. Eine Besonderheit hierbei ist, dass der Nutzer die Sätze in festgelegte Einheiten gruppieren kann. So ist sichergestellt, dass in jeder Einheit sich die selbe Anzahl von Sätzen befindet. Während der Nutzer mit der Maus entlang der Timeline fährt, wird immer für die aktuelle Einheit oder der aktuelle Satz der absolute Score, der normalisierte Score, die Anzahl von Wörtern und der dazugehörige Text dargestellt. Klickt der Nutzer nun eine Stelle in der Timeline an, so scrollt die Evaluation des Texten direkt zur entsprechenden Stelle. Auf diese Weise kann der Nutzer auch Ausreiser erfassen und gegebenenfalls korrigieren. Im nachfolgenden Beispiel wurden jeweils fünf Sätze zu einer Einheit zusammengefasst und der normalisierte Score hinzugezogen.

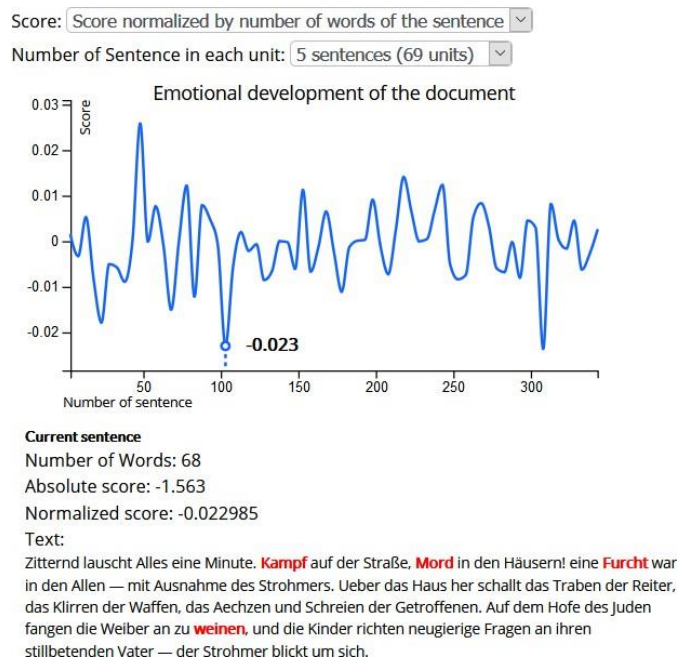


Abbildung 61: Visualisierung - Timeline des Sentiments eines Dokuments

Die letzte Visualisierung, die dem Nutzer für einzelne Dokumente nutzen kann, ist eine tabellarische Auflistung der an den stärksten positiven und negativ gewichteten Sätzen (Abbildung 62). Auch hier kann der Anwender die Sätze nach Ihrem normalisierten und nach ihrem absoluten Sentiment Score ordnen. Um eine Übersichtlichkeit beizubehalten, werden hierbei jeweils nur die zehn stärksten berücksichtigt. Nachfolgend werden die Sätze anhand ihres absoluten Scores dargestellt Jedoch ist aus Platzgründen dies nur ein Teilausschnitt. Klickt der Nutzer auf einen Satz, so scrollt das System direkt zur entsprechenden Stelle im Text.

Score:

positive sentences	value	negative sentences	value
Und der Lohn für Beide sollte nicht schlecht sein; waren ja in Arbeit gepreßt, dachten zu fordern deßhalb, daß es dem Juden ans Leben gehen solle.	0.7666	— Er fühlte sich geschlagen wie mit einem Beile, sich wollte er wehren davor und konnte es nicht; er hörte sich ächzen, und es ächtzte sein Meister aus ihm; vor der Wucht auf ihm fühlte er sich in die Tiefe gelastet — des Ofens Gewölbe brach ein unter der Last — seine Flammen umwirbelten ihn —	-1.5576
Unsre Feierabendsgesellen — nur für die Tage der Noth waren sie in Arbeit gesprochen — der Strohmer also und der rüstige Bursch waren bald gute	0.7529	tiefer und tiefer in eine unabschbare Tiefe sank er hinunter, er und der	

Abbildung 62: Visualisierung – Sätze mit der größten Polaritätsstärke

Die Darstellung des Close-Reading für Ordner ist übersichtlich gestaltet, da der Nutzer keinen Text betrachten kann (Abbildung 63). Und die Timeline des Sentiments auf der Satzebene und die stärksten Sätze auf der Satzebene dem Anwender nicht zur Verfügung stehen.

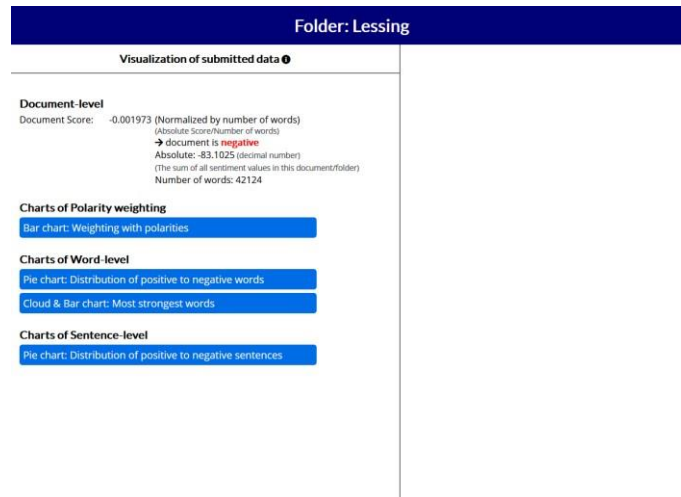


Abbildung 63: Close-Reading des Ordners „Lessing“

Der Nutzer kann sowohl die Ergebnisse Visualisierungen von Textdateien als auch die erstellten Grafiken von Ordner herunterladen. Um die Analyse des Textes herunterzuladen, stehen dem Benutzer die Dateiformate .xml und .csv zur Verfügung. Für Visualisierungen wird die Dateiendung .png angeboten. Um ein Ergebnis herunterzuladen bewegt der Nutzer seine Maus über das Save Icon des zugehörigen Elements im Distant-Reading Bereich. Anschließend erscheint ein Tooltip, mithilfe der User das Dateiformat wählen kann (Abbildung 64). Zusätzlich steht dem Nutzer das Herunterladen der Visualisierungen als Zip-Datei durch einen Rechtsklick auf eine Grafik der entsprechenden Datei zur Verfügung.

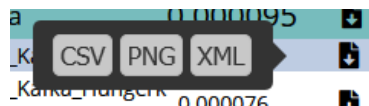


Abbildung 64: Ergebnisse herunterladen

Zuletzt wird der Bereich erklärt, mithilfe der Nutzer die einzelnen Dokumente und Ordner miteinander vergleichen kann. Um zu dieser Funktion zu gelangen kann der Anwender auf den grünen Button im Distant-Reading Bereich klicken. Er kann darüber hinaus auch die kleineren grünen Buttons nutzen, welche sich

6. Entwurf und Implementierung der Software

neben den Grafiken befinden. Auf diese Weise werden direkt die zugehörigen Darstellungen ausgewählt. Insgesamt stehen dem Nutzer zwölf Informationsvisualisierungen zum Vergleich zur Verfügung. Hierbei handelt es sich die Informationen, die sowohl von Textdateien als auch von Ordnern bereitgestellt werden. In der nachfolgenden Abbildung wird der Vergleich der Polaritätsgewichtungen des absoluten Scores der erstellten Ordner Lessing, Schnitzler, Goldhammer und Kafka durchgeführt (Abbildung 65).

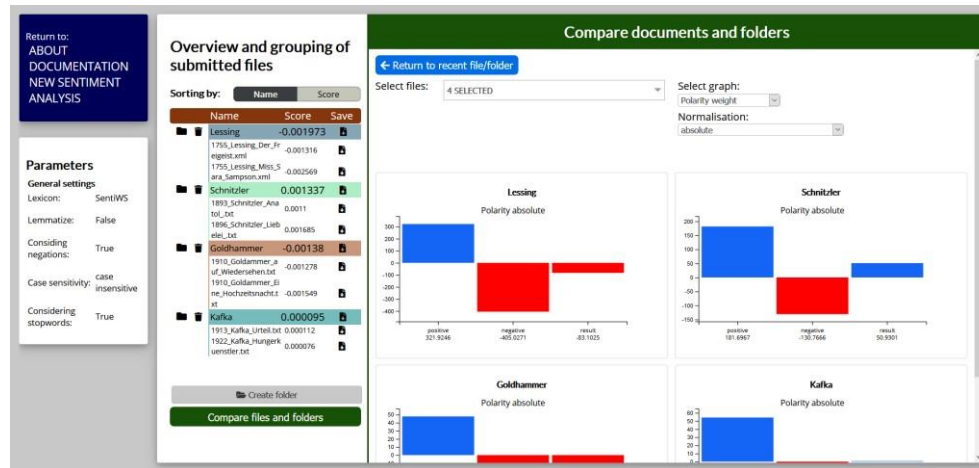


Abbildung 65: Vergleich der Polaritätsgewichtungen

7 Diskussion

Die in dieser Arbeit wurde eine Webanwendung^{29 30} zur lexikonbasierten Sentiment Analysis im Rahmen des User-Centered-Design Prozess entwickelt und evaluiert. Wobei Negationen in einem Bereich von vier Wörtern erkannt werden. Aber auch Negationspräfixe und -Suffixe werden berücksichtigt. Zusätzlich besteht die Anwendung dadurch, dass nicht nur die Funktionalität und Darstellung der Analyse entwickelt wurden. Vielmehr stehen dem Nutzer Informationen zur Sentiment Analysis und dessen Aspekte in Form einer kleinen Informationsseite zur Verfügung. Die summative Studie konnte zeigen, dass die entwickelte Anwendung verfügt über eine sehr gute Usability und eine überdurchschnittliche Attraktivität. Bei der Entwicklung der Anwendung stellte vor allem die verschiedenen Browser eine Problemantik dar. Da je nach Browser und dessen Version teilweise die Interaktion nicht sehr flüssig ist, Elemente unterschiedlich dargestellt werden oder die Funktionalität beeinträchtigt wird. Ebenfalls stellte die Recherche für den Algorithmus zum Aufspüren von Negationen eine Schwierigkeit dar, da sich viele Studien auf englischsprachige Texte beziehen. Momentan wurde die Anwendung für eher kurze Texte von bis zu rund 40.000 Wörter und vergleichsweise kleinen Textsammlungen mit maximal 20 Texten getestet. Für sehr lange Texte kann es deshalb zu Performanceeinbußen seitens der Interaktion kommen. Gut dagegen lief die Implementierung allgemein. Auch wenn die Implementierung einen relativ großen Umfang besitzt, konnten bereits vieles aus dem Studium gewonnen Kenntnisse zurückgegriffen werden.

²⁹ Eine Demonstration der Anwendung (Video) findet sich in den elektronischen Daten unter /10_Sonstiges.

³⁰ Die entwickelte Anwendung ist verfügbar unter:

8 Limitations

Wie bereits beschrieben wird bei der entwickelten Anwendung das Sentiment unabhängig vom Kontext ermittelt. Aus diesem Grund kann beispielsweise Ironie nicht erkannt werden und zu falschen Werten führen (Abbildung 66). Für die Analyse ist es zudem wichtig, dass das zu Überprüfende Wort im Lexikon vermerkt ist. Aus diesem Grund kann beispielsweise eine implizit geäußerte Meinung nicht erkannt werden, da dort kein entsprechenden SBW vorhanden ist (Abbildung 66). So ist es nicht verwunderlich, dass die Genauigkeit der Anwendung stark mit der Genauigkeit des verwendeten Sentiment-Lexikons verknüpft ist.

Beispiel 1 (Ironie):
 Julia balanciert mehrere gestapelte Bücher. Jedoch **fallen** ihr diese Bücher zu Boden.
 „**Toll** gemacht!“, ruft ihr die Bibliothekarin zu.

Beispiel 2 (implizit geäußerte Meinung):
 Als Lea´s Blick Tom streift, bekommt sie einen roten Kopf.

Abbildung 66: Beispiele

Um Negationen zu erkennen, wird ein simpler Algorithmus angewandt. Dieser erkennt die Negationen innerhalb eines festgelegten Bereiches. Dies kann jedoch befolgenden Szenarien zu Schwierigkeiten führen:

1. Die Negation steht nahe eines SBW, bezieht sich jedoch nicht auf dieses.
2. Negation und SBW stehen weit voneinander entfernt.
3. Zwischen Negation und SBW befindet sich ein Satzzeichen.

Zudem besteht die Gefahr, dass sollte eine Lemmatisierung genutzt werden, diese nicht immer korrekt abläuft. Anbei sind zwei Beispiele, in denen Textblob_de nicht akkurat ist (Tabelle 2).

Input	Output	Erwartet
humorvolle	Humorvolle	humorvoll
Begünstigungen	Begünstigungen	Begünstigung

Tabelle 2: Textblob_de

9 Future Work

Die entwickelte Anwendung lässt sich noch weiter ausgebaut und verbessert werden. So bietet sich an, den Algorithmus zum Erkennen der SBW und der sich darauf beziehenden Negationen noch weiter zu optimieren. Ebenfalls ist eine Verbesserung der Identität und die Stimulation der hedonischen Qualität sicher von Relevanz. Neben Optimierungen gibt es noch weitere Funktionalitäten, um die man die Anwendung erweitern kann. So bietet es sich an, eine weitere Sprache bereitzustellen oder ein weiteres Wörterbuch hinzuzufügen. Ebenfalls fehlt noch eine passende Benennung der Anwendung. Passende Namen könnten dabei „Motis“ (abgeleitet aus dem lateinischen Wort für „Emotion“) oder „SenText“ (Schmelzwort aus „Sentiment“ und „Text“) sein. Da die Namensgebung der Anwendung jedoch ein wichtiger und endgültiger Vorgang ist, ist es hilfreich die Meinung anderer Personen einfließen zu lassen. Im Augenblick können DH-Forscher mithilfe dieser Anwendung eine Sentiment Analysis auf Basis eines Lexikons durchführen. Wobei hierbei der Fokus auf die sich im Korpus oder Text befindenden SBW liegt. Die Anwendung ist vor allem dafür geeignet, sich einen Überblick über die Texte zu verschaffen und den Text anhand seiner SBW zu explorieren. Im weitesten Sinne ist das Tool ebenfalls verwendbar, um einen Text anhand von maximal drei Wordklassifikationen (positiv, neutral, negativ oder a, b, c) zu analysieren und dort das Aufkommen der Gruppen zu betrachten.

10 Ausblick

Der nächste Schritt wird das Deployen der Anwendung sein, um diese der Öffentlichkeit zugänglich zu machen. Während des summativen Usability-Tests fand die Anwendung großen Anklang, sodass das Tool in einem Masterkurs der Universität Hamburg im Bereich der Digital Humanities soll das Tool daraufhin evaluiert und auf dessen Forschungsnutzen untersucht werden.

Literaturverzeichnis

- Abran, A., Moore, J. W., Bourque, P., & Dupuis, R. (Eds.). (2004). *SWEBOK. Guide to the Software Engineering Body of Knowledge. 2004 Version*. Los Alamitos, California: IEEE Computer Society.
- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment Analysis of Twitter Data. *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, 30–38. Portland, Oregon: Association for Computational Linguistics.
- Aggarwal, C. C., & Zhai, C. (2012a). A Survey of Text Classification Algorithms. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 163–222). Boston, Massachusetts: Springer US.
- Aggarwal, C. C., & Zhai, C. (2012b). A Survey of Text Clustering Algorithms. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 77–128). Boston, Massachusetts: Springer US.
- Aggarwal, C. C., & Zhai, C. X. (2012c). An introduction to text mining. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 1–10). Boston, Massachusetts: Springer US.
- Althoff, G. (1978). Möglichkeiten und Grenzen Elektronischer Datenverarbeitung bei der Erforschung der Geschichte des Mittelalters. *Computers and the Humanities*, 12, 97–107.
- Andreasen, M. S., Nielsen, H. V., Schröder, S. O., & Stage, J. (2007). What Happened to Remote Usability Testing? An Empirical Study of Three Methods. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1405–1414. New York, New York: Association for Computing Machinery.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. Beijing: O'Reilly Media Inc.
- Blake, C. (2011). Text mining. *Annual Review of Information Science and Technology*, 45(1), 121–155.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. In P. W. Jordan, I. L. McClelland, B. Thomas, & B. A. Weerdmeester (Eds.), *Usability evaluation in*

- industry*. London, United Kingdom: Taylor & Francis Ltd.
- Buechel, S., Hellrich, J., & Hahn, U. (2016). Feelings from the Past—Adapting Affective Lexicons for Historical Emotion Analysis. *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, 54–61. Osaka, Japan: The COLING 2016 Organizing Committee.
- Burghardt, M., & Wolff, C. (2014). Humanist-Computer Interaction: Herausforderungen für die Digital Humanities aus Perspektive der Medieninformatik. In *DHd Workshop: Informatik und die Digital Humanities*. Leipzig.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2), 15–21.
- Crain, S. P., Zhou, K., Yang, S.-H., & Zha, H. (2012). Dimensionality Reduction and Topic Modeling: From Latent Semantic Indexing to Latent Dirichlet Allocation and Beyond. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 129–161). Boston, Massachusetts: Springer US.
- D’Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, 125(3), 26–33.
- Dadvar, M., Hauff, C., & de Jong, F. (2011). Scope of negation detection in sentiment analysis. *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR 2011)*, 16–20. Amsterdam: University of Amsterdam.
- Ding, X., Liu, B., & Yu, P. S. (2008). A Holistic Lexicon-based Approach to Opinion Mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 231–240. New York, NY, USA: Association for Computing Machinery Association for Computing Machinery Association for Computing Machinery.
- Eberhard-Yom, M. (2010). *Usability als Erfolgsfaktor. +Grundregeln +User Centered Design +Umsetzung*. Berlin: Cornelsen Verlag.
- Farooq, U., Mansoor, H., Nongailard, A., Ouzrout, Y., & Qadir, M. A. (2017). Negation Handling in Sentiment Analysis at Sentence Level. *Journal of Computers*, 12(5), 470–478.

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37.
- Feldman, R., & Dagan, I. (1995). Knowledge Discovery in Textual Databases (KDT). *KDD-95 Proceedings*, 112–117. AAAI Press.
- Flüh, M. (2019). Sentimentanalyse. Retrieved February 24, 2020, from forTEXT. Literatur digital erforschen website: <https://fortext.net/routinen/methoden/sentimentanalyse>
- Gibbs, F., & Owens, T. J. (2012). Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly*, 6(2).
- Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *International Conference on Web and Social Media Papers*, 7(21), 219–222. Boulder, Colorado.
- Godsay, M. (2015). The Process of Sentiment Analysis: A Study. *International Journal of Computer Applications*, 126(7).
- Hammerschall, U., & Beneken, G. H. (2013). *Software requirements*. München: Pearson Studium.
- Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge, Massachusetts: MIT Press.
- Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung* (pp. 187–196). Stuttgart: B. G. Teubner.
- Hearst, M. A. (1999). Untangling Text Data Mining. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, 3–10. USA: Association for Computational Linguistics.
- Heyer, G., Quasthoff, U., & Wittig, T. (2006). *Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*. Herdecke: W3L-Verlag.
- Hippner, H., & Rentzmann, R. (2006). Text Mining. *Informatik-Spektrum*, 29(4), 287–290.

- Hotho, A., Nürnberger, A., & Paaß, G. (2005). A Brief Survey of Text Mining. *LDV-Forum*, 20(1), 19–62.
- Jacobsen, J., & Meyer, L. (2017). *Praxisbuch Usability und UX*. Bonn: Rheinwerk Computing.
- Jia, L., Yu, C., & Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 1827–1830. New York, NY, USA: Association for Computing Machinery.
- Jindal, N., & Liu, B. (2008). Opinion Spam and Analysis. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 219–230. New York, New York: Association for Computing Machinery.
- Kim, S.-M., & Hovy, E. (2004). Determining the Sentiment of Opinions. *Proceedings of the 20th International Conference on Computational Linguistics*, 1367–1373. USA: Association for Computational Linguistics.
- Kim, S.-M., & Hovy, E. (2006). Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, 1–8. USA: Association for Computational Linguistics.
- Kodratoff, Y. (1999). Knowledge discovery in texts: A definition, and applications. In Z. W. Raś & A. Skowron (Eds.), *Foundations of Intelligent Systems* (pp. 16–29). Berlin: Springer.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter Sentiment Analysis: The Good the Bad and the OMG! *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. USA: Association for the Advancement of Artificial Intelligence.
- Kumar, A., & Sebastian, T. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, 4(10), 1–14.
- Lemke, M., & Wiedemann, G. (2016). Einleitung Text Mining in den Sozialwissenschaften. In M. Lemke & G. Wiedemann (Eds.), *Text Mining in den Sozialwissenschaften: Grundlagen und Anwendungen zwischen qualitativer*

- und quantitativer Diskursanalyse* (pp. 1–13). Wiesbaden: Springer Fachmedien Wiesbaden.
- Liu, B. (2010a). Sentiment analysis: A multi-faceted problem. *IEEE Intelligent Systems*, 25(3), 76–80.
- Liu, B. (2010b). Sentiment analysis and subjectivity. In N. Indurkha & F. J. Damerau (Eds.), *Handbook of Natural Language Processing* (2nd ed., pp. 627–666). Boca Raton, Florida: CRC Press.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. San Rafael, California: Morgan & Claypool Publishers.
- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. New York, NY, USA: Cambridge University Press.
- Liu, B. (2017). Many Facets of Sentiment Analysis. In E. Cambria, D. Das, S. Bandyopadhyay, & A. Feraco (Eds.), *A Practical Guide to Sentiment Analysis. Socio-Affective Computing* (Vol. 5, pp. 11–39). Cham: Springer.
- Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 415–463). Boston, Massachusetts: Springer US.
- Losiewicz, P., Oard, D. W., & Kostoff, R. N. (2000). Textual Data Mining to Support Science and Technology Management. *Journal of Intelligent Information Systems*, 15, 99–119.
- Lowdermilk, T. (2013). *User-Centered Design. A Developer's Guide to Building User-Friendly Applications*. USA: O'Reilly Media, Inc.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
- Mehler, A., & Wolff, C. (2005). Einleitung: Perspektiven und Positionen des Text Mining. *LDV-Forum*, 20(1), 1–18.
- Mohammad, S. (2011). From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales. *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 105–114. Portland, Oregon: Association for Computational Linguistics.

- Moreno-Ortiz, A. (2017). Lingmotif: Sentiment analysis for the digital humanities. *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 73–76. Valencia, Spain: Association for Computational Linguistics.
- Nalisnick, E. T., & Baird, H. S. (2013). Character-to-Character Sentiment Analysis in Shakespeare's Plays. *Proceedings Of the 51st Annual Meeting Of the Association for Computational Linguistics*, 479–483. Association for Computational Linguistics.
- Nenkova, A., & McKeown, K. (2012). A Survey of Text Summarization Techniques. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 43–76). Boston, MA: Springer US.
- Nielsen, J. (1994). 10 Usability Heuristics for User Interface Design. Retrieved February 24, 2019, from <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Patzold, S. (2013). Personen, Namen, Lemmatisierung: Namenforschung zwischen Geschichts- und Sprachwissenschaft am Beispiel des Projekts „Nomen et Gens“. *Namenkundliche Informationen*, 101/102, 22–48.
- Petz, G. (2019). *Opinion Mining im Web 2.0: Ansätze, Methoden, Vorgehensmodell*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS-A Publicly Available German-language Resource for Sentiment Analysis. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 1168–1171. Valletta, Malta: European Language Resources Association.
- Rockwell, G., Sinclair, S., & Radzikowska, M. (n.d.). About TAPoR 3.0. Retrieved February 24, 2020, from http://tapor.ca/pages/about_tapor
- Schlaps, C., & Schares, T. (2008). Lemmatisierungspraxis und -problematik im Autorenwörterbuch am Beispiel des Goethe-Wörterbuchs. In B. Elisenda & J. DeCesaris (Eds.), *Proceedings of the 13th EURALEX International Congress* (pp. 599–606). Barcelona, Spain: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

- Schmidt, T., & Burghardt, M. (2018). An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing. *Proceedings of Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 139–149. Santa Fe, New Mexico: Association for Computational Linguistics.
- Schmidt, T., Burghardt, M., & Dennerlein, K. (2018). „Kann man denn auch nicht lachend sehr ernsthaft sein?“ – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen. In *Book of Abstracts*. DHd.
- Schmidt, T., Burghardt, M., & Wolff, C. (2018). Herausforderungen für Sentiment Analysis-Verfahren bei literarischen Texten. In M. Burghardt & C. Müller-Birn (Eds.), *INF-DH 2018*. Bonn: Gesellschaft für Informatik e.V.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1631–1642. Seattle, Washington: Association for Computational Linguistics.
- Solka, J. L. (2008). Text Data Mining: Theory and Methods. *Statistics Surveys*, 2, 94–112.
- Soper, D. (n.d.). Free Sentiment Analyzer. Retrieved November 7, 2019, from <https://www.danielsoper.com/sentimentanalysis/default.aspx>
- Sprugnoli, R., Tonelli, S., Marchetti, A., & Moretti, G. (2016). Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities*, 31(4), 762–772.
- Stanford Visualization Group. (2011). Wrangler. Retrieved March 8, 2020, from <http://vis.stanford.edu/wrangler/>
- Still, B., & Crane, K. (2017). *Fundamentals of user-centered design: A practical approach* (1st ed.). Boca Raton: CRC Press.
- Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, 36, 10–25.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-

- based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
- Thelwall, M. (2017). The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength. In J. Holyst (Ed.), *Cyberemotions. Understanding Complex Systems*. Cham: Springer.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.
- Varghese, R., & Jayasree, M. (2013). A survey on sentiment analysis and opinion mining. *International Journal of Research in Engineering and Technology*, 2(11), 312–317.
- Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538.
- Volkova, E. P., Mohler, B., Meurers, D., Gerdemann, D., & Bülthoff, H. H. (2010). Emotional Perception of Fairy Tales: Achieving Agreement in Emotion Annotation of Text. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 98–106. Los Angeles, California: Association for Computational Linguistics.
- Waltinger, U. (2010). GERMANPOLARITYCLUES: A Lexical Resource for German Sentiment Analysis. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*. Valletta, Malta: electronic proceedings.
- Weinrich, H. (2007). *Textgrammatik der deutschen Sprache* (4th ed.). Hildesheim: Georg Olms.
- Wiegand, M., Wolf, M., & Ruppenhofer, J. (2018). Negation Modeling for German Polarity Classification. In G. Rehm & T. Declerck (Eds.), *Language Technologies for the Challenges of the Digital Age* (pp. 95–111). Cham: Springer International Publishing.
- Wilson, C. (2013). *Interview techniques for UX practitioners: A user-centered design method*. Elsevier Science.

- Zehe, A., Becker, M., Hettinger, L., Hotho, A., Reger, I., & Jannidis, F. (2016). Prediction of Happy Endings in German Novels based on Sentiment Information. In P. Cellier, T. Charnois, A. Hotho, S. Matwin, M.-F. Moens, & Y. Toussaint (Eds.), *3rd Workshop on Interactions between Data Mining and Natural Language Processing* (pp. 9–16). Riva del Garda, Italy.
- Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. *HP Laboratories*, 89.

Elektronische Daten

Aufgrund der aktuellen Umstände werden elektronische Daten über folgenden Link zugänglich gemacht:

Sobald es möglich ist wird ein gedrucktes Exemplar der Arbeit zur Archivierung nachgereicht, wobei die elektronischen Daten mittels eines Datenträgers gesichert sind.

Der Inhalt des freigebenden Ordners ist folgendermaßen strukturiert:

/1_Ausarbeitung	Schriftliche Ausarbeitung als PDF
/2_Code	Der Quellcode der entwickelten Anwendung
/3_Studie	Formative und Summative Studie
/3_Studie/1_Formativ	Formative Studie und Aufnahmen als MP4
/3_Studie/2_Summativ	Summative Studie und Aufnahmen als MP4
/4_Quellen	In der Arbeit zitierte Quellen im PDF-Format
/5_Bilder	Alle erstellen Bilder
/6_Vortrag	Folien des Antrittsvortrag
/7_Anforderungsanalyse	Leitfaden für das Interview, Aufnahmen der Interviews als MP3 und Anforderungsdokumentation
/8_Lexika	Verwendete Lexika für die Anwendung
/9_Testdaten	Texte, mit denen die Anwendung getestet wurde
/10_Sonstiges	Informationen bezüglich des deployen der Anwendung, sowie eine Demonstration der Anwendung als MP4 und die ersten Scribbles der Anwendung

Erklärung zur Urheberschaft

Ich habe die Arbeit selbständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie alle Zitate und Übernahmen von fremden Aussagen kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt.

Die vorgelegten Druckexemplare und die vorgelegte digitale Version sind identisch.

Ort, Datum

Unterschrift

Erklärung zur Lizenzierung und Publikation dieser Arbeit

Name: Johanna Dangel

Titel der Arbeit: Entwicklung und Evaluation eines Tools zur lexikonbasierten Sentiment Analysis für die Digital Humanities

Hiermit gestatte ich die Verwendung der **schriftlichen Ausarbeitung** zeitlich unbegrenzt und nicht-exklusiv unter folgenden Bedingungen:

- Nur zur Bewertung dieser Arbeit
- Nur innerhalb des Lehrstuhls im Rahmen von Forschung und Lehre
- Unter einer Creative-Commons-Lizenz mit den folgenden Einschränkungen:
 - BY – Namensnennung des Autors
 - NC – Nichtkommerziell
 - SA – Share-Alike, d.h. alle Änderungen müssen unter die gleiche Lizenz gestellt werden.

(An Zitaten und Abbildungen aus fremden Quellen werden keine weiteren Rechte eingeräumt.)

Außerdem gestatte ich die Verwendung des im Rahmen dieser Arbeit erstellen

Quellcodes unter folgender Lizenz:

- Nur zur Bewertung dieser Arbeit
- Nur innerhalb des Lehrstuhls im Rahmen von Forschung und Lehre
- Unter der CC-0-Lizenz (= beliebige Nutzung)
- Unter der MIT-Lizenz (= Namensnennung)
- Unter der GPLv3-Lizenz (oder neuere Versionen)

(An explizit mit einer anderen Lizenz gekennzeichneten Bibliotheken und Daten werden keine weiteren Rechte eingeräumt.)

Ich willige ein, dass der Lehrstuhl für Medieninformatik diese Arbeit – falls sie besonders gut ausfällt - auf dem Publikationsserver der Universität Regensburg veröffentlichen lässt.

Ich übertrage deshalb der Universität Regensburg das Recht, die Arbeit elektronisch zu speichern und in Datennetzen öffentlich zugänglich zu machen. Ich übertrage der Universität Regensburg ferner das Recht zur Konvertierung zum Zwecke der Langzeitarchivierung unter Beachtung der Bewahrung des Inhalts

Erklärung zur Lizenzierung und Publikation dieser Arbeit

Ich erkläre außerdem, dass von mir die urheber- und lizenzrechtliche Seite (Copyright) geklärt wurde und Rechte Dritter der Publikation nicht entgegenstehen.

Ja, für die komplette Arbeit inklusive Anhang

Ja, für eine um vertrauliche Informationen gekürzte Variante (auf dem Datenträger beigefügt)

Nein

Sperrvermerk bis (Datum)

Ort, Datum

Unterschrift