

Gottfried Wilhelm
Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Praktische Informatik
Fachgebiet Software Engineering

Anforderungen verschiedener Stakeholdergruppen an die Stimmungsanalyse in Softwareprojekten

Masterarbeit

im Studiengang Informatik

von

Natalie Dehn

Prüfer: Prof. Dr. Kurt Schneider
Zweitprüfer: Dr. Jil Klünder
Betreuer: Jakob Droste, Alexander Specht

Hannover, 11.04.2023

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig und ohne fremde Hilfe verfasst und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keinem anderen Prüfungsamt vorgelegen.

Hannover, den 11.04.2023

Natalie Dehn

Zusammenfassung

Die Softwareentwicklung ist eine Disziplin, die immer komplexer wird. Da sie so gut wie ausschließlich in Teams stattfindet, die nicht selten auch auf mehrere Standorte oder Länder verteilt sind, ergeben sich neue Herausforderungen für die Zusammenarbeit innerhalb der Entwicklerteams, vor allem für die Kommunikation. Wie die Wissenschaft gezeigt hat, sind auch soziale Aspekte ein entscheidender Faktor für den Projekterfolg. Relevant ist dabei vor allem die Stimmung des Entwicklungsteams. Hier stellt sich die Frage, wie Stimmungen oder Emotionen gemessen werden können. Sogenannte Stimmungsanalysetools bieten die Möglichkeit, durch die Analyse von Texten und Aussagen positive und negative Stimmungen in Teams zu identifizieren und zu verbessern. Es existieren verschiedene Stimmungsanalysetools mit unterschiedlichen Eigenschaften und Verwendungszwecken. Bisher konnte allerdings noch kein Beleg dafür gefunden werden, dass solche Stimmungsanalysetools in der Industrie verwendet werden. Im Rahmen dieser Arbeit wurde sich daher ein Überblick über die verschiedenen Personengruppen verschafft, die ein Interesse daran haben könnten, solche Analysetools einzusetzen. Nach einer umfangreichen Literaturrecherche haben sich folgende Stakeholdergruppen ergeben: Softwareentwickler, Softwareprojektleiter, Management, Forscher, Requirements Engineer und Qualitätssicherung. Anschließend wurde eine Methode des CrowdRE, eine iterative Interviewstudie, angewandt, um zu bestimmen, welche Anforderungen diese Stakeholdergruppen an ein Stimmungsanalysetool im Kontext eines Softwareprojekts stellen. Auch erfolgte eine Priorisierung der Anforderungen seitens der Teilnehmer und es wurde verglichen, welche Stakeholder welche Anforderungen stellen und wie sie diese priorisieren. Die Ergebnisse dieser Arbeit zeigen, dass die Anforderungen der Teilnehmer sehr vielfältig sind und von den bisher existierenden Stimmungsanalysetools nicht abgedeckt werden. Die Anforderungen, die aus dieser Arbeit hervorgegangen sind, können als Grundlage für die Weiterentwicklung von Stimmungsanalysetools verwendet werden. Auch werden Unterschiede zwischen den verschiedenen Stakeholdergruppen aufgezeigt und verglichen, was insbesondere für die (Weiter-) Entwicklung von Tools mit bestimmten Stakeholdern als primäre Zielgruppe eingesetzt werden kann.

Abstract

Software development is a discipline that is becoming increasingly complex. Since it takes place almost exclusively in teams, which are often distributed across several locations or countries, new challenges arise in regards of collaboration within the development teams, especially in terms of communication. As science has shown, social aspects are also a decisive factor for project success. The mood of the development team is particularly relevant here. This raises the question of how moods or emotions can be measured. So-called sentiment analysis tools offer the possibility to identify and improve positive and negative moods in teams by analysing texts and statements. Various sentiment analysis tools exist with different characteristics and purposes. So far, however, no evidence has been found that these sentiment analysis tools are used in the industry. In the context of this work, therefore, an overview of the different groups of people who might have an interest in using such analysis was obtained. After an extensive literature review, the following stakeholder groups emerged: software developers, software project managers, management, researchers, requirements engineers and quality assurance. Subsequently, a method of CrowdRE, an iterative interview study, was used to determine what requirements these stakeholder groups have for a sentiment analysis tool in the context of a software project. A prioritisation of the requirements by the participants was also carried out. This way, we determined which stakeholders raise which requirements and how they prioritize them. The results of this work show that the requirements of the participants are very diverse and are not entirely covered by the existing sentiment analysis tools. The requirements that emerged from this work may be used as a basis for the further development of sentiment analysis tools. Differences between the various stakeholder groups are also highlighted and compared, which can be used in particular for the further development and maintenance of tools with specific stakeholders as the primary target group.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung	1
1.2	Lösungsansatz	2
1.3	Struktur der Arbeit	3
2	Grundlagen und Verwandte Arbeiten	5
2.1	Stimmungsanalyse	5
2.2	Beispiele für Stimmungsanalysetools	6
2.2.1	SentiStrength	6
2.2.2	SentiStrength-SE	7
2.2.3	SentiCR	7
2.2.4	EmoTxt	7
2.2.5	SEntiAnalyzer	7
2.2.6	BERT	8
2.3	Stakeholder in Softwareprojekten	8
2.3.1	Methoden zur Stakeholder-Identifikation	9
2.4	CrowdRE	11
2.4.1	Multi-Level-Approach	12
2.4.2	StakeNet	12
2.4.3	StakeRare	13
2.4.4	Iterativer Ansatz	13
2.5	Verwandte Arbeiten	14
3	Literaturrecherche	17
3.1	Auswahlkriterien für Literatur	17
3.2	Startpunkt	18
3.3	Snowballing	19
3.4	Ergänzende Datenbanksuche	19
3.4.1	Erste Suche	20
3.4.2	Zweite Suche	20
3.5	Daten aus Studie	20

4	Forschungsmethode	23
4.1	Forschungsfragen	23
4.1.1	RQ1: Identifikation von Stakeholdern	23
4.1.2	RQ2: Anforderungsanalyse und Priorisierung	23
4.1.3	RQ3: Priorisierung der Anforderungen und Unterschiede	24
4.1.4	RQ4: Vergleich mit aktuellen Tools	24
4.1.5	Abdeckung	24
4.2	Auswahl einer geeigneten Forschungsmethode	24
4.2.1	Gewünschte Eigenschaften	24
4.3	Vorstellung der gewählten Methodik	25
4.3.1	Iterative Interviewstudie	25
4.3.2	Ideen-Generierung	25
4.3.3	Ideen-Bewertung	27
4.3.4	In Vivo und Pattern Coding	28
4.4	Abgrenzung von anderen Methoden	28
4.5	Teilnehmerakquise	28
4.6	Video	28
5	Ergebnisse	31
5.1	RQ1: Identifizierte Stakeholder	31
5.1.1	Identifikation relevanter Stakeholder	31
5.1.2	Forschende	32
5.1.3	Entwickler	32
5.1.4	Projektleitung	33
5.1.5	Management	33
5.1.6	Requirements Engineer	33
5.1.7	Qualitätssicherung	33
5.1.8	Ausschluss anderer Stakeholder	34
5.2	RQ2: Anforderungen an ein Stimmungsanalysetool	34
5.2.1	Anforderungen aus der unveröffentlichten Studie	34
5.2.2	Ergebnisse aus dem ersten Interview	37
5.2.3	Ergebnisse aus dem zweiten Interview	40
5.3	RQ3: Priorisierung innerhalb der Stakeholdergruppen	41
5.3.1	Ergebnisse aus dem dritten Interview	42
5.3.2	Priorisierungen der verschiedenen Teilnehmer	51
5.4	RQ4: Abgleich mit existierenden Stimmungsanalysetools	56
5.4.1	SentiStrength	56
5.4.2	EmoTxt	60
5.4.3	SEnti-Analyzer	60
5.4.4	Zusammenfassung des Vergleichs	60

6 Diskussion	61
6.1 Interpretation der Ergebnisse	61
6.1.1 Identifizierte Stakeholdergruppen	61
6.1.2 Interpretation von Anforderungen	62
6.1.3 Interpretation von Priorisierungen	64
6.1.4 Interpretation des Abgleichs mit Stimmungsanalysetools	66
6.2 Einschränkungen und Threads to Validity	68
7 Zusammenfassung und Ausblick	71
7.1 Zusammenfassung	71
7.2 Ausblick	72
A Baumstrukturen	75
B Anforderungskatalog	97
C Video-Skript	103
D Interview-Skript	105

Kapitel 1

Einleitung

Da die Softwareentwicklung immer komplexer wird und vor allem in Teams stattfindet, ist ein wichtiger Aspekt des Erfolgs eines Softwareprojekts das soziale Miteinander aller Beteiligten. Diverse Studien haben bereits gezeigt, dass schlechte Stimmungen innerhalb eines Softwareprojekts das Ergebnis negativ beeinflussen können und dass es wichtig ist es, für ein positives Miteinander bzw. eine positive Kommunikation zwischen den Teilnehmern des Projekts zu sorgen [15][14][10]. Gemäß psychologischer Forschung haben affektive Zustände - Emotionen und Stimmungen - einen hohen Einfluss auf die kognitiven Verarbeitungsfähigkeiten, wie etwa analytisches Problemlösen, die für die Softwareentwicklung benötigt werden [13]. Um die Stimmung des Entwicklerteams zu überprüfen und anschließend auch zu verbessern, kann die sogenannte Stimmungsanalyse eingesetzt werden [29]. Die Stimmungsanalyse ist ein Mittel, um Stimmungen von Fließtexten oder Aussagen zu analysieren. Dabei wird angegeben, ob die vermittelte Stimmung eines Textes oder Satzes positiv, negativ oder neutral zu bewerten ist. Einige Tools verwenden auch eine Skala von -1 bis +1, wobei -1 als negativ, +1 als positiv und 0 als neutral bewertet wird [20]. Den Stimmungspolaritäten werden oft bestimmte Emotionen zugeordnet. Dies hat den Grund, dass diese von Menschen besser zugeordnet werden können. Positive Emotionen sind beispielsweise Freude oder Enthusiasmus - Beispiele für negative Emotionen dagegen sind Wut oder Angst [36]. Ist eine Aussage keiner bestimmten Emotion zuzuordnen, wird sie als neutral bewertet.

1.1 Problemstellung

Es wurden bereits verschiedene Stimmungsanalysetools implementiert, als Beispiel seien hier unter anderem BERT [43] oder SEnti-Analyzer [18] genannt. Bisher wurde jedoch kein Beweis gefunden, dass die Tools in der Industrie bzw. der freien Wirtschaft verwendet werden. Da sie aber so vielfältige Einsatzmöglichkeiten bieten, stellt sich die Frage, wieso das der Fall ist.

Eine Möglichkeit könnte sein, dass die Bedürfnisse und Anforderungen der potenziellen Anwender nicht erfüllt werden und das Tool daher nicht zu ihren täglichen Aufgaben und Routinen passt und sich nicht in den Arbeitsalltag integrieren lässt. Dabei ist zu erwarten, dass Anwender unterschiedliche Anforderungen oder Bedürfnisse an ein Stimmungsanalysetool stellen, je nachdem, welche Rolle sie im Softwareprojekt einnehmen. Daher sollen nun Anforderungen erhoben werden, die verschiedene Stakeholder an das Tool haben könnten. Anschließend kann evaluiert werden, ob die Anforderungen mit dem Ist-Zustand verfügbarer Stimmungsanalysetools übereinstimmen. Weiterhin können die Ergebnisse dieser Arbeit dafür verwendet werden, die Tools weiter- oder auch neu zu entwickeln und diese an die spezifischen Bedürfnisse bestimmter Stakeholdergruppen anzupassen.

1.2 Lösungsansatz

In dieser Arbeit werden Anforderungen verschiedener Stakeholdergruppen an Stimmungsanalysetools in Softwareprojekten ermittelt und bewertet. Die Arbeit startet mit einer Literaturrecherche zum Thema Stimmungsanalyse, um mögliche Stakeholdergruppen zu bestimmen, die ein Interesse haben könnten, Stimmungsanalysetools im Kontext von Softwareprojekten einzusetzen. Weiterhin werden Möglichkeiten des CrowdRE recherchiert und eine geeignete Methode zur Anforderungserhebung ausgewählt. CrowdRE steht für crowd-based Requirements Engineering. Methoden des CrowdRE sind dafür geeignet, Anforderungen einer großen, heterogenen Gruppe zu erheben, auch wenn Mitglieder dieser Gruppe nicht präsent sind. Ein Beispiel dafür sind beispielsweise die Nutzer sozialer Netzwerke. In diese Arbeit fließen außerdem Interviews aus einer bisher unveröffentlichten Studie ein, die von den Betreuern dieser Arbeit bereitgestellt wurden. Mit dem gewonnenen Wissen aus der Recherche wird anschließend eine iterative Interviewstudie mit verschiedenen Stakeholdern aus der Industrie durchgeführt. Das bedeutet, dass Teilnehmer der Studie mehrmals interviewt werden und die Ergebnisse aus einer Stufe in die nächste Stufe mit einfließen. Die Teilnehmer nennen ihre Anforderungen an ein Stimmungsanalysetool nicht nur, sondern priorisieren sie auch. Mithilfe dieser Studie kann sich ein erster Überblick darüber verschafft werden, welche Bedürfnisse potenzielle Anwender an ein Stimmungsanalysetool stellen und auch aus welchen Gründen sie dies tun. Die Studie zeigt, dass mögliche Stakeholder vielfältige Anforderungen an Stimmungsanalysetools stellen, diese sich je nach Rolle, die sie im Projekt einnehmen, unterscheiden, und dass aktuelle Stimmungsanalysetools diese bis auf wenige Ausnahmen nicht erfüllen.

1.3 Struktur der Arbeit

Diese Arbeit ist wie folgt strukturiert: Dieses Kapitel dient als Einleitung für die Inhalte dieser Arbeit. In Kapitel 2 wird das zum Verständnis dieser Arbeit benötigte Hintergrundwissen vermittelt. Dazu gehören vor allem Erklärungen zur Stimmungsanalyse, aber auch zu Stakeholdern und Methoden des CrowdRE. Weiterhin werden verwandte Arbeiten diskutiert. Kapitel 3 beschäftigt sich mit der Literaturrecherche, die im Rahmen dieser Arbeit durchgeführt wurde. Schließlich werden in Kapitel 4 die Forschungsfragen dieser Arbeit erläutert und es wird erklärt, welche Forschungsmethode verwendet wurde und aus welchen Gründen sich dafür entschieden wurde. Der größte Teil dieser Arbeit besteht aus Kapitel 5, in welchem die Ergebnisse der durchgeführten Studie präsentiert und erläutert werden. Anschließend werden diese in Kapitel 6 diskutiert. Die Arbeit schließt mit Kapitel 7 ab, in dem ein Fazit gezogen wird und ein Ausblick auf mögliche weitere Forschung gegeben wird.

Kapitel 2

Grundlagen und Verwandte Arbeiten

In diesem Kapitel werden Grundlagen zu den Themen Stimmungsanalyse, Stimmungsanalysetools, Stakeholdern und CrowdRE erläutert. Des Weiteren werden am Ende des Kapitels verwandte Arbeiten vorgestellt und diskutiert, wie sich diese Arbeit von ihnen abhebt. Eine genauere Beschreibung, wie die verwendete Literatur recherchiert wurde, ist in Kapitel 3 zu finden.

2.1 Stimmungsanalyse

Die Stimmungsanalyse befasst sich mit der Stimmung von Texten, Sätzen, oder Aussagen. Dabei soll untersucht werden, welche Stimmung dem Empfänger vermittelt wird. Die Analyse beruht auf Techniken des Text-Minings, also Analyseverfahren, die darauf abzielen, Informationen aus Texten zu gewinnen und diese beispielsweise zu kategorisieren. Im Rahmen der Stimmungsanalyse werden Aussagen meistens in folgende drei Kategorien eingeteilt: positiv, neutral oder negativ, wobei die neutrale Stimmung vergeben wird, wenn der Text oder die Aussage weder positiv noch negativ einzuordnen ist. Den Bewertungen positiv und negativ werden meistens noch bestimmte Emotionen zugeordnet. Dies hat den Grund, dass diese von Menschen besser zugeordnet werden können. Positive Emotionen sind beispielsweise Freude oder Enthusiasmus - Beispiele für negative Emotionen dagegen sind Wut oder Angst [36].

Dabei ist das Ziel bei der Verwendung von Stimmungsanalysetools, ein gesamtes Stimmungsbild zu erzeugen. Ursprüngliche Anwendungsgebiete dafür waren Nachrichten oder Kommentare in sozialen Netzwerken oder Kundenrezensionen [32]. Hier können Stimmungsanalysetools eingesetzt werden, um einen Überblick zu erhalten, beim Beispiel Kundenrezensionen also über die Gesamtstimmung, die über ein Produkt herrscht [3]. Dank der Analyse können dann Verbesserungsmaßnahmen ergriffen werden, sodass das

Produkt für die Kunden in Zukunft attraktiver ist. In sozialen Netzwerken hingegen ist es möglich, hasserfüllte oder bösartige Kommentare gezielt zu identifizieren und gegebenenfalls zu entfernen [6]. Ein weiteres typisches Anwendungsgebiet sind Open-Source-Projekte, in denen sich ein Überblick über die Gesamtstimmung verschafft werden soll [9]. Ursprünglich wurde die Stimmungsanalyse also nicht zwangsläufig für den Einsatz in Softwareprojekten entwickelt. Für die Stimmungsbewertung existieren verschiedene Methodiken. Einige Stimmungsanalysetools arbeiten keyword-basiert. Das heißt, dass die Stimmungspolarität durch spezielle Wörter oder Ausdrücke innerhalb eines Satzes oder Textes bestimmt wird. Hier wird durch bestimmte Wörter oder Ausdrücke (Keywords) eine positive oder negative Stimmungspolarität ausgelöst. Ein Beispiel für ein keywordbasiertes Stimmungsanalysetool ist Senti4SD [4]. Neben der keywordbasierten Variante existiert die lexikonbasierte Variante. Die Berechnung eines lexikonbasierten Stimmungsanalysetools erfolgt mithilfe eines Lexikons, welches Begriffe enthält, die klassifiziert wurden, d.h. zu einem Wort ist jeweils hinterlegt, ob die damit vermittelte Stimmung positiv oder negativ zu bewerten ist. Wenn nun ein Satz bewertet wird, wird jedes Wort im Lexikon nachgeschlagen und die dort hinterlegten Stimmungspolaritäten anschließend zusammengefasst, sodass sich für einen Satz eine Gesamtbewertung ergibt. Ein Beispiel für ein lexikonbasiertes Stimmungsanalysetool ist SentiStrength [39]. Nicht alle Stimmungsanalysetools basieren jedoch auf linguistischer Analyse. Einige Tools wurden auf Basis von Machine-Learning-Algorithmen implementiert, so zum Beispiel das Tool SentiCR. Das Tool wird mit einem Datensatz trainiert und klassifiziert dann weitere Datenpunkte mithilfe des Gelernten [1].

2.2 Beispiele für Stimmungsanalysetools

Im Folgenden sollen existierende Stimmungsanalysetools, die im Rahmen der Literaturrecherche gefunden wurden, kurz erläutert werden.

2.2.1 SentiStrength

Bei SentiStrength handelt es sich um ein lexikonbasiertes Stimmungsanalysetool. Jedem Wort wird hierbei durch Lexika ein Wert zwischen (+1) - (+5) (sehr positiv) und (-1) - (-5) (sehr negativ) zugeordnet, wobei eine 0 die neutrale Stimmung repräsentiert. Die Bewertung der Stimmung eines Satzes ergibt sich schließlich aus der Zusammenfassung der Maxima und Minima der einzelner Wörter. Das heißt, dass die negativste Bewertung eines Wortes innerhalb eines Satzes mit der positivsten addiert wird. Beispielsweise ergibt sich bei einem Maximum von 3 und einem Minimum von -4 eine Bewertung von -1 für den gesamten Satz [39].

2.2.2 SentiStrength-SE

SentiStrength wurde nicht domänenspezifisch entwickelt. Infolgedessen kam es bei der Verwendung von SentiStrength im Kontext der Softwareentwicklung zu Ungenauigkeiten und Fehlern kommen. Typische Begriffe, die in der Softwareentwicklung verwendet werden, wie beispielsweise 'eliminate', wurden oft zu Unrecht als negativ bewertet. Daher wurde ergänzend für die Domäne der Software Entwicklung von Islam und Zibran [19] ein angepasstes Lexikon entwickelt. Nach einem Test mit einem Datensatz von Jira Issue Comments und einem Vergleich mit domänen-unabhängigen Tools wurde festgestellt, dass die Performance von SentiStrength hier um einiges besser war als die der domänen-unabhängigen Tools.

2.2.3 SentiCR

SentiCR ist ein Stimmungsanalysetool, das explizit für Code Reviews innerhalb von Softwareprojekten entwickelt wurde, da existierende Stimmungsanalysetools in diesem Bereich schlecht performt haben. Das Tool basiert dabei auf einem Machine-Learning-Algorithmus [1].

2.2.4 EmoTxt

Bei EmoTxt handelt es sich um ein Tool, welches speziell nicht nur Stimmungspolaritäten analysiert, d.h. einen Text mit negativer, positiver oder neutraler Stimmung bewertet, sondern dabei auch explizit Emotionen wie Freude oder Ärger angibt. Dafür sind Wörtern bestimmte Emotionen zugeordnet (z.B. Enthusiasmus, Traurigkeit). Auf Basis der vorkommenden Wörter werden Eingaben Label (Emotionen) zugeordnet [5].

2.2.5 SEntiAnalyzer

Mit dem Stimmungsanalysetool SEntiAnalyzer wurde erstmals ein Versuch gestartet, nicht nur textbasierte, sondern auch sprachbasierte Kommunikation, z.B. in Meetings, auszuwerten. Dafür wurden Methoden der Stimmungsanalyse mit Spracherkennungstools kombiniert und es existieren verschiedene Bearbeitungsstufen. Der Bearbeitungsprozess ist in Abbildung 2.1 veranschaulicht. Es existieren Input Layer, Pre Processing Layer, Classification Layer und Output Layer. In der Classification Layer werden verschiedene Methoden zur Stimmungsanalyse kombiniert, unter anderem BERT, SentiStrength-SE und SEnti4SD. Nach der Spracheingabe erfolgt im Laufe des Meetings Feedback auf Basis mehrerer Stimmungsanalysetools. Auch ist es möglich, explizit nach jedem gesprochenen Satz direktes Feedback zu erhalten [18].

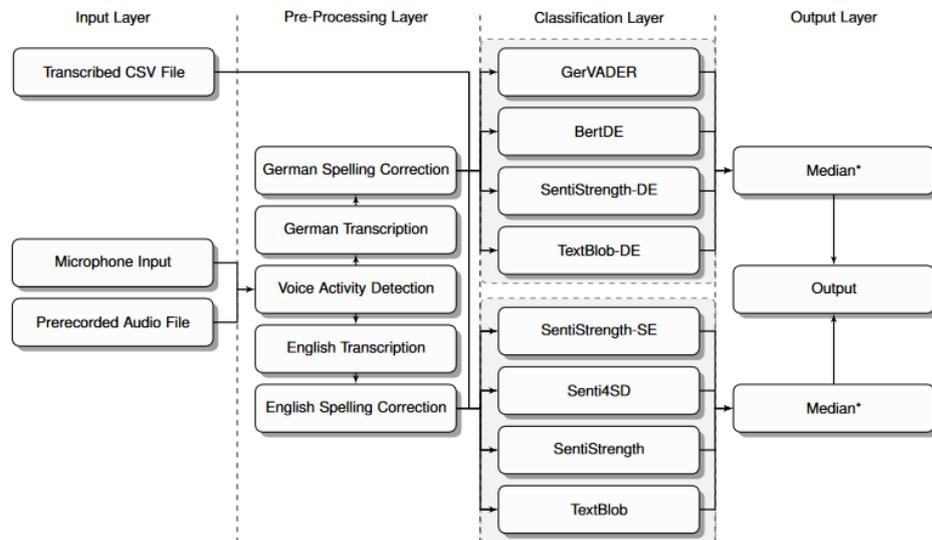


Abbildung 2.1: Bearbeitungspipeline des Senti-Analyzers[18]

2.2.6 BERT

Beim Stimmungsanalysetool BERT (Bidirectional Encoder Representations from Transformers) handelt es sich um ein vortrainiertes Modell zur Verarbeitung bzw. Klassifizierung natürlicher Sprache [8]. Das heißt, das Tool wurde mithilfe eines Trainings-Datensets trainiert. Wu et al. testeten die Performance von BERT im Kontext von Softwareentwicklung an vier verschiedenen Datensätzen und verglichen diese mit der Performance sechs anderer Stimmungsanalysetools, unter anderem mit SentiStrength und SentiCR. Bei den Datensätzen, die für dieses Experiment verwendet wurden, handelte es sich um Jira Issue Comments, App Reviews, Code Reviews und ein Datensatz aus Stack Overflow-Sätzen. BERT erzielte in drei Datensätzen, mit Ausnahme vom App Review-Datensatz, das beste Ergebnis [43]. Besser war lediglich das Tool Stanford CoreNLP [24].

2.3 Stakeholder in Softwareprojekten

Stakeholder im Kontext von Softwareprojekten sind alle Personen, die in dem Projekt involviert sind oder die Interesse an der entwickelten Software haben [31]. Im Folgenden werden verschiedene Methoden zur Stakeholder-Identifikation vorgestellt.

2.3.1 Methoden zur Stakeholder-Identifikation

In Pacheco et al. werden verschiedene Methoden zur Identifikation von Stakeholdern in Softwareprojekten beschrieben [31]. Als Einschränkung wird dabei genannt, dass die meisten Methoden eher eine Liste als ein Rahmenwerk zur Verfügung stellen. Das heißt, es werden Listen von Stakeholdern vorgeschlagen, die durchgearbeitet werden können, aber es werden keine Methoden vorgeschlagen, die darüber hinaus gehen. Die verschiedenen genannten Methoden bzw. Listen aus der Arbeit sollen im Folgenden durchgegangen werden, um möglichst viele Stakeholder von Softwareprojekten zu identifizieren.

SEI Technical Report

Laut Christel et al. [7] werden bei der Identifikation von Stakeholdern in Softwareprojekten Stakeholder aus mindestens sechs Kategorien berücksichtigt. Diese sind:

- Kunden
- Benutzer
- Entwickler
- Qualitätssicherung
- Sicherheitspersonal
- Requirements Engineers

Capability Maternity Model Integration

Dieses Model schlägt explizit bestimmte Stakeholdergruppen zur Berücksichtigung in Softwareprojekten vor [38]. Darunter fallen:

- Kunden
- Endnutzer
- Entwickler
- Produzenten
- Tester
- Lieferanten
- Marketing
- Wartung

- und jeder, der in irgendeiner Art und Weise von Softwareprojekten beeinflusst wird oder wiederum selbst Einfluss nimmt.

ISO Standards: Software Life Cycle Processes

Hier wird eine Guideline zur Identifikation von Stakeholdern und deren Verantwortung und Rollen beschrieben. Dabei wird sich am typischen Software Life Cycle orientiert. Gruppen, die genannt werden, sind unter anderem:

- Kunden
- Qualitätssicherung
- Entwickler
- Auslieferung

Sommerville

Sommerville [37] ordnet die Identifizierung der Stakeholder in die Phase der Analyse von Softwareanforderungen ein. Zu den identifizierten Stakeholdern gehören:

- Endnutzer und jeder, der in der Organisation vom System beeinflusst wird.
- Entwickler
- Entwickler von Systemen, zu denen eine Schnittstelle besteht.
- Business Managers
- Arbeitnehmervertreter
- Domänenexperten

Pressman

Roger Pressman [33] argumentiert, dass die Stakeholdergruppen zu Beginn des Requirements-Engineering-Prozesses identifiziert werden müssen, da viele verschiedene Teilnehmer in dieser Phase beteiligt sind. Pressman nennt folgende Stakeholder:

- Geschäftsleiter
- Markenmanager
- Marketingmitarbeiter

- externe und interne Kunden
- Berater
- Produktingenieure
- Softwareentwickler
- Support- und Wartung
- Ingenieure

Rational Unified Process

Stakeholdergruppen, die laut RPU [21] zu einem Softwareprojekt gehören, sind folgende:

- Endnutzer
- Entwickler
- Käufer/Kunde
- Projektleiter
- jeder, der am Softwareprojekt interessiert ist oder jeder, der das Ergebnis des Projekts für seine Bedürfnisse benötigt.

2.4 CrowdRE

Requirements Engineering ist ein essentieller Faktor, der die Qualität von Software und Systemen beeinflusst [16]. Techniken des Requirements Engineerings setzen oft voraus, dass der Stakeholder präsent bzw. erreichbar ist. Das Aufkommen von Apps für Mobilgeräte, beispielsweise im Rahmen von sozialen Netzwerken oder auch Cloud-Apps, beinhaltet die Herausforderung, dass Stakeholder nun aus einer extrem großen und diversen Gruppe, der sogenannten 'Crowd' stammen und auch verschiedene Bedürfnisse haben. Eine Co-Präsenz ist oft nicht möglich, bzw. es gestaltet sich schwierig, die Heterogenität der Crowd damit abzubilden. Mithilfe des CrowdRE-Ansatzes werden bekannte Ansätze zur Anforderungsanalyse mit neuen Konzepten ergänzt, um Lücken, die durch die Problematik der Crowd entstehen, zu schließen [16]. Im Folgenden sollen einige Methoden des CrowdRE kurz beschrieben werden, damit sich ein Eindruck über mögliche Methoden gemacht werden kann.

2.4.1 Multi-Level-Approach

Einige Konzepte, die von Groen et al. speziell für das CrowdRE vorgestellt werden, sind beispielweise Text Mining, Usage Mining (Tracken von Benutzerverhalten) und Event Logging (Identifikation von Verhaltensmustern). In diesem Rahmen wird ein Multi-Level-Ansatz vorgeschlagen, das heißt, verschiedene Verfahren werden miteinander kombiniert [16]. Wie Abbildung 2.2 zeigt, teilt sich der Prozess in zwei Feedback-Stränge, in qualitatives und quantitatives Feedback. Das quantitative Feedback wird durch Event Logging und Usage Mining geschaffen, d.h. Benutzerverhalten wird getrackt und beobachtet und Verhaltensmuster identifiziert. Das qualitative Feedback wird geschaffen durch Motivational Instrument und Text Mining, d.h. Benutzer werden zum Feedback geben motiviert und es werden Probleme, Bedürfnisse und Ideen entdeckt.

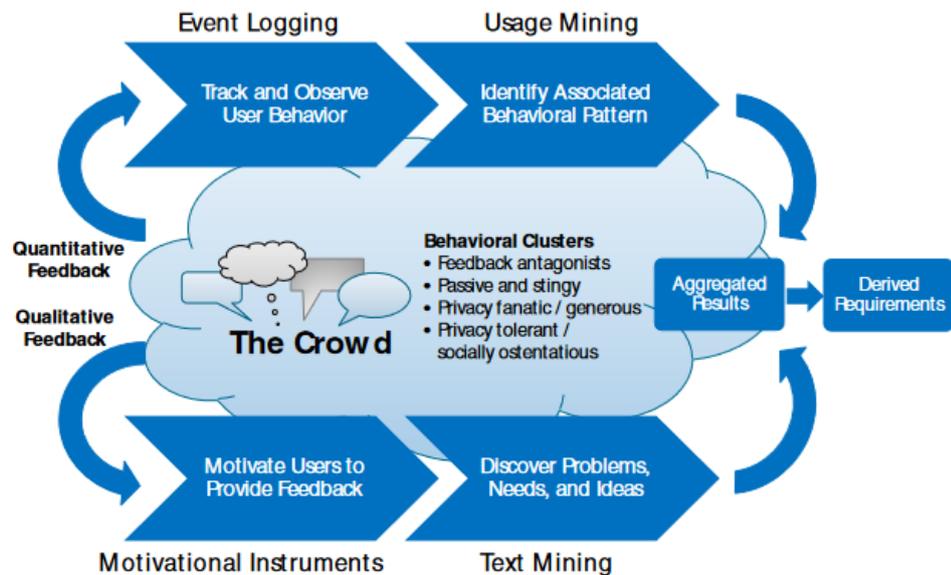


Abbildung 2.2: CrowdRE: Multi-Level-Approach [16]

2.4.2 StakeNet

StakeNet ist eine Methode, die dazu dient, Stakeholder, vor allem im Rahmen von großen Softwareprojekten, zu identifizieren und auch zu priorisieren, da davon ausgegangen werden muss, dass nicht alle Stakeholder gleich zu gewichten sind. Zunächst sollen Stakeholder identifiziert werden und anschließend darum gebeten werden, weitere mögliche Stakeholdergruppen vorzuschlagen. Somit entsteht ein Netzwerk von Stakeholdern, wobei die

identifizierten Stakeholder die Knoten des Netzwerks darstellen und die Empfehlungen die Links [23].

2.4.3 StakeRare

StakeRare ist eine Methode, die ähnlich zu StakeNet ist. Auch hier wird das Stakeholder-Netzwerk aufgebaut und identifizierte Stakeholder werden gebeten, weitere Stakeholder zu empfehlen. Stakeholder werden anschließend gebeten, eine Liste von initialen Anforderungen zu bewerten. Mithilfe von Collaborative Filtering (Auswertung von Verhaltensmustern) werden ihnen weitere Anforderungen vorgeschlagen. Anschließend werden die Bewertungen der verschiedenen Stakeholder noch einmal gewichtet, dabei wird der Projekteinfluss der Stakeholder berücksichtigt, d.h. die Anforderungen und Bewertungen der Stakeholdergruppe(n), die über den größten Einfluss auf das Softwareprojekt verfügen, sollen am stärksten gewichtet werden. Abbildung 2.3 veranschaulicht den Prozess im Detail.

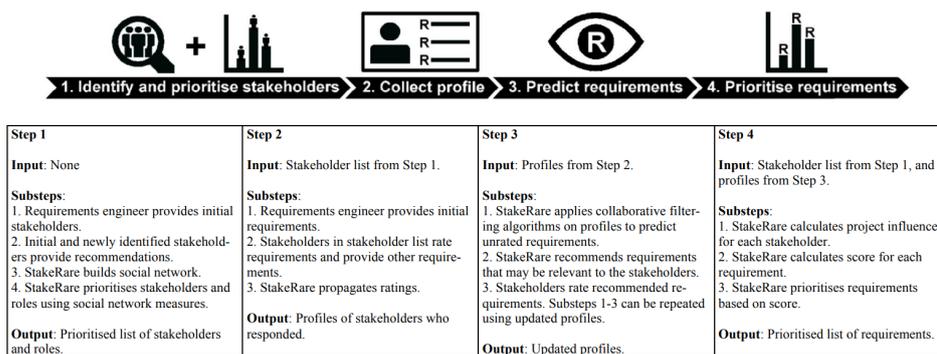


Abbildung 2.3: StakeRare-Prozess [22]

2.4.4 Iterativer Ansatz

Der iterative Ansatz stammt von Murukannaiah et al. [26]. Der Hintergrund ist ein Szenario, in dem zum gewünschten Thema/Projekt etc. noch keine oder nur sehr wenige Anforderungen vorhanden sind. Es werden Stakeholder befragt, wobei zunächst initiale Anforderungen gesammelt werden. In weiteren Stufen werden den Befragten die Anforderungen, die andere Teilnehmer genannt haben, präsentiert und sie sollen diese ergänzen. Dies soll zu einer kognitiven und kreativen Stimulation und somit schlussendlich zu einer erfolgreichen Anforderungserhebung führen. Im zugehörigen Paper besteht der Ansatz aus zwei Phasen. Der Prozess ist in Abbildung 2.4 abgebildet.

Phase 1 besteht aus insgesamt drei Befragungen bzw. Stufen, wobei eine die initiale Stufe ist und die gesammelten Anforderungen einer Stufe

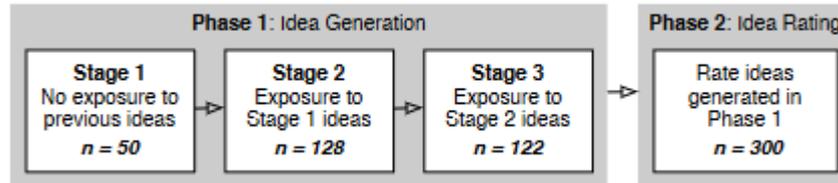


Abbildung 2.4: Iterative Interviewstudie [26]

jeweils in die nächste übernommen werden. Die zweite Phase ist die Anforderungsbewertung. Die in der vorangegangenen Phase gesammelten Anforderungen sollen nun von den Teilnehmern bewertet bzw. priorisiert werden. In einer weiteren Veröffentlichung wird von Murukannaiah et al. beschrieben, wie dieser Prozess automatisiert werden kann, was in Hinblick auf großen Datenmengen, die mit dieser Methode entstehen, wenn sie auf eine große Zielgruppe angewandt wird, in Hinblick auf Zeit- und Kostenprobleme erforderlich ist [27].

2.5 Verwandte Arbeiten

In diesem Unterkapitel werden verwandte Arbeiten beschrieben und diese Arbeit von ihnen abgegrenzt. Die Basis für die in dieser Arbeit durchgeführte Literaturrecherche bildet das Paper von Obaidi und Klünder [29]. Obaidi und Klünder führen eine systematische Literaturrecherche zum Thema Stimmungsanalyse durch. Dabei soll evaluiert werden, welche Anwendungsszenarien sich für die Stimmungsanalyse ergeben, welchem Zweck die Verwendung von Stimmungsanalysetools in den untersuchten Studien dient, auf welchen Datensets verschiedene Stimmungsanalysetools basieren, welche unterschiedlichen Ansätze es für die Stimmungsanalyse gibt und welche Herausforderungen sich für diese Ansätze ergeben. Aufbauend auf diesem Paper führen Obaidi et al. [30] eine Systematic Mapping Study zum Thema Stimmungsanalyse durch. Ziel war es, typische Anwendungsszenarien von Stimmungsanalyse zu identifizieren. Dazu gehört auch die Identifikation von Anwendungsdomänen. Es konnten drei verschiedene Domänen ermittelt werden: Open-Source-Software, Forschung, und Industrie. So wurde ermittelt, dass mehr als zwei Drittel der untersuchten Arbeiten sich auf die Open-Source-Software-Domäne beziehen. Hermann und Klünder [17] testeten mit dem SEntiAnalyzer die Bewertung von Stimmung in Live-Meetings. Weiterhin existieren einige Studien, die die Performance diverser Stimmungsanalysetools miteinander oder auf verschiedenen Datensets vergleichen. Im Folgenden sind einige davon aufgeführt. Islam et al. [20] testeten die Performance drei verschiedener Stimmungsanalysetools in Bezug auf drei Datensets im Kontext von Softwareentwicklung. Damit sollte geprüft werden,

wie präzise diese Tools im domainenspezifischen Kontext der Softwareentwicklung sind. Zhang et al. [44] untersuchten die Frage, wie präzise und effizient vortrainierte Modelle wie etwa BERT im Vergleich zu anderen Tools wie SentiCR, SentiStrength, etc. sind. Novielli et al. [28] führten ebenfalls eine Studie durch, um drei verschiedene Stimmungsanalysetools, Senti4SD, SentiStrength-SE und SentiCR, miteinander zu vergleichen. Weiterhin werden einige Herausforderungen diskutiert, die sich aus der Analyse und eventuellen Falschbewertung von Texten im Rahmen der Stimmungsanalyse ergeben. Einige Veröffentlichungen befassen sich mit der Auswirkung der Stimmung von Projektteilnehmern, hauptsächlich von Entwicklern, auf die Produktivität bzw. die Arbeitsleistung. Graziotin et al. [13] untersuchten den Einfluss von affektiven Zuständen bzw. Stimmungen auf die Fähigkeit von Softwareentwicklern, kreativ und analytisch zu denken. Dazu wurde eine Studie mit 42 Teilnehmern durchgeführt. In einer weiteren Veröffentlichung von Graziotin et al. [14] wird beschrieben, dass die Stimmung von Softwareentwicklern vielen Schwankungen unterliegt und sich in sehr kurzer Zeit dramatisch ändern kann. Für die Produktivität innerhalb von Softwareprojekten sei eine positive Grundstimmung förderlich. Weiterhin beschreiben Graziotin et al. [11] dass psychologische Zustände wie etwa Burnout, Angst oder Sorge durch die gezielte Elimination negativer Stimmung oder Vorkommnisse in Softwareprojekten verhindern werden können. Murukannaiah et al. [26] erheben Anforderungen an Smart-Home-Devices. In dieser Arbeit wurde sich an der von Murukannaiah et al. vorgestellten Methode, einer iterativen Interviewstudie orientiert, es existieren jedoch einige Unterschiede. Murukannaiah et al. arbeiteten nicht mit Interviews, sondern mit Fragebögen. Der Fokus dieser Arbeit liegt daher auf qualitativen, und nicht auf quantitativen Daten.

Kapitel 3

Literaturrecherche

In diesem Kapitel wird beschrieben, wie die in dieser Arbeit verwendete Literatur recherchiert und analysiert wurde. Als primäre Methode wurde sich für den Snowballing-Prozess nach Wohlin [41] entschieden. Das Paper von Obaidi und Klünder [29] wurde hierfür als Startpunkt gewählt, da es sich um eine systematische Literaturrecherche handelt und somit bereits einen guten Ausgangspunkt darstellt. Nach der Durchführung des Snowballing-Prozesses wurde die Recherche anschließend um eine gezielte Suche mit Suchstrings mittels fünf gängiger wissenschaftlicher Suchmaschinen (Science Direkt, IEEE Xplore, ACM Digital Library, Springer Link, Google Scholar) ergänzt, um die gefundene Literatur mit Literatur zu den Themen CrowdRE und Stakeholdern in Softwareprojekten zu erweitern. Für die erste Stufe der in dieser Arbeit durchgeführten Studie wurden außerdem Daten einer bisher unveröffentlichten Interviewstudie zum Thema Stimmungsanalyse verwendet, aus denen Anforderungen an ein Stimmungsanalysetool hervorgingen. Dieser Prozess ist in Abbildung 3.1 veranschaulicht.

3.1 Auswahlkriterien für Literatur

Um in die Literaturliste aufgenommen zu werden, mussten Arbeiten alle der folgenden Inklusionskriterien und keines der folgenden Exklusionskriterien erfüllen:

Inklusionskriterien:

- (1) Die Arbeit muss sich mit einem für diese Arbeit relevanten Thema befassen (Sentiment Analysis, Requirements, Stakeholder...)
- (2) Die Arbeit muss kostenfrei zugänglich sein.
- (3) Die Arbeit muss peer-reviewed sein.

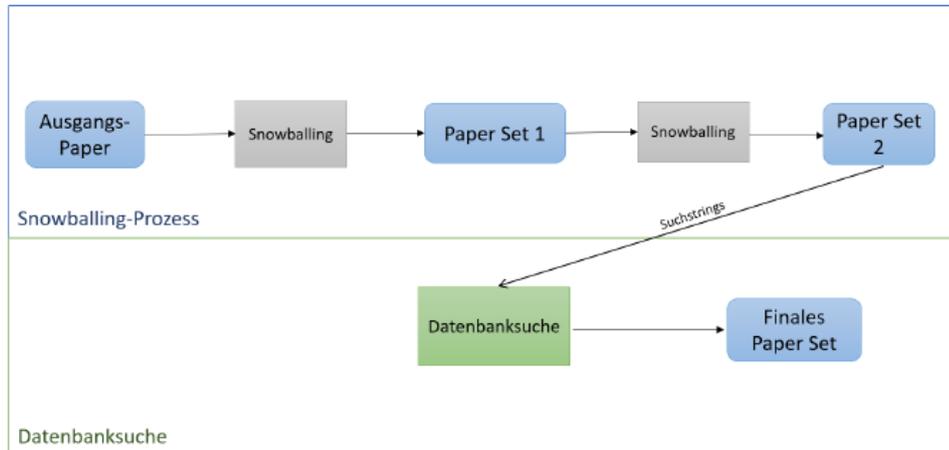


Abbildung 3.1: Prozess der Literaturrecherche

Die oben genannten Inklusionskriterien wurden aus folgenden Gründen gewählt: die Arbeit soll kostenfrei zugänglich sein, damit die Ergebnisse beziehungsweise die Literaturrecherche von jedem reproduziert und nachvollzogen werden können. Die Eigenschaft, dass die Arbeiten peer-reviewed sind, ist nötig, um die Qualität der Arbeiten sicherzustellen. Dafür ist es notwendig, dass die Arbeit vor Veröffentlichung von anderen Wissenschaftlern aus dem gleichen Fachgebiet einer Qualitätskontrolle unterzogen wurde. Die Arbeit soll sich schlussendlich auch mit einem für diese Arbeit relevantem Thema befassen, damit der Inhalt für diese Arbeit von Bedeutung ist.

Exklusionskriterien:

- (1) Die Arbeit ist nicht in Englisch oder Deutsch geschrieben.
- (2) Es existieren mehrere Versionen (es wird nur die neueste berücksichtigt).

Die Exklusionskriterien wurden aus folgenden Gründen gewählt: es wird jeweils nur die neueste Version einer Arbeit verwendet, um den aktuellen Stand der Forschung abzubilden. Weiterhin soll die Arbeit in Deutsch oder Englisch verfasst sein, da sonst sprachliche Barrieren existieren würden und die Arbeit nicht verstanden werden könnte.

3.2 Startpunkt

Bei dem Paper, das als Ausgangspunkt für die Literaturrecherche diente, handelt es sich um eine systematische Literaturrecherche von Obaidi und

Klünder [29] zum Thema Stimmungsanalysetools. Es wurden 80 Arbeiten untersucht und die Ergebnisse zusammengefasst nach Anwendungsbe-
reich, Zweck, verwendeten Datensätzen, Ansätze zur Stimmungsanalyse und
Schwierigkeiten, mit denen Forscher bei der Anwendung von Stimmungsana-
lyse konfrontiert werden. Es eignete sich daher gut als Ausgangspunkt für
die Literaturrecherche.

3.3 Snowballing

Der Snowballing-Prozess, der in dieser Arbeit verwendet wird, wird abge-
leitet von Wohlin, der ihn formal beschreibt [41]. Ausgehend von einem
Ausgangset von Arbeiten, werden sowohl die zitierten Arbeiten als auch
die Arbeiten, die diese Ausgangsarbeiten bereits zitieren, ausfindig gemacht.
Das Identifizieren von Arbeiten, die die Ausgangsarbeit zitiert, wird als
Backward Snowballing bezeichnet. Forward Snowballing bezeichnet dann
das Auffinden von Arbeiten, die die Ausgangsarbeit bereits zitieren. Aus
dem Backward und Forward Snowballing ergibt sich ein weiteres Set
von Arbeiten. Arbeiten werden in das Set aufgenommen, nachdem sie
auf Relevanz untersucht wurden. Dieser gesamte Prozess wird wiederholt,
bis keine weiteren Arbeiten gefunden werden. Wolfswinkel et al. [42]
beschreibt allerdings, dass der Prozess bereits vorzeitig beendet werden
kann, falls die gefundenen Arbeiten sich wiederholen und sich eine Sättigung
einstellt. Nachdem das Paper von Obaidi und Klünder [29] komplett gelesen
wurde, wurden alle Arbeiten identifiziert, die im Paper zitiert werden.
Außerdem wurden mithilfe der Suchmaschine Google Scholar alle Arbeiten
ermittelt, die dieses Paper zitieren. Von den gefundenen Arbeiten wurde
sich jeweils Titel und Abstract angeschaut und unter der Berücksichtigung
der Inklusions- und Exklusionskriterien ermittelt, ob die Arbeit in die
Literaturliste aufgenommen werden kann. Es wurden somit, zum Zeitpunkt
der Recherche, 52 Arbeiten in diesem ersten Schritt in die Literaturliste
aufgenommen, wobei 44 Titel aus den Referenzen stammen und acht Titel
das Startpaper zitieren. Bei der erneuten Durchführung des Snowballing-
Prozesses mit diesen Titeln als neues Ausgangsset zeigte sich bereits eine
Erschöpfung und die Titel wiederholten sich. Daher wurde, wie oben
beschrieben, nach Wolfswinkel et al. [42] der Snowballing-Prozess beendet.
Durch den Snowballing- Prozess wurden insgesamt 175 Titel identifiziert, die
für diese Arbeit relevant sein könnten.

3.4 Ergänzende Datenbanksuche

Da die Arbeiten, die im Snowballing-Prozess ermittelt wurden, sich thema-
tisch auf die Stimmungsanalyse beziehen, es aber bisher keine Verbindung
zu verschiedenen Stakeholdern oder CrowdRE-Techniken gibt, wurde eine

ergänzende Datenbanksuche durchgeführt, um Arbeiten zu diesen Themen zu identifizieren. Verwendet wurden die Datenbanken von Science Direkt, IEEE Xplore, ACM Digital Library, Springer Link, Google Scholar.

3.4.1 Erste Suche

Ziel der ersten Datenbanksuche war es, mögliche Stakeholder im Rahmen von Softwareprojekten zu identifizieren. Der erste Suchstring wurde wie folgt definiert:

stakeholder **AND** (identification **OR** classification) **AND** (software engineering **OR** software development **OR** software project)

Der erste Term inkludiert den Begriff des Stakeholders in die Suche, der zweite Term bezieht die Identifikation bzw. Klassifikation dieser in die Suche mit ein. Der dritte Term ist nötig, um den Kontext, in diesem Fall die Softwareentwicklung bzw. ein Softwareprojekt, zu definieren. Die Ergebnisse der Datenbanksuche wurden auf Relevanz überprüft. Es haben sich weitere neun Titel ergeben, die der Literaturliste hinzugefügt wurden.

3.4.2 Zweite Suche

Ziel der zweiten Datenbanksuche war das Ermitteln von Titeln, die sich mit möglichen CrowdRE-Techniken beschäftigen, die für diese Arbeit verwendet werden könnten. Hier wurden die gleichen Datenbanken zur Suche verwendet wie oben beschrieben. Der zweite Suchstring wurde wie folgt definiert:

(crowdRE **OR** crowd requirements engineering **OR** crowd) **AND** (software engineering **OR** software development **OR** software project)

Der erste Term wurde verwendet, um Ergebnisse zu erhalten, die CrowdRE-Techniken beschreiben und der zweite Term wurde eingesetzt, um auch hier den Kontext im Bereich der Softwareentwicklung zu definieren. Nach der Relevanzprüfung wurden fünf weitere Arbeiten in die Literaturliste aufgenommen.

3.5 Daten aus Studie

Im Vorfeld dieser Arbeit wurde bereits eine noch unveröffentlichte Studie durchgeführt, in der mögliche Anforderungen von Beschäftigten in der Industrie an ein Stimmungsanalystool gesammelt wurden. Die Teilnehmenden bekamen die Möglichkeit, ein Tool zu testen und wurden anschließend in Interviews nach ihrer Meinung und weiteren bisher nicht implementierten Anforderungen, die sie an das Tool haben, gefragt. Die Aufnahmen und

Transkripte dieser Interviews wurden von den Betreuern dieser Arbeit zur Verfügung gestellt. Diese wurden durchgesehen und als Ausgangspunkt für die iterative Interviewstudie dieser Arbeit, wie unter Kapitel 4 beschrieben, verwendet.

Kapitel 4

Forschungsmethode

4.1 Forschungsfragen

4.1.1 RQ1: Identifikation von Stakeholdern

Wie in Kapitel 2 erläutert, sind in der Literatur bestimmte Anwendungsfälle von Stimmungsanalysetools beschrieben, aber es wurde noch nicht zusammenfassend untersucht, welche Stakeholdergruppen ein Interesse am Einsatz von Stimmungsanalysetools in Softwareprojekten haben könnten. Daher sollen diverse in der gefundenen Literatur analysierten Anwendungsfälle in dieser Arbeit zusammengeführt werden und auch erläutert werden, wieso ein Stakeholder in einer bestimmten Rolle, die er in Softwareprojekten innehält, im Rahmen seiner Arbeit Stimmungsanalysetools verwenden sollte. Daher ist die erste Forschungsfrage wie folgt:

RQ1: Welche Stakeholder haben ein Interesse daran, Stimmungsanalysetools in Softwareprojekten zu verwenden?
--

4.1.2 RQ2: Anforderungsanalyse und Priorisierung

Um herauszufinden, welche Anforderungen die ermittelten Stakeholder an ein Stimmungsanalysetool stellen, muss eine Anforderungsanalyse stattfinden. Es soll ermittelt werden, welche Anmutung, Eigenschaften und Funktionen ein Tool haben muss, um den Bedürfnissen der ermittelten Stakeholder gerecht zu werden, sodass sie sich eine regelmäßige Nutzung vorstellen können. Die zweite Forschungsfrage lautet daher:

RQ2: Welche Anforderungen stellen verschiedene Stakeholder an ein Stimmungsanalysetool, damit es für sie in ihren Arbeitsalltag integrierbar und nützlich ist?

4.1.3 RQ3: Priorisierung der Anforderungen und Unterschiede

Der nächste Schritt besteht darin, die ermittelten Anforderungen zu priorisieren. Das ist nötig, um herauszufinden, welche Anforderungen essentiell sind und welche Anforderungen die Stakeholder zwar als nützlich, jedoch nicht als notwendig erachten, um das Tool zu nutzen. Wichtig ist das auch, um zu ermitteln, welche Anforderungen bzw. Funktionen als erstes von einem Tool implementiert werden sollten, bzw. um welche Anforderungen ein bestehendes Tool vorrangig erweitert werden sollte. Zu erwarten ist, dass Stakeholder aufgrund ihrer Rolle, die sie im Softwareprojekt besetzen, die ermittelten Anforderungen unterschiedlich priorisieren und dabei die für ihre Rolle relevanten Aspekte und Arbeitsschritte berücksichtigen. Um Stimmungsanalysetools z.B. explizit auf die Rolle Projektleiter anzupassen, ist es notwendig, zu analysieren, welche Anforderungen von Befragten dieser Rolle als essentiell erachtet wurden. Die dritte Forschungsfrage lautet also:

RQ3: Wie werden die ermittelten Anforderungen von den Stakeholdern priorisiert und welche Unterschiede ergeben sich zwischen den verschiedenen Stakeholdergruppen?

4.1.4 RQ4: Vergleich mit aktuellen Tools

Um abschließend zu ermitteln, ob aktuelle Stimmungsanalysetools die erhobenen Anforderungen nicht, teilweise oder sogar ganz erfüllen, soll ein Vergleich mit verfügbaren Tools angestellt werden. So kann auch bestimmt werden, ob ein möglicher Grund für die bisher ausbleibende Verwendung von Stimmungsanalysetools ist, dass wichtige Anforderungen von potenziellen Nutzern bisher nicht implementiert sind. Die vierte Forschungsfrage lautet:

RQ4: Inwiefern erfüllen aktuell verfügbare Stimmungsanalysetools die erhobenen Anforderungen?

4.1.5 Abdeckung

RQ1 wird durch die Literaturrecherche beantwortet. RQ4 sowohl von der Literaturrecherche als auch von der geplanten Studie. Die anderen Forschungsfragen werden durch die Studie beantwortet.

4.2 Auswahl einer geeigneten Forschungsmethode

4.2.1 Gewünschte Eigenschaften

Im Folgenden sollen einige Eigenschaften definiert werden, die die Forschungsmethode haben sollte, damit die vorher erläuterten Forschungsfragen

beantwortet werden können.

- Sie sollte dafür geeignet sein, Anforderungen zu erheben, für den konkreten Fall, dass bisher keine oder nur sehr wenig Anforderungen zum untersuchten Thema erhoben wurden.
- Sie sollte eine Technik des CrowdRE sein, um sicherzustellen, dass anforderungsrelevante Informationen von großen Benutzergruppen erhoben werden können.
- Sie sollte entweder eine Interviewstudie, eine Umfragestudie oder einen Workshop beinhalten, damit Anforderungen der Teilnehmenden erhoben werden können.
- Sie sollte im Rahmen dieser Arbeit durchführbar sein.

4.3 Vorstellung der gewählten Methodik

Auf Basis der Literaturrecherche und unter Berücksichtigung der oben genannten erforderlichen Eigenschaften wurde sich für eine Methodik entschieden, die im Folgenden erläutert werden soll.

4.3.1 Iterative Interviewstudie

Die in dieser Arbeit verwendete Methode wird von Murukannaiah et al. [26] beschrieben. Es handelt sich um eine iterative Studie, die aus zwei Phasen besteht, wovon die erste Phase noch einmal in drei Stufen gegliedert ist. Das Design ist besonders geeignet für die Anforderungserhebung, falls im Vorfeld noch keine Anforderungen bekannt sind. Dieser Vorgang ist in Abbildung 4.1 veranschaulicht.

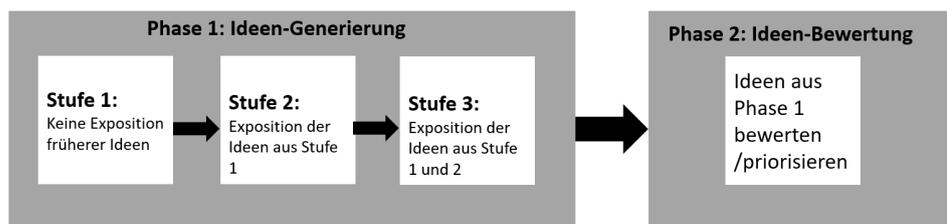


Abbildung 4.1: Iterative Interviewstudie, angepasst aus [26]

4.3.2 Ideen-Generierung

Wie in Abbildung 4.1 zu sehen, besteht die Phase der Ideen-Generierung aus drei Stufen und somit aus drei Befragungen. In der ersten Stufe wurden

im Vorfeld noch keine Anforderungen gesammelt. Teilnehmer werden nun gebeten, ihre Anforderungen zu nennen. Die Anforderungen aus dieser Stufe bilden das initiale Anforderungsset. In der zweiten Stufe werden den Teilnehmern die Anforderungen aus der ersten Stufe präsentiert. Das beinhaltet sowohl ihre eigenen Anforderungen als auch die der anderen Teilnehmer. Dies hat das Ziel, durch die Präsentation der fremden Anforderungen eine kognitive Stimulation bei den Teilnehmern zu erreichen, was zu kreativerem Denken und weiteren Anforderungen führt. Das Kreative Denken hat hier einen hohen Stellenwert, da Anforderungen für ein Tool gesammelt werden sollen, für das noch keine oder nur sehr wenig Anforderungen bekannt sind. Von Vorteil ist es demnach, wenn Teilnehmer viele unterschiedliche Aspekte berücksichtigen und bei der Erhebung ihrer möglichen Anforderungen in verschiedene Richtungen denken. Aus dem zweiten Schritt ergibt sich das zweite Anforderungsset, was aus dem ersten Set plus den neu gewonnenen Anforderungen besteht. In der dritten Stufe wird dieser Prozess noch einmal wiederholt.

Umsetzung in dieser Arbeit

Im oben genannten Paper, aus dem diese Methodik stammt, wurden Umfragen verwendet. Aus mehreren Gründen wurden in dieser Arbeit jedoch Interviews verwendet. Die Gründe dafür sind unter Punkt 4.4 erläutert. Im Rahmen dieser Phase wurden in dieser Arbeit lediglich zwei statt drei Interviews geführt. Grund ist, dass die Daten der unter Kapitel 3 erwähnten bisher unveröffentlichten Studie als initiales Anforderungsset dienen. Das hat den Vorteil, dass die Teilnehmer der Studie ein Stimmungsanalysetool ausprobieren konnten. Gerade im Hinblick auf das Thema Stimmungsanalysetool, das in der Industrie bisher eher unbekannt ist, ist es denkbar, dass Teilnehmer überfordert wären, wenn sie ohne Hintergrundwissen Anforderungen zu einem unbekanntem Thema nennen sollten. Als Vorbereitung für die Interviewstudie wurden also die Daten aus der unveröffentlichten Studie analysiert und die daraus resultierenden Anforderungen in drei Kategorien eingeteilt:

- **Erkennen von Stimmungen:** Bezieht sich auf Funktionen des Tools bezüglich der korrekten Erkennung von Stimmungen. Welche Aspekte muss das Tool hierfür implementieren?
- **Oberfläche:** Wie soll die grafische Benutzeroberfläche des Stimmungsanalysetools aussehen? Was muss sie für Eigenschaften haben?
- **Verständnis:** Beschreibt, sowohl welche Funktionen nötig sind, damit der Anwender die Ergebnisse bzw. die Stimmungsbewertungen, die das Tool ausgibt, versteht, als auch, welche Funktionen implementiert werden müssen, damit der Anwender versteht, wie er das Tool verwenden kann.

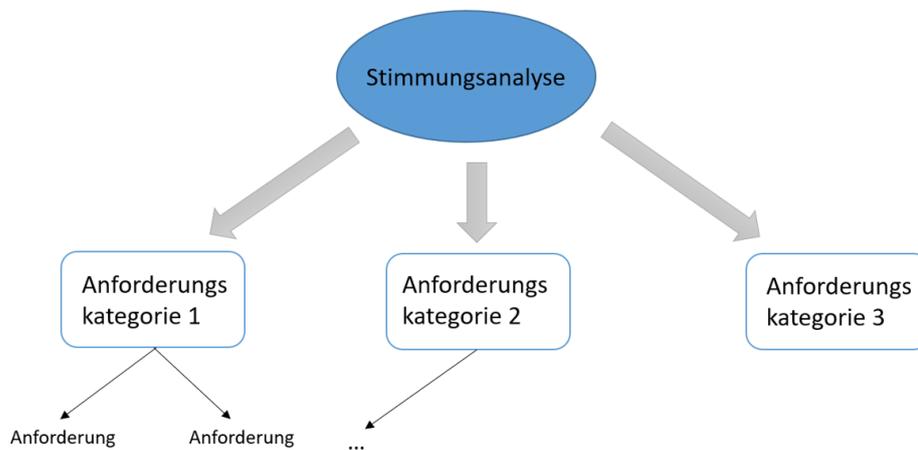


Abbildung 4.2: Beispiel für verwendete Baumstruktur

Die komplette Liste der Anforderungen inklusive Erklärungen sind im Anhang zu finden. Um die Anforderungen zu visualisieren und den Teilnehmern ansprechend zu präsentieren, wurden Baumstrukturen verwendet. Beispielhaft ist dies in Abbildung 4.2 dargestellt. Aus Gründen der Übersichtlichkeit und des Vermeidens von Informationsüberflutung wurde den Teilnehmern in dieser Arbeit zusätzlich die Möglichkeit gegeben, Anforderungen zu löschen, falls sie nicht wünschen, dass diese implementiert werden. So kann auch verglichen werden, welche Anforderungen für bestimmte Stakeholder nicht relevant sind.

4.3.3 Ideen-Bewertung

In der Phase der Ideen-Bewertung findet eine weitere Befragung der Teilnehmer statt. Hier werden Teilnehmer gebeten, mittels einer Skala die ermittelten Anforderungen zu priorisieren.

Skala zur Priorisierung

In dieser Arbeit wurde für die Bewertung der Anforderungen folgende Likert-Skala verwendet:

- 1 – keine Priorität
- 2 – geringe Priorität

- 3 – etwas Priorität
- 4 – neutral
- 5 – moderate Priorität
- 6 – hohe Priorität
- 7 – höchste Priorität

4.3.4 In Vivo und Pattern Coding

Teilnehmer wurden gebeten, falls sie eine Anforderung mit einer 1 oder einer 7, also der niedrigsten oder höchsten Priorität bewerten, eine kurze Begründung dieser Bewertung abzugeben. Anschließend wurden zwei Coding-Zyklen durchgeführt. Im ersten Zykel wurde in Vivo-Coding verwendet, da dieses sich gut für qualitative Studien eignet und mit vielen anderen Coding-Methoden im nächsten Zykel verwendet werden kann. Anschließend wurde Pattern-Coding durchgeführt. [34].

4.4 Abgrenzung von anderen Methoden

Es wurde sich für eine Interviewstudie statt einer Umfragestudie entschieden, da der Fokus dieser Arbeit vor allem auf qualitativen statt quantitativen Ergebnissen liegen soll. Weiterhin ist es sinnvoll, im Hinblick auf das eher unbekanntes Thema Stimmungsanalyse, den Teilnehmern die Möglichkeit zu geben, Verständnisfragen zu stellen. Weiterhin kann im Rahmen eines Interviews nach Gründen für die gestellten Anforderungen gefragt werden.

4.5 Teilnehmerakquise

Um die Qualität der in dieser Arbeit gesammelten Daten sicherzustellen, kamen für die Teilnahme an der Studie nur Teilnehmer in Betracht, die mehrjährige Berufserfahrung (mindestens vier) in ihrer jeweiligen Rolle besitzen. Um Teilnehmer zu gewinnen, wurde eine Nachricht an den Newsletter des Fachbereichs Requirements Engineering der Gesellschaft für Informatik geschickt. Darin wurde die geplante Studie erläutert und Empfänger konnten sich melden, wenn sie an einer Teilnahme interessiert waren. Von den Teilnehmern wurde das Einverständnis eingeholt.

4.6 Video

Die Teilnehmer, die an der Studie dieser Arbeit teilgenommen haben, besaßen, ausgenommen dem Forscher, kein Hintergrundwissen zum Thema

Stimmungsanalyse. Um alle Teilnehmer auf den gleichen Wissensstand zu bringen und da nicht erwartet werden kann, dass Anforderungen zu einem unbekanntem Thema genannt werden, wurde im Vorfeld ein Video erstellt, was den Teilnehmern die Stimmungsanalyse und mögliche Anwendungsfälle neutral erklärt. Das Video wurde den Teilnehmern im Rahmen des ersten Interviews vorgespielt. Anschließend bekamen sie die Möglichkeit, Fragen dazu zu stellen. Erst danach wurden ihnen die initialen Anforderungen präsentiert.

Kapitel 5

Ergebnisse

5.1 RQ1: Identifizierte Stakeholder

Die erste Forschungsfrage lautet, welche Stakeholdergruppen ein Interesse an der Stimmungsanalyse im Kontext von Softwareprojekten haben. Diese Frage lässt sich durch die Literaturrecherche beantworten. Der Prozess der Literaturrecherche ist in Kapitel 3 beschrieben. Zunächst soll identifiziert werden, welche Stakeholder in Softwareprojekten vorkommen. Anschließend soll evaluiert werden, welche davon auch ein Interesse an der Stimmungsanalyse haben könnten. Dazu wird explizit nach Anwendungsszenarien von Stimmungsanalyse im Kontext von Softwareprojekten gesucht. Aus den beschriebenen Szenarien geht hervor, welche Rolle die Person, die die Stimmungsanalyse verwendet, im Softwareprojekt einnimmt. Im Folgenden werden zunächst Stakeholder ermittelt, die generell in Softwareprojekten vorkommen.

5.1.1 Identifikation relevanter Stakeholder

In Kapitel 2 wurden sechs Methoden zur Identifikation von Stakeholdern in Softwareprojekten vorgestellt. Vorgeschlagene Stakeholdergruppen sind dort ebenfalls in Listenform angegeben. Daraus lässt sich folgende, zusammengefasste Liste bilden, die mögliche Stakeholder in Softwareprojekten benennt:

- Kunde (SEI, CMMI, RUP, Pressman)
- Benutzer (SEI, CMMI, RPU, Sommerville)
- **Entwickler** (SEI, CMMI, RPU, Pressman)
- **Qualitätssicherung und Testung** (SEI)
- Sicherheitspersonal (SEI)
- **Requirements Engineer/ Analyst** (SEI, LifeCycle)

- Produzenten (CMMI)
- Lieferanten (CMMI)
- Marketing (CMMI, Pressman)
- Design / Produkt-Ingenieure (LifeCycle, Pressmann)
- Integration (LifeCycle)
- Delivery (LifeCycle)
- **Wartung und Support** (CMMI, LifeCycle, Pressman)
- **Projektleiter** (RPU)
- **Business Managers** (Pressman, Sommerville)
- Brand Managers (Pressman)
- Berater (Pressman)
- **Domänenexperte**
- jeder, der interessiert oder beteiligt ist (CMMI, RPU, Sommerville)

Die Guidelines, in denen die Stakeholder jeweils erwähnt werden bzw. aus denen sie stammen, stehen jeweils im Klammern hinter dem jeweiligen Stakeholder. Die fett geschriebenen Stakeholder sind jene, für die die Literaturrecherche ergeben hat, dass sie auch für die Stimmunganalyse relevant sind, da Anwendungsszenarien beschrieben sind, d.h. es wird in der Literatur erläutert, welche Rolle mit welchem Ziel und in welchem Abschnitt eines Softwareprojekts Stimmungsanalyse verwenden könnte. Im Folgenden werden die identifizierten Stakeholdergruppen im einzelnen erläutert.

5.1.2 Forschende

Sommerville [37] beschreibt als mögliche Stakeholdergruppe Domänenexperten. Im Fall von Stimmungsanalyse sind dies Forscher, die Stimmungsanalysetools entwickeln und an ihnen forschen. Das Interesse von Forschenden erklärt sich auch bereits durch die sehr große Anzahl an Arbeiten, die im Rahmen der Literaturrecherche ausfindig gemacht wurden. Daher soll diese Stakeholdergruppe in die Studie integriert werden.

5.1.3 Entwickler

Softwareentwickler sind direkt von positiver oder negativer Stimmung innerhalb des ihnen zugewiesenen Projekts betroffen. Verschiedene Studien zeigen, dass sich die im Projekt herrschende Stimmung nicht nur auf die Zufriedenheit von Entwicklern, sondern auch auf deren mentale Gesundheit, Produktivität und kognitiven Leistungen auswirkt [12] [25] [15].

5.1.4 Projektleitung

Die Projektleitung eines Softwareprojektes ist verantwortlich für den Projekterfolg. Wie bereits oben beschrieben, wirkt sich die Stimmung innerhalb eines Softwareprojekts auch auf die Leistung der Entwickler aus. Für den zufriedenstellenden Abschluss eines Projekts ist es daher sinnvoll, sicherzustellen, dass die Stimmung des Projektteams vorwiegend positiv ist. Schneider et al. [35] beschreiben auch, wie wichtig angemessene und konstruktive Kommunikation in Meetings für den Projekterfolg ist.

5.1.5 Management

Für das betriebswirtschaftliche Management kommen verschiedene Aspekte und Anwendungsmöglichkeiten in Frage. Der Erfolg eines Softwareprodukts hat auch direkt Einfluss auf den Unternehmenserfolg, daher gelten auch hier die gleichen Aspekte wie bereits bei der Projektleitung erläutert. Mitunter ist auch das Personalwesen betroffen, z.B. kann durch Stimmungsanalyse Unzufriedenheit von allen am Projekt beteiligten Personen frühzeitig erkannt werden [25]. Somit besteht die Möglichkeit, die Stimmung zu verbessern, damit es nicht zu Kündigungen oder Ausfällen kommt. Außerdem kann mittels Analyse von Kundenreviews Rückschluss auf die Marktposition eines Produkts gezogen werden [2].

5.1.6 Requirements Engineer

In der Literatur wird beschrieben, dass Requirements Engineers Stimmungsanalyse verwenden könnten, um Dokumente, die Anforderungen enthalten bzw. beschreiben, auf Neutralität hin zu überprüfen [40]. Außerdem kann eine Stimmungsanalyse von Kundenreviews stattfinden, um mögliche neue Features oder Verbesserungen zu identifizieren [3] [45]. In der Liste möglicher Stakeholder, die oben zu finden ist, ist außerdem 'Wartung und Support' als relevante Gruppe aufgeführt. Die in der Literatur beschriebenen Anwendungsbereiche decken sich allerdings mit dem des Requirements Engineers und beschränken sich auf die Untersuchung von Kundenreviews. Aus diesem Grund und auch, damit das Ergebnis nicht verzerrt wird, wenn zwei deckungsgleiche Stakeholdergruppen in der Studie vorkommen, wird 'Wartung und Support' für die Interviews nicht berücksichtigt.

5.1.7 Qualitätssicherung

Beschäftigte in der Qualitätssicherung können ebenfalls von der Verwendung von Stimmungsanalysetools profitieren. Graziotin et al. [10] erläutern, dass die Qualität des von Entwicklern verfassten Codes sich durch eine positive Stimmung innerhalb des Softwareprojekts verbessert. Da die Aufgabe der Qualitätssicherung ist, die Qualität des Codes möglichst hoch zu halten,

trägt die Verwendung von Stimmungsanalysetools zu diesem Ziel bei. Weiterhin wird in der Literatur die Anwendung von Stimmungsanalysetools im Rahmen von Code Reviews beschrieben, an denen die Qualitätssicherung unmittelbar beteiligt ist [1].

5.1.8 Ausschluss anderer Stakeholder

In der oben zu findenden Auflistung existieren noch weitere mögliche Stakeholder. Die Aufnahme dieser Stakeholder wurde jedoch im Kontext von Softwareprojekten nicht als sinnvoll erachtet. Beispielweise erschließt sich für den Anwender bzw. Benutzer kein sinnvolles Anwendungsszenario. Und bis auf die Anforderungsanalyse ist er auch nicht an der Softwareentwicklung beteiligt- er verwendet lediglich das darauf entstehende Produkt. Kundenreviews können auf Stimmungen hin untersucht werden, dies wird aber von am Softwareprojekt beteiligten Personen durchgeführt, z.B. vom Requirements Engineer, und nicht vom Kunden selbst. Einige Stakeholder sind auch sehr spezifisch. So könnten Brand Manager oder Mitarbeiter des Marketings auch zum betriebswissenschaftlichen Management gezählt werden. Da sich in dieser Arbeit ein erster Überblick verschafft werden sollte, wäre eine zu große Differenzierung nicht sinnvoll und daher wurde sich auf die oben genannten Stakeholdergruppen beschränkt.

5.2 RQ2: Anforderungen an ein Stimmungsanalysetool

5.2.1 Anforderungen aus der unveröffentlichten Studie

Nachdem die zur Verfügung gestellten Dokumente aus der unveröffentlichten Studie untersucht wurden, ergaben sich die unter Kapitel 4 beschriebenen drei Anforderungskategorien (Erkennung von Stimmungen, Oberfläche, Verständnis). Wie bereits in Kapitel 4 erläutert, wurden die Anforderungen für die Interviews in einer Baumstruktur organisiert. Die Baumstruktur ist in Abbildung 5.1 dargestellt. Die Anforderungen werden im Folgenden erläutert.

Erkennen von Stimmungen

Teilnehmer der Studie äußerten mehrfach den Wunsch, dass **Ironie und Sarkasmus** von einem Stimmungsanalysetool als solche erkannt werden. Eine ironische oder sarkastische Aussage sollte also nicht zwingend negativ bewertet werden. Das gleiche gilt für **doppelte Verneinungen**. Auch diese sollen vom Tool erkannt und nicht negativ bewertet werden. Des Weiteren wurde die Anforderung geäußert, dass typische Äußerungen aus der **Umgangssprache** vom Stimmungsanalysetool erkannt werden. Auf die

5.2. RQ2: ANFORDERUNGEN AN EIN STIMMUNGSANALYSETOOL35

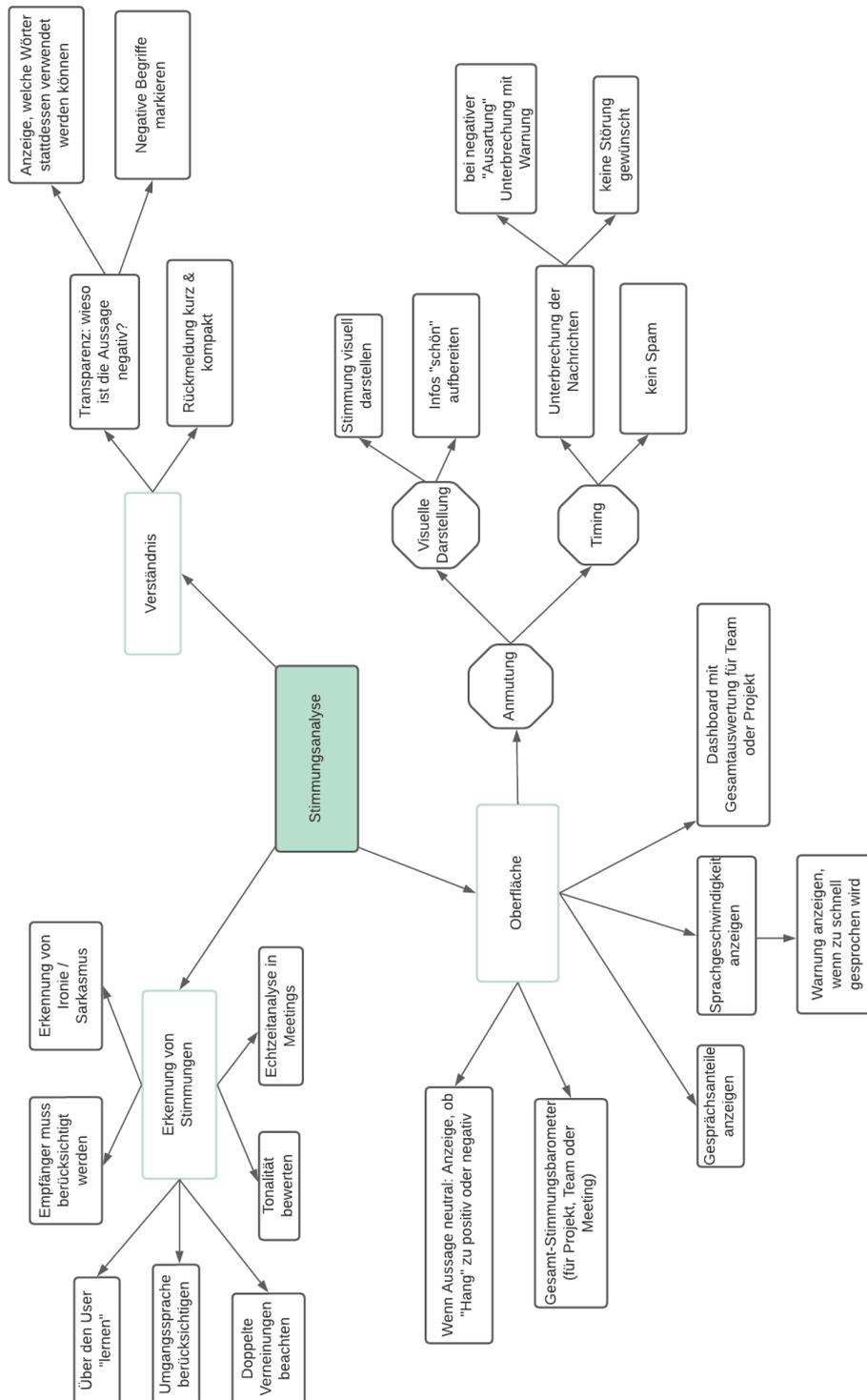


Abbildung 5.1: Anforderungen aus unveröffentlichter Studie

Bewertung soll auch der **Empfänger**, an den die Aussage oder der Text gerichtet ist, Einfluss nehmen. So soll also berücksichtigt werden, ob beispielsweise mit dem Chef, einem Kollegen oder einen Kunden kommuniziert wird. Weiterhin wurde der Wunsch geäußert, dass das Stimmungsanalysetool über den **Benutzer lernt**. Damit ist gemeint, dass durch geschriebene Nachrichten oder getätigte Aussagen ein Profil über den Benutzer erstellt wird. Es könnte somit ermittelt werden, ob der Benutzer oft selbst Sarkasmus benutzt. Daraus könnte unter Umständen gefolgert werden, dass der Erhalt einer sarkastischen Nachricht ihm weniger negative Stimmung vermittelt als jemanden, der selbst nie Sarkasmus oder Ironie benutzt. Eine weitere Anforderung besteht darin, dass Stimmungsanalysetools eine **Echtzeitanalyse in Meetings** ermöglichen, das heißt, dass sie während eines Meetings eingesetzt werden können und in Echtzeit Feedback über die Stimmung des Gesagten ausgeben. Bei verbalen Äußerungen sollte dann auch die **Tonalität** Einfluss auf die Bewertung nehmen.

Oberfläche

Teilnehmer der Studie nannten den Wunsch, dass die grafische Benutzeroberfläche eine Art **Dashboard** aufweist, damit eine Übersicht über die Stimmung innerhalb eines Projekts oder Teams gewährleistet ist. Eine weitere Anforderung besteht darin, dass die Benutzeroberfläche ein **Gesamt-Stimmungsbarometer** bietet, beispielsweise in Form eines Thermometers. Dies kann sich entweder auf die Stimmung eines Projekts, eines Teams oder ein Meeting beziehen. Generell wurde die Anforderung genannt, dass die ermittelte Stimmung **visuell dargesellt** wird. Beispielsweise könnten als negativ formulierte Aussagen rot markiert werden. Informationen, die der Benutzer wahrnimmt, sollten weiterhin **'schön'** aufbereitet werden, sodass die Darstellung vom Benutzer als angenehm empfunden wird. Oft wurde auch kritisiert, dass die Stimmung entweder als negativ, neutral oder positiv bewertet wird. In diesem Zug wurde sich gewünscht, insbesondere wenn ein Satz als neutral bewertet wird, dass das Tool eine Art **Skala** anzeigt, um herauszufinden, ob die Aussage eher einen **Hang** zum Positiven oder Negativen aufweist. Alternativ wäre auch eine differenziertere Bewertungsskala denkbar, beispielsweise von -5 bis 5, wobei 0 als neutral einzustufen ist. Wichtig war den Teilnehmern auch, dass durch die Ergebnisausgabe des Stimmungsanalysetools kein Spam produziert wird, sondern die Ergebnisse minimal und nicht störend dargestellt werden.

Die folgenden Anforderungen beziehen sich auf die Echtzeitanalyse in Meetings. So soll bei Meetings oder Konferenzen die **Sprachgeschwindigkeit** des Sprechenden angezeigt werden. Ist diese zu schnell und könnte dies zu Verständnisproblemen führen, soll außerdem eine **Warnung** angezeigt werden, falls zu schnell gesprochen wird. Das gleiche gilt für Gesprächsanteile. Die **Gesprächsanteile der beteiligten Teilnehmer** an dem Meeting

5.2. RQ2: ANFORDERUNGEN AN EIN STIMMUNGSANALYSETOOL³⁷

sollen angezeigt werden, damit ermittelt werden kann, ob ein Teilnehmer das Meeting dominiert. Falls während des Meetings eine extrem negative Gesamtstimmung ermittelt wird, gibt es zwei verschiedene Möglichkeiten, was passieren soll. Einige Teilnehmer wünschten sich eine **Unterbrechung des Meetings** mit einer Warnung. Andere Teilnehmer dagegen wiesen darauf hin, dass sie **keine Störung** wünschen.

Verständnis

Teilnehmer äußerten die Anforderung, dass **Transparenz** darüber herrschen muss, wieso eine Aussage vom Stimmungsanalysetool positiv oder negativ bewertet wird. Denkbar wäre es daher, **negative Begriffe zu markieren**. Nützlich sei es auch, wenn das Tool anzeigen könne, welche **Wörter als Alternative** verwendet werden könnten, damit die Aussage positiver bewertet wird. Des Weiteren wurde die Anforderung genannt, dass die Rückmeldung des Stimmungsanalysetools **kurz und kompakt** ist, d.h. dass der Nutzer ein schneller und verständliches Feedback erhält, damit er sich einen raschen Überblick über die Bewertung verschaffen kann.

5.2.2 Ergebnisse aus dem ersten Interview

In den durchgeführten Interviews wurde für jeden Teilnehmer ein eigener Baum erstellt. Im ersten Interview bekam jeder Teilnehmer die gleichen Anforderungen, die oben beschrieben werden, in Baumstruktur organisiert vorgelegt. Anforderungen konnten anschließend gestrichen, ergänzt oder neu hinzugefügt werden. Als Ergebnis dieser Interviews sind somit sechs individuelle Baumstrukturen hervorgegangen, die im Anhang eingesehen werden können. Im Folgenden werden alle neu hinzugefügten und ergänzend genannten Anforderungen erläutert. In der ersten Interviewrunde wurden von den Teilnehmern der Studie zwei neue Kategorien, in die sich Anforderungen einordnen lassen, hinzugefügt. Es handelt sich um die Kategorien Integration und Administration.

Administration

Hier wurde die Anforderung genannt, explizit **positiv oder negativ zu bewertende Begriffe festzulegen**. Außerdem soll es die Möglichkeit geben, **wahlweise nur auf negative oder positive Formulierungen zu analysieren**. Auch sollen **Schwellwerte** z.B. für Sarkasmus und Ironie konfigurierbar sein. Das bedeutet, dass beispielsweise ein geringer Anteil an Ironie nicht als negativ zu bewerten ist. Falls dies sich allerdings häuft, kann die Ironie negativ bewertet werden. Eine weitere Anforderung besteht darin, dass **Logfiles** zur Verfügung stehen und die Bewertung der Stimmung darin dokumentiert wird. Ferner wurden **Entwickleroptionen** als Anforderung genannt, sich allerdings bisher nicht näher dazu geäußert.

Integration

Teilnehmer erklärten, dass es für die Verwendung eines Stimmungsanalyse-tools nötig sei, dass dieses in typische **Software- und Kommunikations-tools** integrierbar ist. Darunter fallen Ticketing-Systeme wie Jira, Email-Clients wie Outlook, Kollaborationstools wie Slack oder Teams, sowie Content-Management-Systeme oder Wikis. Weiterhin sei es für die Verwendung innerhalb eines Präsenz-Meetings erforderlich, dass die Möglichkeit einer Sprach- oder Videoaufnahme über ein Mikrofon und/oder eine Kamera besteht, also eine Schnittstelle zu **Hardware-Devices** existiert. Dazu würde dann auch eine zugehörige **Audio-zu-Text-Verarbeitung** erforderlich sein.

Erkennen von Stimmungen

Zusätzlich oder ergänzend zu den aus der unveröffentlichten Studie ermittelten Anforderungen, die oben genannt wurden, wurden in der ersten Interviewrunde folgende Anforderungen genannt: Ergänzend zu der Anforderung des Berücksichtigens der Tonalität wurden **weitere Erkennungsmöglichkeiten** genannt. Als Beispiel wurden der Kontext, die Erkennung von Gesichtsausdrücken oder die Messung von biologischen Daten, wie z.B. die Ermittlung von Stress durch ein Armband, angeführt. Weiterhin wurde die Anforderung, der Empfänger solle berücksichtigt werden, ausgeweitet. So soll eine Bewertung nach verschiedenen **Typen** erfolgen. Beispielsweise sollen Benutzer dahingehend eingeteilt werden, ob sie eher introvertiert oder extrovertiert sind. Auf Grundlage dessen soll dann besser bewertet werden können, wie eine Aussage von einem bestimmten Empfänger aufgefasst werden könnte. Weiterhin sollen auch **Floskeln und Anrede** einer Mail daraufhin geprüft werden, ob sie adressatengerecht formuliert sind. Zur Anforderung der Berücksichtigung von Umgangssprache wurde hinzugefügt, dass auch **Faktoren wie Dialekte, Fachjargon und Jugendsprache** erkannt werden sollen. Es folgen nun die Anforderungen, die zur dieser Kategorie neu hinzugefügt wurden. Zum einen sollen auch **Rechtschreibung und Grammatik** geprüft werden. Das hat den Hintergrund, dass analysiert werden soll, ob der geschriebene Text verständlich ist, denn falls es sich um schwerverständliche Formulierungen handelt, könnten diese negative Stimmungen beim Empfänger auslösen. Als weitere Anforderungen wurde genannt, dass das Tool **effizient** ist, d.h. dass es bei der Bewertung einer Aussage oder eines Textes zu keiner Verzögerung bzw. Wartezeit kommt. Ein Teilnehmer äußerte auch, er wünsche sich, dass das Tool **keinen expliziten Start** erfordere, sondern dass es dauerhaft laufe.

Oberfläche

Folgende Anforderungen wurden in Bezug auf die Oberfläche ergänzt: Für das **Dashboard** wurden zusätzliche Funktionen definiert. So soll es

5.2. RQ2: ANFORDERUNGEN AN EIN STIMMUNGSANALYSETOOL³⁹

möglich sein, Stimmungen nach Projektmitgliedern oder Teilnehmern zu **gruppieren**. Weiterhin soll es ermöglicht werden, dass Benutzer jeweils am Anfang und Ende des Arbeitsalltags ihre **eigene Stimmung angeben**. Zum einen können so die Bewertungen des Stimmungsanalysetools verifiziert werden, andernfalls kann der Projektleiter sehen, wenn ein Entwickler beispielsweise morgens eine gute Stimmung aber am Tagesende eine schlechte Stimmung aufwies und die Ursachen dafür untersuchen. Eine weitere Funktion, die im Rahmen des Dashboards genannt wurde, war, dass im Dashboard **Trendlinien mit Peaks** angezeigt werden. Zur Anforderung 'Gesprächsanteile anzeigen' wurde hinzugefügt, dass sich gewünscht wird, nicht nur die Gesprächsanteile anzuzeigen, sondern auch den **Anteil von Fachbegriffen und Abkürzungen**, da ein hoher Anteil eventuell zu Verstimmungen führen konnte, falls diese nicht verstanden werden. Zur visuellen Darstellung von Stimmungen wurde angemerkt, dass diese Markierungen im Fall einer Videokonferenz **nicht öffentlich angezeigt** werden sollen, sondern nur für die jeweilige Person, die die Aussage getätigt hat, oder eventuell den Moderator, zu sehen ist. Weiterhin wurde die visuelle Darstellung als **Stimmungs-EKG** gewünscht. In Live-Meetings soll also, beispielsweise in einer Bildschirmecke, eine Kurve ähnlich der eines medizinischen EKGs (Elektrokardiogramm) angezeigt werden, die die Stimmung des Meetings visualisiert. Ein Teilnehmer merkte zur Anforderung, dass Informationen bezüglich ermittelter Stimmungen 'schön' dargestellt werden sollen, an, dass für ihn **Usability** wichtiger ist. Die letzte Ergänzung wurde für die Anforderung 'keine Störung in Live-Meetings' vorgenommen. Hier wurde vorgeschlagen, statt das Meeting mit einer Warnung im Fall einer negativen Stimmungsausartung zu unterbrechen, **Alternativen** zu implementieren. Diese könnten beispielsweise aus einer Nachricht an den Moderator oder einer späteren Rückmeldung bestehen. Komplette neu hinzugefügt wurden zwei Anforderungen. Zum einen wurde sich gewünscht, dass eine Funktion implementiert wird, die es ermöglicht, eine **einzelne Aussage oder einen einzelnen Satz zu 'ignorieren'**, d.h. ihn aus der Stimmungsbewertung herauszunehmen. Schlussendlich wurde noch die Anforderung geäußert, dass die grafische Benutzeroberfläche ein **minimalistisches, nicht aufdringliches Design** aufweisen sollte.

Verständnis

Zur Anforderung, dass die Bewertung transparent sein sollte, wurde hinzugefügt, dass die **Erklärung der Bewertung** weiter aufgeschlüsselt werden sollte, also beispielsweise die Informationen enthalten sollte, was als Sarkasmus, Umgangssprache, etc. bewertet wurde. Zur 'kompakten Rückmeldung' wurde hinzugefügt, dass dies zwar der Standard sein sollte, es auf Wunsch jedoch auch die Möglichkeit geben sollte, eine **detaillierte Rückmeldung** zu erhalten. Neu hinzugefügt wurden folgende Punkte: zum einen soll

ein **Tutorial** vorhanden sein, welches den Benutzern die Anwendung des Stimmungsanalysetools erklärt. Die zweite neue Anforderung bezieht sich auf die **Funktionsweise** des Tools. So soll ersichtlich sein, wie die Bewertung erfolgt. Beispielweise soll klar sein: handelt es sich um eine wortbasierte Bewertung oder existiert ein Machine-Learning-Algorithmus?

5.2.3 Ergebnisse aus dem zweiten Interview

In die Baumstruktur jedes Teilnehmers aus der ersten Interviewrunde wurden die Anforderungen der anderen Teilnehmer gemäß der unter Kapitel 4 erklärten iterativen Interviewstudie ergänzt. Im Rahmen der zweiten Interviewrunde tat sich eine weitere Kategorie auf: Datenschutz. Im Folgenden werden die neuen Anforderungen aus dieser Runde erläutert.

Administration

Zur Anforderung, dass speziell eine Filterung nach positiver oder negativer Stimmung möglich sein soll, wurde hinzugefügt, dass es ebenfalls einen **Filter** geben sollte, der es ermöglicht, nach bestimmten **Emotionen** (z.B. Wut) zu suchen. Weiterhin soll ein Filter realisiert werden, der es ermöglicht, explizit nach **Mobbing oder Sticheleien** gegen eine Person zu suchen. Es wurde ergänzt, dass zur Funktion, positive oder negative Begriffe selbst festzulegen, ergänzend die Möglichkeit vorhanden sein sollte, diese Listen explizit für bestimmte **Gruppen oder Profile** anzupassen und anzulegen. Zur Anforderung der Logfiles wurde hinzugefügt, dass auch **technische Logfiles** existieren sollen, um beispielsweise die verwendete Hardware (welches Mikrofon wurde benutzt?) oder Software (z.B. Microsoft Teams) identifizieren zu können.

Drei neue Anforderungen wurden in der zweiten Interviewrunde für die Kategorie Administration genannt. Angemerkt wurde, dass für die Einrichtung eines Administrationsbereichs ein **Rollen- und Berechtigungskonzept** erforderlich ist. Weiterhin soll es eine **benutzerfreundliche Administrationsoberfläche** geben. Eine weitere Anforderung besteht darin, dass die Bewertungen von Aspekten wie Sarkasmus, Fachbegriffen Umgangssprache, etc. **ein- und ausschaltbar** sind.

Datenschutz

Es wurde mehrfach angemerkt, dass bei der Verwendung eines Stimmungsanalysetools rechtliche Vorgaben eingehalten werden müssten. So sei es erforderlich, **Daten nach einem festgelegten Zeitraum** löschen zu können. Weiterhin sei es nötig, gesammelte **Daten zu anonymisieren**, sodass keine personenbezogenen Daten gesammelt werden.

Integration

Diese Kategorie wurde um die Anforderung ergänzt, dass die **Installation des Tools einfach** sein soll, d.h. dass es zum Beispiel per Klick herunterladbar oder verbunden mit einem Installationsassistenten ist.

Erkennen von Stimmungen

Diese Kategorie wurde um zwei neue Anforderungen ergänzt. Zunächst soll beim ersten Start des Stimmungsanalysetools ein **Fragebogen** erscheinen, der Informationen über die Persönlichkeit des Benutzers sammelt, d.h. es wird beispielsweise abgefragt, ob der Benutzer eher introvertiert oder extrovertiert ist. Auch Fragen zum Kommunikationsverhalten und zum Team seien denkbar. Auch sollen bei der Bewertung der Stimmung **unterschiedliche Kulturen** und deren Höflichkeitsempfinden beachtet werden. Bei der Gesichtserkennung sei dabei zu beachten, dass auch Gesten (z.B. Nicken) in bestimmten Kulturkreisen anders zu interpretieren sind.

Oberfläche

Hier wurde eine Anforderung ergänzt. Die **Benutzerfreundlichkeit bzw. Usability** der grafischen Benutzeroberfläche soll hoch sein, sodass diese einfach und intuitiv zu bedienen ist.

Verständnis

In dieser Kategorie wurde die Anforderung des Tutorials ergänzt. Hier wurden unterschiedliche Ausprägungen genannt. Einige Teilnehmer wünschten sich ein **Video-Tutorial**, eine **Schritt-für-Schritt-Anleitung**, ein **Wiki / README** oder ein ausführlicheres **Benutzerhandbuch** für spezifische Fragen. Ein Teilnehmer äußerte, dass es für ihn wichtig sei, dass ein Tutorial nicht automatisch startet, sondern nur bei Bedarf, also verbunden mit einem **Info- oder Helpbutton**.

5.3 RQ3: Priorisierung innerhalb der Stakeholdergruppen

Die Priorisierung der Anforderungen erfolgte im dritten Interview. Dafür stand den Teilnehmern die unter Kapitel 4 erläuterte Bewertungsskala zur Verfügung. Die Baumstrukturen können im Anhang eingesehen werden. Zu Übersichtszwecken wurden die vergebenen Priorisierungspunkte jedoch in Tabellenform übertragen.

5.3.1 Ergebnisse aus dem dritten Interview

Tabellen

Die gesammelten Priorisierungen aus der dritten Interviewrunde wurden in Tabellenform übertragen. Das dient der Übersichtlichkeit, aber auch der Vergleichbarkeit zwischen verschiedenen Teilnehmern. Im Folgenden werden die Tabellen abgebildet und Auffälligkeiten beschrieben.

Datenschutz	Entwickler	Forscher	Mngt.	PL	RE	QM	Summe	Mittelwert	Arith. M.	Std.-Abw.
Löschen von Daten	7	1	3	7	6	6	30	5	6	2,24
Anonymisierung	6	1	3	7	6	7	30	5	6	2,24
								5		

Abbildung 5.2: Priorisierungen zu Anforderungen der Kategorie Datenschutz

Hier fällt auf, dass die meisten Teilnehmer (bis auf Management und Forschung) die Anforderungen aus der Kategorie Datenschutz sehr hoch, also entweder mit 6 oder 7, bewertet haben. Der Teilnehmer aus der Forschung bewertete die Anforderungen aus dem Datenschutz nur mit einer 1, sie sind für ihn nicht relevant. Hieraus ergibt sich, dass die Standardabweichung bei beiden Anforderungen mit 2,24 sehr hoch ist. Zu beachten ist auch, dass kein Teilnehmer sich dazu entschied, eine Anforderung aus dieser Kategorie zu entfernen.

Integration	Entwickler	Forscher	Mngt.	PL	RE	QM	Summe	Mittelwert	Arith. M.	Std.-Abw.
SW und Kommunikationstools	7	7	4	7	5	7	37	6,17	7	1,21
HW-Devices für Spracheingabe	4	6	4	7	6	6	33	5,50	6	1,12
Weiterverarbeitung Audio-zu-Text	4	6	4	6	6	6	32	5,33	6	0,94
Installation per Klick	6	6	4	4	4	4	28	4,67	4,33	0,94
								5,42		

Abbildung 5.3: Priorisierungen zu Anforderungen der Kategorie Integration

Die am höchsten priorisierte Anforderung ist die, dass ein Stimmungsanalysetool in vorhandene Software- und Kommunikationstool integrierbar sein müsste. Vier von sechs Teilnehmern bewerteten diese Anforderungen mit der höchsten Punktzahl. Generell wurden die Anforderungen aus dieser Kategorie sehr hoch bewertet. Die geringste Punktzahl, die vergeben wurde, ist die 4, die für eine neutrale Bewertung steht. Der Teilnehmer aus dem Management bewertete alle Anforderungen als neutral. Die Anforderung, die im Allgemeinen als am wenigsten wichtig bewertet wurde, ist die, dass das Tool einfach und per Klick installierbar sein sollte. Jedoch ergibt sich auch hier ein Mittelwert von 4,67. Auch in dieser Kategorie wurde von keinem Teilnehmer eine Anforderung entfernt.

5.3. RQ3: PRIORISIERUNG INNERHALB DER STAKEHOLDERGRUPPEN⁴³

Administration	Entwickler	Forscher	Mngt.	PL	RE	QM	Summe	Mittelwert	Arith. M.	Std.-Abw.
benutzerfreundliche Adminoberfläche	4	6	2	2	5	5	24	4,00	4,50	1,53
neg. oder pos. Begriffe festlegen	3	6	7	6	4	6	32	5,33	6,00	1,37
anpassbar an Gruppen / Profile	3	6	6	5	4	5	29	4,83	5,00	1,07
Optionen wie Sarkasmus etc. ein- und ausschaltbar	3	7	4	6	3	6	29	4,83	5,42	1,57
Logfiles	3	6	4	1	2	5	21	3,50	3,75	1,71
technische Logfiles	6	7	4	1	2	5	25	4,17	4,58	2,11
Rollen- und Berechtigungskonzept	7	6	3	6	2	5	29	4,83	5,50	1,77
Schwellwerte für Sarkasmus etc. konfigurierbar	3	7	6	6	5	6	33	5,50	6,00	1,26
auf neg. oder pos. Formulierung analysieren	3	7	2	/	3	/	15	2,50	3,00	1,92
Filter (Emotionen, Mobbing)	3	7	2	7	3	5	27	4,50	4,75	1,98
								4,40		

Abbildung 5.4: Priorisierungen zu Anforderungen der Kategorie Administration

Auffällig ist hier, dass der Forscher alle Anforderungen dieser Kategorie sehr hoch priorisiert hat, er vergab Punktzahlen von 6 oder 7. Die Person, die das Requirements Engineering vertritt, bewertete die Kategorie Administration insgesamt am geringsten. Auffällig ist, dass die Standardabweichungen für die Anforderungen dieser Kategorie relativ hoch sind, bei einer Anforderung liegt sie bei 2,11. Bei weiteren Anforderungen liegt die Standardabweichung geringfügig unter 2. In dieser Anforderungskategorie entfernten zwei Teilnehmer, nämlich diejenigen aus dem Management und aus der Qualitätssicherung, die Anforderung 'wahlweise auf positive oder negative Stimmung analysieren'. Gleichzeitig ist dies mit einer durchschnittlichen Bewertung von 2,5 die am niedrigsten priorisierte Anforderung. Ansonsten wurde keine weitere Anforderung entfernt. Im Durchschnitt am höchsten priorisiert wurde die Anforderung, dass es die Möglichkeit geben soll, für Bewertungsfaktoren wie etwa Sarkasmus bestimmte Schwellwerte festzulegen. Der Mittelwert beträgt hier 5,5.

Verständnis	Entwickler	Forscher	Mngt.	PL	RE	QM	Summe	Mittelwert	Arith. M.	Std.-Abw.
Tutorial zur Anwendung	7	6	7	7	6	7	40	6,67	7,00	0,47
Readme/FAQ/Wiki	7	6	4	2	4	5	28	4,67	4,83	1,60
Benutzerhandbuch	5	5	4	1	5	7	27	4,50	5,00	1,80
Schritt-für-Schritt-Anleitung	5	6	4	6	6	4	31	5,17	5,58	0,90
Video-Tutorial	4	6	7	7	6	5	35	5,83	6,00	1,07
Help-Button	7	3	4	7	3	4	28	4,67	4,33	1,70
Transparenz	7	7	6	7	5	6	38	6,33	6,67	0,75
Erklärung Bewertung	7	7	6	7	3	6	36	6,00	6,50	1,41
Negative Begriffe markieren	2	5	5	6	2	5	25	4,17	5,00	1,57
Anzeige Alternative	5	/	6	/	4	5	20	3,33	5,00	0,71
Rückmeldung kurz & kompakt	6	3	6	6	7	6	34	5,67	6,00	1,25
detailliertere Rückmeldung auf Wunsch	5	5	6	7	/	5	28	4,67	5,00	0,80
								5,14		

Abbildung 5.5: Priorisierungen zu Anforderungen der Kategorie Verständnis

Die Kategorie Verständnis wurde insgesamt sehr hoch bewertet. Der Mittelwert über alle Teilnehmer und alle Anforderungen beträgt 5,14. Die

am höchsten priorisierte Anforderung ist die, dass ein Tutorial existieren sollte, das es den Teilnehmern ermöglicht, eine Erklärung zur Verwendung des Tools zu bekommen. Der Mittelwert der Priorisierungen dieser Anforderungen beträgt 6,67. Somit ist diese Anforderung auch die am höchsten priorisierte aus dem gesamten Anforderungskatalog, nicht nur aus dieser Kategorie. Die Standardabweichung ist sehr gering und beträgt 0,47. Die Standardabweichung dieser Anforderung ist auch die niedrigste des gesamten Anforderungskatalogs. Die am geringsten priorisierte Anforderung ist die, dass eine Anzeige einer Alternative zu negativ markierten Begriffen erfolgen soll. Hier ergibt sich ein Mittelwert von 3,33, bei einer geringen Standardabweichung von 0,71. Die höchste Standardabweichung wiederum ergibt sich bei der Anforderung, dass ein Benutzerhandbuch existieren sollte, welches ausführlich die Benutzung des Tools beschreibt und das bei spezifischeren Fragen als Nachschlagemöglichkeit dient. Hier ergibt sich eine Standardabweichung von 1,80. Entfernt wurden folgende Anforderungen: der Teilnehmer des Requirements Engineerings entfernte die Anforderung der detaillierteren Rückmeldung auf Wunsch. Forscher und Projektleiter entschieden sich dazu, die Anforderung der Anzeige von Alternativen bei negativ markierten Begriffen zu entfernen. Ansonsten wurden keine weiteren Anforderungen gelöscht. Insgesamt ergibt sich bei dieser Kategorie über

Erkennen von Stimmungen	Entwickler	Forscher	Mngt.	PL	RE	QM	Summe	Mittelwert	Arith. M.	Std.-Abw.
Echtzeitanalyse in Meetings	4	6	4	6	5	5	30	5,00	5,00	0,82
Erkennung Sarkasmus	5	5	6	5	3	4	28	4,67	5,00	0,94
Kontext berücksichtigen	5	5	7	6	6	5	34	5,67	5,83	0,75
läuft dauerhaft, kein Start erforderlich	5	5	2	/	3	4	19	3,17	4,00	1,17
effizient	7	6	2	4	2	3	24	4,00	4,00	1,91
Umgangssprache berücksichtigen	5	2	6	5	5	4	27	4,50	5,00	1,26
dabei verschiedene Faktoren berücksichtigen	5	2	6	5	5	4	27	4,50	5,00	1,26
über den User lernen	3	5	6	3	4	/	21	3,50	4,00	1,17
Fragebogen Persönlichkeit	2	6	5	1	6	5	25	4,17	5,00	1,95
unterschiedliche Kulturen	5	2	5	6	6	6	30	5,00	5,50	1,41
Empfänger muss berücksichtigt werden	4	6	4	6	7	5	32	5,33	5,67	1,11
Bewertung nach verschiedenen Typen	/	/	4	/	7	/	11	1,83	5,50	1,50
Flaskeln und Anrede überprüfen	5	6	4	/	7	5	27	4,50	5,00	1,02
Weitere Erkennungsmöglichkeiten	2	5	6	6	6	6	31	5,17	6,00	1,46
Rechtschreibung, Grammatik und Klarheit prüfen	/	/	2	/	4	/	6	1,00	3,00	1,00
								4,13		

Abbildung 5.6: Priorisierungen zu Anforderungen der Kategorie Erkennen von Stimmungen

alle Anforderungen und Teilnehmern ein Mittelwert von 4,13. Auffällig ist, dass hier relativ viele Anforderungen entfernt wurden. Der Projektleiter entfernte die Anforderung, dass das Stimmungsanalysetool dauerhaft laufen sollte und das kein expliziter Start erforderlich ist. Der Teilnehmer aus der Qualitätssicherung nahm die Anforderung, dass das Tool über den User lernen sollte, heraus. Die Anforderung, nach welcher verschiedener Adressatentypen für die Bewertung herangezogen werden sollten, wurde sowohl

5.3. RQ3: PRIORISIERUNG INNERHALB DER STAKEHOLDERGRUPPEN⁴⁵

vom Entwickler, Forscher, Projektleiter als auch von der Qualitätssicherung entfernt. Der Projektleiter löschte zusätzlich die Anforderung, dass Floskeln und Anrede bei Schriftverkehr geprüft werden sollten. Die Anforderung der Rechtschreibprüfung wurde sowohl von der Qualitätssicherung, vom Entwickler, Forscher und vom Projektleiter entfernt. Die am höchsten priorisierte Anforderung ist die, dass der Kontext beim Erkennen von Stimmungen berücksichtigt werden sollte. Der Mittelwert beträgt hier 5,67 und die Standardabweichung ist mit einem Wert von 0,75 vergleichsweise gering. Die höchste Standardabweichung findet sich bei der Anforderung, dass beim ersten Start der Anwendung ein Fragebogen erscheint, um den Charakter des Anwenders zu ermitteln. Die Standardabweichung beträgt hier 1,95. Die höchste Bewertung vergab hier der Forscher mit einer Punktzahl von 6, die geringste vergab der Projektleiter mit einer Punktzahl von 1.

Der Mittelwert der Anforderungen aus der Kategorie Oberfläche beträgt

Oberfläche	Entwickler	Forscher	Mngt.	PL	RE	QM	Summe	Mittelwert	Arith. M.	Std.-Abw.
minimalistisches Design	6	2	4	6	4	4	26	4,33	4,17	1,37
Option, zu ignorieren	2	/	4	/	5	6	17	2,83	4,50	1,48
Benutzerfreundlichkeit / Usability	5	5	6	6	7	5	34	5,67	5,83	0,75
Wenn Aussage neutral, Hang anzeigen	6	5	5	3	/	/	19	3,17	5,00	1,09
Gesamt-Stimmungsbarometer	3	5	4	7	2	5	26	4,33	4,67	1,60
Gesprächsanteile anzeigen, evtl. warnen	/	/	5	7	3	4	19	3,17	4,50	1,48
Anteil von Fachbegriffen anzeigen	/	/	5	6	4	2	17	2,83	4,50	1,48
Sprachgeschwindigkeit anzeigen	/	/	4	/	3	4	11	1,83	4,00	0,47
Dashboard mit Gesamtauswertung	3	7	3	7	5	5	30	5,00	5,00	1,63
Trendlinien mit Peaks	3	5	3	7	3	5	26	4,33	4,67	1,49
Teilnehmer geben eigene Stimmung an	2	7	3	6	/	/	18	3,00	4,50	2,06
Gruppierung der Stimmungen nach Teilnehmern	3	6	3	7	5	/	24	4,00	5,00	1,60
Stimmung visuell darstellen	3	7	5	6	3	6	30	5,00	5,50	1,53
Markierungen nicht öffentlich	3	7	2	7	3	6	28	4,67	5,33	2,05
Stimmungs-EKG in Live-Meetings	3	6	5	6	/	5	25	4,17	5,00	1,10
Infos 'schön' aufbereiten	/	6	2	/	/	/	8	1,33	4,00	2,00
bei Live-Meetings: keine Störung gewünscht	6	6	2	6	4	6	30	5,00	6,00	1,53
bei neg. Ausartung: Unterbrechung mit Warnung	5	/	2	1	/	/	8	1,33	2,00	1,70
kein Spam	7	4	4	/	5	6	26	4,33	5,00	1,17
								3,70		

Abbildung 5.7: Priorisierungen zu Anforderungen der Kategorie Oberfläche

insgesamt 3,70. In dieser Kategorie wurden außerdem viele Anforderungen entfernt. Die Option, Sätze zu ignorieren wurde vom Projektleiter herausgenommen. Die Anforderung, den 'Hang' oder eine Skala bei einer neutralen Aussage anzuzeigen, wurden sowohl vom Requirements Engineer als auch von der Qualitätssicherung entfernt. Entwickler und Forscher äußerten, dass es für sie nicht relevant sei, Gesprächsanteile verschiedener Teilnehmer in Meetings anzuzeigen. Dazu gehört auch der Anteil von Fachbegriffen. Diese beiden Gruppen und zusätzlich der Projektleiter entfernten auch die Anforderung, nach welcher die Sprachgeschwindigkeit der Teilnehmer angezeigt werden sollte. Diese Anforderung hat mit 0,47 auch die geringste Standardabweichung in dieser Kategorie und mit einer Anforderung aus der

Kategorie Verständnis zusammen auch die geringste Standardabweichung des gesamten Anforderungskatalogs. Requirements Engineer und Qualitätssicherung nahmen die Anforderung, dass Teilnehmer am Tagesanfang und Tagesende die eigene Stimmung angeben, heraus. Zu dieser Anforderung gehört auch die höchste Standardabweichung dieser Kategorie. Sie beträgt 2,06. Die Qualitätssicherung löschte zusätzlich die Anforderung, dass auf dem Dashboard Stimmungen nach Teilnehmern oder Projekten gruppiert werden sollten.

In Vivo Coding

Nach der Durchführung des in Vivo-Codings wie unter Kapitel 4 beschrieben, ergaben sich für die Kommentare zur Bewertung die im Folgenden dargestellten Ergebnisse. Die Anforderung, auf die sich die Aussage bezieht, ist dem Kommentare vorangestellt und in fetter Schrift markiert.

Entwickler

Tutorial zur Anwendung: 'Es muss klar sein, wie das Tool funktioniert'.

Erklärung der Bewertung: 'Auch hier muss deutlich sein, wie die Bewertung erfolgt, sonst kann man diese überhaupt nicht nachvollziehen, gerade, wenn man seine Aussage z.B. selbst nicht als negativ ansieht'.

Help-Button: 'Das Tutorial sollte jederzeit über den Help-Button aufrufbar sein, wenn man nicht weiter weiß'.

Löschen von Daten nach Zeitraum: 'Das ist eine gesetzliche Vorgabe und ohne kann das Tool gar nicht legal benutzt werden'.

Software und Kommunikationstools: 'Um die Benutzbarkeit zu gewährleisten, muss das Tool in übliche Software wie MS Teams, Outlook etc. eingebunden werden können'.

Forscher

Transparenz UND Erklärung der Bewertung: 'Beides muss vorhanden sein, damit das Tool und die Ergebnisse nachvollziehbar und erklärbar sind'.

Datenschutz: 'Für mich in meiner Tätigkeit als Forscher nicht relevant'.

Integration in Software und Kommunikationstools: 'Eine Integration wäre sehr wichtig für mich für weitere Studien'.

ALLE Konfigurationsmöglichkeiten: 'Die Möglichkeit, bestimmte Einstellungen am Tool und der Bewertung vornehmen zu können ist äußerst wichtig für meine weiteren Studien und für gewisse Forschungsfragen'.

Logfiles: 'Benötige ich zum Debuggen'.

Stimmung visuell darstellen: 'Das ist für mich wichtig, weil die visuelle Darstellung von Stimmungen ein Teil meiner Forschung ist'.

Markierungen nicht öffentlich: 'Die Priorisierung als besonders wichtig erfolgt aus meinen Erkenntnissen aus Forschung und Anwendung'.

Dashboard: 'Für meine Studien erforderlich'.

5.3. RQ3: PRIORISIERUNG INNERHALB DER STAKEHOLDERGRUPPEN47

Teilnehmer geben eigene Stimmung an: 'Wichtig, damit ein Abgleich erfolgen kann'.

Management

Kontext berücksichtigen: 'Ohne Berücksichtigung des Kontexts kann keine korrekte Bewertung erfolgen, da gar nicht klar ist, in welcher Situation sich der Anwender befindet'.

Tutorial zur Anwendung: 'Ohne Tutorial weiß der Nutzer nicht, wie das Tool angewandt werden soll und es entsteht Ablehnung'.

Video-Tutorial: 'Ich lasse mir gerne etwas per Video erklären, das ist für mich persönlich am komfortabelsten'.

Negative oder positive Begriffe festlegen: 'Das ist für mich unerlässlich, es muss für mich unbedingt die Möglichkeit geben, z.B. negative Begriffe für mein Unternehmen, Projekt oder Team selbst festzulegen'.

Projektleiter

Fragebogen zu Persönlichkeit: 'Für mich nicht so wichtig. Die meisten Leute wissen gar nicht, ob sie eher introvertiert oder extrovertiert sind, bzw. wie sie sich da einordnen sollen'.

Tutorial zur Anwendung: 'Wenn keiner weiß, wie das Tool verwendet wird, führt das nur zu Ablehnung'.

Help-Button: 'Das Tutorial sollte nur über den Button aufrufbar sein, ich mag es gar nicht, wenn ich eine Anwendung starte und mir direkt Tutorials aufgedrängt werden'.

Transparenz: 'Wenn das Tool nicht transparent ist, dann führt das zur Ablehnung durch die Benutzer'.

Erklärung der Bewertung: 'Wird benötigt, um zu validieren, um das System funktioniert'.

detaillierte Rückmeldung: 'Das ist besonders wichtig für Benutzer, um zu verstehen, was gemacht und bewertet wurde. Eine kurze Rückmeldung ist da nicht immer ausreichend, manchmal muss es einfach ausführlicher erklärt werden'.

Löschen von Daten / Anonymisierung: 'Ohne die Implementierung von Datenschutzvorgaben kann das Tool nicht verwendet werden, da gibt es sehr große Probleme mit dem Betriebsrat'.

Integration in Software und Kommunikationstools / Hardware-Devices: 'Das Tool muss integrierbar sein, sonst wäre es nicht effektiv nutzbar'.

Filter nach bestimmten Emotionen / Mobbing: 'Als Projektleiter möchte ich herausfinden, ob jemand z.B. dauerhaft schlechte Laune hat. Außerdem möchte ich Mobbing extrahieren und dem entgegenwirken'.

Logfiles: 'Logfiles interessieren mich in meiner Rolle nicht, sie sind für mich einfach nicht relevant'.

Markierungen nicht öffentlich: 'Sonst hätte man Datenschutz-Probleme

und Arbeitnehmer würden sich beschweren’.

Unterbrechen der Nachrichten bei Live-Meetings: ’Das ist gefährlich, denn es könnte als Verletzung der Meinungsfreiheit gewertet werden’.

Dashboard: ’Brauche ich zur Steuerung, damit nicht alles nur eine Momentaufnahme ist’.

textbfGesprächsanteile anzeigen: ’Zur Steuerung von Meetings wichtig’.

Gesamt-Stimmungsbarometer: ’Zur Steuerung von einem Projekt, Team oder Meeting’.

Requirements Engineer

Empfänger muss berücksichtigt werden: ’Die einzelne Person macht für die Bewertung viel aus, und wenn nicht bekannt ist, an welche Person die Aussage gerichtet ist, kann die Bewertung nachhinten losgehen und wird nicht richtig eingeordnet.’

Rückmeldung kurz und kompakt: ’Benötige ich, um schnell und richtig reagieren zu können, vor allem in Live-Meetings’.

Benutzerfreundlichkeit/Usability: ’Muss besonders hoch sein, sonst kann man sich innerhalb des Tools nicht zurechtfinden und das Tool zu nutzen würde keinen Spaß bereiten’.

Qualitätssicherung

Tutorial: ’Es muss eine Erklärung erfolgen, ansonsten erfolgt wahrscheinlich eine Fehl- oder Nichtbenutzung’.

Benutzerhandbuch: ’Es muss auch ein ausführliches Benutzerhandbuch geben, in dem ich spezifischere Fragen nachschlagen kann’.

Integration in Software und Kommunikationstools: ’Die verwendete Software ist in unserem Unternehmen gesetzt, und das Tool müsste dort integriert werden’.

Pattern Coding

Im zweiten Coding-Zyklus wurden die Aussagen nach Pattern geordnet. Als Pattern wurden hier die verschiedenen Anforderungskategorien ausgewählt, also Erkennen von Stimmungen, Oberfläche, Verständnis, Datenschutz, Administration und Integration. Des Weiteren wurde ein weiteres Pattern definiert: Sorgen und Bedenken. Im Folgenden werden Ergebnisse des in Vivo Codings noch einmal nach Pattern geordnet dargestellt.

Erkennen von Stimmungen

Empfänger muss berücksichtigt werden: ’Die einzelne Person macht für die Bewertung viel aus, und wenn nicht bekannt ist, an welche Person die Aussage gerichtet ist, kann die Bewertung nach hinten losgehen und wird nicht richtig eingeordnet’.

Kontext berücksichtigen: 'Ohne Berücksichtigung des Kontexts kann keine korrekte Bewertung erfolgen, da gar nicht klar ist, in welcher Situation sich der Anwender befindet'.

Fragebogen zu Persönlichkeit: 'Für mich nicht so wichtig. Die meisten Leute wissen gar nicht, ob sie eher introvertiert oder extrovertiert sind, bzw. wie sie sich da einordnen sollen'.

Verständnis

Rückmeldung kurz & kompakt: 'benötige ich, um schnell und richtig reagieren zu können, vor allem in Live-Meetings.'

Tutorial zur Anwendung 'Ohne Tutorial weiß der Nutzer nicht, wie das Tool angewandt werden soll und es entsteht Ablehnung', 'es muss klar sein, wie das Tool funktioniert', 'es muss eine Erklärung erfolgen, ansonsten erfolgt wahrscheinlich eine Fehl- oder Nichtbenutzung', 'wenn keiner weiß, wie das Tool verwendet wird, führt das nur zu Ablehnung'.

Video-Tutorial: 'Ich lasse mir gerne etwas per Video erklären, das ist für mich persönlich am komfortabelsten'.

Erklärung der Bewertung: 'auch hier muss deutlich sein, wie die Bewertung erfolgt, sonst kann man diese überhaupt nicht nachvollziehen, gerade, wenn man seine Aussage z.B. selbst nicht als negativ ansieht', 'wird benötigt, um zu validieren, um das System funktioniert'.

Help-Button: 'Das Tutorial sollte jederzeit über den Help-Button aufrufbar sein, wenn man nicht weiter weiß', 'das Tutorial sollte nur über den Button aufrufbar sein, ich mag es gar nicht, wenn ich eine Anwendung starte und mir direkt Tutorials aufgedrängt werden'.

Transparenz: 'Wenn das Tool nicht transparent ist, dann führt das zur Ablehnung durch die Benutzer.'

detaillierte Rückmeldung: 'das ist besonders wichtig für Benutzer, um zu verstehen, was gemacht und bewertet wurde. Eine kurze Rückmeldung ist da nicht immer ausreichend, manchmal muss es einfach ausführlicher erklärt werden'.

Benutzerhandbuch: 'Es muss auch ein ausführliches Benutzerhandbuch geben, in dem ich spezifischere Fragen nachschlagen kann'.

Transparenz UND Erklärung der Bewertung: 'Beides muss vorhanden sein, damit das Tool und die Ergebnisse nachvollziehbar und erklärbar sind'.

Datenschutz

Löschen von Daten nach Zeitraum: 'Das ist eine gesetzliche Vorgabe und ohne kann das Tool gar nicht legal benutzt werden'.

Löschen von Daten /Anonymisierung: 'Ohne die Implementierung von Datenschutzvorgaben kann das Tool nicht verwendet werden, da gibt es sehr große Probleme mit dem Betriebsrat'.

Datenschutz: 'Für mich in meiner Tätigkeit als Forscher nicht relevant'.

Markierungen nicht öffentlich: 'Sonst hätte man Datenschutz-Probleme

und Arbeitnehmer würden sich beschweren’.

Administration

negative oder positive Begriffe festlegen: ‘Das ist für mich unerlässlich, es muss für mich unbedingt die Möglichkeit geben, z.B. negative Begriffe für mein Unternehmen, Projekt oder Team selbst festzulegen’.

Rollen- und Berechtigungskonzept: ‘Ist notwendig, um überhaupt einen Administrations- und Konfigurationsbereich zu betreiben’.

Filter nach bestimmten Emotionen /Mobbing: ‘Als Projektleiter möchte ich herausfinden, ob jemand z.B. dauerhaft schlechte Laune hat. Außerdem möchte ich Mobbing extrahieren und dem entgegenwirken’.

Logfiles: ‘Logfiles interessieren mich in meiner Rolle nicht, sie sind für mich einfach nicht relevant’, ‘benötige ich zum Debuggen’.

ALLE Konfigurationsmöglichkeiten: ‘„Die Möglichkeit, bestimmte Einstellungen am Tool und der Bewertung vornehmen zu können ist äußerst wichtig für meine weiteren Studien und für gewisse Forschungsfragen.“

Integration

Integration in Software und Kommunikationstools: ‘Um die Benutzbarkeit zu gewährleisten, muss das Tool in übliche Software wie MS Teams, Outlook etc. eingebunden werden können’, die verwendete Software ist in unserem Unternehmen gesetzt, und das Tool müsste dort integriert werden’.

Integration in Software und Kommunikationstools / Hardware-Devices: ‘Das Tool muss integrierbar sein, sonst wäre es nicht effektiv nutzbar’.

Oberfläche

Benutzerfreundlichkeit/Usability: ‘Muss besonders hoch sein, sonst kann man sich innerhalb des Tools nicht zurechtfinden und das Tool zu nutzen würde keinen Spaß bereiten’.

Help-Button: ‘Das Tutorial sollte jederzeit über den Help-Button aufrufbar sein, wenn man nicht weiter weiß’, ‘das Tutorial sollte nur über den Button aufrufbar sein, ich mag es gar nicht, wenn ich eine Anwendung starte und mir direkt Tutorials aufgedrängt werden’.

kein Spam: ‘Wenn durch die Ergebnisse bzw. Benachrichtigungen darüber Spam entsteht, dann ist das Tool für mich nicht benutzbar, denn ich muss auch noch woanders hingucken können’.

Markierungen nicht öffentlich: ‘Sonst hätte man Datenschutz-Probleme und Arbeitnehmer würden sich beschweren’, ‘die Priorisierung als besonders wichtig erfolgt aus meinen Erkenntnissen aus Forschung und Anwendung’.

Unterbrechen der Nachrichten bei Live-Meetings: ‘Das ist gefährlich, denn es könnte als Verletzung der Meinungsfreiheit gewertet werden’.

Dashboard: ‘Brauche ich zur Steuerung, damit nicht alles nur eine

5.3. RQ3: PRIORISIERUNG INNERHALB DER STAKEHOLDERGRUPPEN⁵¹

Momentaufnahme ist', 'Für meine Studien erforderlich'.

Gesprächsanteile anzeigen: 'Zur Steuerung von Meetings wichtig'.

Gesamt-Stimmungsbarometer: 'Zur Steuerung von einem Projekt, Team oder Meeting'.

Stimmung visuell darstellen: 'Das ist für mich wichtig, weil die visuelle Darstellung von Stimmungen ein Teil meiner Forschung ist'.

Teilnehmer geben eigene Stimmung an: 'Wichtig, damit ein Abgleich erfolgen kann'.

5.3.2 Priorisierungen der verschiedenen Teilnehmer

Im Folgenden soll noch einmal explizit für jede Stakeholdergruppe die wichtigsten und die weniger stark priorisierten Anforderungen durchgegangen werden. Weiterhin werden die Anforderungen genannt, die von einem Stakeholder entfernt wurden. Der Unterschied zwischen einer 1 und dem Entfernen einer Anforderungen ist wie folgt: eine Punktzahl von 1 bedeutet, dass der Teilnehmer keinen Wert auf die Anforderung legt, es ihn aber nicht stört, sollte sie implementiert werden. Das Entfernen einer Anforderung ist gleichbedeutend mit einer 0, es bedeutet, dass der Teilnehmer diese Anforderung explizit nicht implementiert haben will und diese ihn unter Umständen sogar stören würde.

Entwickler

Auffällig ist die hohe Priorisierung der Kategorie Datenschutz, die beiden Anforderungen, die zu dieser Kategorie gehören, wurden mit den Punktzahlen 6 und 7 bewertet. In der Kategorie Integration wurde ebenfalls eine Anforderung mit der höchsten Punktzahl bewertet, es handelt sich um die Anforderung, dass ein Stimmungsanalysetool in bestehende Software- und Kommunikationstools integrierbar sein müsste. Weiterhin wurde in der Kategorie Administration die Anforderung, dass es für den Administrationsbereich ein Rollen- und Berechtigungskonzept geben müsste, mit einer 7 bewertet. In der Kategorie Verständnis wurden auffällig viele Anforderungen mit der höchsten Punktzahl bewertet. Es handelt sich um folgende Anforderungen: Es soll ein Tutorial geben, speziell eine Art Readme oder FAQ. Weiterhin soll das Tutorial explizit über einen Help-Button aufrufbar sein. Das Stimmungsanalysetool soll eine hohe Transparenz aufweisen, speziell soll die erfolgte Bewertung erklärt werden. Bezüglich der Kategorie 'Erkennen von Stimmungen' wurde die Anforderung, das Tool müsse effizient sein, mit der höchsten Punktzahl bewertet. In der Kategorie Oberfläche wurde auch eine Anforderung mit einer 7 bewertet, nämlich, dass durch das Tool kein Spam in der Darstellung der Ergebnisse entstehen sollte. Tatsächlich bewertete der Entwickler keine Anforderung mit der niedrigsten Punktzahl, einer 1, aus diesem Grund werden nun die

Anforderungen genannt, die mit der zweitniedrigsten Punktzahl bewertet wurden. In der Kategorie Verständnis wurde eine Anforderung mit einer 2 bewertet. Es handelt sich um die Anforderung, dass negativ bewertete Begriffe markiert werden. Zwei Anforderungen aus der Kategorie 'Erkennen von Stimmungen' wurden ebenfalls mit einer 2 bewertet. Das ist zum einen die Anforderung, des Fragebogens zur Charaktereinschätzung, zum anderen um weitere Erkennungsmöglichkeiten wie z.B. ein Armband um Stress zu messen. In der Kategorie Oberfläche wurden ebenfalls zwei Anforderungen mit einer 2 priorisiert. Zum einen erscheint dem Entwickler die Option, einzelne Sätze oder Aussagen zu ignorieren, als weniger wichtig. Zum anderen legt er wenig Wert auf die Option, dass Teilnehmer am Tagesanfang und Tagesende ihre eigene Stimmung angeben. Insgesamt wurden vom Entwickler sechs Anforderungen entfernt, die er in einem Stimmungsanalysetool nicht implementiert haben möchte. Zwei davon stammen aus der Kategorie 'Erkennen von Stimmungen'. Es handelt sich um die Anforderungen 'Bewertung nach verschiedenen Typen' und 'Rechtschreibung, Grammatik und Klarheit prüfen'. Die restlichen vier entfernten Anforderungen stammen aus der Kategorie Oberfläche. Die betreffenden Anforderungen lauten: 'Gesprächsanteile anzeigen, evtl. warnen', 'Anteil von Fachbegriffen anzeigen', 'Sprachgeschwindigkeit anzeigen' und 'Infos *schön* aufbereiten'.

Forscher

Aus der Kategorie Integration wurde eine Anforderung mit der höchsten Punktzahl bewertet. Es handelt sich um die Anforderung, dass das Stimmungsanalysetool in bestehende Software- und Kommunikationstools integrierbar sein sollte. Weiterhin priorisierte der Forscher einige Anforderungen aus der Kategorie Administration mit einer 7. Die betreffenden Anforderungen sind folgende: Bewertungsoptionen wie Sarkasmus etc. sind ein- und ausschaltbar, es existieren technische Logfiles, Schwellwerte sind für Bewertungsinhalte wie Sarkasmus etc. konfigurierbar, es kann wahlweise explizit auf negative oder positive Formulierungen analysiert werden und es existiert eine Art Filter, mit dem Aussagen auf bestimmte Emotionen wie Wut oder Angst analysiert werden können, aber auch die Möglichkeit besteht, Mobbing gezielt zu erkennen. Zwei Kategorien aus der Kategorie Verständnis, zum einen die Transparenz des Tools als auch die Erklärung der Bewertung, wurden ebenfalls mit einer 7 bewertet. In der Kategorie Oberfläche wurde abschließend noch an vier weitere Anforderungen die Höchstpunktzahl vergeben. Explizit sind dies die Anforderungen 'Dashboard mit Gesamtauswertung', 'Teilnehmer geben eigene Stimmung an', 'Stimmung visuell darstellen' und 'Markierungen nicht öffentlich'. Mit der niedrigsten Punktzahl wurden die beiden Anforderungen aus der Kategorie Datenschutz bewertet. Insgesamt wurden acht Anforderungen entfernt. In der Kategorie Verständnis wurde

5.3. RQ3: PRIORISIERUNG INNERHALB DER STAKEHOLDERGRUPPEN⁵³

die Anforderung, dass zur negativ bewerteten Wörtern eine positivere Alternative angezeigt wurde, als nicht relevant erachtet. 'Bewertung nach verschiedenen Typen' und 'Rechtschreibung, Grammatik etc. und Klarheit prüfen' sind die beiden Anforderungen aus der Kategorie 'Erkennen von Stimmungen', die herausgenommen wurden. Die restlichen fünf entfernten Anforderungen stammen aus der Kategorie Oberfläche. Es handelt sich um die 'Option, zu ignorieren', die 'Anzeige von Gesprächsanteilen' verbunden mit der 'Anzeige von Fachbegriffen', der Anforderung 'Sprachgeschwindigkeit anzeigen' und die 'Unterbrechung von Nachrichten' im Fall einer negativen Stimmungsausartung.

Management

Insgesamt wurden vom Management vier Anforderungen mit der höchsten Punktzahl bewertet. Eine davon stammt aus der Kategorie Administration - es soll die Möglichkeit geben, positiv und negativ zu bewertene Wörter explizit festzulegen. Zwei weitere Anforderungen gehören zur Kategorie Verständnis. Auch der Manager bewertete hier die Existenz eines Tutorials mit der Punktzahl 7. Außerdem würde er dabei ein Video-Tutorial bevorzugen. Die letzte mit einer 7 priorisierte Anforderung stammt aus der Kategorie 'Erkennen von Stimmungen': es handelt sich um die Berücksichtigung des Kontexts bei der Bewertung. Der Manager ist der einzige Teilnehmer, der sich dazu entschied, keine Anforderung zu entfernen, die ihm präsentiert wurde. Auch wurde keine Anforderung mit der niedrigsten Punktzahl bewertet, daher werden im Folgenden die Anforderungen genannt, die mit der zweitniedrigsten Punktzahl bewertet wurden. Einige von ihnen stammen aus der Kategorie Administration. Zum einen handelt es sich um die grafische Administrationsoberfläche, die Möglichkeit, gezielt nur auf negative oder positive Formulierungen zu analysieren, sowie der Filter nach bestimmten Emotionen oder Mobbing. Aus der Kategorie 'Erkennen von Stimmungen' wurde nicht als wichtig erachtet, dass das Tool dauerhaft läuft, dass es effizient ist und dass eine Prüfung auf Rechtschreibung, Grammatik und Klarheit des Geschriebenen stattfindet. Weiterhin wurden vier Anforderungen aus der Kategorie Oberfläche mit einer 2 bewertet. Es handelt sich um die Anforderungen, dass Markierungen, die aus der Stimmungsbewertung stammen, nicht öffentlich sein sollen, dass Informationen *schön* dargestellt werden sollen, dass bei Meetings keine Störung bzw. Unterbrechung erfolgt, und auch die andere Ausprägung dieser Anforderung, dass eine Störung erfolgen soll.

Projektleiter

Der Projektleiter bewertete beide Anforderungen aus der Kategorie Datenschutz mit der höchsten Punktzahl. Auch ist dies bei den Anforderungen 'Integration in Software- und Kommunikationstools' und 'Hardware-Devices für Spracheingabe' aus der Kategorie Integration

der Fall. Aus der Kategorie Administration wurde die Anforderung eines Filters nach Emotionen oder Mobbing ebenfalls mit einer 7 bewertet. Außerdem wurden einige Anforderungen aus der Kategorie Verständnis mit der Höchstpunktzahl versehen. Dabei handelt es sich um die Anforderung des Tutorials, explizit eines Video-Tutorials. Dieses soll über den Help-Button aufrufbar sein. Auch wurden die Transparenz und die Erklärung der Bewertung mit einer 7 priorisiert. Die letzte Anforderung, die mit der Höchstpunktzahl bewertet wurde, ist die, dass es die Möglichkeit einer detaillierteren Rückmeldung gibt. Fünf Anforderungen wurden mit der niedrigsten Punktzahl versehen. Davon stammen zwei aus der Kategorie Administration, hier wurden die Existenz von Logfiles, und auch von technischen Logfiles, als nicht relevant erachtet. Bei der Kategorie Verständnis wurde das Benutzerhandbuch mit einer 1 bewertet. Der Fragebogen zur Persönlichkeit aus der Kategorie 'Erkennen von Stimmungen' wurde ebenfalls mit der niedrigsten Punktzahl versehen. Das trifft auch auf die Anforderung, dass bei einer negativen Stimmungslage Meetings unterbrochen werden sollten, aus der Kategorie Oberfläche, zu. Auch der Projektleiter entfernte einige Anforderungen aus dem Anforderungskatalog. Aus der Kategorie Administration wurde die Anforderung, nach der explizit auf negative oder positive Formulierungen analysiert werden kann, herausgenommen. Auch wurde die Anforderung der alternativen Anzeige zu negativ bewerteten Begriffen aus der Kategorie Verständnis entfernt. Aus der Kategorie 'Erkennen von Stimmungen' wurden vier Anforderungen gelöscht, diese sind: 'läuft dauerhaft', 'Bewertung nach verschiedenen Typen', 'Floskeln und Anrede überprüfen' und 'Rechtschreibung, Grammatik und Klarheit prüfen'. Ebenfalls vier Anforderungen wurden aus der Kategorie Oberfläche entfernt. Es handelt sich um folgende Anforderungen: die Option, bestimmte Sätze oder Aussagen explizit zu ignorieren, die Anzeige der Sprachgeschwindigkeit, die Anforderung, dass Infos *schön* aufbereitet werden sollen sowie die Anforderung, dass das Tool keinen Spam produzieren sollte.

Requirements Engineer

Vom Requirements Engineer wurden insgesamt fünf Anforderungen mit der Höchstpunktzahl bewertet. Darunter fällt die Anforderung, dass es die Möglichkeit einer detaillierteren Rückmeldung geben soll aus der Kategorie Verständnis. Drei weitere Anforderungen stammen aus der Kategorie 'Erkennen von Stimmungen'. Es handelt sich um die Anforderungen 'Empfänger muss berücksichtigt werden', 'Bewertung nach verschiedenen Typen', sowie 'Floskeln und Anrede überprüfen'. Die letzte Anforderung, die mit einer 7 bewertet wurde, ist die, dass ein Stimmungsanalysetool Benutzerfreundlichkeit bzw. Usability aufweisen sollte, und stammt aus der Kategorie Oberfläche. Insgesamt wurden sechs Anforderungen entfernt. Dabei handelt es sich vor allem um Anforderungen aus der

5.3. RQ3: PRIORISIERUNG INNERHALB DER STAKEHOLDERGRUPPEN⁵⁵

Kategorie Oberfläche. Hier ist nicht gewünscht, dass das Tool bei sehr negativer Stimmungslage die Kommunikation in Meetings unterbricht, dass verschiedene Abstufungen bei positiver oder negativer Bewertung angegeben werden, dass Teilnehmer jeweils am Tagesanfang und Tagesende ihre eigene Stimmung angeben und dass ein Stimmungs-EKG in Meetings angezeigt wird. Auch sei es nicht nötig, Infos 'schön' aufzubereiten. Eine weitere Anforderung aus der Kategorie Verständnis wurde herausgenommen. Es handelt sich um die Anforderung, nach welcher es möglich sein soll, auf Wunsch eine detailliertere Rückmeldung zur Bewertung zu erhalten. Vom Requirements Engineer wurde keine Anforderung mit der niedrigsten Punktzahl bewertet, aber viele Anforderungen mit einer 2, also der zweitniedrigsten Punktzahl. Aus dem Bereich Administration handelt es sich um die Anforderungen dass (technische) Logfiles existieren sollen und dass ein Rollen- und Berechtigungskonzept existieren sollte. In der Kategorie Verständnis sei es nicht so wichtig, dass negativ bewertete Begriffe markiert werden. Eine weitere niedrig priorisierte Anforderung stammt aus der Kategorie 'Erkennen von Stimmungen'. Hier wurde die Anforderung, das Tool müsse effizient sein, mit einer 2 bewertet. Die letzte niedrig priorisierte Anforderung ist die, dass ein Gesamtstimmungsbarometer angezeigt werden sollte und stammt aus der Kategorie Oberfläche.

Qualitätssicherung

Der Teilnehmer aus der Stakeholdergruppe Qualitätssicherung bewertete insgesamt vier Anforderungen mit der höchsten Punktzahl. Dabei handelt es sich unter anderem um die Datenschutz-Anforderung, dass Ergebnisse nur anonymisiert dargestellt werden und nicht einem Anwender zugeordnet werden können. Auch sei es zwingend erforderlich, dass es, im Bezug auf die Integration, möglich ist, dass Stimmungsanalysetools mit bestehenden Software- und Kommunikationstool kompatibel sind. Die übrigen zwei Anforderungen stammen aus der Kategorie Verständnis. Ein Tutorial zur Anwendung sollte zwingend vorhanden sein, und auf jeden Fall sollte diesbezüglich ein Benutzerhandbuch für spezifische Fragen existieren. Neun Anforderungen wurden aus der Baumstruktur entfernt. Die meisten davon stammen aus der Kategorie Oberfläche. Es handelt sich um die Anforderungen 'Hang anzeigen', 'Teilnehmer geben eigene Stimmung an', 'Gruppierungen der Stimmungen nach Mitgliedern/ Teilnehmern', 'Infos schön aufbereiten', 'bei negativer Ausartung: Unterbrechung mit Warnung'. Drei weitere entfernte Anforderungen gehören zur Kategorie 'Erkennen von Stimmungen'. Hier sei es nicht gewünscht, dass Rechtschreibung und Grammatik überprüft werden und dass das Tool über den Benutzer lernt. Die letzte Anforderung stammt aus der Kategorie Administration. Es handelt sich um die Anforderung, nach der es dem Benutzer möglich sein soll, Aussagen explizit nur aus positive oder negative Formulierungen hin zu untersuchen. Auch der Teilnehmer aus der Qualitätssicherung vergab

an keine Anforderung die niedrigste Punktzahl. Es wurde lediglich eine Anforderung mit der zeitniedrigsten Punktezahl, einer 2, bewertet. Hierbei handelt es sich um die Anforderung, dass der Anteil von Fachbegriffen angezeigt werden sollte.

5.4 RQ4: Abgleich mit existierenden Stimmungsanalysetools

Im Folgenden sollen Stimmungsanalysetools, die in Kapitel 2 vorgestellt wurden, daraufhin geprüft werden, welche in dieser Studie ermittelten Anforderungen sie bereits erfüllen und welche nicht. Da im Rahmen dieser Arbeit nicht alle existierenden Stimmungsanalysetools untersucht und verglichen werden können, wurden hierfür die Tools ausgewählt, die nach der Literaturrecherche als besonders relevant erachtet wurden.

5.4.1 SentiStrength

Das Stimmungsanalysetool SentiStrength kann kostenlos für akademische Zwecke heruntergeladen werden. Alternativ kann auch eine Lizenz für kommerziellen Gebrauch erworben werden. Die Anwendung kann heruntergeladen und anschließend mit einem Doppelklick gestartet werden. Die Startansicht ist in Abbildung 5.8 dargestellt. Eine beispielhafte Ausgabe bzw. Auswertung kann Abbildung 5.9 entnommen werden. Die Ausgabe

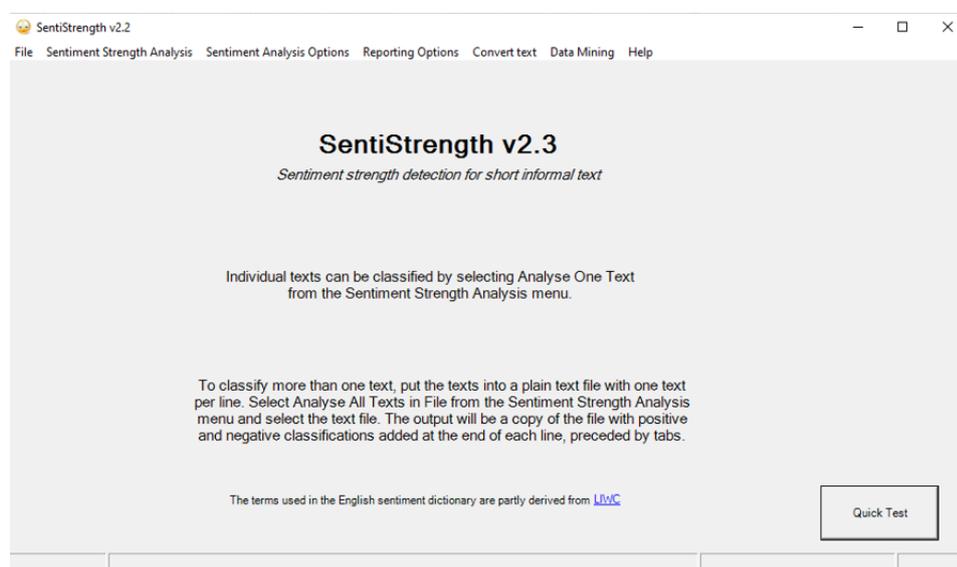


Abbildung 5.8: Startansicht SentiStrength

5.4. RQ4: ABGLEICH MIT EXISTIERENDEN STIMMUNGSANALYSETOOLS57

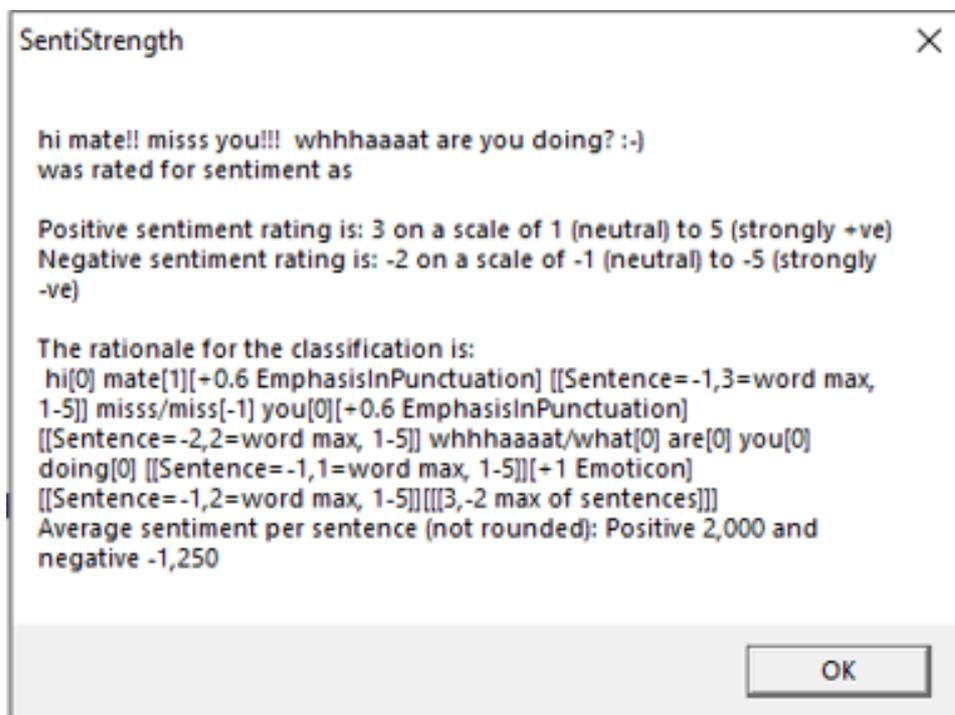


Abbildung 5.9: Auswertung SentiStrength

kann erzeugt werden, wenn in der Startansicht auf 'Quick Test' geklickt wird. Der verwendete Text wird standardmäßig vorgeschlagen. Zu erkennen ist, dass die Auswertung visuell nicht angenehm dargestellt wird. Wörter werden nicht farblich hervorgehoben und die detaillierte Auswertung im dritten Absatz ist sehr unübersichtlich dargestellt und dadurch schwierig nachvollziehbar. Damit wird dargestellt, wie die Bewertung erfolgt, also sind die Anforderungen diesbezüglich also zumindest teilweise erfüllt. Allerdings erfolgt diese Darstellung so wenig visuell ansprechend, dass es möglich sein könnte, dass Nutzer diese Erklärung nicht lesen oder verstehen, selbst wenn sie gerne eine Erklärung hätten.

SentiStrength-SE

Die Installation ist relativ einfach, da es sich um ein Executable Jar File handelt, welches mit Doppelklick gestartet werden kann. Allerdings ist es nicht möglich, SentiStrength-SE in andere Software- und Kommunikations-tool zu integrieren. Auch ist es nicht möglich, Meetings in Echtzeit zu analysieren, und somit besteht auch keine Schnittstelle zu Hardware, wie z.B. zu Mikrofonen. Auch alle anderen Anforderungen, die im Zusammenhang mit einer Echtzeitanalyse stehen, werden somit nicht erfüllt.

Beim Start bietet sich folgende Ansicht:

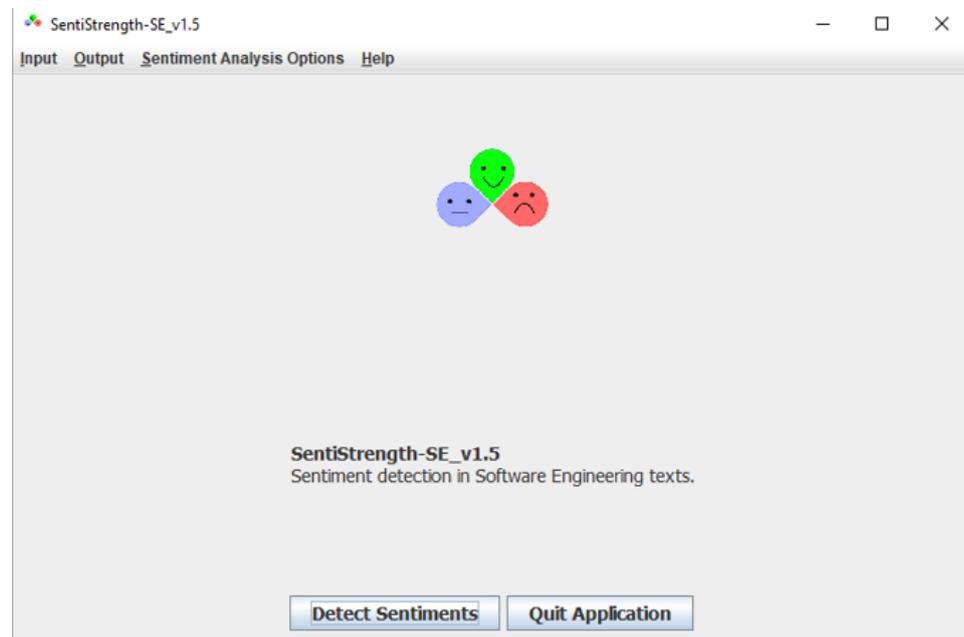


Abbildung 5.10: Startansicht SentiStrengthSE

Die beiden Buttons 'Detect Sentiments' und 'Quit Application' fallen

5.4. RQ4: ABGLEICH MIT EXISTIERENDEN STIMMUNGSANALYSETOOLS 59

sofort ins Auge. Jedoch wird nicht sofort klar, wie ein Dokument für die Stimmungsanalyse eingelesen werden kann. Im Menü existiert eine Help-Option. Es wird beschrieben, wie eine Datei eingelesen werden kann. Demnach muss unter 'Input' ein Dokument ausgewählt werden. Allerdings stellt sich beim Lesen der Anleitung heraus, dass dieses zwei Reihen enthalten muss, eine Zeilennummer und der Inhalt der Zeile, getrennt durch ein Tab. Hier lässt sich bereits feststellen, dass das Tool die Anforderung der Benutzerfreundlichkeit bzw. Usability nicht erfüllt. Der Nutzer muss sich vor Verwendung des Tools bereits die Mühe machen, die Datei wie beschrieben zu formatieren, was im Arbeitsalltag hinderlich und zeitraubend ist. Die Anleitung wurde befolgt und es wurde auch die mitgelieferte Beispiel-Datei verwendet. Trotz der Eingaben aus der Help-Seite gelang es nicht, Output zu generieren, sondern es erscheint lediglich die Fehlermeldung 'Please check your input and configuration files'. Weitere Hilfemöglichkeiten, wie beispielsweise ein Video oder eine Schritt-für-Schritt-Anleitung werden nicht angeboten. Zur Kategorie Administration lässt sich folgendes sagen: Es existieren einige Optionen, die sich auf die Bewertung von Stimmungen beziehen, wie beispielsweise das Aufteilen von Wörtern bei Apostrophen. Allerdings werden die in der Studie genannten Administrationsoptionen nicht abgedeckt bzw. sind nicht vorhanden. Weiterhin kann jeder Nutzer die Optionen selbst konfigurieren, d.h. es ist kein Rollen- und Berechtigungskonzept vorhanden. Viele Anforderungen aus der Kategorie Erkennen von Stimmungen werden ebenfalls nicht erfüllt. So werden weitere Erkennungsmöglichkeiten wie Gesichtsausdrücke, Stress, Kontext, etc. nicht berücksichtigt. Auch findet keine Rechtschreib- und Grammatikprüfung statt. Der Empfänger der Nachricht wird bei der Analyse nicht beachtet. Auch findet sich kein Hinweis darauf, dass Umgangssprache bzw. Dialekte berücksichtigt werden. Dazu muss gesagt werden, dass das SentiStrength-SE bisher nur für die englische Sprache ausgelegt ist und deutsche Sprache ohnehin ohne eine Erweiterung nicht analysiert werden kann. Auch die Oberflächen-Anforderungen werden weitestgehend nicht erfüllt. So existiert kein Dashboard und somit wird diese Anforderung sowie die dazugehörigen Anforderungen nicht erfüllt. Auch existiert keine visuelle Darstellung der Stimmung in Form eines Stimmungs-EKGs, eines Barometers oder farblich markierten Begriffen. Wichtig ist außerdem, dass in den Bedingungen von SentiStrength-SE steht, dass die Verwendung des Tool nur für akademische Zwecke gestattet ist. Das schließt ohnehin sämtliche Stakeholdergruppen, mit Ausnahme des Forschers, aus, sodass das Tool in der Industrie bisher gar nicht verwendet werden kann. Auch eine Lizenz für kommerziellen Gebrauch, sowie bei SentiStrength, kann bisher nicht erworben werden.

5.4.2 EmoTxt

Die Installation von EmoText ist nicht einfach und erfordert Wissen im Umgang mit der Kommandozeile. Zunächst muss Python 2.7.X auf dem Gerät installiert sein, was nicht mehr der aktuellen Version entspricht, die zum Zeitpunkt dieser Arbeit 3.11.2 ist. Außerdem müssen verschiedene Bibliotheken installiert werden, nämlich 'nlTK', 'numpy', 'scikit-learn', 'scipy', 'pattern'. Das Git-Repository von EmoTxt muss geclont werden, anschließend müssen eine Reihe von Kommandozeilen-Befehlen ausgeführt werden. Hier bietet sich viel Potenzial für Fehler und Frustration beim Nutzer- auch in dieser Arbeit kam es zu einer Fehlermeldung, sodass das Stimmungsanalysetool nicht installiert werden konnte.

5.4.3 SEnti-Analyzer

Für die Installation des SEnti-Analyzers sind ebenfalls Kenntnisse über die Verwendung der Kommandozeile nötig. Auch existiert keine grafische Benutzeroberfläche, das Tool muss auch über die Kommandozeile gestartet werden. Bei der Installation muss zunächst eine große Anzahl von Libraries installiert werden. Hier hat sich bei der Installation das Problem ergeben, dass mehrere Libraries nicht mehr auf die aktuelle Python-Version ausgelegt bzw. nicht mehr mit ihr kompatibel sind. Da davon auszugehen ist, dass die Verwendung der aktuellsten Python-Version von Anwendern gewünscht ist, kann das Tool zur Zeit nicht verwendet werden.

5.4.4 Zusammenfassung des Vergleichs

Zusammengefasst lässt sich sagen, dass nur wenige Anforderungen, die in der Studie gesammelt wurden, bereits in existierenden Stimmungsanalysetools implementiert sind. Der Großteil der Anforderungen, die die Teilnehmer stellen, werden nicht erfüllt. Vor allem betrifft dies die Punkte Integration, Datenschutz und Administration. Einige Tools sind auch lediglich über die Kommandozeile bedienbar und verfügen somit über keine grafische Benutzeroberfläche. Die Erkennungsmöglichkeiten zur Stimmungsanalyse sind beschränkt und erfassen, bis auf den SEnti-Analyser, lediglich textbasiert und berücksichtigen den Teilnehmern wichtige Aspekte wie Umgangssprache, Kontext, Tonalität etc. nicht.

Kapitel 6

Diskussion

6.1 Interpretation der Ergebnisse

Im Folgenden sollen die Ergebnisse dieser Arbeit diskutiert werden. Dazu gehören die Interpretation der verschiedenen Stakeholdergruppen, die im Rahmen der Literaturrecherche ausfindig gemacht wurden, die Anforderungen, die von den Teilnehmern an dieser Studie definiert wurden, die festgelegte Priorisierung dieser, sowie die Frage, inwieweit existierende Stimmungsanalysetools die Anforderungen der Teilnehmer erfüllen.

6.1.1 Identifizierte Stakeholdergruppen

Die Vielfalt der gefundenen Stakeholdergruppen zeigt, dass Stimmungsanalysetools vielfältige Verwendungsmöglichkeiten bieten und innerhalb eines Softwareprojekts auf unterschiedlichste Weise eingesetzt werden können. Dazu gehören typische Aufgaben eines Projektleiters, wie das Überwachen der Gesamtstimmung innerhalb des Projektteams oder das Identifizieren von Sticheleien bzw. Mobbing. Auch das betriebswirtschaftliche Management kann das Tool, z.B. mithilfe eines Dashboards, bei täglichen Aufgaben unterstützen. Qualitätsbeauftragte können das Tool im Rahmen von Code Reviews einsetzen. Weiterhin könnte die Verwendung von Stimmungsanalysetools die Qualität des Codes verbessern, was die Arbeit der Qualitätssicherung vereinfacht. Requirements Engineers können Stimmungsanalysetools verwenden, um neue Anforderungen, beispielsweise aus User-Reviews, zu identifizieren. Außerdem können die Tools benutzt werden, um Anforderungen oder dazugehörige Dokumente möglichst neutral zu formulieren. Entwickler wiederum können von einer verbesserten Stimmung innerhalb des Projektteams profitieren. Wie in Kapitel 5 erläutert, kann dies dazu beitragen, die Konzentration und Arbeitsleistung zu steigern. Für Forscher existieren verschiedene Forschungsfragen in Bezug auf Stimmungsanalysetools. Zum einen Forschung zu Stimmungsanalysetools selbst, als auch Forschungsfragen, die unter Zuhilfenahme von Stimmungsanalysetools

beantwortet werden können. Alle Stakeholdergruppen wiederum können das Tool in ihrer Kommunikation verwenden und diese damit positiver gestalten. Eventuell sind weitere Verwendungsmöglichkeiten zu prüfen, die sich in dieser Arbeit nicht aus der Literatur ergeben haben.

6.1.2 Interpretation von Anforderungen

Im Folgenden sollen die ermittelten Anforderungen diskutiert werden. Es zeigt sich, dass die Anforderungen, die von den Teilnehmern genannt wurden, sehr vielfältig sind und sich auf verschiedene Bereiche beziehen. Zum einen existieren Anforderungen an die Analyse an sich, bzw. daran, welche Aspekte beim Erkennen von Stimmungen berücksichtigt werden sollten. Nicht nur der Text an sich, sondern auch Parameter wie Sarkasmus, Umgangssprache, Kontext und Tonalität sollen bei der Bewertung mit einbezogen werden. Das zeigt, dass es wichtig ist, diese Aspekte zu berücksichtigen, da ansonsten die alltägliche Kommunikation nur unzureichend abgebildet wird. Schließlich besteht Kommunikation nicht nur aus den gewählten Worten selbst, sondern auch aus Gesten, Ton und zugehörigem Adressat. Würde lediglich der Text berücksichtigt werden, dann könnte die Bewertung verzerrt werden, weil nicht alle Parameter allumfassend inkludiert werden. Eine weitere Kategorie, die bereits aus der unveröffentlichten Studie hervorging, ist die **Oberfläche**. Das heißt, dass auch die optische Darstellung des Tools den Teilnehmern wichtig ist. Das bedeutet vor allem, dass erst einmal eine grafische Benutzeroberfläche vorhanden sein muss. Bei einigen Stimmunganalysetools, die in Kapitel 5 in Bezug auf die Anforderungen analysiert wurden, war nur eine Verwendung über die Kommandozeile möglich. Der Start über die Kommandozeile erfordert gleichzeitig Wissen über die Kommandozeile und zugehörige Befehle. Selbst wenn dieses vorhanden ist, kann es zu Fehlern kommen, die eventuell nicht gelöst werden können. Weiterhin ist der Einsatz der Kommandozeile nicht für alle Stakeholder sinnvoll bzw. alltagstauglich und kann nicht von Anwendenden erwartet werden. Es dauert länger, als das Programm einfach per Doppelklick zu starten und erfordert technischen Wissen. Jedes professionelle Programm, welches in der Industrie von einem breiten Anwenderspektrum benutzt wird, verfügt normalerweise über eine grafische Benutzeroberfläche. Das Fehlen dieser würde wahrscheinlich zu Ablehnung führen. Wären Nutzer zur Nutzung gezwungen, sind erhebliche Probleme in der Nutzbarkeit und Effizienz zu erwarten. An die Benutzeroberfläche an sich stellen die Anwender verschiedene Anforderungen. Zum einen soll beispielweise für den Projektleiter ein Dashboard existieren, das eine Übersicht über Teams oder Projekte bietet. Weiterhin soll die Oberfläche die Ergebnisse farblich hervorheben, z.B. negativ bewertete Begriffe rot markieren. Auch soll eine Gesamtübersicht, z.B. in Form eines Barometers oder eines EKGs vorhanden sein. Auch einige zusätzliche Anzeigen, wie

beispielsweise die der Sprachgeschwindigkeit bei Meetings sollen vorhanden sein. In der Gesamterscheinung stellen die Anwender sich die Oberfläche minimalistisch und einfach vor, es soll durch die Oberfläche bzw. die Präsentation der Ergebnisse kein Spam entstehen. Die dritte Kategorie, die aus der unveröffentlichten Studie stammte, ist die Kategorie **Verständnis**. Anforderungen daraus sind beispielsweise die Erklärung der Bewertung, die Anzeige von Alternativen, eine detaillierte Rückmeldung etc. Das zeigt, dass es den potenziellen Anwendern wichtig ist, zu verstehen, wie die Bewertung der eingegebenen Texte oder Sprache stattfindet. Ist dies nicht klar, kann das Fehlen der Erklärung zur Ablehnung des Tools führen und die Bewertung könnte willkürlich erscheinen. Ein weiterer wichtiger Punkt innerhalb dieser Kategorie ist die Existenz eines Tutorials. Die Tatsache, dass Teilnehmer hier verschiedene Arten von Tutorials anführten, wie z.B. Videos, FAQs, Wikis etc. zeigt, dass unterschiedliche Anwender verschiedene Arten von Tutorials bevorzugen und es sinnvoll wäre, das Tutorial in mehreren Formen zu Verfügung zu stellen. Wenn den Nutzern kein Tutorial geboten wird, ist es wahrscheinlich, dass auch hier eine Ablehnung entsteht oder die Teilnehmer das Tool sogar falsch verwenden. Der hohe Mittelwert über alle Anforderungen und Teilnehmer in dieser Kategorie von 5,14 zeigt, wie wichtig den Teilnehmern diese Anforderungen sind. Im Laufe der Studie kamen drei weitere Anforderungskategorien hinzu. Eine Kategorie davon ist die **Integration**. Dazu gehört die Möglichkeit, ein Stimmungsanalysetool in bereits verwendete Software- und Kommunikationstools zu integrieren. Somit ist es für die Teilnehmer wichtig, dass es in ihrem Arbeitsalltag und damit verbundene Software passt. Das hat den Vorteil, dass nicht weitere Programme geöffnet werden müssen, sondern, wenn beispielsweise das Programm Outlook bereits geöffnet ist, das Tool damit auch. Die hohe Bewertung der Anforderung, dass das Tool einfach und per Klick installierbar sein soll, zeigt, dass die Teilnehmer keinen hohen Zeitaufwand für die Installation betreiben wollen und dass die Installation auch für Personen ohne technischen Hintergrund durchführbar sein sollte. Ist diese Anforderung nicht erfüllt, ist zu erwarten, dass die Installation des Tools zu Frustration führt und schließlich abgebrochen wird. Falls das Tool kommerziell erwerbbar ist, ist es denkbar, dass viel Hilfestellung seitens des Anbieters eingefordert werden muss, dass der Kauf reklamiert wird und aufgrund der gemachten Erfahrung zukünftig keine Stimmungsanalysetools mehr eingesetzt werden sollen. Die Kategorie Integration ist somit besonders wichtig für die Akzeptanz eines Stimmungsanalysetools. Die zweite Kategorie, die von den Teilnehmern ergänzt wurde, ist **Datenschutz**. Als Begründung wurde hier angegeben, dass die Einhaltung des Datenschutzes obligatorisch ist und bei Nichtbeachtung rechtliche Konsequenzen drohen könnten. Das zeigt, dass Stimmungsanalysetools ohne die Implementierung von Datenschutzvorgaben in der Industrie überhaupt nicht eingesetzt werden können. Sinnvoll wäre es daher, bei einer Weiter- oder Neuentwicklung

eines Tools diese Anforderungen sehr hoch zu priorisieren bzw. als erstes zu implementieren. Die letzte Kategorie ist **Administration**. Die hohe Anzahl an genannten Anforderungen zur Anpassung und Konfiguration der Stimmungsanalyse zeigt, dass die Teilnehmer ein großes Bedarf haben, das verwendete Stimmungsanalysetool zu analysieren, um es an ihre Bedürfnisse anzupassen. Damit dies überhaupt möglich ist, ist auch die Implementierung eines Rollen- und Berechtigungskonzepts nötig, damit einheitliche Vorgaben für ein Team oder Projekt eingestellt werden können und nicht jede Person einzeln die Bewertungsparameter ändern kann. Zuletzt soll noch darauf hingewiesen werden, dass einige Anforderungen widersprüchlich bzw. nicht oder schwer miteinander vereinbar sind. Beispielsweise betrifft dies die Datenschutz-Anforderungen und Anforderungen, die besagen, dass über den User gelernt oder sogar Persönlichkeitsprofile per Fragebogen angelegt werden sollen. Denn die Methode, Informationen über den Benutzer zu speichern widerspricht den genannten Datenschutzerfordernungen, nach welchen keine personenbezogenen Daten gesammelt werden sollen. Hier müsste evaluiert werden, ob dieser Widerspruch aufgelöst werden könnte. Vielleicht wäre es möglich, dass Nutzer explizit der Verwendung eines Persönlichkeitsfragebogens- oder Profils zustimmen, dies aber auch ablehnen können. Ist dies rechtlich nicht möglich, sind die Datenschutz-Anforderungen zu priorisieren. Zum einen, weil es sich um gesetzliche Vorgaben handelt, zum anderen, weil sie, wie in Kapitel 5 dargestellt, von den Teilnehmern als wichtiger eingestuft wurden.

6.1.3 Interpretation von Priorisierungen

Es lässt sich feststellen, dass die Anforderungen zwischen den verschiedenen Stakeholdergruppen variieren. Im Folgenden sollen die verschiedenen Kategorien noch einmal durchgegangen und auf Unterschiede untersucht werden. In der Kategorie **Datenschutz** existieren deutliche Unterschiede in der Priorisierung. Projektleiter, Qualitätssicherung, Entwickler und Requirements Engineer haben die Anforderungen aus der Kategorie Datenschutz jeweils mit einer 6 oder 7, also sehr hoch bzw. am höchsten bewertet. Wenn die abgegebenen Begründungen berücksichtigt werden, nämlich dass die Beachtung des Datenschutzes gesetzlich vorgegeben ist, lässt sich feststellen, dass bei der Verwendung von Stimmungsanalysetools in der Industrie die Beachtung des Datenschutzes essentiell ist und ohne die Implementierung dieser Anforderungen die Benutzung eines Tools nicht möglich ist. Der Forscher wiederum bewertete die Anforderungen aus dieser Kategorie mit der niedrigsten Punktzahl. Hier kann man vermuten, dass Datenschutz bei der bisherigen Forschung über Stimmungsanalysetools noch nicht übermäßig relevant war. Dies könnte sich jedoch zukünftig ändern, falls die Forschung in diesem Bereich ausgeweitet werden soll. Auch Forscher müssen Datenschutzvorgaben einhalten, bzw. sich Einverständniserklärungen der

Teilnehmer einholen und diese über Abbruchmöglichkeiten laut Datenschutz-Grundverordnung (DSGVO) informieren. Das bedeutet auch, dass sicher gestellt werden muss, dass per Stimmungsanalysetool gesammelte Daten gelöscht werden können. Folgendes lässt sich zur Kategorie **Integration** sagen. Generell wurden die Anforderungen dieser Kategorie mit einem Mittelwert von 5,42 über alle Teilnehmer und Anforderungen hinweg als hoch erachtet. Das zeigt, dass es den Teilnehmer wichtig ist, dass mögliche Stimmungsanalysetools problemlos in den Arbeitsalltag eingefügt werden können, und das kann dadurch erreicht werden, dass sie in bereits verwendete Software integriert werden können. Das hat den Vorteil, dass nicht weitere Programme geöffnet werden müssen, sondern, wenn beispielsweise das Programm Outlook bereits geöffnet ist, das Tool damit auch. Die Anforderungen aus der Kategorie **Administration** wurden im Durchschnitt am höchsten vom Forscher bewertet. Begründet wurde dies damit, dass verschiedene Einstellungsmöglichkeiten für die Forschung sinnvoll sind. Das zeigt also, dass das Tool für die Arbeit des Forscher umso nützlicher ist, je mehr Funktionen bezüglich der Einstellung es bietet, da es dann möglich ist, in mehreren Bereichen zu forschen bzw. weiteren Forschungsfragen nachgegangen werden kann. Besonders relevant für den Projektleiter ist im Bereich Administration die Funktion, die es ermöglicht, explizit auf bestimmte Emotionen wie z.B. Wut oder Angst zu untersuchen oder gezielt nach Mobbing zu filtern. Das lässt sich damit begründen, dass die Position des Projektleiters auch Verantwortung für die Projektteilnehmer beinhaltet und das Tool helfen könnte, Spannungen und negative Entwicklungen der Stimmungen frühzeitig zu erkennen, frühzeitig zu intervenieren und die Stimmung im Projekt zukünftig zu verbessern. Ähnliches gilt auch für das Management, hier wurde es als besonders wichtig empfunden, dass es explizit festlegbare Listen für Begriffe gibt, die entweder immer positiv oder immer negativ bewertet werden sollen und dass es die Möglichkeiten geben sollte, z.B. für Sarkasmus Schwellwerte festzulegen. Im Gegensatz dazu bewertete beispielsweise der Entwickler die meisten Anforderungen aus dieser Kategorie lediglich mit einer 3, was vermutlich daran liegt, dass sie für seine täglichen Aufgaben nicht relevant sind, sondern die Vorgaben wie Blacklists etc. dann voraussichtlich von den Vorgesetzten definiert werden würden. Ähnliches gilt für einige Anforderungen der **Oberfläche**. Hier gab der Projektleiter an, dass er ein Dashboard am höchsten priorisiert, da er für seine Aufgaben eine Übersicht über die Stimmungslage im Team oder Projekt benötigt. Im Gegensatz dazu bewertete der Entwickler diese Anforderung mit einer 3. Viele Anforderungen aus der Kategorie **Verständnis** haben eine vergleichsweise geringe Standardabweichung in der Bewertung, d.h. sie wurden ähnlich bewertet. Grundsätzlich lässt sich sagen, dass das Verständnis über die Verwendung und Bewertung von Stimmungsanalysetools für sämtliche Stakeholdergruppen relevant ist, da das Tool ohne Wissen darüber nicht oder zumindest falsch verwendet werden würde. Nutzer möchten verstehen, wie

eine Bewertung zustande kommt und wie sie das Tool ihren Bedürfnissen entsprechend verwenden können. Insgesamt lässt sich sagen, dass Nutzer jene Anforderungen besonders hoch priorisieren, die sie in ihren täglichen Aufgaben unterstützen könnten. Beispielsweise benötigt ein Projektleiter ein Dashboard, ein Manager verschiedene Administrationsmöglichkeiten, aber ein Entwickler benötigt beides nicht zwangsläufig, weil er es einfach nicht verwenden muss. Dafür legt er Wert auf die Effizienz und die Transparenz. Oft sind die Gründe für die Priorisierung verschieden. Auch wenn mehrere Gruppen eine Anforderung mit der höchsten Punktzahl bewerten, kann dies unterschiedliche Gründe haben. Zum Beispiel legt der Entwickler einen besonderen Wert auf die Implementierung von Datenschutz-Anforderungen, da sie ihn arbeitsrechtlich betreffen und es ihm wichtig ist, dass keine negativen Konsequenzen aus rückverfolgbaren Stimmungsbewertungen für ihn entstehen. Im Gegensatz dazu äußerte der Projektleiter, dass er vor allem rechtliche Konsequenzen vom Betriebsrat fürchte, falls Datenschutz-Anforderungen nicht umgesetzt werden. Überraschend ist, dass von keinem Teilnehmer die Anforderung genannt wurde, dass die Ergebnisse des Tools präzise bzw. korrekt sein sollen. Eine naheliegende Vermutung ist, dass die Teilnehmer diese Anforderung voraussetzten oder dass sie diese als so essentiell erachteten, dass es ihnen nicht einfiel, dass das Tool auch falsche Ergebnisse ausgeben könnte.

6.1.4 Interpretation des Abgleichs mit Stimmungsanalysetools

Beim Abgleich mit existierenden Stimmungsanalysetools lässt sich feststellen, dass vorhandene Stimmungsanalysetools die erhobenen Anforderungen nahezu überhaupt nicht erfüllen. Das könnte daran liegen, dass vorgestellte Stimmungsanalysetools vorwiegend im Rahmen der Forschung entwickelt wurden. Es zeigt sich somit, dass sie im Kontext von Softwareprojekten eher ungeeignet sind. Datenschutz-Anforderungen sind bei keinem der untersuchten Tools implementiert. Nur wenige Tools haben eine grafische Benutzeroberfläche und wenn, dann ist sie nicht benutzerfreundlich. Man könnte den potenziellen Anwendern bzw. Käufern von Stimmungsanalysetools nicht zumuten, das Tool über die Kommandozeile zu starten. Anwender sind an grafische, und auch benutzerfreundliche Oberflächen gewöhnt. Ist diese nicht vorhanden, würde dies höchstwahrscheinlich zur Ablehnung des Tools führen. Einige Tools konnten überhaupt nicht installiert werden, da es zu Kompatibilitätsproblemen zwischen den verwendeten Libraries und der aktuellen Python-Version kommt. Auch hier gilt wieder, dass das Tool einfach und per Klick, oder auch mit einem Installationsassistenten installierbar sein muss. Es kann nicht vorausgesetzt werden, dass Anwender über Kenntnisse über die Verwendung der Kommandozeile verfügen. Weiterhin möchten sich Kunden oder Benutzer nicht mit der Installation von

Libraries und damit verbundenen Kompatibilitätsproblemen beschäftigen. Die Stimmungsanalysetools SentiStrength bzw. SentiStrength-SE bieten einige Einstellungsmöglichkeiten bezüglich der Analyse von Stimmungen, jedoch ist keine der erhobenen Anforderungen aus der Kategorie Administration darunter. Auch fehlt ein Rollen- und Berechtigungskonzept, da davon auszugehen ist, dass es gewünscht wird, dass die Einstellungen im Rahmen eines Projekt oder Teams gleich sind und nicht von jedem einzelnen Nutzer angepasst werden sollen. Bei den meisten Tools ist eine kurze Anleitung vorhanden, entweder als mitgelieferte Datei oder verbunden mit einer Help-Option im Menü. Leider ist diese meistens nicht ausführlich genug und insbesondere bei SentiStrength bzw. SentiStrength-SE wurde die Anleitung befolgt, es konnte jedoch trotzdem kein Output generiert werden, da es, obwohl die Anleitung genau nachgearbeitet wurde, zu Fehlermeldungen kam. Die Anleitung ist die einzige Anforderung aus der Kategorie Verständnis, die implementiert ist. Alle anderen Anforderungen, wie beispielsweise die genaue Erklärung der Bewertung, andere Anleitungsförmlichkeiten wie Erklärvideos oder die Anzeige von Alternativen. Wenn eine grafische Benutzeroberfläche vorhanden ist, dann ist dennoch keine der Anforderungen, die die Teilnehmer dieser Studie an die Oberfläche stellen, implementiert. Es existieren keine farblichen Markierungen, kein Dashboard, keine Gesamtübersicht in Form eines EKGs oder Barometers, etc. Insbesondere sind die vorhandenen Oberflächen überhaupt nicht benutzerfreundlich, beispielweise sind Buttons seltsam platziert und die wichtigsten Funktionen, unter anderem die Möglichkeit, Texte einzulesen, sind im Menü versteckt und beim Start des Programms überhaupt nicht ersichtlich. Auch die Anforderungen aus der Kategorie 'Erkennen von Stimmungen' sind kaum implementiert. Teilweise ist die Erkennung von doppelten Verneinungen möglich, aber Sarkasmus, Dialekte oder Umgangssprache können dennoch nicht erkannt werden. Auch werden Parameter wie der Adressat, Tonalität, Kontext, Stressdaten etc. nicht berücksichtigt. Besonders wichtig ist hier die Tatsache, dass die Echtzeitanalyse in Meetings nicht implementiert wird und die Tools den Fokus auf die Analyse von Texten legen. Allein der Senti-Analyzer verfügt über diese Funktion [18]. Die meisten Teilnehmer sprachen ganz selbstverständlich von der Analyse in Meetings, daher ist davon auszugehen, dass ihnen die Anwendung in diesem Kontext besonders viel nutzen würde und dass sie diese Funktion in einem Tool auch erwarten würden. Zusammenfassend lässt sich sagen, dass es sehr wahrscheinliche erscheint, dass einer der Hauptgründe, weshalb Stimmungsanalysetools in der Industrie nicht verbreitet sind, darin besteht, dass nahezu keine der erhobenen Anforderungen in existierenden Stimmungsanalysetools implementiert sind und Tools den Nutzern aufgrunddessen keinen Mehrwert bieten. Im Gegenteil, die Installation und Verwendung von Stimmungsanalysetools wäre aufgrund fehlender Benutzeroberflächen, Einstellungsmöglichkeiten, Kompatibilitätsproblemen, der Notwendigkeit, Dokumente in bestimmter Weise

zu formatieren und mangelnder Erklärung mit viel Frustration und einem extrem hohen Zeitaufwand verbunden. Bezieht man dann auch noch in Betracht, dass eine Lizenzgebühr erforderlich ist und die Software erworben werden muss, kommt man zu dem Schluss, dass höchstwahrscheinlich kein Unternehmen für die gebotene Software in ihrem jetzigen Zustand bezahlen würde.

6.2 Einschränkungen und Threads to Validity

Diese Arbeit und die durchgeführte Studie unterliegen einigen Einschränkungen und Bedrohungen der Validität. Zu erwähnen ist, dass der Snowballing-Prozess nicht komplett durchgeführt wurde, sondern lediglich, bis eine theoretische Sättigung nach Wolkswinkel et al. [42] erreicht war. Dies bedeutet, dass es möglich ist, dass relevante Literatur nicht gefunden wurde und damit nicht in dieser Arbeit berücksichtigt wurde. Die Durchführung eines kompletten Snowballing-Prozesses wäre im Rahmen dieser Arbeit jedoch nicht möglich gewesen. Weiterhin wurde ergänzend eine Datenbanksuche durchgeführt. Eine weitere mögliche Einschränkung besteht darin, dass der Coding-Prozess, der in Kapitel 4 erläutert und für die Kommentare des dritten Interviews durchgeführt wurde, nur alleine, und nicht wie üblich, durch mehrere Personen ausgearbeitet wurde. Daher ist es möglich, dass andere Personen zu einem anderen Ergebnis bzw. einer anderen Zuordnung gekommen wären. Dadurch, dass diese Masterarbeit jedoch alleine bearbeitet werden musste, bestand dahingehend keine andere Möglichkeit. Ein weiterer Punkt betrifft das Erklärvideo, welches den Teilnehmern vor Beginn des ersten Interviews vorgespielt wurde. Bei der Erstellung des Videos wurde darauf geachtet, möglichst neutral die Funktionsweise und Anwendungsmöglichkeiten von Stimmungsanalysetools darzustellen. Es könnte jedoch sein, dass mit der Beschreibung des Funktionsweise des Tools unbewusst Anforderungen vorgegeben wurden. Dagegen spricht, dass das Video lediglich grundlegende Funktionen erklärt und diese auch von keinem Teilnehmer als Anforderung genannt wurden. Weiterhin war es unerlässlich, den Teilnehmern vorab eine Erklärung zum Thema Stimmungsanalyse bereitzustellen, da der Einsatz von Stimmungsanalysetools nicht weit verbreitet ist, und kein Teilnehmer, außer dem Forscher, vor der Studie über Berührungspunkte mit Stimmungsanalysetools verfügte. Zu bedenken ist auch, dass es möglich sein kann, dass Stakeholder übersehen bzw. nicht identifiziert wurden. Dem entgegenzusetzen ist die große Anzahl an gelesener und einbezogener Literatur zum Thema Stakeholder in Softwareprojekten. Der letzte Punkt besteht darin, dass die Anzahl der Teilnehmer, die an dieser Studie teilnahmen, begrenzt ist. Dadurch könnte die Repräsentativität der Studie eingeschränkt sein. Im Rahmen dieser Arbeit wäre es aufgrund des Arbeitsaufwands nicht möglich gewesen, mehr Teilnehmer je Stakeholdergruppe zu interviewen.

Weiterhin wurde darauf geachtet, dass die Teilnehmer wirklich repräsentativ für die dazugehörige Stakeholdergruppe sind, d.h. sie verfügen alle über mehrjährige Berufserfahrung in ihrer Position.

Kapitel 7

Zusammenfassung und Ausblick

In diesem letzten Kapitel soll noch einmal ein abschließender Überblick über die Arbeit und die wichtigsten Erkenntnisse dieser erfolgen, sowie ein Ausblick auf mögliche künftige Forschung gegeben werden.

7.1 Zusammenfassung

Die Motivation bzw. Problemstellung dieser Arbeit war folgende: Es existieren diverse Stimmungsanalysetools zum Erkennen von Stimmungen, aber es gibt keine Hinweise darauf, dass diese in der Industrie verwendet werden. Daher sollten mit der in dieser Arbeit durchgeführten Studie Anforderungen verschiedener Stakeholdergruppen an Stimmungsanalysetools im Rahmen von Softwareprojekten erhoben werden. Dafür wurden zunächst eine Literaturrecherche nach dem Snowballing-Verfahren von Wohlin [41] durchgeführt, um sich einen Überblick über existierende Stimmungsanalysetools, Anwendungsfälle im Rahmen von Softwareprojekten und mögliche Stakeholder zu verschaffen. Als Stakeholder wurden die Gruppen Entwickler, Projektleiter, betriebswirtschaftliches Management, Requirements Engineer, Qualitätssicherung und Forschung identifiziert. Zur Anforderungserhebung wurde sich für eine iterative Interviewstudie mit zwei Phasen und insgesamt vier Stufen entschieden. Diese Methode wurde ausgewählt, da sie sich, wie in der Arbeit erläutert, besonders gut eignet, falls noch keine oder nur sehr wenige Anforderungen zu einem Themenbereich bekannt sind und durch den Prozess die Kreativität der Teilnehmer angeregt wird. Die Interviews der ersten Stufe wurden durch die Daten aus einer bisher unveröffentlichten Studie ersetzt, in der Teilnehmer ein Stimmungsanalysetool getestet und schließlich Anforderungen definiert haben. Die Anforderungen wurden in die Kategorien Erkennen von Stimmungen, Oberfläche und Verständnis eingeteilt und in eine Baumstruktur übertragen. Als Vorbereitung für die

Interviews wurde ein Video erstellt, was den Teilnehmern die Funktion von Stimmungsanalyse erklärt, da nicht davon ausgegangen werden kann, dass diese bekannt sind. In den Interviews wurde den Teilnehmern schließlich die erstellte Baumstruktur präsentiert und sie wurden gebeten, Anforderungen hinzuzufügen, zu erweitern oder auch zu löschen, falls sie diese in einem Tool nicht implementiert haben wollen. Daraus entstand für jeden Teilnehmer eine neue Baumstruktur, die um die neuen Anforderungen anderer Teilnehmer ergänzt wurde. Im zweiten Interview wurde dieser Prozess wiederholt. In diesen beiden Stufen kamen die drei Kategorien Datenschutz, Administration und Integration hinzu. Festhalten lässt sich, dass Teilnehmer vielfältige Anforderungen an Stimmungsanalysetools stellen. Für die Erkennung und Bewertung von Stimmungen sollen zahlreiche Faktoren (Sarkasmus/Ironie, Umgangssprache, Pulsfrequenz) aber auch Einflussgrößen wie Grammatik, Gesprächsanteile, Kontext etc. berücksichtigt werden. Auch Datenschutz ist ein sehr wichtiges Thema in der Industrie, und die Verwendung des Tools ist an rechtliche Vorgaben gebunden, die unbedingt beachtet werden müssen. Schließlich muss das Stimmungsanalysetool mit einer angenehmen Benutzeroberfläche überzeugen und einfach installierbar und integrierbar sein. Sehr wichtig ist den Teilnehmern darüber hinaus eine Erklärung zur Verwendung und eine klare und transparente Bewertung. Im dritten Interview erfolgte eine Priorisierung der präsentierten Anforderungen unter Zuhilfenahme einer siebenstufigen Likert-Skala. Beachtenswert ist, dass es durchaus Unterschiede zwischen den verschiedenen Stakeholdergruppen gibt, z.B. für den Projektleiter ist ein Dashboard zur Steuerung seines Projekts oder Teams notwendig, ein Entwickler benötigt dies nicht zwangsläufig. Um herauszufinden, inwieweit aktuelle Stimmungsanalysetools die ermittelten Anforderungen abdecken, wurden diverse Tools analysiert. Es hat sich herausgestellt, dass es bei vielen Tools zu Installationsproblemen aufgrund von Kompatibilitätsproblemen zwischen Libraries und aktueller Python-Version kam und die Installation daher nicht abgeschlossen werden konnte. Bei einigen Tools ist die Oberfläche nicht benutzerfreundlich genug, und obwohl die Anleitung befolgt wurde, konnte keine Ausgabe erzeugt werden. Existierende Stimmungsanalysetools decken die vielfältigen Anforderungen der Teilnehmer in keinsten Weise ab und es wird einen extrem hohen Entwicklungsaufwand erfordern, die Anforderungen zu implementieren. Anzunehmen ist daher, dass dies einer der Hauptgründe dafür ist, weswegen Stimmungsanalysetools in der Industrie nicht verwendet werden.

7.2 Ausblick

Es existieren noch einige ungelöste Fragen oder Bereiche, in denen weitergeforcht werden könnte. Zum einen wäre es denkbar, weitere Stakeholdergruppen und Anwendungsfälle zu identifizieren, da diese Arbeit keinen

Anspruch auf Vollständigkeit erhebt. Da auch jeweils nur ein Teilnehmer pro Stakeholdergruppe interviewt wurde und somit die Repräsentativität eingeschränkt ist, wäre es sinnvoll, weitere Repräsentanten dieser Gruppen zu Anforderungen an Stimmungsanalysetools zu befragen. Einige der erhobenen Anforderungen müssen zudem daraufhin geprüft werden, ob sie überhaupt umsetzbar sind, bzw. ob sie mit einem realistischen Arbeitsaufwand umsetzbar sind. In dieser Arbeit wurden die Anforderungen lediglich gesammelt, aber nicht auf den Aspekt der Umsetzbarkeit hin untersucht. Das wäre aber insbesondere von Interesse, wenn eine Implementierung der Anforderungen im Rahmen einer Neu- oder Weiterentwicklung von Stimmungsanalysetools gewünscht ist. Weiterhin müssten die Datenschutz-Anforderungen darauf geprüft werden, welche juristischen Vorgaben tatsächlich existieren und welche Unterschiede zwischen verschiedenen Ländern bestehen, damit gegebenenfalls eine Anpassung daran stattfinden kann.

Anhang A

Baumstrukturen

Hinweis: Für die bessere Lesbarkeit der Baumstrukturen kann auf die digitale Version zurückgegriffen werden.

Erklärung: Grün umrandete Kästen beschreiben die unterschiedlichen Kategorien. In Interviewrunde zwei wurde für die Anforderungen mit unterschiedlichen Farben gearbeitet. Blau sind die 'ursprünglichen' Anforderungen aus der Metanoy-Studie. Gelb markiert sind die Anforderungen, die in Runde 1 von anderen Teilnehmern hinzugefügt wurden. Schwarz sind eigene Anforderungen des Teilnehmers. In Interviewrunde 3 wurden alle Anforderungskästen wieder schwarz eingefärbt, um bei der Priorisierung durch farbliche Markierungen keine Anforderung unbewusst hervorzuheben.

Anmerkung: In Abbildung A.20 wurde die in einer vorherigen Runde entfernte Anforderung 'keine Unterbrechung der Nachrichten, stattdessen Alternative' wieder aufgenommen, da der Projektleiter die Anforderung erneut äußerte. Dies ist in der handschriftlichen Notiz vermerkt.

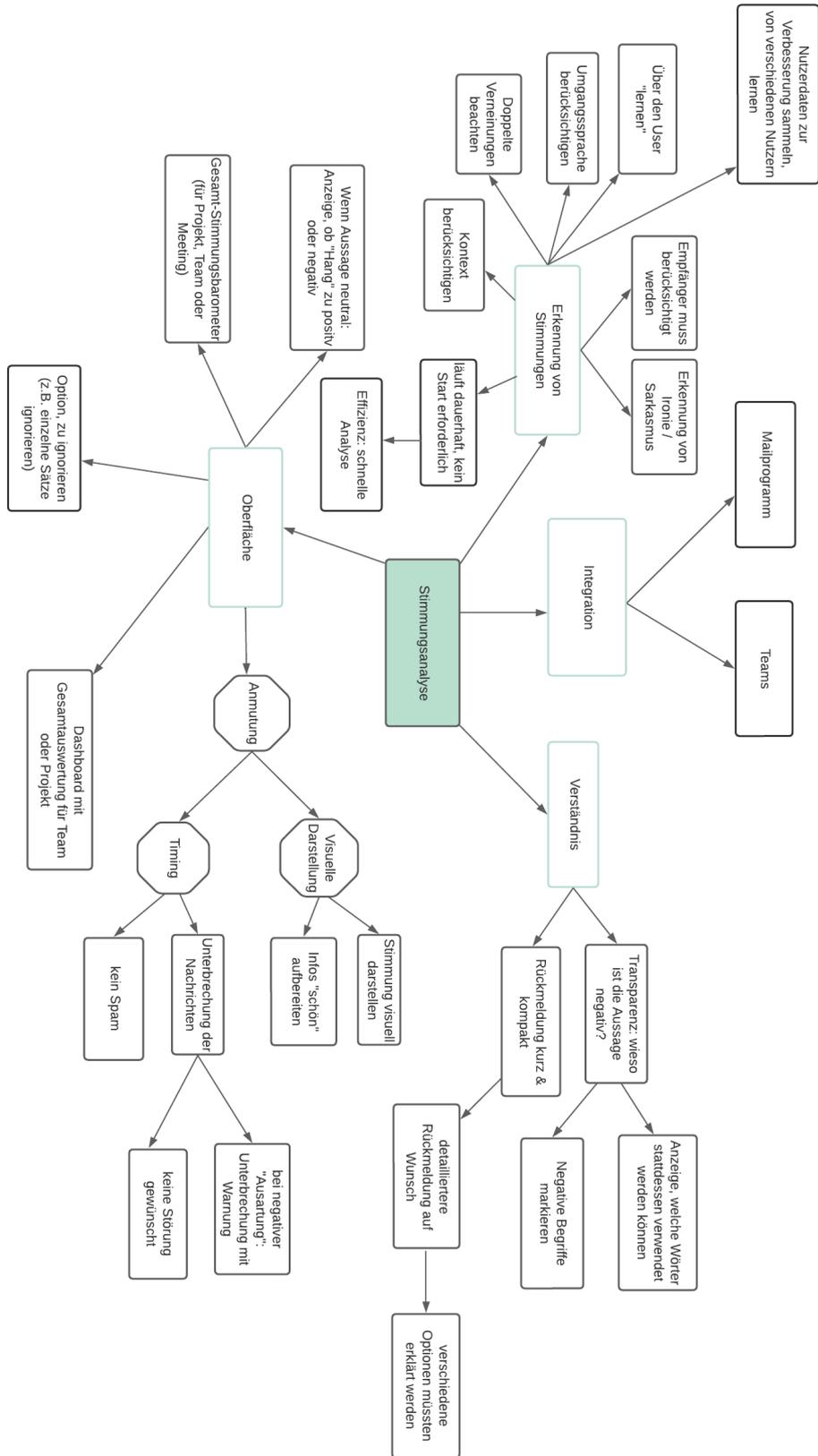


Abbildung A.1: Ergebnisse 1. Interview - Entwickler

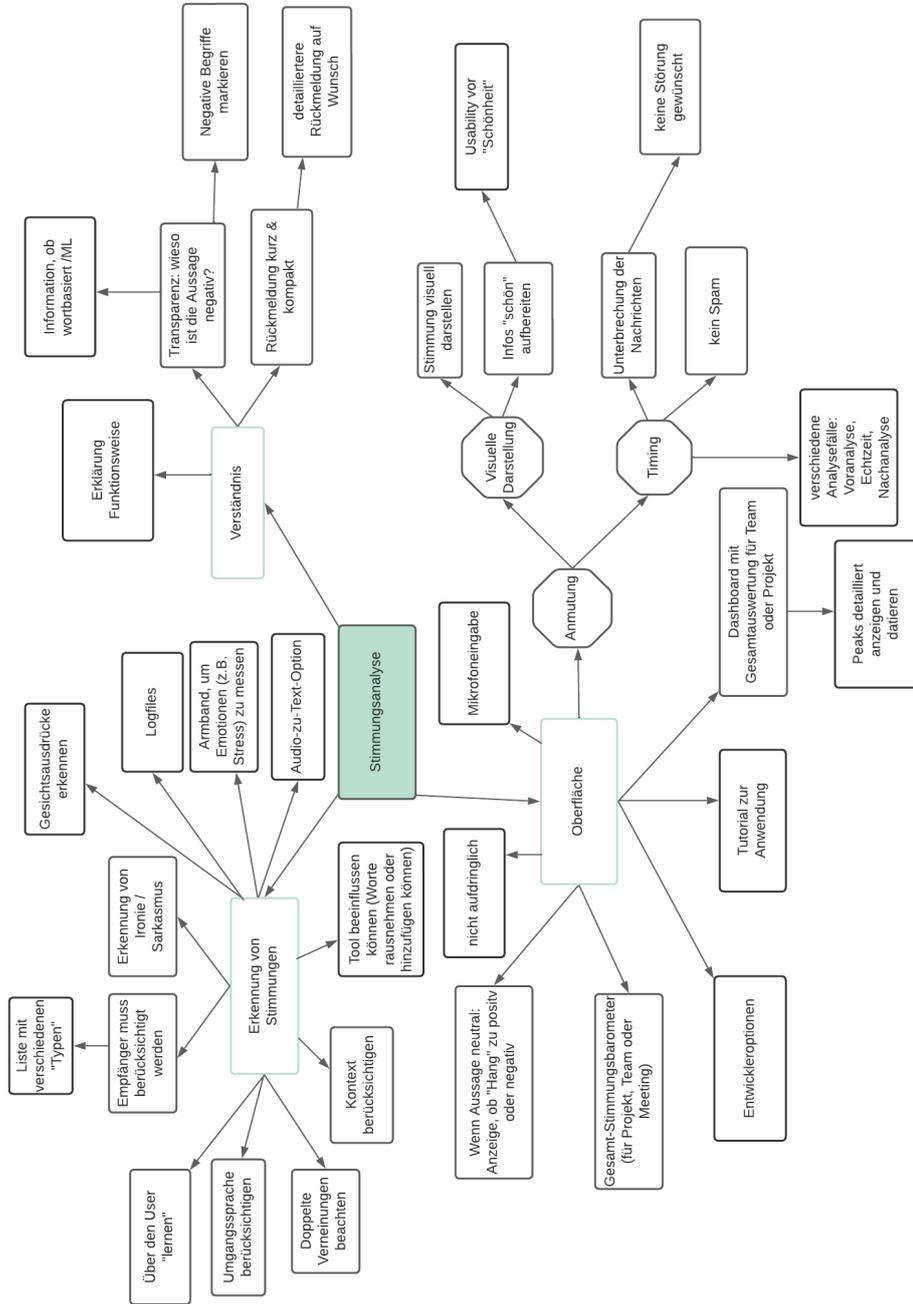


Abbildung A.2: Ergebnisse 1. Interview - Forscher

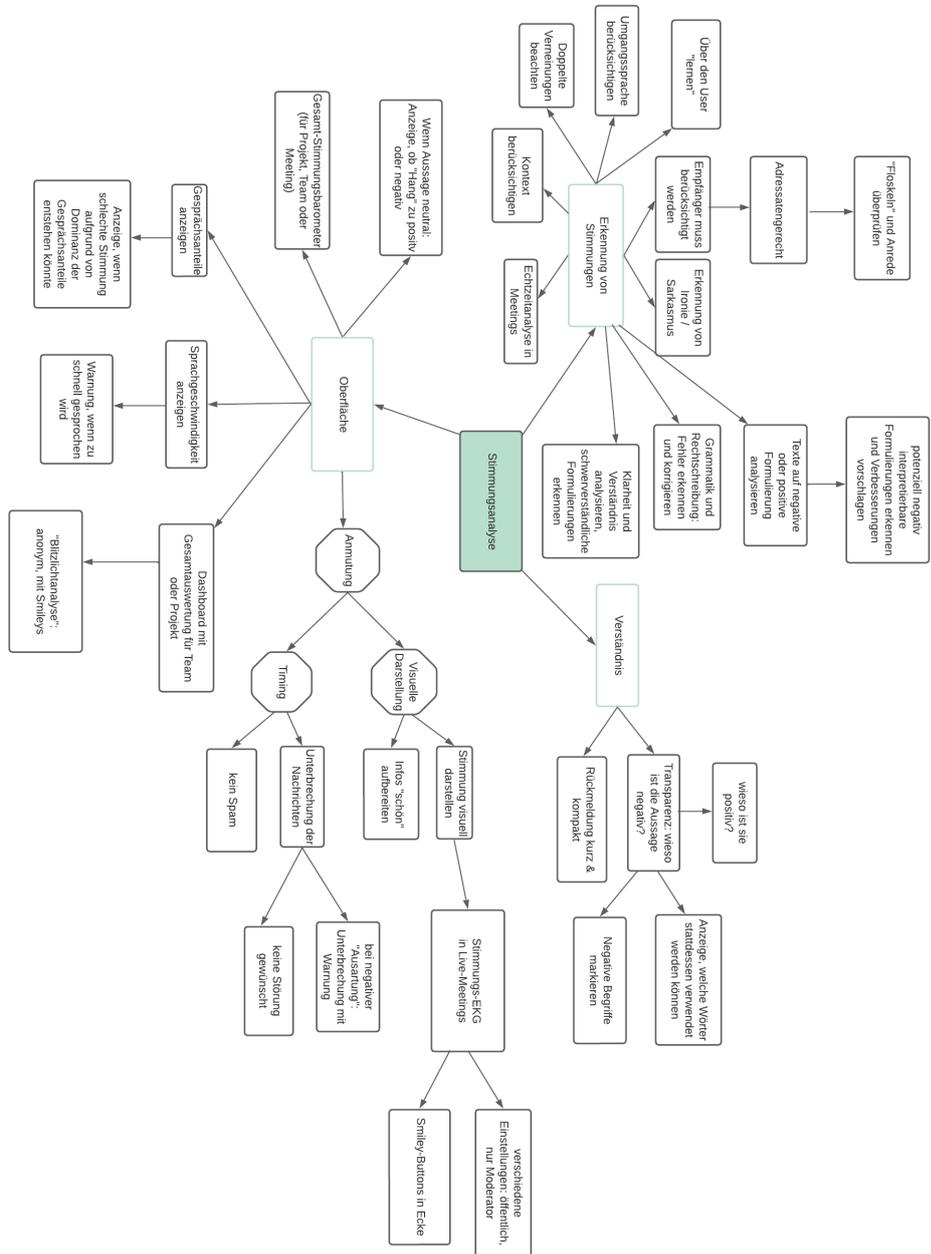


Abbildung A.3: Ergebnisse 1. Interview - Management

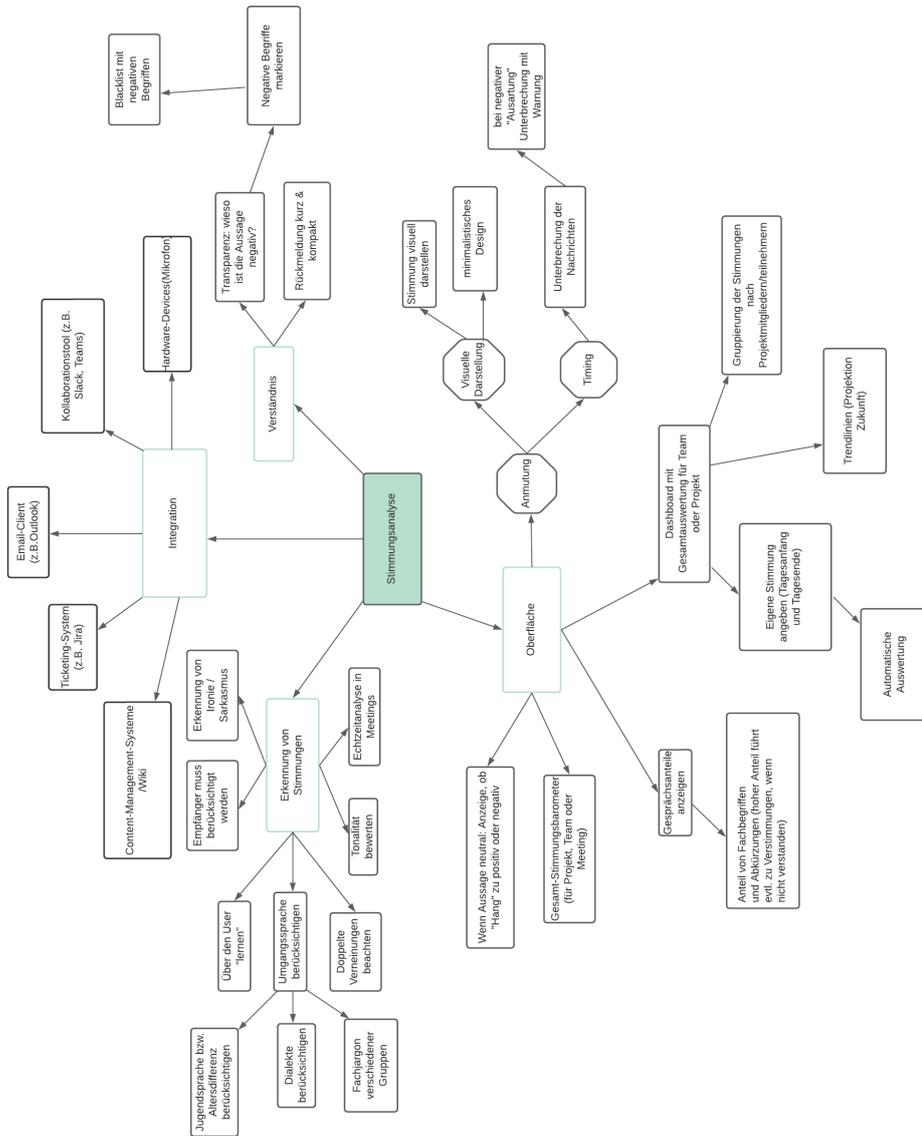


Abbildung A.4: Ergebnisse 1. Interview - Projektleiter

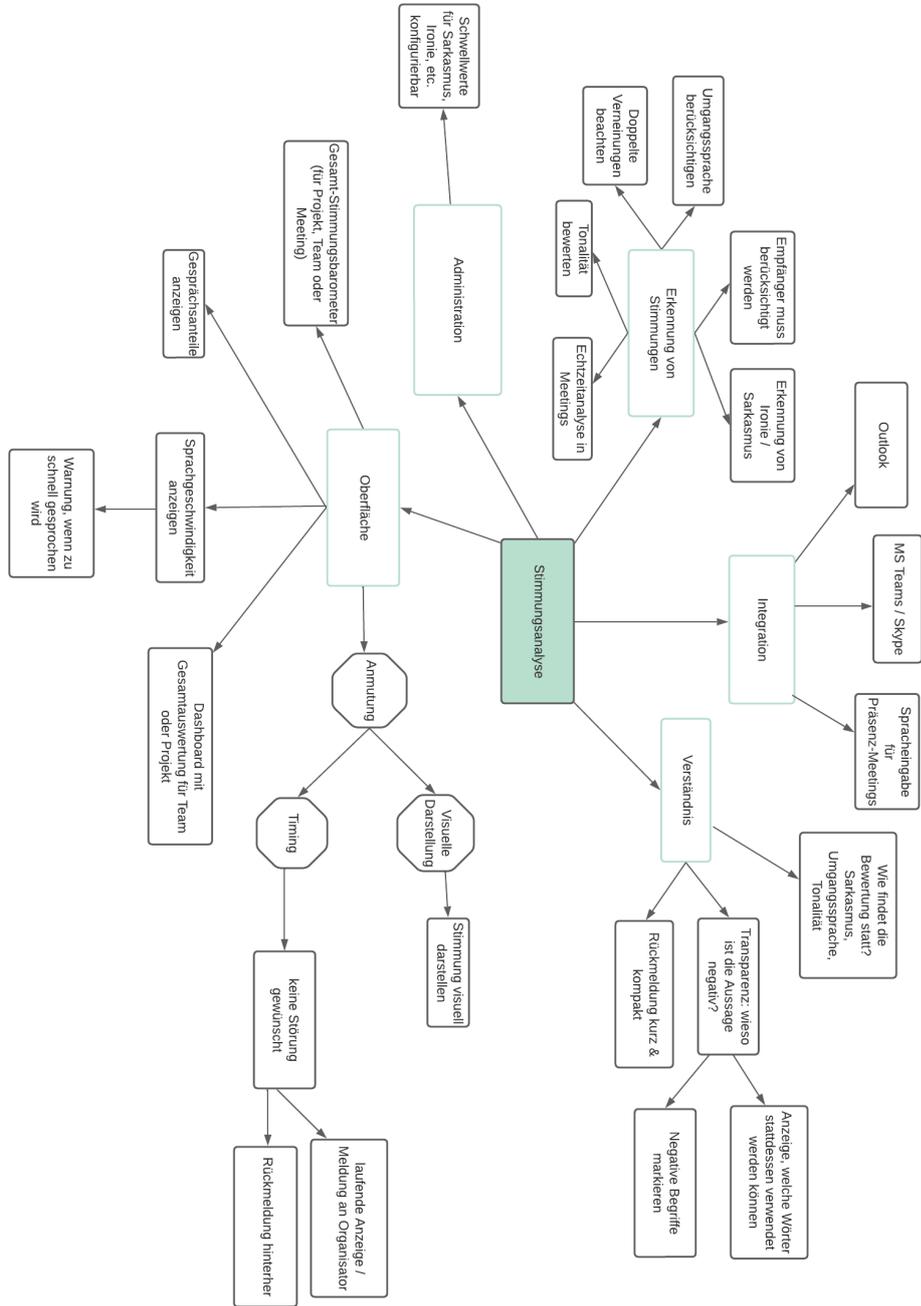


Abbildung A.5: Ergebnisse 1. Interview - Qualitätssicherung

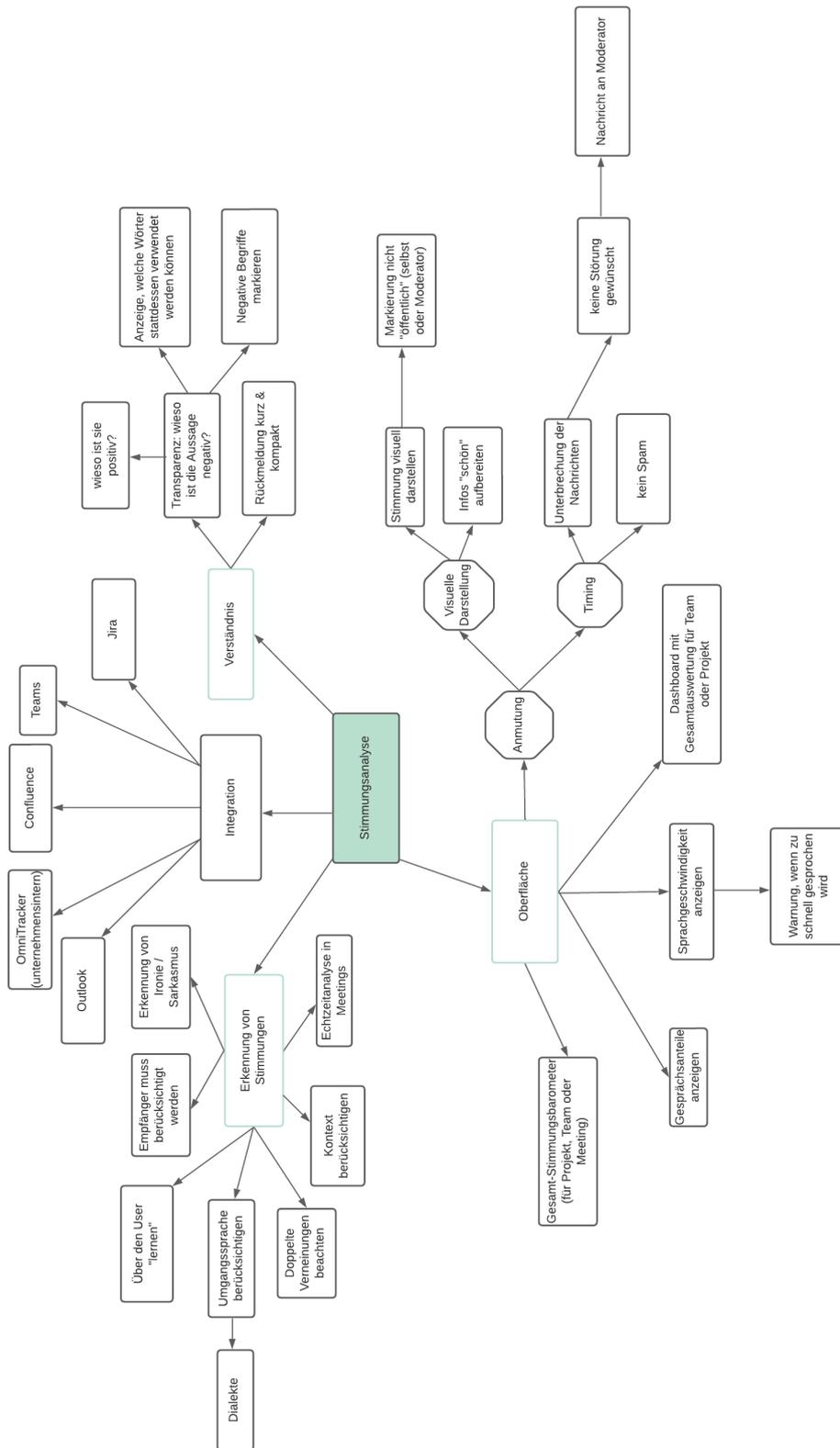


Abbildung A.6: Ergebnisse 1. Interview - Requirements Engineer

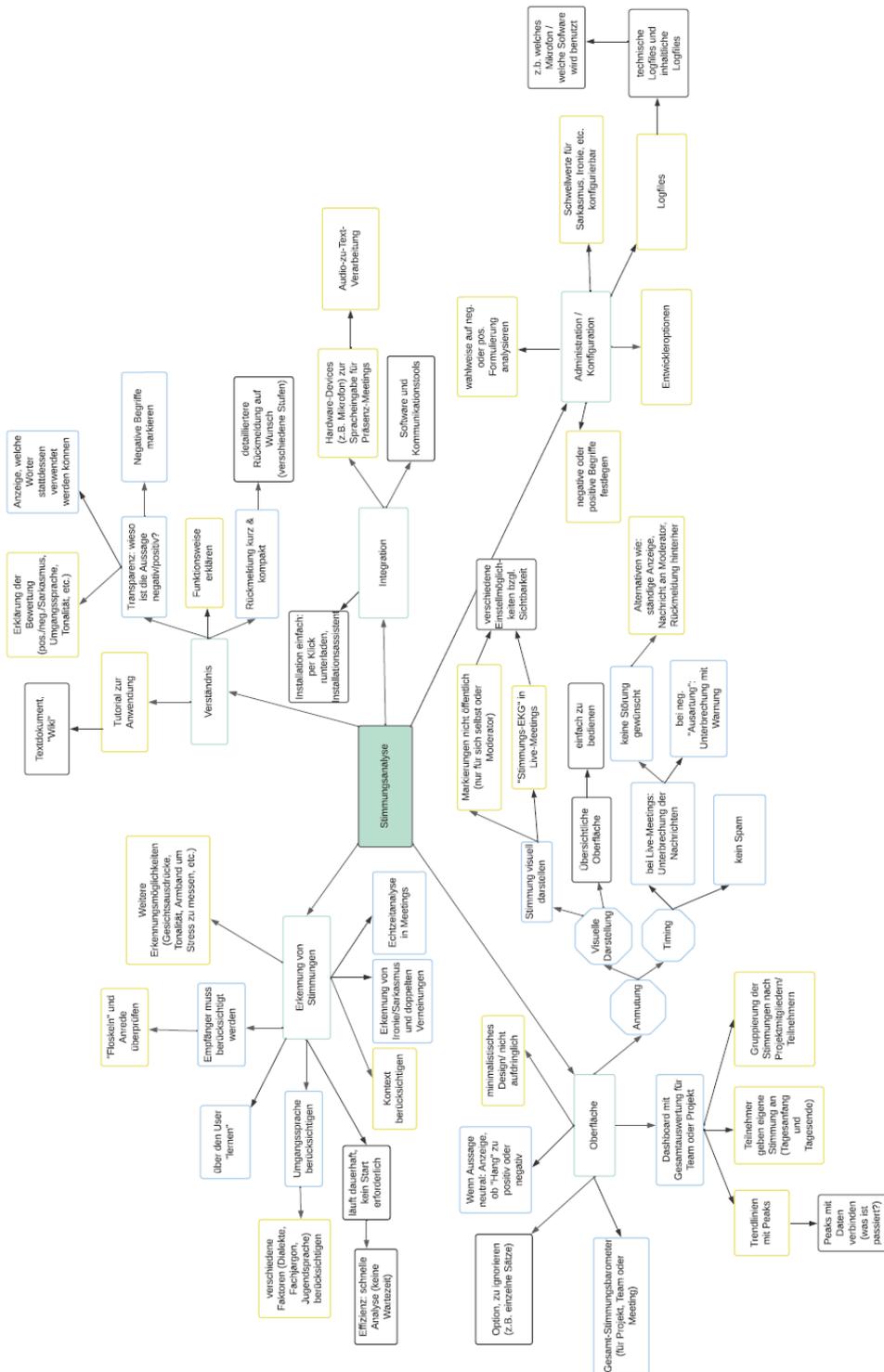


Abbildung A.8: Ergebnisse 2. Interview - Entwickler

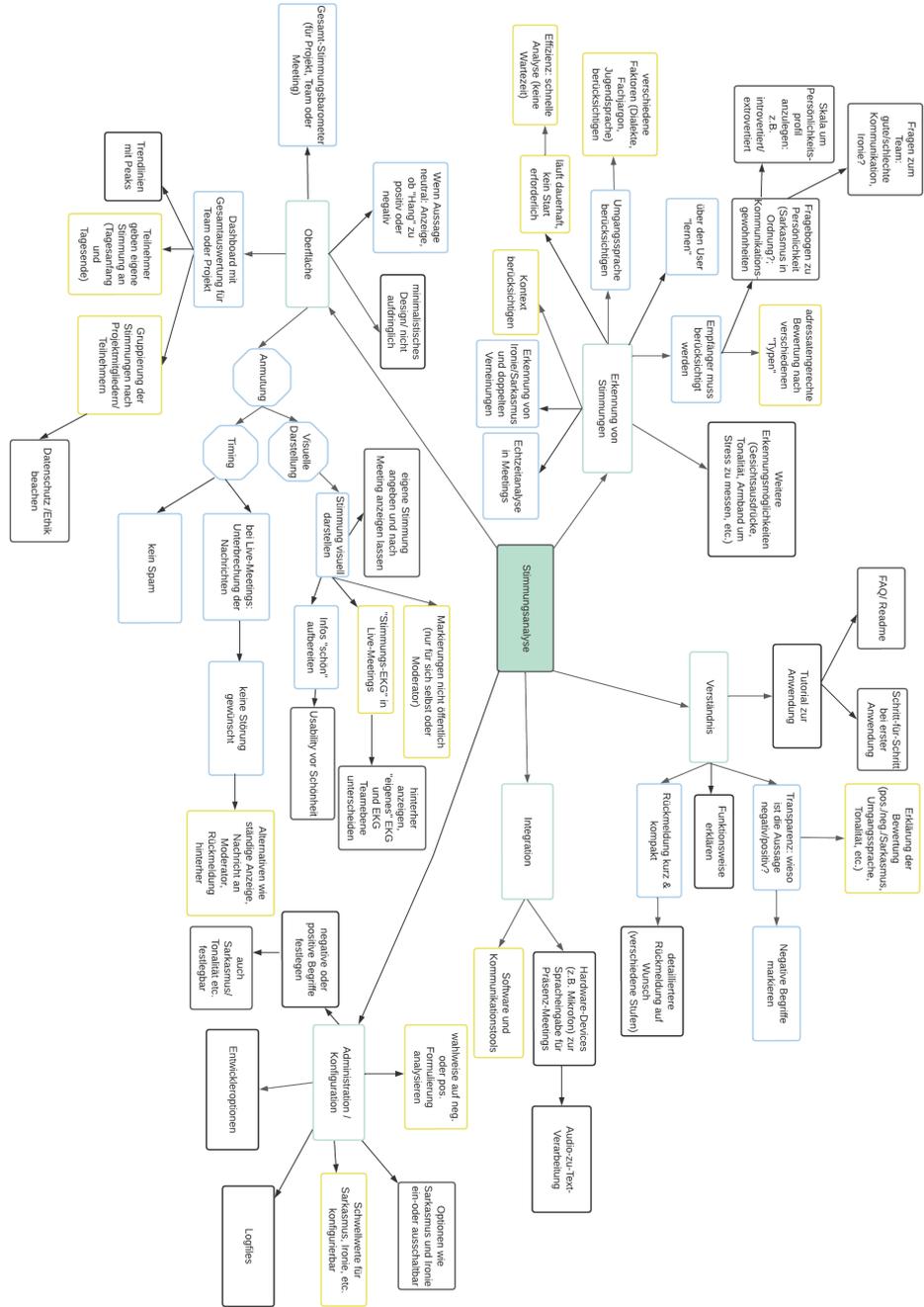


Abbildung A.9: Ergebnisse 2. Interview - Forscher

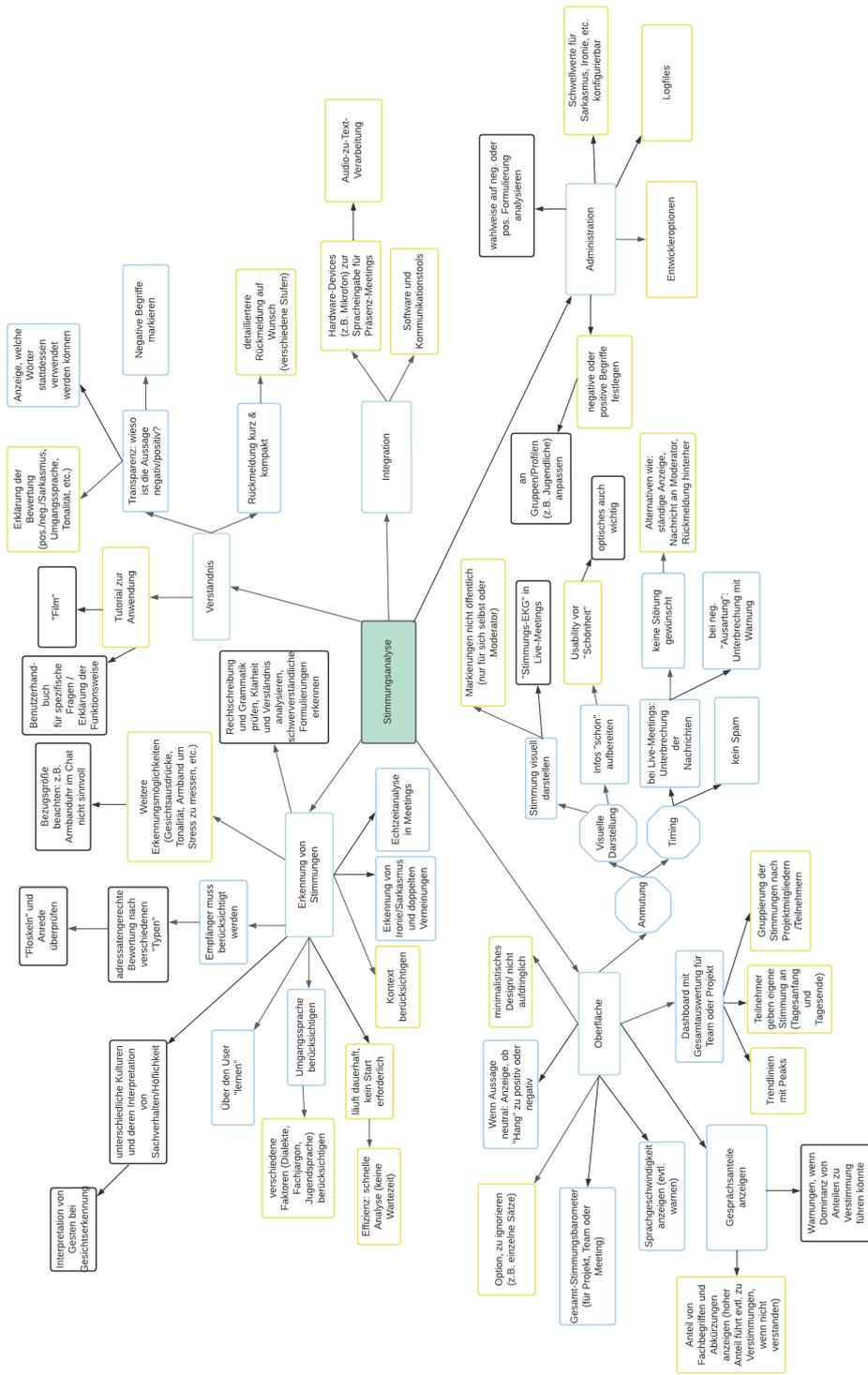


Abbildung A.10: Ergebnisse 2. Interview - Management

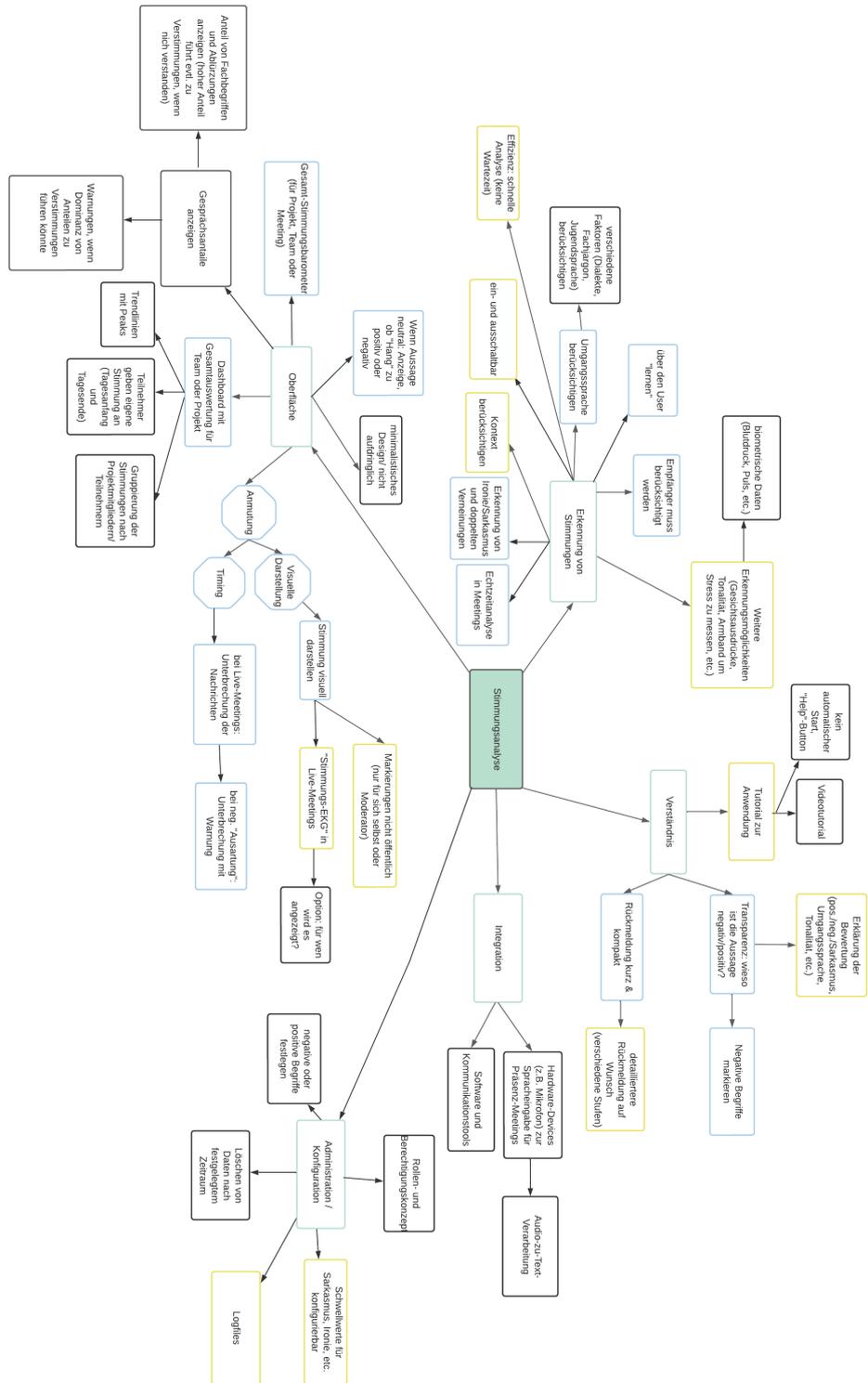


Abbildung A.11: Ergebnisse 2. Interview - Projektleiter

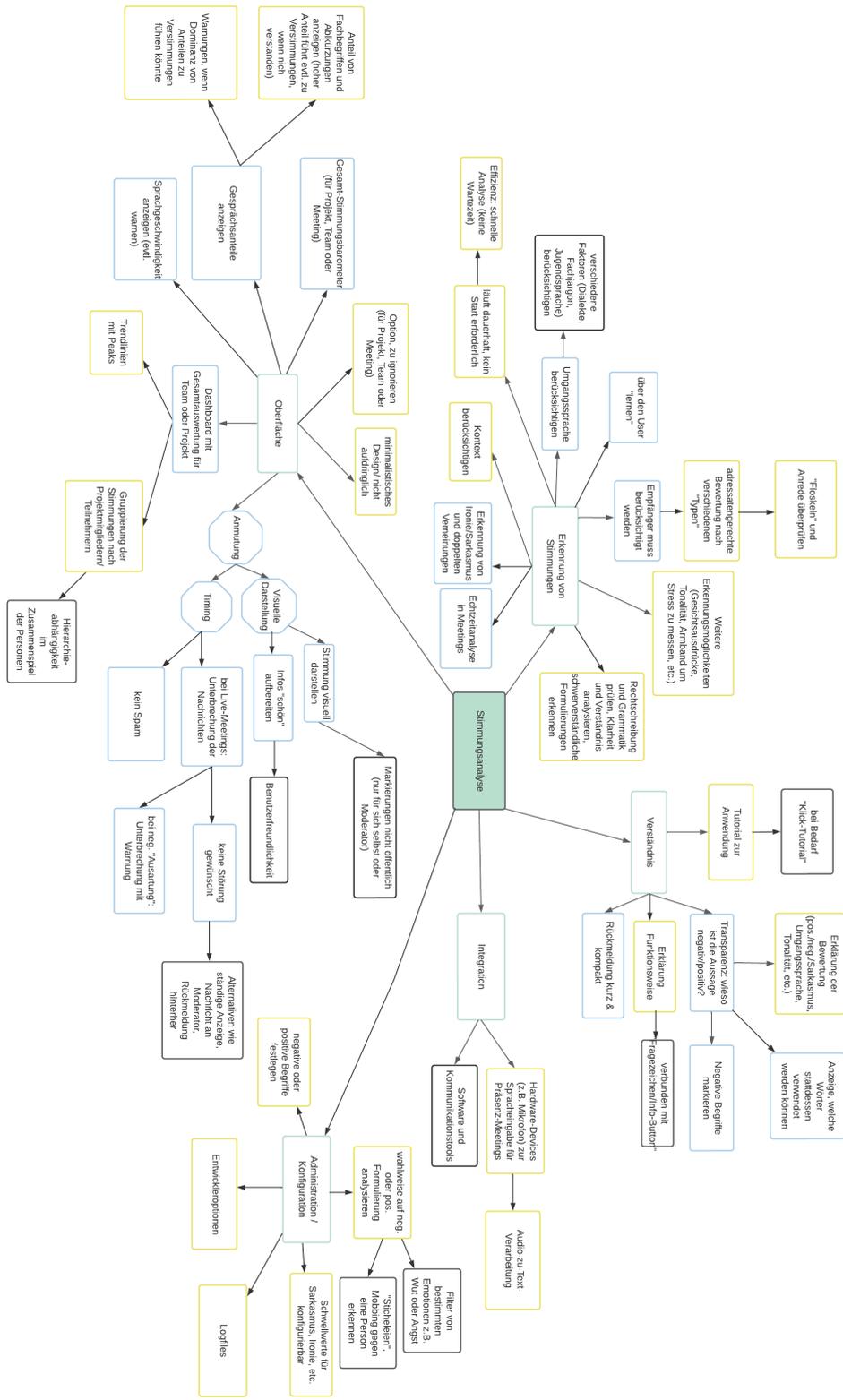


Abbildung A.13: Ergebnisse 2. Interview - Requirements Engineer

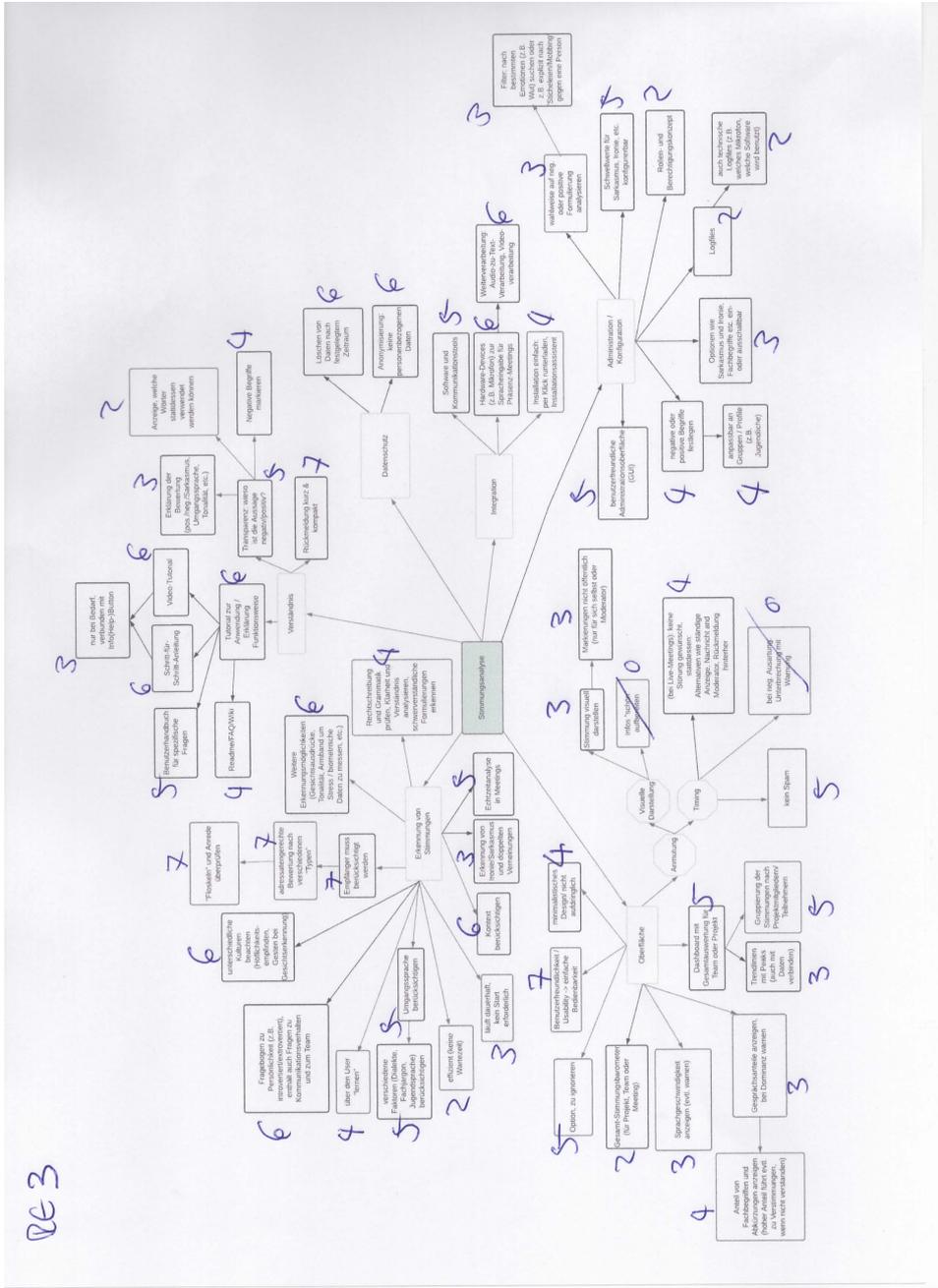


Abbildung A.18: Ergebnisse 3. Interview - Requirements Engineer

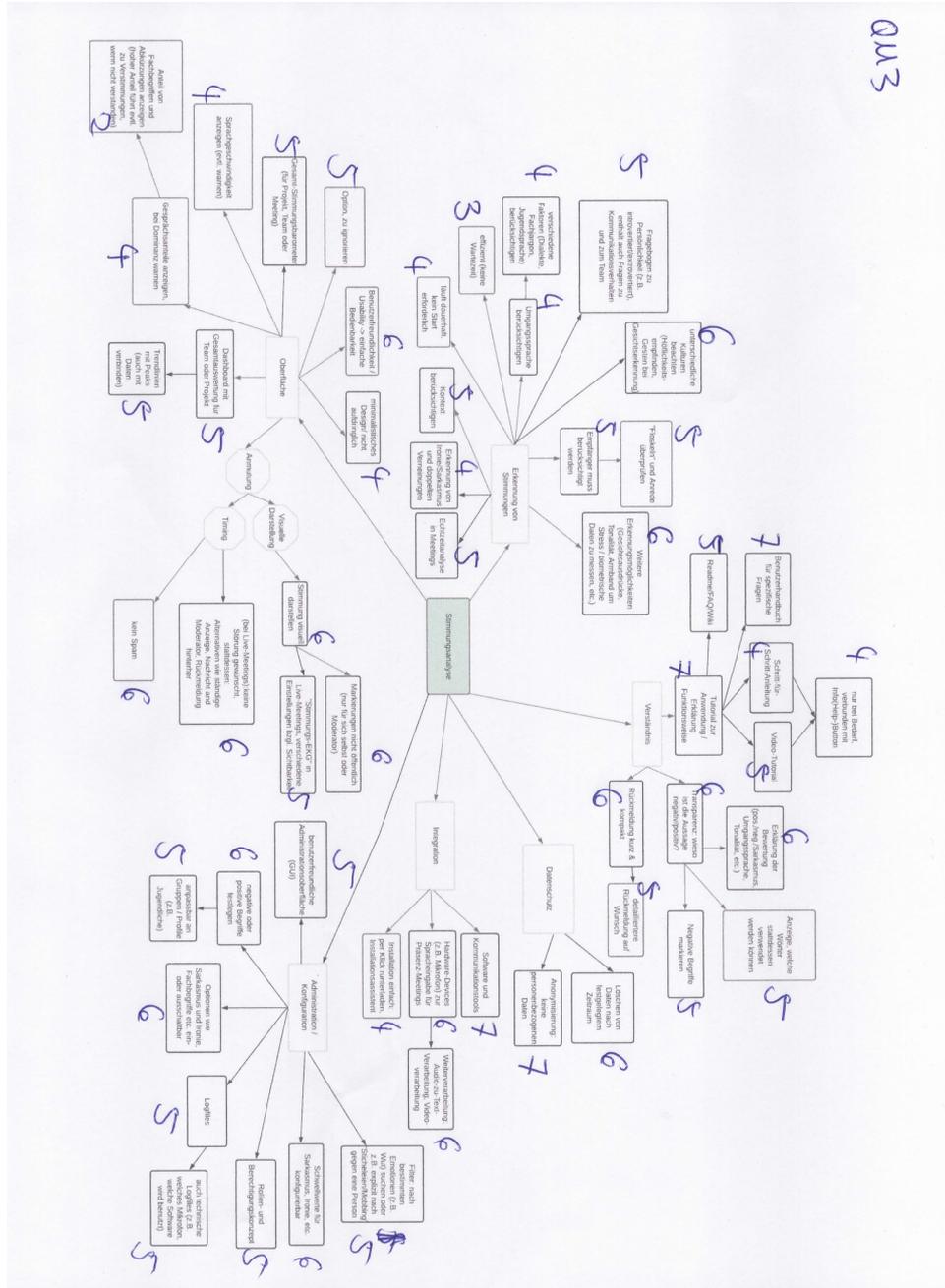


Abbildung A.19: Ergebnisse 3. Interview - Qualitätssicherung

Anhang B

Anforderungskatalog

Erklärung: Kursive Punkte sind Unterpunkte des vorherigen Punkts.

Erkennen von Stimmungen:

- **Echtzeitanalyse in Meetings:** Das Tool läuft in Echtzeit bei z.B. einer Videokonferenz mit und gibt direktes Feedback.
- **Erkennung von Sarkasmus/Ironie und doppelten Verneinungen:** Es sollen vom Tool sowohl Ironie und Sarkasmus als auch doppelte Verneinungen zuverlässig erkannt werden.
- **Kontext berücksichtigen:** die Situation bzw. der Kontext in dem die Konversation stattfindet, soll berücksichtigt werden. Zum Beispiel: Handelt es sich um ein tägliches Meeting mit Kollegen, ein Kundengespräch oder ein Gespräch mit Vorgesetzten?
- **läuft dauerhaft:** es ist kein Start des Tools erforderlich, sondern es ist in bereits verwendete Tools integriert und startet mit dem Starten der anderen Anwendung.
- **Umgangssprache berücksichtigen:** Auch typische Umgangssprache soll in die Bewertung mit einfließen.
- *verschiedene Faktoren dabei berücksichtigen:* nicht nur die Umgangssprache an sich soll berücksichtigt werden, sondern auch Faktoren wie Dialekte, Fachjargon und Jugendsprache sollen erkannt werden.
- **über den User „lernen“:** durch geschriebene Nachrichten / getätigte Aussagen eine Art Profil über den Nutzer erstellen. Gemeint ist z.B. so etwas wie: benutzt der User oft selbst Sarkasmus? Dann macht es ihm wohl nicht so viel aus, selbst eine sarkastische Nachricht zu bekommen, als jemand, der nie Ironie verwendet.

- **Fragebogen zu Persönlichkeit:** soll beim ersten Starten der Anwendung ausgefüllt werden müssen. Hiermit soll bestimmt werden, ob der Anwender z.B. eher extrovertiert oder introvertiert ist. Außerdem sollen Fragen zum Team und zum Kommunikationsverhalten gestellt werden. Damit soll erreicht werden, dass vom System besser eingeschätzt werden kann, wie eine Aussage beim Anwender ankommt.
- **unterschiedliche Kulturen berücksichtigen:** hier gibt es teilweise große Unterschiede zwischen dem Höflichkeitsempfinden. Bei Gesichtserkennung ist außerdem zu beachten, dass unterschiedliche Kulturen verschiedene Gesten verwenden (z.B. Nicken bedeutet nicht immer das gleiche).
- **Empfänger muss berücksichtigt werden:** Es soll berücksichtigt werden, an wen die Nachricht/Aussage gerichtet ist. Wird mit dem Chef, dem Kollegen oder einem Kunden gesprochen?
- *adressatengerechte Bewertung nach verschiedenen Typen:* ähnlich wie Fragebogen, es soll berücksichtigt werden, ob der Empfänger eher „empfindlich“, schüchtern, robust etc. ist.
- *Floskeln und Anrede überprüfen:* Die Anrede und der Abschluss bei Mails sollen überprüft werden und es soll bewertet werden, ob diese adressatengerecht sind.
- **Weitere Erkennungsmöglichkeiten:** Faktoren wie Gesichtsausdrücke und Tonalität sollen bewertet werden. Denkbar ist auch eine Messung von biologischen Daten wie z.B. Stress / Pulsfrequenz durch Armbänder/Smartwatches etc.
- **Rechtschreibung und Grammatik prüfen:** Eine integrierte Rechtschreib- und Grammatikprüfung soll vorhanden sein. Dabei soll außerdem geprüft werden, ob Formulierungen schwerverständlich sind, also z.B. zu viele Schachtelsätze vorhanden sind.

Verständnis

- **Tutorial zur Anwendung:** Es soll irgendeine Form von Tutorial geben, die dem Nutzer vermittelt, wie das Tool zu verwenden ist und die ihm eine Hilfestellung gibt, wenn er nicht weiterweiß. Dabei gibt es verschiedene denkbare Formen:
- *Benutzerhandbuch:* besonders auch für spezifische Fragen.
- *Schritt-für-Schritt-Anleitung:* zum „Durchklicken“, Benutzer wird mit Erklärungen durch das Tool geführt.

- *Video-Tutorial*: die Funktion /Verwendung wird anschaulich über ein Video erklärt.
- **Help-Button**: das Tutorial ist jederzeit über einen Help/Info-Button aufrufbar. Das Tutorial soll nicht „störend“ beim ersten Mal zwangsweise starten.
- **Transparenz**: es soll klar sein, wieso eine Aussage als positiv oder negativ deklariert wurde
- *Erklärung der Bewertung*: zusätzlich enthält die Bewertung auch Erklärungen, ob Sarkasmus / Umgangssprache etc. erkannt wurde und wo.
- *Anzeige, welche Wörter stattdessen verwendet werden können*: Bei negativer Bewertung wird angezeigt, wie man die Aussage umformulieren könnte, sodass sie positiver wirkt.
- *Negative Begriffe werden markiert*: Negativ bewertete Begriffe/Wörter/Passagen werden z.B. rot markiert, damit direkt erkannt werden kann, dass diese vom Gegenüber negativ aufgenommen werden könnten.
- **Rückmeldung kurz & kompakt**: ein kurzes Feedback soll angezeigt werden, um eine schnelle Übersicht über die Bewertung zu bekommen, auf Wunsch auch:
- *detailliertere Rückmeldung*: z.B. wie unter „Erklärung der Bewertung“. Wenn die Bewertung z.B. nicht verstanden wird, kann hier genauer nachgeprüft werden, wie sie zustande gekommen ist.

Datenschutz

- **Löschen von Daten**: nach einem festgelegten / gesetzlich vorgegebenen Zeitraum werden alle gesammelten Daten gelöscht.
- **Anonymisierung**: Daten werden nicht personenbezogen gesammelt.

Integration

- **Software & Kommunikationstools**: Das Tool ist wie ein Plugin in gängige Software, die von den Unternehmen verwendet werden, integrierbar. Dazu zählen MS Teams, MS Outlook, Confluence, Jira, etc. und unternehmensspezifische Tools.

- **Die Installation des Tools ist einfach:** es ist per Klick runterladbar und/oder verbunden mit einem Installationsassistenten.
- **Weiterverarbeitung von Audio-zu-Text oder Videos:** Wenn das Tool in einem Präsenz-Meeting verwendet werden soll, sollte die Möglichkeit einer Spracheingabe/Aufnahme über ein Mikrofon bestehen.

Administrations/Konfiguration:

- **Benutzerfreundliche Administrationsoberfläche:** für den Admin soll eine GUI bereitstehen, hier sollen die im Folgenden erläuterten verschiedenen Einstellmöglichkeiten vorgenommen werden können.
- **negative oder positive Begriffe festlegen:** Es soll die Möglichkeit geben, explizit eine Liste von Begriffen festzulegen, die immer positiv oder immer negativ bewertet werden sollen, unabhängig von der Logik des Tools
- *anpassbar an Gruppen/Profile:* möglich soll auch sein, diese Listen bestimmten Gruppen zuzuordnen.
- **Optionen wie Sarkasmus ein- oder ausschaltbar:** es soll explizit eingestellt werden können, ob Faktoren wie Sarkasmus / Fachbegriffe / Umgangssprache / Tonalität etc. bei der Bewertung berücksichtigt werden sollen.
- **Logfiles:** Bewertungen sollen mit Datum und Uhrzeit gespeichert werden und zum späteren Abruf bereitstehen.
- *Technische Logfiles:* auch technische Daten wie das verwendete Mikrofon oder die verwendete Software (z.B. Teams) sollen gespeichert werden.
- *Rollen- und Berechtigungskonzept:* um die Administrationsoberfläche überhaupt umsetzen zu können, braucht es ein Rollen- und Berechtigungskonzept. Es soll die Möglichkeit geben, bestimmten Nutzern Rollen zuzuweisen, z.B. die des Admins oder Projektleiters.
- *Schwellwerte konfigurierbar:* für Faktoren wie Sarkasmus soll festgelegt werden können, wie viel Sarkasmus „angemessen ist“. Z.B. soll eine Nachricht negativ bewertet werden, wenn mehr als fünf Sätze sarkastisch sind.

- *Wahlweise auf positive oder negative Formulierungen analysieren:* es soll die Möglichkeit geben, einzustellen, dass z.B. NUR negative Formulierungen angezeigt werden und positive oder neutrale Sätze ignoriert/ nicht angezeigt werden.
- *Filter:* es soll möglich sein, Konversationen auf eine bestimmte erkannte Emotion (z.B. Wut, Angst) zu untersuchen. Weiterhin soll es die Möglichkeit geben, explizit nach Mobbing/Sticheleien zu suchen.

Oberfläche

- **minimalistisches Design:** das Design sollte nicht aufdringlich sein, um nicht von eigentlichen Aufgaben abzulenken:
- **Benutzerfreundlichkeit / Usability:** das Design ist darauf ausgelegt, möglichst benutzerfreundlich und einfach bedienbar zu sein.
- **Option, zu ignorieren:** Es soll, ähnlich wie bei „Word Rechtschreibprüfung“ die Option geben, z.B. per Rechtsklick einen Satz aus der Bewertung herauszunehmen.
- **Anzeige des „Hangs“, wenn Aussage neutral:** Es sollte nicht nur positiv, negativ und neutral als Ergebnis geben, sondern viel mehr eine Art Skala, z.B. von -5 bis 5, wobei 0 neutral, -5 extrem negativ und 5 extrem positiv ist.
- **Gesamt-Stimmungsbarometer:** die Oberfläche enthält ein Gesamt-Stimmungsbarometer, z.B. wie ein Thermometer, und dieses bezieht sich auf die gesamte Stimmung im Team, Projekt oder Meeting.
- **Sprachgeschwindigkeit anzeigen:** bei Live-Meetings wird die Sprachgeschwindigkeit gemessen. Es wird gewarnt, wenn ein Teilnehmer zu schnell spricht und daher z.B. schwer verständlich ist.
- **Gesprächsanteile anzeigen:** bei Live-Meetings soll gemessen werden, wie hoch der Gesprächsanteil der verschiedenen Teilnehmer ist. Es soll gewarnt werden, wenn jemand dominiert (außer es handelt sich um einen Vortrag) und die anderen nicht zu Wort kommen.
- *Anteil von Fachbegriffen:* angezeigt werden soll nicht nur der Gesprächsanteil, sondern auch der Anteil von Fachbegriffen. Je nach Kontext soll gewarnt werden. Zum Beispiel dann, wenn es sich um ein Kundengespräch handelt und der Kunde die Begriffe nicht versteht und daher verstimmt werden könnte.

- **Dashboard mit Gesamtauswertung für Team / Projekt:** als Übersicht über die Stimmung in einem Projekt oder Team. Enthält verschiedene Daten, beispielweise:
- *Trendlinien mit Peaks:* Trendlinien für Stimmungen im Projekt, extrem negative oder positive „Ausreißer“ oder Häufungen werden mit Daten verbunden und angezeigt.
- *Teilnehmer geben eigene Stimmung an:* Am Anfang und Ende eines Arbeitstages soll es für Teilnehmer die Möglichkeit geben, ihre eigene Stimmung anzugeben (beispielweise mit grünen, gelben, oder roten Smileys). Das dient zum einen dazu, das Tool zu verifizieren, aber andererseits kann die Führungskraft auch erkennen, wenn z.B. jemand mit guter Stimmung ankommt, aber mit schlechter wieder nach Hause geht und die Ursache (z.B. mithilfe der Daten) untersuchen.
- *Gruppierung der Stimmungen nach Projektmitgliedern /Teilnehmern:* Für ein Projekt oder Team werden z.B. Pie-Charts angezeigt, die den Anteil von positiven/neutralen/negativen Aussagen und detaillierter auch den Anteil bestimmter Emotionen (Wut/Angst/etc.) daran anzeigen.
- **kein Spam:** die Ergebnisse ploppen nicht für jeden Satz/jede Aussage „nervig“ auf, es gibt keine Pop-Ups, sondern sie werden minimal und nicht störend dargestellt.
- bei Live Meetings soll bei einer negativen Ausartung der Stimmung eine **Unterbrechung mit Warnung** stattfinden. **ODER**
- es findet **keine Unterbrechung** statt, stattdessen gibt es Alternativen, wie z.B. die Benachrichtigung nur an den Moderator senden, ständige Anzeige oder Rückmeldung hinterher.
- **Stimmung visuell darstellen:** die ermittelte Stimmung wird optisch dargestellt, z.B mit:
- *Stimmungs-EKG in Live-Meetings:* es soll z.B. oben links eine kleine „EKG“-Kurve angezeigt werden, die den Stimmungsverlauf verdeutlicht. Dafür soll es verschiedene Einstellungen der Sichtbarkeit geben: z.B. für das gesamte Team angezeigt, gesamte Stimmung nur für Moderator sichtbar, jeder Teilnehmer bekommt nur Stimmung der eigenen Aussagen angezeigt, etc. *Markierungen nicht öffentlich:* die Bewertung der eigenen Aussagen soll in Live-Meetings nicht allen Teilnehmern, sondern nur dem, der die Aussage getan hat, angezeigt werden.

Anhang C

Video-Skript

Bild	Text
Dokument und Kommentar mit Pfeilen auf Analysetool (Blackbox), aus der Box heraus drei Pfeile die auf positive, negative und neutrale Stimmung (und dazugehörige Emotionen) zeigen, die nacheinander erscheinen.	Die Stimmungsanalyse ist ein Mittel, um Stimmungen von Fließtexten oder Aussagen zu analysieren. Dabei wird angegeben, ob die vermittelte Stimmung eines Textes oder Satzes positiv, negativ oder neutral zu bewerten ist. Jeder Stimmungspolarität werden verschiedene Emotionen zugeordnet. Emotionen einer positiven Stimmung sind beispielweise Freude, Aktivität oder Interesse. Negative Emotionen dagegen Wut, Angst oder Nervosität. Ist eine Aussage keiner bestimmten Emotion zu zuordnen, wird sie als neutral bewertet.
Mehrere Personen im Meeting, Denkblasen (repräsentieren Stimmungen) noch leer, weil unbekannt. Pfeil auf Analysetool.	Für die Stimmungsanalyse ergeben sich innerhalb eines Softwareprojekts vielfältige Einsatzmöglichkeiten.
Gleiches Meeting, Denkblasen enthalten nun negative Emotionen. Kreuz um zu verdeutlichen, dass dies „falsch“ ist.	Sie kann dazu eingesetzt werden, die Stimmung im Meeting zu analysieren,
Zwei Personen am Tisch, die ein Gespräch führen	zu besprechen sowie Mitglieder zu schulen,
Wieder das Meeting, dieses Mal mit positiven Emotionen, Haken um zu verdeutlichen, dass es nun besser/ „richtig“ ist	und anschließend zu verbessern.
Symbole für Mail, Dokument und Kommentar	Sie kann dazu eingesetzt werden, um vor dem Abschicken einer Mail, eines Dokuments, oder eines Kommentars, den Inhalt auf negative Stimmungen zu kontrollieren und ggf. anzupassen, um die Stimmung im Team zu verbessern.
Zwei Personen, zwischen denen eine Mail ausgetauscht wird, welche negative Emotionen beim Empfänger verursacht.	Oft ist dem Absender gar nicht bewusst, welche Stimmung er dem Empfänger des Textes vermittelt.
Bevor Person 1 die Mail sendet, wird sie mit einem Analysetool bewertet, daraufhin wird eine verbesserte Mail an Person 2 gesendet, die beim Empfänger positive Emotionen hervorruft.	Ein Stimmungsanalysetool soll dem Absender helfen, sich vor dem Verschicken seines Textes einer möglicherweise negativen Wirkung bewusst zu werden und gibt ihm die Möglichkeit, diesen anzupassen.
Nacheinander erscheinen Symbole für Händedruck, Produktivität (Zahnräder), Aufgabenverteilung und ein „Daumen hoch“	Denn positive Stimmungen betreffen nicht nur die Beziehung zwischen zwei Personen, sondern können sich auch direkt auf die Produktivität des Projekts, die Synchronisation von Aufgaben und letztendlich auch auf die Jobzufriedenheit auswirken.

Abbildung C.1: Skript zum Erklärvideo

Anhang D

Interview-Skript

Hinweis: Ein Skript existiert nur für die erste Interviewrunde, da die Baumstrukturen sich bei jedem Teilnehmer nach der ersten Runde unterscheiden, auf jeden Teilnehmer und deren Anforderungen individuell eingegangen wird und dies nicht vorher geplant werden kann.

Begrüßung Interviewteilnehmer / Vorbereitung:

„Guten Tag Herr/Frau X, danke, dass Sie sich dazu bereit erklärt haben, an der Studie teilzunehmen. Mit der Aufnahme sind Sie einverstanden?“

„Vielen Dank. Ich starte zunächst einmal mit ein paar allgemeinen Fragen.“

„Sie sind zurzeit im Unternehmen X als X tätig, ist das richtig?“

„Wie viele Jahre Berufserfahrung haben Sie in dieser Position?“

„Haben Sie vorher bereits in einer anderen Position gearbeitet?“

„Das wäre erstmal alles. In diesem Interview geht um mögliche Anforderungen an die Stimmungsanalyse in Softwareprojekten. Dazu würde ich Ihnen nun zunächst ein kurzes Video zur Erklärung vorspielen. Anschließend zeige ich Ihnen mögliche Anforderungen an die Stimmungsanalyse, die aus einer vorangegangenen Studie stammen. Sie werden dann gebeten, die Anforderungen nach Ihren Bedürfnissen zu ergänzen.“

Video zeigen

„Haben Sie dazu Fragen oder möchten Sie das Video nochmal sehen?“

„Dann würden wir jetzt mit dem Hauptteil starten. Sollten im weiteren Verlauf Fragen, Unklarheiten oder Verständnisprobleme auftreten, dürfen Sie natürlich jederzeit nachfragen.“

Hauptteil

Aushändigung der Mindmap

„Verschaffen Sie sich zunächst in Ruhe einen Überblick über die Mindmap. Die Anforderungen sind zunächst in drei Kategorien eingeteilt (Oberfläche, Erkennen von Stimmungen und Verständnis). Wenn Sie später gerne weitere Kategorien hinzufügen möchten, steht es Ihnen frei dies zu tun.“

Zeit lassen

„Konnten Sie sich einen Überblick verschaffen? Haben Sie Fragen zu bestimmten Anforderungen?“ „Dann würde ich Sie bitten, zunächst einige Anforderungen zu streichen, die für Sie weniger relevant sind und auf die Sie bei einem Stimmungsanalysetool verzichten könnten.“

Zeit lassen

„Schauen Sie sich als nächstes die übrig gebliebenen Anforderungen an. Können Sie diese genauer konkretisieren oder erweitern?“

Zeit lassen

„Vielen Dank bis hierhin. Was fehlt nun noch, wäre aber für Sie relevant? Denken Sie dabei an Ihre täglichen Aufgaben und Routinen in Ihrem Unternehmen. Berücksichtigen Sie dabei sowohl Funktionen und Aussehen des Tools aber auch was z.B. technisch nötig wäre, um ein Stimmungsanalysetool in Ihren Arbeitsalltag zu integrieren.“

Abschluss

„Vielen Dank, das war es auch schon. Haben Sie noch Fragen oder Anmerkungen?“

„Dann bedanke ich mich noch einmal für Ihre Teilnahme und würde die Aufnahme nun beenden.“

Abbildungsverzeichnis

2.1	Bearbeitungspipeline des SEnti-Analyzers[18]	8
2.2	CrowdRE: Multi-Level-Approach [16]	12
2.3	StakeRare-Prozess [22]	13
2.4	Iterative Interviewstudie [26]	14
3.1	Prozess der Literaturrecherche	18
4.1	Iterative Interviewstudie, angepasst aus [26]	25
4.2	Beispiel für verwendete Baumstruktur	27
5.1	Anforderungen aus unveröffentlichter Studie	35
5.2	Priorisierungen zu Anforderungen der Kategorie Datenschutz	42
5.3	Priorisierungen zu Anforderungen der Kategorie Integration	42
5.4	Priorisierungen zu Anforderungen der Kategorie Administration	43
5.5	Priorisierungen zu Anforderungen der Kategorie Verständnis	43
5.6	Priorisierungen zu Anforderungen der Kategorie Erkennen von Stimmungen	44
5.7	Priorisierungen zu Anforderungen der Kategorie Oberfläche	45
5.8	Startansicht SentiStrength	56
5.9	Auswertung SentiStrength	57
5.10	Startansicht SentiStrengthSE	58
A.1	Ergebnisse 1. Interview - Entwickler	76
A.2	Ergebnisse 1. Interview - Forscher	77
A.3	Ergebnisse 1. Interview - Management	78
A.4	Ergebnisse 1. Interview - Projektleiter	79
A.5	Ergebnisse 1. Interview - Qualitätssicherung	80
A.6	Ergebnisse 1. Interview - Requirements Engineer	81
A.7	Ergebnisse 1. Interview - Zusammenfassung	82
A.8	Ergebnisse 2. Interview - Entwickler	83
A.9	Ergebnisse 2. Interview - Forscher	84
A.10	Ergebnisse 2. Interview - Management	85
A.11	Ergebnisse 2. Interview - Projektleiter	86
A.12	Ergebnisse 2. Interview - Qualitätssicherung	87

A.13 Ergebnisse 2. Interview - Requirements Engineer	88
A.14 Ergebnisse 2. Interview - Zusammenfassung	89
A.15 Ergebnisse 3. Interview - Entwickler	90
A.16 Ergebnisse 3. Interview - Management	91
A.17 Ergebnisse 3. Interview - Forscher	92
A.18 Ergebnisse 3. Interview - Requirements Engineer	93
A.19 Ergebnisse 3. Interview - Qualitätssicherung	94
A.20 Ergebnisse 3. Interview - Projektleiter	95
C.1 Skript zum Erklärvideo	104

Literaturverzeichnis

- [1] T. Ahmed, A. Bosu, A. Iqbal, and S. Rahimi. Senticr: A customized sentiment analysis tool for code review interactions. In *Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering*, ASE '17, page 106–111. IEEE Press, 2017.
- [2] A. M. Alkalbani, A. M. Ghamry, F. K. Hussain, and O. K. Hussain. Sentiment analysis and classification for software as a service reviews. In *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, pages 53–58, 2016.
- [3] J. Buchan, M. Bano, D. Zowghi, and P. Volabouth. Semi-automated extraction of new requirements from online reviews for software product evolution. In *2018 25th Australasian Software Engineering Conference (ASWEC)*, pages 31–40. IEEE, 2018.
- [4] F. Calefato, F. Lanubile, F. Maiorano, and N. Novielli. Sentiment polarity detection for software development. *Empirical Software Engineering*, 23(3):1352–1382, sep 2017.
- [5] F. Calefato, F. Lanubile, and N. Novielli. Emotxt: A toolkit for emotion recognition from text. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 79–80, 2017.
- [6] L. V. G. Carreño and K. Winbladh. Analysis of user comments: an approach for software requirements evolution. In *2013 35th international conference on software engineering (ICSE)*, pages 582–591. IEEE, 2013.
- [7] M. Christel and K. Kang. Issues in requirements elicitation. Technical Report CMU/SEI-92-TR-012, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 1992.
- [8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1, page 2, 2019.

- [9] J. Ferreira, M. Glynn, D. Hunt, J. Babu, D. Dennehy, and K. Conboy. Sentiment analysis of open source communities: An exploratory study. In *Proceedings of the 15th International Symposium on Open Collaboration*, pages 1–5, 2019.
- [10] D. Graziotin, F. Fagerholm, X. Wang, and P. Abrahamsson. Consequences of unhappiness while developing software. In *2017 IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion)*, pages 42–47, 2017.
- [11] D. Graziotin, F. Fagerholm, X. Wang, and P. Abrahamsson. On the unhappiness of software developers. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering, EASE’17*, page 324–333, New York, NY, USA, 2017. Association for Computing Machinery.
- [12] D. Graziotin, X. Wang, and P. Abrahamsson. Do feelings matter? on the correlation of affects and the self-assessed productivity in software engineering. *CoRR*, abs/1408.1293, 2014.
- [13] D. Graziotin, X. Wang, and P. Abrahamsson. Happy software developers solve problems better: Psychological measurements in empirical software engineering. *PeerJ*, 2:e289, 03 2014.
- [14] D. Graziotin, X. Wang, and P. Abrahamsson. Software developers, moods, emotions, and performance. *Software, IEEE*, 31, 05 2014.
- [15] D. Graziotin, X. Wang, and P. Abrahamsson. How do you feel, developer? an explanatory theory of the impact of affects on programming performance. *PeerJ Computer Science*, 1, 05 2015.
- [16] E. C. Groen, J. Doerr, and S. Adam. Towards crowd-based requirements engineering a research preview. In S. A. Fricker and K. Schneider, editors, *Requirements Engineering: Foundation for Software Quality*, pages 247–253, Cham, 2015. Springer International Publishing.
- [17] M. Herrmann and J. Klünder. From textual to verbal communication: Towards applying sentiment analysis to a software project meeting. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, pages 371–376, 2021.
- [18] M. Herrmann, M. Obaidi, and J. Klünder. Senti-analyzer: joint sentiment analysis for text-based and verbal communication in software projects. *arXiv preprint arXiv:2206.10993*, 2022.
- [19] M. Islam and M. Zibran. Sentistrength-se: Exploiting domain specificity for improved sentiment analysis in software engineering text. *Journal of Systems and Software*, 145, 08 2018.

- [20] M. R. Islam and M. F. Zibran. A comparison of software engineering domain specific sentiment analysis tools. In *2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 487–491, 2018.
- [21] P. Kruchten. *The Rational Unified Process: An Introduction*. Addison-Wesley Longman Publishing Co., Inc., USA, 3 edition, 2003.
- [22] S. L. Lim and A. Finkelstein. Stakerare: Using social networks and collaborative filtering for large-scale requirements elicitation. *IEEE Transactions on Software Engineering*, 38(3):707–735, 2012.
- [23] S. L. Lim, D. Quercia, and A. Finkelstein. Stakenet: using social networks to analyse the stakeholders of large-scale software projects. In *Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering-Volume 1*, pages 295–304, 2010.
- [24] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.
- [25] S. C. Müller and T. Fritz. Stuck and frustrated or in flow and happy: Sensing developers’ emotions and progress. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, volume 1, pages 688–699. IEEE, 2015.
- [26] P. K. Murukannaiah, N. Ajmeri, and M. P. Singh. Acquiring creative requirements from the crowd: Understanding the influences of personality and creative potential in crowd re. In *2016 IEEE 24th International Requirements Engineering Conference (RE)*, pages 176–185. IEEE, 2016.
- [27] P. K. Murukannaiah, N. Ajmeri, and M. P. Singh. Toward automating crowd re. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pages 512–515, 2017.
- [28] N. Novielli, D. Girardi, and F. Lanubile. A benchmark study on sentiment analysis for software engineering research. In *Proceedings of the 15th International Conference on Mining Software Repositories*, pages 364–375, 2018.
- [29] M. Obaidi and J. Klünder. Development and application of sentiment analysis tools in software engineering: A systematic literature review. In *Evaluation and Assessment in Software Engineering, EASE 2021*, page 80–89, New York, NY, USA, 2021. Association for Computing Machinery.

- [30] M. Obaidi, L. Nagel, A. Specht, and J. Klünder. Sentiment analysis tools in software engineering: A systematic mapping study. *Information and Software Technology*, 151:107018, 2022.
- [31] C. Pacheco and I. Garcia. Stakeholder identification methods in software requirements: Empirical findings derived from a systematic review. In *2008 The Third International Conference on Software Engineering Advances*, pages 472–477, 2008.
- [32] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall. How can i improve my app? classifying user reviews for software maintenance and evolution. In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 281–290, 2015.
- [33] R. S. Pressman. *Software Engineering: A Practitioner’s Approach (McGraw-Hill Series in Computer Science)*. McGraw Hill College Division, 2004.
- [34] J. Saldaña. The coding manual for qualitative researchers. *The coding manual for qualitative researchers*, pages 1–440, 2021.
- [35] K. Schneider, J. Klünder, F. Kortum, L. Handke, J. Straube, and S. Kauffeld. Positive affect through interactions in meetings: The role of proactive and supportive statements. *Journal of Systems and Software*, 143:59–70, 2018.
- [36] C. Smith and R. Lazarus. *Emotion and Adaptation*, volume 21, pages 609–637. 01 1990.
- [37] I. Sommerville. *Software engineering*. America: Pearson Education Inc, 2011.
- [38] C. P. Team. Cmmi for development, version 1.2. Technical Report CMU/SEI-2006-TR-008, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2006.
- [39] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61:2544–2558, 12 2010.
- [40] C. Werner, Z. S. Li, and N. Ernst. What can the sentiment of a software requirements specification document tell us? In *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, pages 106–107, 2019.

- [41] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. *ACM International Conference Proceeding Series*, 05 2014.
- [42] J. F. Wolfswinkel, E. Furtmueller, and C. P. Wilderom. Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, 22:45–55, 2013.
- [43] J. Wu, C. Ye, and H. Zhou. Bert for sentiment classification in software engineering. In *2021 International Conference on Service Science (ICSS)*, pages 115–121, 2021.
- [44] T. Zhang, B. Xu, F. Thung, S. A. Haryono, D. Lo, and L. Jiang. Sentiment analysis for software engineering: How far can pre-trained transformer models go? In *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 70–80, 2020.
- [45] L. Zhao and A. Zhao. Sentiment analysis based requirement evolution prediction. *Future Internet*, 11:52, 02 2019.

