Institut für Informatik
Lehrstuhl für Programmier- und Modellierungssprachen

LUDWIG–MAXIMILIANS–UNIVERSITÄT MÜNCHEN

Diplomabeit

Implementierung einer Auswertungsplattform für Reputationsalgorithmen

Michael Weisbein

Informatik Diplom

Aufgabensteller: Prof. Dr. Francois Bry
Betreuer: Alexander Pohl und Niels Heller
Abgabetermin: 17. Dezember 2015

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

Ich erkläre mich damit einverstanden, dass meine Abschlussarbeit ganz oder in Teilen im Rahmen des Internet- Auftritts des Instituts für Informatik der LMU unter Angabe meines Namens veröffentlicht werden darf.

München, den 17. Dezember 2015

.....

(Unterschrift des Kandidaten)

Zusammenfassung

Im Mittelpunkt dieser Arbeit steht die Konzipierung und Realisierung einer webbasierten Auswertungsplattform für Reputationsalgorithmen. Das Konzept sieht eine vergleichende Nutzerstudie vor, die einen Chat beinhaltet, in dem jeder Proband jede Nachricht der anderen Probanden bewerten kann. Diese Bewertungen bilden die Basis zur Berechnung von Reputationswerten des zu testenden Reputationsalgorithmus. Während des Chats bildet sich bei den Probanden eine Meinung über die anderen Probanden, die in Form einer subjektiven Rangfolge dokumentiert wird. Über den Vergleich zwischen der Meinung der Probanden und berechneten Reputationswerten können Rückschlüsse über die Qualität des Reputationsalgorithmus gezogen werden. Im Anschluss an das Konzept folgt die Beschreibung der technischen Umsetzung eines ersten Prototyps der Auswertungsplattform. Die Arbeit wird mit einem Ausblick abgeschlossen, in dem mögliche Erweiterungen der Auswertungsplattform vorgeschlagen werden.

Abstract

The main concern of this thesis is the design and realization of a web-based evaluation platform for reputation algorithms. The design is based on a comparative user study that includes a chat offering to each participant the possibility to rate the messages of other participants. These ratings are used to calculate reputation values of the reputation algorithm that should be evaluated. During a chat every participant gathers an opinion about the other participants that is documented as a subjective ranking. By comparing the opinion of the participants with the calculated reputation values it is possible to draw conclusions about the quality of the reputation algorithm. Referring to the concept a technical realization of a prototype is proposed. Finally, an outlook about possible improvements of the evaluation platform is given.

Danksagung

An dieser Stelle möchte ich mich zuallererst bei Prof. Dr. François Bry für die Möglichkeit bedanken, diese Diplomarbeit am Lehrstuhl für Programmier- und Modellierungssprachen anfertigen zu dürfen.

Ganz besonders möchte ich mich bei meinen beiden Betreuern Alexander Pohl und Niels Heller bedanken, die in zahlreichen Besprechungen viel Geduld an den Tag gelegt haben und mich stets super unterstützt haben.

Außerdem bedanke ich mich bei meiner Familie und meinen Freunden, die sich des Öfteren meinen Informatikmonologen unterziehen mussten. Zu guter Letzt danke ich meiner Freundin Sabine für das Korrekturlesen und die Unterstützung in dieser nicht immer einfachen Periode.

Inhaltsverzeichnis

1 Einleitung **1**

1.1 Backstage . . . 2
1.2 Gliederung . . . 2

2 Forschungsstand **5**

2.1 Begriffsabgrenzung . . . 5
2.1.1 Reputation . . . 6
2.1.2 Trust . . . 6
2.1.3 Status . . . 7
2.2 Merkmale von Reputationssystemen . . . 8
2.2.1 Metriken der Reputationsberechnung . . . 9
2.2.2 Zentrale vs. Dezentrale Architekturen . . . 10
2.2.3 Reputation von Personen vs. Ressourcen . . . 12
2.2.4 Globale vs. Personalisierte Reputation . . . 12
2.3 Eigenschaften und Vorteile von Reputationssystemen . . . 12
2.3.1 Eigenschaften . . . 12
2.3.2 Vorteile . . . 14
2.4 Auswertung von Reputationssystemen . . . 16
2.4.1 Simulationen . . . 17
2.4.2 Existierende Datenquellen . . . 17

3 Konzept der Auswertungsplattform **19**

3.1 Studiendesign . . . 20
3.1.1 Gesprächsrunde zur Meinungsbildung . . . 21
3.1.2 Bewertungen während der Gesprächsrunde . . . 21
3.1.3 Dilemma als Diskussionsgrundlage . . . 22
3.1.4 Subjektive Rangfolgen als Referenzdaten . . . 22
3.1.5 Auswertung der Nutzerstudie . . . 23
3.2 Webbasierte Auswertungsplattform . . . 23
3.2.1 Vorbereitungsphase des Studienleiters . . . 24
3.2.2 Vorbereitungsphase der Probanden . . . 27
3.2.3 Durchführungsphase des Studienleiters . . . 27
3.2.4 Durchführungsphase der Probanden . . . 28
3.2.5 Abschlussphase . . . 29
3.3 Zusammenfassung . . . 30

4 Implementierung der Auswertungsplattform **31**
4.1 Datenmodell . . . 31
4.2 Phasenmodell . . . 33
4.2.1 Studienleiter Panel . . . 33
4.2.2 Technische Realisierung . . . 34
4.3 WebSocket Protokoll . . . 34
4.4 Sichten des Studienleiters . . . 36
4.4.1 Studienverwaltung . . . 36
4.4.2 Auswertung . . . 40
4.5 Sichten der Probanden . . . 41
4.5.1 Persönliche Daten . . . 41
4.5.2 Chat . . . 42
4.5.3 Nachbewertungen . . . 43
4.5.4 Subjektive Rangfolge . . . 43
4.6 Fehlerbehandlung . . . 44
4.7 Zusammenfassung . . . 45

5 Resümee und Ausblick **47**
5.1 Nicht realisierte Funktionalitäten . . . 49
5.2 Ausblick . . . 49

Anhang A Datenbankmodell **51**

Abbildungsverzeichnis **53**

Tabellenverzeichnis **55**

Zur Arbeit gehörendes **57**

Literaturverzeichnis **59**

Kapitel 1

Einleitung

In den meisten erfolgreichen sozialen Netzwerken werden mittlerweile Reputationssysteme[1] eingesetzt. In einigen Anwendungsgebieten, wie Frage- und Antwortportalen oder Filesharingbörsen, können sie dabei helfen gewünschtes Verhalten zu motivieren [Jøs07] [TSP10] [Vas12]. Im Bereich E-Commerce sind Reputationssysteme ein unverzichtbarer Faktor des Erfolgs [Puj06]. Bei dem Onlineversandhaus eBay[2] sorgt das Reputationssystem für Vertrauen und Verlässlichkeit zwischen Käufern und Verkäufern [RZ02]. In Informationsportalen, wie zum Beispiel StackOverflow[3] oder gutefrage.net[4], soll es bei der Einschätzung der Glaubwürdigkeit und Expertise von Teilnehmern helfen. Im Lernumfeld, wie zum Beispiel bei *Backstage*[5] (vgl. Kapitel 1.1), wird mit Hilfe des Reputationssystems die Relevanz von Beiträgen ermittelt [BP14].

Übliche Reputationssysteme haben die Gemeinsamkeit, dass sie die Meinung von Anderen, meist in Form von numerischen Bewertungen, zu einem Reputationswert zusammenfassen [JIB07]. Allerdings ist diese Meinung stets an einen Kontext gebunden, in dem sie verschiedene Einschätzungen repräsentieren kann [Vas12], wie beispielsweise Glaubwürdigkeit, Vertrauenswürdigkeit, Relevanz oder Expertise. Daraus folgt, dass es kein einheitliches Reputationssystem gibt, dass für alle Anwendungen passend ist. In der Literatur werden zahlreiche Verfahren zur Berechnung von Reputationswerten in verschiedenen Anwendungsgebieten vorgeschlagen [JIB07]. Oft werden diese nur durch Simulationen evaluiert, in denen beispielsweise Bewertungen synthetisch erzeugt werden. Simulationen können wichtig sein, da sie meist mit geringem Ressourcenaufwand verwirklicht werden können und erste Erkenntnisse liefern. Trotzdem kann auf die Evaluation mit menschlicher Beteiligung nicht verzichtet werden, wenn die Wirkung von berechneten Reputationswerten auf diese untersucht werden soll. Folglich bietet es sich an Nutzerstudien zur Evaluation von Reputationsalgorithmen einzusetzen.

Der wesentliche Ansatz dieser Arbeit ist die Entwicklung einer webbasierten Auswertungsplattform zur Durchführung von Nutzerstudien. Dabei sollen Reputationswerte auf Plausibilität überprüft werden. Sie können als plausibel angesehen werden, wenn sie anwendungsbezogene Eigenschaften widerspiegeln.

[1] Als Reputationssystem wird die Berechnung und Darbietung von Reputationswerten verstanden
[2] http://www.ebay.com
[3] http://www.stackoverflow.com
[4] http://www.gutefrage.net
[5] http://backstage.pms.ifi.lmu.de/

Der Fokus der Auswertungsplattform liegt auf der Evaluation des Reputationsalgorithmus von *Backstage*. Dabei werden Referenzdaten mit Hilfe von subjektiven Einschätzungen von Probanden aufgestellt. Die subjektiven Einschätzungen werden auf der Basis einer Gesprächsrunde innerhalb eines Chats gebildet. Da das Reputationssystem von *Backstage* Bewertungen zur Berechnung von Reputationswerten benötigt, sollen diese während des Chats vergeben werden können. Im Studiendesign wird Wert darauf gelegt, möglichst vergleichbare Bedingungen zu schaffen, wie sie bei *Backstage* vorgefunden werden.

1.1 Backstage

Das Forschungsumfeld dieser Arbeit umfasst die anwendungsspezifische *Backchannel Software Backstage* zur Verbesserung der Qualität und Interaktivität von Massenvorlesungen [BP14]. Als *Backchannel* wird computergestützte Kommunikation verstanden, die während einer Präsentation parallel ausgeübt werden kann [Atk09].

Massenvorlesungen der Größenordnung von mehreren Hundert Personen haben den Nachteil, dass sich die Zuhörer meist passiv verhalten [BP14]. Dadurch kann die Effektivität der Vorlesung beeinträchtigt werden [For03]. Auf eine Beteiligung mittels Wortmeldungen wird oftmals verzichtet, da dies als aufdringlich aufgefasst und den Fortlauf der Vorlesung stören kann [BP14].

Backstage bietet die Möglichkeit, sich auf eine möglichst nicht störende Art und Weise in die Vorlesung einzubringen. Eine Alternative zu Wortmeldungen sind Beiträge, die auf Vorlesungsfolien als Frage, Antwort oder Hinweis platziert werden können. Genauer sollen die Beiträge an der Stelle der Vorlesungsfolie gesetzt werden, zu der sie sich beziehen. Durch diesen anwendungsspezifischen Bezug sollen themafremde oder unpassende Beiträge vermieden werden. Damit sich die Teilnehmer trauen am System teilzuhaben und sich somit die Beteiligung erhöht, können Beiträge in *Backstage* anonymisiert veröffentlicht werden. Außerdem ist ein privater Nachrichtenaustausch möglich [BP14].

Die Teilnehmer können Feedback zu Beiträgen geben, indem sie diese positiv oder negativ bewerten. Damit geben sie an, ob sie mit dem Beitrag einverstanden sind beziehungsweise ob der Betrag aktuell relevant ist. Infolge dessen können die Beiträge mit der meisten Relevanz herausgefiltert und gewinnbringend für die Vorlesung genutzt werden [BP14].

Auf der Basis von Bewertungen seiner Beiträge erlangt ein Teilnehmer Reputation. Bei *Backstage* bedeutet ein hoher Reputationswert, dass der Teilnehmer zahlreiche nützliche und relevante Beiträge verfasst hat, die häufig von anderen Teilnehmern positiv bewertet wurden [BP14].

1.2 Gliederung

Im folgenden Kapitel 2 wird der Begriff Reputation von ähnlichen Begriffen, wie Trust und Status, abgegrenzt. Weiter wird der Forschungsstand rund um das Thema Reputation und Reputationssysteme beleuchtet. Es wird diskutiert, welche Rolle sie im Internet spielen, welche Vorteile sie im Einzelfall haben, die gängigen Ansätze der Reputationsberechnung und mögliche Probleme.

Kapitel 3 beschäftigt sich mit dem Konzept der webbasierten Auswertungsplattform. Dabei wird zuerst auf das Studiendesign eingegangen. Danach wird die Umsetzung des Studiendesigns beschrieben, die aus der Vorbereitungsphase, der Durchführungsphase und der Auswertungsphase besteht.

Anschließend werden in Kapitel 4 die Aspekte der Implementierung erläutert. Dazu gehört das Datenmodell, das Phasenmodell, die Sichten von Probanden und Studienleiter und die Behandlung technischer Probleme. Des Weiteren wird die Implementierung eines Chats über WebSockets erläutert.

Im abschließenden Kapitel 5 werden die Ergebnisse dieser Arbeit zusammengefasst und mögliche Anwendungsfälle der Auswertungsplattform vorgeschlagen, die im Kontext dieser Arbeit interessant sind.

Kapitel 2

Forschungsstand

Gegenstand dieses Kapitels ist das Forschungsumfeld rund um Reputationssysteme und wie sie ausgewertet werden können. Zunächst wird der Unterschied zwischen Reputation und ähnlichen Begriffen, wie Trust und Status, abgegrenzt. Anschließend werden Merkmale und mathematische Modelle von Reputationssystemen im Internet vorgestellt. Schließlich wird auf existierende Ansätze zur Auswertung von Reputationssystemen eingegangen.

2.1 Begriffsabgrenzung

In der Literatur herrscht nicht immer Einigkeit bezüglich der Begriffe Reputation, Trust[6] und Status [JIB07]. Oft werden Reputation und Status miteinander vermischt [Vas12]. Einige Autoren, wie zum Beispiel *Tang* [TSP10] benutzen Reputation und Trust im gleichen Kontext. Amazon[7] bezeichnet die Anzahl von Rezensionen als Reputation, wobei es laut der in dieser Arbeit verwendeten Definitionen eher dem Statusbegriff zuzuordnen ist [Vas12].

Reputationswerte reflektieren die Messung der anwendungsbezogenen kollektiven Meinung über eine Entität in einer Gemeinschaft [WV07]. Im Gegensatz dazu reflektieren Trustwerte die persönliche subjektive Meinung eines Individuums über eine Entität [WV07]. Vereinfacht gesprochen entsprechen Reputationswerte der Meinung von Allen während Trustwerte die Meinung des Einzelnen repräsentieren. Infolge dessen kann Trust als eine spezielle Form der Reputation angesehen werden.

Ferner ist Trust einerseits die Beurteilung, ob jemand oder etwas vertrauenswürdig ist, und andererseits die tatsächliche Entscheidung, zu vertrauen. Nach *Jøsang* spielen Trustwerte eine wichtigere Rolle als Reputationswerte, wenn es darum geht eine Entscheidung für oder gegen etwas zu treffen, wie zum Beispiel bei Käufen in Onlineversandhäusern [JIB07].

Während bei Reputations- und Trustsystemen andere Personen beziehungsweise die Gemeinschaft involviert ist, kann der Status einer Person selbstständig durch diese beeinflusst werden [Vas12].

[6] Gemeint ist nicht Vertrauen, sondern ein Index, der die Messung von Vertrauen symbolisiert.
[7] http:// www.amazon.de

Im Folgenden wird auf die Begriffe Reputation, Trust und Status näher eingegangen.

2.1.1 Reputation

Reputation ist die Meinung oder der Glaube der Allgemeinheit über eine Person oder Sache [Jøs07]. Ein Reputationssystem sammelt Informationen, die Aufschluss über die Meinung der Anderen geben können und leitet davon einen Reputationswert ab, der von der Allgemeinheit eingesehen werden kann. Nach *JØSANG* kann Reputation folgendermaßen definiert werden:

> **Reputation (Definition)** [Jøs07, Seite 6]
> Reputation is „the overall quality or character as seen or judged by people in general."

2.1.2 Trust

Obwohl in der Literatur des öfteren von Trust die Rede ist, wird dieser Begriff auf vielerlei Art und Weise benutzt [MC96]. *JØSANG* definiert Trust mittels einer Unterteilung in *Reliability Trust* und *Decision Trust*, die folgendermaßen formuliert wird:

> **Reliability Trust (Definition)** [Jøs07, Seite 4]
> „Trust is the subjective probability by which an individual, A, expects that antother individual, B, perform a given action on which its welfare depends."

> **Decision Trust (Definition)** [Jøs07, Seite 5]
> „Trust is the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible."

Eine Veranschaulichung dieser beiden Aspekte von Trust bietet sein Seilbeispiel in Abbildung 2.1. Auf dem Bild ist ein Seil dargestellt, dass in einem schlechten Zustand ist und somit kein *Reliability Trust* verspricht. Zu dieser Einschätzung würden üblicherweise die meisten Menschen kommen, die schon mal ein Seil in der Hand hatten. *Decision Trust* hängt jedoch von der zugrunde liegenden Situation ab. In einer Feueralarmübung würde man die Entscheidung treffen das Seil nicht zu benutzen. Bei einem echten Feueralarm gäbe es keine andere Wahl als das Seil zu verwenden.

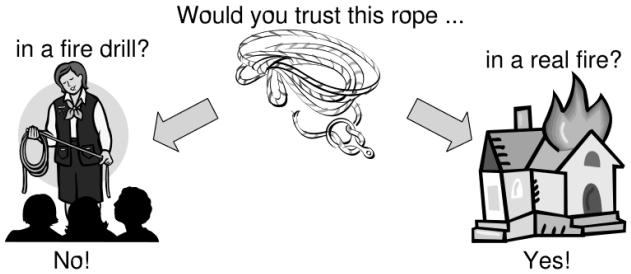

Abbildung 2.1: Gleicher *Reliability Trust*, unterschiedlicher *Decision Trust* (Bildquelle [Jøs07])

Eine weitere Möglichkeit Trust zu gewährleisten ist Transitivität, deren Prinzip in Abbildung 2.2 illustriert ist. Wenn Alice Bob vertraut und Bob Claire vertraut, vertraut

Alice auch Claire aufgrund der Empfehlung von Bob. Allerdings muss bei transitivem Trust beachtet werden, dass keine Trust-Kaskaden gebildet werden, da sonst Abhängigkeiten und Schleifen entstehen können. Angenommen es liegt folgende Situation vor. A vertraut B, B vertraut C und C vertraut D. Daraus resultiert nicht, dass A Vertrauen zu D hat. Würde D auch noch A vertrauen, so entsteht eine Schleife bei der jeder jedem vertraut [Jøs07]. In der Literatur gibt es zahlreiche Vorschläge für transitive Trust- und Reputationssysteme [TSP10]. Ein wichtiger Aspekt dabei ist die Überlegung, ob direkter (vgl. trust auf Abbilung 2.2) und indirekter (vgl. derived trust auf Abbilung 2.2) Trust das gleiche Gewicht haben sollten.

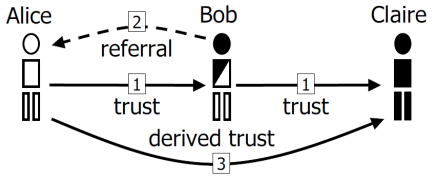

Abbildung 2.2: Prinzip des transitiven Vertrauens (Bildquelle [Jøs07])

Darüber hinaus können Reputationswerte eine Grundlage für Trust schaffen, insbesondere in Onlineversandhäusern. Ein hoher Reputationswert resultiert demzufolge in einer hohen Vertrauenswürdigkeit für zukünftige Transaktionen [JIB07].

Außerdem muss beachtet werden, dass Vertrauensbeziehungen auf dem gleichen Kontext basieren sollten, damit sie gültig bleiben [WV07]. Angenommen A vertraut B in Bezug auf sein Fußballwissen, da B Fußballtrainer ist. Daraus folgt nicht, dass A Vertrauen in B hat, wenn es zum Beispiel um Literatur geht.

Das Ergebnis von Vertrauenswürdigkeit kann das Eingehen einer Abhängigkeit mit der anderen Partei sein. Im Beispiel des eBay-Verkäufers würde der Käufer auf der Basis seines Vertrauens etwas kaufen und bezahlen. Damit ist der Käufer vom Verkäufer dahingehend abhängig, ob dieser die geschuldete Leistung tatsächlich erbringt.

Wenn es um *Decision Trust* geht, wird das Ganze aus einer allgemeineren Perspektive betrachtet. Hierbei ist die Bereitschaft entscheidend eine Abhängigkeit einzugehen unabhängig von der Vertrauenswürdigkeit der anderen Partei. Ziehen wir das Beispiel des eBay Verkäufers heran. Angenommen ein unbekannter Verkäufer hat als einziger eine Ware anzubieten, die der Käufer dringend braucht. Obwohl der Käufer keinen Grund hat dem Verkäufer zu vertrauen, wird er möglicherweise trotzdem die Entscheidung treffen das Risiko einzugehen und die Ware zu kaufen.

2.1.3 Status

Ein bestimmter Status kann im Gegensatz zu einem Reputations- und Trustwert ausschließlich durch eigene Handlungen, wie zum Beispiel Beiträge verfassen oder Dateien hochladen, erreicht werden [Vas12]. Reputations- und Trustwerte resultieren zwar auch aus eigenen Handlungen, jedoch besteht eine Abhängigkeit von anderen Parteien. Nach *Vassileva* kann Status folgendermaßen beschrieben werden:

> **Status (Definition)** [Vas12, Seite 183]
> By contrast to reputation „status can be earned by the user in isolation, by

performing certain actions.“

Der Status wird zum Beispiel häufig in Internetforen verwendet. Abhängig von der Anzahl der geschriebenen Beiträge bekommt jeder Autor einen Status zugewiesen, wie beispielsweise Visitor, Observer, Prototype, Decorator, Implementor und Builder. In der Abbildung 2.3 wird die Anzeige des Status im Forum die-informatiker.net[8] illustriert.

Ein weiteres Beispiel ist die Onlineplattform *Comtella*, die zum Austausch von akademischen Texten entwickelt wurde. Nutzer dieser Plattform werden für nützliche Aktionen mit Punkten belohnt. Das aufaddieren der Punkte ergibt dabei den Status Gold, Silber und Bronze. Je nach Status werden den Nutzern Privilegien innerhalb der Plattform gewährt. Untersuchungen dieser Plattform haben ergeben, dass durch den Status ein kurzzeitiger Anstieg der Beteiligung zu verzeichnen war. Die Schwachstelle von Anreizsystemen, die auf Status basieren, ist die fehlende Einbeziehung der Qualität von Beiträgen. Böswillige Nutzer könnten eine Fülle von qualitativ schlechten Beiträgen verfassen. Dadurch könnte der Effekt des *information overload* auftreten. Das bedeutet, dass Nutzer von zu vielen mangelhaften Beiträgen überschwemmt werden und ihre Beteiligung an der Plattform beenden [CV05].

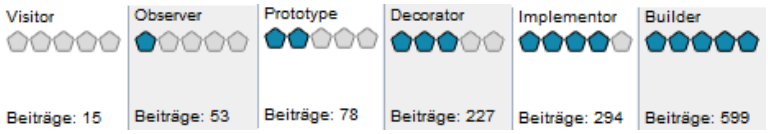

Abbildung 2.3: Übersicht der Status im Forum *die-informatiker.net*

2.2 Merkmale von Reputationssystemen

Reputationssysteme können nach der Art der Informationsgewinnung, -berechnung und -verteilung untersucht werden [VPZ14].

Als Informationsgewinnung wird das zugrunde liegende mathematische Modell und die benötigten Eingabedaten verstanden [VPZ14]. Eingabedaten sind Informationen, die zur Berechnung von Reputationswerten verwendet werden, üblicherweise numerische Bewertungen. Außer Bewertungen gibt es auch andere Informationsquellen auf deren Basis Reputation ermittelt werden kann, wie zum Beispiel der PageRank Algorithmus von *Google*[9]. Hierbei werden Hyperlinks benutzt, um die Wichtigkeit einer Webseite zu berechnen [PBMW99].

Die Informationsberechnung betrifft die technische Umsetzung des Algorithmus zur Reputationsberechnung [VPZ14]. Implementierungsdetails wie die verwendetem Programmierparadigmen oder Datenbankschemata spielen hier eine Rolle. Die technischen Aspekte, die im Rahmen dieser Arbeit von Interesse sind werden in Kapitel 4 dargestellt.

Zur Reputationsverteilung gehören Aspekte, die die Struktur des Reputationssystems betreffen [VPZ14]. Es sind Fragestellungen bezüglich der Verwaltung, Speicherung und

[8] http://www.die-informatiker.net/
[9] http://www.google.de

Darbietung von Reputationswerten von Interesse. Die Struktur von Reputationssystemen kann Zentral/ Dezentral (siehe Unterkapitel 2.2.2), Personen-basiert/ Ressourcenbasiert (siehe Unterkapitel 2.2.3) oder Global/ Personalisiert (siehe Unterkapitel 2.2.4) aufgebaut sein.

In den folgenden Unterkapiteln wird auf Metriken und Strukturen von Reputationssystemen näher eingegangen.

2.2.1 Metriken der Reputationsberechnung

Jedes Reputationssystem benötigt zur Berechnung von Reputationswerten eine Metrik. Dabei handelt es sich um ein mathematische Modell, um Bewertungsinformationen zu Reputationswerten zu aggregieren [VPZ14]. Da die jeweiligen Anforderungen an ein Reputationssystem sehr unterschiedlich ausfallen können, wurden verschiedene Metriken entwickelt [Vas12]. Während im Internethandel Vertrauen zwischen Käufer und Verkäufer von entscheidender Bedeutung ist, kommt es in Frage- und Antwortportalen vor allem auf die Qualität der Antworten an. In Internettauschbörsen sollen hingegen böswillige Nutzer identifiziert werden [VPZ14].

Die einfachste Möglichkeit zur Ermittlung von Reputationswerten aus Bewertungen ist die einfache Aufsummierung. Dabei wird die Differenz aller positiven und negativen Bewertungen eines Nutzers berechnet. Der Vorteil ist die einfache Berechenbarkeit und Verständlichkeit des Prinzips. Der Nachteil ist, dass die Aussagekraft von aufsummierten Bewertungen aus folgenden Gründen nur begrenzt ist. Wichtige Faktoren, wie zum Beispiel die Verlässlichkeit des Bewerters, werden nicht berücksichtigt. Es spielt keine Rolle, ob ein Bewerter lange an der Gemeinschaft teilnimmt oder neu ist. Jede Bewertung hat das gleiche Gewicht [JIB07] [VPZ14].

Einer der bekanntesten Vertreter, der Reputationswerte mittels einfacher Aufsummierung ermittelt, ist das Internetaktionshaus eBay[10]. Käufer und Verkäufer können sich gegenseitig nach jeder Transaktion positive, neutrale oder negative Bewertungen geben. Das Ergebnis davon wird neben dem Alias des Nutzers eingeblendet [RZ02] [WV07].

Dabei ist die Tatsache interessant, dass Bewertungen als Käufer und Verkäufer in die gleiche Reputationsberechnung aufgenommen werden [WV07]. Angenommen ein Nutzer erhält für fünfzig Käufe jedes Mal eine positive Bewertung. Wenn dieser Nutzer nun etwas verkauft, so wird potenziellen Käufern vorgegaukelt, dass es sich hierbei um einen zuverlässigen Verkäufer handelt. In Wirklichkeit ist keine Einschätzung über die Zuverlässigkeit mangels Bewertungen möglich.

Ein weiterer Nachteil bei eBay ist die mangelnde Robustheit gegenüber Täuschungsversuchen. Zwei Nutzer könnten sich gegenseitig so oft positiv bewerten, bis beide einen ungerechtfertigt hohen Reputationswert erreichen. Diese Problematik wird allerdings dadurch eingedämmt, dass jeder Transaktion für den Verkäufer Gebühren verursacht [WV07].

Dadurch, dass Nutzer Bewertungen genau nachverfolgen können, sind negative Bewertungen sehr selten. Die Befürchtung ist, dass sich negativ bewertete Nutzer mit einer

[10] http://www.ebay.com

negativen Bewertung revanchieren. Bei eBay sind nur ein Prozent aller Bewertungen negativ [RZ02].

Empirische Untersuchungen zeigen jedoch, dass es für die Zwecke des Internethandels ausreichend sein kann einfache Reputationssysteme einzusetzen, um Transaktionen zwischen Käufern und Verkäufern zu motivieren [RZ02]. Einen Käufer bei eBay interessieren naturgemäß hauptsächlich zwei Aspekte. Wie zuverlässig verschickt der Verkäufer die Ware und ob die Ware der Produktbeschreibung entspricht. Das Beispiel eBay zeigt, dass es dem Käufer ausreicht, wenn der Verkäufer eine ausreichende Anzahl an erfolgreichen Transaktionen vorweisen kann.

Ein etwas fortschrittlicherer Ansatz ist die Berechnung des Durchschnitts aller Bewertungen [RKZF00], wie es vom Online-Versandhaus Amazon[11] praktiziert wird. Es können Bewertungen von einem bis fünf Sternen vergeben werden deren Durchschnitt die Gesamtbewertung des jeweiligen Produkts ergibt. Eine weitere Verbesserung dieses Ansatzes kann erzielt werden, wenn Faktoren wie zum Beispiel das Alter oder die zugrunde liegende Thematik der Bewertung berücksichtigt werden. Dies würde einen gewichteten Durchschnitt der Bewertungen ergeben.

Die fortschrittlicheren Metriken können mit dem Begriff *Soft Computing* zusammengefasst werden. Darunter werden Ansätze verstanden, in denen Präzision und Fakten eine untergeordnete Rolle spielen. Vielmehr wird Raum für Ungenauigkeiten, Ungewissheiten und Schätzungen gelassen. Dieses Vorgehen ist dem Vorbild des menschlichen Verstands nachempfunden [Zad94].

Im Detail können Bayes'sche Modelle, Belief Modelle, Fuzzy Modelle und Flow Modelle zu dieser Kategorie gezählt werden [JIB07]. Bayes'sche Modelle verwenden die Betaverteilung, um aus binären Bewertungen Reputationswerte abzuleiten. Belief Modelle basieren auf der Evidenztheorie, in der Raum für Ungewissheit gelassen wird. Als Vertreter des Fuzzy Konzepts kann das REGRET Reputationssystem von *Sabater und Sierra* genannt werden. Sie unterteilen Reputation in die drei Dimensionen individuelle Reputation, soziale Reputation und kontextabhängige Reputation. Dabei gilt die individuelle Reputation als verlässlichste. Stehen zu wenige Informationen der individuellen Reputation zu Verfügung, wird auf die ungenaueren anderen beiden Reputationsarten erweitert [SS02]. Ein bekanntes Beispiel für Flow Modelle ist der PageRank Algorithmus von Google [PBMW99]. Dabei wird grob gesprochen die Wichtigkeit einer Webseite anhand der Qualität und Quantität von ankommenden Hyperlinks bestimmt. PageRank kann als Reputationssystem angesehen werden, da Hyperlinks als Meinung der Anderen interpretiert werden kann. Eine Verlinkung auf eine Webseite wird als positive Bewertung interpretiert [JIB07].

2.2.2 Zentrale vs. Dezentrale Architekturen

Reputationssysteme lassen sich in zentrale und dezentrale Architekturen aufteilen. Bei zentralen Architekturen existiert eine zentrale Sammelstelle, die alle Bewertungsinformationen entgegennimmt [WV07] [JIB07]. Auf Abbildung 2.4 ist dieser Sachverhalt unter a) dargestellt. Die Bewertungen (Ratings) beziehen sich auf die aktuelle Transaktion, die eine anwendungsspezifische Beziehung zwischen Parteien darstellt. Die zentrale

[11] http://www.amazon.com

Sammelstelle nimmt alle Bewertungen auf, leitet daraus anhand eines Berechnungsalgorithmus Reputationswerte für jeden Teilnehmer ab und veröffentlicht diese. Mit Hilfe der Reputationswerte können die Teilnehmer entscheiden, ob sie eine Interaktion eingehen möchten [JIB07]. (vgl. b) in Abbildung 2.4) Gegenwärtig kommen in den meisten existierenden Reputationssystemen zentrale Architekturen zum Einsatz [IBJR03] [WV07]. Das liegt daran, dass sie eine überschaubare Komplexität aufweisen und somit gut implementiert werden können. Auf der anderen Seite wird ein leistungsstarker Server benötigt, auf dem die zentrale Sammelstelle eingerichtet werden kann [WV07].

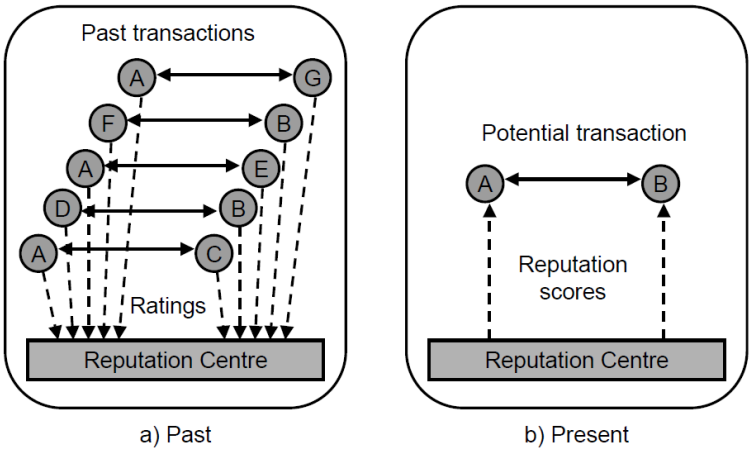

Abbildung 2.4: Veranschaulichung zentral gesteuerter Reputationssysteme (Bild aus [JIB07], Seite 14)

Eine andere Herangehensweise wird bei dezentralen Architekturen angewandt. Jeder Teilnehmer der Gemeinschaft sammelt Erfahrungen über Transaktionen mit anderen Teilnehmern, wie in Teil a) von Abbildung 2.5 veranschaulicht. Auf Anfrage stellt er diese Informationen zur Verfügung. Wenn Teilnehmer A mit Teilnehmer B eine Transaktion eingehen möchte, so muss A alle anderen Teilnehmer suchen, die mit B bereits direkten Kontakt hatten. (vgl. b) in Abbildung 2.5) Von diesen Teilnehmern werden die Bewertungen über B angefragt und anhand eines Berechnungsalgorithmus ein Reputationswert berechnet. Verteilte Architekturen kommen oftmals in Filesharing Netzwerken zum Einsatz [JIB07].

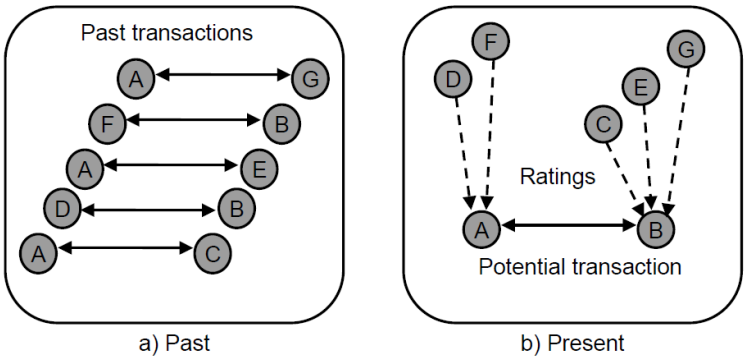

Abbildung 2.5: Veranschaulichung verteilter Reputationssysteme (Bild aus [JIB07], Seite 15)

2.2.3 Reputation von Personen vs. Ressourcen

Eine weitere Möglichkeit Reputationssysteme zu unterscheiden ist der Bezug der Reputationswerte auf Personen oder Ressourcen [WV07]. In Frage- und Antwortportalen ist es zum Beispiel praktikabel Reputationswerte von Nutzern zu ermitteln. Beiträge, die in diesem Fall als Ressourcen angesehen werden können, werden üblicherweise nur von einer kleinen Interessengruppe von Nutzern gelesen und verlieren oft mit der Zeit an Relevanz. Aus diesem Grund und wegen der meist hohen Anzahl von Beiträgen macht es keinen Sinn Reputationswerte für Beiträge zu ermitteln. Im Gegensatz dazu macht es Sinn Reputationswerte für Nutzer zu ermitteln, da sie sich fortführend an der Gemeinschaft beteiligen.

Auf der anderen Seite kann es zum Beispiel in Onlineversandhäusern sinnvoll sein Reputationswerte von Ressourcen zu ermitteln. In diesem Fall werden Waren oder Dienstleistungen als Ressourcen angesehen. In Onlineversandhäusern ist es typisch, dass die gleichen Waren mehrfach verkauft werden und somit die ganze Zeit relevant bleiben. Potenzielle Kunden können anhand von geeigneten Reputationsmechanismen gleichartige Produkte miteinander vergleichen.

2.2.4 Globale vs. Personalisierte Reputation

Die dritte Unterscheidungsmöglichkeit bezieht sich auf Reputationsinformationen, die für die Reputationsberechnung herangezogen werden. Sie können auf der Basis der Meinung der Gesamtheit (Global) oder einer bestimmten Gruppe (Personalisiert) beruhen. Während global berechnete Reputationswerte von allen Mitgliedern der Gemeinschaft eingesehen werden können, sind personalisierte Reputationswerte nur aus der jeweiligen Perspektive verfügbar [WV07].

2.3 Eigenschaften und Vorteile von Reputationssystemen

Es gibt eine Reihe von wünschenswerten Eigenschaften beziehungsweise Anforderungen, die bei Reputationssystemen relevant sind. Je nach Anwendungsgebiet kann die Wichtigkeit einzelner Eigenschaften variieren. Außerdem bieten Reputationssysteme viele Vorteile, sofern sie passend zur Umgebung modelliert werden. In diesem Kapitel werden Eigenschaften und Vorteile von Reputationssystemen dargeboten und erklärt.

2.3.1 Eigenschaften

Nachfolgend wird eine Auflistung einiger relevanter Eigenschaften von Reputationssystemen nach *Vavilis* dargeboten und im Folgetext erläutert [VPZ14]:

- Plausibilität der Reputationswerte
- Themenspezifische Reputationsbewertung
- Resistenz gegen Manipulation
- Beeinflussung der eigenen Reputationswerte
- Einfluss der Informationsmenge
- Berücksichtigung des Informationsalters

- Behandlung neuer Nutzer
- Implementierbarkeit des Reputationsalgorithmus

Plausible Reputationssysteme sind in der Lage Einschätzungen über gewünschte Eigenschaften von Nutzern, wie Vertrauenswürdigkeit, Glaubwürdigkeit, Expertise oder Relevanz, sinnvoll zu reflektieren [WV07]. Mangelnde Plausibilität kann am Beispiel von eBay verdeutlicht werden. Wenn ein Verkäufer 10.000 positive Bewertungen bekommen hat und ein zweiter 20.000, so müsste der zweite als deutlich vertrauenswürdiger eingeschätzt werden. In Wirklichkeit ist das nicht der Fall, denn beide weisen eine hohe Anzahl an erfolgreichen Transaktionen auf und würden ähnlich vertrauenswürdig eingestuft werden [RZ02]. Damit wird die Einschätzung von potenziellen Käufern in Bezug auf die Vertrauenswürdigkeit der beiden Verkäufer nicht sinnvoll von den Reputationswerten reflektiert.

Ein weiteres Merkmal eines plausiblen Reputationssystems ist die Stimmigkeit von Reputationswerten. Das bedeutet, dass das Zustandekommen eines Reputationswertes annähernd reproduzierbar beziehungsweise wiederholbar sein soll. Wenn sich zwei verschiedene Nutzer in einem System ähnlich verhalten, so sollten ihnen vergleichbare Reputationswerte zugeordnet werden.

Als Beispiel für ein plausibles Reputationssystem kann PageRank von Google genannt werden. Laut der Webseite seo-united.de[12] hat die Google-Suchmaschine einen Marktanteil von 95 Prozent. Daraus kann man schließen, dass die Google-Suchmaschine von den meisten Nutzern als plausibel angesehen wird.

Reputationswerte sollten sich auf eine bestimmte Thematik beziehen [VPZ14]. Besonders in Frage- und Antwortportalen mit unterschiedlichen Themengebieten, wie zum Beispiel gutefrage.net[13], ist es wünschenswert Bewertungen themenspezifisch zu behandeln. Bei gutefrage.net werden Nutzer zu sogenannten *Experten der Community*, wenn sie in einem Themenbereich besonders gut bewertete Antworten verfasst haben.

Des Weiteren können themenspezifische Techniken genutzt werden, um Web Spam[14] besser bekämpfen zu können. Eine Möglichkeit ist der *TrustRank* Algorithmus, der auf Linkanalyse basiert. Dabei handelt es sich um einen halb-automatisierten Algorithmus, bei dem im ersten Schritt einige wenige gute Webseiten manuell selektiert werden. Anhand der Verlinkungsstruktur des Internets werden aus den selektierten Seiten weitere gutartige Webseiten abgeleitet [GGMP04]. *Wu, Goel und Davison* zeigen, dass TrustRank erheblich bessere Ergebnisse liefert, wenn im ersten Schnitt themenspezifische Informationen berücksichtigt werden [WGD06].

Eine weitere bedeutende Eigenschaft von Reputationssystemen ist die Resistenz gegen Täuschungsversuche [VPZ14]. Um eine Verfälschung von Reputationswerten zu vermeiden, sollte das System bösartiges Verhalten erkennen, egal ob dieses aus Versehen oder absichtlich entstanden sind. Böswillige Nutzer könnten mit gezielten Falschbewertungen versuchen Einfluss auf Reputationswerte von Kontrahenten zu nehmen

[12] http://www.seo-united.de/suchmaschinen.html abgerufen am 08.10.2015
[13] http://www.gutefrage.net
[14] Maßnahmen, damit Webseiten von Suchmaschinen besser bewertet werden als sie sollten [GGM05].

[SŠZ12]. Außerdem könnten sich mehrere Nutzer zusammenschließen, um durch gegenseitige Bewertungen ungerechtfertigt hohe Reputationswerte zu erlangen. Eine andere Manipulationsmöglichkeit wäre es, sich eine neue Identität anzulegen, um damit einen schlechten Reputationswert ungeschehen zu lassen.

Nachdem Reputationssysteme die Meinung der Anderen widerspiegeln sollten, kann ein Einfluss auf den eigenen Reputationswert nicht erlaubt werden. Dies beinhaltet sowohl direkten Einfluss auf den Reputationswert, die Reputationsberechnung oder auch indirekten Einfluss durch Selbstbewertungen [VPZ14].

Ein Reputationssystem sollte eine gewisse Mindestmenge an Bewertungen voraussetzen, um plausible Reputationswerte ermitteln zu können. Stehen zu wenig Bewertungen zu buche, so kann oft keine aussagekräftige Berechnung erfolgen [VPZ14].

Die Reputation kann sich mit Fortlauf der Zeit verändern, wenn Nutzer neue Transaktionen eingehen oder sich länger nicht mehr an der Gemeinschaft beteiligen. Üblicherweise sind aktuelle Transaktionen und Bewertungen bedeutender als weit in der Vergangenheit liegende. Beiträge könnten mit der Zeit veraltet oder irrelevant sein. In vielen Anwendungsfällen kann es sinnvoll sein diesem Umstand in die Reputationsberechnung einzubeziehen [WV07].

Eine weitere wichtige Eigenschaften ist die Behandlung neuer Nutzer. Auf der einen Seite sollte neuen Nutzern ermöglicht werden relativ schnell einen guten Reputationswert zu erreichen und sich somit in die Gemeinschaft zu integrieren. Auf der anderen Seite wäre es bedenklich neue Nutzer besser dastehen zu lassen als welche mit relativ schlechten Reputationswerten.

Schließlich nutzt der beste Reputationsalgorithmus nichts, wenn er nicht technisch verwirklicht werden kann. Daher muss die Komplexität der Reputationsberechnung an die zur Verfügung stehenden Ressourcen, wie zum Beispiel Server, angepasst werden.

2.3.2 Vorteile

Reputationssysteme können sich auf vielseitige Art und Weise positiv in computergestützten Umgebungen, wie zum Beispiel *Backstage* oder in zahlreichen Internetanwendungen, auswirken. Eine Übersicht der Vorteile von Reputationssystemen kann der Abbildung 2.6 entnommen werden und wird im Folgenden genauer erklärt.
Nach *Vassileva* [Vas12] können Reputationssysteme in Onlinegemeinschaften die Aktivität von Nutzern anregen. Durch das Einführen von Privilegien und den Einsatz von spielähnlichen Mechanismen in nicht-spielerischem Kontext, *Gamification* genannt, können Nutzer motiviert werden Beiträge zu verfassen oder Beiträge anderer zu bewerten. Ein gutes Beispiel für reputationsbedingte Privilegien bietet *StackOverflow*[15], ein Onlineforum für Computerprogrammierung. Hier wird Reputation sowohl aus der Meinung anderer als auch durch eigene Aktivität ermittelt. Mit steigender Reputation bekommt man zum Beispiel die Möglichkeit Bilder in eigene Beiträge einzufügen, einen Beitrag eines anderen Nutzers positiv zu bewerten oder als beste Antwort zu markieren. Ein positiver Seiteneffekt von hoher Aktivität ist eine genauere Reputationsberechnung, da sie auf der Grundlage der Bewertungen ausgeführt wird. Je mehr

[15] http://www.stackoverflow.com

Vorteile	Anreiz zur Aktivität	Anreiz zur Qualität	Anreiz zur Kooperation	Ausgleich asynchroner Information	Möglichkeit zu Filtern
Begründung	- Privilegien - Gamification	Streben nach positiven Bewertungen	Vermeiden von negativen Bewertungen	nach Literatur (Spreman & Josang)	Filtern auf der Basis von Bewertungen
Folge	- mehr Beiträge - mehr Bewertungen - genauere Reputationswerte	Qualitativ hochwertige Beiträge	Schutz vor Spam und aggressiven Beiträgen	Vertrauen unter Fremden	- schnellerer Zugang zu wichtigen Inhalten - Aussortieren von Doppelten Inhalten

Abbildung 2.6: Vorteile von Reputationssystemen

Beiträge, desto mehr Bewertungen. Je mehr Bewertungen, desto akkurater die berechneten Reputationswerte.

Die Qualität von Beiträgen kann mit Hilfe von Reputationssystemen ebenfalls erhöht werden. Nachdem die Reputation eines Nutzers Rückschlüsse auf seine Verlässlichkeit und Fachkompetenz ermöglicht [HYD09], wird grundsätzlich das Erreichen eines hohen Reputationswertes angestrebt. Infolge dessen werden Nutzer motiviert ein hohes Maß an Sorgfalt beim Verfassen von Beiträgen zu gewährleisten um möglichst positive Bewertungen zu erhalten.

Außer der Quantität und der fachlichen Qualität der Beiträge kann sich Reputation positiv auf sozial erwünschtes Verhalten auswirken [Jøs07]. Zur Vermeidung von drohenden schlechten Bewertungen werden Nutzer Abstand davon nehmen themafremde oder beleidigende Beträge zu verfassen. *Jøsang* spricht in diesen Zusammenhang von *Soft Security*:

> **Soft Security (Definition)** [Jøs07, Seite 3]
> „Soft security is the collaborative enforcement of, and adherence to common ethical norms by participants in a community."

Typischerweise interagieren Parteien beziehungsweise Personen miteinander, die sich nicht kennen [JIB07]. Wenn beispielsweise ein Student in einer Massenvorlesung eine Frage stellt und ein anderer sie beantwortet, so ist nicht bekannt ob die Antwort verlässlich ist. Es ist keine Einschätzung über die Fachkundigkeit des antwortenden Studenten möglich. Analog dazu verhält es sich in Transaktionen bei Onlineversandhäusern. Der Käufer hat initial keinerlei Informationen darüber, wie verlässlich und zuverlässig der Verkäufer ist. Daraus ergibt sich, dass der Wissensstand von Parteien über eine bestimmte Angelegenheit ungleich verteilt ist, auch *asymmetrische Information* genannt [JIB07]. Im Falle der Studenten trägt der fragende Student das Risiko eine falsche Antwort zu bekommen, während der Käufer in einem Onlineversandhaus nach dem übersenden des Geldes auf die Fairness des Verkäufers angewiesen ist.

Nach *Jøsang* [JIB07] und *Spremann* [Spr88] können Reputationssysteme dabei behilflich sein die *asymmetrische Informationsverteilung* zu verringern. Anders gesprochen können sie dabei helfen Vertrauen zwischen Fremden aufzubauen, indem vergangenes Verhalten als Anhaltspunkt für zukünftiges herangezogen wird [Puj06]. *Spremann* spricht in diesem Kontext von *Qualitätsunsicherheit* und geht noch einen Schritt weiter.

Die ausgleichende Wirkung von Reputation in Bezug auf *asymmetrische Information* ist umso größer, je weniger Informationen vorliegen [Spr88]. Nach *Pujari* kommt der Internethandel zwischen fremden Parteien ausschließlich mit der Unterstützung von Reputationssystemen zustande, da sonst Vertrauen der Parteien zueinander fehlen würde. [Puj06].

Zusätzlich zu ihrer Funktion als Grundlage der Berechnung von Reputationswerten, können Bewertungen zum Filtern von Inhalten verwendet werden, beispielsweise in Frage- und Antwortportalen. Durch die Benutzung eines bewertungs-basierten Reputationssystems müssen Bewertungen nicht extra zum Filtern gesammelt werden. Aufgrund der Fülle von Daten ist es oft schwierig passende Informationen zu finden [HYD09]. In Frage- und Antwortportalen bekommt der Fragende oft sehr viele verschiedene Antworten. Alle Antworten genau durchzulesen kann gegebenenfalls unpraktikabel sein. Wenn die besten Antworten von anderen Nutzern positiv bewertet werden, so können diese ganz oben in der Antworthistorie angezeigt werden. Wenn der Fragesteller die für ihn beste Antwort identifiziert hat, kann er sie markieren. Damit wird zukünftigen Nutzern, die die gleiche Frage haben, schnell und effizient geholfen, da sie die beste Antwort sofort sehen. Außerdem werden doppelte und wenig hilfreiche Beiträge automatisch aussortiert, indem sie in der Antworthistorie nach unten rutschen. Diese Vorgehensweise wird beispielsweise von *StackOverflow* praktiziert.

Ein weiteres Beispiel, wie Bewertungen zum Filtern von Inhalten genutzt werden können, bietet *Backstage*. In Massenvorlesungen verfassen potenziell viele Studenten zahlreiche Beiträge beziehungsweise Fragen. Der Dozent kann unmöglich auf alle Fragen eingehen, da dies den zeitlichen Rahmen der Vorlesung sprengen würde. Infolge dessen könnten die am besten bewerteten und somit relevantesten Fragen an den Dozent weitergeleitet werden [BP14].

2.4 Auswertung von Reputationssystemen

Dieses Unterkapitel beschäftigt sich mit der Frage, welche Methoden bisher vorgeschlagen wurden, um Reputationsalgorithmen auszuwerten. Grundsätzlich wird für die Auswertung eines Reputationsalgorithmus ein Graph benötigt, der Informationen über die Interaktionen zwischen Nutzern enthält [TSP10]. Diese Informationen können beispielsweise in Form von Bewertungen zwischen Nutzern oder einem erfolgreichen oder nicht erfolgreichem Ausgang einer Interaktion vorliegen. Außerdem können Referenzdaten und ein Vergleichsmaß benötigt werden. Referenzdaten werden für den Vergleich mit den berechneten Reputationswerten des zu testenden Reputationsalgorithmus benötigt. Folglich muss bei Referenzdaten sichergestellt sein, dass sie von einer vertrauenswürdigen Quelle stammen. Das Vergleichsmaß wird durch eine Metrik repräsentiert, mit der die Referenzdaten mit den berechneten Reputationswerten des zu testenden Reputationsalgorithmus verglichen werden. Im Folgenden werden zwei Lösungen zur Auswertung von Reputationsalgorithmen dargeboten, und zwar Simulationen [TSP10][YS02] und Auswertung mit Hilfe von existierenden Datenquellen [PSD02][HYD09].

2.4.1 Simulationen

Bei dieser Art der Auswertung wird sowohl ein Graph als auch die Referenzdaten wie folgt simuliert. Zunächst werden Nutzertypen generiert, die in mehreren Durchgängen miteinander interagieren. Der Erfolg einer Interaktion zwischen zwei Nutzern hängt vom Nutzertyp ab, der durch zufällig zugeteilte Wahrscheinlichkeitswerte repräsentiert wird. Diese Wahrscheinlichkeitswerte fungieren als Referenzdaten. Anschließend wird die Simulation eine gewisse Anzahl von Runden durchgeführt, in denen jeder Nutzer eine Interaktion mit einem zufällig ausgewählten anderen Nutzer eingeht. Durch Aggregation der Interaktionen zwischen Nutzern kann ein Graph abgeleitet werden, mit dessen Hilfe Reputationswerte des zu testenden Reputationsalgorithmus berechnet werden können. Im letzten Schritt der Auswertung werden die berechneten Reputationswerte mit den Nutzertypwerten verglichen. Der Vergleich erfolgt meist mit Metriken der Korrelationsberechnung. Die eben beschriebene Vorgehensweise der Auswertung wird zum Beispiel von *Tang u.a.* [TSP10] und *Yu und Singh* [YS02] angewandt.

Ein Problem bei der Auswertung mittels Simulationen könnte die zufällige Ermittlung der Nutzertypen, auf der das gesamte Konzept aufbaut. Die reale Verteilung von Nutzertypen ist unbekannt und wird durch eine stochastische Verteilung nur geschätzt. Dadurch können unglaubwürdige Resultate entstehen.

2.4.2 Existierende Datenquellen

Ein anderer Ansatz der Auswertung ist die Zuhilfenahme existierender Daten, um einen Graph mit bewertenden Informationen oder Referenzdaten zu erhalten. Ein Beispiel dafür beschreibt *Pujol u.a.* [PSD02] wie folgt. Die digitale wissenschaftliche Bibliothek *CiteSeer*[16] wurde zur Auswertung von Reputationswerten herangezogen. Bei *CiteSeer* kann eingesehen werden, wie oft ein wissenschaftlicher Artikel zitiert worden ist. Wenn eine bestimmte Anzahl von Artikeln zufällig ausgewählt und eine Rangfolge anhand der Anzahl der Zitate erstellt wird, so kann diese Rangfolge nach der Auffassung *Pujol u.a.* als Referenzrangfolge angesehen werden, die nah an die wahren Reputationswerte herankommt [PSD02]. Grundsätzlich ist eine Plattform wie *CiteSeer* gut geeignet, um als Datenquelle verwendet zu werden, da Zitate zwischen Autoren als eine Art positive Bewertung aufgefasst werden können. Allerdings darf angezweifelt werden, ob eine Rangfolge aus der Anzahl der Zitate von Artikeln als Referenz dienen kann. Die Anzahl der Zitate eines Artikels kann ebenfalls dazu verwendet werden, um Reputationswerte entsprechend eines Reputationsalgorithmus zu berechnen. Schließlich können zu testende Reputationsalgorithmen mit der Referenzrangfolge von *CiteSeer* verglichen werden, indem der Korrelationskoeffizient berechnet wird. Zusätzlich dazu geben *Pujol u.a.* einen Vergleich zu den weit verbreiteten Algorithmen PageRank [PBMW99] und HITS [Kle99].

Ein weiteres Beispiel für die Verwendung von existierenden Datenquellen zur Auswertung von Reputationsalgorithmen ist die Auskunftsplattform *Yahoo!Answers*[17]. Darin können Fragen gestellt werden, die üblicherweise mehrere Antworten erhalten. Der fragende hat die Möglichkeit eine der Antworten als beste Antwort zu markieren [HYD09]. *Hong u.a.* haben rund 60.000 Beiträge auf *Yahoo!Answers* als Grundlage für die Auswertung herangezogen, davon rund 14.000 Fragen und rund 46.000 Antworten. Rund

[16] http://citeseerx.ist.psu.edu
[17] http://answers.yahoo.com

13.000 Beiträge wurden als beste Antworten markiert. Fragen ohne beste Antwort wurden nicht berücksichtigt [HYD09]. Anhand der Frage- und Antwortstruktur von *Yahoo!Answers* lässt sich ein Graph aufbauen, mit dem Reputationswerte der zu testenden Reputationsalgorithmen berechnet werden können. *Hong u.a.* vergleichen insgesamt vier Algorithmen, PageRank, SimpleRank, HITS und den von ihnen vorgeschlagenen Topical PageRank, wobei jeweils eine Rangfolge aus den Reputationswerten gebildet wird [HYD09]. Die 13.000 besten Antworten fungieren als Referenzdaten, indem untersucht wird, ob die beste Antwort einer Frage sich an einer oberen Position in der jeweiligen Rangfolge befindet.

Kapitel 3

Konzept der Auswertungsplattform

Der Grundgedanke dieser Arbeit ist die Auswertung von Reputationsalgorithmen mit Hilfe von Nutzerstudien und Fragebögen. Genauer sollen berechnete Reputationswerte auf Plausibilität überprüft werden. Im Kontext dieser Arbeit können sie als plausibel angesehen werden, wenn die Eigenschaften der Akzeptanz und Konsistenz erfüllt sind.

Die Akzeptanz ist nach *Hilbig* definiert als „eine mehr oder weniger zustimmende Einstellung eines Individuums oder einer Gruppe gegenüber einem Objekt, Subjekt oder sonstigem Sachverhalt“ [Hil84, Seite 120]. In dieser Arbeit wird als Akzeptanz eine zustimmende Einstellung von Probanden in Bezug auf berechnete Reputationswerte verstanden. Die Messung der Akzeptanz kann beispielsweise mit einem Fragebogen erfolgen. Die Fragen des Fragebogens sollen so ausgelegt sein, dass Rückschlüsse gemacht werden können, ob der Proband die Bereitschaft hat berechnete Reputationswerte mit deren Kontext zu verknüpfen.

Die Überprüfung auf Konsistenz beschäftigt sich mit der Fragestellung, ob berechnete Reputationswerte stimmig sind. Genauer soll herausgefunden werden, inwiefern sie mit der Meinung eines Menschen übereinstimmen und ob sich diese reproduzieren lässt. Es wird von der Hypothese ausgegangen, dass berechnete Reputationswerte eine umso höhere Konsistenz aufweisen, je mehr sie mit der Meinung von Probanden zusammenhängen. Die Messung der Konsistenz wird mittels einer vergleichenden Nutzerstudie mit Probanden als Kontrollgruppe realisiert.

Zur einfacheren und reproduzierbaren Durchführung von Nutzerstudien wird eine webbasierte Auswertungsplattform als technische Unterstützung entwickelt und realisiert. Durch den Einsatz von Datenspeicherungsstrukturen, wie zum Beispiel Datenbanken, werden Daten der Nutzerstudien automatisch gespeichert und ausgewertet. Mit Hilfe eines geeigneten Modells zum verwalten von Nutzerstudien sollen Rahmenbedingungen für gleichartige Durchläufe mehrerer Nutzerstudien geschaffen werden, deren Ergebnisse miteinander verglichen werden können. Eine webbasierte Modellierung der Plattform hat den Vorteil, dass Nutzerstudien über das Internet durchgeführt werden können und die Probanden somit nicht zwingend am gleichen Ort sein müssen.

Die Auswertungsplattform soll die Meinung von Probanden messen und mit den berechneten Reputationswerten vergleichen können. Daraus folgt, dass Probanden Repu-

tationswerte subjektiv angeben müssten. Hierbei wird von der Hypothese ausgegangen, dass es für Menschen natürlicher ist Rangordnungen anzugeben als einen numerischen Wert. Infolge dessen sollen in der Auswertungsplattform Rangfolgen und nicht Reputationswerte aufgestellt und verglichen werden. Durch einfaches sortieren lassen sich aus berechneten Reputationswerten Rangfolgen erzeugen.

Infolge dessen muss die Auswertungsplattform so modelliert werden, dass Bewertungen und subjektive Rangfolgen von Probanden gesammelt werden können. Die Bewertungen sind notwendig, da sie als Übergabeparameter für den auszuwertenden Reputationsalgorithmus gebraucht werden. Subjektive Rangfolgen dienen als Referenzdaten zur Messung der Konsistenz von berechneten Reputationswerten. Der Vergleich zwischen berechneten und subjektiven Rangfolgen kann durch ein geeignetes statistisches Verfahren durchgeführt werden, zum Beispiel mittels Berechnung des Rangkorrelationskoeffizienten [Liu14].

Im folgenden Kapitel 3.1 und den dazugehörigen Unterkapiteln wird das Studiendesign zur Gewinnung von Bewertungen und subjektiven Rangfolgen erläutert. Komplettiert wird das Konzept mit dem Kapitel 3.2, dass den konzeptuellen Aufbau der webbasierten Auswertungsplattform beschreibt.

3.1 Studiendesign

Ein möglicher Ansatz für das Studiendesign kann es sein, Probanden in einer Gesprächsrunde miteinander kommunizieren zu lassen und ihnen die Möglichkeit zu geben die Beiträge der anderen Probanden in Echtzeit zu bewerten. Damit hat die Gesprächsrunde zwei Funktionen, zum einen das Erzeugen von Bewertungen und zum anderen die Bildung eines Eindrucks über die Aussagen der Probanden. Der Eindruck, den jeder Proband über die anderen Teilnehmer gewinnt, wird Grundlage für die Aufstellung der subjektiven Rangfolge sein.

Das Thema der Gesprächsrunde wird vorgegeben. Um einen flüssigen Ablauf zu ermöglichen, muss ein Gesprächsthema gewählt werden, dass kontroverse Meinungen zulässt und somit genügend Gesprächsstoff bietet. Eine Möglichkeit dazu bieten Dilemmata, auf die im Abschnitt 3.1.3 eingegangen wird.

Nach dem Ende der Gesprächsrunde werden alle Probanden gebeten eine subjektive Rangfolge der anderen Probanden anzugeben. Diese Rangfolge soll sich nach den Gesamteindrücken richten, die ein Proband von den anderen Probanden während der Gesprächsrunde gebildet hatte. Nach der Angabe der subjektiven Rangfolge sind die Probanden mit der Nutzerstudie fertig, da sowohl Bewertungen als auch subjektive Rangfolgen gesammelt wurden, die für die Auswertung des vorliegenden Reputationsalgorithmus nötig sind.

Im Folgenden wird auf die Gesprächsrunde, Bewertungen, Diskussionsgrundlage subjektive Rangfolgen und die Auswertung genauer eingegangen.

3.1.1 Gesprächsrunde zur Meinungsbildung

Neben Ihrer eigentlichen Funktion als Lieferant für Bewertungen und Grundlage der Meinungsbildung über die anderen Teilnehmer der Nutzerstudie sollten weitere Kriterien der Gesprächsrunde beachtet werden. Bevor die Gesprächsrunde starten kann, müssen alle Probanden in Ruhe das vorgegebene Gesprächsthema kennenlernen, damit sie darüber diskutieren können. Die Dauer der Gesprächsrunde wird vom Studienleiter bestimmt und den Probanden beispielsweise mit einer Stoppuhr oder einem Timer dargeboten. So können die Probanden abschätzen, wie rasch und detailliert sie ihre Nachrichten verfassen sollten. Grundsätzlich orientiert sich die Dauer der Gesprächsrunde an der Anzahl der Probanden, an der Eignung des Gesprächsthemas und den Zielen der Auswertung. Einige Reputationsalgorithmen sind auf eine ausreichend große Anzahl der Bewertungen angewiesen, da die Reputationswerte sonst zu wenig Aussagekraft haben. Die letztendliche Dauer kann beispielsweise experimentell über mehrere Durchläufe der Nutzerstudie ermittelt oder abgeschätzt werden. Darüber hinaus kann der Studienleiter während der Gesprächsrunde jederzeit die Dauer verkürzen oder verlängern, wenn er dies für nötig hält. Allerdings muss im diesem Fall beachtet werden, dass die Vergleichbarkeit zwischen mehreren gleichen Durchläufen der Nutzerstudie mit initial gleicher Gesprächsdauer beeinträchtigt werden könnte.

Ebenso wie die Gesprächsdauer sollte das Thema der Gesprächsrunde zu jeder Zeit einsehbar sein. Damit wird den Probanden eine Stütze geboten, falls sie den Bezug zu dem Thema verlieren oder Details des Themas nochmal nachlesen wollen.

Außerdem sollte der Studienaufbau so gewählt werden, dass Anonymität gewährleistet ist. Dies kann beispielsweise mit Sichtschutzwänden erreicht werden, wenn sich alle Probanden in einem Raum befinden. Im Rahmen dieser Arbeit wird der Computer als technisches Hilfsmittel verwendet, um Anonymität sowohl in Bezug auf das Sehen als auch auf das Hören zu ermöglichen. Auf die genaue Umsetzung wird in Kapitel 4 eingegangen. Probanden sollen ihre Impressionen ausschließlich aus den Nachrichten gewinnen und nicht auf der Basis von Freundschaften, Sympathie oder sonstigen zwischenmenschlichen Einflüssen.

3.1.2 Bewertungen während der Gesprächsrunde

Während der Gesprächsrunde hat jeder Teilnehmer die Möglichkeit die Nachrichten der anderen Teilnehmer zu bewerten. Selbstbewertungen sind nicht erlaubt, da Reputation nach Definition aus der Meinung anderer gebildet werden soll [Jøs07]. Somit würden Selbstbewertungen das Ergebnis verfälschen.

Das Bewertungsschema wird vor Beginn der Gesprächsrunde vom Studienleiter festgelegt und wird währenddessen nicht mehr verändert, um die Integrität aufrecht zu erhalten. Es wäre denkbar positive und negative Bewertungen zu erlauben. Eine andere Möglichkeit wäre es zum Beispiel nur positive Bewertungen zuzulassen oder ein Bewertungsmodell mit mehr als zwei Optionen zu wählen. Die Wahl des Bewertungsschemas ist letztendlich stark von den Zielen der Auswertung abhängig.

Um die Aussagekraft von Bewertungen aufrechterhalten zu können, darf eine Person keine konträren Bewertungen zu einer Nachricht vergeben. Das heißt, dass eine bestimmte Nachricht nicht von der gleichen Person sowohl positiv als auch negativ bewertet wer-

den darf. Da Probanden mit Bewertungen implizit ihre Meinung kundtun, muss eine Präferenz erkennbar sein. Das wäre bei konträren Bewertungen nicht immer möglich.

Außerdem muss die Vergabe von Bewertungen analog zu den Ausführungen in vorigen Unterkapitel 3.1.1 anonym erfolgen, um Revanchebewertungen und zwischenmenschliche Einflüsse, wie Sympathien oder Freundschaften, zu vermeiden.

3.1.3 Dilemma als Diskussionsgrundlage

Als Diskussionsgrundlage ist es sinnvoll ein kontroverses Thema zu wählen, dass ausreichend Gesprächsstoff liefert. Die Gefahr bei ungünstig gewählten Themen könnte sein, dass die Gesprächsrunde nicht richtig in Gang kommt. Es wäre beispielsweise nachteilig, wenn alle Probanden die gleiche Meinung zu einem Sachverhalt haben und daher nichts mehr zu sagen haben. Der Studienleiter könnte versuchen durch eigene Kommentare Anreize für den Fortlauf der Diskussion zu sorgen. Allerdings ist es nicht möglich die Kommentare des Studienleiters zu bewerten, da er kein aktiver Teilnehmer der Nutzerstudie ist.

Ein möglicher Ansatz für eine Diskussionsgrundlage können Dilemmata sein. Nachfolgend werden zwei Beispiele in den Abbildungen 3.1 und 3.2 dargeboten, dass Möbelpacker-Dilemma[18] und das Sportunterricht-Dilemma[18]:

Ein Team von fünf Möbelpackern führt einen Umzug durch. Sie stehen unter Zeitdruck und der Platz im Wagen ist knapp, so dass sie sich gut überlegen müssen, wie sie die Möbel verstauen. Sie bekommen für den Umzug 1000 Euro. Die fünf Männer haben sich bei der Arbeit folgendermaßen verhalten:

- A ist kräftig, strengt sich an und schafft daher sehr viel.
- B ist nicht so kräftig, strengt sich an und erbringt eine normale Arbeitsleistung
- C strengt sich nicht an, ist kräftig und erbringt eine normale Arbeitsleistung
- D ist weder kräftig, noch strengt er sich besonders an und schafft dementsprechend wenig
- E schließlich ist körperlich fürs Möbelpacken nicht mehr geeignet, kann das Verstauen von Möbeln gut planen. Er steht auf dem Wagen und dirigiert sehr geschickt das Aufladen.

Wie würden SIE die 1000 Euro auf die fünf Arbeiter verteilen? Welche der folgenden Gesichtspunkte sind für Ihre Entscheidung ausschlaggebend gewesen: Kraft, Arbeitsmoral, Arbeitsklima, Tagesform, Gruppengeist, Erfahrung? In welcher Reihenfolge?

Abbildung 3.1: Beispiel 1 für ein kontroverses Dilemma - Möbelpacker

3.1.4 Subjektive Rangfolgen als Referenzdaten

Im Allgemeinen kann am Reputationswert von zwei zufälligen Nutzern eines sozialen Netzwerks oder ähnlichen Portalen abgelesen werden, wer die höhere Reputation hat

[18] http://wikis.zum.de/zum/dilemma

Die Tochter möchte nicht am Sportunterricht teilnehmen, weil sie mit den Jungens zusammen sein müsste. Die Mädchen-Sportlehrerin ist verhindert, deswegen sind die beiden Gruppen zusammengeworfen worden. Davor fürchtet sich das Mädchen, denn sie ist nicht sehr gut im Sport vor allem nicht in den Mannschaftssportarten wie Völkerball und Volleyball, welche bei solchen Gelegenheiten gerne gespielt werden, und sie fürchtet die blöden Bemerkungen der Jungs.
Die Tochter möchte nun, dass ihre Eltern eine Entschuldigung für den ganzen Schultag schreiben, damit es nicht auffällt, dass sie lediglich den Sportunterricht vermeiden möchte.
Die Eltern haben nun folgende Alternativen:

1. Die gewünschte Entschuldigung schreiben (Lügen)
2. Nur für die Sportstunden entschuldigen und irgendwelche Unpässlichkeiten andeuten (Mogeln)
3. Den Sportlehrer anrufen und die Lage auf den Tisch bringen (Ehrlich währt am längsten!) - was die Tochter unter keinen Umständen möchte.
4. Schließlich gar nichts tun und die Tochter in die Schule schicken, damit sie selbst damit fertig wird.

Gibt es noch andere Lösungswege? Welchen Weg würdest du - als Elternteil - einschlagen?

Abbildung 3.2: Beispiel 2 für ein kontroverses Dilemma - Sportunterricht

und wie groß der Unterschied zwischen den beiden Reputationswerten ist. Das Studiendesign dieser Arbeit beschäftigt sich lediglich mit der Fragestellung, wer die höhere Reputation hat und nicht wie groß der Reputationsunterschied ist. Der Grund für diese Abstraktion ist die Annahme, dass Menschen Rangordnungen verlässlicher abschätzen können, als numerische Werte. Daraus resultiert die Vorgabe, dass Reputationsrangfolgen miteinander verglichen werden statt Reputationswerte.

3.1.5 Auswertung der Nutzerstudie

Nach dem Ende einer Nutzerstudie stehen die Bewertungen und die subjektiven Rangfolgen zur Auswertung zur Verfügung. Als nächstes müssen alle subjektiven Rangfolgen aggregiert werden. Es sind grundsätzlich zwei Alternativen denkbar. Zum einen können alle subjektiven Rangfolgen zu einer aggregiert werden, um diese anschließend mit der berechneten zu vergleichen. Zum anderen ist es denkbar jede einzelne subjektive Rangfolge mit der berechneten zu vergleichen, um anschließend alle Vergleiche zu einem zusammenzufassen. Die Auswertungsplattform soll beide Formen der Aggregierung von subjektiven Rangfolgen ermöglichen.

3.2 Webbasierte Auswertungsplattform

Das Studiendesign soll mit einer webbasierten Auswertungsplattform umgesetzt werden, mit deren Hilfe vergleichende Nutzerstudien durchgeführt werden können. Eine Nutzerstudie wird durch einen kompletten Durchlauf der Auswertungsplattform modelliert, bestehend aus den drei Phasen Vorbereitungsphase, Durchführungsphase und

Abschlussphase. Eine Übersicht der Phasen bietet die Abbildung 3.3.

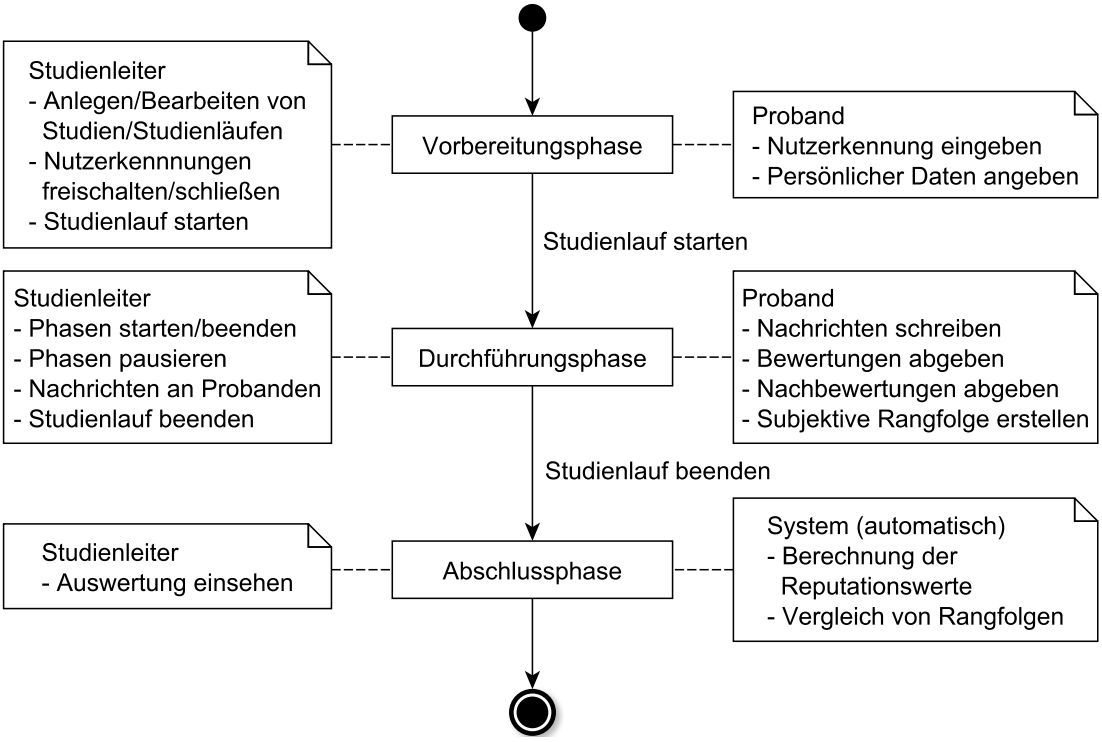

Abbildung 3.3: Phasen der Auswertungsplattform

In der Vorbereitungsphase werden die Rahmenbedingungen für die Nutzerstudie festgelegt. Die Durchführungsphase besteht aus einem Chat mit Bewertungsfunktionalität, einer optionalen Nachbewertungsrunde im Anschluss an den Chat und der Angabe einer subjektiven Rangfolge. In der Abschlussphase werden die gesammelten Daten aggregiert, ausgewertet und veranschaulicht zur weiteren Interpretation dargestellt.

Es gibt zwei Arten von Nutzern, Studienleiter und eine von ihm festgelegte Anzahl von Probanden. Naturgemäß haben Studienleiter und Probanden unterschiedliche Aufgaben innerhalb der Phasen. Der Ablauf der Nutzerstudie soll so gestaltet werden, dass der Studienleiter jederzeit die volle Kontrolle hat. Jede Phase wird explizit von ihm gestartet und beendet, so dass alle Probanden jederzeit in der gleichen Phase sind.

Nachfolgend werden die einzelnen Phasen aus Studienleitersicht und Probandensicht in Abbildung 3.3 näher beschrieben.

3.2.1 Vorbereitungsphase des Studienleiters

Die Vorbereitungsphase des Studienleiters besteht aus der Verwaltung von Nutzerstudien, wobei neue angelegt und vorhandene bearbeitet werden können. Dabei werden alle relevanten Parameter für eine Nutzerstudie festgelegt, die wie folgt spezifiziert sind:

- Anzahl der Probanden
- Anzahl erlaubter Zeichen pro Chatnachricht
- Mit oder ohne Nachbewertungsrunde

- Dauer des Chats
- Bewertungsschema des Chats
- Chatthema

Die Angabe der Anzahl der Probanden wird benötigt, damit die Auswertungsplattform entsprechend Probanden anlegen kann. Jeder Proband ist eindeutig identifizierbar und nimmt an genau einer Nutzerstudie teil. Jedem Proband wird ein zufällig generierter neutraler Alias zugeordnet, der aus einer Kombination von kurzen einprägsamen Vornamen und einer zweistelligen Zahl, wie zum Beispiel *anna23* oder *hans69*, besteht. Die Aliase werden vorgegeben, um initiale Sympathien zu verringern und die Anonymität zu erhöhen. Wenn sich zum Beispiel zwei StarTrek-Fans in einer Nutzerstudie begegnen würden und einer davon den Alias *Captain_Kirk* wählen würde, könnte eine ungewollte Sympathie zwischen den beiden entstehen und die Bewertungen beeinflussen. Dadurch, dass alle Aliase eine vergleichbare Länge aufweisen, sind sie ähnlich gut einprägsam. Das kann beim aufstellen der subjektiven Rangfolge hilfreich sein.

Mit der Anzahl erlaubter Zeichen pro Chatnachricht können zu lange Chatnachrichten vermieden werden. Probanden sind somit gezwungen sich prägnanter auszudrücken oder mehrere Nachrichten abzuschicken. Je mehr separate Nachrichten geschrieben werden, desto mehr potenzielle Bewertungen.

Im Kontext der Reputationsberechnung ist es wünschenswert möglichst viele Nachrichten und somit potenzielle Bewertungen zu bekommen. Eine Nachbewertungsrunde findet direkt nach Beendigung des Chats statt. Probanden haben hierbei die Möglichkeit die noch nicht bewerteten Chatnachrichten nachträglich in Ruhe zu bewerten. Nachbewertungen sind optional und können vom Studienleiter in Betracht gezogen werden, falls die durch den Chat generierte Anzahl der Bewertungen unzureichend ist.

Über die Chatdauer lässt sich ebenfalls die Menge der Nachrichten und Bewertungen regulieren. Es liegt auf der Hand, dass eine längere Chatdauer höchstwahrscheinlich mehr Daten generiert. Auf der anderen Seite sollte keine zu lange Chatdauer gewählt werden, um die Probanden nicht zu stark zu strapazieren.

Das Bewertungsschema und Chatthema beziehungsweise Gesprächsgrundlage verhalten sich analog zu den Ausführungen in den Kapiteln 3.1.2 und 3.1.3.

Grundsätzlich wird zwischen Studien und Studienläufen unterschieden, wobei eine Studie einen oder mehrere Studienläufe kapselt. In einer Studie werden Parameter festgelegt, die für alle Studienläufe konstant bleiben sollen. Somit müssen diese Parameter nicht nochmals in den einzelnen Läufen angegeben werden. In den Läufen müssen genau diejenigen Parameter spezifiziert werden, die in der Studie nicht berücksichtigt wurden. Im Gegensatz zu Studien dürfen hier keine Parameter mehr leer gelassen werden. Daraus ergibt sich, dass alle Parameter mit einer Studie und einem Studienlauf dieser Studie abgedeckt sind.

Die Aufteilung in Studien und Studienläufe ist sinnvoll, da im Normalfall in mehreren Studienläufen Parameter verwendet werden, die sich nur in einem oder zwei Merkmalen unterscheiden. Angenommen der Studienleiter möchte testen, wie gut bestimmte

Chatthemen die Kommunikation zwischen Probanden anregen. Infolge dessen könnte er eine Studie und mehrere Studienläufe anlegen, wobei in den Studienläufen nur das Gesprächsthema, dass möglicherweise jedes Mal ein anderes ist, angegeben wird. Alle anderen Parameter werden in der Studie angegeben.

Neben den oben aufgelisteten Parametern, die sowohl für Studien als auch für Studienläufe verwendet werden können, gibt es zusätzlich spezifische Parameter. So hat jede Studie einen Studiennamen, der vergeben werden muss. Damit kann der Studienleiter die einzelnen Studien bequem unterscheiden. Für jeden Studienlauf muss ein Termin angegeben werden, an dem dieser stattfindet.

Nachdem benötigte Studien und Studienläufe angelegt worden sind, hat der Studienleiter die Möglichkeit diese für Probanden freizuschalten und ihnen somit zu ermöglichen, sich in den Studienlauf einzuloggen. Dazu benötigen die Probanden lediglich die Nutzerkennung, die ihnen mitgeteilt werden muss. Die Vorbereitungsphase des Studienleiters ist beendet, wenn er den Studienlauf startet. Dies ist nur möglich, wenn der dieser vorher freigeschaltet worden ist.

Abbildung 3.4 zeigt die Vorbereitungsphase des Studienleiters. Jeder Knoten entspricht einer Sicht und eine Kante dem Übergang zwischen zwei Sichten. Zusätzlich zu den bereits beschriebenen Aktionen ist es möglich Parameter angelegter Studien und Studienläufe zu bearbeiten. So kann zum Beispiel ein Parameter einer Studie in einem Studienlauf dieser Studie bearbeitet werden anstatt eine neue Studie anzulegen. Außerdem können mit dieser Funktion mögliche Tippfehler ausgemerzt werden.

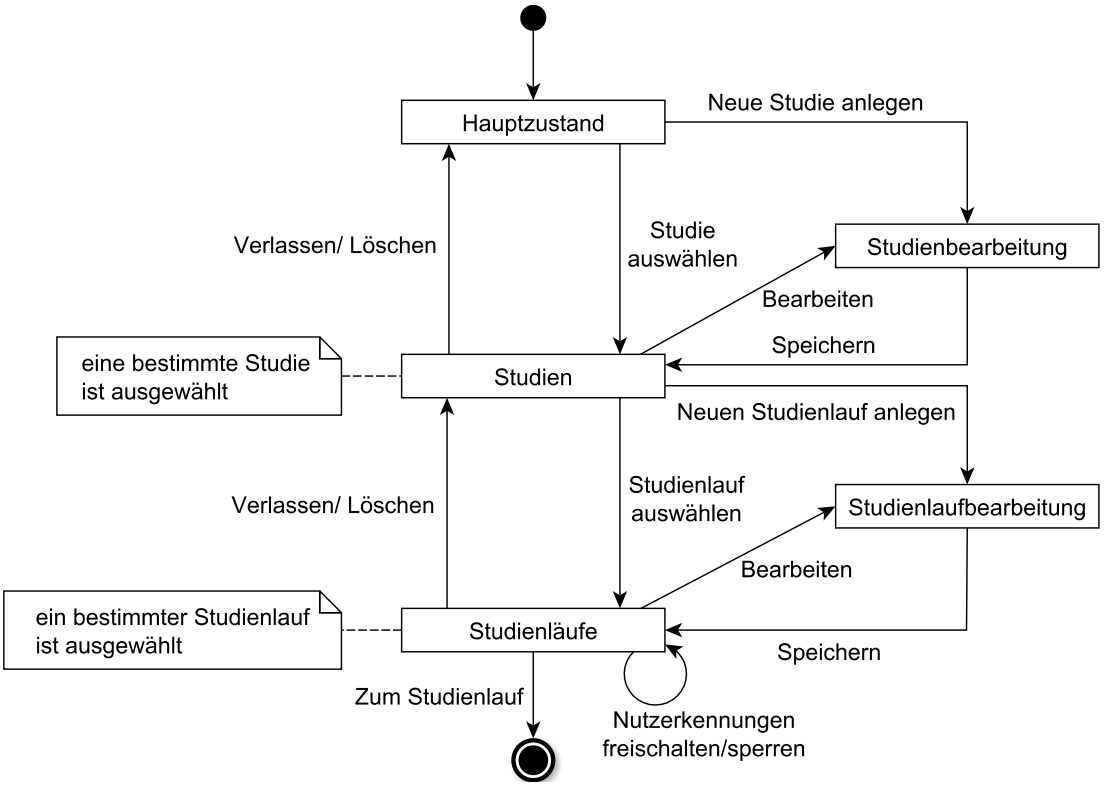

Abbildung 3.4: Vorbereitungsphase des Studienleiters

3.2.2 Vorbereitungsphase der Probanden

Für den Probanden besteht die Vorbereitungsphase lediglich aus zwei Schritten. Zum einen muss er sich mit seiner Nutzerkennung in den Studienlauf einloggen. Diese erhält er zusammen mit der Webadresse des Studienlaufs vom Studienleiter. Zum anderen soll jeder Proband nach dem einloggen persönliche Daten angeben. Der Studienleiter bestimmt im Vorfeld, welche persönlichen Daten Pflichtangaben sind und welche nur optional. Nach dem Abschicken der persönlichen Daten ist die Vorbereitungsphase für den Probanden zu Ende. Er muss warten, bis der Studienleiter die nächste Phase startet. In der Abbildung 3.5 ist die Vorbereitungsphase des Probanden illustriert.

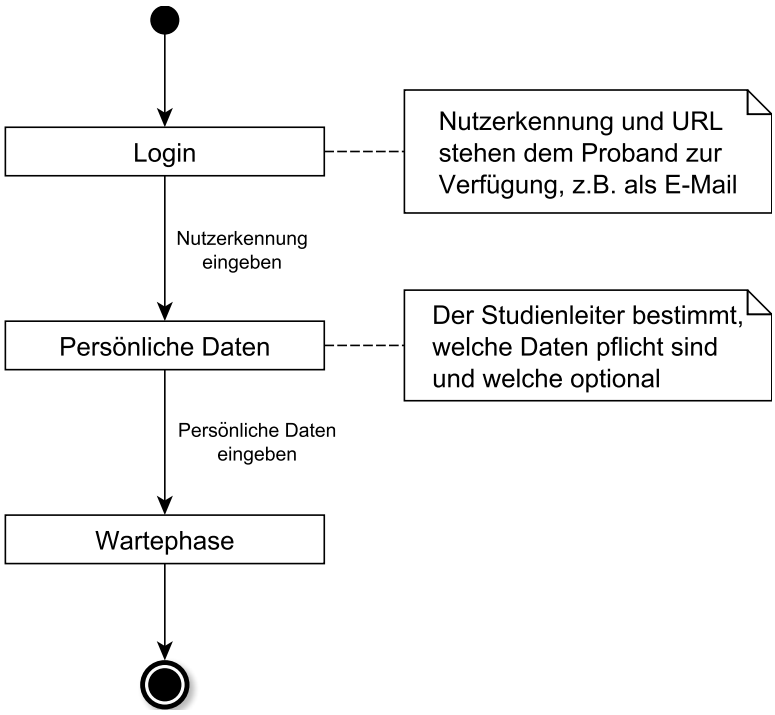

Abbildung 3.5: Vorbereitungsphase der Probanden

3.2.3 Durchführungsphase des Studienleiters

Während der Durchführungsphase nimmt der Studienleiter eine passive beziehungsweise beobachtende Rolle ein. Er greift nur ein, wenn Probleme auftreten sollten. Grundsätzlich hat er die Möglichkeit Hinweise an einen oder alle Probanden zu geben. Hinweise an alle Probanden könnten zum Beispiel Anreize zur Fortführung des Chats sein, falls die Chatfreudigkeit abnehmen sollte. Es könnten neue kontroverse Fragestellungen in den Raum geworfen werden. Sollte der Studienleiter zum Entschluss kommen, die Chatrunde verlängern oder verkürzen zu wollen, so kann er dies ebenfalls mit einem Hinweis kundtun.

Im Gegensatz dazu sind Nachrichten des Studienleiters an einen Probanden gut geeignet, um technische Probleme zu lösen und sonstige Hilfestellungen zu geben. Die restlichen Probanden werden dadurch nicht gestört.

Nach dem Verstreichen der vorgegebenen Zeit oder wenn der Studienleiter es für an-

gemessen hält, beendet er den Chat. Falls für den aktuellen Studienlauf eine Nachbewertungsrunde vorgesehen ist, beginnt sie an dieser Stelle automatisch. Jeder Proband bestätigt explizit, wenn er mit der Nachbewertungsrunde abgeschlossen hat. Sobald der Studienleiter feststellt, dass alle fertig sind, kann er die Durchführungsphase mit der Angabe der subjektiven Rangfolge fortsetzen.

Falls keine Probleme auftreten, kann der Studienleiter die Beendigung der Durchführungsphase feststellen, wenn alle Probanden die subjektive Rangfolge abgegeben haben. Abbildung 3.6 zeigt die eben beschriebene Durchführungsphase des Studienleiters.

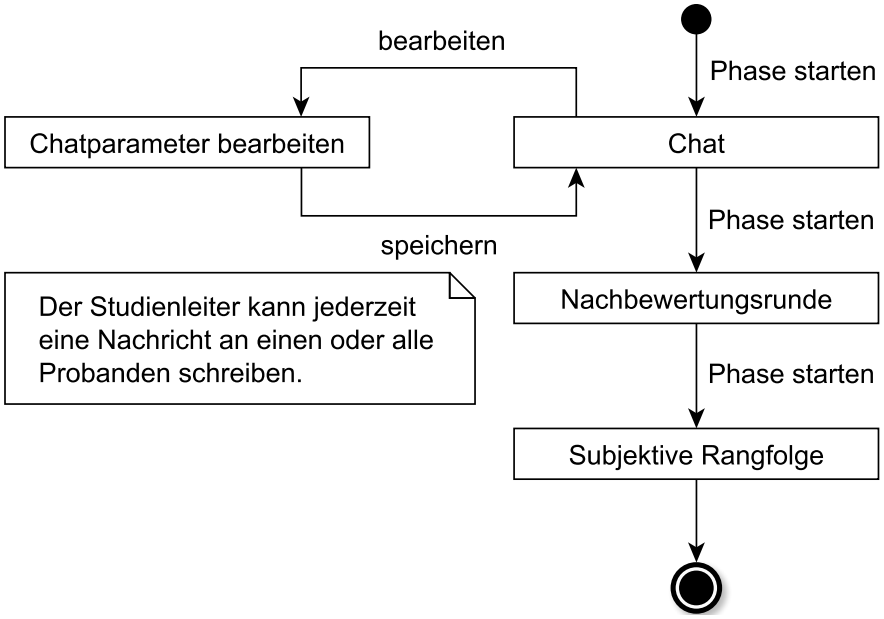

Abbildung 3.6: Durchführungsphase des Studienleiters

3.2.4 Durchführungsphase der Probanden

Sobald die Durchführungsphase beginnt, haben alle Probanden die Möglichkeit über einen Chat zu kommunizieren und Bewertungen der Chatnachrichten der anderen Probanden abzugeben. Des Weiteren werden alle Teilnehmer des Chats aufgelistet. Dadurch bekommt jeder Proband einen besseren Eindruck über die Größe der Chatrunde und kann sich die Aliase besser einprägen. Die Anzahl der erlaubten Zeichen pro Chatnachricht wird den Probanden ebenfalls angezeigt. Außerdem wird während der Chatrunde das Chatthema angezeigt, damit es bei Bedarf jederzeit ins Gedächtnis gerufen werden kann.

Im Falle einer Nachbewertungsrunde wird den Probanden die gesamte Historie des Chats nochmals dargeboten. Die noch nicht bewerteten Nachrichten können nun nachträglich in Ruhe bewertet werden. Dadurch sollen zusätzliche Bewertungen generiert werden, die je nach Reputationsalgorithmus zu genaueren Reputationswerten führen können.

Im Anschluss an die Nachbewertungsrunde soll die subjektive Rangfolge aufgestellt werden. Dabei soll die Liste der zu sortierenden Probanden möglichst neutral darge-

stellt werden, um Beeinflussung zu verhindern. Hierzu bietet es sich an die Probandenliste alphabetisch zu sortieren und auf diese Sortierung hinzuweisen. Der sortierende Proband kommt in der Probandenliste naturgemäß nicht vor. Abbildung 3.7 zeigt die Durchführungsphase des Probanden.

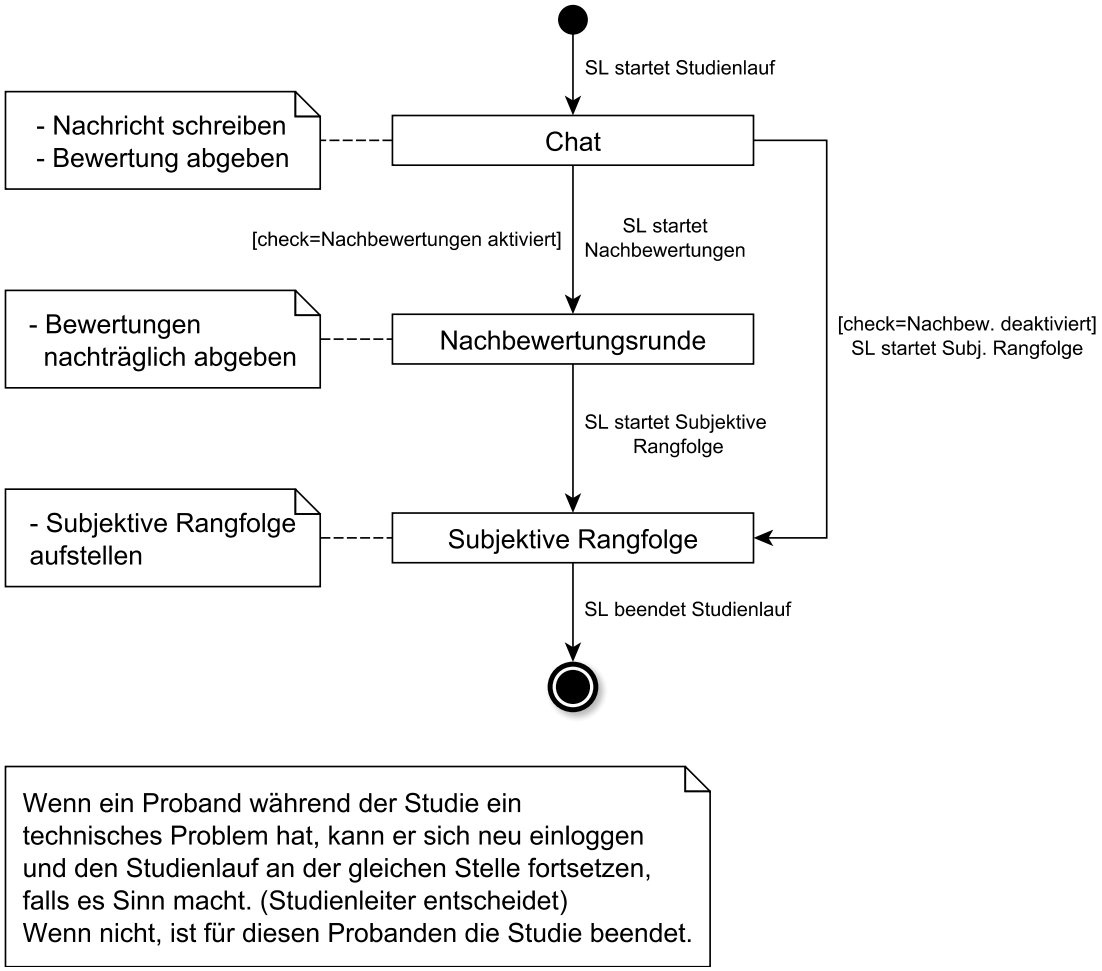

Abbildung 3.7: Durchführungsphase des Probanden

3.2.5 Abschlussphase

Die Abschlussphase beginnt mit der automatischen Auswertung der Daten, nachdem der Studienleiter den Studienlauf beendet hat. Die genaue Modellierung der Auswertung ist von den Zielen der Auswertung abhängig und kann sehr vielfältig ausfallen. Die Auswertungsplattform dient lediglich als technische Hilfestellung, um Nutzerstudien effizient durchführen und passende Daten sammeln zu können.

Eine Möglichkeit der Auswertung könnte grob gesprochen folgendermaßen erfolgen. Da jeder Proband eine subjektive Rangfolge abgegeben hatte, stehen mehrere davon zur Verfügung. Durch geeignete Aggregation können alle subjektiven Rangfolgen zu einer zusammengefasst werden. Die berechnete Rangfolge wird mit Hilfe der Bewertungen entsprechend dem Reputationsalgorithmus erstellt.

Zum Vergleich der beiden Rangfolgen kann der Rangkorrelationskoeffizient verwendet werden. Dabei handelt es sich um den zahlenmäßigen Ausdruck des Zusammenhangs zweier Rangfolgen [Liu14].

3.3 Zusammenfassung

In diesem Kapitel wurde das Konzept einer webbasierten Auswertungsplattform für Reputationsalgorithmen vorgestellt, mit der Nutzerstudien durchgeführt werden können. Genauer wird die Konsistenz von berechneten Reputationswerten und subjektiven Rangfolgen von Probanden geprüft. Die Auswertungsplattform ist in drei Phasen aufgebaut, in denen relevante Daten gesammelt und ausgewertet werden. Die zentrale Phase ist dabei die Durchführungsphase, in der die Probanden kontrovers miteinander unterhalten, Chatnachrichten der anderen Probanden bewerten und eine subjektive Rangfolge angeben sollen.

Des Weiteren soll die Auswertungsplattform dem Studienleiter die technische Basis liefern, um Studien und Studienläufe unterscheiden zu können und Parameter einstellen zu können. Eine Auflistung der Anforderungen an die Auswertungsplattform wird im Folgenden dargeboten:

- Realisierung eines Chats
- Bewertungsmöglichkeit entsprechend eines Bewertungsschemas
- Anzeige von Chatthema und Chatdauer
- Gewährleistung von Anonymität zwischen Probanden
- Mechanismus zum aufstellen subjektiver Rangfolgen
- Aufteilung von Nutzerstudien in Studien und Studienläufe
- Möglichkeit Parameter zu spezifizieren

Kapitel 4

Implementierung der Auswertungsplattform

Dieses Kapitel beschäftigt sich mit der technischen Umsetzung des in Kapitel 3 beschriebenen Konzepts. Dabei wird auf die wichtigsten Aspekte der Implementierung eingegangen. Zunächst wird das Datenmodell erläutert, insbesondere wie die Daten gespeichert werden und wie sie zusammenhängen. Anschließend folgt ein Unterkapitel zum Phasenmodell, dass für einen kompletten Durchlauf eines Studienlaufs verantwortlich ist. Daraufhin wird die Umsetzung der Server-Client Kommunikation während eines Studienlaufs mittels WebSockets erklärt. Weiter wird die Auswertungsplattform aus der Sicht der Studienleiters und des Probanden betrachtet und dabei die einzelnen Sichten erläutert. Schließlich werden mögliche Probleme betrachtet, die bei der Nutzung der Auswertungsplattform entstehen können, und Lösungen vorgestellt.

4.1 Datenmodell

Zur dauerhaften Speicherung relevanter Daten wird die relationale Datenbank MySQL[19] verwendet, wobei grundsätzlich auch andere Datenspeicherungsmodelle vorstellbar wären. Das Datenmodell wird im Folgenden anhand des Entity-Relationship Diagramms in Abbildung A.1 im Anhang erläutert.

Die blauen rechteckigen Knoten symbolisieren Tabellen und die grünen ovalen entsprechende Attribute der Datenstruktur. Gerichtete Kanten beschreiben eine m-zu-1 Beziehung. Schlüsselattribute sind unterstrichen.

Eine Studie wird durch die Tabelle *studies* dargestellt und kann einen oder mehrere Studienläufe beinhalten, die durch *iterations* beschrieben werden. Weiter hat jede Studie Studienparameter, die in der Tabelle *studyparams* gekapselt sind. Aus Gründen der Übersichtlichkeit bekommt jede Studie einen Studiennamen zugewiesen.

Studienläufe haben ebenfalls Parameter, die in den Tabellen *iterparams* und *liveparams* gekapselt sind. Grundsätzlich haben sowohl *studyparams* als auch *iterparams* und *liveparams* bis auf den Schlüssel die gleichen Attribute. Die Aufteilung in *studyparams* und *iterparams* wurde gewählt, um Parameter sowohl Studien als auch Studienläufen zuordnen zu können. Die Tabelle *liveparams* ist notwendig, um nachträgliche Änderungen

[19] http://www.mysql.de

Attributname	**Bedeutung**
userscount	Anzahl der Probanden im Studienlauf
charscount	Anzahl erlaubter Zeichen pro Chatnachricht
extrarating	Mit oder ohne nachträgliche Bewertungsrunde
chatduration	Dauer des Chats
ratingscheme	Bewertungsschema
topic_id	Fremdschlüssel zum Chattema

Tabelle 4.1: Erklärung der Parameter

von Parametern festhalten zu können ohne die alten Werte zu überschreiben. Schließlich wird in der Tabelle *iterations* das Attribut *iterdate* benötigt, dass die Startzeit eines Laufs festlegt.

Beim anlegen einer Studie können zunächst alle Parameter angegeben werden. Sie werden in der Tabelle *studyparams* gespeichert und gelten für alle Studienläufe der Studie. Damit lassen sich gleich bleibende Parameter wiederverwenden und müssen nicht für jeden neuen Studienlauf neu angegeben werden. Leer gelassene Parameter bekommen automatisch den Wert *null* zugewiesen. Beim anlegen eines Studienlaufs werden genau diejenigen Parameter abgefragt, die in *studyparams* den Wert *null* haben. Somit wird jeder Parameter genau einmal gesetzt.

Falls nach dem Start eines Studienlaufs Änderungen an Parametern nötig sind, so werden diese in der Tabelle *liveparams* gespeichert. Das System ist so modelliert, dass Parameter von *liveparams* entsprechende Parameter in *studyparams* und *iterparams* überschatten. Wenn beispielsweise ein Studienlauf für zehn Probanden angelegt wird und nur acht davon erscheinen, so wird in *liveparams* beim Attribut *userscount* der Wert 8 eingetragen. Der ursprüngliche Wert 10, der entweder in studyparams oder in iterparams angegeben wurde, wird nicht überschrieben. Der Vorteil dieser Vorgehensweise ist, dass alle Schritte im Nachhinein nachvollzogen werden können, da Parameter nie überschrieben werden sondern in einer separaten Tabelle gespeichert werden. Dies könnte bei der Auswertungen der Ergebnisse hilfreich sein. In der Tabelle 4.1 werden die einzelnen Parameterattribute erklärt, deren Notwendigkeit sich aus den Ausführungen in Unterkapitel 3.2.1 ergibt.

In der Tabelle *topics* werden die Chatthemen gespeichert, die in den drei Parametertabellen verwendet werden. Ein Chatthema hat einen Namen und einen Text als Attribute.

Jedem Studienlauf werden entsprechend des *userscount* Parameters Probanden zugeordnet, die in der Tabelle *users* gespeichert werden. Ein Datensatz kann in genau einem Studienlauf verwendet werden, spezifiziert durch den Fremdschlüssel *iteration_id*. Das Attribut *status* gibt über die aktuelle Phase Auskunft, in der sich ein Proband befindet. Genauere Erklärungen zu Phasen folgen in Kapitel 4.2. Des Weiteren werden in der Tabelle *users* der Alias und die persönlichen Daten des Probanden spezifiziert.

Komplettiert wird das Datenmodell von Tabellen zum Speichern von Auswertungsdaten, die während der Nutzerstudie von Probanden generiert werden sollen. Dazu gehören Chatnachrichten (*messages*), Bewertungen (*ratings*) und subjektive Rangfol-

Phase	Zustände im Diagramm
Login	Loginmaske
Daten	„Persönliche Daten“ und „Wartephase Persönliche Daten abgeschickt“
Chat	Chat (keine Wartephase nötig)
Nachbewertungen	„Nachbewertungen“ und „Wartephase Nachbewertungen abgeschickt“
Rangfolge	„Rangfolge“ und „Wartephase Rangfolge abgeschickt“

Tabelle 4.2: Gegenüberstellung von Hauptphasen und Zuständen (grün: Vorbereitungsphase, gelb: Durchführungsphase)

gen (*rankings*). Da eine Nachricht, Bewertung oder subjektive Rangfolge genau einem Probanden zugeordnet werden soll, beinhalten alle drei Tabellen den Fremdschlüssel *user_id*. Zu jeder Nachricht wird außerdem ein Zeitstempel und der Nachrichtentext gespeichert. Die Informationen können bei der Auswertung behilflich sein und ermöglichen die Wiederherstellung der Chathistorie, falls ein Proband ungewollt aus dem Chat ausscheidet, zum Beispiel durch technische Probleme oder versehentliches Schließen des Browserfensters. Damit Bewertungen eindeutig zugeordnet werden können, bietet die Tabelle *ratings* die Attribute *ratingto*, *rating* und *message_id*. Dabei symbolisiert *ratingto* den Empfänger der Bewertung und *message_id* die Chatnachricht, die bewertet wurde. Bei *rating* handelt es sind um die Ausprägung der Bewertung entsprechend dem vorliegenden Bewertungsschema. Die Tabelle *rankings* speichert die subjektive Rangfolge von Probanden.

4.2 Phasenmodell

Das Phasenmodell realisiert im Wesentlichen die Vorbereitungs- und Durchführungsphase des Probanden basierend auf den Abbildungen 3.5 auf Seite 27 und 3.7 auf Seite 29 aus dem Kapitel 3. Dabei wird jeder Zustand im Diagramm als separate Ansicht im Browser verstanden. Eine Phase fasst einen zusammengehörigen Aktivitäts- und Warteabschnitt zusammen. In einem Aktivitätsabschnitt soll der Proband etwas tun, zum Beispiel persönliche Daten eingeben, chatten oder eine subjektive Rangfolge angeben. Sobald er mit der jeweiligen Aktivität fertig ist, bestätigt er dies und gelangt in den Warteabschnitt. Abbildung 4.2 zeigt eine Gegenüberstellung der Hauptphasen und entsprechender Zustände aus den Diagrammen. Grün hinterlegte Zeilen gehören zur Vorbereitungsphase und gelb hinterlegte zur Durchführungsphase. Der Wechsel von einer Phase in die nächste kann nur vom Studienleiter eingeleitet werden. Dies gewährt ihm Kontrolle über die Nutzerstudie und sorgt für einen synchronen Ablauf, da sich alle Nutzer immer in der gleichen Hauptphase befinden. Der Warteabschnitt ist zwingend notwendig, da er dem Studienleiter den Status der Probanden anzeigt. Ohne Warteabschnitt hätte der Studienleiter keine Kenntnis darüber, ob ein Proband mit einer Aktivität fertig ist.

4.2.1 Studienleiter Panel

Sobald ein Studienlauf vom Studienleiter gestartet worden ist, steht ihm ein Bereich zum Überwachen des Studienlaufs zur Verfügung, der in Abbildung 3.3 dargeboten wird und als *Studienleiter Panel* bezeichnet wird. Mit dessen Hilfe kann ein Studienlauf und nächste Phasen innerhalb des Studienlaufs gestartet werden, indem entsprechende Buttons geklickt werden. Damit der Studienleiter keine Aktion aus Versehen ausführt, muss

jeder Buttonklick in einem separaten Fenster bestätigt werden. Diese beiden Buttons sind allerdings nur anklickbar, wenn der aktuelle Zustand des Studienlaufs es erlaubt.

Ein Studienlauf kann nur gestartet werden, wenn alle Probanden eingeloggt sind. Wenn das nicht der Fall ist, muss der Studienleiter dafür sorgen, dass Ersatz gefunden wird oder die Anzahl der Probanden entsprechend anpassen. Aliase eingeloggter Probanden werden grün dargestellt und nicht eingeloggter rot. Wenn ein Proband zu einem Studienlauf eingeteilt ist und sich nicht zur vereinbarten Zeit einloggt, kann er vom Studienleiter kurzfristig aus dem Studienlauf entfernt werden. Dazu dient das entsprechende Symbol neben dem Alias. Allerdings muss hier beachtet werden, dass mindestens drei Probanden für die Durchführung eines Laufs notwendig sind. Bei nur zwei Probanden wäre es nicht sinnvoll eine subjektive Rangfolge anzugeben, da die Probandenliste aus nur einem Probanden bestehen würde. Das Konzept sieht vor, dass Probanden sich selbst nicht in der subjektiven Rangfolge berücksichtigen sollen.

Der Button zum starten der nächsten Phase kann nur betätigt werden, wenn alle Probanden im Warteabschnitt sind.

Unter den beiden Buttons befindet sich ein Fortschrittsbalken, der dem Studienleiter die aktuelle Phase anzeigt. In der Abbildung 3.3 befindet sich der Studienlauf aktuell in der Phase Chat.

Die Tabelle unter dem Fortschrittsbalken realisiert Aktivitäts- und Warteabschnitte. Die Bezeichnung *Im Gange* symbolisiert eine Aktivität während *Fertig* einen Warteabschnitt anzeigt. Mit Hilfe der Häkchen kann der Studienleiter nachvollziehen, welcher Proband sich in welchem Abschnitt befindet. In der Phase Chat ist naturgemäß kein Warteabschnitt nötig, da eine feste Chatdauer vorgegeben ist.
Des Weiteren werden im Studienleiter Panel Bewertungen der Probanden dokumentiert. Sobald eine Bewertung abgegeben wurde, erscheint unter dem Punkt *Bewertungen* eine Benachrichtigung der Form *alias1* $\rightarrow$ *alias2* in rot oder grün. Das bedeutet, dass *alias1* eine Chatnachricht von *alias2* bewertet hat. Dabei steht rot für negative und grün für positive Bewertungen. Ein Log von Bewertungen kann für den Studienleiter nützlich sein, um rasch erste Rückschlüsse über Probanden ziehen zu können.

4.2.2 Technische Realisierung

Jedem Probanden ist in der Datenbanktabelle *users* ein Datensatz zugeordnet, der das Attribut *phase* enthält. Jedes mal wenn ein Proband in eine neue Phase oder Abschnitt kommt, wird dieses Attribut aktualisiert. Es besteht aus zwei Zeichenketten, die mit einem Bindestrich verbunden sind. Der erste Teil symbolisiert die aktuelle Phase, der zweite den aktuellen Abschnitt. Die Häkchen aus der Abbildung 4.1 resultieren aus diesem Attribut. Zur Illustration folgt zeigt Tabelle 4.3 alle möglichen Ausprägungen des Attributs *phase* mit kurzer Erklärung.

4.3 WebSocket Protokoll

Die Kommunikation zwischen Webserver und Client/ Webbrowser während eines Studienlaufs ist mit dem WebSocket Protokoll realisiert. Dabei handelt es sich um ein auf TCP basierendes Netzwerkprotokoll, dass bidirektionale Kommunikation zwischen

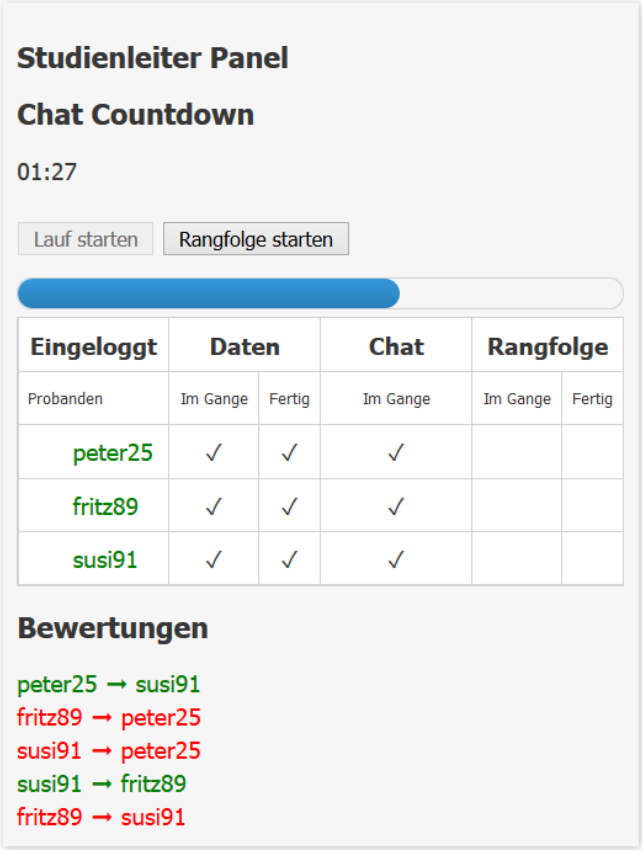

Abbildung 4.1: Studienleiter Panel

Phase Datenbank	**Zustände im Diagramm**
null	Proband ist nicht eingeloggt
logged-ready	Proband ist eingeloggt und wartet
userdata-running	Proband gibt persönliche Daten an
userdata-ready	Proband hat persönliche Daten abgeschickt und wartet
chat-running	Proband chattet
extrarating-running	Proband vergibt Nachbewertungen
extrarating-ready	Proband ist fertig mit Nachbewertungen und wartet
rank-running	Proband gibt die subjektive Rangfolge an
rank-ready	Proband hat die subjektive Rangfolge abgeschickt und wartet
done-running	Proband ist fertig mit der Nutzerstudie

Tabelle 4.3: Mögliche Ausprägungen des Attributs *phase*

Webserver und Client/ Webbrowser ermöglicht. Im Gegensatz zu konventionellen Methoden über HTTP-Anfrage und HTTP-Antwort ist die Verbindung zum WebSocket ständig aktiv. Dadurch ist ein mehrfaches Aufbauen und Schließen der Verbindung nicht mehr nötig. Infolge dessen können HTTP-Header-Informationen eingespart werden [Abt15]. Es gibt zwar auch die Möglichkeit mit dem HTTP-Standard 1.1 eine persistente Verbindung zwischen Webserver und Client/ Webbrowser herzustellen, jedoch folgt diese ebenfalls dem Anfrage-Antwort-Muster [FR14]. WebSockets ermöglichen hingegen auch den Datenaustausch von Server zum Client ohne vorherige Anfrage durch den Client [Abt15].

Das WebSocket Protokoll eignet sich hervorragend, um Chat-Applikationen, wie bei der Chatphase eines Studienlaufs, zu realisieren. Bei konventioneller Kommunikation weiß die Client-Anwendung nicht, ob neue Chatnachrichten abgeschickt wurden. Also müssten in regelmäßigen Abschnitten Anfragen an den Server gestellt werden. Bei zu kurzem Anfrageintervall könnten unter Umständen viele unnötige Anfragen gesendet werden, obwohl keine Chatnachrichten verfasst wurden. Bei zu langem Anfrageintervall könnten einige Chatnachrichten erst mit Verzögerung angezeigt werden. Durch die dauerhaft aktive WebSocket-Verbindung wird dieses Problem elegant umgangen. Sobald der Server eine neue Chatnachricht erhält, wird diese an alle Clients kommuniziert.

Außerdem helfen WebSockets in dieser Auswertungsplattform dabei, die Problematik des Zurückbuttons zu verringern. Üblicherweise würde jede Phase eines Studienlaufs das laden der entsprechenden Ansicht erfordern. Der Proband hätte die Möglichkeit mit dem Zurückbutton des Browsers zu einer bereits abgeschlossenen Phase zu wechseln und Manipulationen an Daten durchzuführen. Dieses nicht gewünschte Verhalten müsste durch zusätzlichen Code abgefangen werden. Bei einer Umsetzung mit WebSockets wird keine neue Seite oder Ansicht beim wechseln der Phasen geladen, da sonst das WebSocket-Objekt verloren gehen würde. Alle Ansichten werden über die WebSocket Verbindung generiert und benötigen kein Neuladen der Seite. Infolge dessen ist die Reichweite des Zurückbuttons stark eingeschränkt. Es kann damit lediglich auf die Indexseite des Studienlaufs gewechselt werden. Dies stellt allerdings kein großes Problem dar und wird in Kapitel 4.6 näher erläutert.

4.4 Sichten des Studienleiters

Dem Studienleiter stehen im Gegensatz zu Probanden die Sichten *Studienverwaltung* und *Auswertung* zur Verfügung. Sie können über ein Navigationsmenü aufgerufen werden. In den folgenden Unterkapiteln folgt eine Beschreibung dieser beiden Sichten.

4.4.1 Studienverwaltung

Die Studienverwaltung ist die Umsetzung des Diagramms in Abbildung 3.4 auf Seite 26. Sie ermöglicht dem Studienleiter das Anlegen, Bearbeiten und Löschen von Studien und Studienläufen, die noch nicht abgeschlossen worden sind. Abgeschlossene Studienläufe werden hier nicht angezeigt und stattdessen in der Auswertung berücksichtigt. Dadurch wird die Übersichtlichkeit erhöht. Zusätzlich können Studienläufe freigeschaltet werden, damit sich entsprechende Probanden einloggen können.

Die drei Zustände *Hauptzustand*, *Studie* und *Studienlauf* werden mit Hilfe von jQuery

UI Accordion[20] modelliert. Dabei handelt es sich um ein dynamisches Skript, dass das ein- und ausklappen von Inhalten ermöglicht. Die *Studienverwaltung* befindet sich im Hauptzustand, wenn alle Studien eingeklappt sind (siehe Abbildung 4.2) oder wenn noch keine angelegt sind.

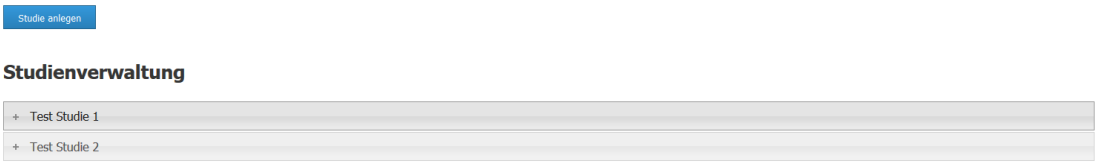

Abbildung 4.2: Studienverwaltung im Hauptzustand

Aus dem Hauptzustand können entweder neue Studien angelegt werden, indem der Button *Studie anlegen* geklickt wird, oder vorhandene Studien ausgewählt werden. Falls eine neue Studie angelegt werden soll, wird mit der Ansicht *Studienbearbeitung* fortgefahren, auf die im Fortlauf des Textes eingegangen wird. Eine vorhandene Studie wird durch das öffnen eines Accordionelements ausgewählt, illustriert in Abbildung 4.3. Dies symbolisiert einen Übergang in den Zustand *Studie*, da sich der Studienleiter explizit für eine Studie entscheidet. Durch das schließen des Accordionelements oder das betätigen des *Studie löschen* Buttons kann der Hauptzustand wiederhergestellt werden.

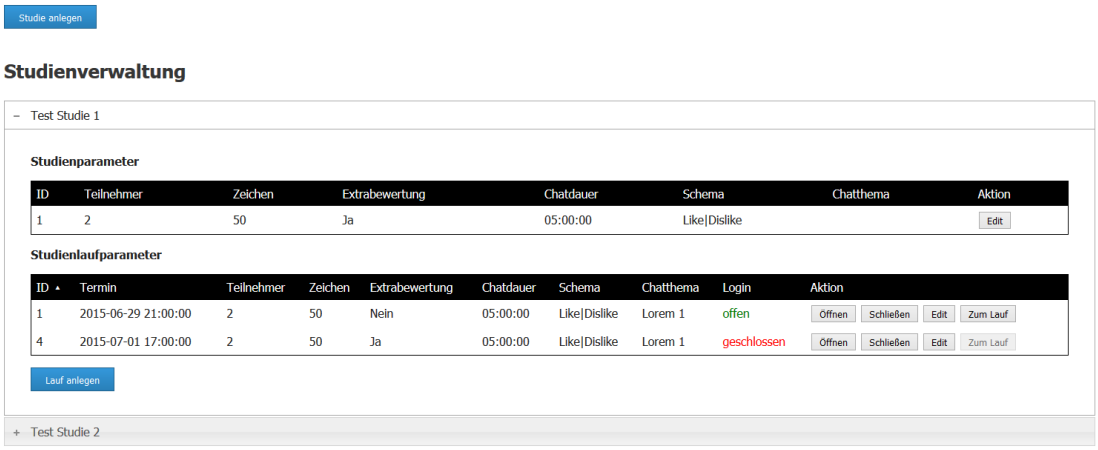

Abbildung 4.3: Studienverwaltung im Zustand *Studien*

Nachdem eine Studie ausgewählt ist, können neue Studienläufe innerhalb der Studie angelegt werden und vorhandene Studienläufe verwaltet werden. Ein neuer Studienlauf wird durch das anklicken des entsprechenden Buttons *Studienlauf anlegen* angelegt. Dies passiert analog zum anlegen einer neuen Studie in einer separaten Ansicht *Studienlaufbearbeitung*, auf die ebenfalls noch näher eingegangen wird.

Jedes aufgeklappte Studienelement beinhaltet die beiden Tabellen *Studienparameter* und *Studienlaufparameter*. Erstere zeigt Parameter, die beim anlegen der Studie spezifiziert worden sind. Nicht spezifizierte Parameter sind leer. Mit dem Button *Bearbeiten* lassen sich Studienparameter nachträglich verändern. Dies passiert ebenfalls über die Studienbearbeitung. Allerdings muss hier beachtet werden, dass Studienparameter von Studienlaufparametern überschattet werden, wie es im Datenmodell (Kapitel 4.1) be-

[20] http://jqueryui.com/accordion

schrieben ist. Zusätzlich gibt es einen Button *Löschen*, mit dem die Studie gelöscht werden kann. Dabei muss beachtet werden, dass Studien nur gelöscht werden können, wenn keine Studienläufe vorhanden sind. Wenn der Studienleiter versucht eine Studie zu löschen, zu der es Studienläufe gibt, so erscheint eine entsprechende Fehlermeldung.

In der Tabelle *Studienlaufparameter* befinden sich außer den Parametern Buttons zum verwalten eines Studienlaufs, und zwar *Öffnen*, *Schließen*, *Bearbeiten*, *Zum Studienlauf* und *Löschen*. Durch das Betätigen dieser Buttons wählt der Studienleiter implizit einen Studienlauf und befindet sich im Diagramm im Zustand *Studienlauf*. Die beiden Buttons *Öffnen* und *Schließen* ermöglichen es dem Studienleiter Läufe freizuschalten und bei Bedarf zu sperren. Somit kann kontrolliert werden, wann sich Probanden in einen Studienlauf einloggen dürfen. Manchmal kann es notwendig sein einen geöffneten Studienlauf vor Beendigung zu schließen, zum Beispiel wenn zu wenig Probanden erscheinen. In diesem Fall kann der Studienlauf neu terminiert und nochmals geöffnet werden. Analog zum Bearbeiten von Studienparametern lassen sich auch Studienlaufparameter mittels *Bearbeiten* verändern. Bereits geöffnete Studienläufe können jedoch nicht mehr bearbeitet werden, da deren Integrität gestört werden könnte. Angenommen es loggen sich fünf Probanden in einen geöffneten Studienlauf ein. Wenn der Studienleiter die Teilnehmeranzahl heruntersetzt, würde ein inkonsistenter Datenbankzustand entstehen, da es Datensätze von Probanden geben kann, die zu keinen Studienlauf zugeordnet sind. Mit dem Button *Zum Studienlauf* wird die Studienverwaltung verlassen und die entsprechende Studienlaufansicht angezeigt. Naturgemäß lässt sich dieser Button nur anklicken, wenn der dazugehörige Studienlauf geöffnet worden ist. Schließlich können nicht benötigte oder fälschlicherweise angelegte Läufe mit dem Button *Löschen* entfernt werden.

Die Ansichten für die Zustände *Studienbearbeitung* und *Studienlaufbearbeitung* sind ähnlich aufgebaut. Sie bieten die Möglichkeit Parameter in Formularfelder einzugeben. In ersterer gibt es ein Feld *Name der Studie* während in zweit genannter der *Termin des Studienlaufs* angegeben wird. Bei Parametern, die eine feste Anzahl an Auswahloptionen vorschreiben, kommen Auswahlboxen zum Einsatz. Dazu zählen Parameter für das Bewertungsschema, das Chatthema und die Nachbewertungsrunde. Bei den Parametern für Termin und Chatdauer ist das jQuery[21] Plugin *datetimepicker*[22] eingebaut, um die Eingabe zu erleichtern (siehe Abbildung 4.4). Zusätzlich sorgt das Plugin dafür, dass Zeit- und Datumseingaben passend formatiert an das System übergeben werden.

In der Studienbearbeitung können außerdem neue Chatthemen angelegt werden. Mit dem Button *Neues Chatthema anlegen* wird eine entsprechende Ansicht aufgerufen, in der Name und Text des Chatthemas angegeben werden können. Zur komfortablen Eingabe des Textes wird das jQuery Plugin *TinyMCE*[23] verwendet, dass in Abbildung 4.5 illustriert ist.

Für den Fall, dass sehr viele Läufe in einer Studie angelegt werden, kann das jQuery Plugin *sorttable*[24] helfen die Übersichtlichkeit zu erhöhen. Mit einem Klick auf den Kopf einer Tabellenspalte der Studienlaufparameter werden diese wahlweise aufsteigend oder absteigend sortiert.

[21] http://jquery.com
[22] http://xdsoft.net/jqplugins/datetimepicker
[23] http://www.tinymce.com
[24] http://www.kryogenix.org/code/browser/sorttable

Abbildung 4.4: Auswahl des Termins beim anlegen eines Studienlaufs

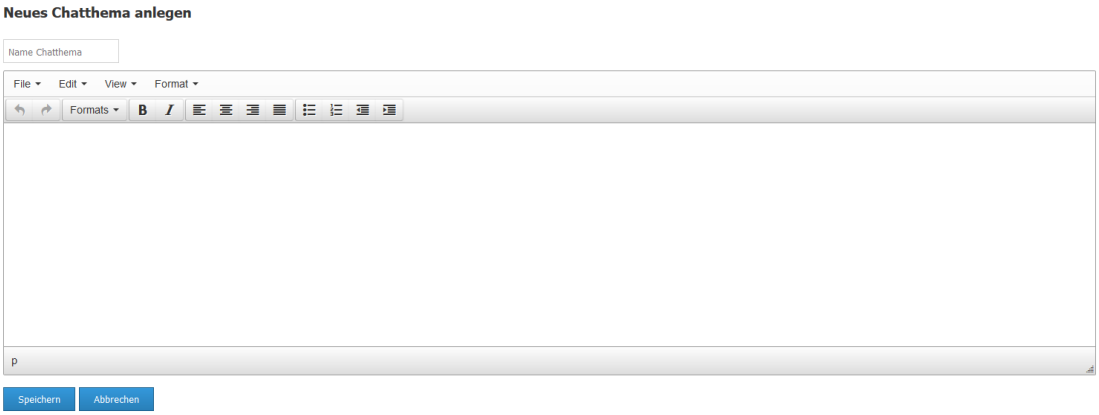

Abbildung 4.5: TinyMCE Editor

4.4.2 Auswertung

Im Rahmen dieser Arbeit ist die Auswertung von Reputationsalgorithmen interessant. Eine mögliche Ausprägung wird im Folgenden erläutert. Unter dem Menüpunkt *Auswertung* des Navigationsmenüs kommt analog zur Studienverwaltung das Accordion zum Einsatz, wobei jedes aufklappbare Element die Auswertung eines Studienlaufs beinhaltet. Eine solche Auswertung besteht aus allgemeinen Informationen zum Studienlauf und einer detaillierten Tabelle, die mit Beispieldaten in Abbildung 4.6 gezeigt wird. In der Tabelle sind die Teilnehmer des Studienlaufs in x- und y-Richtung aufgelistet. In den Zellen der entsprechenden Kombinationen sind die Bewertungen eingetragen. So hat *hans23* zum Beispiel *peter92* zehn Mal positiv bewertet und zwei Mal negativ. Die Differenz ergibt somit einen Gesamtbewertungswert von acht, der für die berechnete Rangfolge berücksichtigt wurde. Analog dazu geben die Werte der Spalten *Summe geschriebener Nachrichten* und *Summe der Bewertungen* an, wie viele davon der Proband in der entsprechenden Zeile getätigt hat. Bei Bewertungen wird zwischen positiven und negativen unterschieden. Zusätzlich lässt sich über der Tabelle ablesen, wie viele Teilnehmer am Studienlauf teilnahmen sowie die Summen der Nachrichten und Bewertungen des Studienlaufs.

Neben Nachrichten und Bewertungen ist für jeden Probanden die jeweilige Platzierung angegeben. *Rang subjektiv* gibt die aggregierte subjektive Rangfolge an. *Rang berechnet* repräsentiert die berechnete Rangfolge. So wurde hans23 beispielsweise von seinem Mitprobanden subjektiv auf Platz 1 geschätzt während die Berechnung des Reputationsalgorithmus Platz 2 ergab.

Der interessanteste und aussagekräftigste Wert ist der Rangkorrelationskoeffizient, da er den Vergleich zwischen der beiden Rangfolgen berechnet [Liu14]. Er wird zusammen mit den Parametern des Studienlaufs über der Tabelle unter *Rangkorrelation* dargeboten.

Die Ansicht einer Studienlaufauswertung wurde so gewählt, da sie dem Studienleiter weitere Rückschlüsse außer dem Rangkorrelationskoeffizient ermöglichen soll. Man könnte zum Beispiel die Anzahl der Nachrichten oder Bewertungen eines Probanden seiner Platzierung in den Rangfolgen gegenüberstellen. Möglicherweise lassen sich hierbei Zusammenhänge feststellen, ob besonders aktive Probanden besser beurteilt werden, als eher passive oder ob die Qualität der Nachrichten ein einschlägige Rolle spielt. Interessant könnte auch sein, ob es einen Zusammenhang zwischen Nachrichtenanzahl und Bewertungsfreudigkeit gibt. Am Beispiel der Abbildung 4.6 kann man ablesen, dass viele geschriebene Nachrichten zu einer besseren Platzierung in beiden Rangfolgen führt.

Weiter könnte man die Bewertungen näher betrachten. Wie schneiden Probanden mit zum Beispiel 20 positiven und 11 Negativen Bewertungen im Vergleich zu Probanden mit zehn positiven und einer negativen ab. Beide hätten in dem Fall eine Bewertungsdifferenz von neun. Eine Gegenüberstellungen der Platzierungen dieser beiden Probanden in Bezug auf die Rangfolgen wäre von Interesse. An der Bewertungsstruktur der einzelnen Probanden könnte man ableiten, ob ein bestimmter Proband vielleicht grundsätzlich gerne negativ oder positiv bewertet. Die Parameter eines Studienlaufs können dem Studienleiter helfen zukünftige Läufe zu optimieren. Die Studiendauer könnte zum Beispiel optimiert werden, wenn mehrere Läufe mit sonst identischen Pa-

rametern miteinander verglichen werden. Wenn ein zehn minütiger Studienlauf zu wenig Nachrichten generiert und ein zwanzigminütiger zu viele, so können 15 Minuten möglicherweise optimal sein. Analog dazu könnte das Chatthema oder die Anzahl der Probanden optimiert werden.

Die Auswertung kann um die persönlichen Daten der Probanden erweitert werden. Dann könnten Fragen beantwortet werden, ob beispielsweise das Alter oder der Bildungsgrad von Probanden Einfluss auf deren Verhalten während der Nutzerstudie hat.

Auswertung

Chatdauer	**10:00 Minuten**	∑ Probanden	**4**	Rangkorrelation	**0.85**
Chatthema	**Bauarbeiter Dilemma**	∑ Nachrichten	**126**		
Ratingschema	**Positiv | Negativ**	∑ Bewertungen	**127**		
Zeichenanzahl	**100**				

↘	hans23 (+/−)	peter92 (+/−)	sepp31 (+/−)	anna10 (+/−)	Rang subjektiv	Rang berechnet	∑ geschriebener Nachrichten	∑ Bewertungen (+/−)
hans23	-	8 (10/2)	5 (6/1)	6 (6/0)	1	2	41	24 (32/8)
peter92	6 (7/1)	-	6 (8/2)	7 (10/3)	2	1	36	20 (27/7)
sepp31	10 (15/5)	4 (6/2)	-	1 (6/5)	3	3	33	13 (18/5)
anna10	8 (10/2)	8 (11/3)	2 (4/2)	-	4	4	16	14 (22/8)

Abbildung 4.6: Mögliches Auswertungsbeispiel

4.5 Sichten der Probanden

In den folgenden Unterkapiteln wird der Durchlauf eines Studienlaufs aus der Sicht des Probanden dargestellt, bestehend aus der Eingabe persönlicher Daten, dem Chat, der Nachbewertungsrunde und der Angabe der subjektiven Rangfolge. Dabei wird besonders auf die technische Umsetzung der im Konzept vorgestellten Aspekte eingegangen.

4.5.1 Persönliche Daten

Sobald der Studienleiter den Studienlauf gestartet hat, gelangen alle eingeloggten Probanden zu einer Ansicht, in der persönlicher Daten eingegeben und das Thema des Chats durchgelesen werden soll. (vgl. Abbildung 4.7) Pflichtfelder werden als solche gekennzeichnet, sofern sie vorgesehen sind. Sollte ein Proband ein Pflichtfeld leer lassen, wird eine passende Fehlermeldung angezeigt. Zusätzlich zu den persönlichen Daten wird dem Probanden das Thema des Chats dargeboten. Die Eingabe der persönlichen Daten und das Lesen des Chatthemas werden durch einen Klick auf den Button *Weiter* bestätigt.

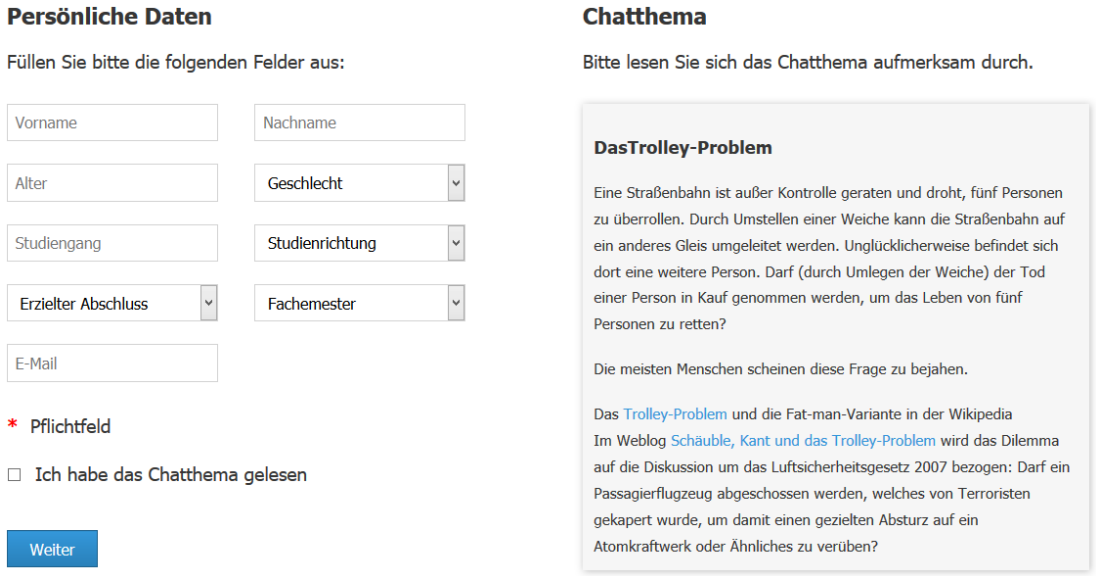

Abbildung 4.7: Ansicht zur Angabe persönlicher Daten und zum Lesen der Chatthemas

4.5.2 Chat

Die Chatansicht bietet den Probanden die Möglichkeit zu einem vorgegebenen Thema zu chatten und Chatnachrichten der Anderen zu bewerten. Eine Illustration des Chatfensters kann in Abbildung 4.8 eingehen werden. Vor jeder Chatnachricht befinden sich Bewertungsbuttons analog zum vorgegebenen Bewertungsschema. Die Bewertungsbuttons von eigenen und bereits bewerteten Beiträgen werden ausgegraut dargestellt und können nicht geklickt werden. Da keine konträren Bewertung möglich sein sollen, werden alle Bewertungsoptionen einer Nachricht ausgegraut, sobald eine davon geklickt worden ist. Des Weiteren wird das Symbol des nicht gewählten Bewertungsbuttons entfernt, damit im Nachhinein erkennbar ist, welche Bewertung abgegeben wurde. Angenommen das Bewertungsschema bietet die beiden Optionen *positiv (+)* und *negativ (-)*. Nun wird eine Chatnachricht positiv bewertet. Infolge dessen werden beide Buttons ausgegraut und das Minussymbol im entsprechenden Button entfernt. Somit ist es ersichtlich, dass diese Chatnachricht positiv bewertet wurde. Jedem Proband wird zufällig eine Farbe zugeordnet, die für seinen Alias in den Chatnachrichten verwendet wird. Dies fördert die Wiedererkennung und sorgt für Übersichtlichkeit im Chat. In einem Extrabereich des Chatfensters wird eine Liste aller im Studienlauf befindlichen Probanden dargeboten. Damit hat jeder Proband ein Gefühl für die Chatumgebung und weiß, wie viele Mitprobanden am Chat teilnehmen. Neben dem Eingabefeld für eigene Nachrichten ist der Alias des Probanden positioniert. Da der Alias zufällig generiert wird und keinen Bezug zum Proband hat, ist es für diesen hilfreich einen visuellen Hinweis zu haben, wer er im Chat ist. Zwischen dem Eingabefeld und dem Sendenbutton befindet sich ein Zeichenzähler. Er zeigt dem Proband die restliche zur Verfügung stehende Zeichenanzahl für die aktuelle Chatnachricht. Sie verringert sich dynamisch nach dem eintippen eines jeden Zeichens. Bei zehn verbleibenden Zeichen färbt sich die Zeichenanzeige rot, um die Aufmerksamkeit des Probanden auf den auslaufenden Zeichenzähler zu richten. Weiter kann die erlaubte Zeichenanzahl nicht überschritten werden, da das System keine weiteren Zeichen akzeptiert. In einem weiteren Fenster wird das Chatthema dargeboten, damit die Probanden jederzeit die Möglichkeit ha-

ben es bei Bedarf nochmals zu lesen. Komplettiert wird die Chatansicht von einem Countdown, der über die verbleibende Chatdauer informiert.

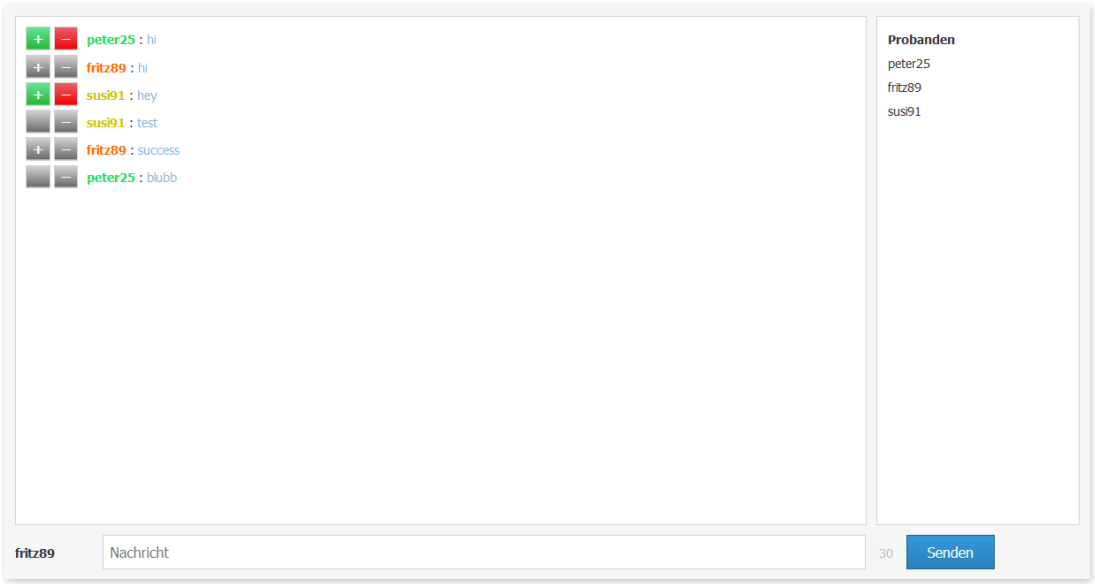

Abbildung 4.8: Das Chatfenster

4.5.3 Nachbewertungen

Die Ansicht für eine Nachbewertungsrunde ist genau aufgebaut wie die Chatansicht mit dem Unterschied, dass es nicht mehr möglich ist zu chatten. Das Feld zum eingeben von Chatnachrichten wird nicht mehr angezeigt. Neu hinzu kommt ein erklärender Text, was der Proband in dieser Ansicht tun soll. Sobald der Proband mit den Nachbewertungen fertig ist, klick er auf den Button *Weiter*. Eine Veranschaulichung dieser Ansicht findet sich in Abbildung xx.

4.5.4 Subjektive Rangfolge

Die Ansicht für die Angabe einer subjektiven Rangfolge besteht aus einem erklärenden Text und dem jQuery UI Skript *Sortable*[25]. Mit Sortable können Listen per Drag-and-Drop sortiert werden. Abbildung 4.9 zeigt den Aufbau dieser Ansicht, bestehend aus zwei Boxen. In der linken Box sind alle Probanden in alphabetischer Rangfolge aufgelistet. Durch die alphabetische Sortierung soll sichergestellt werden, dass keine Beeinflussung des bearbeitenden Probanden durch eine zufällige vorgegebene Rangfolge passiert. Da der bearbeitende Proband nicht die Möglichkeit bekommen soll sich selbst zu bewerten, wird dessen Alias nicht in der Probandenliste aufgeführt. Die rechte Box ist am Anfang noch leer. Nun soll der bearbeitende Proband Elemente von der linken Box in die rechte verschieben. Elemente in der rechten Box können beliebig oft neu positioniert werden. Sobald der Proband mit der subjektiven Rangfolge fertig ist, schickt er sie mit einem Klick auf den *Rangfolge abschicken* Button ab.

[25] http://jqueryui.com/sortable

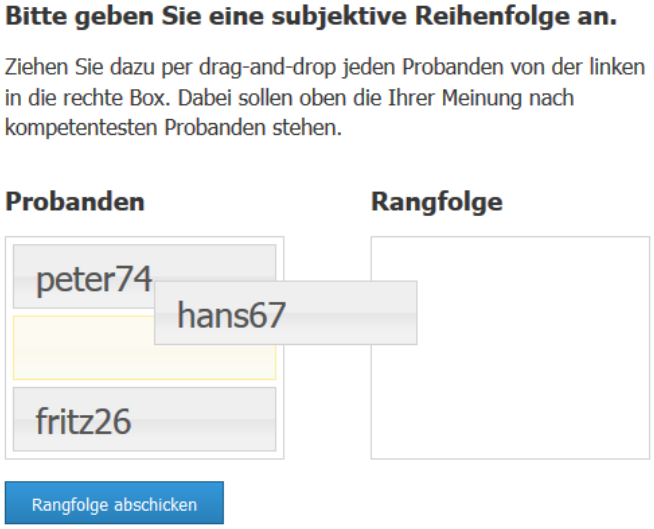

Abbildung 4.9: Ansicht zum sortieren von Probanden

4.6 Fehlerbehandlung

Während der Durchführung eines Studienlaufs können grundsätzlich Fehler passieren, entweder technischer Natur oder durch falsche Aktionen von Probanden. Es kann zum Beispiel vorkommen, dass die Internetverbindung verloren geht oder dass ein Rechner aufgrund eines Hardwaredefekts abstürzt. Ein Proband könnte versehentlich das Browserfenster schließen oder den Zurückbutton betätigen. In allen genannten Fällen geht die Verbindung zum WebSocket naturgemäß verloren und muss neu aufgebaut werden, wenn der Proband zurückkehrt.

Da für jeden Probanden die aktuelle Phase in der Datenbank festgehalten wird, ist ein unkomplizierter Wiedereintritt in die Nutzerstudie möglich. Dazu muss die URL der Nutzerstudie nochmals aufgerufen und die Nutzerkennung eingegeben werden. Das System leitet den Probanden automatisch an die richtige Ansicht weiter, wenn der Studienleiter dies explizit erlaubt. Das Wiedereintreten in den Studienlauf muss vom Studienleiter autorisiert werden, da Situationen eintreten können, in denen ein Wiedereintritt nicht sinnvoll wäre. Wenn ein Proband zum Beispiel zu Beginn des Chats ausscheidet und zum Ende des Chats zurückkehren will, so hat er nahezu die gesamte Konversation verpasst. Infolge dessen hat er sich keine ausreichende Meinung über die anderen Probanden bilden können und wird nicht in der Lage sein eine subjektive Rangfolge verlässlich abzugeben.

Wenn Probanden den Zurückbutton des Browsers betätigen, dürfen sie nicht die Möglichkeit erhalten bereits abgeschickte Daten nochmal abzuschicken oder zu überschreiben. Dies wird durch WebSockets implizit verhindert, da die Seite beim wechseln von Ansichten nicht neu geladen wird. Die einzige Ansicht, in die ein Proband mit dem Zurückbutton gelangen kann, ist die Indexseite des Programms, auf der die Nutzerkennung eingegeben wird. Mit erneuter Eingabe der Nutzerkennung kann der Proband in die aktuelle Phase des Studienlaufs zurückkehren.

4.7 Zusammenfassung

Dieses Kapitel beschäftigte sich mit der technischen Umsetzung der webbasierten Auswertungsplattform anhand des Studiendesigns und der damit verbundenen Anforderungen in Kapitel 3. Dabei werden vom Datenmodell alle Parameter der jeweiligen Nutzerstudie, analog zur Tabelle 4.1 auf Seite 32, erfasst. Des Weiteren werden alle Angaben der Probanden gespeichert und zwar Chatnachrichten, Bewertungen, Rangfolgen und Persönliche Daten. Das Datenmodell realisiert zusätzlich die Unterteilung in Studien und Studienläufe sowie die Speicherung der Parameter in unterschiedlichen Tabellen, um Überschreibungen zu verhindern.

Im Unterkapitel 4.2 zum Phasenmodell wurde erläutert, wie der Studienleiter mit Hilfe des Studienleiter Panels (vgl. Abbildung 4.1) einen Studienlauf überwachen kann. Weiter wurde auf die technische Realisierung der Phasen mit Hilfe des Attributs *phase* eingegangen.

Im darauf folgenden Kapitel 4.3 wurden WebSockets vorgestellt und deren besondere Eignung und Vorteile für Chatanwendungen mittels bidirektionaler Kommunikation erklärt.

In den beiden Unterkapiteln 4.4 und 4.5 wurden die Sichten des Studienleiters und der Probanden innerhalb der Auswertungsplattform beschrieben. Der Studienleiter hat Zugriff auf die beiden Sichten *Studienverwaltung* und *Auswertung*. In der *Studienverwaltung* können Studien und Studienläufe angelegt und verwaltet werden. Hierbei kommen mehrere JavaSkript-basierte Skripts zum Einsatz, um die Übersichtlichkeit und Benutzerfreundlichkeit zu erhöhen. In der Sicht *Auswertung* wird eine Möglichkeit gezeigt, wie die gesammelten Daten für die Auswertung eines Reputationsalgorithmus genutzt werden können.

Schließlich wurde in Unterkapitel 4.6 auf mögliche Fehlerszenarien und deren Behandlung während eines Studienlaufs eingegangen. Genauer wurde erklärt, wie Probanden nach einem technischen Problem in den Studienlauf zurückkehren können und warum das Betätigen des Zurückbuttons des Browsers zu keinen Problemen führt.

Kapitel 5

Resümee und Ausblick

Der zentrale Bestandteil dieser Arbeit war die Konzipierung und Realisierung einer webbasierten Auswertungsplattform für Reputationsalgorithmen. Die Plattform ermöglicht die Durchführung von vergleichenden Nutzerstudien, in denen Meinungen von Probanden als Referenzdaten gesammelt werden, um mit numerischen Reputationswerten verglichen werden zu können.

Das Konzept der Auswertungsplattform zielt auf die Überprüfung von Reputationswerten auf Plausibilität. Reputationswerte können als plausibel angesehen werden, wenn die Eigenschaften der Akzeptanz und Konsistenz erfüllt sind. Die Messung der Akzeptanz erfolgt über Fragebögen im Anschluss an die Nutzerstudie. Im Rahmen dieser Arbeit ist die Messung der Konsistenz von größerer Bedeutung als die Messung der Akzeptanz. Im Zuge dessen wurde auf die technische Umsetzung von Fragebögen innerhalb der Auswertungsplattform verzichtet. Des Weiteren gibt es im Internet eine Vielzahl von Anbietern für die Erstellung von Online-Fragebögen, wie zum Beispiel UmfrageOnline[26], Survio[27] oder SurveyMonkey[28], so dass eine separate Implementierung eines Fragebogen-Moduls nicht zwingend erforderlich ist.

Die zentrale Komponente der Auswertungsplattform ermöglicht die Messung der Konsistenz von numerischen Reputationswerten in Bezug auf Meinungen von Probanden. Anders gesprochen soll mit Hilfe der Auswertungsplattform ermittelt werden können, inwiefern berechnete Reputationswerte eines Reputationsalgorithmus mit der subjektiven Einschätzung eines Menschen übereinstimmen.

Dabei sieht das Konzept eine Unterteilung der Auswertungsplattform in die drei Phasen Vorbereitungsphase, Durchführungsphase und Auswertungsphase vor. Jede Phase kann aus der Sicht des Studienleiters oder eines Probanden betrachtet werden. In der Vorbereitungsphase legt der Studienleiter die gewünschten Nutzerstudien an und startet sie. Die Probanden haben in der Vorbereitungsphase die Aufgabe sich mit der Nutzerkennung in die Nutzerstudie einzuloggen, persönliche Daten einzugeben und das Thema des Chats zu lesen.

In der Durchführungsphase werden die zur Auswertung nötigen Daten gesammelt und zwar Bewertungen und subjektive Rangfolgen, die von den Probanden abgegeben wur-

[26] http://www.umfrageonline.com
[27] http://www.survio.com
[28] http://de.surveymonkey.com

den. Zunächst sollen die Probanden zu einem kontroversen Thema, wie beispielsweise einem Dilemma, miteinander kommunizieren. Dabei kann jeder Proband Nachrichten der anderen Probanden bewerten. Die Bewertungen sind notwendig, um Reputationswerte zu berechnen. Im Anschluss an die Chatphase kann der Studienleiter optional eine Nachbewertungsphase einstellen. Dabei wird den Probanden die Chathistorie nochmals dargeboten und die Möglichkeit gegeben bisher nicht bewertete Chatnachrichten nachträglich zu bewerten. Während der Chatphase gewinnt der Proband anhand der Chatnachrichten der anderen Probanden eine subjektive Einschätzung über jeden Probanden. Nach einer optionalen Nachbewertungsrunde sollen die Probanden anhand der gewonnenen Einschätzung eine subjektive Rangfolge der anderen Probanden angeben. Bei auftretenden Problemen, wenn zum Beispiel zu wenig Nachrichten geschrieben werden oder wenn Probanden technische Probleme haben, entscheidet er über den weiteren Fortlauf. Bei stockendem Chatverlauf könnte er das Chatthema austauschen. Sollte ein Proband aus der Nutzerstudie ausscheiden, so entscheidet der Studienleiter, ob dieser Proband zurückkehren darf. Im letzten Schritt der Durchführungsphase sieht das Konzept die Bearbeitung des Fragebogens vor.

Die Auswertungsphase verläuft ohne die Beteiligung von Probanden. Hierbei werden anhand der Bewertungen Reputationswerte berechnet und entsprechend dem Vergleichsmaß mit den subjektiven Rangfolgen verglichen.

Aus technischer Sicht ist die Auswertungsplattform in Studien und Studienläufe unterteilt, wobei eine Studie mehrere Studienläufe kapselt. Dem Studienleiter steht der Bereich *Studienverwaltung* zur Verfügung, in dem er Studien und Studienläufe anlegen, verwalten und löschen kann. Beim anlegen von Studien und Studienläufen werden die Parameter für die Anzahl der Probanden, die Chatdauer, das Bewertungsschema, die erlaubte Zeichenanzahl pro Chatnachricht und die Erfordernis einer Nachbewertungsrunde spezifiziert.

Der Ablauf eines Studienlaufs berücksichtigt die Notwendigkeit, dem Studienleiter jederzeit die volle Kontrolle über diesen zu gewähren. Somit muss der Studienleiter den Studienlauf, den Chat, die optionale Nachbewertungsrunde und die Angabe der subjektiven Rangfolgen explizit starten. Dies wird mit der Hilfe von Phasen realisiert. Eine Phase besteht aus einem Aktivitätszustand, in dem die Probanden etwas tun müssen, und einem Wartezustand, in dem die Probanden auf den Beginn der nächsten Phase warten.

Ein Studienlauf ist mit WebSockets realisiert, einem TCP-basierten Netzwerkprotokoll zur Umsetzung bidirektionaler Kommunikation zwischen Webserver und Client/ Webbrowser. WebSockets sind für die Umsetzung von Onlinechats sehr gut geeignet, da sie den Datenaustausch von Server zum Client ohne vorherige Anfrage durch den Client ermöglichen [Abt15]. Der Webbrowser eines Probanden muss somit keine Anfragen stellen, ob es neue Chatnachrichten gibt. Sobald eine neue Chatnachricht den WebSocket-Server erreicht, werden alle Clients beziehungsweise Probanden benachrichtigt.

Die webbasierte Auswertungsplattform ist eine technische Hilfe für die Durchführung von Nutzerstudien zur Auswertung von Reputationsalgorithmen. Die tatsächliche Auswertung und Interpretation der gesammelten Daten erfolgt individuell durch der Nutzer

der Auswertungsplattform. Es bleibt ihm zum Beispiel überlassen, ob persönliche Daten der Probanden in der Auswertung berücksichtigt werden sollen. Ebenso kann eine passende Methode gewählt werden, um berechnete und subjektiven Rangfolge miteinander zu vergleichen.

5.1 Nicht realisierte Funktionalitäten

Im Folgenden wird auf Funktionalitäten eingegangen, die im Konzept vorgesehen sind und in der technischen Umsetzung aus Zeitgründen nicht realisiert wurden. Während des Chats ist es aktuell nicht möglich diesen zu pausieren, um Änderungen an den Parametern vorzunehmen. Der Studienleiter hat gegenwärtig nur im Chat die Möglichkeit, Nachrichten an die Probanden zu verschicken. Es gibt keine Funktion, mit der Nachrichten an einen bestimmten Probanden oder außerhalb der Chatphase möglich sind. Des Weiteren können Probanden in der aktuellen Version der Auswertungsplattform nach einem technischen Problem auch ohne die Bestätigung des Studienleiters in den Studienlauf zurückkehren. Die Auswertung von Studien und Studienläufen ist nicht realisiert, da der erste Prototyp der Auswertungsplattform hauptsächlich als technisches Hilfsmittel zur Erfassung der benötigten Daten zu sehen ist.

5.2 Ausblick

Für zukünftige Arbeiten wäre die Erweiterung der Auswertungsplattform um eine flexible Auswertungskomponente interessant. Idealerweise sollte der Anwender der Auswertungsplattform die Möglichkeit erhalten genau angeben zu können, welche Daten zum Vergleich zwischen berechneter und subjektiver Rangfolge herangezogen werden. Konkreter wäre es interessant nur eine Teilgruppe der Probanden eines Studienlaufs zu berücksichtigen, wie beispielsweise nur Männer/ Frauen, nur Probanden einer bestimmten Altersgruppe oder nur Naturwissenschaftler. Die Auswertung von solchen Teilgruppen könnte hilfreich sein, um mehrere davon miteinander zu vergleichen. Dadurch wären Rückschlüsse möglich, ob ein Reputationsalgorithmus in Bezug auf eine Teilgruppe konsistentere Ergebnisse liefert als auf eine andere Teilgruppe. Ein weiterer Teil dieser Erweiterung könnte die Implementierung von gängigen Reputationsalgorithmen sein. Dadurch könnte der zu testende Reputationsalgorithmus zusätzlich zu Meinungen von Probanden mit den gängigen Reputationsalgorithmen verglichen werden. Schließlich könnte die Art und Weise, wie berechnete und subjektive Rangfolgen miteinander verglichen werden, parametrisiert werden. So könnte der Anwender der Auswertungsplattform zum Beispiel wählen, ob die Rangkorrelationsberechnung oder ein anderes Vergleichsverfahren verwendet werden soll.

Des Weiteren wäre es interessant Testläufe mit Probanden zu machen. So könnten für Parameter, wie Chatdauer, Chatthema oder die erlaubte Zeichenanzahl, sinnvolle Werte gefunden werden. Außerdem könnten Umfragen auf Metaebene durchgeführt werden, um die Benutzerfreundlichkeit der Auswertungsplattform zu verbessern.

Anhang A

Datenbankmodell

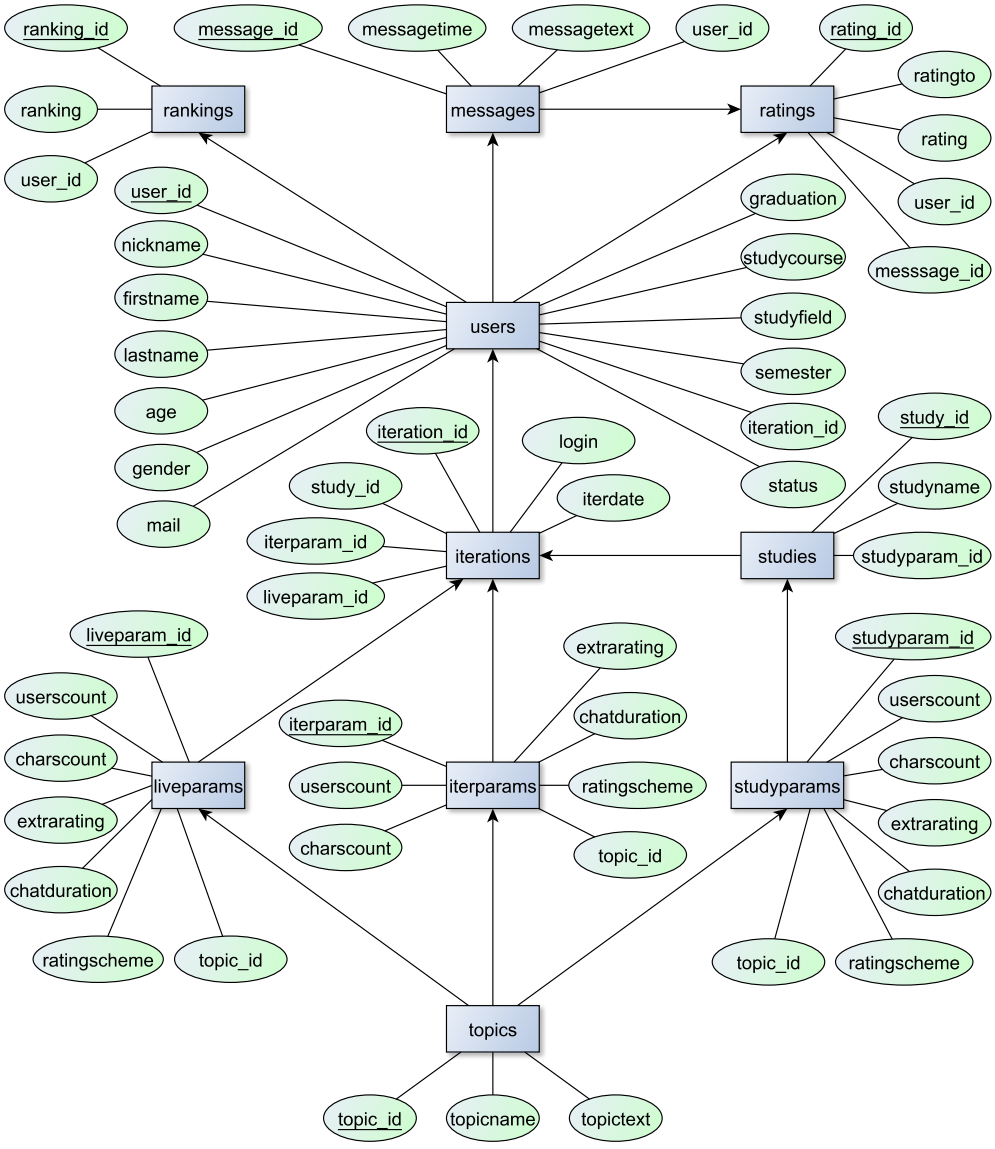

Abbildung A.1: Datenbankmodell der Auswertungsplattform

Abbildungsverzeichnis

2.1 Gleicher *Reliability Trust*, unterschiedlicher *Decision Trust* (Bildquelle [Jøs07]) . . . 6
2.2 Prinzip des transitiven Vertrauens (Bildquelle [Jøs07]) . . . 7
2.3 Übersicht der Status im Forum *die-informatiker.net* . . . 8
2.4 Veranschaulichung zentral gesteuerter Reputationssysteme (Bild aus [JIB07], Seite 14) . . . 11
2.5 Veranschaulichung verteilter Reputationssysteme (Bild aus [JIB07], Seite 15) 11
2.6 Vorteile von Reputationssystemen . . . 15

3.1 Beispiel 1 für ein kontroverses Dilemma - Möbelpacker . . . 22
3.2 Beispiel 2 für ein kontroverses Dilemma - Sportunterricht . . . 23
3.3 Phasen der Auswertungsplattform . . . 24
3.4 Vorbereitungsphase des Studienleiters . . . 26
3.5 Vorbereitungsphase der Probanden . . . 27
3.6 Durchführungsphase des Studienleiters . . . 28
3.7 Durchführungsphase des Probanden . . . 29

4.1 Studienleiter Panel . . . 35
4.2 Studienverwaltung im Hauptzustand . . . 37
4.3 Studienverwaltung im Zustand *Studien* . . . 37
4.4 Auswahl des Termins beim anlegen eines Studienlaufs . . . 39
4.5 TinyMCE Editor . . . 39
4.6 Mögliches Auswertungsbeispiel . . . 41
4.7 Ansicht zur Angabe persönlicher Daten und zum Lesen der Chatthemas 42
4.8 Das Chatfenster . . . 43
4.9 Ansicht zum sortieren von Probanden . . . 44

A.1 Datenbankmodell der Auswertungsplattform . . . 51

Tabellenverzeichnis

4.1 Erklärung der Parameter 32
4.2 Gegenüberstellung von Hauptphasen und Zuständen (grün: Vorbereitungsphase, gelb: Durchführungsphase) 33
4.3 Mögliche Ausprägungen des Attributs *phase* 35

Zur Arbeit gehörendes

Die digitale Version dieser Diplomarbeit befindet sich im GitLab, dass vom PMS-Lehrstuhl verwaltet wird. Das zugehörige Verzeichnis ist *alex/da_michael_weisbein*. Neben dem Programmcode der Auswertungsplattform und einem SQL-Dump der Datenbank befinden sich dort alle LaTeX-Dateien.

Installationsanleitung

1. XAMPP installieren (Version 5.6.8 oder neuer)
   beinhaltet bereits PHP, MySQL und phpMyAdmin

2. Ordner *platform* ins *htdocs*-Verzeichnis von XAMPP kopieren

3. Neue Datenbank im phpMyAdmin erstellen (Standardeinstellungen)

4. SQL-Dump im phpMyAdmin in die erstellte Datenbank importieren
   (Der SQL-Dump befindet sich ebenfalls in GitLab und heißt *platform-dump.sql*)

5. Datenbankkonfiguration einstellen in *platform/lib/DatabaseFunctions.php*
   In dieser Arbeit verwendete Konfiguration:
   servername: localhost
   username: root
   password: *leer*
   database: platform

6. Apache und MySQL Module im XAMPP Control Panel starten

7. WebSocket-Server starten

   (a) Console öffnen und ins *platform/app*-Verzeichnis wechseln

   (b) *php -q server.php* eingeben

8. Plattform aufrufen mit *http://localhost/platform*

Verzeichnisstruktur der Auswertungsplattform

Verzeichnis	**Erklärung**
app	
controllers	Steuerung der Studienverwaltung und der Indexseite
css	Alle CSS-Dateien, insb. *style.css* als Haupt-Stylesheet
img	Alle Bilder, die in der Plattform benötigt werden
js	Alle JavaScript-Dateien zur Steuerung der Plattform
layout	Wiederverwendbare HTML-Konstrukte
lib	Plattform-eigene Funktionen für WebSocket und DB-Zugriff
vendor	Externe Skripte, die in der Plattform benötigt werden
views	Ansichten, die während des Studienlaufs injiziert werden
views/server	Fallback-Ansichten bei Problemen
Wichtigste Dateien	
app/server.php	Der WebSocket-Server
controllers/FrontendController.php	Controller für die Index-Seite
controllers/StudiesController.php	Controller für die Studienverwaltung
js/platform.js	Realisiert alle Interaktionen im Studienlauf
js/platform-cp.js	Realisiert alle Studienleiter-Interaktionen im Studienlauf
platform.php	Hauptansicht eines Studienlaufs
platform-cp.php	Hauptansicht eines Studienlaufs für den Studienleiter

Literaturverzeichnis

[Abt15] Abts, Dietmar: Bidirektionale Kommunikation mit WebSocket. In: *Masterkurs Client/Server-Programmierung mit Java.* Springer Fachmedien Wiesbaden, 2015, S. 189–205

[Atk09] Atkinson, Cliff: *The backchannel: How Audiences Are Using Twitter and Social Media and Changing Presentations Forever.* New Riders, 2009

[BP14] Bry, François ; Pohl, Alexander: Backstage: A Social Medium for Large Classes. Version: 2014. `http://www.pms.ifi.lmu.de/publikationen/#PMS-FB-2014-2`. In: Keuper, Frank (Hrsg.) ; Arnold, Heinrich (Hrsg.): *Campus Transformation – Education, Qualification and Digitalization.* Berlin : Logos Verlag, 2014, 255–280

[CV05] Cheng, Ran ; Vassileva, Julita: Adaptive Reward Mechanism for Sustainable Online Learning Community. In: *AIED*, 2005, S. 152–159

[For03] Foreman, Joel: Generation. In: *EDUCAUSE review* 500 (2003)

[FR14] Fielding, Roy ; Reschke, Julian: *Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing.* Internet Engineering Task Force (IETF), 2014

[GGM05] Gyongyi, Zoltan ; Garcia-Molina, Hector: Web spam taxonomy. In: *Proceedings of the first international workshop on adversarial information retrieval on the web (AIRWeb 2005)*, 2005

[GGMP04] Gyöngyi, Zoltán ; Garcia-Molina, Hector ; Pedersen, Jan: Combating web spam with trustrank. In: *Proceedings of the Thirtieth international conference on Very large data bases, Volume 30,* VLDB Endowment, 2004, S. 576–587

[Hil84] Hilbig, Winfried: Akzeptanzforschung neuer Bürotechnologien. Ergebnisse einer empirischen Fallstudie. In: *Office Management* 32 (1984), Nr. 4, S. 320–323

[HYD09] Hong, Liangjie ; Yang, Zaihan ; Davison, Brian D.: Incorporating participant reputation in community-driven question answering systems. In: *Proceedings of the computational Science and Engineering, 2009. CSE'09. International Conference on* Bd. 4 IEEE, 2009, S. 475–480

[IBJR03] Ismail, Roslan ; Boyd, Colin ; Josang, Audun ; Russell, Selywn: A security architecture for reputation systems. In: *E-Commerce and Web Technologies.* Springer, 2003, S. 176–185

[JIB07] JØSANG, Audun ; ISMAIL, Roslan ; BOYD, Colin: A survey of trust and reputation systems for online service provision. In: *Decision support systems* 43 (2007), Nr. 2, S. 618–644

[Jøs07] JØSANG, Audun: Trust and reputation systems. In: *Foundations of security analysis and design IV*. Springer, 2007, S. 209–245

[Kle99] KLEINBERG, Jon M.: Authoritative sources in a hyperlinked environment. In: *Journal of the ACM (JACM)* 46 (1999), Nr. 5, S. 604–632

[Liu14] LIU, Brandon: *Better than PageRank: Hitting Time as a Reputation Mechanism*. Harvard College, 2014

[MC96] MCKNIGHT, D. H. ; CHERVANY, Norman L.: The meanings of trust / Carlson School of Management, University of Minnesota. Citeseer, 1996 (96-04). – MISRC Working Paper Series

[NHC08] NOBARANY, Syavash ; HARATY, Mona ; COSLEY, Dan: GePuTTIS: General Purpose Transitive Trust Inference System for Social Networks. In: *AAAI Spring Symposium: Social Information Processing*, 2008, S. 66–71

[PBMW99] PAGE, Lawrence ; BRIN, Sergey ; MOTWANI, Rajeev ; WINOGRAD, Terry: *The PageRank citation ranking: bringing order to the Web*. Stanford InfoLab, 1999

[PSD02] PUJOL, Josep M. ; SANGÜESA, Ramon ; DELGADO, Jordi: Extracting reputation in multi agent systems by means of social network topology. In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* ACM, 2002, S. 467–474

[Puj06] PUJARI, Pinak: *Independent Study Report on Reputation Systems*. John Hopkins University, USA, 2006

[RKZF00] RESNICK, Paul ; KUWABARA, Ko ; ZECKHAUSER, Richard ; FRIEDMAN, Eric: Reputation systems. In: *Communications of the ACM* 43 (2000), Nr. 12, S. 45–48

[RZ02] RESNICK, Paul ; ZECKHAUSER, Richard: Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. In: *The Economics of the Internet and E-commerce* 11 (2002), Nr. 2, S. 23–25

[Spr88] SPREMANN, Klaus: Reputation, garantie, information, Schweiz. Institut für Banken und Finanzen, 1988, S. 613–629

[SS02] SABATER, Jordi ; SIERRA, Carles: Reputation and social network analysis in multi-agent systems. In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* ACM, 2002, S. 475–482

[SŠZ12] SIMONE, Antonino ; ŠKORIĆ, Boris ; ZANNONE, Nicola: Flow-based reputation: more than just ranking. In: *International Journal of Information Technology & Decision Making* 11 (2012), Nr. 03, S. 551–578

[TSP10] Tang, Jie ; Seuken, Sven ; Parkes, David C.: Hybrid transitive trust mechanisms. In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1* International Foundation for Autonomous Agents and Multiagent Systems, 2010, S. 233–240

[Vas12] Vassileva, Julita: Motivating participation in social computing applications: a user modeling perspective. In: *User Modeling and User-Adapted Interaction* 22 (2012), Nr. 1-2, S. 177–201

[VPZ14] Vavilis, Sokratis ; Petković, Milan ; Zannone, Nicola: A reference model for reputation systems. In: *Decision Support Systems* 61 (2014), S. 147–154

[WGD06] Wu, Baoning ; Goel, Vinay ; Davison, Brian D.: Topical trustrank: Using topicality to combat web spam. In: *Proceedings of the 15th international conference on World Wide Web* ACM, 2006, S. 63–72

[WV07] Wang, Yao ; Vassileva, Julita: Toward trust and reputation based web service selection: A survey. In: *International Transactions on Systems Science and Applications* 3 (2007), Nr. 2, S. 118–132

[YS02] Yu, Bin ; Singh, Munindar P.: An evidential model of distributed reputation management. In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* ACM, 2002, S. 294–301

[Zad94] Zadeh, Lotfi A.: Fuzzy logic, neural networks, and soft computing. In: *Communications of the ACM* 37 (1994), Nr. 3, S. 77–84