Universität Zürich[UZH]

Informatics and Sustainability Research

Informationsaggregation durch Textmining als Mittel zur Unterstützung bei Kaufentscheidungen im Onlinehandel

Bachelorarbeit am Institut für Informatik

Lars-Eric Windhab
Ankum, Deutschland
09-919-754

Zuständiger Professor: Prof. Dr. Lorenz Hilty

Betreuer: Yuliyan Maksimov

Abgabedatum: 31.01.2017

Danksagung

Mein erster Dank geht an Herrn Prof. Lorenz Hilty, der es mir ermöglicht hat dieses spannende Thema zu untersuchen und somit eine Forschungsmöglichkeit bot, die mir neben ein paar Nachtschichten auch viel Freude bereitet hat.
Besonderer Dank gilt ferner meinem Betreuer Yuliyan Maksimov. Er hat mich von der Auswahl des Themas bis hin zur Abgabe, wann immer nötig mit seiner wertvollen Expertise unterstützt und war darüberhinaus zu jeder Zeit für mich zu erreichen.
Last but not least danke ich meiner Familie, welche mich in allen Belangen des Studiums und schliesslich auch bei der Erstellung dieser Arbeit massgeblich unterstützt hat.

Summary

Textmining for information extraction from unstructured information is of growing importance with the permanent increase in computing capacity. In the area of customer review evaluation for products from on-line stores there are some initial research approaches reported. Such aim for feedback on product related properties. The significance of this-like extracted information still lacks sufficient interpretation of human language due to limitations of computers and software. Related research attempts to address these shortcomings by improving the computer's ability to interpret human language by improving text-mining methods. The presented work develops and tests a different approach for bridging the gaps by combining several different existing textmining analyses methods but also enable the user to apply interpretation corrections interactively, which is expected to enable improved product analysis information. To approve this a functional prototype was developed, programmed in the computer language "R", tested and evaluated. The motivation was to gain valuable insight into such textmining approaches in order to derive ideas for improved method development, which in future could improve practicability for end users and not being restricted to the use of larger enterprises.

Zusammenfassung

Textmining zur Informationsrückgewinnung aus unstrukturierter Information gewinnt mit der permanenten Zunahme der Computerleistung immer mehr an Bedeutung. Auch im Bereich der Auswertung von Kundenbewertungen zu Produkten von Onlinekaufhäusern gibt es dazu bereits einige erste Forschungsansätze. Dabei geht es darum aus unstrukturierten Kundenbewertungen Rückschlüsse auf das Produkt zu ziehen und dieses klar zu bewerten. Die Aussagekraft der extrahierten Information aus den Kundenbewertungen leidet aber bislang unter den Limitierungen des Computers menschliche Sprache zu interpretieren. Die Forschung versucht diese Unzulänglichkeiten durch eine Verbesserung der Interpretationsgabe des Computers, z.B durch eine Verbesserung der Textmining-Methoden anzugehen. Durch eine praktische, auf den Kunden bezogene Betrachtung verfolgt diese Arbeit einen anderen Ansatz. Um die Unzulänglichkeiten zu umgehen wird eine Kombination verschiedener Textmining Techniken sowie die Möglichkeit des Anwenders interaktiv einzugreifen genutzt. Damit können Fehlinterpretationen von Information korrigiert werden. Hierzu wurde ein voll funktionsfähiger Prototyp in der Programmiersprache R entwickelt, getestet und evaluiert. Die Motivation bestand darin wertvolle Einsichten aus der Nutzung von Textmining Methodik in die zukünftige Verbesserung der entsprechenden Methoden einfliessen zu lassen. Ferner sollte ein Ansatz aufgezeigt werden wie Textmining in Zukunft praktikabel und vor allem aus Endverbraucher-Sicht vereinfacht handhabbar eingesetzt werden kann, um solche Anwendungen des Textmining nicht nur auf grössere Unternehmen zu beschränken.

Inhaltsverzeichnis

Tabellenverzeichnis **vii**

Abbildungsverzeichnis **viii**

1 Einführung **1**

1.1 Ausgangslage 1

1.2 Stand der Forschung 2

1.2.1 Textmining und Kundenbewertungen 2

1.2.2 Sentimentanalyse 4

1.2.3 Der Kunde und Information 6

1.3 Die Forschungsfrage 7

2 Prototyp **8**

2.1 Werkzeuge 9

2.1.1 Die Programmiersprache: R vs Python 9

2.1.2 Bibliotheken 10

2.2 Funktionsweise 10

2.2.1 Webscraping 11

2.2.2 Textmining 12

2.2.3 Visualisierung 19

2.2.4 Wichtige Funktionen im Überblick 23

2.2.5 Sequenzdiagram 24

2.3 Schwierigkeiten bei der Implementierung 25

2.3.1 Webscraping 25

2.3.2 Produktaspekt-Browser 27

2.3.3 UTF8 Kodierung 27

2.3.4 R und Java 27

3 Diskussion **29**

3.1 Kompaktheit 29

3.1.1 Aspektextraktion 29

3.1.2 Assoziationsanalyse 29

3.1.3 Sentimentanalyse 29

3.2 Aussagekraft 30

3.2.1 Abbildung relevanter Sachverhalte 30

3.2.2 Qualität der Information 31

3.3 Limitierungen 32

4 Evaluierung 34
4.1 Evaluierungsmethodik 34
4.1.1 Referenzkunde 35
4.1.2 Testkunden 35
4.1.3 Vergleich 35
4.2 Testpersonen 35
4.3 Testprodukt 36
4.4 Auswertung 36
4.4.1 Referenzkundenauswertung 36
4.4.2 Testkundenauswertung 36
4.4.3 Synthese 38
4.5 Evaluierungsfazit 38

5 Fazit und Ausblick 40

6 Literatur 42

7 Anhänge 44
7.1 Fragebogen zur Evaluierung des Prototyps 44
7.2 USB Stick 45

Tabellenverzeichnis

1 Funktionsübersicht des Prototyps. 23
2 Ergebnisse der Referenzkundenauswertung 36
3 Ergebnisse der Testkundenauswertung: Aspektextraktion 36
4 Ergebnisse der Testkundenauswertung: Sentimentanalyse 37
5 Ergebnisse der Testkundenauswertung: Zeitaufwand 37

Abbildungsverzeichnis

1 Schematischer Aufbau des Prototyps mit den Funktionskategorien 9
2 Eingabefeld für die URL von Amazon-Produktbewertungen, die analysiert werden sollen . 11
3 Generierte Nomen-Wortwolke aus den ersten zehn Reviews des Amazon Tap bei einem minimalen Wortvorkommen von fünfzehn. 13
4 Nomenextraktion aus unstrukturiertem Text im Detail 13
5 Zusammenhang zwischen confidence und support bei der Assoziationsanalyse. 14
6 Zusammenhang zwischen lift und support bei der Assoziationsanalyse. 14
7 Korrelierende Wörter in Kundenbewertungen über einen Staubsauger, bezogen auf den Aspekt “hose” aus 100 Kundenbewertungen. 14
8 Sentimentanalyse von Sätzen, die sich auf den Aspekt *battery life* beim Amazon Tap beziehhen 16
9 Die Extraktion von Meinungssätzen im Detail 16
10 Produktaspekt-Browser: Übersicht über sechsundvierzig Kundenbewertungen zum Amazon Tap, bezogen auf den Produktaspekt “bass”, inklusive der Darstellung einiger Sätze, die den Datenpunkten entsprechen 17
11 Feedback der reaktiven Funktionen 19
12 Übersicht des User-Interface 20
13 Detaillierte Schematische Darstellung der Funktionen des Prototyps, sowie deren Interaktionen und Aufteilungen 22
14 Sequenzdiagramm, welches einen beispielhaften Ablauf einer Nutzung des Prototyps zeigt. 24
15 Der ursprüngliche Code zum direkten Einlesen einer HTML Seite mittels der nativen *rvest* Funktion *read_html()* 26
16 Der angepasste Code zum Einlesen einer HTML Seite mit Hilfe der *rSelenium* Bibliothek 26
17 Bennenung von Datenreihen und Rollen zur Vorbereitung der Darstellung durch die *googleVis* Bibliothek 27
18 *iconv* Funktion zur Kodierung von Zeichen 27

1 Einführung

1.1 Ausgangslage

Im Digitalzeitalter liegt gespeicherte Information in verschiedenen Formen vor. Ein Hauptunterscheidungsmerkmal ist dabei die Strukturiertheit der Information. Während strukturierte Information z.B in Datenbanksystemen existiert, ist das unstrukturierte Pendant dazu etwa ein Fliesstext oder eine Audioaufnahme in natürlicher Sprache. Je weniger strukturiert Information vorliegt desto schwieriger ist deren Auswertung durch einen Computer. Schätzungen basierend auf einem Zitat von Merrill Lynch [Shilakes and Tylman, 1998] gehen davon aus, dass etwa achtzig Prozent von geschäftsrelevanter Information unstrukturiert ist, allerdings ohne dass dafür quantitativ gesicherte Quellen existieren. Eine Studie des Data Warehousing Institute (TDWI) greift diese achtzig Prozent erstmals in einer wissenschaftlichen Arbeit auf. Das Ergebnis der Studie, beziffert den Wert von unstrukturierten bzw. semi-strukturierten Daten (z.B XML Dokumente), im Verhältnis zu strukturierten Daten, auf 53%. Der Autor merkt an, dass die Studie auf Befragungen von professionellen Datenverwaltern basiert [Russom, 2007]. In jedem Fall herrscht Einigkeit über die Relevanz von unstrukturierter Information.

Die ersten Versuche in Form von Text vorliegendene unstrukturierte Information durch manuelle Methoden zu analysieren, wurden bereits Ende der 1980er / Anfang der 1990er Jahre unternommen [Schulman et al., 1989]. Durch den fortwährenden Anstieg der Computerleistung eröffneten sich schon wenige Jahre später neue Möglichkeiten immer grössere Datenmengen mit der Hilfe von Computern zu analysieren und auch das kommerzielle Interesse daran stieg im Zuge der Digitalisierung rapide an. Eines der bekanntesten Beispiele dafür ist IBM's Watson, ein Computersystem, das Fragen, gestellt in natürlicher Sprache, beantworten kann. So hat sich im letzten Jahrzehnt ein interdisziplinäres Feld rund um diese Fragestellung der Analyse von unstrukturierter textueller Information unter dem Sammelbegriff des "Textmining" etabliert. Dazu zählen unter anderem Informationsrückgewinnung, Data Mining, Maschinelles Lernen, Statistik und die Computerlinguistik. Die Vorgehensweise beim Textmining ist dabei üblicherweise, das Einlesen und Strukturieren von Information, die Erkennung von Zusammenhängen und schließlich deren Interpretation.

Eine Form in der unstrukturierte Information, heutzutage mehr denn je vorkommt, sind nutzergenerierte Webinhalte. Dazu zählen unter anderem Produktbewertungen auf entsprechenden online Portalen wie etwa Amazon, die potentiellen Käufern den Austausch über Produkte ermöglichen und somit Informationen darstellen, welche die Kaufentscheidung beeinflussen bzw. bestimmen können.
Wie wichtig Produktbewertungen für Kunden inzwischen geworden sind zeigen zwei amerikanische Studien mit über 2000 Teilnehmern [Pang and Lee, 2008]. Diesen sind folgende Angaben zu entnehmen:

1. 81% der Internetnutzer haben sich bereits einmal über ein Produkt im Internet informiert.
2. 20% davon tun dies täglich.
3. Zwischen 73% und 87% geben an, dass Produktbewertungen einen signifikanten Einfluss auf ihre Kaufentscheidung haben.
4. 32% geben an, dass sie bereits selbst eine Produktbewertung geschrieben haben.

Die Menge an unstrukturierter Information kann dabei, vor allem bei beliebten Produkten, für den Endverbraucher schnell unüberschaubar werden. So berichtet Horrigan [Horrigan, 2008], dass 58% der amerikanischen Internetnutzer bereits auf Probleme bei der Informationsbeschaffung zu Produkten gestossen sind, aufgrund von fehlenden, unmöglich zu findenden oder überwältigenden Mengen an Inhalten.

Es erstaunt daher nicht, dass auf Seiten verschiedenster Onlinekaufhäuser bereits darauf zurückgegriffen wird die Informationen durch verschiedene Methoden, wie z.B der Aggregation von numerischen Bewertungskriterien (z.B 5-Sterne Bewertung), oder der gezielten Filterung nach Relevanz (Wie hilfreich war diese Kundenbewertung?) effizienter zugänglich zu machen, um dem Kunden eine bessere Einkaufserfahrung zu ermöglichen. Ein Nebeneffekt der bei der Nutzung zusammenfassender numerischer Bewertungskriterien als Barometer für die Produktqualität durch den Kunden auftritt, ist dessen eher oberflächliche Beschäftigung mit dem Produkt hinsichtlich wertvoller Informationen, die in den beschriebenen Texten enthalten sind [Flanagin et al., 2014].
Vor diesem Hintergrund erscheint es naheliegend, die Nutzung von textbasierter Datenverarbeitung verstärkt in Betracht zu ziehen, um eine zielführende Informationsextraktion im Sinne des Kunden weiter zu verbessern.

1.2 Stand der Forschung

1.2.1 Textmining und Kundenbewertungen

Bereits aus dem Jahre 2002 stammen erste Versuche Produktbewertungen mit Hilfe von Textmining Methodik zu analysieren [Morinaga et al., 2002]. Seinerzeit war dies allerdings ein sehr genereller Ansatz bei dem es darum ging die Stimmung der Kunden im Zusammenhang mit einem Produkt zu evaluieren, gestützt auf eine Analyse von Wörtern die im Zusammenhang mit dem Produktnamen genannt werden und die Stimmung des Kunden ausdrücken. Im Prinzip also die Art von Information, welche sich heute oft in aggregierten numerischen Bewertungskriterien widerspiegelt.

In 2004 stellten Hu und Liu [Hu and Liu, 2004] einen Textmining-Ansatz vor, der in praktisch allen nachfolgenden Forschungsarbeiten in diesem Gebiet gewürdigt wird. Dieser Ansatz erlaubt Produktbewertungen zusammenzufassen und zwar nach folgenden Kriterien:

1. Produktaspekte[1], über die Kunden sprechen

2. Die Art und Weise wie sie darüber sprechen (Positiv oder Negativ)

Um dieses Ziel zu erreichen greifen Hu und Liu auf verschiedene Methoden des Textmining zurück. Für den ersten Schritt werden die Produktbewertungen in eine Datenbank geladen und vorbereitet d.h Schreibfehler korrigiert und Stoppwörter[2] gelöscht. Dann wird auf das sogenannte Part-of-Speech

[1] Unter Produktaspekten, im englischen “product features” kombinieren Hu und Liu zum einen die Produktkomponenten an sich (z.B battery, flash, motor) und zum anderen Eigenschaften des Produkts oder einer Komponente (z.B weight, life, length)

[2] Stoppwörter nennt man in der Informationsrückgewinnung Wörter, die bei einer Volltextindizierung nicht beachtet werden, da sie sehr häufig auftreten und gewöhnlich keine Relevanz für die Erfassung des Dokumentinhalts besitzen. [Koch, 2007]

Tagging (POS) zurückgegriffen, wobei jedem Wort in den Produktbewertungen eine Bezeichnung zugeordnet wird, welche Auskunft über die Wortart gibt. Die Autoren gehen davon aus, dass relevante Informationen zu Produktaspekten bevorzugt in Nomen enthaltenden Sätzen zu finden sind. Diese Idee basiert auf der Annahme, dass Produktbewertungen zwar sehr verschieden sein können, bei den adressierten Produktaspekten jedoch bestimmte Nomen verdichtet auftreten und daher solche Nomen und Nomenkombinationen die gehäuft vorkommen im Umkehrschluss mit grösserer Wahrscheinlichkeit Produktaspekte beschreiben (dabei beschränken Hu und Liu sich auf explizit benannte Eigenschaften). Um Nomenkombinationen herauszufiltern welche gehäuft vorkommen, greifen sie auf die sogenannte Assoziationsanalyse zurück. Die Assoziationsanalyse bezeichnet die Suche nach starken Regeln. Diese beschreiben Korrelationen zwischen gemeinsam auftretenden Dingen. Hu und Liu verwenden dazu eine leicht angepasste Version des Apriori-Algorithmus.[3] Um die Analysegenauigkeit weiter zu erhöhen, berücksichtigen die genannten Autoren bei einem gemeinsamen Vorkommen von Nomen auch deren Positionen im Text, um sicherzustellen, dass Nomenkombinationen wie z.B 'battery life' auch zusammenhängend genannt werden. Weiterhin löschen sie redundante Eigenschaften. So wird beispielsweise das Nomen 'life' zugunsten der Nomenkombination 'battery life' verworfen.

Für das zweite Kriterium, die bewertende Einschätzung der Kundenmeinung der Produktaspekte, suchen Hu und Liu nach sogenannten Meinungssätzen. Meinungssätze bestehen für sie aus einem Produktaspekt in Form eines Nomens sowie mindestens einem dazugehörigen Adjektiv, welches von ihnen in diesem Zusammenhang als Meinungswort definiert wird. Gefundene Meinungssätze werden mit Hilfe eines Lexikons, nach positiver oder negativer Ausdrucksweise des Verfassers bewertet. Das Lexikon enthält eine Reihe von Adjektiven für die Bewertung der ausgedrückten Stimmungslage. Der Umfang des Lexikons wird durch den Einbezug von Synonymen und Antonymen der entsprechenden Wörter vervielfacht. Falls mehrere gegensätzliche Adjektive in einem Meinungssatz vorkommen entscheidet welche Stimmungsorientierung häufiger vorkommt über die Auslegung des Meinungssatzes. Bei einer gleichen Anzahl von positiven und negativen Adjektiven, wird die Ausrichtung des vorherigen Meinungssatzes, welcher sich auf den selben Produktaspekt bezieht, berücksichtigt. Um mit Negationen wie 'no' und 'not' umzugehen, untersuchen die genannten Autoren, ob entsprechende negierende Wörter nahe (im Abstand von 5 Wörtern) eines Meinungswortes vorkommen und kehren falls nötig, die entsprechende Auslegung in das Gegenteil um.
Eine Besonderheit besteht darin, dass Hu und Liu über die Meinungsworte rückwirkend auch wenig genannte Eigenschaften identifizieren können, indem sie wenig genannte Nomen, die allerdings häufig nahe Meinungswörtern vorkommen, als potentielle Produktaspekte mitbetrachten.

In der anschliessenden Evaluation stellen diese Autoren fest, dass ihre Vorgehensweise Produktaspekte mit einer Trefferquote von 67% und einer Wirksamkeit von 79% extrahiert. In Sätzen die korrekt als Meinungssätze über Produktaspekte erkannt werden, liefert ihre Vorgehensweise eine 84.2% Genauigkeit bei der Einschätzung der Stimmung. Die Trefferquote für richtig erkannte Meinungssätze liegt allerdings bei lediglich 69.3% mit einer Wirksamkeit von 64.2%. Dazu merken sie an, dass die Einteilung, ob es sich um einen Meinungssatz oder nicht handelt, auch einer gewissen Subjektivität unterliegt. So beschreiben Kunden Produktaspekte in Geschichten, in denen sie Adjektive verwenden, die oft aufzeigen wie das Produkt genutzt wurde. Diese werden vom Algorithmus als Meinungssätze erkannt, von den Forschern allerdings nicht als Meinungssätze eingestuft, obwohl sie trotzdem nützliche Information

[3] Der Apriori-Algorithmus stellt eine Weiterentwicklung des ursprünglich im Jahre 1993 entwickelten Algorithmus zur Assoziationsanalyse dem AIS-Algorithmus dar. Der Apriori-Algorithmus wird heute immer mehr vom effizienteren FPGrowth-Algorithmus abgelöst. [Agrawal et al., 1993] [Han et al., 2004]

enthalten können. Für die Evaluation benutzen Hu und Liu zu fünf Produkten je 100 Produktbewertungen.

Nachfolgende Arbeiten in diesem Bereich bauen auf den Grundlagen von Hu und Liu auf. So erweiterten Carillo et al, die Funktionen der Eigenschaftserkennung durch maschinelles Lernen, wodurch sie versuchen, zusätzlich die subjektive Wichtigkeit von Eigenschaften eines Produktes für die Kunden, zu bewerten und einzustufen [de Albornoz et al., 2011].
Popescu und Etzioni [Popescu and Etzioni, 2007] erreichen mit ihrem OPINE System bei der Eigenschaftserkennung eine Steigerung der Wirksamkeit verglichen mit den Resultaten von Hu und Liu um 22% auf Kosten einer Verringerung der Trefferquote von 3%. Diese Verbesserung erreichen sie durch die Berechnung eines Wertes der Auskunft darüber gibt, wie Nomen sich aufeinander beziehen, um daraus eine Hierarchie von Produkt und Eigenschaften abzuleiten (z.B 'scanner has x', 'scanner comes with x', 'x of scanner'). Bei der Erkennung von Meinungssätzen und deren Bewertung liegen die Unterschiede im einstelligen Prozentbereich.
Zhang und Narayanan [Zhang and Narayanan, 2010] stellten ihrerseits ein Verfahren vor Produkte der selben Produktkategorie, anhand ihrer Eigenschaften und wie Kunden sich darüber äussern zu vergleichen. Dafür wählten sie Produktaspekte manuell aus und überprüften zunächst deren produktübergreifendes Vorkommen. Für die Sentimentanalyse achteten sie nebst der subjektiven Einschätzung ausserdem explizit auf komparative Aussagen zu den Eigenschaften (Eigenschaft X in Produkt A ist besser als in Produkt B). Mithilfe der so gewonnen Informationen erstellten sie einen gewichteten und gerichteten Graphen, um die relative Qualität von Produktaspekten von Produkten der selben Kategorie zu bestimmen.

Zusammenfassend ist der bisherigen Forschung im hier adressierten Bereich ein grundlegender Konsens zu entnehmen Kundenbewertungen mit Hilfe von Textmining, hinsichtlich der Sinnhaftigkeit und Zielgerichtetheit aufzubereiten. Bezeichnend ist dabei auch die einheitliche Vorgehensweise die zum einen aus der Erkennung von Produktaspekten zum anderen aus deren Bewertung besteht.
Obwohl in Bezug auf die Genauigkeit der vorgestellten Systeme bereits grosse Fortschritte erzielt wurden, sind sie weit davon entfernt fehlerfrei zu funktionieren oder gar als alleiniges Bewertungsinstrument hinzureichen. Es bleibt abzuwarten wie sich dieses Feld der Computer-Sciences entwickelt. Fakt ist, dass im Kontext sich weitergehend steigernder Computerleistung bereits bestehende Textmining-Systeme den Anfang einer neuen Ära der Informationsgewinnung einläuten. Ich bin optimistisch hinsichtlich weiterer zielführender Entwicklungen in diesem Bereich in naher Zukunft. Mit der hier beschriebenen Arbeit kann gegebenenfalls zum kleinen Teil in dieser Entwicklungsrichtung beigetragen werden.

1.2.2 Sentimentanalyse

Eine der grössten Hürden in Bezug auf Textmining, spiegelt sich in der Sentimentanalyse, also der Stimmungserkennung respektive der Interpretation von menschlich kommunizierter Information wieder. Die bisherigen programmtechnischen Ansätze sind von unserer menschlichen Interpretationsgabe noch weit entfernt. Das Forschungsfeld der Stimmungserkennung, welches mit zunehmendem Interesse erforscht wird, hat sich längst als eigene Disziplin etabliert und ist dabei eng mit der Erforschung von Lernprozessen verknüpft [Pang et al., 2002]. Im Anbetracht der Tatsache, dass es selbst Menschen schwerfallen kann, die subjektive Stimmung von anderen Menschen einzuschätzen, wird ein weitergehend zufriedenstellendes Ergebnis wohl erst mit der Erforschung der Künstlichen Intelligenz Einzug halten.

Eine umfassende Aufstellung über den Stand der Forschung im Bereich der Stimmungserkennung liefern beispielsweise Collomb et. al [Collomb et al., 2014] in einer Studie, welche die gängigsten Methoden der Sentimentanalyse wie folgt kategorisiert:

1. Der auf maschinellem Lernen basierende Ansatz benutzt Algorithmen die auf Basis bestimmter Datensets trainiert werden, um so zu lernen, Stimmungen einzuschätzen.

2. Der lexikonbasierte Ansatz berechnet die Stimmung anhand der semantischen Orientierung von Wörtern oder Sätzen, die bereits bekannt sein müssen. Die semantische Orientierung misst dabei die subjektive Orientierung einer Meinung.

3. Der regelbasierte Ansatz bezieht sich auf Vorgehensweisen welche zur Sentimentanalyse Regeln wie etwa zur Negation, gemischten Meinungen, Idiomen, verstärkenden Wörtern, Smilies usw. nutzen.

4. Der statistische Ansatz nutzt Modelle die numerische Bewertungskriterien (etwa eine 5-Sterne Bewertung) zur Bewertung von Sentiment miteinbeziehen.

Die vorab beschriebene Methodik zur Sentimentanalyse nach Hu und Liu, basiert wie vermittelt auf einem vordefinierten Lexikon sowie einem Regelwerk und ist hinsichtlich erzielbarer Genauigkeit (84.2% s.o) im Vergleich zu anderen Methoden welche die Sentimentanalyse auf Satzebene mittels maschinellem Lernen (74% Genauigkeit) [Khan et al., 2011] oder basierend auf statistischen Ansätzen unter Einbezug numerischer Bewertungskriterien (73% Genauigkeit) adressieren, bis jetzt noch überlegen [Collomb et al., 2014].

Die Komplexen Probleme [Pang and Lee, 2008] die Forscher zu lösen versuchen um die Genauigkeit der Sentimentanalyse weiter zu steigern, beziehen sich beispielsweise darauf die richtigen Begriffe in einem Lexikon zu definieren oder Emotionen besser einzuschätzen. Als Beispiel sei nachfolgend ein Zitat von Mark Twain angeführt, in dem deutlich eine bewertende Meinung geäussert wird, ohne dass Wörter die positiven oder negativen Emotionen zugeordnet werden können, benutzt werden:

> “Everytime I read ‘Pride and Prejudice’ I want to dig her up and beat her over the skull with her own shin-bone.”

Ohne dass hier offensichtlich negative behaftete Wörter (z.B bad, hurt, problem etc.) benutzt werden, drückt der Author dennoch eine starke negative emotionale Position aus.

Eine weitere Problematik ist die Erkennung bei Zusammenhängen von Reihenfolgen. So kann z.B ein Satz die Inhalte vorheriger Sätze relativieren:

> “This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance.

However, it can't hold up."

Dies sind wohlgemerkt nur einige Beispiele der KI-Entwicklung, der ich mich in dieser Arbeit nicht hauptsächlich zuwenden möchte. Am Beispiel der Sentimentanalyse soll jedoch gezeigt werden, inwiefern sich die Grenzen des heute technisch Möglichen auswirken und die Limitierungen bedingen, die im vorangestellten Abschnitt aufgezeigt wurden.

1.2.3 Der Kunde und Information

Die Kernfrage die sich aus meiner Sicht zum Thema dieser Arbeit stellt bezieht sich darauf, wie die automatisierte Informationsbeschaffung in den Kaufentscheidungsprozess eines Kunden passt bzw. mit diesem kompatibilisert werden kann. Dazu ist es zunächst wichtig zu verstehen, wie ein Kunde einkauft. Nach Kotler und Keller [Kotler and Keller, 2012] lässt sich der Kaufprozess eines Kunden in folgenden Phasen darstellen:

1. Need of recognition
2. Information search
3. Evaluation of alternatives
4. Purchase decision
5. Purchase
6. Post-purchase evaluation

Aus dieser Abfolge an Vorgängen war es im Kontext zur hier vorgenommenen Bearbeitung naheliegend insbesondere die Phase der Informationsakquise (Information search) durch den Kunden genauer zu analysieren. Der Kunde möchte sich in dieser Phase beispielsweise mittels Durchsicht von Kundenbewertungen ein genaueres Bild von einem Produkt bzw. dessen Eigenschaften machen. Wie er dabei genau vorgeht, kann produktspezifisch stark variieren.
So vertraut er in aller Regel beispielsweise bei Gütern deren Qualität erst nach dem Kauf bestimmt werden kann (z.B. Wein, Musik etc.) bevorzugt auf ausgeglichene Kundenbewertungen, da er extreme Bewertungen hierzu als unglaubwürdiger einstuft. Erstaunlicherweise gilt diese Regel nicht für Bücher. Dort werden extreme Bewertungen sogar bevorzugt aufgenommen. Bei Produkten die bereits vor dem Kauf qualitativ bewertet werden können (z.B. Kamera, Fernseher) werden längere Kundenbewertungen oft als hilfreicher eingeschätzt auf Grund des Umfangs der gegebenen Information. [Mudambi and Schuff, 2010] [Willemsen et al., 2011]. Anhand dieser Charakteristiken lässt sich erkennen, dass die subjektive Aussagekraft von unstrukturierter Information auf vielen nicht einfach ersichtlichen Zusammenhängen basiert. Es ist daher nicht verwunderlich, dass auch zur Güte von unstrukturierter Information, Meta-Information von Seiten einiger Onlinehändler angeboten wird, z.B in Form eines Zählers der misst, wie viele Personen eine Bewertung hilfreich fanden.

Man kann also festhalten, dass die Aussagekraft von unstrukturierter Information über ein Produkt bzw. dessen Eigenschaften nach verschiedenen Kriterien bewertet werden kann und für den Kunden Relevanz besitzt.
Weiterhin scheint es logisch, vorauszusetzen, dass der Kunde Interesse hat „effizient“ zu sein in der Entscheidungsfindung. Der Zeitaufwand bzw. der kognitive Aufwand diese Information zu sichten spielen dabei eine Rolle. Er möchte klare Informationen möglichst schnell finden.

Daraus lassen sich zwei zentrale Anforderungen eines Kunden an Information im Bezug auf ein Produkt ableiten:

1. Die Kompaktheit der Information

2. Die Aussagekraft der Information

Diese Rahmenbedingungen bieten grundsätzlich eine ideale Umgebung für das Textmining, denn es liegt in dessen intrinsischem Anspruch bislang zumindest die erste der vorgenannten Anforderungen aus Kunden-/Benutzersicht zu erfüllen. Der zweiten Anforderung soll in dieser Arbeit verstärkt Beachtung gelten.

1.3 Die Forschungsfrage

Die Forschungsfrage leitet sich aus einer Kombination der Zielsetzung des Kunden bei der Informationssuche und dem bisherigen Stand der Forschung zum Textmining bei Kundenbewertungen ab.

Gemäss den im vorherigen Kapitel gemachten Ausführungen sind Kompaktheit (für den Kunden gleichbedeutend mit Zeitaufwand) und Aussagekraft die massgeblichen Kriterien für die Güte der Kundeninformation. Eine Analyse der bisherigen Forschung zum Produkt-Textmining zeigt, dass diese Anforderungen bisher vor allem in Bezug auf die Aussagekraft der Information noch nicht ausreichend befriedigt werden.
Betrachtet man beispielsweise die Sentimentanalyse als eine der Textmining-Methoden für die Analyse von kundenrelevanten Produktaspekten, so hat die Durchsicht bislang erzielter Forschungsergebnisse gezeigt, dass diese Art Produktinformation zu verdichten noch mit erheblichen Fehlern behaftet ist. Das heisst, diese Analysemethode bietet für sich alleine betrachtet, wenn dann höchstens eine Teil-Grundlage, um daraus sinnvolle Aussagen für den Kunde abzuleiten. Abgesehen davon, hat ein durchschnittlicher Kunde heute noch keine Möglichkeit die beschriebene Form des Textmining zu betreiben.

Aus diesem Kontext wird die folgende Forschungsfrage abgeleitet:

„Wie kann unter Einsatz von Textmining-Methodik ein Werkzeug geschaffen werden, das aus unstrukturierten Kundenbewertungen zu einem bestimmten Produkt, dem interessierten Dritten bei minimiertem Zeitaufwand eine aussagekräftige Übersicht zu dessen Komponenten und Eigenschaften sowie deren Bewertung liefert ?“

Es steht ausser Frage, dass die heutige Forschung sich bereits mit dieser Thematik auseinandersetzt. Das Augenmerk liegt dabei bislang insbesondere darauf, genannte Unzulänglichkeiten durch eine Verbesserung der Genauigkeit für einzelne Methoden zu beheben (z.B für die Sentimentanalyse).

In der hier vorgelegten Arbeit wird ein anderer Weg eingeschlagen. Es soll ansatzweise untersucht werden, inwieweit eine Kombination verfügbarer Textmining-Methoden, eingebettet in eine Webumgebung, trotz noch vorhandener Unzulänglichkeiten bereits verbessert zur Unterstützung eines Kunden eingesetzt werden kann. Diese anwendungsorientierte Vorgehensweise kann helfen Potentiale aber auch Schwächen der einzelnen Textmining-Methoden, sowie mögliche Synergismen bei deren Kombination aufzudecken und somit zur weiteren Methodenverbesserung beizutragen.

Zu diesem Vorgehen wird folgende Arbeitshypothese formuliert:

Eine praxisorientierte Textmining-Methoden-Kopplung von (1) Aspektextraktion, (2) Assoziationsanalyse, (3) Sentimentanalysen und einer erweiterten Auswertung der Resultate der Sentimentanalysen, sowie der Einbezug einer interaktiven Zugriffsmöglichkeit durch den Kunden, lassen eine qualitative Steigerung der Aussagekraft gegenüber der Anwendung der jeweils einzelnen oder auch gekoppelten Methoden ohne interaktive Zugriffsmöglichkeit, erwarten.

Zur Bearbeitung der Arbeitshypothese wurde in dieser Arbeit, wie nachfolgend ausgeführt, ein Prototyp entwickelt, der die beschriebenen state-of-the-art Methoden des Textmining kombiniert, um den Anforderungen eines Kunden an den Einkaufsprozess und im speziellen an die Informationssuche verbessert gerecht zu werden.
Basierend auf einer Evaluation dieses Prototyps wurde anschliessend der Versuch unternommen die Arbeitshypothese (teilweise oder in Gänze) zu bestätigen bzw. zu widerlegen. Aus diesem Ergebnis wurde dann eine Antwort auf die Forschungsfrage abgeleitet, die darstellt, inwieweit heute verfügbare Ansätze zum Textmining bereits genutzt werden können, um den Anforderungen eines Kunden bei der Informationsakquise im Onlineshopping möglichst nahe zu kommen.

2 Prototyp

Der Aufbau des Prototyps orientiert sich, der Forschungsfrage entsprechend, an den Anforderungen des Kunden die in Kapitel 1 dargelegt wurden. Das Augenmerk liegt hierbei auf dem zweiten Schritt des Kaufentscheidungsprozesses (2. Information search), da hierbei massgebliche, unstrukturierte Information in Form von nutzergeneriertem Text vom Kunden aufgenommen werden muss. Um den Programmieraufwand in Grenzen zu halten, wurde der hier entwickelte Prototyp beispielhaft auf den Onlineshop von Amazon massgeschneidert. Prinzipiell kann eine Anpassung auf jede andere relevante Website erfolgen. Da die Bibliotheken die zum Textmining verwendet wurden auf die englische Sprache ausgerichtet sind, kann der Prototyp bisher ausschliesslich in Englisch verfasste Kundenbewertungen einlesen und auswerten.

Der Prototyp basiert auf der Programmiersprache R und beinhaltet drei Kategorien von Funktionen: (i) Webscraping, (ii) Textmining, (iii) Visualisierung. (Abbildung 1)

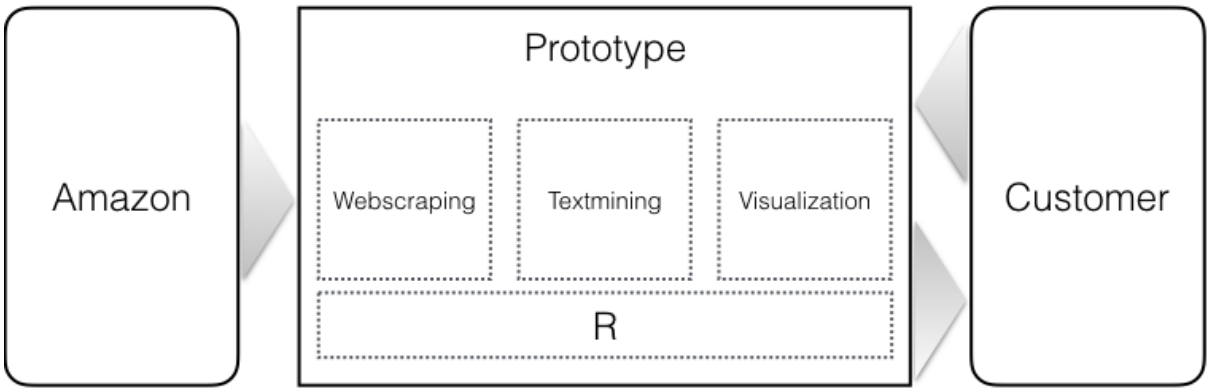

Abbildung 1: Schematischer Aufbau des Prototyps mit den Funktionskategorien

2.1 Werkzeuge

Zunächst werden die technischen Grundlagen des Prototyps beschrieben. Diese beinhalten neben der gewählten Programmiersprache auch die zur Hilfe herangezogenen Bibliotheken. Ausführungen zur Funktionsweise folgen in Kapitel 2.2.

2.1.1 Die Programmiersprache: R vs Python

Wenn es um Textmining geht sind die Auswahlmöglichkeiten an Programmiersprachen, die dafür verwendet werden können und über eine solide Wissensdatenbank rund um den Themenbereich verfügen, nicht sehr umfangreich. Da es sich hier um ein relativ junges Forschungsfeld handelt, in welchem vielfach Forscher aus dem mathematisch-statistischen Bereich involviert sind verwundert es nicht, dass die in deren Expertise- und Anwendungsfeldern beliebte und potente Programmiersprache R auch für das Textmining bereits vielfach erfolgreich appliziert wurde. Als Alternative hat sich inzwischen aber auch die Allzwecksprache Python in der Datenverarbeitung bewährt und gewinnt weiter an Zulauf.

Grundsätzlich eignen sich beide genannten Programme, es gibt allerdings einige im Kontext relevante Unterschiede, auf die kurz eingegangen werden soll:

Python ist eine vollumfängliche Programmiersprache, wohingegen R fast ausschliesslich zur Datenverarbeitung genutzt wird. Die Vorteile bei Python liegen darin, dass der Funktionsumfang als Allzwecksprache deutlich grösser und syntaktisch einfacher zu erfassen ist als bei R. Der Anwendungsbereich von R birgt wiederum den Vorteil, dass viele Optionen zur Datenvisualisierung standardmässig verfügbar sind. R ist zwar beschränkter was den Umfang und möglicherweise auch eine weiterführende Implementierung im Sinne eines Produktes angeht, bildet aber für die explorative Arbeit in einem speziellen Gebiet der Datenverarbeitung, wie dies das Textmining repräsentiert, eine solide Grundlage. Vor allem, da gerade in Forschungskreisen viel mit R gearbeitet wird, existiert eine Vielzahl von bereits verwendbaren Bibliotheken zu Algorithmen, die den neusten Stand der Technik, sofern veröffentlicht, darstellen. Im Rahmen dieser Arbeit wurde aus vorgenannten Gründen entschieden für das Prototyping, R als Programmiersprache zu nutzen.

2.1.2 Bibliotheken

Hinsichtlich Verfügbarkeit von Bibliotheken kommen die Stärken von R klar zum Tragen. Solche Bibliotheken waren für diese Arbeit überaus hilfreich bei der Implementierung der Kernfunktionen. Im Folgenden erfolgt eine Auflistung mit kurzen Erklärungen zu den wichtigsten im Prototyp genutzten R Bibliotheken, katalogisiert nach Funktion.

a) Webscraping:

rSelenium: Automatisiert die Nutzung eines lokalen Webbrowsers in R.

rvest: Ermöglicht das Parsen von HTML Seiten mit Hilfe von CSS Selektoren.

b) Textmining:

tm: Liefert Funktionen zur Vorbereitung von Texten für die weitere Verarbeitung, d.h das Erstellen eines Korpus (Datenobjekt, welches Textdokumente enthält) und die Aufbereitung des Korpus (Löschen von Stoppwörtern, Zeichensetzung, überflüssigen Leerzeichen usw.)

openNLP: Ein auf Maschinenlernen basierter Werkzeugkasten zur Sprachverarbeitung, der es unter anderem ermöglicht mit Hilfe von Part-Of-Speech Tagging, Wortarten zu identifizieren.

syuzhet: Eine Sammlung von lexikon- und regelbasierten Methoden zur Sentimentsanalyse von Texten.

1. Syuzhet [Jockers, 2016]
2. Afinn [Nielsen, 2011]
3. Bing [Hu and Liu, 2004][Liu, 2012]
4. NRC [Mohammad et al., 2013]

c) Visualisierung:

shiny: Eine Schnittstelle zwischen R und einer Netzapplikation. Shiny bietet eine praktische Möglichkeit R Inhalte interaktiv auf einer Website zu visualisieren, ohne dass explizit in HTML, CSS oder JavaScript programmiert werden muss.

googleVis: Bietet eine Vielzahl von Visualisierungsmöglichkeiten von Daten, wie z.B ein Bubblechart.

wordcloud: Ermöglicht eine Wortwolke als Visualisierungsform.

2.2 Funktionsweise

Da sich die Funktionsweise am Kunden orientiert, liegt es nahe zunächst jeweils das Design der Funktionen anhand des Einkaufsprozesses zu erläutern, um damit in der später folgenden Diskussion die

Verbindung zu den Anforderungen herzustellen. In einem zweiten Schritt wird jeweils die Implementierung der Funktionen genauer unter die Lupe genommen. Abschliessend wird anhand eines tabellarischen Überblicks sowie eines Sequenzdiagramms aufgezeigt, wie die Funktionen ineinandergreifen bzw. sich in der Interaktion mit einem Kunden verhalten.

2.2.1 Webscraping

Design: Hat der Kunde ein Produkt auf Amazon gefunden, welches ihn interessiert, kann er im Prototyp der im Browser läuft, die URL, welche zu den Kundenbwertungen des besagten Produktes führt, eingeben und bestimmen wieviele Bewertungen er gerne zur weiteren Analyse erfassen möchte. (Abbildung 2)

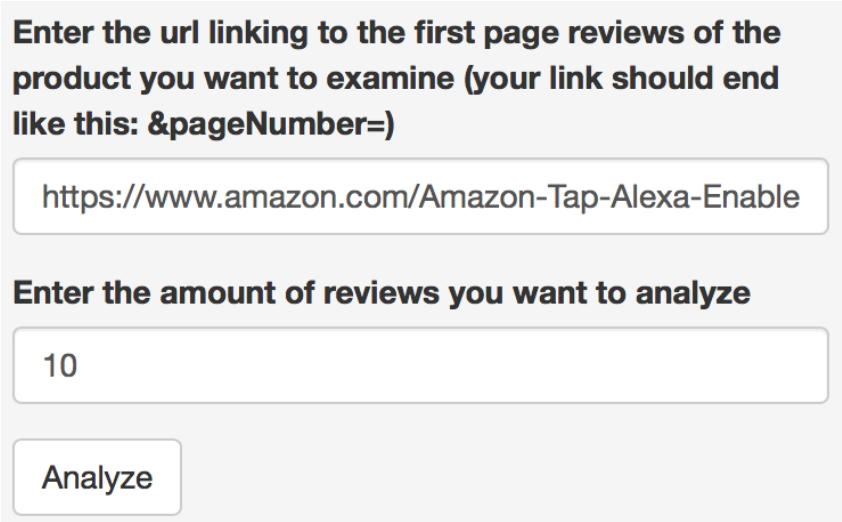

Abbildung 2: Eingabefeld für die URL von Amazon-Produktbewertungen, die analysiert werden sollen

Implementierung: Sobald er die Eingabe bestätigt, greift die Webscraping Funktionalität des Prototyps, die auf einer Kombination der *rSelenium* und der *rvest* Bibliothek basiert. Die entsprechende Website mit den Kundenbewertungen wird zunächst mit Hilfe der *rSelenium* Bibliothek eingelesen. Danach können mit Hilfe der *rvest* Bibliothek Kundenbewertungen über CSS Selektoren extrahiert und für die weitere Verarbeitung als Textvektor zwischengespeichert werden (mehr zu der Begründung dieser Vorgehensweise im Abschnitt zu den Implementierungsschwierigkeiten). Je nach Bedarf verfügt der Prototyp ausserdem über eine Funktion, einen Textvektor zu normalisieren, die auf der *tm* Bibliothek basiert. Dies erlaubt Stoppwörter, Zeichensetzung und überflüssige Leerzeichen aus einem Textvektor zu löschen.

2.2.2 Textmining

Wortwolke

Design: Hat der Kunde sich noch nicht näher über das Produkt informiert, wird ihm jetzt eine Wortwolke präsentiert, welche ihm die am häufigsten genannten Nomen aus den Produktbewertungen anzeigt. Dies hat den Sinn, dass basierend auf der Vermutung von Hu und Liu [Hu and Liu, 2004], Kunden mit ihren Äusserungen zu Produktaspekten konvergieren und Nomen, die häufiger genannt werden, daher relevantere Produktaspekte ((1) Komponenten und/oder (2) Eigenschaften) des zu bewertenden Objektes bezeichnen, welche in der Bewertung für den Kunden von Interesse sind. Da in aller Regel auch für Nomen die Produktkomponenten bezeichnen verbundene Eigenschaften von primärem Kundeninteresse sind, sollte eine weiterführende Textmining Analyse dem Anwender (Kunden) die Möglichkeit eröffnen, selektierte Nomen dieser Kategorie derart zu erweitern, dass bezogene Eigenschaften kundenseitig ergänzt werden können (z.B. battery => battery life).
Anhand einer solchen Wortwolke kann sich der Kunde damit im ersten Schritt einen Überblick verschaffen, welche Produktaspekte für andere Kunden von Interesse sind. Die Parameter, die er dabei beeinflussen kann, sind die Anzahl wie oft eine potentielle Eigenschaft (summiert über alle Bewertungen) mindestens genannt werden muss, um in der Wolke aufzutauchen, und wieviele Aspekte maximal in der Wolke, absteigend nach Häufigkeit geordnet, dargestellt werden.

In Abbildung 3 ist beispielhaft eine solche Wortwolke für die ersten zehn Kundenbewertungen des mobilen Bluetooth-Musiklautsprechers *Amazon Tap* zu sehen. Es ist zu erkennen, dass neben dem Produktnamen der ein Nomen enthält auch Nomen wie “music” oder “use” (was möglicherweise auch daran liegt, dass “use” gleichzeitig ein Verb ist und falsch zugeordnet wird) auftauchen, die offensichtlich keine Produktaspekte bezeichnen, jedoch häufig im Zusammenhang mit dem Produkt genannt werden. Nichtsdestotrotz können auch erste nützliche Informationen daraus gezogen werden, denn wie vermutet tauchen in der Wortwolke deutlich überwiegend Produkt bezogene Aspekte auf. In diesem Beispiel sieht der Kunde etwa, dass andere Kunden sich rege über die Batterie oder den Knopf des Gerätes unterhalten. Anhaltspunkte um weiter nachzuforschen.

Implementierung: Programmtechnisch wurde die Implementierung gelöst, indem auf die Funktionalitäten der *openNLP* Bibliothek zurückgegriffen wurde, um Wortarten in einem Text zu identifizieren. Durch einen regulären Ausdruck, welcher die Wortart spezifiziert, werden Wörter (in diesem Falle die Nomen) herausgefiltert und schliesslich, unter Zuhilfenahme der *tm* Bibliothek quantifiziert. Dabei wird eine Matrix erstellt, die Auskunft über die Anzahl der vorkommenden Wörter in einem Vektor gibt (siehe Abbildung 4). Die Ergebnisse werden dann durch die *wordcloud* Bibliothek als Wortwolke abgebildet. Die Grösse, in der ein Wort dargestellt wird, hängt dabei von der Häufigkeit seines Vorkommens ab.

Abbildung 3: Generierte Nomen-Wortwolke aus den ersten zehn Reviews des Amazon Tap bei einem minimalen Wortvorkommen von fünfzehn.

```
#Hilfsfunktion zur extraktion von bestimmten Wortarten aus einem Text
#@function: getPOS(chr vector, string) @return: vector
getPOS <- function(x, marker) {

    #Eingabe Vektor als String einlesen
    x <- as.String(x)

    #Start und Endpunkte jedes Wortes annotieren
    wordAnnotation <- annotate(x, list(Maxent_Sent_Token_Annotator(), Maxent_Word_Token_Annotator()))

    #Die Wortart jedes erkannten Wortes annotieren
    POSAnnotation <- annotate(x, Maxent_POS_Tag_Annotator(), wordAnnotation)

    #Wörter finden bei denen die Wortart dem entsprechenden eingabe String (z.B NN für Nomen) entspricht
    tags <- sapply(POSAnnotation$features, '[[', "POS")
    thisPOSindex <- grep(marker, tags)
    wordVector <- sprintf("%s", x[POSAnnotation][thisPOSindex])
    wordVector

}
```

Abbildung 4: Nomenextraktion aus unstrukturiertem Text im Detail

Assoziationsanalyse

Design: Weiss der Kunde welchen Produktaspekt er sich genauer anschauen möchte, kann er fortfahren und diesen in ein Textfeld eingeben. Als direktes Feedback erhält er dabei sofort eine Auswahl an Wörtern die im Zusammenhang mit diesem Aspekt genannt werden. Dazu werden durch das Finden von Assoziationsregeln die Korrelation anderer Wörter mit dem Produktaspekt aus den eingelesenen Kundenbewertungen berechnet und die fünf am stärksten korrelierenden Wörter angezeigt.

Kenngrössen der Assoziationsanalyse sind:

- Support: Die relative Häufigkeit der Beispiele, in denen die Regel anwendbar ist.
- Confidence: Die relative Häufigkeit der Beispiele, in denen die Regel richtig ist. (Abbildung 5)
- Lift: Die Höhe mit der der Konfidenzwert für die Regel den Erwartungswert übertrifft, also die generelle Bedeutung einer Regel. (Abbildung 6)

$$\text{confidence}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}$$

Abbildung 5: Zusammenhang zwischen confidence und support bei der Assoziationsanalyse.

$$\text{lift}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}$$

Abbildung 6: Zusammenhang zwischen lift und support bei der Assoziationsanalyse.

Diese Idee basiert auf der Vermutung, dass Wörter die nicht trivial sind (z.B Stoppwörter) und oft im Zusammenhang mit einem Produktaspekt genannt werden, mit einer gewissen Wahrscheinlichkeit damit in Verbindung stehen. Diese Überlegung orientiert sich an der Methodik die Hu und Liu [Hu and Liu, 2004] verwenden, um zusammengesetzte Produktaspekte mit Hilfe des Apriori-Algorithmus zu identifizieren. Es soll dem Kunden helfen, erste Rückschlüsse über einen Produktaspekt zu ziehen. Untersucht man auf diese Weise beispielsweise den Aspekt “hose”, also Schlauch, bei einem Staubsauger auf Amazon, erhält man das Ergebnis, welches in Abbildung 7 zu sehen ist.

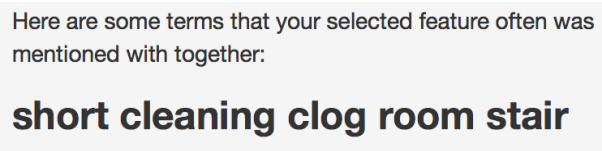

Abbildung 7: Korrelierende Wörter in Kundenbewertungen über einen Staubsauger, bezogen auf den Aspekt “hose” aus 100 Kundenbewertungen.

Oft im Zusammenhang mit “hose” werden demnach bei diesem Staubsauger “short”,“cleaning”,“clog”,“room” und “stair” genannt. Man kann erkennen, dass Wörter wie “room” und “cleaning” begrenzte Aussagekraft über den Produktaspekt haben, aber offensichtlich häufig zusammen in einer Kundenbwertung genannt werden. Daneben könnte man aus “short” und “clog” ableiten, dass Verstopfungen des Schlauches eventuell ein Problem darstellen und gezielt danach suchen. Oder aber die Länge des Schlauches als Problem identifizieren, die möglicherweise auch im Zusammenhang mit Treppen steht.
Die Aussagekraft solcher Information ist schwer zu quantifizieren, natürlich tauchen nicht immer Wörter auf, aus denen der Kunde etwas schliessen kann. Aus Versuchen wurde bestätigt, dass bei Datenmengen aus ca. 100 Bewertungen hinreichend nützliche Informationen auftauchen. Ein typisches weiteres Anwendungsfeld dieser Technik ist, allerdings auf Seiten der Händler, die sogenannte Warenkorbanalyse um gezielt Werbemaßnahmen einzuleiten. Dabei wird analysiert, welche Produkte Kunden häufig zusammen kaufen.

Implementierung: Implementiert wurde diese Funktion durch die Nutzung einer Matrix, wie sie bei der Wortwolke erstellt wurde, allerdings nicht nur auf Nomen bezogen. Die Matrix gibt Auskunft über die Anzahlhäufigkeit von Wörtern in Kundenbewertungen und liefert damit die Grundlage für eine Assoziationsanalyse die in der *tm* Bibliothek implementiert ist.

Sentimentanalyse

Design: Weiterhin wurde eine Sentimentanalyse aller Sätze durchgeführt, welche den gewählten Produktaspekt beinhalten, um dem Kunden Anhaltspunkte über die generelle Stimmung über diesen Aspekt zu liefern. Dazu wurde in dieser Arbeit auf vier Ansätze[4] der Sentimentanalyse (Syuzhet, Afinn, Bing, NRC s.o.) zurückgegriffen, die jeweils den selben Satz analysieren. Der Grund, hier vier Ansätze, die auf zwei Methoden der Sentimentanalyse basieren, anzuwenden, liegt darin, die Genauigkeit der Aussage im Vergleich zur Anwendung nur eines Ansatzes, zu erhöhen.
Da alle vier Ansätze Sentiment in einem Text mit Hilfe eines Vorzeichens ausdrücken, wurde hier die Anzahl der Vorzeichen als Indikator für den Grad der Übereinstimmung gewählt. Daraus ergibt sich ein Balkendiagramm, das Auskunft darüber gibt, wieviele Sätze mit Produktaspekten mit welchem Grad an Übereinstimmung der vier Ansätze als positiv, negativ oder neutral eingestuft werden. Zusätzlich wird die durchschnittliche Bewertung aller Sätze angezeigt und wie häufig ein Produktaspekt genannt wurde (Abbildung 8).

Implementierung: Implementiert wurde mit Hilfe der *syuzhet* Bibliothek, die eine Sammlung von aktuellen Sentimentanalysemethoden anbietet. In einem Vorbereitungsschritt werden zunächst alle Sätze, in denen eine gesuchter Produktaspekt vorkommt, erkannt, um dann eingebettet in einen Vektor auf ihr Sentiment hin überprüft zu werden. Abbildung 9 zeigt die Extraktion der Meinungssätze im Detail. Die Kundenbewertungen werden zunächst durchnummeriert und dann mit Hilfe einer Hilfsfunktion, die auf regulären Ausdrücken basiert (*get_Sentences()*), in ihre einzelnen Sätze zerlegt. Ist das geschehen wird nochmals ein regulärer Ausdruck benutzt, um nur Sätze zu extrahieren, in der sich der gewünschte weiter zu bewertende Produktaspekt befindet. Diese werden dann zur Sentimentsanalyse übergeben.

[4] Wobei alle Ansätze in die Kategorie der lexikon- bzw. regelbasierten Sentimentanalysemethoden fallen, darunter auch die Weiterentwicklung der Algorithmen von Hu und Liu, die am Anfang der Arbeit in Abschnitt 2.2.1 vorgestellt wurden.

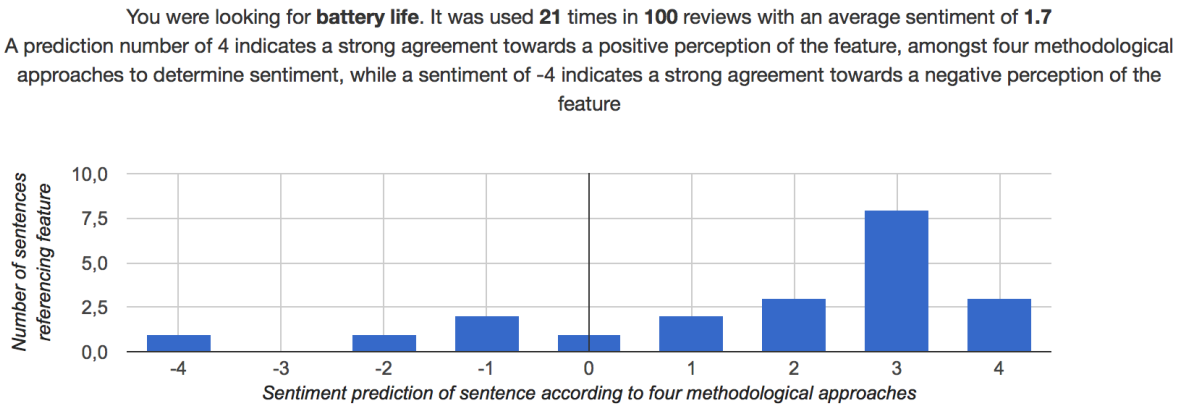

Abbildung 8: Sentimentanalyse von Sätzen, die sich auf den Aspekt *battery life* beim Amazon Tap beziehhen

```
#Die Kundenbewertungen im eingabe Vektor durchnummerieren
reviewsRaw_index <- seq(1,length(vector), 1)
reviewsRaw_filtered <- sapply(reviewsRaw_index, getReviews, vector)

#Die nummerierten Kundenbewertung in ihre Sätze zerlegen
sentences <- getSentences(reviewsRaw_filtered)

#Die Sätze mit hilfe einer RegEx herausfilteren, welche das gesuchte Wort beinhalten

word <- paste(" ",word,sep="")  #Leerzeichen vor das Wort setzen um Fälle zu Vermeiden wo das Gesuchte Wort
                                #in einem anderen Wort vorkommt z.B "hose" => "those"

sentences_filtered <- grep(word, sentences[,2], ignore.case = TRUE)
result <- sentences[c(sentences_filtered),]
result
```

Abbildung 9: Die Extraktion von Meinungssätzen im Detail

Produktaspekte-Browser

Design: Die bei der Sentimentanalyse gewonnenen Daten finden für den Kunden noch eine weitere Verwendung. Dadurch, dass die bewerteten Sätze mit dem Produktaspekt bekannt sind und auch zu den entsprechenden Kundenbewertungen zugeordnet werden können, liegt es nahe auch diese Information an den Kunden weiterzugeben. Um die Nützlichkeit des vorgestellten Balkendiagramms zur Einschätzung der Kundenmeinung weiter zu erhöhen, soll dem Kunden ermöglicht werden, nachzuvollziehen, wie die berücksichtigten Sätze, die zu dieser Bewertung führen aussehen und woher sie stammen. Eine Möglichkeit dies effizient und benutzerfreundlich umzusetzen besteht darin, einen wie nachfolgend dargestellten Graphen zu erstellen. Auf der X-Achse wird, wie schon beim Balkendiagramm, die analysierte "Stimmung" der Sätze eingetragen. Auf der Y-Achse wird eingezeichnet, aus welcher Kundenbewertung (nummeriert nach Reihenfolge des Einlesens) diese Sätze stammen.

Je nach Anzahl der Sätze, die in einer Kundenbewertung vorkommen und die gleichwertige Resultate bei der Einschätzung ihres Sentiments bekommen haben, werden diese als grössere oder kleinere Kreise im entsprechenden Achsenabschnitt eingezeichnet. Fährt der Kunde mit der Maus über eine solchen Kreis, werden die entsprechenden Sätze direkt als Tooltip angezeigt. Man kann also übersichtlich sehen, welche Kunden wie (kategorisiert nach Sentiment) über einen spezifischen Produktaspekt befinden. Dies ist verknüpft mit der Möglichkeit, die Sätze, die zu dieser Bewertung führen anzuschauen und so zu überprüfen, ob die Schlüsse, die der entsprechende Kunde zieht sinnvoll sind und erscheint zielführend angesichts der nicht immer hinreichenden Genauigkeit der Sentimentanalysemethoden. Abbildung 10 zeigt einen wie vorab beschriebenen Graphen über die ersten sechsundvierzig Kundenbewertungen zum Amazon Tap, beispielhaft bezogen auf den Produktaspekt "bass".

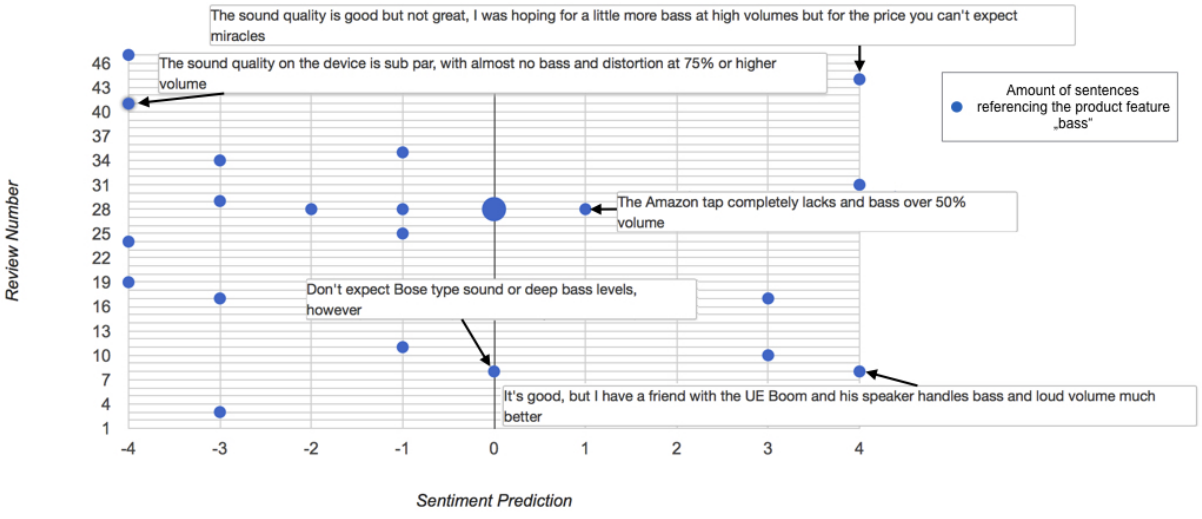

Abbildung 10: Produktaspekt-Browser: Übersicht über sechsundvierzig Kundenbewertungen zum Amazon Tap, bezogen auf den Produktaspekt "bass", inklusive der Darstellung einiger Sätze, die den Datenpunkten entsprechen

Eine Auswertung dieses Graphen bestätigt die Vermutung, dass es sinnvoll ist, dem Kunden den Kontext der Sentimentanalyse aufzuzeigen, um deren Aussagekraft zu verbessern. So gibt es zum Beispiel einen Satz in Kundenbewertung 28, der von mindestens einem Algorithmus als positiv eingestuft wurde, aber einen negativen Punkt der Klangqualität anspricht: “The Amazon tap completely lacks [...] bass over 50% volume”.
Auch kann man erkennen, dass zwei Sätze, die beide eine Einstufung von +4 besitzen (also eine Übereinstimmung aller Algorithmen über die Ausrichtung der Stimmung) zwar nicht negativ über den Produktaspekt berichten, dabei allerdings auch auf Preis und Konkurrenzprodukte bezogen, relativieren. Wie oben bereits aufgezeigt, basieren diese Ungenauigkeiten auf bisher ungelösten Problemen rund um die Sentimentanalyse. Nichtsdestotrotz erhält der Kunde schnell eine Übersicht, in welchem Zusammenhang andere Kunden über einen Produktaspekt reden und kann so auch die Sentimentanalyse in diesen Kontext setzen.
Eine weitere Eigenschaft des Graphen, die auf der Reihenfolge der eingelesenen Kundenbewertungen basiert ist, dass externe Parameter diese Reihenfolge beeinflussen, wie etwa das Sortieren der Kundenbewertungen auf Amazon nach Relevanz oder Datum (was sich im ersten Schritt eingegebenen Link zu den Bewertungen widerspiegelt). So enthält der Graph z.B. ebenfalls eine Information darüber, wie alt oder relevant eine Kundenbewertung im Vergleich zu den anderen ist.

Implementierung: Die Implementierung unterscheidet sich vom vorherigen Schritt lediglich in der Aufbereitung und Darstellung der Daten, die durch die Sentimentanalyse gewonnen wurden.

Direktsuche

Design: Für Fälle, bei denen ein angezeigter Satz ohne Kontext keinen Sinn ergibt (siehe Abbildung 10, Textbeispiel zur Kundenbewertung 7), oder wenn der Kunde durch den Browser auf eine Kundenbewertung gestossen ist, die ihn ganzheitlich interessiert, kann er, indem er die Nummer der Kundenbewertung auswählt, auf den zugrundeliegenden Fliesstext zugreifen. Falls zuvor ein Produktaspekt ausgewählt wurde, werden die Sätze, die diesen beinhalten, ausserdem Fett dargestellt. Damit ist die Verbindung zum Originaltext geschaffen und ermöglicht so dem Kunden, ohne zwischen Fenstern wechseln zu müssen, direkt auf die Kundenbewertungen zuzugreifen. Dies erfolgt eingebettet in eine Analyseumgebung, womit der Wichtigkeit einer kontextuellen Einbindung Rechnung getragen wird.

Implementierung: Implementiert wurde dieser Schritt, indem Sätze mit dem Produktaspekt erkannt und dann in den ursprünglichen Kundenbewertungen mit Hilfe von HTML Tags fett markiert werden. Dies wird dann in der Webapplikation als Text dargestellt.

2.2.3 Visualisierung

Design: Da der Fokus dieser Arbeit sich nicht massgeblich auf das Design des User-Interfaces bezieht, liegt dieses in einer rudimentären Form vor und dient lediglich dazu dem Kunden das Testen der beschriebenen Funktionen zu ermöglichen, um damit später eine Evaluierung der Methoden durchführen zu können (Abbildung 12). Dennoch wurde hier der Versuch unternommen, soweit sinnvoll möglich, an gängige User-Interface Designpraktiken [Constantine and Lockwood, 1999] anzuknüpfen. Dabei wurde berücksichtigt:

- Das Strukturprinzip:
  Das User-Interface wurde so strukturiert, dass der Nutzer/Kunde sinnvoll durch die Funktionalitäten geleitet wird. Auf der linken Seite befinden sich alle notwendigen Eingabefelder dafür. So muss der Kunde zuerst eine Quelle von Kundenbewertungen auf Amazon angeben. Dann hat er die Möglichkeit die daraus resultierende Wortwolke nach seinen Bedürfnissen zu filtern, d.h bestimmte Produktaspekte zur Weiterverfolgung extrahieren. Letztlich erhält er die Möglichkeit eine Kundenbewertung als Volltext darzustellen. Zu jeder entsprechenden Aktion des Kunden werden die Informationen in der gleichen Reihenfolge entsprechend auf der rechten Seite im Visualisierungsteil abgebildet. Eine Ausnahme bildet dabei die Assoziationsanalyse. Diese wird direkt unter dem vom Kunden eingegebenen Produktaspekt dargestellt, um den Kontext zu verdeutlichen. Weiterhin sind alle Ein- bzw. Ausgabefelder zusätzlich beschrieben, um den Zusammenhang zwischen ihnen weiter zu verdeutlichen.

- Das Feedbackprinzip:
  Um dem Kunden zu signalisieren, dass seine Eingaben von den reaktiven Funktionen verarbeitet werden, erscheint in der rechten oberen Ecke des User-Interface je nach Bearbeitungsschritt eine Anzeige (Abbildung 11), welche Funktion gerade ausgeführt wird.

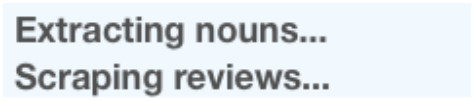

Abbildung 11: Feedback der reaktiven Funktionen

- Das Sichtbarkeitsprinzip:
  Alle, für jede Funktion benötigten Eingabeparameter sind gut ersichtlich. Weiterhin ist jeder Bereich des User-Interface eindeutig beschrieben bzw. beschriftet. Es wird keine redundante Information dargestellt.

- Das Einfachheitsprinzip:
  Es liegt in der Natur dieses Prototyps, ihn kundengerecht auf die wesentlichen Funktionen zu beschränken.

- Das Toleranzprinzip:
  Fehlerhafte Eingaben auf Seiten des Kunden fallen kaum ins Gewicht, da sie sich ohne weiteres korrigieren lassen und keine mehrstufigen, voneinander abhängigen, Eingabesequenzen benötigt werden.

Abbildung 12: Übersicht des User-Interface

Implementierung: Wann immer der Prototyp Information darstellen muss, greift er dabei auf das *Shiny* Framework in R zurück, welches es ermöglicht interaktive Webapplikationen direkt in R zu erstellen und dabei nur minimal in HTML, CSS oder JavaScript zu programmieren. Für die Anforderungen an den Prototypen der auf der Kundeninteraktion basiert, bietet sich dieses Framework optimal an. Weiterhin wird für die Darstellung der Graphen auf die *googleVis* Bibliothek zurückgegriffen, wobei der Produktaspekte-Browser eine durch HTML Anpassungen modifizierte Version eines *ScatterCharts* ist. Dazu zählen u.a. die sich verändernde Grösse der Punkte je nach Anzahl gefundener Produkteigenschaftensätze oder die Tooltips welche die entsprechenden Sätze beim MouseOver abbilden. Für die Wortwolke kommt, wie bereits erwähnt, die *wordcloud* Bibliothek zum Einsatz.
Um den Code übersichtlich zu halten und die Webscraping bzw. Analysefunktionen von den Visualisierungsfunktionen des Prototyps abzugrenzen und damit auch unabhängig von der Visualisierung nutzbar zu machen, wurde entschieden folgende Aufteilung für den Code zu wählen:

- library.R: Enthält alle Analysefunktionen sowie die dazu nötigen Hilfsfunktionen.

- global.R: Dient als Schnittstelle zwischen dem *Shiny* Framework und den Analysefunktionen. Hier werden die Daten der Analysefunktionen zur Visualisierung vorbereitet.

- server.R: Enthält die Funktionen der Webapplikation. Hier werden Nutzereingaben registriert und Ausgabedaten durch das *Shiny* Framework interaktiv dargestellt.

- ui.R: Kontrolliert das Layout des User-Interface der Webapplikation.

In Abbildung 13 wird diese Aufteilung im Detail dargestellt. Man erkennt die Einteilungen in die verschiedenen Funktionsbereiche durch die Rechtecke, die jeweils nach ihren Quelldateien benannt sind. So findet man in *library.R* die Implementierungen der Kernfunktionen, die in den vorherigen Abschnitten beschrieben wurden. Über die *global.R* Schnittstelle werden Informationen aus den Kernfunktionen entweder direkt weitergegeben oder für die Sentimentanalyse bzw. die Volltextdarstellung inklusive highlighting vorbereitet. Ausgelöst werden diese Prozesse durch die reaktiven Funktionen, die das *Shiny* Framework zur Verfügung stellt, und die sich in *server.R* befinden. Verändert der Kunde einen Parameter der Webapplikation, wird dies über die reaktiven Funktionen registriert. So reagiert die reaktive Funktion *reviews()* darauf, wenn der Kunde neue Kundenbewertungen einlesen möchte und einen Knopf drückt. Dabei hat der Kunde eine URL und die Anzahl der auszuwertenden Kundenbewertungen im UI festgelegt, die jetzt von der reaktiven Funktion an die Schnittstellenfunktion *getReviews()* weitergeleitet werden. Da keine spezifische Vorbereitung der Daten nötig ist, ruft *getReviews()* direkt die Kernfunktion *scrapeAmazon()* in *library.R* auf. Dadurch wird ein Vektor erstellt, der die gewünschten Kundenbewertungen enthält. Je nach Bedarf kann dieser Vektor mit Hilfe der *normaliseChrVector()* Funktion normalisiert werden. Der Vektor wird an die reaktive Funktion zurückgegeben und ist damit bereit für weitere Operationen. Im vorliegenden Fall wird der Vektor z.B dazu genutzt durch die reaktive Funktion *getNouns()* Nomen und deren Frequenz zu extrahieren, die dann mit Hilfe von *renderPlot()* als Wortwolke dargestellt werden.

Bei genauerer Betrachtung fällt auf, dass keine Funktion zur Assoziationsanalyse zu finden ist, obwohl diese Bestandteil des Prototyps ist. Dies hat den Grund, dass die Methode zur Assoziationsanalyse zwar eine Kernfunktion darstellt, aber ohne Anpassungen aus der *tm* Bibliothek übernommen werden kann und daher direkt in einer *renderUI()* Funktion aufgerufen wird.

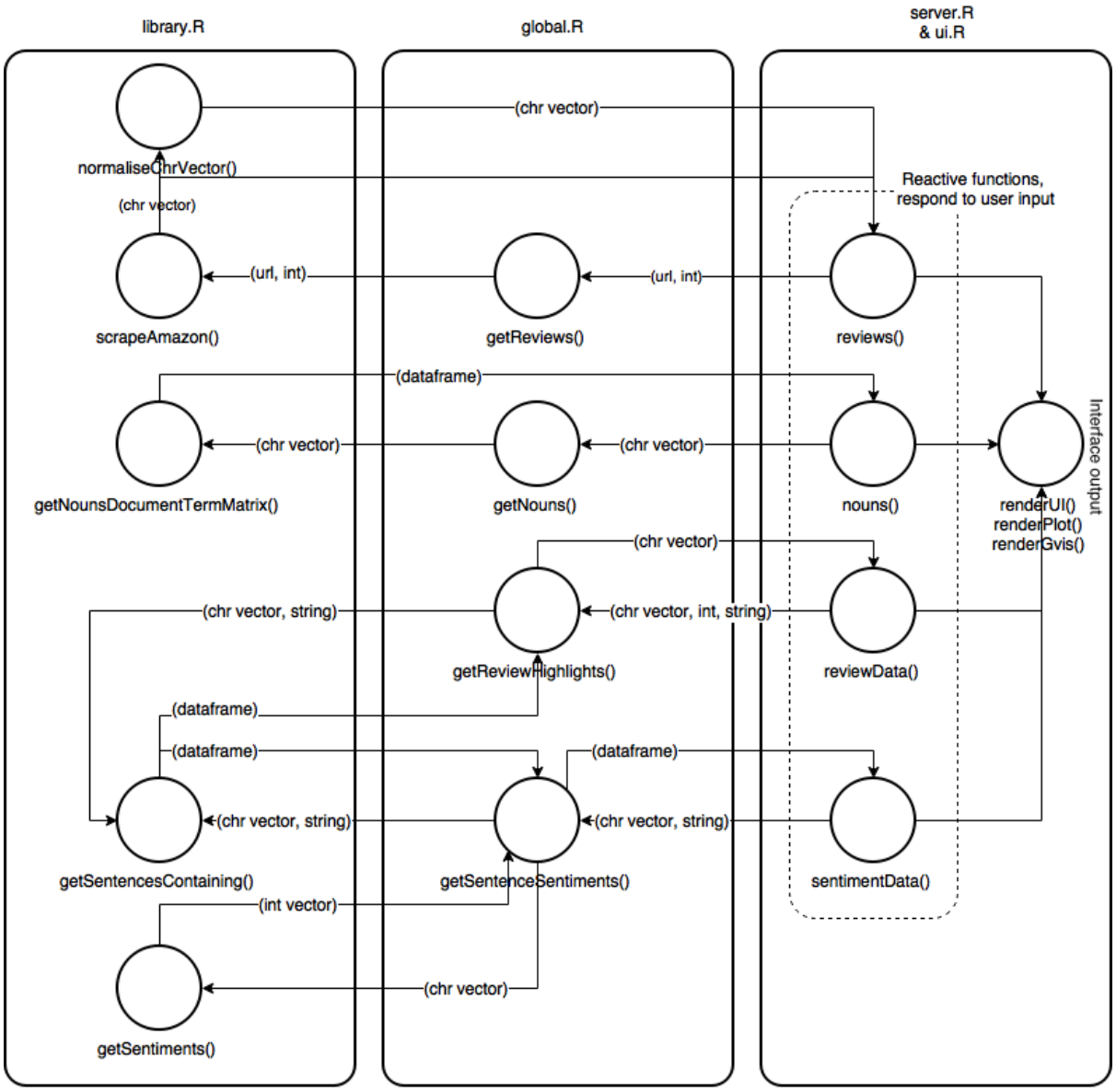

Abbildung 13: Detaillierte Schematische Darstellung der Funktionen des Prototyps, sowie deren Interaktionen und Aufteilungen

2.2.4 Wichtige Funktionen im Überblick

Im Folgenden wird ein tabellarischer Überblick zu den wichtigsten Funktionen des Prototyps gegeben:

Tabelle 1: Funktionsübersicht des Prototyps.

Funktion	Eingabe	Ausgabe	Beschreibung	Ursprung
renderUI()	Text	HTML Code	Funktion zu Generierung von HTML code für Text	Shiny
renderPlot()	Grafik	HTML Code	Funktion zu Generierung von HTML code für Bilder	Shiny
renderGvis()	googleVis Grafik	HTML Code	Funktion zu Generierung von HTML code für die googleVis Bibliothek	Shiny
reviews()	url, int	chr vector	Reaktive Funktion, die eingelesene Kundenbewertungen in Rohform zurückgibt	server.R
nouns()	chr vector	dataframe[5]	Reaktive Funktion, die Nomen und deren Häufigkeit zurückgibt	server.R
reviewData()	chr vector, int, string	chr vector	Reaktive Funktion, die einen mit HTML tags markierten Text zurückgibt	server.R
sentimentData()	chr vector, string	dataframe	Reaktive Funktion, die eine Satzbezogene Sentimentauswertung zurückgibt	server.R
getReviews()	url, int	url, int	Schnittstellenfunktion ohne Umwandlungsschritt	global.R
getNouns()	chr vector	chr vector	Schnittstellenfunktion ohne Umwandlungsschritt	global.R
getReviewHighlights()	chr vector, int, string	chr vector	Schnittstellenfunktion, die Sätze in einer Kundenbewertung mittels HTML Tags fett markiert	global.R
getSentenceSentiments()	chr vector, string	dataframe	Schnittstellenfunktion, die Sätze in einer Kundenbewertung mittels Sentimentanalyse auswertet	global.R
normaliseChrVector()	chr vector	chr vector	Funktion zur Normaliserung eines Textvektors	library.R
scrapeAmazon()	url,int	chr vector	Funktion zum Parsen von Kundenbwertungen	library.R
getNounsDocumentTermMatrix()	chr vector	DTM[6]	Funktion zur Analyse von Nomenhäufigkeit in einem Text	library.R
getSentencesContaining()	chr vector, string	dataframe	Funktion zur Auswahl von Sätzen in einem Text anhand eines Wortes	library.R
getSentiments()	chr vector	int vector	Funktion zur Sentimentanalyse eines Texts	library.R
findAssocs()	DTM, string	dataframe	Funktion zur Assoziationsanalyse eines Wortes in verschiedenen Texten	tm

[5] Ein Dataframe ist ein Datenobjekt der R Programmiersprache und dient dazu tabellarische Daten zu speichern

[6] Eine DocumentTermMatrix ist ein Datenobjekt der *tm* Bibliothek und liefert Auskunft über die Häufigkeit von verwendeten Ausdrücken in einem Text

2.2.5 Sequenzdiagram

Abbildung 14 zeigt eine exemplarische Nutzung des Prototyps und verschafft einen Überblick über das Zusammenspiel der beschriebenen Funktionen zwischen Amazon, Prototyp und Kunde.

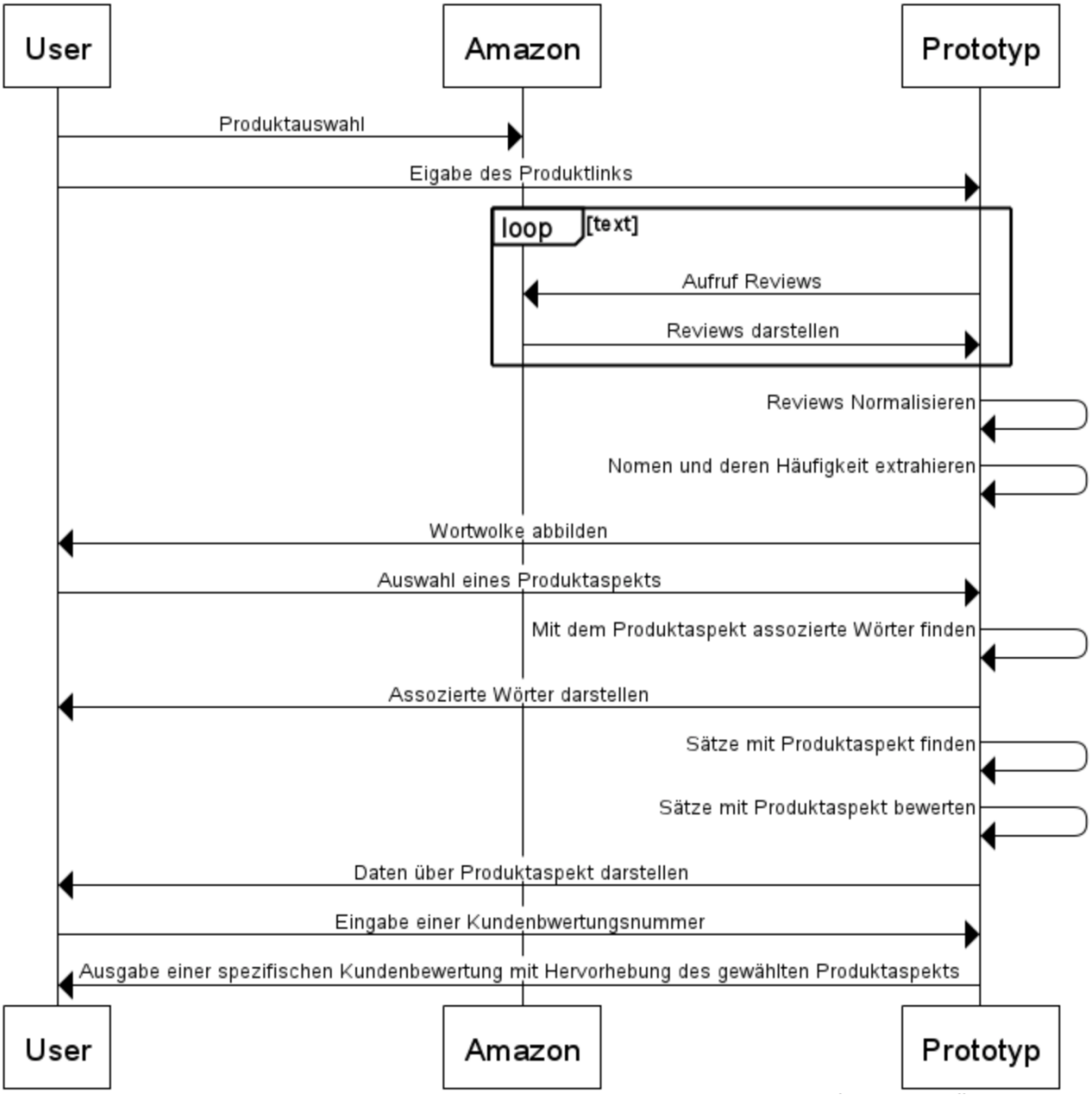

Abbildung 14: Sequenzdiagramm, welches einen beispielhaften Ablauf einer Nutzung des Prototyps zeigt.

2.3 Schwierigkeiten bei der Implementierung

Im Folgenden wird auf einige interessante Problematiken eingegangen, die bei der Implementierung des Prototyps gelöst werden mussten.

2.3.1 Webscraping

Die webscraping Funktionalität des Prototyps, welche anfänglich nur auf der *rvest* Bibliothek basiert war, hat eine entscheidende Schwäche erst im Verlaufe der Auswertung des Prototyps offenbart. Die *rvest* Bibliothek beinhaltet nativ eine Funktion *read_html(url)*, die es erlaubt eine HTML basierte Website zur weiteren Verarbeitung einzulesen (siehe Abbildung 15). Es kann allerdings vorkommen, dass vermutlich aufgrund von Unzulänglichkeiten bei der Datenübertragung oder Restriktionen seitens Amazon bezüglich des automatisierten Einlesen ihrer Daten, die HTML Seite durch diese Funktion nur unvollständig oder gar nicht abgebildet wird. Auch das Einbauen einer Warteschleife, um einen eventuell verfrühten Zugriff auf die eingelesene HTML Seite zu verhindern, hat daran nichts geändert. Dieser Sachverhalt hatte massive Auswirkungen auf die Datenqualität bzw. die grundlegende Nutzbarkeit des Prototyps. Letztlich ist es gelungen diese Unzulänglichkeit durch die Nutzung der *rSelenium* Bibliothek zu beheben, welche es ermöglicht, durch R, über einen lokalen Server einen voll funktionsfähigen Webbrowser anzusteuern. Die so dargestellte Website mit den Kundenbewertungen kann danach problemlos mit den *rvest* Funktionen eingelesen werden (siehe Abbildung 16). Ein Nachteil der sich daraus ergibt ist, dass nun zusätzlich für die Nutzung des Prototyps ein lokaler Server gestartet werden muss, um die *rSelenium* Webdriver API zu nutzen (näheres dazu in der ReadMe Datei des Prototyps).

```
#HTML Seite einlesen
amazon_review <- read_html(url)

#HTML Nodes parsen
reviewRaw <- amazon_review %>%
html_nodes(".review-text") %>%
html_text()
```

Abbildung 15: Der ursprüngliche Code zum direkten Einlesen einer HTML Seite mittels der nativen *rvest* Funktion *read_html()*

```
#HTML Seite einlesen
remDr$navigate(url)

#Warten bis die Seite geladen ist
Sys.sleep(5)

#Quellcode der Seite auslesen
page_source<-remDr$getPageSource()

#HTML Nodes parsen
reviewRaw <- read_html(page_source[[1]]) %>%
html_nodes(".review-text") %>%
html_text()
```

Abbildung 16: Der angepasste Code zum Einlesen einer HTML Seite mit Hilfe der *rSelenium* Bibliothek

2.3.2 Produktaspekt-Browser

Schwierigkeiten bei der Implementierung traten auch beim Produktaspekt-Browser auf. Wie vorab bereits angedeutet, basiert dieser auf einem *ScatterChart* aus der *googleVis* Bibliothek. Bei einer standardmässigen Webentwicklung können diese Graphen mit Hilfe von HTML und CSS gut angepasst werden. Das Problem liegt jedoch darin, dass die Generierung des HTML Codes in R vom *Shiny* Framework übernommen wird. Das heisst, man hat höchstens indirekt Einfluss auf die Umfunktionierung von verfügbaren Graphen. Um die Idee zur Darstellung von unterschiedlich grossen Datenpunkten und interaktiven Tooltips dennoch umzusetzen, wurde eruiert wie diese Informationen an den *googleVis* Graphen übermittelt werden kann. Bei der Durchsicht der relevanten Dokumentation wurde gefunden, dass man Daten, die im Graphen dargestellt werden, sogenannte Rollen zuweisen kann. Rollen ermöglichen es dem Graphen Daten auf verschiedene Weise zu interpretieren. Dabei muss man sich an vordefinierte Namenkonventionen halten damit beim Zeichnen des Graphen verstanden wird, welche Funktion eine Rolle hat. Auf diese Weise kann man zu jeder Datenreihe verschiedene Rollen von Zusatzdaten definieren. Aus Abbildung 17 ist ersichtlich wie die Ausgangsdaten für den Produktaspekt-Browser benannt werden müssen, damit man später zu jedem Datenpunkt eine entsprechende Grösse (sentiment.style) und einen Tooltip (reviewnumber.html.tooltip) zugeordnet erhält.

```
dfrm <- dfrm[c("sentiment", "review_number", "sentiment.style", "review_number.html.tooltip")]
```

Abbildung 17: Bennenung von Datenreihen und Rollen zur Vorbereitung der Darstellung durch die *googleVis* Bibliothek

Diese Vorgehensweise setzt voraus, dass man bevor der Graph erstellt werden kann ein komplett vorbereitetes Datenset zur Verfügung hat.

2.3.3 UTF8 Kodierung

Gerade bei den Textminingfunktionen stösst man oft auf Probleme bezüglich der Kodierung von Texten. Nimmt man beispielsweise die Funktion der *tm* Bibliothek, die ermöglicht bei einem Korpus (d.h einem Textvektor) die Gross-Kleinschreibung zu entfernen, dann erhält man bei falscher Kodierung des Korpusinhalts die Fehlermeldung *Fehler in FUN(content(x), ...) : ungültige multibyte Zeichenkette*. Die Funktionen der *tm* Bibliothek benötigen einen Ausgangstext der UTF-8 kodiert ist. Daher lässt sich der Fehler beheben, indem man die R Funktion *iconv* (Abbildung 18) anwendet, welche es ermöglicht Textkodierungen umzuwandeln.

```
review_text<-iconv(enc2utf8(review_text),sub="byte")
```

Abbildung 18: *iconv* Funktion zur Kodierung von Zeichen

2.3.4 R und Java

Dieser Abschnitt hat nur indirekt etwas mit dem Prototypen zu tun, da die hier zu erläuternde Problematik erst auftritt, wenn man versucht die *coreNLP* Bibliothek [Manning et al., 2014] zu implementieren. Die in Stanford entwickelte Bibliothek enthält neben ähnlichen Funktionen zur *openNLP* Bibliothek auch eine Sentimentanalysefunktion und ist Java basiert. Bei der Implementierung der Bibliothek wurde festgestellt, dass die Funktionen redundant sind und ausserdem im Vergleich zu äquivalenten Funktionen

der anderen Bibliotheken langsamer ausgeführt wurden. Es wird vermutet, dass dieses Phänomen dadurch zustande kommt, weil für deren Ausführung ein Java interface benötigt wird. Da für die Umsetzung des Prototyps die Funktionalitäten der anderen benannten Bibliotheken ausreichend erschienen, wurde die *coreNLP* Bibliothek nicht weiter verwendet.

An dieser Stelle seien dennoch weitere Erkenntnisse zur Implementierung dieser Bibliothek wiedergegeben, da sich die dazu nötige Verknüpfung von R mit Java als trickreich herausgestellt hat. Zunächst muss das Javainterface *rJava* mit Hilfe des Befehls *install.packages* in R heruntergeladen und installiert werden. Dazu muss eine aktuelle Java Version auf dem Rechner vorhanden sein. Falls eine neue Javaversion heruntergeladen wird, nachdem R installiert wurde, kann man unter UNIX mit dem Terminalbefehl *R CMD javareconf* die R Einstellungen bezüglich der Javaversion updaten. Sollte *rJava* mit der Fehlermeldung *"Error: package or namespace load failed for 'rJava'"* den Dienst verweigern, so muss die JVM Bibliothek *libjvm.dylib* richtig verlinkt werden. Unter MacOS, kann dies mit dem Terminalbefehl

sudo ln -f -s $(/usr/libexec/javahome)/jre/lib/server/libjvm.dylib /usr/local/lib

erreicht werden, indem *libjvm.dylib* mit dem Verzeichnis /usr/local/lib verknüpft wird. Danach kann das *rJava* Interface genutzt werden.
Bei sehr speicherintensiven Java Bibliotheken (wie etwa der *coreNLP* Bibliothek) die über dieses Interface verknüpft werden, kann zusätzlich mit Hilfe des R Befehls *options(java.parameters = Xmx4g")* die Heap Grösse der JVM angepasst werden.

3 Diskussion

In diesem Abschnitt soll betrachtet werden, inwiefern das vorgestellte, aus verschiedenen Analyse-Funktionen bestehende Design des Prototyps, als ganzheitliches Konzept den Anforderungen des Kunden an Information hinsichtlich Kompaktheit und Aussagekraft für den Einkaufsprozess gerecht wird, und welche Limitierungen dabei gegebenenfalls zu berücksichtigen sind. Damit wird insbesondere, mit Blick auf die Aussagekraft, die bearbeitete Arbeitshypothese validiert, um nach der Evaluierung (Kapitel 4) im Fazit (Kapitel 5) auf die zu dieser Arbeit gestellte Forschungsfrage abschliessend einzugehen.

3.1 Kompaktheit

Es liegt in der Natur des Textmining, dass Information verdichtet wird. Da die Funktionalität des Prototyps auf dem Textmining basiert, wird die Anforderung an eine Kompaktierung der Information daher zwangsläufig erfüllt. Dabei ist es wichtig, den Zusammenhang zwischen dem Verdichten der unstrukturierten Information und der resultierenden Aussagekraft der verdichteten Information zu erfassen und zu bewerten. Die Verdichtung erfolgt mit dem Ziel, die Aussagekraft zu verbessern. Im Optimalfall wird soweit verdichtet, dass eine maximale Konzentration an zielführender Information ohne Verlust an solcher erfolgt.
In der vorliegenden Ausführung des Prototyp bestehen diesbezügliche Freiheitsgrade des Kunden in der Wahl (i) der Anzahl an Produktaspekten bzw. (ii) der minimalen Anzahl an Nennungen der Aspekte, damit diese in der Wortwolke dargestellt werden und somit für die Auswahl zur Weiterbearbeitung mittels Sentimentanalysen zur Verfügung stehen.
Um die Diskussion zur Aussagekraft (Kapitel 3.2) der erhaltenen Information nach kompletter Bearbeitung durch den Prototyp inklusive interaktiver Einflussnahme durch den Kunden vorzubereiten, wird nachfolgend eine Übersicht zu den in dieser Arbeit durch den Prototyp genutzten Methoden der Informationsverdichtung gegeben. Im Kontext wird eine Beschreibung zu den Resultaten der einzelnen methodischen Schritte gegeben, die der Prototyp als Kundeninformation liefert (siehe 3.1.1 – 3.1.3).

3.1.1 Aspektextraktion

- Die Frequenz der Erwähnung von Produktaspekten.
- Die Frequenz der Erwähnung eines Produktaspekts im Verhältnis zur Anzahl der Kundenbewertungen.
- Sätze, in denen ein Produktaspekt genannt wird.
- Ursprung der Sätze beziffert durch eine Nummer, welche die Verbindung zum Originaltext herstellt.
- Den Quelltext mit Hervorhebung der entsprechenden Sätze.

3.1.2 Assoziationsanalyse

- Häufig im Zusammenhang mit relevanten Produktaspekten genannte Wörter.

3.1.3 Sentimentanalyse

- Bewertung der Sätze, in denen ein relevanter Produktaspekt vorkommt, durch vier unterschiedliche Ansätze, welche zwei Methoden (regel- und lexikonbasiert) der Sentimentanalyse koppeln.

- Ein Balkendiagramm, mit der Verteilung der Bewertung von Sätzen bezogen auf einen ausgewählten Produktaspekt.
- Einen Durchschnittswert, basierend auf der Bewertung der Sätze für einen ausgewählten Produktaspekt.
- Eine Übersicht der Kundenbewertungen inklusive einer Auswertung, wie diese Sentiment über einen ausgewählten Produktaspekt ausdrücken. (Produktaspekte-Browser)

3.2 Aussagekraft

Die Aussagekraft, der vom Prototyp bereitgestellten Information, kann nach zwei Hauptkriterien unterschieden werden: (1) Abbildung relevanter Sachverhalte zur Information über ein Produkt und (2) Qualität der Information.

3.2.1 Abbildung relevanter Sachverhalte

In diesem Abschnitt wird eruiert, welche Information die durch den Prototyp gewonnen wird, aus welchem Grund als sinnvoll für einen Kunden zu betrachten ist.

Produktaspekte: Betrachtet man die Verdichtungsarten für unstrukturierte Information aus Kundenbewertungen wie unter 3.1.1 - 3.1.3 beschrieben, wird ersichtlich, dass diese entweder bezogen auf den Überblick zu relevanten Produktaspekten oder auf einen spezifischen Produktaspekt im Detail erfolgt.
Damit wird dem Umstand Rechnung getragen, dass ein Kunde sich über ein Produkt, welches ihn interessiert, zunächst einen Überblick verschaffen will, bevor er ins Detail geht. Dieser Kundenstrategie folgt die zunächst vorgenommene Aggregation von Information auf Produktebene und liefert einen Überblick über wiederholt genannte und somit als relevant bewertete Produktaspekte. Zu dieser Kategorie der verdichteten Informationen würde auch eine durchschnittliche numerische Bewertung des Produkts z.B mittels einer “Sterne-Bewertung” gezählt werden. Prinzipiell könnte solche Information auch aus Komfortgründen übernommen und im Prototyp zusätzlich abgebildet werden.
Will sich ein Kunde im Detail mit einem Produkt befassen so tut er dies, indem er sich typischerweise genauer mit dessen bestimmten Produktaspekten (sprich Komponenten und Eigenschaften) auseinandersetzt. Die Intuition der bisherigen Forschung, sich im Gebiet der Auswertung von Kundenbewertungen auf Produktaspekte zu beziehen, ist somit als zielführend zu erachten und wird daher auch im Prototyp umgesetzt.

Assoziationen: Die Motivation zur Assoziationsanalyse für einen Produktaspekt dient dazu dem Kunden zu einem in der Aspektanalyse gefundenen Nomen weitere assoziierte Wörter zu ermitteln. Dies dient der möglichen Erweiterung der Aspektauswahl für die anschliessende, auf bestimmte Aspekte fokussierte Bewertung mittels Sentimentanalyse.

Sentiment: Will sich der Kunde näher mit einem Produktaspekt befassen, liegt es nahe darüber Informationen zu sammeln, die spezifische bewertende Auskunft darüber geben. Eine derart bewertende Information, die durch Textmining dabei extrahiert werden kann, entspricht einer Analyse der Stimmung, die zu einem solchen Aspekt ausgedrückt wird.

3.2.2 Qualität der Information

In diesem Abschnitt wird weitergehend diskutiert, welche, das Prototypdesign betreffende Entscheidungen, zu einer Verbesserung der Informationsqualität im Sinne einer Erhöhung der Aussagekraft der komprimierten Information beitragen. Dazu ist es zunächst erforderlich, zu definieren wie die Qualität dieser Information verstanden werden soll. Hierzu werden folgende drei Qualitätsaspekte separiert:

- Präzision: Ist die Information genau?
- Verständlichkeit: Kann die Information verstanden werden?
- Zielgerichtetheit: Lässt sich die Information auf das Produkt beziehen?

Vier Ansätze zur Sentimentanalyse: Die Bewertung des Sentiments wird mittels vier unterschiedlichen Ansätzen, welche sich zweier Methodiken zur Sentimentanalyse bedienen, vorgenommen. Deren Ergebnisse werden gleich gewichtet. Die Kombination der vier unterschiedlichen Ansätze lässt eine Verbesserung der Präzision erwarten.

Der "rote Faden": Der "rote Faden" ist Bestandteil der interaktiven Toolbox für die Kontrolle durch den Kunden und bezeichnet eine ersichtliche Verbindung von der aggregierten Information zurück zu den entsprechenden Teilen/Sätzen in den Originaltexten (unstrukturierte Kundenbewertungen).
Die in Form einer Bewertungsverteilung (Balkendiagramm und Produkt-Aspekt Browser) vorliegende Aussage zu einem Produktaspekt aus der Sentimentanalyse, kann unter Nutzung dieser Rückverbindung, welche der "Rote Faden" zulässt, kontrolliert und gegebenenfalls korrigiert werden. Bislang muss ein derartiger Korrekturschritt vom Kunden noch im Nachgang zur Sentimentanalyse inhaltlich qualitativ vollzogen werden, d.h er stellt die Verbindung zwischen Balkendiagramm und Produktaspekt-Browser her und überprüft so, ob die angezeigte Verteilung im Balkendiagramm sich korrekt aus den entsprechenden Sätzen, die auf die Produktaspekte eingehen, ableiten lässt.
Für die Zukunft ist geplant im Rahmen der Rückverfolgung vom Prototyp "falsch verstandene" Sätze durch Kundeneingabe im Produktaspekt-Browser zu eliminieren und die Auswirkung einer solchen Aktion, in die Bewertungsverteilung aus der Sentimentanalyse, direkt einfliessen und visualisieren zu lassen (siehe auch Fazit und Ausblick, Kapitel 5). In diesem Sinne dient der "rote Faden" dazu die Präzision aber auch die Verständlichkeit der extrahierten Information zu erhöhen.

Information über Information: Der Produktaspekte-Browser erlaubt es einem Kunden noch weitere Rückschlüsse über die Ursprungsinformation also die Kundenbewertung zu ziehen. Zum einen, wird für einen gewählten Produktaspekt angezeigt, welche Kundenbewertungen darüber Auskunft geben und gekoppelt mit der Sentimentanalyse auch die Art und Weise, wie dies bewertend erfolgt.
Wie in Kapitel 1.2.3 aufgezeigt, bewerten Kunden die Relevanz von Information unter anderem auch nach dem Kontext aus dem sie entsteht. Schaut der Kunde sich etwa einen für ihn relevanten Produktaspekt an, so kann er sich beispielsweise ein Bild davon machen, welche Kundenbewertungen darüber ausgeglichen oder eher einseitig berichten. Auch eine Information über die Intensität, mit der sich Kundenbewertungen mit einem Aspekt auseinandersetzten, ist durch die Anzahl sowie die Grösse der Punkte im Produktaspekte-Browser zugänglich. Weiterhin können numerische Bewertungskriterien, die bereits von Amazon zur Verfügung gestellt werden, im Prototyp genutzt werden. Da die Kundenbe-

wertungen in der Einlesereihenfolge erhalten bleiben, werden Kriterien, welche auf der entsprechende Website des Online-Händlers (hier: Amazon) z.B. nach Alter bzw. Relevanz sortiert wurden, in der Reihenfolge übernommen. Hat ein Kunde die Kundenbewertungen beispielsweise nach Relevanz vorsortiert, so wird er auch im Produktaspekte-Browser diese Sortierung vorfinden und hat dadurch ein weiteres Bewertungskriterium für die Information. Dies erhöht die Zielgerichtetheit der extrahierten Information.

Der interaktive Kunde als überprüfendes und präzisierendes Element: Bereits in der Hinleitung zur Forschungsfrage wurde aufgezeigt, dass die einzelnen Textmining-Methoden, sei es die Auswahl von Produktaspekten durch eine Nomenanalyse zu relevanten Aspekten oder deren Bewertung durch die Sentimentanalysefunktionen, immer auch mit Fehlern behaftet sein können und damit ein valider Kritikpunkt zu deren Aussagekraft besteht.
Um mit dieser Problematik umzugehen, wurde versucht, den Einfluss fehlerhafter Information abzuschwächen bzw. eine Korrektur zu ermöglichen, indem der interaktive Kunde als Bindeglied zwischen den Blöcken B1 (Aspektextraktion, Assoziationsanalyse) und B2 (Sentimentanalyse, sowie deren spezifischer Auswertung) eingebunden wird. Durch diesen Einbezug des Kunden in eine Umgebung in der er cross-referencing betreiben und selbst darüber entscheiden kann, welche Informationen aus der Aspektextraktion bzw. der Assoziationsanalyse er (gegebenenfalls auch ergänzt) in die Sentimentanalyse übergibt, erhöht sich die Präzision und somit die Aussagekraft der dargestellten Information. Hinsichtlich der erwähnten Ergänzungsmöglichkeit hat der Kunde auch, z.B via Produktaspekte-Browser, die Freiheit Aspekte zu überprüfen, die nicht im Ergebnis der Aspektextraktion (Wortwolke) auftauchen. Die in dieser Arbeit beschriebene Methodik gewinnt für den Kunden somit an Präzision und Flexibilität gegenüber publizierten Methoden, die Produkteigenschaftserkennung und deren Bewertung im selben Schritt und ohne Interaktionsmöglichkeit durch den Kunden vollziehen.

3.3 Limitierungen

Abschliessend soll in diesem Kapitel auf die Limitierungen des Prototyps eingegangen werden:

Webscraping: Eine Limitierung hierzu ergibt sich aus den Implementierungsschwierigkeiten des Webscraping. Wie aufgezeigt hilft der Umweg über den automatisierten Webbrowser in den meisten Fällen. In seltenen Fällen kann es dennoch vorkommen, dass trotz dieser Massnahme Kundenbewertungen nicht korrekt eingelesen werden. In einem solchen Fall wird dies daraus ersichtlich, dass im Produktaspekt-Browser sich wiederholende Muster zu erkennen sind. Die Funktionen der *rSelenium* Bibliothek laden die HTML Seite offensichtlich neu, falls es beim Laden zu Verbindungsproblemen kommt. Dadurch werden manche Kundenbewertungen mehrfach angezeigt. Abhilfe schafft dann in aller Regel einfach das erneute Einlesen der Kundenbewertungen.

Produktaspekterkennung: Die Erkennung von Produktaspekten berücksichtigt im Vergleich zur Methodik von Hu und Liu keine zusammenhängenden Wörter. Die Qualität der Information kann dadurch etwas zurückgehen. Dies soll wie beschrieben dadurch kompensiert werden, dass der Kunde die Resultate selbst zusätzlich zwischenbewerten kann, bevor diese z.B mittels Sentimentanalyse weiterbearbeitet werden. Die Tatsache, nicht bereits in der initialen allgemeinen Produktaspektanalyse eine Assoziationsanalyse durchzuführen, begründet sich damit, dass die Auswertung der Nomenfrequenz von allen Funktionen des Prototyps am meisten Zeit beansprucht. Die Laufzeit dieser Operation bei n

eingelesenen Wörtern berechnet sich wie folgt:

Einlesen der Wörter:

$$O(n) \tag{1}$$

Markierung der Wörter:

$$O(2n) \tag{2}$$

Filterung von Nomen:

$$O(n) \tag{3}$$

Gesamt:

$$O(n + 2n + n) = O(4n) \tag{4}$$

Das heisst zusammengefasst liegt eine lineare Laufzeiterhöhung, abhängig von der Anzahl der eingelesen Wörter, vor. Würde zusätzlich eine Assoziationsanalyse für jedes Nomen unternommen, veränderte sich die Laufzeit wie folgt:
Die Laufzeit des Apriori-Algorithmus [Kumar et al., 2006], liegt für jedes Objekt *n*, das in Bezug auf ein Datenset überprüft werden muss, bei:

$$O(n^2) \tag{5}$$

Würde man also für jedes Nomen überprüfen wollen mit welchen anderen Nomen es im Zusammenhang genannt wird, ergäbe sich daraus eine Laufzeit von:

$$O(4n^2) \tag{6}$$

Dieses quadratische Wachstum der Laufzeit würde schon bei wenigen Reviews deutliche Zeiteinbussen bedeuten.

Assoziationsanalyse: Bei der Assoziationsanalyse im Rahmen der Darstellung von Wörtern die häufig mit einem Produktaspekt genannt werden, wird die Position der Wörter im Text nicht berücksichtigt. Es ist allerdings unklar, ob eine Berücksichtigung die Aussagekraft der Information verbessern würde. Anders als Hu und Liu die Assoziationsanalyse verwenden, um Produktaspekte, die aus mehreren Wörtern bestehen, zu identifizieren (und daher deren Stellung im Text berücksichtigen), wird sie im Prototyp dazu genutzt generelle Aussagen über einen Produktaspekt abzuleiten. Es ist nicht ersichtlich, ob relevante Wörter, die etwas mit einem Produktaspekt zu tun haben, stets nah bei diesem Produktaspekt zu finden sein müssen. In diesem Sinne ist dies weniger eine Limitierung als ein Hinweis. Weiterhin kann bei zusammengesetzen Produktaspekten (Produktkomponente und Produkteigenschaft, z.B “battery life”) durch den Prototyp bisher keine Assoziationsanalyse durchgeführt werden.

4 Evaluierung

In diesem Kapitel soll geprüft werden, inwieweit mit Hilfe der im Prototypen vorgenommenen praxisorientierten Textmining-Methoden Kopplung, unter Einbezug der zusätzlich integrierten interaktiven Zugriffsmöglichkeit durch den Kunden, verbesserte Aussagekraft bei minimalem Zeitaufwand realisiert werden kann.

Dazu wurde betrachtet, wie Test-Kunden, die den Prototyp nutzen, im Vergleich zu einem Referenzkunden, der die Kundenbewertungen manuell auswertet, abschneiden.

4.1 Evaluierungsmethodik

Ein Produkt aus der Amazon Datenbank mit einer hinreichend hohen Zahl an Kundenbewertungen wurde vorgegeben. Aus hundert aus Kundensicht am meisten hilfreich bewerteten Kundenbewertungen wurde von Seiten eines Referenzkunden der erste Abschnitt der Evaluation vorgenommen, indem er (1) diese Reviews durchlas, (2) massgebliche Aspekte (Eigenschaften und Komponenten) des Produkts in Form von Nomen identifizierte und (3) jedem der identifizierten Aspekte basierend auf jeder Erwähnung in den Kundenbwertungen, eine gemäss seiner Textinterpretation subjektive Sentiment-Bewertung (positiv, negativ, ausgewogen) zuordnete.

Das Ergebnis einer derartigen Referenzkunden-Bewertung wurde mit den Ergebnissen einer repräsentativen Anzahl an (Testkunden) Bewertungen mittels Prototyp verglichen. Dabei erfolgte im Einzelnen:

(a) ein Vergleich der vom Referenzkunden ermittelten relevanten Produktaspekte mit den von den Testkunden unter Nutzung des Prototyps extrahierten Produkt Hauptaspekte

(b) ein Vergleich vom Referenzkunden zu den von ihm ermittelten relevanten Produktaspekten vorgenommenen Bewertungen, mit den von den Testkunden aus dem Balkendiagramm der Sentimentanalysenauswertungen mittels Prototyp, zum selben Produktaspekt, herausgelesenen positiv-/neutral-/negativ-Tendenzen. Letztere können von Testkunden gegebenenfalls interaktiv korrigiert worden sein, basierend auf der Detaileinsicht in einzelne Kundenbewertungen über den Produktaspekte Browser.

(c) ein Vergleich der Bearbeitungszeit für die Produktevaluierung gemäss (a,b) auf Basis von 100 Kundenbewertungen zwischen dem Referenzkunden und den Testkunden.

Den Testkunden war dabei freigestellt im Test mit dem Prototypen an der Schnittstelle zwischen Aspektanalyse/Assoziationsanalyse und Sentimentanalyse zu interagieren. Die Grundform solcher Interaktion bestand in der Extraktion der aus Testkundensicht als massgeblich bewerteten Nomen. Zu diesen konnte der Testkunde gegebenenfalls Zusatznomen generieren. Diese konnten aus der Aspekt- bzw. Assoziationsanalyse entnommen oder frei gewählt werden. Danach folgte die Weitergabe an die Sentimentanalyse.

Eine Quantifizierung der Evaluierungsergebnisse aus (a) und (b) bewertet (1) den Prozentsatz der übereinstimmenden Produktaspekte relativ zur Referenzkunden-Analyse und (2) den Prozentsatz der

zwischen Referenz- und Testkundenbewertung übereinstimmenden Sentimentbewertungen.
Die Vorgehensweise in Einzelschritten, wie sie zu Anleitung/Training von Referenz- und Testkunden realisiert wurde, wird nachfolgend beschrieben:

4.1.1 Referenzkunde

1. Lesen der 100 als am hilfreichsten deklarierten Kundenbewertungen zum Produkt X, gemäss Angaben der Website (Amazon)
2. Extraktion der massgeblichen Produktaspekte in Nomen-Form, basierend auf der Häufigkeit ihrer Nennung.
3. Bewertung jeder Nennung eines Produktaspekts auf einer Bewertungsskala mit den Kategorien - (=negativ), o (= neutral) , + (= positiv)
4. Ableitung einer Tendenz aus den Bewertungen eines Produktaspekts

4.1.2 Testkunden

1. Erklärung des Prototyps durch den Testleiter
2. Einlesen von hundert als am hilfreichsten deklarierten Kundenbewertungen zum ausgewählten Produkt
3. Analyse der Wortwolke durch den Testkunden
4. Auswahl der 6 wichtigsten Produktaspekte
5. Gegebenenfalls Ergänzung der sechs wichtigsten Produktaspekte mit einem zusätzlichen, präzisierenden Nomen (aus Assoziationsanalyse oder frei gewählt)
6. Eingabe der Produktaspekte in Sentimentanalyse
7. Auswertung der Sentimentanalyse (Balkendiagramm, Produktaspekte-Browser) zu jedem Produktaspekt in Form einer Aussage über die wahrgenommene Tendenz anhand der einfachen Bewertungsskala (-/o/+) des Referenzkunden.

4.1.3 Vergleich

Quantitativer Vergleich der Ergebnisse aus Referenzkundenschritt 2 mit Testkundenschritt 4 und Referenzkundenschritt 4 mit Testkundenschritt 7

4.2 Testpersonen

Als Testkunden wurden Personen ausgewählt, die alle bereits Erfahrung mit Onlineshopping gemacht hatten und über grundlegende Computerkenntnisse, sowie Kenntnisse der englischen Sprache verfügten.

4.3 Testprodukt

Als Testprodukt diente ein Staubsauger.

4.4 Auswertung

4.4.1 Referenzkundenauswertung

Tabelle 2: Ergebnisse der Referenzkundenauswertung

Aspekt	Totale Anzahl der Erwähnungen	+	o	-	Abgeleitete Tendenz
canister	44	12	16	16	o
weight	40	35	4	1	+
suction	37	25	11	1	+
hose	27	8	8	11	o
cord	26	10	5	11	o
price	24	23	1	0	+

Um zu diesem Ergebnis zu kommen, benötigte der Referenzkunde 181 Minuten.
Zur Ermittlung der “abgeleiteten Tendenz” erfolgte Differenzenbildung der Anzahlen für die “+” und “-“ Bewertungen. Die resultierende Differenz wurde mit den neutralen “o“ Bewertungen verglichen.

4.4.2 Testkundenauswertung

Tabelle 3: Ergebnisse der Testkundenauswertung: Aspektextraktion

Proband	Aspekt 1	Aspekt 2	Aspekt 3	Aspekt 4	Aspekt 5	Aspekt 6
1	canister	suction	cord	hose	price	handle
2	canister	suction	hose	cord	price	handle
3	canister	suction	cord	hose	handle	price
4	canister	suction	hose	cord	price	handle
5	canister	suction	hose	cord	handle	price
6	canister	hose	suction	cord	handle	price
7	canister	hose	use	cord	suction	handle
8	canister	suction	hose	cord	handle	price
9	canister	hose	cord	suction	price	handle
10	canister	hose	suction	cord	handle	price

Tabelle 4: Ergebnisse der Testkundenauswertung: Sentimentanalyse

Aspekt	Proband 1	Proband 2	Proband 3	Proband 4	Proband 5	Proband 6	Proband 7	Proband 8	Proband 9	Proband 10
canister	o	o	o	o	o	-	o	o	+	o
suction	+	+	+	+	+	+	+	+	+	+
cord	o	-	o	o	o	-	+	-	+	o
hose	o	-	o	-	o	o	+	o	+	-
handle	+	o	+	+	+	+	+	+	+	o
price	+	+	+	+	+	+	x	+	+	+
use	x	x	x	x	x	x	+	x	x	x

Tabelle 5: Ergebnisse der Testkundenauswertung: Zeitaufwand

Proband Nr.	Zeitaufwand
1	9 Minuten 52 Sekunden
2	14 Minuten 25 Sekunden
3	8 Minuten 20 Sekunden
4	10 Minuten 40 Sekunden
5	12 Minuten 15 Sekunden
6	9 Minuten 11 Sekunden
7	4 Minuten 35 Sekunden
8	9 Minuten 12 Sekunden
9	6 Minuten 45 Sekunden
10	12 Minuten 08 Sekunden
Durchschnitt	9 Minuten 44 Sekunden

4.4.3 Synthese

Aspektextraktion: Bei der Extraktion von relevanten Produktaspekten fällt unmittelbar auf, dass eine hohe Übereinstimmung zwischen den Referenzkunden- und Testkundenbewertungen vorliegt. Alle Testkunden haben mittels Prototyp die selben Produktaspekte in nahezu gleicher Reihenfolge gefunden. Die Unterschiede in der Reihenfolge liegen vermutlich daran, dass Aspekte, die gleich oder ähnlich häufig vorkommen, in der Wortwolke entsprechend gleich oder nahezu gleich gross abgebildet sind. Auffällig ist, dass im Gegensatz zum Referenzkunden keiner der Testkunden den Produktaspekt “weight” zu den sechs wichtigsten Aspekten zählt. Bei einer genaueren Betrachtung wird ersichtlich, dass viele Kundenbewertungen im Zusammenhang mit dem Gewicht von “lightweight” sprechen. Da dies kein Nomen ist, kommt es in der Wortwolke nicht vor und wird daher von den Testkunden nicht erkannt.
Es gibt natürlich auch Kundenbewertungen, die anstatt des genannten Adjektivs “lightweight”, das Nomen “weight” benutzten. Durch diese Unterteilung kam dieses jedoch deutlich weniger oft, als viele andere Aspekte, in der Wortwolke vor. Das heisst, letztlich wird auch das Gewicht als Produktaspekt erkannt, allerdings wird die Häufigkeit der Nennungen falsch eingeschätzt. Anstelle von “weight” reihte sich so bei der Auswertung mittels Prototyp durch Testkunden in mehreren Fällen der Aspekt “handle” unter die sechs wichtigsten Produktaspekte. Auch wurde von einem Probanden “use” (=Nutzen) als Produktaspekt ausgelegt, allerdings fällt diese Nennung in der Wortwolke in die Kategorie eines falsch erkannten Nomen.
In diesem Fall werden also von allen, ausser einem Probanden, 5 von 6 der wichtigsten Aspekte des Prototyps erkannt, was einer Genauigkeit von 81.48% entspricht.

Sentimentanalyse: Was die bewertende Einschätzungen der Testkunden bezüglich der Produktaspekte betrifft, so wurden auch hierbei gute Übereinstimmungen mit dem Referenzkunden erzielt. In 75.51% der Fälle stimmten die Testkunden mit dem Referenzkunden in der Bewertung der gleich erkannten Produktaskpekte überein. Die schlechtesten Werte der Übereinstimmung tauchten bei Produktaspekten auf, die auch vom Referenzkunden nur schwer einzuschätzen waren, und für welche die Meinungen der zugrundeliegenden, unstrukturierten Kundenbewertungen teilweise stark auseinander gingen. So gab es z.B Kundenbewertungen, welche die Länge des Staubsaugerkabels als völlig ausreichend oder sogar gut befanden, während andere sich über das zu kurze Kabel beschwerten. Dies liegt an der Subjektivität dieser Einschätzungen.

Zeitaufwand: Im Vergleich zum Referenzkunden, der 181 Minuten benötigte um die vorab beschriebene Analyse durchzuführen, brauchten die Probanden mit Unterstützung durch den Prototypen zur Auswertung im Schnitt lediglich 9 Minuten 44 Sekunden.

4.5 Evaluierungsfazit

Im Rahmen dieser ersten semi-quantitativen Evaluierung des Prototyps, lassen sich solide positive Ergebnisse feststellen. Bei der Aspektextraktion liegt die Extraktionsquote der Testkunden durch den Prototyp für die wichtigsten Produktaspekten bei einer Genauigkeit von 81.48%. Die Einschätzungen zur Qualität der Produktaspekte durch Testkunden und Referenzkunde deckte sich mit einer Genauigkeit von 75.51%.

Der wesentliche Unterschied lag im Zeitaufwand, der für die Auswertung entstand. So waren die Testkunden gegenüber dem Referenzkunden im Schnitt um ca. Faktor 19 schneller, um 100 Kundenbewertungen auszuwerten.

5 Fazit und Ausblick

Zum Abschluss der Arbeit soll zunächst eine Stellungnahme zur Arbeitshypothese erfolgen. Hierzu wird diese zunächst einer “Merkmalsanalyse” (Untergliederung nach Merkmalseinheiten M1-M4) unterzogen und danach eine Beschreibung zu den Merkmalen M1 bis M3 in der Wechselbeziehung zum “Leistungs-Merkmal” M4 gegeben.

M1: Eine praxisorientierte Textmining-Methoden-Kopplung von (1) Aspektextraktion, (2) Assoziationsanalyse, (3) Sentimentanalysen und

M2: einer erweiterten Auswertung der Resultate der Sentimentanalysen, sowie

M3: der Einbezug einer interaktiven Zugriffsmöglichkeit durch den Kunden,

M4: lassen eine qualitative Steigerung der Aussagekraft gegenüber der Anwendung der jeweils einzelnen oder auch gekoppelten Methoden ohne interaktive Zugriffsmöglichkeit erwarten.

M1 bis M3 waren dabei in Bezug auf M4 zu bewerten. Hierzu wurden folgende Aussagen aus den Resultaten der Arbeit abgeleitet:

M1-M4: Ohne korrektive Kunden-interaktive Möglichkeiten ist zu erwarten, dass M1 vergleichbare Resultate zum Stand der Technik liefert (z.B. [Hu and Liu, 2004][Liu, 2012])

M2-M4: Die Berücksichtigung von vier Varianten der Sentimentanalyse: (1) Syuzhet [Jockers, 2016]; (2) Afinn [Nielsen, 2011]; (3) Bing [Hu and Liu, 2004][Liu, 2012]; (4) NRC [Mohammad et al., 2013], welche sich durch die verwendeten Lexika und Regeln unterscheiden und deren erweiterte Auswertung (z.B durch Nutzung des “roten Fadens” oder der “Inormation über Information”) steigern die Aussagekraft.

M3-M4: der Einbezug einer interaktiven Zugriffsmöglichkeit durch den Kunden gibt eine wesentliche Erweiterung der Flexibilität zur Korrektur falsch verstandener Sätze hinsichtlich deren Produkt-Aspekt und dessen Bewertung. Wenngleich der Zeitaufwand gegenüber einer Textmining-Methodik ohne interaktiven Eingriff des Kunden damit voraussichtlich ansteigt, gewinnt die Aussagekraft in Folge Eliminierungsmöglichkeit falsch verstandener Sätze.

Zur Arbeitshypothese kann somit zusammenfassend angemerkt werden, dass insgesamt eine Steigerung der Aussagekraft mit dem in dieser Arbeit verfolgten Ansatz erreicht werden sollte.

Die soweit vorgenommene Evaluierung (siehe Kapitel 4) stellt nur einen ersten Schritt dar, hinsichtlich einer quantitativen Untermauerung dieser Schlussfolgerung zur Arbeitshypothese. Erweiterte Evaluierungsansätze sollten in der Weiterverfolgung der Themenstellung zu dieser Arbeit generiert werden (siehe Ausblick).

Aus diesen Folgerungen zur Arbeitshypothese kann folgende Ableitung zur Beantwortung der Forschungsfrage generiert werden:

Textmining-Methodik kann als Werkzeug, das aus unstrukturierten Kundenbewertungen zu einem bestimmten Produkt dem interessierten Dritten eine aussagekräftige bewertende Übersicht liefert, verbessert werden durch Kopplung sich ergänzender Textmining-Methoden, deren erweiterter Auswertung, sowie durch interaktive Eingriffsmöglichkeiten des Kunden, welche Unzulänglichkeiten beim automatisierten Textverständnis korrigieren lassen. Dabei muss jedoch vorraussichtlich mit einer Erhöhung des Zeitaufwandes gerechnet werden, im Vergleich zu vollautomatisierten Ansätzen ohne interaktiven Kundeneingriff.

Darüberhinaus hat die erfolgte Evaluierung des in dieser Arbeit verfolgten Ansatzes für einen Prototyp gezeigt, dass die Aussagekraft der mittels dieses Prototyps generierten Resultate sehr nahe an das direkte Lesen und das summarische Bewerten von Kundenbewertungen heranreicht und dabei der Zeitaufwand um ein mehrfaches (ca. Faktor 19, siehe Kapitel 4) minimiert werden kann. Somit entspricht der entwickelte Prototyp den Anforderungen eines Kunden hinsichtlich Kompaktheit und Aussagekraft der Information.

Direkte quantitative Vergleiche zu den in der Forschung vorgestellten Methoden zur Informationsgewinnung aus Kundenbewertungen liessen sich im Rahmen dieser Arbeit nur bedingt realisieren. Die Etablierung eines Referenzkunden liess jedoch eine valable Vergleichsbasis schaffen.

Ausblick: Durch den modularen Aufbau des erstellten Prototyps lässt sich dieser bequem durch neue Funktionen erweitern. So scheint es z.B sinnvoll weitere Sentimentanalysemethoden zu betrachten und einzubeziehen, nämlich solche, die auf statistischen Methoden oder Maschinenlernen basieren, um somit die vier derzeit eingesetzten Methoden zu erweitern.
Weiterhin erscheint es sinnvoll den “roten Faden” wie in Kapitel 3 besprochen auszubauen, indem dem Kunden ermöglicht wird fehlverstandene Sätze bzw. Aspekte zu löschen und dabei gleichzeitig diese Korrektur-Einflussnahme auf das Resultat der Sentimentanalyse direkt verfolgen zu können. Damit steigt der Arbeitsaufwand geringfügig, aber die Aussagekraft der Information sollte damit weiter zu verbessern sein.

Diese Massnahmen sind vor allem auch im Ausblick auf die 3. Phase des Kaufentscheidungsprozesses eines Kunden sinnvoll. Diese 3. Phase, in welcher der Kunde Produkte vergleicht, wurde in dieser Arbeit nur implizit berücksichtigt. Aussagekräftige und kompakte Information ist die Grundlage für einen sinnvollen Vergleich. Im Design des Prototyps besteht Spielraum zur direkten Unterstützung eines Produktevergleichs.

Weiterhin kann man den Prototypen als eigenständiges Browser-AddOn weiterentwickeln, welches z.B automatisch Kundenbewertungen beim Scrollen durch eine Produktliste ausliest und die Informationen kompakt zu jedem Produkt anzeigt.

Ein Fülle weiterer Entwicklungsmöglichkeiten!

6 Literatur

[Agrawal et al., 1993] Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. *SIGMOD '93 Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pages 207–216.

[Collomb et al., 2014] Collomb, A., Costea, C., Joyeux, D., Hasan, O., and Brunie, L. (2014). A study and comparison of sentiment analysis methods for reputation evaluation. Technical report, INSA Lyon.

[Constantine and Lockwood, 1999] Constantine, L. and Lockwood, L. A. D. (1999). *Software for use: a practical guide to the models and methods of usage-centered design*. Google Books.

[de Albornoz et al., 2011] de Albornoz, J. C., Plaza, L., Gervas, P., and Diaz, A. (2011). A joint model of feature mining and sentiment analysis for product review rating. In *Lecture Notes in Computer Science*, pages 55–66. Springer.

[Flanagin et al., 2014] Flanagin, A. J., Metzger, M. J., Pure, R., Markov, A., and Hartsell, E. (2014). Mitigating risk in ecommerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention. *Science+Business Media*.

[Han et al., 2004] Han, J., Pei, J., Yin, Y., and Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8:53–87.

[Horrigan, 2008] Horrigan, J. A. (2008). Online shopping. *Pew Internet and American Life Project Report*.

[Hu and Liu, 2004] Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. Technical report, University of Illinois.

[Jockers, 2016] Jockers, M. (2016). *Package 'syuzhet'*. Nebraska Literary Lab.

[Khan et al., 2011] Khan, A., Baharudin, B., and Khab, K. (2011). Sentiment classification using sentence-level lexical based semantic orientation of online reviews. Technical report, Universiti Teknologi PETRONAS Perak.

[Koch, 2007] Koch, D. (2007). *Suchmaschinen-Optimierung: Website-Marketing für Entwickler*. Pearson Deutschland.

[Kotler and Keller, 2012] Kotler, P. and Keller, K. L. (2012). *Marketing Management*. Pearson, 14 edition.

[Kumar et al., 2006] Kumar, V., Steinbach, M., and Tan, P. (2006). *Introduction to Data Mining*. University of Minnesota.

[Liu, 2012] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Claypool Publishers.

[Manning et al., 2014] Manning, C. D., Surdeanu, M., Bauer, J., and Finkel, J. (2014). The stanford corenlp natural language processing toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations System Demonstrations*.

[Mohammad et al., 2013] Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. Technical report, National Research Council Canada.

[Morinaga et al., 2002] Morinaga, S., Yamanishi, K., Tateishi, K., and Fukushima, T. (2002). Mining product reputations on the web. Technical report, NEC Corporation.

[Mudambi and Schuff, 2010] Mudambi, S. M. and Schuff, D. (2010). What makes a helpful online review? a study of customer reviews on amazon.com. *MIS Quarterly*, 34(1):185–200.

[Nielsen, 2011] Nielsen, F. A. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. Technical report, Technical University of Denmark.

[Pang and Lee, 2008] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. Technical report, Yahoo! Resarch, Cornell University.

[Pang et al., 2002] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 10.

[Popescu and Etzioni, 2007] Popescu, A. and Etzioni, O. (2007). Extracting product features and opinions from reviews. In *Natural Language Processing and Text Mining*.

[Russom, 2007] Russom, P. (2007). Bi search and text analytics. Technical report, The Data Warehousing Institute.

[Schulman et al., 1989] Schulman, P., Castellon, C., and Seligman, M. (1989). Assessing explanatory style: The content analysis of verbatim explanations and the attributional style questionnaire. *Behaviour Research and Therapy*, 27(5):505–512.

[Shilakes and Tylman, 1998] Shilakes, C. C. and Tylman, J. (1998). Enterprise information portals. Technical report, Merrill Lynch.

[Willemsen et al., 2011] Willemsen, L. M., Neijens, P. C., Bronner, F., and de Ridder, J. A. (2011). "highly recommended!" the content characteristics and perceived usefulness of online consumer reviews. *Journal of Computer-Mediated Communication*, 17:19–38.

[Zhang and Narayanan, 2010] Zhang, K. and Narayanan, R. (2010). Voice of the customers: Mining online customer reviews for product feature-based ranking. Technical report, Northwestern University.

7 Anhänge

7.1 Fragebogen zur Evaluierung des Prototyps

Fragebogen zur Evaluierung eines Prototyps zur Einkaufsunterstützung bei Amazon

Bitte entnehmen Sie die sechs wichtigsten Produktaspekte (Komponenten und/oder Eigenschaften) die sie erkennen können, beginnen sie mit dem wichtigsten Aspekt:

Bitte tragen Sie die Produktaspekte in die Nachfolgende Tabelle ein und bewerten sie ihre Einschätzung basierend auf den Informationen die Ihnen der Prototyp liefert (positiv, neutral, negativ):

Aspekt	+	0	-

Gestoppte Zeit:

Name: Datum:

7.2 USB Stick

Inhalte:

- Dokumentierter Programmcode des Prototyps sowie Anleitung zu dessen Ausführung
- Eingescannte Evaluierungsfragebögen
- Die Arbeit als PDF
- Zusammenfassung als Textdatei
- Abstract als Textdatei