

# Empirische Analyse von Bewertungskommentaren des Reputationssystems von eBay

Jörg Botsch, Stefan Luckner

Institut für Informationswirtschaft und –management (IISM)  
Universität Karlsruhe (TH)  
Englerstr. 14  
76131 Karlsruhe  
botsch@bimbly.com  
luckner@iism.uni-karlsruhe.de

**Abstract:** Online-Auktionshäuser wie eBay nutzen Reputationssysteme zur Lösung des Vertrauensproblems. Ein Nachteil heutiger Reputationssysteme ist, dass Händler ihre Bewertungen nicht immer entsprechend ihrer wahren Meinung abgeben. Dieser Beitrag untersucht empirisch anhand von Bewertungskommentaren des Reputationssystems von eBay, ob und in welchem Umfang es Abweichungen zwischen den abgegebenen Bewertungen und den damit verbundenen Bewertungskommentaren gibt. Die Untersuchung zeigt, dass Händler Transaktionen teilweise positiv beurteilen, obwohl sie unzufrieden sind und diese Unzufriedenheit auch durch entsprechende Bewertungskommentare zum Ausdruck bringen. Ausgehend von diesem Ergebnis werden alternative Gestaltungsmöglichkeiten von Reputationssystemen diskutiert, die eine wahrheitsgemäße Bewertungsabgabe erleichtern sollen.

## 1 Einleitung

Vertrauen ist die Basis oder zumindest Bestandteil vieler Handelsbeziehungen im täglichen Leben. Häufig wird Vertrauen durch die persönliche Einschätzung des Gegenübers gebildet. Im Internet kommt der Reputation eine besondere Bedeutung zu, da im Allgemeinen der Handelspartner nicht persönlich bekannt ist und sich Identitäten im weltweiten Netz nur sehr schwer überprüfen lassen. In der wirtschaftswissenschaftlichen Literatur gibt es für den Begriff Reputation keine allgemeingültige Definition. Im Allgemeinen wird darunter jedoch das wahrgenommene Ansehen einer Person, Organisation oder Institution verstanden. Entscheidend bei der Definition des Begriffes Reputation ist, dass die wahre Qualität des Reputationsobjektes nur unvollständig erfasst werden kann [Sc04]. Unter spieltheoretischen Gesichtspunkten ist die Auswirkung von Reputation gut untersucht. Ohne Reputation ist es für beide Handelspartner die dominante Strategie, sich betrügerisch zu verhalten.

Zum Zweck des Austausches von Informationen zur Reputation existieren im Internet verschiedene Formen von Reputationssystemen. Bekanntestes Beispiel dürfte das Reputationssystem von eBay sein. Diekmann und Wyder zeigen, dass die positive Reputation eine signifikant positive Auswirkung auf den Erfolg eines Verkaufs als solchen hat [DW02]. Weiterhin konnte Ockenfels zeigen, dass bei CD-Verkäufen auf half.com der Verkaufserlös bei Händlern, die mit Reputation ausgestattet waren, im Mittel 20% höher war als bei Verkäufern ohne entsprechende Reputation [Oc02]. Der höhere Preis, den ein Verkäufer mit einem hohen Maß an Reputation erzielen kann, wird als Reputationsrendite bezeichnet.

Bestehende Reputationssysteme weisen jedoch Schwächen auf, die von betrügerischen Händlern ausgenutzt werden. In Foren sind zahlreiche Einträge zu finden, in denen sich Betroffene von Betrugsversuchen oder Opfer von Verkäufern mit unlauteren Praktiken über ihre Erfahrungen austauschen. In diesem Beitrag wird die Existenz eines Problems, nämlich die Abgabe nicht wahrheitsgemäßer Bewertungen, anhand von empirischen Daten des Reputationssystems von eBay überprüft. Nach einer kurzen Beschreibung des Reputationssystems von eBay werden die gesammelten Daten beschrieben und die daraus gewonnenen Erkenntnisse diskutiert. Basierend auf den erörterten und auch empirisch beobachteten Problemen werden alternative Gestaltungsmöglichkeiten von Reputationssystemen diskutiert, die eine wahrheitsgemäße Bewertungsabgabe erleichtern sollen. Abschließend werden die wesentlichen Ergebnisse zusammengefasst.

## **2 Das Reputationssystem von eBay**

Im Internet mit seiner Vielzahl an Käufern und Verkäufern sowie der weltweiten geographischen Verteilung wird die Bildung von Reputation zusätzlich erschwert. Die Nutzer verteilen sich auf eine Vielzahl von Händlern, so dass nur schwer ein verlässliches Gesamtbild über einen Händler entstehen kann. Zudem erschwert die Anonymität der Nutzer den Erfahrungsaustausch. Im Internet besteht also nicht das Problem mangelnder Erfahrung mit einzelnen Akteuren, sondern die Frage, wie die Erfahrungen von Nutzern anderen Nutzern verfügbar gemacht werden.

eBay, nach eigenen Angaben auf 33 internationalen Märkten vertreten und Millionen registrierter Mitglieder, versucht dem durch ein Reputationssystem entgegenzuwirken. Die Mitglieder von eBay haben sich im Bewertungsportal bis September 2006 weltweit mehr als 5 Mrd. Bewertungspunkte nach dem Abschluss von Transaktionen gegeben. Voraussetzung für den Handel auf eBay ist ein Benutzerkonto, das sich mit wenig Aufwand anlegen lässt. Neben persönlichen Daten wie Name, Adresse und Geburtsdatum muss eine gültige Emailadresse angegeben werden. Nutzer müssen außerdem einen eBay-Mitgliedsnamen wählen, unter dem sie auf der Plattform handeln möchten. Händler von eBay Deutschland stimmen mit ihrer Anmeldung einer Überprüfung ihres Namens, der Adresse und ihres Geburtsdatums über die Schutzgemeinschaft für allgemeine Kreditsicherung (SCHUFA) zu.

Der typische Ablauf eines Handels auf eBay beginnt mit der Definition eines Angebotes. Dazu erstellt der Verkäufer unter Angabe von Bildern, Beschreibungstext, Kategorien sowie von der Kategorie abhängigen Standardmerkmalen des Artikels in einer Beschreibungsseite ein Angebot. Neben der Artikelbeschreibung werden ein Mindestpreis (im Falle einer Auktion) oder ein Festpreis sowie das Ende des Auktionszeitraums definiert. Käufer bieten bei Auktionen bis zu einem festgelegten Betrag auf den Artikel bzw. kaufen bei Festpreisgeschäften den gewünschten Artikel direkt. Im Falle einer Auktion erhält der Meistbietende nach Auktionsende den Zuschlag. eBay teilt nach Abschluss des Handels beiden Parteien die jeweiligen Kontaktdaten mit, sodass Lieferung und Bezahlung vereinbart werden können. Danach geben sich die Nutzer im Normalfall gegenseitig eine Bewertung. Abgesehen von der Möglichkeit zur Kommentierung und der einvernehmlichen Rücknahme von Bewertungen haben die beiden Handelspartner keine weiteren gegenseitige Rechte bzw. Verpflichtungen.

Mit dem Abschluss einer Transaktion wird ein Zeitfenster von 90 Tagen für eine Bewertungsabgabe geöffnet. Danach ist es unter nicht genauer spezifizierten Umständen weiterhin möglich eine Bewertung abzugeben. Diese Möglichkeit wird aber nicht garantiert. Die Nutzer geben eine Bewertung in der Form „positiv“, „neutral“ oder „negativ“ ab und können diese durch einen einzeiligen Kommentar ergänzen. Sofort nach Abgabe der Bewertung wird diese für alle Nutzer von eBay sichtbar, einschließlich des Handelspartners. Bewertung und Kommentar sind nach der Abgabe nicht mehr änderbar. Die Bewerteten haben jedoch die Möglichkeit, die Bewertungen, welche über sie abgegeben wurden, zu kommentieren. Weiterhin kann der ursprünglichen Bewertung ein Ergänzungskommentar hinzugefügt werden. Die ursprüngliche Bewertung mit Kommentar bleibt jedoch weiterhin sichtbar. Eine einmal abgegebene Bewertung kann nur dann gelöscht werden, wenn sich die beiden Handelspartner darauf einigen („einvernehmlichen Rücknahme und Entfernung von Bewertungen“). Daraufhin wird die Bewertung gelöscht und der Kommentar um einen Hinweis auf die Rücknahme ergänzt; Bewertungspunkte werden danach weder angezeigt noch gezählt. Jedes Mitglied hat durch das Bewertungssystem die Möglichkeit, den Punktestand eines anderen Mitgliedes (also den Wert, der sich aus der Summe der Einzelbewertungen ergibt) um einen Punkt zu verändern. Dies bedeutet, dass ein Nutzer, auch wenn er mehrere Transaktionen mit einem anderen Nutzer durchgeführt hat, die Bewertungspunkte des Handelspartners nur um eine Einheit erhöhen oder vermindern kann. eBay bildet dazu zunächst die Summe der Einzelbewertungen eines Nutzers. Ist das Ergebnis positiv, so erhält der andere einen Bewertungspunkt, ist das Ergebnis negativ, so wird ein Bewertungspunkt abgezogen. Gleichen sich positive und negative Bewertungspunkte aus, so bleibt der Kontostand unverändert.

Im Frühjahr 2007 hat ebay sein Reputationssystem um „Detaillierte Bewertungen“ erweitert. Dabei können Nutzer zusätzlich Bewertungen zu den folgenden Bereichen abgeben: „Genauigkeit der Artikelbeschreibung“, „Kommunikation“, „Versandzeit“ und „Versand- und Verpackungsgebühren“. Im Unterschied zu den klassischen Bewertungen wird für die detaillierten Bewertungen immer nur ein Durchschnitt angezeigt. Der Verkäufer kann die detaillierte Bewertung daher keiner Transaktion oder einem Käufer zuordnen. Somit erfolgt die detaillierte Bewertungsabgabe durch den Käufer anonym.

### 3 Empirische Untersuchung des Reputationssystems von eBay

Ein Problem des Reputationssystems von eBay, über das in vielen Foren berichtet wird und das auch in der Presse diskutiert wurde<sup>1</sup>, ist die teilweise nicht wahrheitsgemäße Bewertung von Handelspartnern. Ein Grund dafür könnte sein, dass gerade professionelle Händler sehr auf ein möglichst einwandfreies Bewertungsprofil achten und sich sowohl juristischer als auch anderer fragwürdiger Methoden bedienen, um Handelspartner einzuschüchtern. Mehrfach wird in Foren berichtet, dass Käufer von Händlern unter Druck gesetzt werden, trotz eigentlich negativer Erfahrung eine positive Bewertung abzugeben oder ihre negative Bewertung zurückzuziehen. Exemplarisch seien hierfür die beiden folgenden Kommentare genannt:

*„(...)Nachdem ich Ihm eine schlechte Bewertung gegeben habe, wurde der Ersatz geliefert, eben falls erhielt ich eine Mail, das ich die Bewertung innerhalb von 5 Tagen zurückziehen soll (passender Link war eingefügt) ansonsten droht eine Klage. Wie sollte man sich jetzt verhalten?(...)“<sup>2</sup>*

*„(...) 5. wenn seine Überredungskünste und zugeschickten Patronen nicht helfen, droht er mit Rachebewertungen (...)“<sup>3</sup>*

Vermutlich lassen sich einige Käufer dadurch tatsächlich einschüchtern. Als häufigste Gründe werden dabei die Androhung juristischer Schritte bzw. die Abmahnung durch einen Anwalt und die Androhung einer negativen Gegenbewertung genannt. Die Regeln des Bewertungssystems von eBay sagen hierzu allerdings eindeutig, dass es verboten ist, Mitglieder mit einer Negativbewertung oder anderweitig zu bedrohen, damit sie eine positive Bewertung abgeben<sup>4</sup>. Daraus lässt sich schließen, dass sich eBay dieser Möglichkeit der Beeinflussung des Reputationssystems durchaus bewusst ist. Dellarocas stellte bereits die Frage, warum der Anteil negativen Feedbacks auf eBay derart gering ist [De03]. Er bezieht sich dabei auf die Daten von Resnick und Zeckhauser, die bei Ihrer Analyse auf 99,1% positive, 0,6% negative und 0,3% neutrale Bewertungen kommen [RZ02]. Dellarocas sieht dafür zwei mögliche Erklärungsansätze. Entweder ist der Mechanismus so erfolgreich, dass sich tatsächlich der Großteil der Teilnehmer ehrlich verhält und es so kaum enttäuschte Nutzer gibt oder aber der Mechanismus ist nicht erfolgreich, aber die Nutzer sind nicht bereit ihre wahre Meinung preiszugeben.

#### 3.1 Ziel der Untersuchung

Ziel der Analyse von Bewertungskommentaren des eBay-Reputationssystems ist herauszufinden, ob es Anzeichen dafür gibt, dass teilweise positive Bewertungen bei eBay erzwungen wurden und nicht etwa aufgrund positiver Erfahrungen zustande kamen.

---

<sup>1</sup> Misstrauen, Angst, Unsicherheit – Spiegel-Online von 18. Mai 2007

<sup>2</sup> eBay-Forum, 4.10.06 (<http://forums.ebay.de/thread.jspa?messageID=1300626892&tstart=0>)

<sup>3</sup> eBay-Blog, 29.08.07 ([http://blogs.ebay.de/d.susi/entry/Patronen-Power/\\_W0QQidZ317161010](http://blogs.ebay.de/d.susi/entry/Patronen-Power/_W0QQidZ317161010))

<sup>4</sup> <http://pages.ebay.de/help/feedback/reputation-policies.html>

Da sich eine negative Bewertung deutlich von der Masse der Bewertungen abhebt, wird diese von einem Großteil der eBay-Nutzer vermutlich nur im Extremfall vergeben. Ebenso ist davon auszugehen, dass einige Teilnehmer versuchen ihre Handelspartner dazu zu bewegen eine Bewertung zu ihren Gunsten abzugeben, obwohl dazu objektiv kein Grund besteht. Dies würde zu ungerechtfertigten positiven Bewertungen führen. Im Folgenden wird untersucht, ob positive Bewertungen gefunden werden können, die im Bewertungskommentar eine eigentlich negative Erfahrung beschreiben und möglicherweise gar kritische Äußerungen von Nutzern enthalten. Anzumerken ist hierbei, dass viele Nutzer in einer Situation, in der sie negativ bewerten würden aber unter Druck gesetzt werden, vermutlich ganz auf eine Bewertungsabgabe verzichten. Die von Resnick und Zeckhauser berechneten Bewertungsraten von 52,1% (Käufer bewerten Verkäufer) und 60,6% (Verkäufer bewerten Käufer) lassen diese Vermutung durchaus zu [RZ02].

### **3.2 Datensammlung**

Um einige der Bewertungskommentare von eBay zu untersuchen wurden im April 2007 Bewertungskommentare mit Hilfe eines Webcrawlers von den eBay-Webseiten ausgelesen. In zwei Nächten wurden im Rahmen von zwei voneinander unabhängigen Läufen so einmal 140 MB und 113 MB an Daten gesammelt, was im ersten Lauf 1.750.133 Bewertungen und im zweiten Lauf 1.463.552 Bewertungen entspricht. Das Vorgehen war dabei wie folgt: Ausgehend von einem vorgegebenen Nutzer wurden im ersten Lauf alle, im zweiten Lauf jeweils nur die max. 200 Bewertungen der ersten Bewertungsseite dieses Nutzers besucht, die dort enthaltenen Bewertungen extrahiert und in eine Textdatei gespeichert. Für jede Bewertung wurden dabei die Nutzernamen von Bewertetem und Bewertendem, der Nutzertyp (Käufer, Verkäufer), die Bewertung (positiv, neutral, negativ) sowie der Bewertungskommentar (Freitext) gespeichert. Von den gesammelten Bewertungen enthielten 224 bzw. 271 Einträge keine Information über die Bewertung (u.a. auch weil diese von eBay entfernt wurde). Diese Einträge wurden von der weiteren Betrachtung ausgeschlossen.

Der Crawler führte intern je eine Liste mit Benutzernamen, die er bereits besucht hatte, und eine mit Benutzernamen, die noch für einen Besuch vorgesehen waren. So konnte sichergestellt werden, dass die Bewertungen eines Nutzers nie doppelt erfasst wurden. Sobald alle Bewertungsseiten eines Nutzers analysiert waren, fuhr der Crawler mit dem nächsten Namen in der Liste der zu besuchenden Nutzer fort. Die Liste wurde jeweils um die Nutzernamen der gegenwärtig analysierten Seite ergänzt, sofern sie nicht schon auf einer der Listen standen. Nach dem Ende einer Nacht wurde der Crawler manuell angehalten. Die Vorgehensweise der Datensammlung erzeugt damit einen Baum von Nutzern. Die beiden Nutzer der Anfangsknoten wurden zufällig gewählt, jedoch unter Beachtung der folgenden Vorgaben. Für den ersten Lauf wurde ein Nutzer mit einer hohen Anzahl an Bewertungen (> 10.000) gesucht, um gleich zu Beginn einen möglichst großen Stamm an Nutzern zu erhalten. Der erste Nutzer, der bei einer zufällig durchgeführten Suche den genannten Kriterien entsprach, war ein Händler für Outdoorartikel mit rund 10.636 Bewertungen (davon 99,9 % positiv) von 8.816 Nutzern.

Für den zweiten Lauf wurde ein professioneller Verkäufer, der mehrere Artikel im Bereich „Computer“ eingestellt hatte und eine sehr hohe Anzahl an Bewertungen besaß (rund 35.000, davon 98,5% positiv), gewählt. Grund hierfür war vor allem, dass der Handel mit Computern allgemein als Bereich gilt, in dem es häufiger zu Problemfällen kommt als in anderen Produktkategorien. Demzufolge wurde angenommen, dass hier auch eine potentiell größere Menge an unzufriedenen Nutzern existiert. Weiterhin agieren in diesem Umfeld eine große Menge kleiner, aber durchaus professioneller Händler, die mit gebrauchten Computern Handel betreiben. Dieser Gruppe wurde wegen des starken Wettbewerbes eine höhere Bereitschaft zugetraut, ihre Reputation mit zweifelhaften Methoden positiv zu beeinflussen. Durch die Wahl eines Händlers im Bereich „Computer“ als Anfangsknoten kann allerdings nur die Wahrscheinlichkeit dafür, dass die Bewertungen aus dem Umfeld des Computerhandels kommen, erhöht werden. Da die einzelnen Nutzer, die anschließend vom Crawler besucht wurden, wurden nämlich nicht auf ihr Tätigkeitsgebiet hin untersucht.

Im ersten Lauf wurden jeweils alle Bewertungsseiten der Nutzer besucht und diese gespeichert. Für den zweiten Lauf wurde dieses Vorgehen geändert und die Erfassung von Bewertungen aus zwei Gründen auf die jeweils erste Bewertungsseite beschränkt. Erstens sollten möglichst aktuelle Daten gesammelt werden. Da eBay die Bewertungen antichronologisch darstellt, wurden somit jeweils die letzten max. 200 Bewertungen gesammelt. Zweitens sollte so der Einfluss des ersten Knotens reduziert werden. Während im ersten Lauf die Bewertungen des ersten Nutzers 0,61% der Gesamtzahl der Bewertungen ausmachen, sind es im zweiten Lauf nur 0,01%. Außerdem konnte auf diese Weise die Anzahl der Nutzer fast verdoppelt werden und auch die Tiefe des Baumes wird entsprechend größer. Beim ersten Lauf besaß der Ursprungsknoten Bewertungen von 8.816 Nutzern; insgesamt wurden aber nur Bewertungen von 5.312 Nutzern gesammelt. Dies bedeutet, dass alle Nutzer direkte Kunden des ersten Knoten waren. Der zweite Datensatz unterscheidet sich in diesem Punkt deutlich. Die ersten 200 Bewertungen stammten von 161 Nutzern. Insgesamt wurden 9.499 Nutzerprofile besucht.

### **3.3 Auswertung**

Eine Übersicht über die gesammelten Daten ist in Abbildung 1 zu finden. Da im ersten Datensatz das Profil des Anfangsknotens eine hohe Anzahl an Bewertungen besaß (10636), aber nur Bewertungen von 5312 Nutzern gesammelt wurden, waren alle besuchten Profile von Nutzern, die direkte Kunden des Anfangsknotens sind. Da diese Kunden vermutlich überwiegend Konsumenten sind und diese daher auch überwiegend als Käufer auftreten, würde dies den hohen Anteil von über 61% der Bewertungen im Datensatz erklären, die von Verkäufern für Käufer abgegeben wurden. Umgekehrt ist die Situation im zweiten Datensatz, in dem knapp zwei Drittel der Bewertungen von Käufern stammen. Die besuchten Nutzerprofile waren somit überwiegend solche von Händlern oder Privatpersonen mit hoher Verkaufsaktivität. Die Annahme, dass es im Bereich des Computerhandels eher zu Schwierigkeiten kommt, erfährt durch den Vergleich der beiden Datensätze eine Unterstützung.

Bei Betrachtung der negativen und neutralen Bewertungen zeigt sich im zweiten Datensatz ein deutlich erhöhtes Konfliktpotential. Betrachtet man nur die negativen Bewertungen, die von Käufern abgegeben wurden, ist deren Anteil im zweiten Datensatz mehr als dreimal so hoch. Auch die deutlich höhere Anzahl entfernter Bewertungen könnte ein Zeichen dafür sein, dass Bemühungen bestehen, sich meist schlechter Bewertungen zu entledigen. Allerdings muss man hierzu anmerken, dass in den Datensätzen keinerlei Informationen über die gehandelten Artikel vorhanden sind. Die Wahl eines PC-Händlers als Anfangsknoten erhöht damit lediglich die Wahrscheinlichkeit, dass Daten in diesem Umfeld gesammelt worden sind.

	Datensatz 1		Datensatz 2	
Anzahl Bewertungen Startknoten	10636		200	
Nutzerprofile (Reputationsobjekte)	5312		9499	
Bewertende Nutzer	1320902		1234725	
<b>Handelspartner je Nutzer</b>				
Min	1		1	
Max	29263		201	
<b>erhaltene Bewertungen je Nutzer</b>				
Min	1		1	
Max	44114		200	
Durchschnitt	329,43		154,05	
<b>abgegebene Bewertung je Nutzer</b>				
Durchschnitt	1,32		1,19	
<b>Bewertungen</b>				
<b>Gesamtzahl abgegeben</b>	1749909	100,000%	1463281	100,000%
davon positiv	1741453	99,517%	1443054	98,618%
davon neutral	4224	0,241%	8955	0,612%
davon negativ	4232	0,242%	11272	0,770%
von Verkäufern	1076853	61,538%	493270	33,710%
von Käufern	645835	36,907%	965503	65,982%
Unbekannt	27221	1,556%	4508	0,308%
<b>abgegeben von Verkäufern</b>				
	1076853	100,000%	493270	100,000%
davon positiv	1073361	99,676%	490650	99,469%
davon neutral	1258	0,117%	748	0,152%
davon negativ	2234	0,207%	1872	0,380%
<b>abgegeben von Käufern</b>				
	645835	100,000%	965503	100,000%
davon positiv	641203	99,283%	948020	98,189%
davon neutral	2798	0,433%	8149	0,844%
davon negativ	1834	0,284%	9334	0,967%
<b>von eBay entfernt</b>	39	0,002%	108	0,007%

Abbildung 1: Struktur der bei eBay gesammelten Daten

Aufgrund der großen Datenmenge mussten die gesammelten Daten automatisch bzw. mit geringerem manuellen Aufwand analysiert werden. Grundidee war, die gesammelten Bewertungskommentare nach Schlagwörtern zu durchsuchen, die Hinweise auf eine Kritik bzw. negative Bewertung geben. Die ersten Versuche dieser Form der Auswertung waren einfache Abfragen nach Wörtern, die eine solche negative Aussage andeuten. Allerdings ist die Suche nach einzelnen Wörtern höchst unscharf.

Ein gutes Beispiel hierfür ist das Wort „nicht“. Die Verneinung lässt zuerst vermuten, dass etwas nicht so ist wie der Nutzer es erwartet hätte. Korrekte Treffer sind bspw. „nicht so schön“ oder „nicht wie beschrieben“. Problematisch ist jedoch, dass Verneinungen auch zum Ausdruck von positiven Meinungen verwendet werden. Die Verwendung von „nicht“ mit positiver Aussage tritt besonders häufig in den Formen „nicht schlecht“ und „besser geht es nicht“ auf. Neben dieser grundlegenden Herausforderung, korrekte Treffer zu finden, machen die Verwendung von Verzierungen durch Satz- und Sonderzeichen, dialektische Schreibweisen, orthographische Fehler sowie fehlende Leerzeichen bei manchen Kommentaren die automatische Erkennung von Worten nahezu unmöglich. Zumindest das Problem der fehlenden Leerzeichen lässt sich mit Regular Expressions lösen. Die entsprechende Regular Expression für das Wort „leider“ lautet somit: `*([\^a-z^A-Z]\|A)+leider([\^a-z^A-Z]\|Z)+.*` Mit dieser Regular Expression werden nun zwar Kombinationen der Form „kaputt,leider keine Reaktion“ gefunden, Wörter mit anderer Bedeutung wie „Kleider“ jedoch nicht.

Somit stellt sich die Frage, wie aus allen Kommentaren die gesuchten Kommentare automatisiert herausgefiltert werden können. Um die Hypothese bestätigen zu können, dass es eine größere Anzahl positiver Bewertungen gibt, die aufgrund des Bewertungskommentars eigentlich negativ sein müssten, muss eine ausreichend große Anzahl dieser Kommentare gefunden werden. Unsere Auswertung erfolgte auf Basis von Worten, die häufig in den Bewertungskommentaren vorkommen. Aus diesem Grund wurde vor der eigentlichen Analyse der Kommentare eine Worthäufigkeitsanalyse durchgeführt. Für die Worthäufigkeitsanalyse wurden die Kommentare zuerst in Kleinschreibung umgewandelt und danach Sonderzeichen durch Leerzeichen ersetzt. Die Kommentare wurden anschließend anhand der Leerzeichen in einzelne Worte unterteilt.

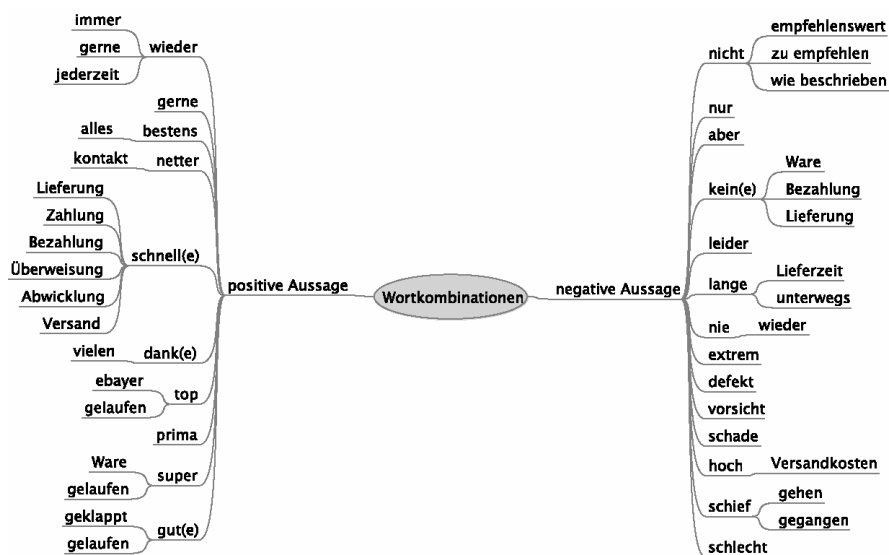


Abbildung 2: Karte möglicher Wortkombinationen

Die Top 20 der häufigsten Wörter der beiden Datensätze enthalten jedoch keinerlei Worte, mit denen eine Kritik verbunden werden kann. Worte, die Hinweise auf eine kritische Äußerung geben, wurden deshalb manuell aus der Gesamtliste der Worte, die mindestens 100-mal in den Kommentaren vorkommen (im ersten Datensatz sind dies 1.714, im zweiten 1.938 Worte), ausgewählt. Zu den negativ besetzten Worten wurden zusätzlich anhand der Worthäufigkeitsliste Wörter gesucht, die zum einen häufig in den Kommentaren vorkommen, zum anderen in Zusammenhang mit dem jeweiligen Wort gebracht werden können. Dasselbe wurde auch mit Worten deutlich positiver Aussage vorgenommen. Die daraus entstandene Karte zeigt Abbildung 2. Um die Anzahl an kritischen Kommentaren quantifizieren zu können, wurden Abfragen mit den gefundenen Schlagwörtern bzw. Schlagwortkombinationen gegen die Datenbank durchgeführt und die Anzahl der treffenden Bewertungskommentare gezählt.

Für diese Datenbankabfragen wurden die Schlagwörter in zwei Gruppen eingeteilt, 'GO' und 'STOP'-Wörter. 'GO' Wörter sind dabei solche, die geeignet sind, kritische Kommentare zu identifizieren. Diese müssen im Kommentar vorhanden sein, damit dieser als potentiell kritisch gewertet wird. 'STOP'-Wörter im Gegensatz sind solche, die eindeutig eine positive Aussage ausdrücken und in Negativaussagen nicht anzutreffen sind. Diese dürfen in Kommentaren, die als kritisch gewertet werden, nicht vorkommen. Für die Auswertung wurden Wortlisten bestehend aus Regular Expressions gebildet (s. Abbildung 3, analog dazu 'STOP'-Wörter). Wie schon bei der Bestimmung der Worthäufigkeiten wurden nur durch Satz-, Sonder- und Leerzeichen getrennte Wörter zugelassen. Direkt zusammenhängende Worte wie „SuperToll“ wurden als ein Wort gezählt und wurden weder auf der Suche nach „Super“ noch „Toll“ als Treffer gewertet. Groß- und Kleinschreibung wurde ebenfalls nicht beachtet. Die in den Regular Expressions vorkommenden Elemente für einzelne Wörter („\*<sup>^</sup>[a-zA-Z]+“) sowie für Wortgruppen („(\\W)\*“) sorgen dafür, dass im ersten Fall keine Buchstaben direkt vor bzw. nach dem Wort auftreten. Danach dürfen beliebige Zeichen folgen. Somit werden Worte nur freistehend und nicht als Bestandteil eines anderen Wortes erkannt. Die zweite Regular Expression erlaubt zwischen den beiden Worten des Wortpaares nur Leer-, Satz- oder Sonderzeichen.

```
String [] goWords = {
    ".*^[a-zA-Z]+defekt^[a-zA-Z]+.*",
    ".*^[a-zA-Z]+vorsicht^[a-zA-Z]+.*",
    ".*^[a-zA-Z]+schade^[a-zA-Z]+.*",
    ".*^[a-zA-Z]+schief^[a-zA-Z]+.*",
    ".*^[a-zA-Z]+schlecht^[a-zA-Z]+.*",
    ".*nicht(\\W)empfehlenswert.*",
    ".*nicht(\\W)zu(\\W)empfehlen.*",
    ".*nicht(\\W)wie(\\W)beschrieben.*",
    ".*aber(\\W)nicht.*",
    ".*kein(e)?(\\W)Ware.*",
    ".*kein(e)?(\\W)Bezahlung.*",
    ".*kein(e)?(\\W)Lieferung.*",
    ".*lange(\\W)Lieferzeit.*",
    ".*hohe(\\W)Versandkosten.*",
    ".*leider(\\W)nur.*",
    ".*nie(\\W)wieder.*"
};
```

Abbildung 3: „GO“-Wörter

### **3.4 Ergebnis**

Mit dem beschriebenen Verfahren konnten rund 335 negative Kommentare im ersten sowie 1454 im zweiten Datensatz identifiziert werden, die mit einer positiven Bewertung verbunden waren. Dabei wurden ausschließlich Kommentare von Käufern, die positiv bewertet hatten, ausgewertet. Die Einschränkung auf Käufer erschien sinnvoll, da eher davon auszugehen ist, dass Verkäufer Druck auf Käufer ausüben als umgekehrt. Diese Vermutung wird durch die gesammelten Daten bestätigt, da unter allen positiven Kommentaren, also sowohl von Käufern als auch Verkäufern, 420 im ersten Datensatz und 1504 im zweiten Datensatz mit den Filterlisten selektiert wurden. Die überwiegende Mehrheit der selektierten Kommentare stammte also von Käufern.

Bei manueller Überprüfung der gefunden Bewertungskommentare finden sich in den Ergebnissen immer noch vereinzelt positive Kommentare. Ziel dieser Untersuchung war es allerdings nicht, eine exakte Anzahl an potentiell falschen Bewertungen im Reputationssystem von eBay zu bestimmen. Ziel war vielmehr festzustellen, ob solche Kommentare überhaupt in größerer Anzahl vorhanden sind. Die 1454 gefundenen Bewertungen entsprechen immerhin knapp 15% der Anzahl an negativen Bewertungen von Käufern im zweiten Datensatz sowie die 335 Bewertungen gut 18% der Anzahl negativer Bewertungen von Käufern aus dem ersten Datensatz. Die Anzahl der negativen Bewertungen ist also im Vergleich zu einer wahrheitsgemäßen Bewertung deutlich zu gering.

Natürlich ist die Anzahl der vermutlich nicht wahrheitsgemäßen Bewertungen im Vergleich zur überwältigenden Anzahl an positiven Bewertungen recht gering. Das ist allerdings nicht überraschend. Andernfalls würde dies nämlich bedeuten, dass das Reputationssystem von eBay nicht funktioniert, was im Gegensatz zu anderen Untersuchungen der Auktionsplattform und deren Erfolg steht. Dennoch kann auf Basis der durchgeführten Untersuchung festgehalten werden, dass es im Falle des Reputationssystems von eBay Anzeichen darauf gibt, dass Nutzer nicht immer entsprechend ihrer wahren Meinung bewerten.

## **4 Reputationssysteme und soziale Netzwerke**

Die empirische Untersuchung liefert Hinweise darauf, dass auf eBay positive Bewertungen abgegeben werden, obwohl dazu kein Grund besteht. Nun stellt sich die Frage, ob und wie dieses Problem durch ein verbessertes Reputationssystem gelöst werden kann. Hogg und Adamic schlagen vor, Reputationssysteme durch die Berücksichtigung sozialer Netzwerke zu verbessern [HA04]. Unter sozialen Netzwerken verstehen sie Social Network Sites, bei denen die Verbindungen zwischen Nutzern manuell erstellt werden. Das Fälschen von Identitäten wird dadurch deutlich erschwert. Ähnlich dem Google PageRank könnte außerdem die Reputation einer Person aufgrund der Struktur und Anzahl der Verbindungen im sozialen Netzwerk festgestellt werden.

Außerdem könnte das soziale Netzwerk als Filtermechanismus für abgegebene Bewertungen anderer Nutzer zum Einsatz kommen. Beispielsweise indem die Beschränkung bei eBay, dass ein Nutzer die Reputation eines anderen nur einmalig beeinflussen kann auf das direkte soziale Umfeld dieser Nutzers erweitert wird. Bewertungen durch Personen im gleichen sozialen Umfeld würden somit zusammengefasst und damit schwächer gewichtet, Kollusion dadurch erschwert.

Sabater und Sierra beschreiben das Reputationssystem „ReGreT“, das auf sozialen Verbindungen basiert [SS02]. Die Reputation nach diesem System setzt sich aus mehreren Dimensionen zusammen, so dass die direkten Erfahrungen mit dem Reputationsobjekt sowie die Informationen, die im sozialen Netz vorhanden ist, berücksichtigt werden. Problematisch ist jedoch, dass diese Methode relativ komplex ist und somit fraglich ist, ob auf Nutzerseite nicht Vorbehalte gegen eine Reputation aus einer „Black Box“ entstehen würden.

Sowohl Hogg und Adamic als auch Sabater und Sierra nutzen also soziale Netzwerke, um mit ihren Ansätzen die Funktion von Reputationssystemen zu verbessern. Aus Nutzersicht wäre es jedoch vermutlich wünschenswert, die Reputation ausschließlich auf Basis der von den Nutzern abgegebenen Bewertungen zu berechnen. Das soziale Netzwerk könnte dann zur Filterung von Bewertungen genutzt werden. Grundsätzliche Idee unseres erweiterten Systems ist also primär nicht die Berechnung eines alternativen Reputationswertes, sondern eine Erweiterung der Mechanismen zur Informationsgewinnung und -verteilung.

Entsprechend der Unterscheidung in nahe und ferne Kontakte in einem sozialen Netzwerk existieren im erweiterten Reputationssystem auch zwei Arten von Bewertungen, nämlich öffentliche und private. Jeder Nutzer kann, je nachdem mit wem er handelt, bis zu zwei Bewertungen abgeben. Öffentliche Bewertungen sind solche Bewertungen, die von allen Nutzern eingesehen werden können. In den gängigen Reputationssystemen wie bspw. auf eBay kann jeder Nutzer die Bewertungen einsehen. Private Bewertungen dagegen sind nur den nahen Kontakten des die Bewertung abgebenden Nutzers zugänglich. Die bewertete Person selbst kann die Bewertung nicht einsehen und folglich auch nicht entfernen lassen oder mit Rachebewertungen drohen. Im Fall von privaten Bewertungen werden vermutlich mehr und genauere Informationen ausgetauscht, da es sich um privat kommunizierte Informationen handelt. Man wird bemüht sein, seinem Bekannten möglichst nicht zu schaden und ihm daher bestmögliche Informationen bieten. Zur Verminderung der Gefahr des Erschleichens einer guten Reputation durch Kollusion wird die Abgabe öffentlicher Bewertungen beschränkt. So ist es nur möglich, dass Personen, die zueinander keine nahen Kontakte sind, sich gegenseitig öffentlich bewerten. Zudem besteht bei den privaten Bewertungen kein Grund mehr, von einer wahrheitsgemäßen Bewertung des Handelspartners abzuweichen. Insofern könnte ein derart erweitertes Reputationssystem geeignet sein, das in den vorangehenden Abschnitten beschriebene Problem des bisherigen eBay-Reputationssystems zu lösen.

## 5 Zusammenfassung und Ausblick

In diesem Beitrag wurde untersucht, inwieweit sich das eBay-Reputationssystem dafür eignet, möglichst ehrliches Feedback der Nutzer zu einer Handelstransaktion zu erhalten. Dazu wurde in den Bewertungskommentaren positiver Bewertungen nach Hinweisen auf kritische Äußerungen gesucht. Die so gefundenen Bewertungen zeigen klar, dass es durchaus positive Bewertungen gibt, obwohl die Nutzer im Bewertungstext eher eine negative Bewertung abgeben.

Die Gründe hierfür konnten in dieser Untersuchung nicht erschlossen werden. Es ist aber davon auszugehen, dass Nutzer bspw. negative „Rachebewertungen“ befürchten. Der Umfang der gefundenen Bewertungen (15 bzw. 18% zusätzlich zu den negativen Bewertungen) lässt dabei keinesfalls den Schluss zu, dass bisherige Reputationssysteme grundsätzlich in Frage gestellt werden müssen. Dennoch sind die Ergebnisse ein Zeichen dafür, dass das Problem der ehrlichen Abgabe von Bewertungen durchaus existiert. Das vorgeschlagene erweiterte Reputationssystem setzt soziale Netzwerke ein, um Bewertungen an verschiedene Zielgruppen abzugeben. Dadurch lassen sich vermutlich wertvollere und ehrlichere Bewertungen von den Handelsteilnehmern gewinnen.

Im nächsten Schritt soll nun ein derart erweitertes Reputationssystem realisiert und auf einer Handelsplattform eingesetzt werden, um das tatsächliche Verhalten der Nutzer empirisch untersuchen zu können.

## Literaturverzeichnis

- [De03] Dellarocas, C. N.: The Digitization of Word of Mouth: Promise and Challenges of Online Reputation Mechanisms. *Management Science* 49 (10), 2003; S. 1407-1424.
- [DW02] Diekmann, A.; Wyder, D.: Vertrauen und Reputationseffekte bei Internetauktionen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54 (4), 2002; S. 674-693.
- [HA04] Hogg, T.; Adamic, L.: Enhancing Reputation Mechanisms via Online Social Networks. In: *Conference on Electronic Commerce (EC'04)*, New York, 2004.
- [Oc02] Ockenfels, A.: Reputationsmechanismen auf Internet-Marktplattformen. In: *III. Symposium zur ökonomischen Analyse der Unternehmung*, 2002.
- [RZ02] Resnick, P.; Zeckhauser, R.: Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. *The Economics of the Internet and E-Commerce* 11, 2002; S. 127-157.
- [Sc04] Schwalbach, J.: Reputation. *Forschungsbericht 2004-2*, Humboldt-Universität Berlin, Institut für Management, 2004.
- [SS02] Sabater, J.; Sierra, C.: Social ReGreT, a reputation model based on social relations. *SIGecom Exch.* 3 (1), 2002; S. 44-56.