

Funktionale Gestaltungsoptionen von Online-Bewertungssystemen

Jörg Becker¹, Sebastian Herwig¹, Jens Pöppelbuß¹, Daniel Tiebe¹, Axel Winkelmann²

1: European Research Center for Information Systems (ERCIS)

Universität Münster

Leonardo-Campus 3

D-48149 Münster

joerg.becker@ercis.uni-muenster.de

sebastian.herwig@ercis.uni-muenster.de

jens.poeppelbuss@ercis.uni-muenster.de

daniel.tiebe@ercis.uni-muenster.de

2: Institut für Wirtschafts- und Verwaltungsinformatik

Universität Koblenz

Universitätsstrasse 1

D-56070 Koblenz

winkelmann@uni-koblenz.de

Abstract: Bewertungen sind ein einflussreiches Medium, um Vertrauen zwischen Anspruchsgruppen zu schaffen. Dies gilt nicht nur für die physische, sondern auch für die virtuelle Welt. Seit Beginn der Internet-Ära in den 1990er-Jahren entstanden unzählige Bewertungssysteme. Als einer der ersten Anbieter etablierte bspw. Amazon.com 1995 ein Bewertungssystem für Produkte. Derzeit erfahren Bewertungssysteme eine weitreichende Verbreitung, bspw. in Onlineshops, sozialen Netzwerken und auf Informations- oder Empfehlungsportalen. In den meisten Fällen generieren die Nutzer den Inhalt dieser Systeme selbst, der aggregiert entsprechenden Empfängern zur Verfügung gestellt wird. Die zunehmende Relevanz dieser Bewertungssysteme führt zu einem neuen Forschungsfeld in Bezug auf Gestaltung, Effekte und Validität derartiger Systeme. Durch eine Erhebung des Status-Quo existierender Bewertungssysteme trägt dieser Artikel hierzu bei. Mithilfe von 237 Kriterien werden 102 Systeme insbesondere hinsichtlich funktionaler Entwurfsentscheidungen analysiert. Der Beitrag präsentiert Auszüge aus den Erkenntnissen sowie eine Systematisierung funktionaler Entwurfsentscheidungen von Bewertungssystemen.

1 Einleitung

Bewertungen werden insbesondere für den Aufbau von Vertrauen zwischen Geschäftspartnern geschätzt. In der realen Welt erfolgen Bewertungen zumeist traditionell mündlich und informal über Einschätzungen oder Gerüchte [Del01, CLS07]. Darüber hinaus entwickelten sich mit Aufkommen des Internets zunehmend Online-Bewertungssysteme. So kamen [AH95] zu dem Ergebnis, dass Online-Bewertungssysteme potenziellen Kunden dabei helfen, Online-Transaktionen durchzuführen, indem sie auf durch Bewertungen ausgesprochene Empfehlung anderer Kunden vertrauen. Darüber entdeckten sie, dass Online-Bewertungssysteme zu einer verbesserten Kundenbindung führen können [AH95]. Amazon.com hat bspw. bereits seit 1995 Kunden um Produktbewertungen gebeten. Viele andere Onlineshops haben ebenfalls Bewertungsfunktionalität implementiert. Darüber hinaus existieren auch reine Bewertungsportale, deren Hauptfunktion die Bewertung von Personen, Produkten oder Dienstleistungen ist. Das intensive Teilen von Empfehlungen, Reviews und Bewertungen trägt heute als soziale Erfahrung zum digitalen Shopping bei [VH07].

Der Vorteil von Online-Bewertungssystemen ist in der Skalier- und Formalisierbarkeit des Bewertungsverfahrens zu sehen [Del03, BKO04]. Skalierbarkeit bedeutet, dass Bewertungen unabhängig von Zeit und Ort von einer Vielzahl von Anspruchsgruppen erreicht und für diese aufbereitet werden können [RKZ00]. [CP02, TSW08] sehen insbesondere in der schnellen Aktualisierung von Bewertungen einen großen Vorteil. Das Zusammenspiel von Erhebung, Aggregation und Präsentation von Bewertungen gestaltet sich hierdurch nachvollziehbarer und die Akzeptanz von Bewertungen wird dadurch insgesamt gestärkt [RZ01].

Bewertungsfunktionalitäten sind in vielfältiger Form im Internet zu finden, der hauptsächlich Einsatzbereich ist jedoch im Bereich des E-Commerce zu sehen. Eines der wesentlichen Probleme bei Online-Transaktionen zwischen anonymen Partnern ist der Mangel an Vertrauen. Kunden haben oft Probleme, die Qualität und Zuverlässigkeit von Angeboten und Geschäftspartnern einzuschätzen [WL99]. Bewertungssysteme haben sich daher zu einem integralen Bestandteil von E-Commerce-Plattformen entwickelt und tragen insbesondere zum Aufbau von Vertrauen zwischen Transaktionspartnern bei. In Bezug auf Ebay.com haben Analysen ergeben, dass ein Zusammenhang zwischen dem Bewertungsprofil eines Anbieters und dem Erfolg seiner Transaktionen sowie dem Verkaufspreis besteht [AH95, Seb08]. Darüber hinaus können Bewertungsfunktionen auch in anderen Bereichen wie sozialen Netzwerken Anwendung finden. Ein Beispiel hierfür findet sich bei LinkedIn.com, wo Empfehlungen für verschiedene Personenkreise abgegeben werden können. Diese Empfehlungen beziehen sich auf bestimmte Fähigkeiten und werden entsprechend in den Nutzerprofilen aufgelistet.

Die zunehmende Relevanz und Verbreitung von Bewertungssystemen schafft einen neuen Forschungsbereich. [PR08] konstatiert einen zunehmenden Forschungsbedarf in Bezug auf Untersuchungen zu Ausprägungen, Effekten und zur Validität von Bewertungssystemen im Internet. Relevante Fragestellungen sind bspw. die sinnvolle Aggregation von Bewertungen und die Repräsentation valider Bewertungsergebnisse. Obwohl die Gestaltung von Bewertungsfunktionalitäten in starkem Maße die Ergebnisse beein-

flusst, fehlen Studien hierüber. Um diesem Umstand zumindest ein Stück weit Rechnung zu tragen, wird in dem vorliegenden Artikel der Status-Quo bei der Ausgestaltung bestehender Bewertungssysteme erhoben, analysiert und darauf aufbauend eine Systematisierung grundlegender Gestaltungsoptionen von Bewertungssystemen abgeleitet. Hierzu werden in Kapitel 2 theoretische Grundlagen zu Bewertungssystemen erörtert. In einem zweiten Schritt werden die der Arbeit zugrundeliegende Forschungsmethode sowie das Vorgehen bei der Analyse dargelegt. Auszüge der Ergebnisse werden in Kapitel 4 dargestellt und diskutiert. Eine Systematisierung der identifizierten Gestaltungsoptionen wird in Kapitel 5 vorgestellt. Der Beitrag schließt mit einer Darlegung von Einschränkungen der Analyse und einer Diskussion des zukünftigen Forschungsbedarfs (Kapitel 6).

2 Grundlagen

Die grundlegende Idee von Online-Bewertungssystemen ist die Bewertung von Entitäten durch Nutzer von Webanwendungen und somit das Sammeln, Aggregieren und Aufbereiten von Bewertungen [RKZ00]. Die aggregierten Bewertungen können dazu genutzt werden, einen Score oder eine Vertrauenseinschätzung zu generieren, die Nutzern zur Verfügung steht und einen Beitrag zur Entscheidungsfindung über die Durchführung von Transaktionen leistet [JRB07]. Typischerweise existiert eine zentrale Instanz zur Sammlung und Aufbereitung von Bewertungen.

Ein Bewertungssystem kann als eine Abbildung von Einschätzungen verstanden werden, inwieweit eine bestimmte Entität nutzbringend für eine verfolgte Zielstellung ist [MC01]. Die Kriterien für diese Bewertung sind im Regelfall durch das Ziel der Bewertung determiniert, wohingegen die den Kriterien zugeordneten Werte durch die Entität bestimmt werden [Sch04]. Eine Bewertung findet für gewöhnlich subjektiv statt und hängt von dem bewertenden Nutzer sowie dessen Einstellungen und Erfahrungen ab. Sie ist daher nur eine situations- und zeitbezogene Einzelmeinung [SM07].

Die sinnvolle Verwendung von Online-Bewertungssystemen setzt eine adäquate Ausgestaltung der darunter liegenden Funktionen voraus. [CHW08] gehen davon aus, dass eine der bedeutendsten Aspekte der Ausgestaltung von Bewertungssystemen die Informationssammlung. Betreiber eines Dienstes müssen auch festlegen, welche Nutzer welche Entitäten bewerten können. Dabei ist insbesondere eine Vermeidung von Manipulationen anzustreben [Del03]. Um die Privatsphäre zu respektieren, sollten Nutzer darüber hinaus explizit der Möglichkeit der eigenen Bewertung zustimmen (Opt-In-Verfahren), bevor sie durch andere Benutzer bewertet werden können [Zie08].

3 Forschungsmethode

Das zentrale Forschungsziel der hier vorgestellten Untersuchung war die Analyse der funktionalen Ausgestaltung von Online-Bewertungssystemen, mit der Absicht eine generelle Systematisierung von Gestaltungsoptionen zu entwickeln und zu diskutieren. Der hierbei zugrundeliegende Forschungsprozess gliedert sich die folgenden sechs Schritte:

- Auswahl von für die Analyse geeigneten Webseiten
- Identifizierung von Bewertungssystemen als Teil der ausgewählten Webseiten
- Iterative Entwicklung eines Kriterienkatalogs
- Auswertung der identifizierten Bewertungssysteme auf Basis des Kriterienkatalogs
- Analyse der Datenbasis und Konsolidierung der Kriterien

Auswahl von für die Analyse geeigneten Webseiten

Aus einer initialen Literaturrecherche ging hervor, dass Referenzstudien, die Bewertungssysteme in Bezug auf ihre funktionale Ausgestaltung untersuchen, bisher nicht verfügbar sind. Bei der Auswahl der für die Untersuchung geeigneten Webseiten musste daher ein eigener Weg gefunden werden. Die Auswahl der Stichprobe sollte transparent und objektiv mit einer hinreichend großen Anzahl an Bewertungssystemen sein. Um diesen Anforderungen gerecht zu werden, wurden zunächst die Top 50-Webseiten des Alexa Traffic Ranking (vgl. Tabelle 1) herangezogen. Dieses populäre Ranking wird aus Daten von Nutzern gespeist, die die Alexa-Toolbar in ihrem Browser installiert haben. Obwohl das Traffic-Ranking nur eine Näherung darstellt, liefert es zumindest einen Anhaltspunkt zur Identifikation der populärsten Webseiten. Webseiten mit pornografischem Inhalt und Angebote, die nicht in deutscher oder englischer Sprache verfügbar waren, wurden ausgeschlossen.

#	Webseite	Art des Dienstes	Ausschlussgrund	#	Webseite	Art des Dienstes	Ausschlussgrund
1	Yahoo.com	Webportal		26	Photobucket.com	Bilderdienst	
2	Google.com	Suchmaschine		27	Google.com.br	Suchmaschine	
3	Youtube.com	Videportal		28	Amazon.com	E-Commerce-Portal	
4	Live.com	Suchmaschine		29	Imdb.com	Filmportal	
5	Msn.com	Webportal		30	Vkontakte.ru	Soziales Netzwerk	Sprache
6	Myspace.com	Soziales Netzwerk		31	Google.it	Suchmaschine	
7	Wikipedia.org	Online-Enzyklopädie		32	Google.es	Suchmaschine	
8	Facebook.com	Soziales Netzwerk		33	Google.cn	Suchmaschine	Sprache
9	Blogger.com	Blog-Community		34	Imageshack.us	Bilderdienst	
10	Yahoo.co.jp	Webportal	Sprache	35	Youporn.com	Videportal	Inhalt
11	Orkut.com	Soziales Netzwerk		36	Wordpress.com	Blogportal	
12	Rapidshare.com	Online Storage Dienst		37	Google.co.jp	Suchmaschine	
13	Baidu.com	Suchmaschine	Sprache	38	Yandex.ru	Webportal	Sprache
14	Microsoft.com	Firmenhomepage		39	Flickr.com	Bilderdienst	
15	Google.co.in	Suchmaschine		40	Friendster.com	Soziales Netzwerk	
16	Google.de	Suchmaschine		41	Skyrock.com	Soziales Netzwerk	
17	Qq.com	Webportal	Sprache	42	Adultfriendfinder.com	Soziales Netzwerk	Inhalt
18	Ebay.com	Auktionsportal		43	Go.com	Webportal	
19	Hi5.com	Soziales Netzwerk		44	Odnoklassniki.ru	Soziales Netzwerk	Sprache
20	Google.fr	Suchmaschine		45	Goggle.com.mx	Suchmaschine	
21	Aol.com	Webportal		46	Bbc.co.uk	Newsportal	
22	Mail.ru	Webportal	Sprache	47	Craigslist.org	Werbportal	
23	Google.co.uk	Suchmaschine		48	Dailymotion.com	Videportal	
24	Sina.com.cn	Webportal	Sprache	49	Redtube.com	Videportal	Inhalt
25	Fc2.com	Online Storage Dienst	Sprache	50	Cnn.com	Newsportal	

Tabelle 1: Überblick über Alexa Top50 (April 2008)

Identifizierung von Bewertungssystemen als Teil der ausgewählten Webseiten

Bei der Identifikation von Bewertungsfunktionalitäten galt es zu berücksichtigen, dass in den Alexa Top 50 Portale wie Yahoo.com enthalten sind, die eine große Anzahl von separaten Online-Diensten bieten. Die Suche auf den Webseiten wurde daher auf die Dienste beschränkt, die direkt auf der Startseite verlinkt sind. Insgesamt wurden durch dieses Vorgehen 102 Bewertungssysteme identifiziert.

Iterative Entwicklung eines Kriterienkatalogs

Um die Charakteristika jedes Bewertungssystems systematisch erfassen und alle Systeme vergleichen zu können, wurden iterativ 237 Kriterien erarbeitet. Da kein zweckmäßiger Kriterienkatalog identifiziert werden konnte, wurde zunächst eine Literaturrecherche vorgenommen. Der Katalog wurde im Laufe der Untersuchung erweitert, wobei hinzukommende Kriterien auch bei bereits untersuchten Bewertungssystemen nachträglich berücksichtigt wurden.

Auswertung der identifizierten Bewertungssysteme auf Basis des Kriterienkatalogs

Jedes Bewertungssystem wurde anhand der Kriterien überprüft. In wenigen Fällen war es nicht möglich zu entscheiden, ob ein System eine bestimmte Charakteristik besitzt oder nicht. Daher wurden zusätzlich die Eigenschaften *nicht anwendbar* und *nicht prüfbar* eingeführt. Ein Kriterium ist nicht anwendbar, wenn die Voraussetzungen dafür nicht vorliegen. Nicht prüfbar wurde dann vergeben, wenn sich das Verhalten des Systems z. B. aufgrund fehlender Dokumentation nicht genau bestimmen ließ.

Analyse der Daten und Konsolidierung der Kriterien

In diesem Artikel können nur Auszüge der Studie dargestellt. Eine Systematisierung in Form eines morphologischen Kastens [Zwi69] wurde durch kritische Diskussion und Konsistenzprüfungen abgeleitet, um die Ergebnisse der Studie zu aggregieren und zu strukturieren [Bai94].

4 Analyse des Status-Quo

Die im Rahmen der Analyse von Bewertungssystemen erlangten Ergebnisse werden im Folgenden anhand von sechs Kerndimensionen der funktionalen Gestaltung dargestellt. Sie stellen eine Kategorisierung der identifizierten Gestaltungsoptionen dar und sind implizit wiederum als Ergebnis der Analyse anzusehen. Ausgehend von der Vielzahl unterschiedlicher Arten von Diensten und Plattformen können unterschiedliche Anforderungen an die Möglichkeiten der Gestaltung von Bewertungssystemen resultieren. Die hieraus resultierenden Kriterien werden in der Dimension *Dienstbetreiber* zusammengefasst. In einer weiteren Dimension werden die Gegenstände der Bewertung (*Entitäten*) hinsichtlich resultierender Einflussparameter analysiert. Aus Sicht des Bewertungsgebers werden weiterhin die Dimensionen der *Abgabe*, *Aggregation* und *Präsentation* von Be-

wertungen untersucht. Abschließend werden Gestaltungsoptionen zur Schaffung von *Anreizen* für die Abgabe von Bewertungen betrachtet.

4.1 Dienstbetreiber

Mit dem breiten Spektrum an analysierten Dienst Anbietern (Produzenten, Händler, Vergleichsdienste, Bewertungsportale, Communities etc.) gehen unterschiedliche Intentionen einher, die sich in sechs Zielstellungen unterscheiden lassen (vgl. Tabelle 2). Die Absatzförderung bezieht sich auf Bewertungsverfahren, die Händler oder Produzenten im Rahmen des Verkaufs ihrer Produkte und Leistungen verwenden. Bewertungsverfahren zur Partnerselektion treten auf Marktplätzen auf, um potenzielle Transaktionspartner zu identifizieren. Vergleichsdienste verfolgen mit Bewertungssystemen zumeist das Ziel der Produktbewertung.

Intention zur Bereitstellung	# Systeme
Produktbewertung bei Vergleichsdiensten	5
Absatzförderung	8
Auswahl des Transaktionspartners	9
Informationsbewertung	16
Neutrale Bewertungsfunktionalität	27
Community-bezogene Funktionen	37

Tabelle 2: Intentionen zur Bereitstellung von Bewertungssystemen aus Betreibersicht

Die Unterteilung der Bewertung von Produkten nach Absatzförderung, Vergleichsdienste und neutraler Bewertungsfunktionalität ist in der Existenz unterschiedlicher Geschäftsmodelle begründet. So ist anzunehmen, dass vor allem Nutzer mit einer überdurchschnittlichen Preissensitivität auf Vergleichsdienste zurückgreifen. Dies lässt den Schluss zu, dass bei Vergleichsdiensten die Bewertungen anders wahrgenommen werden als bei einem Händler oder Produzenten. Die vermehrte Präsenz neutraler Bewertungsfunktionalität ist darauf zurückzuführen, dass Bewertungssysteme nicht ausschließlich in Bewertungsportalen Einsatz finden. Dies kann als Beleg für die zunehmende Beliebtheit von Bewertungssystemen angesehen werden. Der Erfüllung einer Community-Funktion kommt mit einer Anzahl von 37 die größte Bedeutung zu.

Die Betrachtung der Einbindung der Bewertungssysteme in den jeweiligen Dienst zeigt, dass 77 der untersuchten Bewertungssysteme sich auf den vom Dienst fokussierten Gegenstand beziehen (bspw. Autos bei einem Autoportal). Bei 12 dieser 77 Systeme sind Bewertungen elementarer Gegenstand des Dienstes. So zielt bspw. Pixnay.bebo.com ausschließlich auf die Bewertung der Benutzer anhand eingestellter Bilder ab.

4.2 Entität

Die Entität ist das zentrale, zu bewertende Element eines Bewertungssystems und hat folglich erheblichen Einfluss auf die Ausgestaltung von Bewertungssystemen. Die im Rahmen der Analyse identifizierten Entitäten können anhand der drei Typen *Subjekt*, *abstraktes Subjekt* und *Objekt* charakterisiert werden (vgl. Abbildung 1).

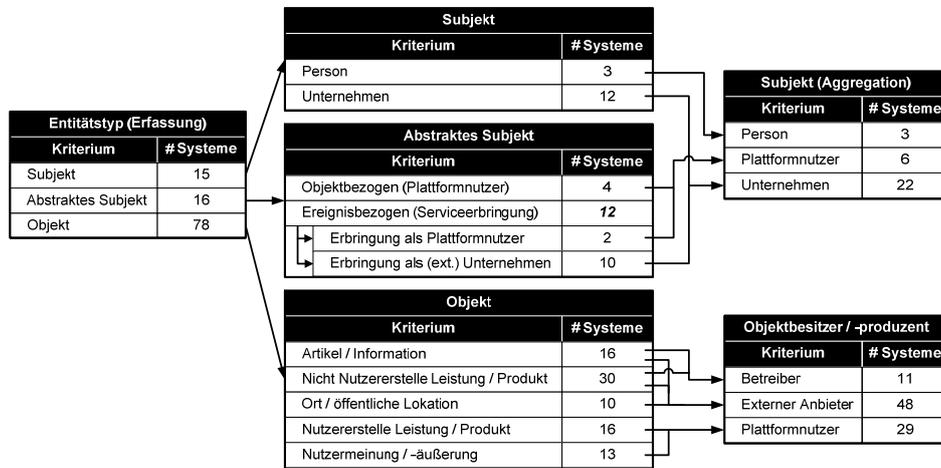


Abbildung 1: Klassifikation von Entitäten

Unter Subjekten werden selbstständig handelnde Entitäten verstanden. Objekte stellen hingegen passive Entitäten wie bspw. Produkte oder Informationen dar. Ein Subjekt wird als abstraktes Subjekt betrachtet, wenn ein Objekt bzw. ein Ereignis bewertet wird, die Verdichtung der Bewertungen jedoch auf Basis eines Subjekts erfolgt. Als Beispiel lässt sich Ebay.com nennen, wo Verkäufer durch Käufer über abgewickelte Transaktionen bewertet werden. Die Bewertungen sind wiederum mit dem Profil des Käufers verknüpft. Hieraus lässt sich nur beurteilen, wie das handelnde Subjekt in Bezug zur Abwicklung einer Transaktion zu betrachten ist. Folglich ergeben sich aus dieser Bewertungsform Einschränkungen bei der Interpretation und Anwendbarkeit. Eine Aufstellung der Anzahl von Systemen mittels derer die einzelnen Entitätstypen bewertet werden können findet sich in Abbildung 1.

Subjekte werden anhand von objektbezogenen und ereignisbezogenen Bewertungen weiterhin in *Personen* und *Unternehmen* sowie *abstrakte Subjekte* unterschieden. Die 12 ereignisbezogenen Systeme können dahingehend differenziert werden, ob ein Plattformnutzer oder ein (externes) Unternehmen betrachtet wird. Ein abstraktes Subjekt kann letztlich als Subjekt auf einer Plattform dargestellt werden. Folglich können diese zu den drei Subjekttypen *Person*, *Plattformnutzer* und *Unternehmen* aggregiert werden. Die 78 objektbezogenen Systeme werden ebenfalls anhand von fünf Objektarten differenziert betrachtet (vgl. nochmals Abbildung 1). Bei einem Teil der Bewertungsverfahren besteht die Möglichkeit der Bewertung von verschiedenen Objektarten. Es wird ferner betrachtet, wer der Besitzer bzw. Produzent der Objekte ist. Bei 11 Bewertungssystemen werden vom Betreiber erstellte Objekte bewertet, bei 48 Systemen hingegen Objekte von externen Anbietern. Bei 29 Systemen sind dies vom Plattformnutzer erstellte Objekte.

4.3 Erfassung von Bewertungen

Hinsichtlich der Erfassung von Bewertungen wird zunächst analysiert, wer zur Abgabe einer Bewertung berechtigt ist. Hierbei wurden 84 Systeme identifiziert, die eine Regist-

rierung voraussetzen. Hiervon wird bei 62 Systemen die Angabe einer E-Mail-Adresse verlangt. Bei 11 Systemen werden für die Bewertungsabgabe gewissen Qualifikationen vorausgesetzt, bspw. durch Gruppenmitgliedschaften (2 Systeme) oder einen spezifischen Mitgliedsstatus (3 Systeme). Bei 6 Systemen ist die Abwicklung einer Transaktion Voraussetzung für eine Bewertungsabgabe. Mit der Vielzahl an Betreibern variiert die Form der Bewertungserfassung. Unterschiede zeigen sich vor allem im Detaillierungsgrad und im Umfang der abgefragten Informationen. Weiterhin kann zwischen einer direkten Bewertung und der Abgabe einer Bewertung in Relation zu einer Vergleichsentität unterschieden werden, wobei die direkte Bewertung dominiert. Abschließend wurde analysiert, inwiefern die Möglichkeit der Mehrfachbewertung einer Entität durch einen Benutzer besteht, wobei 67 Systeme dies in Verbindung mit einer vorherigen Registrierung ausschließen.

4.4. Verdichtung von Bewertungen

Die Aggregation von Bewertungen sowie deren Darstellung liegen zumeist im Einflussbereich des Betreibers und können somit als bedeutende Erfolgsfaktoren von Bewertungssystem betrachtet werden. 93 der analysierten 102 Systeme aggregieren Einzelbewertungen einer Entität zu einem Punktwert (vgl. Tabelle 3). Dies erscheint insofern problematisch, als dass Informationen verloren gehen. Einige Dienste erlauben die Abgabe von Einzelbewertungen (bspw. Bewertung des *Karma* bei orkut.com über die Eigenschaften „zuverlässig“, „cool“ oder „sexy“) ohne diese zu aggregieren. Weiterhin geben 8 Systeme die Anzahl abgegebener Bewertungen pro Eigenschaft an.

Bei 78 Bewertungssystemen erfolgt die Verdichtung der Bewertungen über die Bildung eines Mittelwerts. 22 Systeme verfolgen bei der Aggregation der Bewertungen einen kumulativen Ansatz wie das Zählen von Bewertungen (12 Systeme), Kumulieren pro Skalenwert (5 Systeme) bzw. Kumulieren unter Ausschluss negativer Bewertungen (4 Systeme) sowie das Zählen von Stimmen in Relation zur Gesamtzahl (1 System). Einige Dienste beziehen weitere Informationen wie bspw. die Häufigkeit der Betrachtung eines Artikels in die Berechnung mit ein. Um offenbar Manipulationen vorzubeugen, erfolgt mehrheitlich keine exakte Beschreibung der Verdichtungsmethode.

Kriterium	# Systeme				Total
	Ja	Nein	Nicht anwendbar	Nicht prüfbar	
Verdichtung je Entität	93	9	0	0	102
Durchschnittswert	78	22	1	1	102
Kumulativ	22	78	1	1	102
→ Kumulierung pro Skalenwert	5	96	1	0	102
→ Stimmen für Skalenwert in Relation zu Gesamtzahl	1	100	1	0	102
→ Kumulierung ohne negative Bewertungen	4	96	1	1	102
→ Zählen von Stimmen	12	89	1	0	102

Tabelle 3: Verdichtung von Bewertungen

4.5 Darstellung von Bewertungen

Die Form der Darstellung determiniert ebenfalls die Aussagekraft eines Bewertungssystems [SM07] und wird durch die erhobenen Daten bedingt [DW08]. In Tabelle 4 wird analysiert, welche der erhobenen Daten dem Nutzer angezeigt werden und in welcher Form dies erfolgt. Bei 89 Bewertungssystemen sind die Bewertungen für alle Plattformnutzer sichtbar. Von den 13 resultierenden Systemen stellen 4 Systeme die Bewertungen nur für registrierte Nutzer zur Verfügung, 9 Systeme setzen eine Qualifikation voraus. In der Mehrheit aller Fälle (85 Systeme) wird die Gesamtzahl aller Bewertungen auf Profil- (ebenfalls bei 85 Systemen) oder Übersichtsseiten (30 Systemen) angegeben. Bei 28 Systemen wird die Anzahl der Bewertungen je Skalenwert angegeben (bspw. bewerteten 17 Nutzer einen Artikel mit "sehr gut", 24 Nutzer mit "gut" etc.). 4 Systeme stellen zusätzliche Kennzahlen zu den Bewertungen zur Verfügung.

Kriterium	# Systeme				
	Ja	Nein	Nicht anwendbar	Nicht prüfbar	Total
Sichtbarkeit von Bewertungen					
Sichtbar für alle Plattformnutzer	89	13	0	0	102
↳ Sichtbar nur für registrierte Nutzer	4	9	89	0	102
↳ Sichtbar nur bei bestimmten Qualifikationen	9	0	93	0	102
Gesamtzahl abgegebener Bewertungen					
Angabe der Gesamtzahl	85	16	1	0	102
↳ Angabe auf Entitätsebene / Profiseite	85	0	17	0	102
↳ Angabe bei Suchergebnissen / Übersichtsseiten	30	68	4	0	102
Angabe weiterer Informationen zur Bewertung					
Anzahl der Bewertung pro Skalenwert	28	70	4	0	102
Prozentsatz positiver Bewertungen	7	94	0	1	102
Bereitstellung zusätzlicher Kennzahlen	4	98	0	0	102
Aktivierung / Freischaltung von Bewertungen					
Aktivierung verzögert / durch Betreiber	32	69	0	1	102
↳ Sofortige Sichtbarkeit trotz offizieller Verzögerung	12	14	69	7	102
Sofortige Sichtbarkeit	69	32	0	1	102
Sichtbarkeit von Einzelbewertungen					
Einzelne Bewertungen sichtbar	54	48	0	0	102
Angabe zusätzlicher Daten für interne Zwecke	8	8	83	3	102
Standardreihenfolge für Einzelbewertungen					
Aktuelle Bewertung zuerst	38	13	48	3	102
Reihenfolge basiert auf Algorithmus	3	48	48	3	102
Sortierung / Selektion von Bewertungen					
Anzeige TopRated / selektierbar	38	64	0	0	102
Sortierung auf Basis von Kriterien möglich	28	26	48	0	102

Tabelle 4: Darstellung von Bewertungen

Die Betreiber von 32 Bewertungssystemen gaben an, dass die Bewertungen einer Prüfung unterzogen und daraufhin verzögert freigeschaltet werden. Bei Expedia.com erscheinen bspw. Bewertungen erst nach einer Prüfung mit einer Verzögerung von bis zu 10 Tagen. Bei 12 der 32 Systeme wird angegeben, dass Bewertungen geprüft und verzögert freigeschaltet würden, obwohl diese mit der Abgabe direkt sichtbar werden. 69 Systeme erlauben eine Sichtbarkeit von Bewertungen direkt nach deren Abgabe. Darüber hinaus erlauben 54 Systeme den Einblick in Einzelbewertungen, bei 48 Systemen ist

dies indes nicht möglich. Zusatzdaten werden nicht von allen Betreibern veröffentlicht bzw. sie werden nicht direkt mit der Bewertung angezeigt. Letzteres ist bspw. bei Ebay.com der Fall, wo ein Käufer neben der Gesamtbeurteilung auch in Bezug auf Kriterien wie Lieferzeit beurteilt werden kann, wobei diese Bewertungsteile nur in aggregierter Form veröffentlicht werden.

Einige Betreiber bieten Sortier- und Selektionsmechanismen für Bewertungssysteme an. Bei 38 Systemen werden Einzelbewertungen stets nach der Aktualität sortiert ausgegeben. Bei 3 Systemen wird ein spezieller Sortieralgorithmus angewandt. Darüber hinaus erlauben eine Reihe von Systemen eine individuelle Sortierung und Auswahl von Bewertungen. So sind Bewertungen bei 38 Systemen gezielt selektierbar. Bei 28 Systemen ist darüber hinaus eine individuell Sortierung und Selektion von Bewertungen auf Basis vorgegebener Kriterien möglich.

4.6 Anreizmechanismen

Anreizmechanismen dienen der Anregung von Nutzern, Bewertungen zu Entitäten abzugeben. Sie beeinflussen somit die Akzeptanz und den Erfolg von Bewertungssystemen. Bewertungen können hierbei nicht nur für den Bewertungsgeber und -nehmer Nutzen stiften, sondern vielmehr für Konsumenten von Bewertungen. Die Darstellung von Benutzernamen oder realen Namen wird von 51 Bewertungssystemen genutzt, um Bewertungsgeber hervorzuheben (vgl. Tabelle 5). 26 Systeme erlauben eine direkte Verknüpfung mit dem Benutzerprofil des Bewertungsgebers. Eine aggregierte Darstellung der abgegebenen Bewertungen im Benutzerprofil des Gebers erlauben 22 Systeme. Darüber hinaus wird eine Bewertung von Bewertungsgebern und die Vergabe von speziellen Auszeichnungen von einer geringen Anzahl von Systemen genutzt, um Anreize zur Bewertungsabgabe zu schaffen. Nur 2 Systeme offerieren monetäre Anreize zur Abgabe von Bewertungen.

Kriterium	# Systeme				
	Ja	Nein	Nicht anwendbar	Nicht prüfbar	Total
Sichtbarkeit des Bewertungsgebers pro Bewertung					
Sichtbarkeit des Benutzernamens	25	26	48	3	102
Sichtbarkeit des realen Namens	26	25	48	3	102
Möglichkeiten der Profilbildung mittels Bewertungen					
Anzeige abgegebener Bewertungen im Nutzerprofil	22	4	76	0	102
Ranking von Bewertungsgeber	5	17	80	0	102
Verwendung von Labels	9	72	21	0	102
↳ Expertenlabels	3	6	93	0	102
↳ Charakterlabels	6	3	93	0	102
Monetäre Vergütung von Bewertungen	2	100	0	0	102

Tabelle 5: Anreize zur Abgabe von Bewertungen

5. Diskussion der Forschungsergebnisse

Die Anwendung von 237 Kriterien auf 102 Bewertungssysteme ergab eine breite Vielfalt an Eigenschaften von Bewertungssystemen. Die Identifikation und Vorgabe eines für jeden Anwendungsfall gültigen und anwendbaren Gestaltungsansatzes ist folglich nicht möglich. Jeder Anwendungsbereich hat spezielle Anforderungen an die Ausgestaltung von Bewertungssystemen. Mit dem Ziel, die Gestaltung von Bewertungssystemen anzuleiten, wird im Folgenden eine Systematisierung der identifizierten Gestaltungsoptionen vorgeschlagen. Identifizierte Gestaltungsoptionen wurden kritisch diskutiert, unter wechselseitiger Konsistenzprüfung [Bai94] aggregiert und schließlich in einem morphologischen Kasten [Zwi69] zusammengefasst (vgl. Abbildung 2).

Die untersuchten Dienste und Plattformen weisen entweder eine kommerzielle oder eine nicht kommerzielle, gemeinnützige Ausrichtung auf. Die Mehrheit der untersuchten Dienste hat einen kommerziellen Hintergrund, wobei das Bewertungssystem eng mit dem Geschäftsmodell verbunden ist und daher ein Erfolgsfaktor für diese Plattformen darstellt. Aber auch Dienste mit gemeinnützigem Hintergrund weisen kaum Unterschiede in der verfolgten Zielstellung und Art der bewerteten Entitäten auf. Bei Entitäten kann grundlegend zwischen Subjekten, abstrakten Subjekten und Objekten unterschieden werden. Bewertungen von Entitäten werden nicht nur auf den jeweiligen Dienst bezogen sondern können auch in anderen Diensten eines Betreibers und über Plattformen hinweg eingebunden und interpretiert werden. Hierbei ist die Integrität der Bewertungen zu beachten. So herrscht bspw. bei Amazon.com Integrität über die komplette Domain hinweg, indem Bewertungsdaten zu Produkten stets mit der Detailseite der Entität verknüpft sind.

	Parameter	Ausprägungen			
Betreiber	Ausrichtung	Kommerziell		Nicht kommerziell	
	Typ	Subjekt	Abstraktes Subjekt	Objekt	
Entität	Bezug zum Dienst	Spezifisch		Unspezifisch	
	Integrität (Kapselung)	Dienstbereich	Domain	Domain-übergreifend	Betreiber-übergreifend
	Nutzerkreis	Unbeschränkt		Beschränkt	
Erfassung	Voraussetzung	Keine	Einmalige Qualifikation	Qualifikation pro Bewertung	
	Bewertungsart	Direkt		Relativ	
	Bewertungsrichtung	Unilateral		Bilateral	
	Urteilslevel	Gesamturteil		Teildimension	
	Mehrfachbewertung	Unterbunden	Konditional	Unbegrenzt möglich	
	Einbezug	Ausschließlich Bewertungen durch Nutzer		Einbezug mehrerer Bewertungen	
Aggregation	Berechnungsmethode	Keine	Kumulativ	Durchschnittsbildung	
	Gewichtung	Gleichgewichtet	Zeitabhängig	Qualitätsabhängig	Abhängig von Bewertungsgeber
Darstellung	Konsumentenkreis	Beschränkt		Unbeschränkt	
	Sichtbarkeit	Unmittelbar	Verzögert	Nach Prüfung	
	Abbildung der Bewertung	Keine	Beschränkt		Unbeschränkt
	Kommentierung	Nicht möglich	Unbeschränkt	Beschränkt	Indirekt
Anreize	Angabe des Bewerbers	Ohne Angabe	Anonym	Name	Profil
	Anreize	Keine	Indirekt	Direkt	

Abbildung 2: Systematisierung der Gestaltungsoptionen von Bewertungssystemen

Die Verknüpfung von Restriktionen mit der Bewertungsabgabe kann die Bewertungsqualität fördern [Del00]. Die meisten Bewertungssysteme weisen eine Beschränkung des Nutzerkreises auf. Infolgedessen ist eine Validierung der Bewertungsberechtigung möglich. Die Einführung eines restringierten Bewertungssystems wird sehr stark von der Nutzerakzeptanz, dem Geschäftsmodell des Dienstes und der Bedeutung der Bewertungen abhängig gemacht. Wenn die Nutzung eines Dienstes bereits eine Anmeldung voraussetzt, ist ein beschränktes Bewertungssystem zu empfehlen. In sensiblen Anwendungsfällen (z. B. Auktionen) ist die Einbindung von weiteren Voraussetzungen (z. B. Abwicklung einer Transaktion) sinnvoll. Zusätzliche Qualifikationsmechanismen wie bspw. Freundschaftsstatus können als Voraussetzung für Bewertung von Mitgliedern in sozialen Netzwerken dienlich sein.

Die Mehrheit der Bewertungssysteme ermöglichen es, Meinungen in einer direkten Art auszudrücken. Hierdurch wird die Behandlung der Bewertungen bei der Verdichtung sehr stark vereinfacht. Wie wissenschaftliche Studien belegen, sind derartige Bewertungssysteme jedoch sehr anfällig für Verfälschungen [GS06]. Ferner können bilaterale, direkte Bewertungen zum Phänomen der Rachebewertungen führen. Aus diesem Grund wurde bspw. das Bewertungssystem auf Ebay.com zu einem eher unilateralen System umgeformt. In wenigen Fällen sind Bewertungssysteme auf eine relative Bewertung von Entitäten (in Bezug auf eine andere Entität) ausgelegt. Ein derartiger Ansatz ist hingegen nur in Szenarien mit vergleichbaren und in hohem Grad verknüpften Entitäten möglich [BL08]. Grundsätzlich können relative Bewertungssysteme zur Unterbindung von Verfälschungen beitragen, diese jedoch nicht vollkommen verhindern.

Betreiber von Bewertungssystemen tendieren zur Verdichtung von Einzelbewertungen zu einer generellen Bewertungsaussage, riskieren hierbei jedoch den Verlust individueller Informationen. Die meisten Berechnungsmethoden zur Verdichtungen erweisen sich als sehr einfach und basieren auf einer Gleichgewichtung der Bewertungen. Hierbei berechnen einfache Systeme je Entität den Durchschnitt der Bewertungen, summieren die Werte der Einzelbewertungen auf oder liefern die Anzahl abgegebener Bewertungen zurück. Erweiterte Systeme berechnen einen gewichteten Mittelwert der Bewertungen und beziehen hierbei Faktoren wie Alter der Bewertung, Anzahl der Betrachtungen einer Entität oder Vertrauen des Bewertungsgebers als Gewichtungsfaktor mit ein. Die Anwendung anspruchsvollerer statistischer Methoden erfolgt nicht, wäre jedoch denkbar. Im Hinblick auf die Häufigkeit der Bewertungen schließt zirka die Hälfte der Systeme eine wiederholte Bewertung einer Entität durch den gleichen Bewertenden aus.

Die Betrachtung abgegebener Einzelbewertungen sowie verdichteter Gesamtbewertungen ist gewöhnlich für jeden Nutzerkreis möglich. Die Vielzahl an sichtbaren Bewertungen deutet generell auf eine hohe Bedeutung der Bewertungen für die Geschäftsmodelle der Dienstbetreiber hin. In diesem Zusammenhang ist auch der mehrfach von Betreibern genutzte Kontrollmechanismus von Bewertungen zu sehen. Hierbei werden Bewertungen explizit durch den Betreiber freigeschaltet und werden somit erst verzögert sichtbar. Hierdurch kann die Qualität und Seriosität von Bewertungen erheblich gesteigert werden. Darüber hinaus bieten Bewertungssysteme unterschiedliche Möglichkeiten die bestehenden Bewertungen zu kommentieren.

Das Setzen von Anreizen zur Abgabe einer Bewertung steht im Verhältnis zu den Zielen, die mit dem Bewertungssystem verfolgt werden. Die Ausprägungen können Einfluss auf die Anzahl und die Qualität der Bewertung haben. Bei der Angabe des Bewertungsgebers kann unterschieden werden, ob dieser anonym angegeben wird, mit seinem Namen, mit Link zu seinem Profil oder ob darauf gänzlich verzichtet wird. Die Anreizstimulierung kann unsystematisch sein, d. h. es wird auf Instrumente zur Stimulierung verzichtet, oder es werden Mittel wie Auszeichnungen und Rankings verwendet.

6. Limitationen und weiterer Forschungsbedarf

Wenngleich die durchgeführte Analyse auf einer großen Stichprobe von Webseiten und zugehörigen Bewertungssystemen aufbaut, konnten aufgrund sprachlicher Barrieren nur deutsch- und englischsprachige Webseiten untersucht werden. Darüber hinaus ist die Analyse auf die 50 populärsten Webseiten beschränkt. In zukünftige Forschungsaktivitäten sollen hingegen auch weniger bekannte Webseiten einbezogen werden, um mögliche Gestaltungsunterschiede zwischen Bewertungssystemen von hoch- und weniger frequentierten Webseiten verstehen und in die vorgeschlagene Systematisierung einfließen zu lassen. Die Betrachtung von Webseiten und Bewertungsmechanismen aus anderen Kulturkreisen liefert möglicherweise weitere Beiträge zur bisher aufgespannten Systematisierung von Gestaltungsoptionen. Mit 237 Kriterien ist davon auszugehen, einen umfangreichen und umfassenden Kriterienkatalog geschaffen zu haben. Dennoch ist der Umstand nicht auszuschließen, dass die Analyse weiterer Bewertungssysteme neue, bisher nicht berücksichtigte Kriterien aufbringt. Die bisherigen Analysen konnten leider nur in der Rolle des Nutzers durchgeführt werden, ohne einen tieferen Einblick in die Funktionsweise der Bewertungssysteme zu erhalten.

Die bisherigen Bewertungssysteme sind mehrheitlich auf die Bewertung von Objekten und weniger von Subjekten ausgerichtet. Mit der zunehmenden Verbreitung von sozialen Netzwerken und anderen sozialen Anwendungen ist zukünftig eine zunehmende Verschiebung der Ausrichtung hin zur Beurteilung von Nutzern anzunehmen. In Verbindung mit sozialen Netzwerken werden in den nächsten Jahren neue Bewertungssysteme aufkommen, die weiteren Fragestellungen wie Privatsphäre, Selbstbestimmung, Datenschutz etc. mit sich bringen. Technologische Entwicklungen wie OpenSocial (Opensocial.org) eröffnen neue Wege der Interaktion zwischen Plattformen und ermöglichen so den Entwurf von plattformübergreifenden Bewertungssystemen. Die hier präsentierte Systematisierung bietet ein erstes Verständnis von Bewertungssystem und kann als Ausgangspunkt für zukünftige Forschungsaktivitäten angesehen werden.

Literaturverzeichnis

- [AH95] Armstrong, A.; Hagel III, J.: Real profits from virtual communities. *The McKinsey Quarterly* (3) 1995; S. 126-141.
- [Bai94] Bailey, K.: *Typologies and Taxonomies - An Introduction to Classification Techniques*. Sage University Papers: Sage Publications, Thousand Oaks, 1994.

- [BKO04] Bolton, G.; Katok, E.; Ockenfels, A.: How Effective are Online Reputation Mechanisms? An experimental investigation. *Management Science*, Volume 50 (11) 2004; S. 1587-1602.
- [BL08] Botsch, J.; Luckner, S.: Empirische Analyse von Bewertungskommentaren des Reputationssystems von eBay. In Bichler, M.; Hess, T.; Krcmar, H.; Lechner, U.; Matthes, F.; Picot, A.; Speitkamp, B.; Wolf, P., (Hrsg.): *Proceedings of the Multikonferenz Wirtschaftsinformatik, München, 2008*; S. 1199-1210.
- [CHW08] Chen, K.-Y.; Hogg, T.; Wozny, N.: Experimental study of market reputation mechanisms. In: *Proceedings of the Proceedings of the 5th ACM conference on Electronic commerce, New York, USA, 2004*.
- [CLS07] Cheung, C.M.Y.; Luo, C.; Sia, C.L.; Chen, H.P.: How do People Evaluate Electronic Word-of-Mouth? Informational and Normative Based Determinants of Perceived Credibility of Online Consumer Recommendations in China. In: *Proceedings of the Pacific Asia Conference on Information Systems (PACIS 2007), Auckland, Neuseeland, 2007*.
- [CP02] Conte, R.; Paolucci, M.: Reputation in artificial societies: social beliefs for social order In (Weiss, G., Hrsg.): *Multiagent systems, artificial societies, and simulated organizations*. Kluwer Academic Publishers, Boston u. a., 2002.
- [Del00] Dellarocas, C.: Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior. In Jhingran, A.; Mason, J.M.; Tygar, D., (Hrsg.): *Proceedings of the 2nd ACM Conference on Electronic Commerce, Minneapolis, 2000*; S. 150-157.
- [Del01] Dellarocas, C.: Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. In: *Proceedings of the Proceedings of the 3rd ACM conference on Electronic Commerce, Tampa, Florida, USA, 2001*.
- [Del03] Dellarocas, C.: The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, 49 (10) 2003; S. 1407-1424.
- [DW08] Dellarocas, C.; Wood, C.A.: The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. *Management Science*, 54 (3) 2008; S. 460-476.
- [GS06] Gregg, D.G.; Scott, J.E.: The Role of Reputation Systems in Reducing On-Line Auction Fraud. *International Journal of Electronic Commerce*, 10 (3) 2006; S. 95-120.
- [JRB07] Josang, A.; Roslan, I.; Boyd, C.: A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems*, 43 (2) 2007; S. 618-644.
- [MC01] Miceli, M.; Castelfranchi, C.: The role of evaluation in cognition and social interaction. In Dautenhahn, K., (Hrsg.): *Human cognition and agent technology*, John Benjamins, Amsterdam 2000; S. 225-261.
- [PR08] Peters, R.; Reitzenstein, I.: Reputationssysteme im eCommerce - Funktionsweise, Anwendung und Nutzenpotenziale. *HMD - Praxis der Wirtschaftsinformatik*, 45 (261) 2008; S. 43-50
- [RKZ00] Resnick, P.; Kuwabara, K.; Zeckhauser, R.; Friedman, E.: Reputation systems. *Communications of the ACM*, 43 (12) 2000; S. 45-48.
- [RZ01] Resnick, P.; Zeckhauser, R.: Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. Working Paper for the NBER workshop on empirical studies of electronic commerce. In, 2001.
- [SM07] Sabater-Mir, J.; Paolucci, M.: On representation and aggregation of social evaluations in computational trust and reputation models. *International Journal of Approximate Reasoning*, 46 (3) 2007; S. 458-483.
- [Sch04] Schuler, H.: Der Prozess der Urteilsbildung und die Qualität der Beurteilungen. In Schuler, H.H., (Hrsg.): *Beurteilung und Förderung beruflicher Leistung*, Hogrefe-Verlag, Göttingen u.a., 2004; S. 33-60.

- [Seb08] Sebralla, M.-L.: Über 875 Millionen Menschen shoppen online. 40 Prozent mehr als noch vor zwei Jahren, 2008. <http://de.nielsen.com/news/pr20080206.shtml>. Abruf am 2008-07-29.
- [TSW08] Töpfer, A.; Silbermann, S.; William, R.: Die Rolle des Web 2.0 im CRM. In Töpfer, A., (Hrsg.): Handbuch Kundenmanagement, Springer, Berlin, Heidelberg, 2008; S. 651-675.
- [VH07] Vossen, G.; Hagemann, S.: Unleashing Web 2.0. Morgan Kaufmann Publishers, Burlington, 2007.
- [WL99] Ward, M.R.; Lee, M.J.: Internet shopping, consumer search and product branding. Journal of Product and Brand Management 1999.
- [Zie08] Ziegler, P.-M.: Datenschützer verhängt Bußgeld gegen Bewertungsportal meinprof.de, 2008. <http://www.heise.de/newsticker/Datenschuetzer-verhaengt-Bussgeld-gegen-Bewertungsportal-meinprof-de-/meldung/107123>. Abruf am 2008-06-23.
- [Zwi69] Zwicky, F.: Discovery, Invention, Research - Through the Morphological Approach. The Macmillian Company Toronto, 1969.