

Partizipieren statt Konsumieren, oder: Lohnt sich Engagement in Q&A-Communities?

Philipp Herrmann
Universität Paderborn
Warburger Straße 100
33098 Paderborn
0049 5251 60 5511

philipp.herrmann@wiwi.uni-paderborn.de

Dennis Kundisch
Universität Paderborn
Warburger Straße 100
33098 Paderborn
0049 5251 60 5533

dennis.kundisch@wiwi.uni-paderborn.de

ZUSAMMENFASSUNG

In den letzten Jahren haben sich online Question & Answer Communities wie Yahoo! Answers oder CosmiQ für viele Internetnutzer als bedeutende Quelle von Informationen etabliert. Es ist jedoch fraglich, welche Faktoren die Qualität der Informationen, die ein Nutzer in einer solchen Community erhält, beeinflussen. Zur Beantwortung dieser Frage nutzen wir die Tauschtheorie und die Uncertainty Reduction Theory. Anhand dieser Theorien entwickeln wir ein Modell zur Vorhersage der Qualität der Antworten, die ein Nutzer einer solchen Community auf seine Fragen erhält. Dieses Modell testen wir anhand von 295.500 Fragen, die in den Jahren 2007 und 2008 in einer deutschsprachigen Question & Answer Community gestellt wurden. Die Ergebnisse zeigen, dass das allgemeine Interesse an einer Frage, vorausgegangenes, qualitativ hochwertiges Engagement des Fragestellers in der Community sowie die Preisgabe bestimmter persönlicher Informationen einen positiven Einfluss auf die Qualität der Antworten haben.

Schlüsselwörter

Soziale Netzwerke, Q&A-Community, Sozialkapital, Tauschtheorie, Uncertainty Reduction Theory

1. Einleitung

Immer mehr Menschen nutzen das Internet, um ihren Informationsbedarf zu befriedigen. Dabei stellen Suchmaschinen einen Hauptzugang zu den im Internet verfügbaren Informationen dar. Neben diesen haben sich in den letzten Jahren verschiedene *Question & Answer Communities* (Q&A-Communities) als wichtige Quelle von Informationen etabliert [25]. Die Anzahl der Seitenaufrufe der Top 5 Q&A-Communities ist zwischen 2007 und 2009 um fast 900% gestiegen [24] und Yahoo! Answers (answers.yahoo.com) ist heute, nach Wikipedia, die zweit

populärste Internetseite in der Kategorie *Reference*¹ [2]. In Q&A-Communities stellen Millionen registrierter Nutzer weltweit täglich zehntausende von Fragen [16] aus den unterschiedlichsten Lebensbereichen und erhalten – meist öffentlich einsehbare – Antworten auf diese Fragen.

Harper et al. [16] definieren eine Q&A-Community als „[...] a Web site [...] purposefully designed to allow people to ask and respond to questions on a broad range of topics“. Die Autoren unterteilen Q&A-Communities in *Digital Reference Services*, *Ask an Expert Services* und *Community Q&A Sites*. Dabei stellen Digital Reference Services die online Analogie zu traditionellen Bibliotheksauskunftsdiensten dar [23]. In diesen beantworten professionelle Rechercheure die Online-Anfragen von Bibliotheksbenutzern. Ask an Expert Services ermöglichen es den Nutzern Fachfragen direkt an Experten aus den jeweiligen Fachgebieten zu stellen, während Community Q&A Sites zur Beantwortung der Fragen auf alle Nutzer des Services zugreifen. Dabei bieten Community Q&A Sites neben der Formulierung und Beantwortung von Fragen weitere Funktionalitäten, wie z.B. Kommentar- und Bewertungsfunktionen zu Fragen und Antworten oder die Möglichkeit sich ein persönliches Profil anzulegen.

Die Literatur zu Community Q&A Sites lässt sich in *User-zentrierte* sowie in *Inhalts-zentrierte* Arbeiten einteilen. User-zentrierte Arbeiten untersuchen verschiedene Charakteristika der Nutzer von Community Q&A Sites (Shah et al. [25] geben einen guten Überblick über diese Arbeiten), während Inhalts-zentrierte Arbeiten meist der Frage nachgehen, welche Einflussfaktoren auf die Qualität der Antworten auf eine Frage wirken. Wir fokussieren uns im Folgenden auf diesen Inhalts-zentrierten Aspekt.

Harper et al. [16] untersuchen zum einen, ob sich Digital Reference Services, Ask an Expert Services und Community Q&A Sites in der Antwortqualität unterscheiden und zum anderen, wie ein Fragesteller die Qualität der Antworten auf seine Frage positiv beeinflussen kann. Dabei analysieren die Autoren den Einfluss der für die Beantwortung der Frage ausgesetzten Belohnung sowie der rhetorischen Qualität und des Themas der

¹ In dieser Kategorie werden Nachschlagewerke aller Art zusammengefasst. Diese umfassen unter anderem Enzyklopädien (z.B. wikipedia.org), Wörterbücher (z.B. dict.cc) und Bibliographien (z.B. ebscohost.com)

Frage. Sie zeigen, dass die Qualität der Antworten nur durch die für die Beantwortung der Frage ausgesetzte monetäre Belohnung signifikant beeinflusst werden kann. Weiterhin zeigen die Autoren, dass die durchschnittliche Antwortqualität auf Community Q&A Sites im Vergleich zu den beiden anderen Community Typen am höchsten ist.

Während die Arbeit von Harper et al. [16] Antworten auf die Frage gibt, in welchem Community Typ ein Fragesteller besonders gute Antworten auf seine Fragen erhält und durch welche Handlungen er die Qualität der Antworten auf seine Fragen positiv beeinflussen kann, gehen die im Folgenden vorgestellten Beiträge der Frage nach, wie besonders hochwertige Inhalte auf Community Q&A-Sites identifiziert werden können. Diese Ergebnisse können unter anderem dazu dienen als hochwertig eingestufte Inhalte als passende Suchergebnisse für in Suchmaschinen eingegebene Fragen vorzuschlagen.

Agichtein et al. [1], Liu et al. [22] und Shah und Pomerantz [26] analysieren Fragen und die dazugehörigen Antworten aus der Community Q&A Site Yahoo! Answers. Dabei nutzen die Autoren direkt aus Fragen und Antworten ableitbare Maße (z.B. Rechtschreibfehler), Maße die sich auf den Fragesteller bzw. Antwortgeber beziehen (z.B. die Qualität der bisher gestellten Fragen bzw. gegebenen Antworten) sowie Maße, welche die Attraktivität einer Frage bzw. Antwort erfassen (z.B. wie oft wurde eine Frage bzw. Antwort angeklickt), um qualitativ hochwertige Fragen und Antworten [1] bzw. nur qualitativ hochwertige Antworten [22], [26] zu identifizieren.

Die Autoren wenden Data Mining Verfahren an, um wertvolle Inhalte auf der Community Q&A Site zu identifizieren. Dabei nutzen sie sowohl Informationen über die Frage und den Fragesteller als auch Informationen über die Antwort und den Antwortgeber. Dies erfolgt jedoch ohne theoretische Fundierung, sondern vielmehr aus einer explorativen Sichtweise. Gemein ist den genannten Studien zudem, dass diese nicht die Qualität der Antworten, die ein *bestimmter Fragender* in Zukunft erhalten wird, untersuchen. Vielmehr identifizieren die Autoren potenziell wertvolle, *schon vorhandene* Antworten für *andere* Nutzer mit vergleichbaren Fragen.

Die Ergebnisse von Agichtein et al. [1], Liu et al. [22] sowie Shah und Pomerantz [26] legen außerdem nahe, dass neben den von Harper et al. [16] untersuchten Einflussfaktoren weitere, im Fragesteller und in der Frage begründete Einflussfaktoren, auf die Qualität einer Antwort wirken.

An dieser Stelle setzen wir mit unserer Arbeit an. Wir entwickeln ein theoretisch fundiertes Modell, mit dem die Qualität der Antworten auf eine Frage anhand von in der Frage und im Fragesteller begründeten Charakteristika vorhergesagt werden soll. Dieses Modell testen wir anhand eines Datensatzes einer deutschen Community Q&A Site. Dabei beantworten wir die folgende Forschungsfrage:

Welche in der Frage und im Fragesteller begründeten Einflussfaktoren führen dazu, dass ein Nutzer einer Community Q&A Site besonders hilfreiche Antworten auf seine Fragen erhält?

Unsere Untersuchung leistet dabei drei Beiträge zur bestehenden Literatur: Erstens liefern wir eine theoretische Fundierung der, von uns und anderen, betrachteten Einflussfaktoren auf die

Qualität der Antworten auf einer Community Q&A Site. Zweitens überprüfen wir diese Theorien empirisch und zeigen, welche Faktoren einen Beitrag zur Erklärung der Antwortqualität leisten. Drittens internationalisieren und erweitern wir die bestehende Forschung zu Community Q&A Sites, indem wir einen neuen Datensatz in die Literatur einführen.

Es zeigt sich, dass unterschiedliche Faktoren einen Einfluss darauf haben, ob ein Nutzer wertvolle Antworten auf seine Frage erwarten darf. Zum ersten spielt es eine Rolle, ob sich ein Nutzer insbesondere durch ein hochgeladenes Bild und der Angabe des Wohnsitzlandes freiwillig nicht vollständig anonym in der Community präsentiert. Zum zweiten zeigt sich, dass andere Nutzer nicht quantitatives sondern qualitatives Engagement belohnen. Wer in der Vergangenheit viele sehr hilfreiche Antworten gegeben hat, bekommt auch selbst mit einer höheren Wahrscheinlichkeit sehr wertvolle Antworten. Die Menge an bisher in der Community gegebenen wenig hilfreichen Antworten wirkt sich hingegen eher negativ aus. Zum dritten haben der Frageinhalt und die Fragestellung einen substanziellen Einfluss darauf, ob man wertvolle Antworten erwarten kann. Gut gestellte und für die Community-Teilnehmer interessante Fragen werden häufiger und qualitativ besser beantwortet als andere Fragen. Insgesamt finden sich Hinweise darauf, dass die Tauschtheorie und die Uncertainty Reduction Theory einen Beitrag zur Erklärung der Antwortqualität leisten.

Der Artikel ist wie folgt aufgebaut. In Abschnitt 2 stellen wir die Community Q&A Site vor. Die Hypothesen für das zu testende Modell werden in Abschnitt 3 entwickelt. Im vierten Abschnitt präsentieren wir den analysierten Datensatz. In Abschnitt 5 stellen wir die Ergebnisse vor und diskutieren diese im sechsten Abschnitt. In Abschnitt 7 fassen wir die Ergebnisse zusammen und zeigen Perspektiven für zukünftige Forschungsarbeiten auf.

2. Community Q&A Site

Die in diesem Beitrag untersuchte, kostenlose Community Q&A Site² ermöglicht es ihren Nutzern Fragen, die innerhalb eines festgelegten Zeitraums beantwortet werden können, an die Community zu stellen.

Jeder Nutzer der Seite hat die Möglichkeit, sich kostenlos in der Community zu registrieren und so Zugriff auf verschiedene weitere Funktionalität zu bekommen. Für jeden registrierten Nutzer existiert ein persönlicher Bereich, in dem er unter anderem Übersichten über die selbst gestellten und beantworteten Fragen, über Fragen die ihn möglicherweise interessieren könnten und über seine persönliche Linksammlungen erhält.

Weiterhin nimmt jeder registrierte Benutzer automatisch an einem Community-internen Punktesystem teil. In diesem wird zwischen Status- und Bonuspunkten unterschieden. Die Statuspunkte entscheiden über den sogenannten Rang eines Nutzers. Jeder Nutzer startet bei Registrierung im Rang eines Einsteigers und hat die Möglichkeit über das Sammeln von Statuspunkten im Rang aufzusteigen. Diese können durch folgende Aktivitäten gesammelt werden:

² Auf Wunsch des Praxispartners wird der Name der Community in diesem Beitrag nicht genannt.

- Durch das Stellen von Fragen, die durch andere Nutzer beantwortet oder als hilfreich bewertet werden.
- Durch die Beantwortung von Fragen mit als hilfreich oder als top bewerteten Antworten (eine Bewertung als Top-Antwort stellt die bestmögliche Bewertung durch einen Fragesteller dar).
- Durch die Bewertung von Antworten auf eigene Fragen innerhalb einer Frist von 14 Tagen nach Ablauf der für die Beantwortung der Frage eingestellten Frist.
- Durch das Hinzufügen von Links.
- Durch das Einladen neuer Nutzer in die Community sowie durch Rangaufstiege dieser eingeladenen Nutzern.

Neben den Statuspunkten erhält jeder registrierte Nutzer zu Beginn seiner Mitgliedschaft 1.000 Bonuspunkte. Diese kann er als Belohnung für die Beantwortung für von ihm gestellte Fragen einsetzen. Beantworten mehrere Antwortgeber eine Frage mit einer, mindestens als hilfreich bewerteten, Antwort, werden die ausgesetzten Bonuspunkte zwischen allen Antwortgebern, deren Antwort als mindestens hilfreich bewertet wurde, aufgeteilt. Dabei wird Top-Antworten ein überproportionaler Anteil der eingesetzten Bonuspunkte zugeteilt. Gibt ein Fragesteller innerhalb von 14 Tagen keine Bewertung für die Antworten auf seine Frage ab, werden diese automatisch alle als hilfreich bewertet und die eingesetzten Bonuspunkte werden zu gleichen Teilen auf die Antwortgeber verteilt.

Für jeden registrierten Nutzer wird außerdem ein persönliches Profil freigeschaltet. In diesem Profil hat jeder Nutzer die Möglichkeit, ein Bild hochzuladen sowie weitere persönliche Informationen (z.B. Adresse, Geschlecht, Email, Geburtsdatum, Land) anzugeben. Diese Informationen können für alle Nutzer, nur für Freunde, oder für niemanden im Netzwerk sichtbar gemacht werden. Neben diesen freiwilligen Angaben werden im Profil die Anzahl der gestellten Fragen und gegebenen Antworten, der Rang eines Nutzers, seine Statuspunkte, das Anmeldedatum, Themen, in denen ein Nutzer aktiv ist, sowie die Anteile von gegebenen Top-Antworten, hilfreichen Antworten und nicht-hilfreichen Antworten für alle anderen Nutzer sichtbar angezeigt.

Es werden weiterhin Funktionalitäten zum Aufbau und zur Pflege eines persönlichen Netzwerks angeboten. Ein registrierter Nutzer hat die Möglichkeit, andere Personen in die Q&A-Community allgemein und/oder in sein persönliches Netzwerk innerhalb der Community einzuladen. Zudem hat ein registrierter Nutzer unter anderem Zugriff auf Fragen von Personen aus seinem persönlichen Netzwerk und deren verwaltete Links. Schließlich gibt es auch die Möglichkeit, private Nachrichten an andere registrierte Nutzer zu verschicken.

3. Hypothesen

Zur Beantwortung der Forschungsfrage stützen wir uns auf unterschiedliche theoretische Erklärungsansätze, die wir im Folgenden vorstellen. Basierend auf diesen Theorien entwickeln wir Hypothesen, die wir im weiteren Verlauf dieses Beitrags empirisch überprüfen.

3.1 Sozialkapital und Tauschtheorie

In Anlehnung an Bordieu [6], Coleman [9] und Flap und de Graaf [10] wird *Sozialkapital* als aktuell und in Zukunft aus sozialen

Beziehungen resultierende Ressourcen definiert, welche ein Akteur für eigene Zwecke nutzen kann. Allgemein beruht Sozialkapital dabei auf der Bereitschaft oder Verpflichtung verbundener Akteure einem Akteur eigene Ressourcen zur Verfügung zu stellen.

Für eine Bewertung von Sozialkapital bedarf es einer Betrachtung der individuellen Netzwerkbeschaffenheit eines Akteurs, der so genannten *Sozialstruktur*. Die Sozialstruktur beschreibt, wie ein Akteur innerhalb seines sozialen Netzwerks mit anderen Akteuren verbunden ist. Die sich aus der Sozialstruktur (potenziell) ergebenden Handlungsbegünstigungen werden als *soziale Ressourcen* bezeichnet. Diese Handlungsbegünstigungen beruhen zum einen darauf, dass ein Akteur befähigt wird auf Ressourcen, wie z. B. Kapital und komplexes Wissen anderer, zu ihm in Beziehung stehender Akteure zurückzugreifen. Zum anderen lassen sich aus den Beziehungen selbst Handlungsbegünstigungen, z. B. in Form von Information und Kontrolle über andere Akteure ableiten (vgl. [7], [8] und [11]).

Im Rahmen unserer Analyse fassen wir das gesamte soziale Netzwerk der Q&A-Community als Sozialstruktur auf und wollen untersuchen, ob aus dieser Sozialstruktur bei entsprechendem Verhalten des Akteurs soziale Ressourcen verfügbar gemacht werden können. Diese sozialen Ressourcen sind in unserem speziellen Anwendungsfall fachliche Informationen und Wissen, die einem Akteur in seinem privaten oder beruflichen Handlungskontext nützlich sein können.

Als theoretische Grundlage bedienen wir uns dabei der sozio-ökonomischen Tauschtheorie [9], denn die von uns genutzte Sozialkapitaldefinition impliziert ein tauschtheoretisches Verständnis, da sie explizit auf die Bereitschaft der Akteure abzielt, eigene Ressourcen zur Verfügung zu stellen.

Grundlage der Sozialkapitalbildung oder -realisation ist ein Ressourcentransfer vom Sozialkapitalgeber zum Sozialkapitalempfänger. Ein Sozialkapitalgeber wird eigene Ressourcen ggf. jedoch nur dann offerieren, wenn er dafür eine Gegenleistung erhält. Im Gegensatz zum Markttausch werden Ressourcen hier allerdings ohne Geld und ohne explizite Verträge getauscht. Die Gegenleistung ist im sozialen Tausch also in Form, Höhe und Zeitpunkt nicht spezifiziert und nicht einklagbar. [19]

In der Tauschtheorie werden Akteure als Nutzenmaximierer unterstellt, die rational ihre Eigeninteressen verfolgen. Der Beitrag einer Ressource zur Befriedigung der Bedürfnisse eines Akteurs determiniert dabei das Interesse des Akteurs an einer Ressource. Tauschprozesse führen in einem sozialen System zu einem Gleichgewichtszustand, in dem weitere Tauschhandlungen nicht mehr zu einer individuellen Verbesserung von Akteuren führen würden. Unter den Bedingungen des von Coleman [9] beschriebenen Basismodells entspricht die Ressourcenmacht eines Akteurs im Gleichgewicht seinem Sozialkapital. Ein Akteur kann in einem solchen System auf Ressourcen anderer Akteure in dem Maße zugreifen, in dem er selbst Ressourcen – hier also bspw. für andere Nutzer wertvolle Informationen und Wissen – kontrolliert. Tritt ein Akteur dabei in Vorleistung, dann kann er theoretisch Verpflichtungen in Höhe seiner Ressourcenmacht auf Seiten der anderen Akteure schaffen. [19]

Auf unsere Analyse übertragen leiten wir daraus folgende Hypothese ab:

H1) Je stärker sich ein Fragesteller in der Vergangenheit im Netzwerk engagiert hat, desto qualitativ hochwertigere Antworten erhält dieser Fragesteller auf seine Fragen.

Wir messen das vergangene Engagement eines Fragestellers anhand von acht Variablen. Da wir für jede Variable einen abnehmenden Grenznutzen unterstellen und gleichzeitig alle Variablen auch den Wert Null annehmen können, addieren wir zu jeder Variable eins hinzu und bilden im Anschluss den natürlichen Logarithmus des Ergebnisses.

- FREUNDE bezeichnet die Anzahl der registrierten Nutzer, mit denen ein Fragesteller zum Zeitpunkt der Erstellung seiner Frage befreundet ist.
- EINGELADEN bezeichnet die Anzahl der registrierten Nutzer, die ein Fragesteller zum Zeitpunkt der Erstellung seiner Frage in die Community eingeladen hat.
- ANTWORTEN bezeichnet die Anzahl der weder als top noch als hilfreich bewerteten Antworten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage gegeben hat.
- HILFREICH bezeichnet die Anzahl der als hilfreich bewerteten Antworten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage gegeben hat.
- TOP bezeichnet die Anzahl der als top bewerteten Antworten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage gegeben hat.
- VERSCHICKT bezeichnet die Anzahl der Nachrichten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage an andere registrierte Nutzer verschickt hat.
- ERHALTEN bezeichnet die Anzahl der Nachrichten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage von anderen registrierten Nutzern erhalten hat.
- LINKS bezeichnet die Anzahl der Links, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage erstellt hat.

Ähnliche Variablen zur Messung der früheren Aktivitäten des Fragestellers finden sich auch bei [26] sowie [22].

3.2 Uncertainty Reduction Theory

In direkter Face-to-Face Kommunikation führt das Vorhandensein von zusätzlichen, den Gesprächspartner betreffenden Informationen zu verstärkter positiver Aufmerksamkeit gegenüber diesem [27]. Diese Aussage stimmt mit den Vorhersagen der *Uncertainty Reduction Theory* von Berger und Calabrese [5] für erstmalig stattfindende Face-to-Face Kommunikation überein. Die Theorie besagt, dass zusätzliche Informationen über den Gesprächspartner Unbehagen verringern und die Vorhersagbarkeit sowie die Gewogenheit gegenüber dem Gesprächspartner erhöhen.

Tidwell und Walther [28] zeigen in ihrer Arbeit, dass die Theorie auch auf computer-medierte Kommunikation übertragen werden kann. Dieses Ergebnis wird durch die Arbeiten von Zheng et al. [30] und Walther et al. [29] bestätigt. In beiden Arbeiten finden sich Hinweise darauf, dass das Vorhandensein von Bildern der

Gesprächspartner einen positiven Einfluss auf das Vertrauen [30] bzw. die Gewogenheit und die soziale Attraktivität [29] der Gesprächspartner hat.

Interpretiert man die Formulierung und die Beantwortung von Fragen innerhalb der untersuchten Community Q&A Site als computer-medierte Kommunikation, bieten sich den registrierten Nutzern verschiedene Möglichkeiten Informationen über sich preiszugeben. Registrierte Nutzer haben zum einen die Möglichkeit, ein Bild in ihr Profil einzufügen, und können zum anderen verschiedene Informationen wie z.B. ihren Namen, ihren Wohnort oder ihr Geschlecht angeben (vgl. Abschnitt 2). Gemäß der *Uncertainty Reduction Theory* sollte sich die Preisgabe von zusätzlichen Informationen positiv auf die Gewogenheit gegenüber einem Nutzer auswirken. Wir nehmen außerdem an, dass sich eine erhöhte positive Gewogenheit gegenüber einem Nutzer positiv auf die Qualität der Antworten auf seine Fragen auswirkt. Daraus leiten wir die folgende Hypothese ab:

H2) Je mehr Informationen ein Fragesteller über sich preisgibt, desto qualitativ hochwertigere Antworten erhält dieser Fragesteller auf seine Fragen.

Wir messen die Informationen, die ein Fragesteller über sich preisgibt, anhand von sieben binären Variablen:

- BILD nimmt den Wert eins an, wenn ein registrierter Nutzer ein Bild in sein Profil eingefügt hat.
- GESCHLECHT nimmt den Wert eins an, wenn ein registrierter Nutzer sein Geschlecht angibt.
- NAME nimmt den Wert eins an, wenn ein registrierter Nutzer seinen Vornamen angibt.
- POSTLEITZAHL nimmt den Wert eins an, wenn ein registrierter Nutzer seine Postleitzahl angibt.
- LAND nimmt den Wert eins an, wenn ein registrierter Nutzer sein Herkunftsland angibt.
- GEBURTSDATUM nimmt den Wert eins an, wenn ein registrierter Nutzer sein Geburtsdatum angibt.
- BERUF nimmt den Wert eins an, wenn ein Nutzer seinen Beruf angibt.

3.3 Allgemeines Interesse an einer Frage

Neben den in den Abschnitten 3.1 und 3.2 präsentierten Faktoren unterstellen wir analog zu [26], [22] sowie [1], dass das allgemeine Interesse an einer Frage einen weiteren wichtigen Einflussfaktor auf die Antwortqualität einer Frage darstellt. Hypothese 3 lautet daher:

H3) Je höher das allgemeine Interesse an einer Frage ist, desto qualitativ hochwertigere Antworten erhält dieser Fragesteller auf seine Fragen.

Dieses Interesse kann über vier unterschiedliche Variable gemessen werden. Da wir, wie auch für Hypothese H1, jeweils einen abnehmenden Grenznutzen der betrachteten Variablen unterstellen und gleichzeitig alle Variablen auch den Wert Null annehmen können, addieren wir wiederum zu jeder Variable eins hinzu und bilden den natürlichen Logarithmus des Ergebnisses.

- BONUSPUNKTE bezeichnet die für die Beantwortung einer Frage ausgesetzten Bonuspunkte.
- HITS bezeichnet die Anzahl der Aufrufe eine Frage.

- F_HILFREICH bezeichnet die Anzahl der Markierungen einer Frage als hilfreich.
- F_ANTWORTEN bezeichnet die Antworten, die für eine Frage erstellt wurden.

4. Datensatz

Der zur Verfügung stehende Datensatz beinhaltet alle Aktivitäten die in der Community Q&A Site zwischen dem 09.12.2005 und dem 06.06.2008 durchgeführt wurden.

In Tabelle 1 sind die Fragen und Antworten aggregiert für den gesamten Datensatz sowie für die einzelnen Jahre dargestellt. Die Tabelle zeigt, dass der Datensatz insgesamt 1.033.542 Fragen und 2.996.446 Antworten enthält. Auf jede Frage wurde also im Durchschnitt knapp dreimal geantwortet.

Tabelle 1: Fragen und Antworten

Jahr	# Fragen	# Antworten	# Antworten / # Fragen
Gesamter Datensatz	1.033.542	2.996.446	2,9
2005	343	411	1,2
2006	73.714	243.992	3,3
2007	580.717	1.662.719	2,9
2008	378.767	1.089.324	2,9

Tabelle 2 zeigt die Entwicklung der Nutzerzahlen der untersuchten Community Q&A Site vom 09.12.2005 bis zum 06.06.2008. Von den ca. 460.000 identifizierbaren Nutzern haben sich ca. 20% in der Community registriert. Neben den registrierten Nutzern besteht die Community aus ungefähr 368.000 Gastnutzern sowie 70 Moderatoren. Für unsere Untersuchungen werden im Weiteren nur die durch registrierte Nutzer gestellten Fragen von Relevanz sein.

Da die Community erst am 12.01.2006 für die Öffentlichkeit zugänglich gemacht wurde und um mögliche Verzerrungen durch die relativ geringe Mitgliederanzahl im ersten Jahr der Community zu vermeiden, nehmen wir nur Fragen, die nach dem 01.01.2007 erstellt wurden in unsere Analyse auf.

Tabelle 2: Nutzertypen

Jahr	Nutzer	Davon reg. Nutzer	Davon Moderatoren	Davon Gäste
Gesamter Datensatz	461.321	93.291	70	367.960
2005	89	77	9	3
2006	36.551	12.901	20	23.630
2007	290.770	54.404	39	236.327
2008	133.911	25.909	2	108.000

Von diesen 959.484 Fragen (vgl. Tabelle 1), werden 590.972 Fragen entfernt, da die Antworten auf diese Fragen nicht durch den Fragesteller sondern durch das System bewertet wurden. Dies

kann aus den beiden folgenden Gründen geschehen: Entweder handelt es sich bei dem Fragesteller um einen nicht registrierten Nutzer oder der Fragesteller ist ein registrierter Nutzer, der die Antworten auf seine Frage nicht innerhalb von 14 Tagen bewertet hat.

Weitere 3.659 Fragen werden entfernt, da diese Fragen nicht durch regulär angemeldete Nutzer sondern durch Moderatoren der Community gestellt wurden. Bei diesen Fragen könnten weitere, nicht für unsere Studie relevante Effekte die Bewertung der Antworten durch den Fragesteller beeinflussen.

Da eine stichpunktartige Textanalyse von Fragen mit besonders vielen Antworten ergeben hat, dass es sich bei diesen sehr häufig um tägliche Begrüßungen, Glückwünsche zum Statuswechsel oder Scherzfragen handelt, haben wir außerdem alle Fragen mit mehr als fünf Antworten von der Analyse ausgenommen (insgesamt 69.209). Von den verbleibenden 295.644 Fragen werden abschließend 144 Fragen aufgrund von offensichtlichen Fehlern im Datensatz gelöscht.

Der finale Datensatz besteht somit aus insgesamt 295.500 Fragen mit insgesamt 654.030 Antworten. Diese werden in zwei Kategorien eingeteilt. 145.475 Fragen, bei denen der Fragesteller keine der Antworten als Top-Antwort bewertet hat, ordnen wir der ersten Kategorie zu, wohingegen 150.025 Fragen mit mindestens einer Top-Antwort der zweiten Kategorie zugeordnet werden.

Tabelle 3 zeigt die Mittelwerte für alle metrisch skalierten Variablen sowie die Anteile der jeweiligen Ausprägung für die nominal skalierten Variablen für beide Kategorien. Anhand von χ^2 -Tests für alle nominal skalierten Variablen sowie t-Tests für alle metrisch skalierten Variablen wird untersucht, ob sich die Mittelwerte bzw. die Anteile der jeweiligen Ausprägung zwischen den Kategorien signifikant unterscheiden. Obwohl es sich bei der untersuchten Datenbasis um die Grundgesamtheit aller im Betrachtungszeitraum erstellten Fragen handelt, berechnen wir die χ^2 - und t-Statistiken sowie die dazugehörigen p-Werte.³ Wir interpretieren diese jedoch analog zu Gefen und Carmel [12] als Goodness-of-Fit Indikatoren und nicht als Irrtumswahrscheinlichkeit dafür, dass die Mittelwerte voneinander verschieden sind. Die in Tabelle 3 präsentierten Ergebnisse dieser Tests zeigen, dass sich die Anteile der jeweiligen Ausprägungen beider Kategorien bzw. die Mittelwerte für alle betrachteten Variablen substantiell voneinander unterscheiden. Diese Ergebnisse geben einen ersten Hinweis darauf, dass die betrachteten Variablen die Qualität der Antworten auf eine Frage beeinflussen.

5. Ergebnisse der logistischen Regression

Die Hypothesen werden anhand einer binären, logistischen Regression empirisch überprüft.⁴ Als abhängige Variable wird ein binärer Indikator gewählt, welcher den Wert eins annimmt, sobald

³ Umfasst der untersuchte Datensatz die komplette Grundgesamtheit, entsprechen die geschätzten Regressionskoeffizienten genau den Parametern der Grundgesamtheit [15].

⁴ Wir wählen die logistische Regression, da bei dieser Methode explizit dichotome abhängige Variablen berücksichtigt werden [14], [18].

ein Fragesteller mindestens eine der Antworten auf seine Frage als Top-Antwort bewertet. Wir nehmen an, dass dies der beste

Indikator für eine vom Fragesteller als sehr gut empfundene

Tabelle 3: Deskriptive Statistiken

Variable	Top-Antwort erhalten		
	Nein	Ja	
Anzahl Beobachtungen	145.475	150.025	
<i>Panel A: Sozialkapital und Tauschtheorie</i>			t-Statistic (p-Wert)
FREUNDE	1,15	1,59	87,86 (0,000)
EINGELADEN	0,02	0,03	20,60 (0,000)
ANTWORTEN	1,99	2,45	76,67 (0,000)
HILFREICH	3,88	4,63	101,45 (0,000)
TOP	2,67	3,41	108,37 (0,000)
VERSCHICKT	0,81	1,20	73,13 (0,000)
ERHALTEN	0,46	0,66	43,97 (0,000)
LINKS	0,84	1,23	65,48 (0,000)
<i>Panel B: Uncertainty Reduction Theory</i>			χ^2 -Statistic (p-Wert)
BILD	67,20%	81,23%	7619,37 (0,000)
GESCHLECHT	29,55%	35,13%	1048,49 (0,000)
NAME	15,67%	19,52%	657,74 (0,000)
POSTLEITZAHL	5,04%	5,62%	50,25 (0,000)
LAND	17,76%	23,82%	1643,08 (0,000)
GEBURTSDATUM	15,68%	18,32%	366,17 (0,000)
BERUF	18,98%	22,46%	544,80 (0,000)
<i>Panel C: Allgemeines Interesse an einer Frage</i>			t-Statistic (p-Wert)
BONUSPUNKTE	4,57	4,98	28,96 (0,000)
HITS	3,86	4,29	115,12 (0,000)
F_HILFREICH	0,09	0,20	90,79 (0,000)
F_ANTWORTEN	0,88	1,21	189,90 (0,000)

Beantwortung seiner Frage ist. Die unabhängigen Variablen werden anhand des von Hosmer und Lemeshow [18] empfohlenen Grenzwertes von 0,25 für die in der vierten Spalte in Tabelle 3 präsentierten p-Werte ausgewählt. Da keiner der p-Werte größer als 0,25 ist, werden alle in Tabelle 3 aufgeführten Variablen in das logistische Regressionsmodell aufgenommen.

Tabelle 4 zeigt die Ergebnisse der logistischen Regression. In der Tabelle sind in der ersten Spalte die unabhängigen Variablen, in der zweiten Spalte die anhand unserer Hypothesen vermuteten Wirkzusammenhänge, in der dritten Spalte die Regressionskoeffizienten, in der vierten Spalte die sogenannten Effektkoeffizienten und in der fünften Spalte die p-Werte dargestellt. Analog zu Abschnitt 4 interpretieren wir diese wiederum als Goodness-of-Fit Indikatoren und nicht als Wahrscheinlichkeiten dafür, dass die Regressionskoeffizienten signifikant von Null verschieden sind. Der Tabelle kann man entnehmen, dass mit Ausnahme der Koeffizienten für die Variablen FREUNDE und EINGELADEN alle Koeffizienten einen guten Fit ausweisen. Weiterhin haben alle Koeffizienten,

deren Betrag größer als 0,1 ist, das durch die Theorie vorhergesagte Vorzeichen.⁵

Anders als bei einer klassischen OLS Regression ist eine lineare Interpretation oder ein Vergleich der Regressionskoeffizienten untereinander bei einer logistischen Regression nicht möglich. Anhand der Regressionskoeffizienten kann nur auf die Wirkungsrichtung eines Zusammenhangs geschlossen werden [3]. Eine Angabe der Stärke des Wirkzusammenhanges ist jedoch mit Hilfe der Effektkoeffizienten (vgl. Spalte 4) möglich. Steigt z.B. die Variable TOP um eins, so wächst das Chancenverhältnis (Odds), d.h. der Quotient aus der Wahrscheinlichkeit, mindestens eine Top-Antwort zu erhalten, und der Gegenwahrscheinlichkeit um den Faktor 1,219. Ist der Effektkoeffizient dagegen kleiner als eins so sinkt dieses Chancenverhältnis. Ein Effektkoeffizient von genau eins hat keine Auswirkung auf das Chancenverhältnis.

⁵ Alle Variablen deren Effektkoeffizienten um mehr als 0,1 von 1 verschieden sind werden in Tabelle 4 fett hervorgehoben.

Tabelle 4: Ergebnisse der logistischen Regression

Variable	Vermutete Wirkungsrichtung	Regressionskoeffizient B	Exp^B	p-Wert
<i>Panel A: Sozialkapital und Tauschtheorie</i>				
FREUNDE	+	-0,005	0,995	0,309
EINGELADEN	+	0,009	1,009	0,779
ANTWORTEN	+	-0,135	0,874	0,000
HILFREICH	+	-0,029	0,971	0,001
TOP	+	0,198	1,219	0,000
VERSCHICKT	+	0,047	1,048	0,000
ERHALTEN	+	-0,044	0,957	0,000
LINKS	+	0,006	1,006	0,049
<i>Panel B: Uncertainty Reduction Theory</i>				
BILD	+	0,395	1,484	0,000
GESCHLECHT	+	-0,030	0,971	0,001
NAME	+	0,065	1,067	0,000
POSTLEITZAHL	+	-0,073	0,929	0,000
LAND	+	0,207	1,230	0,000
GEBURTSDATUM	+	-0,064	0,938	0,000
BERUF	+	-0,032	0,968	0,000
<i>Panel C: Allgemeines Interesse an einer Frage</i>				
BONUSPUNKTE	+	0,016	1,016	0,000
HITS	+	0,100	1,106	0,000
F_HILFREICH	+	0,498	1,645	0,000
F_ANTWORTEN	+	1,173	3,231	0,000

Hypothese H1 wird durch den relativ hohen Effektkoeffizienten für die Variable TOP unterstützt. Das negative Vorzeichen sowie der relativ hohe Effektkoeffizient für die Variable ANTWORTEN sprechen jedoch gegen die in Abschnitt 3.1 aufgestellte Hypothese. Die weiteren dieser Hypothese zugeordneten Variablen haben nur einen vernachlässigbaren Einfluss auf die Wahrscheinlichkeit, dass mindestens eine Antwort auf eine Frage als Top-Antwort bewertet wird.

Ein ähnliches Bild ergibt sich für Hypothese H2. Hier haben nur die Effektkoeffizienten für die Variablen BILD und LAND einen substantiellen Einfluss auf die Wahrscheinlichkeit, dass mindestens eine Antwort auf eine Frage als Top-Antwort bewertet wird. Der Einfluss der weiteren dieser Hypothese zugeordneten Variablen ist auch hier vernachlässigbar.

Den größten Einfluss auf die Wahrscheinlichkeit der Beantwortung einer Frage mit mindestens einer Top-Antwort haben die Variablen, die unserer dritten Hypothese zugeordnet sind. Hier sind insbesondere die Effektkoeffizienten für die Variablen F_ANTWORTEN und F_HILFREICH hervorzuheben.

Die Modellgüte wird in einem ersten Schritt anhand eines Likelihood-Ratio-Tests beurteilt. Bei diesem Test wird die Anpassung des vollständig spezifizierten Modells mit einem Modell, welches nur eine Konstante enthält, verglichen. Als Ergebnis dieses Tests kann die Nullhypothese, dass alle

Regressionskoeffizienten im vollständig spezifizierten Modell gleich Null sind, abgelehnt werden. Somit verfügt das vollständig spezifizierte Modell über eine im Vergleich zum Nullmodell verbesserte Vorhersagekraft bezüglich des Auftretens mindestens einer Top-Antwort.

Mit Werten von 0,183 für das Nagelkerke- R^2 und 0,137 für das Cox & Snell R^2 liegen die Werte für die Pseudo- R^2 Statistiken leicht unter dem von der Literatur als akzeptabel empfohlenen Niveau [3].⁶ Dies lässt darauf schließen, dass neben den im Modell betrachteten noch weitere unabhängige Variable existieren, die einen Einfluss auf die abhängige Variable ausüben. Hieraus ergibt sich jedoch kein Widerspruch zum Ziel unserer Studie, den Einfluss der betrachteten unabhängigen Variablen auf die Wahrscheinlichkeit mindestens eine Top-Antwort zu erhalten, zu untersuchen. Weiterhin zeigt der Vergleich mit anderen Studien, dass auch Modelle mit weitaus geringeren Werten für die

⁶ Analog zum Bestimmtheitsmaß der klassischen, linearen Regression zeigen diese Werte, welcher Anteil der Variation der abhängigen Variablen durch das vollständig spezifizierte Modell erklärt wird. Anders als bei der linearen Regression wird jedoch nicht auf den Anteil der erklärten Varianz, sondern auf das Verhältnis der Likelihood-Werte zwischen dem vollständigen und dem Null-Modell zurückgegriffen.

Pseude- R^2 -Statistiken als ausreichend eingestuft wurden. (vgl. z.B. [12] oder [13]).

Die Beurteilung der Klassifikationsergebnisse bietet weitere Hinweise auf einen guten Modellfit. Der Vergleich der Klassifikationsergebnisse des vollständig spezifizierten Modells (63,4% der Beobachtungen werden richtig klassifiziert) mit einer reinen Zufallsklassifikation (50,8% der Beobachtungen werden richtig klassifiziert) zeigt, dass unser Modell einen substantiellen Beitrag zur korrekten Klassifikation der Beobachtungen leistet. Ein berechneter Wert von 21.224 für Press's Q unterstützt dieses Ergebnis.⁷ Wir verzichten auf die Durchführung eines Hosmer-Lemeshow-Tests, da dieser bei nicht perfekt fittenden Modellen sensitiv auf die Größe des Datensatzes reagiert [20].

Zusammengefasst deuten somit alle Maße auf eine akzeptable Güte des Modells hin.

6. Diskussion

Insgesamt können alle in Abschnitt 3 aufgestellten Hypothesen durch die empirische Analyse bestätigt werden, jedoch sind die Ergebnisse differenziert zu betrachten.

Tabelle 4 zeigt, dass nur für Hypothese H3 alle Regressionskoeffizienten einen guten Fit und gleichzeitig das erwartete Vorzeichen aufweisen. Die vergleichsweise hohen Effektkoeffizienten für die Variablen HITS, F_HILFREICH, und F_ANTWORTEN zeigen, dass der größte Teil der durch unser Modell erklärten Wahrscheinlichkeit, mindestens eine Top-Antwort zu erhalten, im Interesse an einer Frage begründet liegt. Dieses Ergebnis ist durchaus plausibel, da ein ausreichendes Interesse für eine Frage notwendige Bedingung dafür ist, überhaupt eine Antwort zu erhalten. Weiterhin signalisiert ein hohes Interesse an einer Frage eine tendenziell höhere Fragequalität, welche sich zusätzlich positiv auf die Qualität der Antworten auswirken könnte. Dagegen hat die Höhe der ausgesetzten Belohnung für die Beantwortung einer Frage im Gegensatz zu der Studie von Harper et al. [16] nahezu keinen Einfluss auf die Qualität der Antworten. Betrachtet man jedoch, dass es sich bei der untersuchten Belohnung bei Harper et al. [16] um reales Geld, in dieser Studie jedoch nur um quasi wertlose Bonuspunkte handelt, ist dieses Ergebnis weniger verwunderlich.

Die Hypothesen H1 und H2 müssen dagegen differenzierter betrachtet werden. Die Effektkoeffizienten der Hypothese H1 zugeordneten Variablen FREUNDE, EINGELADENE, VERSCHICKT, ERHALTEN und LINKS zeigen, dass nicht jede Form der Partizipation in der Community einen positiven Einfluss auf die Qualität der Beantwortung der eigenen Fragen hat. Dieses Bild wird durch den Effektkoeffizienten für die Variable ANTWORTEN noch bestätigt. Dieser zeigt, dass von anderen Nutzern eher negativ bewertete Antworten einen negativen Einfluss auf die Qualität der Antworten auf eine eigene Frage haben, wohingegen sich sehr gute Antworten auf Fragen Anderer positiv auf diese Qualität auswirken. Dieses Ergebnis kann dahingehend interpretiert werden, dass im Rahmen der untersuchten Community Q&A Site ein tatsächlicher Austausch von qualitativ hochwertigen Antworten stattfindet und dass es

nicht möglich ist, andere Formen der Partizipation gegen qualitativ hochwertige Antworten einzutauschen.

Ein ähnliches Bild ergibt sich für die Variablen die unserer zweiten Hypothese zugeordnet werden. Nur das Vorhandensein eines Bildes und die Angabe des Herkunftslandes des Fragestellers haben einen substantiellen Einfluss auf die Qualität der Beantwortung einer Frage. Ein möglicher Erklärungsansatz für den nahezu nicht vorhandenen Einfluss der anderen Variablen könnte der in jedem Fall fragliche, zusätzliche Informationsgehalt dieser Variablen sein. Hierbei ist weiterhin zu berücksichtigen, dass die bei der Registrierung angegebenen Daten nicht durch das Team der untersuchten Community Q&A Site verifiziert werden und somit insbesondere bei sensiblen, persönlichen Daten davon auszugehen ist, dass nicht immer die Wahrheit angegeben wird.

Zusammenfassend zeigen die Ergebnisse der logistischen Regression, dass ein Fragesteller, der eine qualitativ hochwertige Beantwortung seiner Frage wünscht, insbesondere auf eine qualitativ hochwertige und interessante Fragestellung Wert legen sollte. Durch Hinzufügen eines Nutzerbildes sowie der Angabe des Herkunftslandes ist es möglich, die Wahrscheinlichkeit einer qualitativ hochwertigen Beantwortung weiter zu steigern. Aktive Teilnahme in der Community wirkt sich hingegen nur bei sehr hoher Qualität von vergangenen Antworten des Fragestellers positiv auf diese Wahrscheinlichkeit aus.

Vor dem Hintergrund des vergleichsweise hohen Zeitaufwandes für die qualitativ hochwertige Teilnahme an einer solchen Community Q&A Site kann die Frage aus der Überschrift dieses Beitrags daher tendenziell mit Nein beantwortet werden. Allerdings zeigen die Ergebnisse von Krasnova et al. [21], dass neben einer reinen Nutzenüberlegung viele weitere Faktoren auf die Partizipation in Online Communities wirken. Diese könnten einen substantiellen Beitrag zur Erklärung der vielen aktiven Nutzer auf Community Q&A Sites liefern.

7. Zusammenfassung und Ausblick

“A major challenge for IS research lies in making models and theories that were developed in other academic disciplines usable in IS research and practice” [4]. In diesem Beitrag nutzen wir die Uncertainty Reduction Theory aus der Psychologie sowie die sozio-ökonomische Tauschtheorie, um Einflussfaktoren zu identifizieren, die in Q&A-Communities zu besonders wertvollen Antworten auf gestellte Fragen führen. Die Analyse wird erstmals mit einem umfangreichen Datensatz einer deutschsprachigen Community Q&A Site durchgeführt.

Die Ergebnisse dieser Studie sind sowohl aus praktischer als auch aus wissenschaftlicher Sicht von Interesse. Bspw. können Community-Betreiber das Community-Design und die Teilnahme-Incentivierungssysteme dergestalt anpassen, dass vermehrt mit hilfreichen Antworten von Nutzern gerechnet werden kann. Dies kann zu einer Steigerung der Attraktivität und des Wertes einer Community insgesamt führen. Ebenso lassen sich für den individuellen Nutzer Verhaltensempfehlungen ableiten, welche bei Befolgung die Chancen auf wertvolle Antworten erhöhen. Für die Wissenschaft finden sich Hinweise darauf, dass das Verhalten der Nutzer teilweise mit der Tauschtheorie und der Uncertainty Reduction Theory erklärt werden kann.

⁷ Übersteigt Press's Q den kritischen Wert von 3,84, sind die Klassifikationsergebnisse signifikant ($\alpha = 0,05$) von einer zufälligen Zuordnung verschieden [3].

Es gibt noch zahlreiche Richtungen, diese Analyse zu verfeinern und zu erweitern. Zum ersten ist eine automatische Textanalyse der Fragen und Antworten geplant, um treffgenauer tägliche Begrüßungen, Glückwünsche zum Statuswechsel und Scherzfragen aus dem analysierten Datensatz zu entfernen. Hinweise zur Umsetzung finden sich z.B. bei [17]. Zum zweiten ist der Vergleich mit einer alternativen Modellspezifikation von Interesse. Als abhängige Variable wird statt der Bewertung von Antworten durch den Frager selbst die Bewertung von Antworten durch andere Nutzer verwendet. Zum dritten ist eine Separierung von Effekten des persönlichen Netzwerks von den Effekten der gesamten Community vorgesehen.

8. Literatur

- [1] Agichtein, E., Castillo, C., Donato, D., Gionis, A. und Mishne, G. 2008. Finding High-Quality Content in Social Media. In *Proceedings of the 1st ACM Intl. Conf. in Web Search and Data Mining (WSDM)* (Palo Alto, California, February 11-12, 2008).
- [2] Alexa, The Web Information Company. 2010. Top Sites > Reference. Zuletzt abgerufen: August 20, 2010, <http://www.alexa.com/topsites/category/Reference>.
- [3] Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. 2008. *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 12. Aufl., Springer, Berlin.
- [4] Benaroch, M., Kauffman, R. 1999. A Case for Using Real Options Pricing Analysis to Evaluate Information Technology Project Investments. *Information Systems Research*, 10, 1, 70-86.
- [5] Berger, C. und Calabrese, R. 1975. Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication. *Human Communication Research*, 1, 2, 99-112.
- [6] Bourdieu, P. 1983. Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In *Soziale Ungleichheiten, Soziale Welt*, R. Krekel, Ed. Schwartz, Göttingen, 183-198.
- [7] Brass, D. 1995. A Social Network Perspective on Human Resources Management. In *Research in Personnel and Human Resources Management* Ferris, G. Ed. JAI Press, Greenwich, CT, 39-79.
- [8] Coleman, J. 1988. Social Capital and the Creation of Human Capital. *American Journal of Sociology*, 94, 95-120.
- [9] Coleman, J. 1990. *Foundations of Social Theory*. Harvard University Press. Cambridge, MA.
- [10] Flap, H. und de Graaf, ND. 1986. Social Capital and Attained Occupational Status. *The Netherlands Journal of Sociology*, 22, 145-161.
- [11] Gargiulo, M., und Bernassi, M. 1999. The dark side of social capital. In *Corporate social capital and liability* R. Leenders and S. Gabbay Eds. Kluwer, Boston, 298-322.
- [12] Gefen, D. und Carmel, E. 2008. Is the World Really Flat? A Look at Offshoring at an Online Programming Marketplace. *MIS Quarterly*, 32, 2, 367-384.
- [13] Haas, A. 2009. Kann zu viel Kundenorientierung nachteilig sein? Eine Analyse der Wirkung der Kundenorientierung von Verkäufern auf die Kaufentscheidung. *Zeitschrift für Betriebswirtschaft*, 79, 7-29.
- [14] Hair, J., Anderson, R., Tatham, R. und Black, W. 1992. *Multivariate Data Analysis with Readings*. 4th Ed., Prentice Hall, Englewood Cliffs, NJ.
- [15] Hand, D. 1999. Statistics and Data Mining: Intersecting Disciplines. *ACM SIGKDD Explorations*, 1,1, 112-118.
- [16] Harper, M., Raban, D., Rafaeli, S. und Konstan, J. 2008. Predictors of Answer Quality in Online Q&A Sites. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 05-10, 2008). ACM, New York, NY.
- [17] Harper, M., Moy, D. und Konstan, J. 2009. Facts or Friends? Distinguishing Informational and Conversational Questions in Social Q&A Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (Boston, USA, April 4-9, 2009).
- [18] Hosmer, D. und Lemeshow, S. 1989. *Applied Logistic Regression*. Wiley, New York.
- [19] Iseke, A. 2009. Humankapital und Sozialkapital: Gibt es einen Matthäus-Effekt bezüglich der Sozialkapitalbildung von Nachwuchsführungskräften? *Zeitschrift für betriebswirtschaftliche Forschung*, 61, 3, 186-211.
- [20] Kramer, A. und Zimmermann, J. 2007. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Critical Care Medicine*, 35, 9, 2052-2056.
- [21] Krasnova, H., Hildebrand, T., Günther, O., Kovrigin, A. und Nowobilska, A. Why Participate in an Online Social Network: An Empirical Analysis. In *Proceedings of the 16th European Conference on Information Systems (ECIS)* (Galway, Ireland, June 09-11, 2008).
- [22] Liu, Y., Bian, J., Agichtein, E. 2008. Predicting Information Satisfaction in Community Question Answering. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008)* (Singapore, July 20-24, 2008).
- [23] Pomerantz, J., Nicholson, S., Belanger, Y. und Lankis, D. 2004. The Current State of Digital Reference: Validation of a General Digital Reference Model through a Survey of Digital Reference Services. *Information Processing and Management*, 40, 2, 347-363.
- [24] Rosenbaum, H. und Shachaf, P. 2010. A Structuration Approach to Online Communities of Practice: The Case of Q&A Communities. *Journal of the American Society for Information Science and Technology*, im Erscheinen.
- [25] Shah, C., Oh, S. und Oh, J. 2009. Research Agenda for Social Q&A. *Library and Information Science Research*, 11, 4, 205-209.
- [26] Shah, C. und Pomerantz, J. 2010. Evaluating and Predicting Answer Quality in Community QA. In *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010)* (Geneva, Switzerland, July 19 – 23, 2010).

- [27] Storck, J. und Sproull, L. 1995. Through a Glass Darkly What do People Learn in Videoconferences? *Human Communication Research*, 22, 2, 197-219.
- [28] Tidwell, L. und Walther, J. 2002. Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations. *Human Communication Research*, 28, 3, 317-348.
- [29] Walther, J., Slovacek, C. and Tidwell, L. 2001. Is a Picture Worth a Thousand Words? Photographic Images in Long Team and Short Team Virtual Teams. *Communication Research*, 28, 1, 105-134.
- [30] Zheng, J., Veinott, E., Bos, N., Olson, J. and Olson, G. 2002. Trust without Touch: Jumpstarting Long-Distance Trust with Initial Social Activities. In *Proceedings of the SIGCHI Conference in Computing Systems* (Minneapolis, USA, April, 20-25, 2002) ACM, New York, NY.